Standardization of Yeast Growth Curves from Several Curves with Different Initial Sizes

MD. ASADUZZAMAN

Standardization of Yeast Growth Curves from Several Curves with Different Initial Sizes
MD. ASADUZZMAMAN

Division of Mathematical Statistics
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
SE - 412 96 Göteborg
Sweden
Telephone: +46 (0)31 - 772 10 00

To the memory of my Father

Abstract

The main aim of this thesis is to provide a precise method by which the growth of the yeast *Saccharomyces cerevisiae* can be standardized from a specified initial optical density (OD). To apply the proposed method of standardization and also to compare with the previous methods, we use two datasets: one consists of 99 wild-type yeast strains growth data, observed for 24 hours and the other consists of 200 wild-type yeast strains, observed for 44 hours, cultured in normal environment.

We propose a method to standardize the growth of yeast with respect to an initial OD, fitting two observed growth curves at the same time by the modified Chapman-Richards growth model. We check the fitting and standardization of growth by residual plot and calculating the coefficient of determination. To verify the efficiency of the proposed method, we standardize growth curves at the mean initial OD values for each dataset by the proposed method and also by the previous methods, and compare the estimates of the physiologically growth parameters: the lag phase, the maximum relative growth rate and stationary phase OD increment.

The method works well for close observed growth curves. It shows better result for lag phase using distant observed curves but slightly worse results for the other two growth parameters. Weighing the observations has a better effect in this case and it is found to be insensitive with respect to the point of standardization. Finally, a generalization of the method is also discussed.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Yeasts are unicellular fungi. *Saccharomyces cerevisiae* is well-known and commercially significant among the yeasts. This organism has long been utilized to ferment the sugars of rice, wheat, barley, and corn to produce alcoholic beverages and in the baking industry to expand, or raise, dough. *Saccharomyces cerevisiae* is commonly known as baker's yeast and it has also been used as the most ideal eukaryotic microorganism for biological studies in many aspects since a long time ago (Sherman, 1997). The genetics of yeast has become an essential part of those who work with higher eukaryotes. It has also become extremely useful as a reference towards the sequences of human and other higher eukaryotic genes (Sherman, 1997). In fact, *Saccharomyces cerevisiae*, was the first eukaryotic genome that was fully sequenced, annotated, and made publicly available (Goffeau, 1997). The completed yeast genome sequence made it possible towards the development of many novel tools for analyzing all molecular components of the cell and their interactions (Scherens and Goffeau, 2000). It now serves as a model organism for understanding and interpreting eukaryotic cell function like human DNA sequences (Botstein et al., 1997).

Study of the consequences of the gene deletion mutants can be carried out to understand the cellular process performed by various genes. One aspect to study this consequences is the study of the genomewide phenotypic characterization, referred to as phenomics. It is very crucial to detect different physiological features like time to adapt environmental challenges and kinetics and efficiency of growth to determine the cellular role of a gene (Warringer et al., 2003). In recent

years, genomewide large-scale phenotypic characterization of yeast deletion mutants has received a lot of attention. Some quantitative phenotypic analyses have been performed to a large set of yeast strains (Warringer et al., 2003). But to provide consistent and more easily interpreted quantitative phenotypic analysis still requires a lot of works to be done. Quantifying the growth of yeast strains could, therefore, be an attempt towards facilitating the phenotypic profiling of the yeast *Saccharomyces cerevisiae*.

1.1 Growth of yeast

The cell division of yeast occurs by budding in which a daughter is initiated as an out growth from the mother cell, followed by nuclear division, cell-wall formation, and finally cell separation. The sizes of haploid[1] and diploid[2] cells vary with the phase of growth and from strain to strain (Sherman, 1997). Yeast cell grows in three main phases– the lag phase, the exponential phase and the stationary phase. When a culture of yeast cells is inoculated in a fresh growth medium, they enter a brief lag phase where they are biochemically active but not dividing. The lag phase refers the initial growth phase, when number of cells remains relatively constant prior to rapid growth, also referred as adaptation time. During this phase the individual cells are actively metabolizing, in preparation for cell division. The cells usually activate the metabolic pathways to make enough of the essential nutrients to begin active growth. From literature it is seen that the duration and extent of this phase depends on firstly the initial population size and secondly environmental conditions like temperature, pH, alcohol, oxygen, salt concentration, nutrients etc. Once the cell starts actively metabolizing, they begin DNA replication and shortly after the cells divide. This begins the second phase of growth called the exponential phase of growth. This is the period in which the cells grow most rapidly. The time it takes the culture to double is called generation time. This exponential phase depends on several factors: the organism itself, the growth medium, and the temperature are all important fac-

---

[1]A single set of chromosomes (half the full set of genetic material), present in the egg and sperm cells of animals and in the egg and pollen cells of plants.

[2]A full set of genetic material, consisting of paired chromosomes one chromosome from each parental set. Most animal cells except the gametes have a diploid set of chromosomes.

tors in determining the generation time. The third phase in growth of yeast is stationary phase when metabolism slows and the cells stop rapid cell division. The factors that cause cells to enter stationary phase are related to change in the environment typically caused by high cell density. Figure 1.1 illustrates how the yeast cell grows in three main phases.
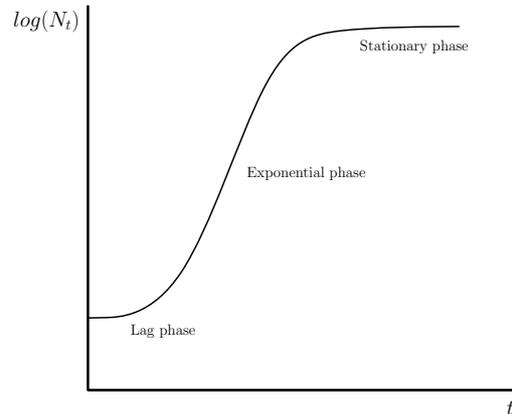
Figure 1.1: A typical logarithmic yeast growth curve where N_t is the population size at time t.

1.2   Mathematical modelling of yeast growth

Mathematical tools in biological system have become important in assisting scientists to understand and explain a lot of biological phenomena like the cellular process, the whole metabolism of organisms, the evolution of species and so forth. Concerning the growth of yeast cells, many physiological models have been proposed in the literature. However, mathematically it is not easy to model the behavior of the growth of yeast considering important growth parameters.

Richards (1928) performed the mathematical analysis of the growth of the yeast *Saccharomyces cerevisiae* a long time ago. He devised that the growth of yeast *Saccharomyces cerevisiae* has as asymmetrical S-shape. The growth reached a maximum yield within 100 hours in the synthetically defined medium in his study. He also measured the effect of temperature by van't Hoff Arrhenius equation. Gani and Saunders (1977) proposed branching process to model the

growth of yeast colonies. They tested their theoretical results on three types of yeast cells. As an approximation, their model reasonably descriptive of the biological phenomena involved in the budding of yeast cells. Szewczyk (1989) demonstrated a model for baker's yeast growth which gives a flexible framework for description of the yeast cultivation under oxidative utilization of glucose, ethanol coassimilation under oxidative conditions, a change from oxidative to oxidoreductive metabolism above a critical growth rate. Parameter estimation method of the model and kinetic expressions both for the saturated respiratory capacity and the repression of respiration were also discussed.

The phenotypic characterization is also becoming popular to the scientists now a days though genomic, proteomic studies of yeast have been conducted by a number researchers during last decade. Warringer et al. (2003) presented gene functional prediction by studying the growth behavior of yeast after the salt sensitive genes are being deleted. They investigated the consequences of loss of individual genes under saline stress. A high correlation to protein-protein data and no correlation between gene dispensability and gene expression were reported. Medawar et al. (2003) proposed a simple non-structured model to assess the lag-phase before yeast growth in alcoholic media. Ethanol is the main yeast growth inhibitor acting in wine. Different cultures were run with varying initial ethanol content and the influence on yeast growth were shown.

Warringer et al. (2003) presented and evaluated a method based on microscale liquid cultivation of some commonly used yeast strains to provide consistent and more easily explained quantitative phenotypic information, which leads to a large-scale automated phenotypic profiling. The concept of phenotypic indices at growth rate and at stationary phase OD increment were introduced and compared for the used strains. Fernandez-Ricaud et al. (2005) announced a publicly available source -PROPHECY- PROfiling of PHEnotypic Characteristics in Yeast– database accessible at http://prophecy.lundberg.gu.se to mine, filter and visualize the phenotypic data which is continuously updated. This is the first online resource to observe the genome-wide growth behavior of mutant strains in the yeast deletion collection during condition of environmental changes. PROPHECY allows us to see the physiologically relevant growth parameters: the time to adopt to the environmental changes - the lag time, the maximum growth rate - the exponential phase and stationary phase cell density - the efficiency of

growth.

Pylvänäinen (2005) compared the fits of the modified logistic, Gompertz and Chapman-Richards models using nonlinear least square method for yeast growth on the datasets extracted from PROPHECY database where altogether 576 *Saccharomyces cerevisiae* mutant strains were cultured in eight different environments. The physiologically important growth parameters - the lag time, the maximum relative growth rate and the efficiency of growth, were estimated. The modified Chapman-Richards model was found to be the best. One of her finding showed that the lag time and the growth rate strongly depends on the initial population size. Methods for standardizing growth curve were also derived with varying initial ODs. The main idea behind standardization was to predict and generalize the growth behavior having a standard initial OD. The methods to construct a summary curve from different standardized growth curves and filtering methods for atypical growth curves were also demonstrated.

Ericson et al. (2006) investigated genetic pleiotropy[3] in *Saccharomyces cerevisiae* by using high-resolution phenotypic profiling method to quantify the fitness contribution of genes on the five smallest yeast chromosomes under six distinct environmental stress. A higher extent of pleiotropy in yeast genes were found and the non-existence of phenotypic islands are reported. Hietala et al. (2006) used a five-parameter polynomial ratio (PR) model to characterize the growth, from lag through stationary phase, of the yeast *Saccharomyces cerevisiae* in response to cadmium toxicity. The PR model used in his study had more advantage over standard mathematical models in the ability to represent the initial cell mortality observed when *Saccharomyces cerevisiae* is exposed to increasing cadmium levels, up to 12 mg/l Cd, as well as following cell recovery and growth to stationary levels.

1.3 Objectives

This study is an attempt to fit and predict simultaneously the growth behavior of the yeast *Saccharomyces cerevisiae* having the population a fixed standard initial size. The specific intentions include:

---

[3]Pleiotropy describes the genetic effect of a single gene on multiple phenotypic traits.

- To fit yeast growth data and standardize it simultaneously using Chapman-Richards model from a standard initial OD for two set of observations, one with low and one with comparatively high initial OD.

- To compare the proposed method with the existing methods of standardization.

- To derive a method for generalization of fitting and standardization for wild-type yeast strains which will facilitate to phenotypic profiling.

This thesis is organized as follows. In chapter 2, the datasets, preprocessing of the data as blank correction, calibration, smoothing are discussed. The Chapman-Richards growth model, introduction of the three part model and the fitting are discussed in chapter 3. The proposed method of standardization and necessary formulae are derived in chapter 4. In chapter 5, the results from the analysis are demonstrated. In chapter 6, discussion from the analysis, some conclusions and the scope of further research are presented.

Chapter 2

Experimental Data

Yeast strains are inoculated in 350 μl synthetically defined (SD) medium[1] and incubated for 24 to 48 hours in a Bioscreen Analyzer with high intensity shaking 60 seconds every other minute. Optical density (OD) is measured every 20 minutes using a wide band (450-580 nm) filter. Each plate contained wild-type strains in randomized (once) positions. Strains are analyzed in duplicate (inhibitory conditions). All data measured by Bioscreen Analyzer are smoothened so that each OD value lower than previous value (i.e. the OD value at the previous time point) is set the previous value. The simplest biological reason for this, could be due to air bubbles. The OD values are then further blank corrected and calibrated to get the actual measurement of cell densities.

2.1 Optical density

Optical density is easy, very quick and widely used method to estimate the number of cell in a culture. It is defined as the logarithm to the base 10 of the ratio of the incident light to the transmitted light. OD is defined as

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

---

[1]The SD medium contains yeast nitrogen base (YNB) without amino acids, ammonium sulphate and succinic acid, buffered to pH 5.8; glucose; histidine, methionine, uracil, lysine and leucine.

where I_0 is the intensity of the incident light and I is the intensity of the transmitted light. Optical density can be measured by a spectrophotometer, a device that passes light through a cell suspension and detects the amount of unscattered light that goes through. Optical density is proportional to the number of cells (in somewhat extent).

2.2 Bioscreen C Analyzer

Bioscreen C is an highly developed instrument which has been designed to create up to 200 growth curves from 200 separate samples. Designed for the automation of microbiology, as micro-organisms grow, they increase the suspension turbidity of their growth medium. The organism of interest and any growth medium may be selected. Specimens are manually dispensed using an Electronic or Digital specially modified multichannel pipette into the wells of a special 10 x 10 microplates in two sets. Each well has a capacity of 400 μl. An optical density (OD) curve is measured by calculating the turbidity of this medium over time. This curve reflects the growth, increase of the organism. Bioscreen C system provides the platform for these measurements to be made.

2.3 Blank correction

A blank represents a measurement of OD in a Bioscreen Analyzer where the OD value measurement of wells contains only sterile water. A blank equal to 0.067 is subtracted from all data used in this thesis. This blank is the average blank of all wells in all five Bioscreens, calculated after running several hundreds of wells with only sterile water. There are differences betweens Bioscreens; the lowest Bioscreens average was 0.063 and the highest being 0.072. The reason of using same blank in all Bioscreens is that in practice it is not possible to measure Bioscreens and well specific blanks for each run. For more details, see Ericson (2004).

2.4 Calibration

While measuring yeast growth curve by optical density, a technical problem of non-linear relationship between OD values measured and number of cell at higher densities, is encountered. The yeast cultures are supposed to be diluted at higher OD values, but it is not possible with current high throughput set up. A calibration curve function is then introduced to transform the non-linear relation to a linear, so that the calibrated OD values will be proportional to to the number of cells. To derive the calibrated curve, pairwise OD measurements of diluted and undiluted samples are taken in all five Bioscreens. The well and Bioscreen specific blanks are subtracted from each of the measured OD values and then multiplied by the dilution factor. In order to reduce the sensitivity of the OD measurements, the well specific averages over all Bioscreen analyzers are taken so that there are a set of OD values of the diluted and undiluted samples. For more details, see Pylvänäinen (2005).

A regression equation is fitted using the least square method with x, the well specific average of the blank corrected undiluted OD as independent and y, the well specific average of the blank corrected diluted OD multiplied by the dilution factor as dependent variable. It is assumed that there is almost a proportional relation between the blank corrected diluted OD values and the blank corrected undiluted OD values approximately up to 0.3. A cubic function is then fitted as

$$y = x + cx^3$$

where c is a constant. Least square method is used to estimate c and the fitted curve becomes

$$y = x + 0.8324057x^3 \tag{2.1}$$

A second degree polynomial term is avoided as it could make the curve too steep so that the y values would not be too high if we extrapolate for higher values of x.

2.5 Datasets

Two datasets have been used for the analysis in this thesis where all of the yeast strains were wild type.

2.5.1 Dataset I

In the first dataset wild type *S. cerevisiae* strains are cultured in synthetically defined (SD) medium. Synthetically defined (SD) medium contains yeast nitrogen base (YNB), ammonium sulphate and succinic acid, glucose; histidine, methionine, uracil, lysine and leucine. All the strains are incubated twice; first for 72 hours, then for 24 hours (in fresh medium) in a Bioscreen Analyzer C, optical density is measured every 20 minute. Each well plate contained wild-type strain in randomized (once) positions. Well 4 does not include cells, it was used just to measure absorbance of well with media. Therefore, the dataset includes 99 wild-type *S. cerevisiae* strains growth data by OD measurement at 73 time points.

2.5.2 Dataset II

Like dataset I, dataset II also contains the growth data points of wild-type strains of the yeast *S. cerevisiae*. Here the strains are also cultured in synthetically defined (SD) medium and incubated exactly as we have done for dataset I. But here the strains are incubated for 44 hours in the Bioscreen Analyzer and optical density is measured every 20 minute. In order to get the robust measurement of OD values, strains with different initial cell density are replicated in the wells of the Bioscreen. Therefore, the dataset contains optical density measurements at 132 time points of 200 wild-type yeast strains with replicated and variable initial cell density.

Chapter 3

Growth Model

Many models have been proposed so far to describe the growth pattern in wide range of disciplines such as crop science, fishery research, forest science and biology. Some of the commonly used growth models are logistic, Gompertz, Richards, Stannard, Schnute etc. Not all the models give adequate fit of the growth behavior of yeast. Also the parameters used in the growth models should have a clear biological meaning. Pylvänäinen (2005) compared most commonly used growth models: logistic, Gompertz and Champman-Richards in yeast growth. She fitted and standardized the growth models for a wide range of start ODs. The Champman-Richards model was found to be the best to describe the growth pattern of the mutant and wild type yeast strains cultured in eight different environmental stresses. Therefore, we propose to fit and standardize the yeast growth by using Champman-Richards model in this thesis. In this chapter we state the Champman-Richards model, its reparameterization, derivation of the growth parameters, three part model and assessing the fit of a model.

3.1 Chapman-Richards model and its reparameterization

Pienaar and Turnbull (1973) described the Chapman-Richards generalization of von Bertalanffy's growth model which contain parameters in the model having at least a gross physiological or biological interpretation. All the models consider

the relative population size $log(N_t/N_0)$, where N_0 is the initial population size and N_t is the population size at time t. The Champman-Richards model (Pienaar and Turnbull, 1973) is

$$v_t = log\left(\frac{N_t}{N_0}\right) = \beta_0\left[1 - \beta_1 e^{-\beta_2 t}\right]^{\frac{1}{(1-\beta_3)}} \tag{3.1}$$

While modelling of the bacterial growth curve Zwietering et al. (1990) noticed the problem of lack of biological parameters in the models and suggested the modification of the logistic, Gompertz, Richards, Schnute and Stannard growth curve functions. The models are reparameterized in such a way that they contain parameters that are microbiologically relevant. The reparameterized Chapman-Richards model is

$$v_t = log\left(\frac{N_t}{N_0}\right) = A_z\left[1 - (1-\beta_3)e^{\frac{\beta_3^{\frac{\beta_3}{\beta_3-1}}\mu}{A_z}(\lambda_z-t)+\beta_3}\right]^{\frac{1}{1-\beta_3}} \tag{3.2}$$

where $A_z = \beta_0$ is the asymptote, the maximum value of the growth (on logarithmic scale), $\mu = \beta_0\beta_2\beta_3^{\frac{\beta_3}{1-\beta_3}}$ is the slope of the tangent of the logarithmic growth curve at the inflection point[1], the maximum relative population growth rate and $\lambda_z = \frac{log\left(\frac{\beta_1}{1-\beta_3}\right)-\beta_3}{\beta_2}$ is the time axis intercept of the tangent at the point of inflection on the logarithmic growth curve, the lag time. The Chapman-Richards model is also known as Richards model and it is very flexible as it can be fitted to both exponential and sigmoidal[2] growth patterns.

Since the number of cells $v_t > 0$ at time $t = 0$ for all t, the models were reparamiterized by introducing a new parameter $D < 0$ which gave the following form

$$g_t = log(N_t) = y_t + D \tag{3.3}$$

where $D = log(N_0) - y_0$. Then the Chapman-Richards reparameterized model

---

[1]An inflection point is a point on a curve at which the sign of the curvature (i.e., the concavity) changes.

[2]A curve which has first a convex shape and then a concave shape, i.e. has the shape like the letter S.

becomes

$$g_t = \beta_0 \Big[1 - \beta_1 e^{-\beta_2 t}\Big]^{\frac{1}{(1-\beta_3)}} + D \qquad (3.4)$$

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

The physiologically important growth parameters– the lag time λ, the maximum growth rate μ and the stationary phase OD increment Y of the model were also derived and estimated by Pylvänäinen (2005). For the analysis purpose of this thesis here we reproduce the growth parameters from Pylvänäinen (2005).

The stationary phase OD increment: The stationary phase OD increment is the final OD minus the initial OD. Mathematically

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

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 time point. The inflection time point t_I can be derived by calculating the second derivative of the function (3.4) with respect to t, equating at zero and solving with respect to t. The first and second derivatives are

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

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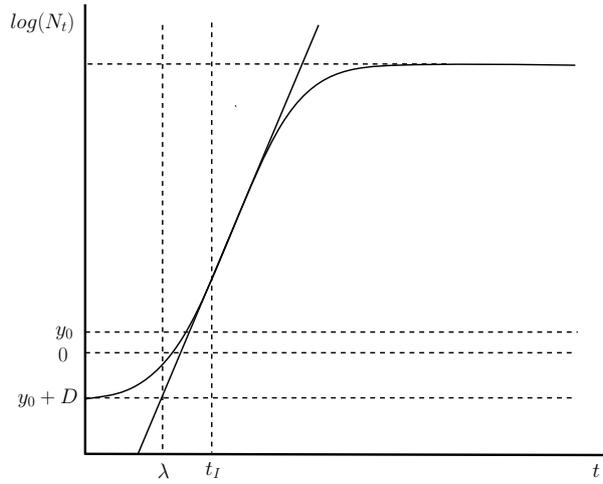\end{aligned}
$$

Figure 3.1: Illustration of the growth parameter calculation in the Chapman-Richards model where N_t is the population size at time t, t_I is the inflection time point, y_0 is the value of the Chapman-Richards function (3.4) at time $t = 0$, $D = log(N_0) - y_0$, and λ is the lag time.

Equating these at zero gives the solution

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

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

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

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

$$y_0 + D = \mu\lambda + \beta_0\beta_3^{\frac{1}{1-\beta_3}} - \mu t_I + D \tag{3.5}$$

Solving (3.5) with respect to λ yields

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

3.3   The three part model

While fitting by Chapman-Richards model, the fit in the very beginning and the transition from the exponential phase to stationary phase was not very good in Pylvänäinen (2005). Therefore, a three part model was suggested to reduce the error in fit. The proposed model consists of three parts: the beginning of the curve until inflection point, the linear part following the inflection point which is modelled as straight line and the rest after the linear part. Mathematically

$$
g_t^* = \begin{cases}
g_t; & t \leq t_I \\
g_{t_I} + \mu(t - t_I); & t_I \leq t \leq t_I + \Delta \\
g_{t-\Delta} + \mu\Delta; & t \geq t_I + \Delta
\end{cases} \tag{3.6}
$$

where Δ is the time span of the linear part ($\Delta \geq 0$) and g_t is the Chapman-Richards function given by equation (3.4). The three part model can be illustrated by Figure 3.2.

3.4   Fitting the model and assessing the fit of the model

The data are blank corrected, smoothened and calibrated before fitting the model as described in chapter 2. Non-linear regression approach are used to fit the
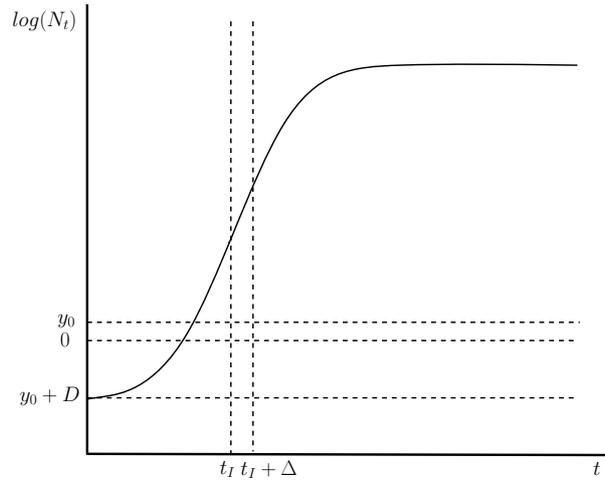
Figure 3.2: Illustration of the three part model where N_t is the population size at time t, t_I is the inflection time point, y_0 is the value of the Chapman-Richards function (3.4) at time $t = 0$, and $D = log(N_0) - y_0$.

three part model to the data by `lsqnonlin` (non-linear least square) function in `Matlab`. To make good comparison with uniformity in all respect and also for simplicity, we keep initial values same for all runs to estimate the parameters and to standardize simultaneously: $\beta_0 = 4.5$, $\beta_1 = -50$, $\beta_2 = 0.3$, $\beta_3 = 3$, $D = -3$.

To assess fits of the model, we plot the residuals and calculate the coefficient of determination. The coefficient of determination can be written as

$$
\begin{aligned}
r^2 &= 1 - \frac{SSE}{SST} \\
&= 1 - \frac{\sum_{t_p=1}^{n} (g_{t_p}^* - x_{t_p})^2}{\sum_{t_p=1}^{n} (x_{t_p} - \bar{x})^2}
\end{aligned}
\tag{3.7}
$$

where $g_{t_p}^*$ is the fitted curve at time point t_p, x_p is the observed OD value at time point t_p, $\bar{x}$ is mean of observed OD values over n time points and n is the total number of time points.

Chapter 4

Standardization Method

The purpose of standardization is to predict the growth behavior of yeast. In large scale experiments it is very hard to keep the initial cell size constant or to start with a prefixed cell size. Therefore, a method is essentially required to predict what would happen if we start with a fixed OD. In the proposed method of standardization we assume that the idealized model of a logarithmic growth curve consisting of a lag phase in the beginning, then an exponential phase and a stationary phase in the end, is true. Our approach allows to start with a simultaneous fitting of the three part model with Chapman-Richards function for two sets of growth data– one started with a low initial OD and the other with a comparatively higher initial OD. Then we are to predict the growth of the yeast starting from an initial OD in the middle between the initial ODs of two observed growth curves. The main idea behind the proposed approach of having data from two growth curves is to extract different pieces of information from two growth curves. We believe that data from the curve with higher initial OD i.e. from the upper observed curve, have greater influence to the beginning and lower influence to the end. And the data from the lower observed curve are less informative in the beginning and provides higher information at the end when we standardize a curve from an initial OD. Compare to Pylvänäinen (2005) methods (for more details, see Appendix) we may then expect a better overall precision incorporating those ideas through weighing the observations. We expect the proposed approach will also reduce the correlation between initial OD and growth rate and will also reduce the sensitivity of the growth.

4.1 Standardizing two curves simultaneously

The standardization method predicts the growth starting from a fixed initial OD, taking information from curves with one higher initial OD and one lower initial OD. First we use the three part model with Chapman-Richards function without a linear part in the middle to fit the observed curve with higher initial OD using non-linear least square and then the three part model is fitted to the lower observed curve also by non-linear least square. Finally we derive the standardized curve starting from an initial OD in the middle between the initial ODs of the observed curves, keeping the growth parameters– the lag time λ, the maximum relative growth rate μ and stationary phase OD increment Y, constant and removing the linear piece from the middle and lifting the curve up from the lower observed curve as illustrated in Figure 4.1. Therefore, the standardized curve is the Chapman-Richards curve. The growth parameters are also kept the same for the fitted curves and the standardized curve.
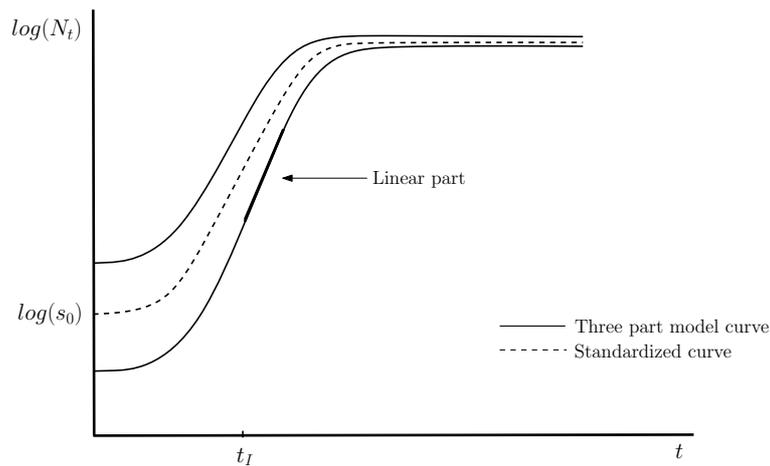

Figure 4.1: Illustration of standardizing curve at s_0 where N_t is the population size at time t, t_I is the inflection time point.

The model of the upper observed curve, is written as

$$
g_t^* = \begin{cases} g_t; & t \leq t_L \\ g_{t+\Delta_{up}} - (g_{t_U} - g_{t_L}); & t \geq t_L \end{cases} \tag{4.1}
$$

where

$$g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{\frac{1}{(1-\beta_3)}} + D_{up} \tag{4.2}$$

is the Chapman-Richards function, and where the derivatives of the curve at t_L and $t_L + \Delta_{up}$ have to be the same. See the Appendix for further details.

The three part model for the lower observed curve is given by

$$g_t^* = \begin{cases} g_t; & t \leq t_I \\ g_{t_I} + \mu(t - t_I); & t_I \leq t \leq t_I + \Delta_{down} \\ g_{t-\Delta_{down}} + \mu \Delta_{down}; & t \geq t_I + \Delta_{down} \end{cases} \tag{4.3}$$

where

$$g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{\frac{1}{(1-\beta_3)}} + D_{down} \tag{4.4}$$

and

$$t_I = \frac{log\left(\frac{\beta_1}{1-\beta_3}\right)}{\beta_2} \tag{4.5}$$

is the inflection time point and

$$\mu \;=\; \beta_0 \beta_2 \beta_3^{\frac{\beta_3}{1-\beta_3}} \tag{4.6}$$

is the maximum relative growth rate and Δ_{down} is the time span of the linear part. We first fit the two observed curves simultaneously using fixed Δ_{up} and Δ_{down}. Consequently the free parameters are β_0, β_1, β_2, β_3, D_{up} and D_{down}.

To also fit Δ_{up} and Δ_{down}, we proceed as follows:

- Δ_{up} is fitted using *'golden section search'* (Press et al., 1992), by minimizing the fitting errors for both observed curves using non-linear least square method, keeping Δ_{down} fixed. During the implementation of the *'golden section search'*, we keep our search space for the linear part, Δ_{up}, 0 to 5 hours as our experience shows that the linear part never exceeds 5 hours.

- Δ_{down} is iteratively recalculated assuming that the stationary phase OD increment of the three part model curve equals the stationary phase OD in-

crement for the standardized curve (Pylvänäinen, 2005). See the Appendix for further details.

Then with the constraints that removing the linear piece from the middle and shifting the lower curve up by $\tau \geq 0$ and taking estimated parameter values from the fit of the two observed curves, we obtain the standardized curve which is denoted by $g_t^\bullet$.

$$g_t^\bullet = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{\frac{1}{(1-\beta_3)}} + D_{down} + \tau \qquad (4.7)$$

Let s_0 be the standard initial OD, the fixed value. Equating (4.7) at $t = 0$, we get $\beta_0[1 - \beta_1]^{\frac{1}{(1-\beta_3)}} + D_{down} + \tau$ which is equal to $log(s_0)$, i.e.

$$\beta_0[1 - \beta_1]^{\frac{1}{(1-\beta_3)}} + D_{down} + \tau = log(s_0) \qquad (4.8)$$

Solving (4.8) for τ gives

$$\tau = log(s_0) - \beta_0[1 - \beta_1]^{\frac{1}{(1-\beta_3)}} \qquad (4.9)$$

4.2 Model fitting procedure

First we fix a standard initial OD, denoted by s_0 somewhere between the initial ODs of the two observed curves. Then according to the Chapman-Richards model, the curve that has initial value s_0 can be obtained by the following procedure:

1. An initial value of Δ_{down} is chosen for the lower curve.

2. Perform the *'golden section search'* to find out the optimum Δ_{up} for the upper curve, minimizing the fitting errors for both observed curves simultaneously by using least square method.

3. The models (4.1) and (4.3) are fitted simultaneously for the upper and lower curves respectively using non-linear least square method using the fixed Δ_{down} for the lower curve and the optimum Δ_{up} obtained by *'golden section search'* for the upper curve.

4. The stationary phase OD increment Y and $Y^\bullet$, are calculated. If $|Y - Y^\bullet| > c$, then a new value of Δ_{down} is calculated as given by (A.1.6) in the Appendix. The constant c is the maximum allowed difference between Y and $Y^\bullet$.

Step 2 to step 4 are repeated until $|Y - Y^\bullet| < c$.
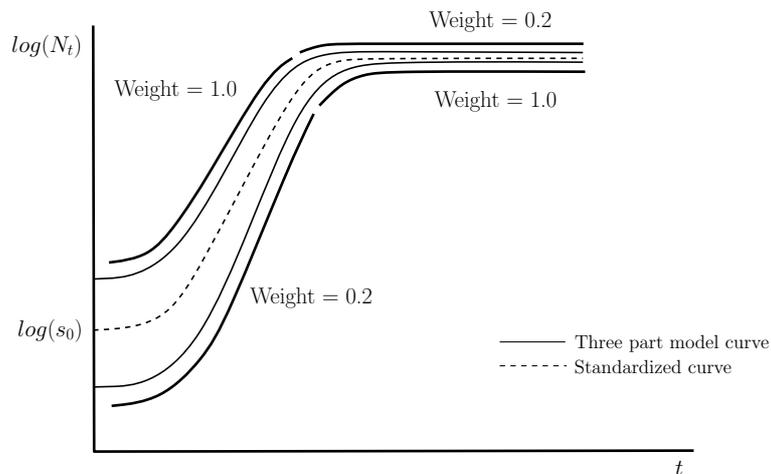
4.3   Weighted non-linear least square

Figure 4.2: Illustration of standardizing curve at s_0 when the weights are used to the observations. Here N_t is the population size at time t.

The reason behind weighing the observations is that we think the observations in the beginning of the upper curve have greater influence in the standardized curve than the observations of the beginning of the lower curve. Similarly the observations at the end of the lower curve, that is, during stationary phase, have greater influence than that of the upper curve. So, we give weight 1.0 for half of the observations in the beginning of the upper curve while the rest gets 0.2 and 0.2 for half of the observations in the beginning of the lower curve whereas 1.0 for the rest. The weighting is illustrated by Figure 4.2. Then we perform the non-linear least square to fit both observed curves and derive the standardize curve simultaneously.

Chapter 5

Results

The main purpose of this study is to predict the growth behavior of yeast strains from a specified initial OD. In the previous chapter, we propose a method by which the growth can be predicted with respect to an initial OD. The initial OD has an obvious influence towards the growth of yeast. If one starts with very small number of cell in the beginning, the lag phase will be longer (than that of higher cell densities). On the other hand the lag phase will be shorter if one starts with higher number of cells in the beginning. Consequently the population will go to the exponential phase very quickly. Considering all these constraints, the proposed approach is applied to the datasets I and II. As mentioned in chapter 2, the dataset I consist of OD values of 99 wild-type yeast strains observed for 24 hours. Since the follow-up for 24 hours is not enough for the cells to go to the stationary phase, we therefore collect another dataset; dataset II which has growth information for 200 wild-type yeast strains observed for 44 hours. Plot of well-wise initial OD values and the histogram of the initial OD values are presented in Figure 5.1. In dataset I, the deviations of the initial OD values are not very large; the lowest OD value is 0.0030 and the highest is 0.1060. Also the data collected for over 24 hours only which is not enough to have the stationary phase. Therefore we collect dataset II where initial OD values vary from 0.0270 to 0.3091 and the growth is observed for 44 hours. The minimum, maximum, mean and coefficient of variation (CV) of the initial OD values in the dataset I and II are presented in Table 5.1.

Table 5.1: The minimum, maximum, mean and coefficient of variation (CV) of the initial OD values in the datasets. The values are blank corrected and calibrated.

Dataset	Minimum	Maximum	Mean	CV
I	0.0030	0.1060	0.0596	47.14
II	0.0270	0.3091	0.1297	57.19

Pylvänäinen (2005) also proposed methods for standardization; one for standardizing up and the other for standardizing down. But her methods are quite sensitive since those show poor fit if the difference of the initial OD between the observed and the standardize curve are high. Therefore, to make a good comparison we proceed as follows: first we sort the observations in the datasets with respect to initial OD values and we divide them into three parts and we take the middle part to standardize by Pylvänäinen methods at the mean value (in logarithmic scale) $s_0 = 0.0513$ for dataset I and $s_0 = 0.1092$ for dataset II. The rest one-third with higher initial OD and one-third with lower initial OD paired randomly, are used to derive standardized curve at the same points by the proposed method. In this way we derive 33 standardized growth curves by Pylvänäinen methods and 33 standardized growth curves at $s_0 = 0.0513$ by proposed method for dataset I. Similarly we derive 66 standardized growth curves by Pylvänäinen methods and 67 standardized growth curves at $s_0 = 0.1092$ by proposed method for dataset II. We also use weights as described in section 4.3. For each of the standardized growth curves, we calculate the growth parameters: adaptation time λ, the maximum relative growth rate μ and stationary phase OD increment Y at the same time. The summary of the results are presented in Table 5.2 and Table 5.3.

Figures 5.2 to 5.6 are derived by the proposed method of standardization when the observed curves have higher and lower initial OD values and their corresponding residual plots are also shown in the right. The standardized curves are also derived by the proposed method when the two observed curves are very close. From the residual plot we see that the fit does not look good when we choose the two observed curves with distant initial ODs. A systematic error occurs in the beginning of all fittings. There could be some biological reasons for this. But the residuals do not vary too much if we standardize at different points
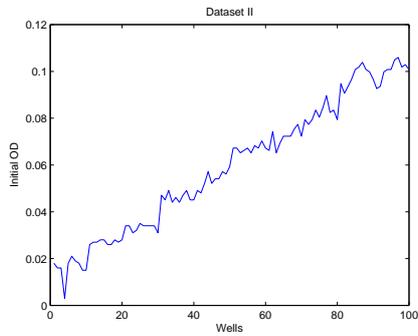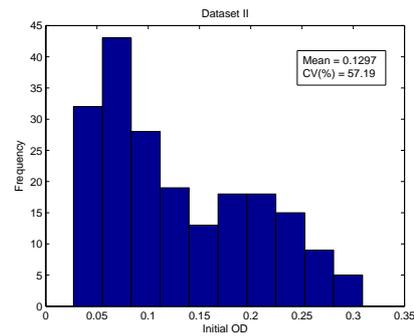
(a) Plot of well-wise initial OD values in dataset I. The values are blank corrected and calibrated.
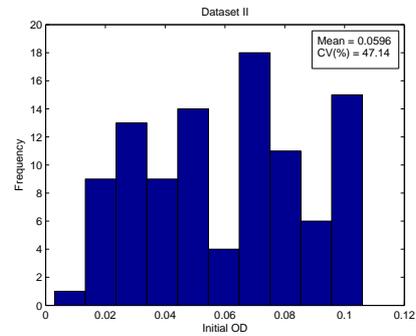
(b) Histogram of the initial OD values of all wild-type in dataset I. The values are blank corrected and calibrated.
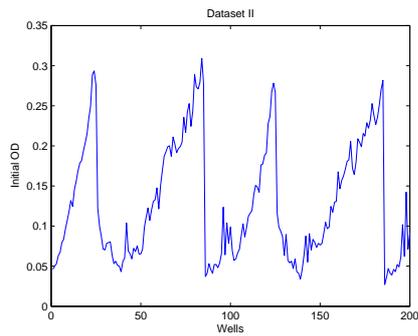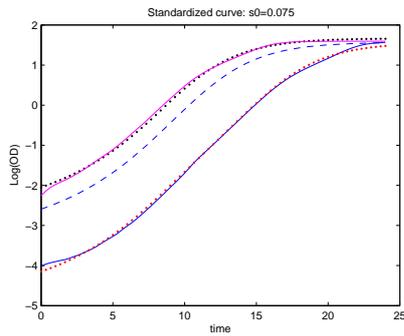

(c) Plot of well-wise initial OD values in dataset II. The values are blank corrected and calibrated.
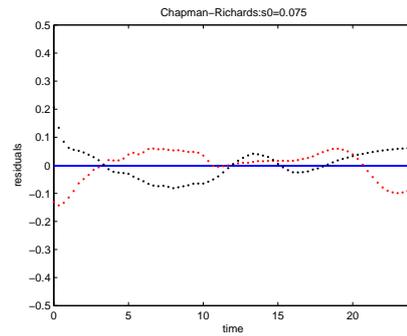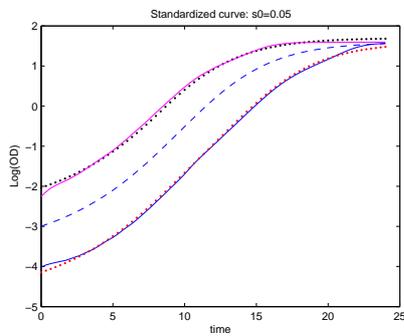

(d) Histogram of the initial OD values of all wild-type in dataset II. The values are blank corrected and calibrated.

Figure 5.1: Plot of well-wise initial OD values and histogram of initial OD values in dataset I and II.

(a) The r-square value for the upper curve is 0.9992 and for the lower curve is 0.9983.

(b) Plot of residuals.

(c) The r-square value for the upper curve is 0.9992 and for the lower curve is 0.9979.
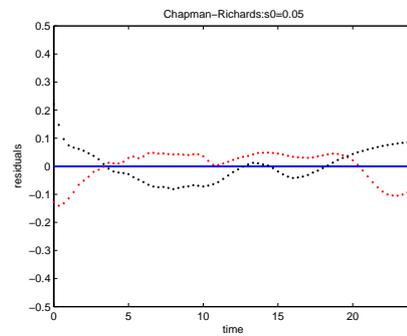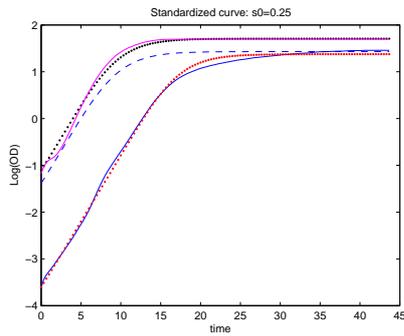
(d) Plot of residuals.
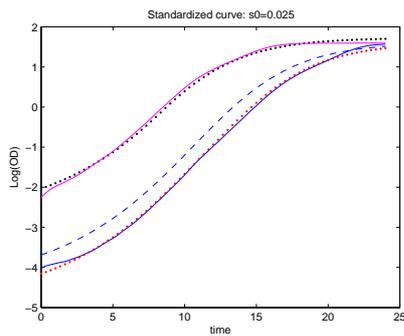
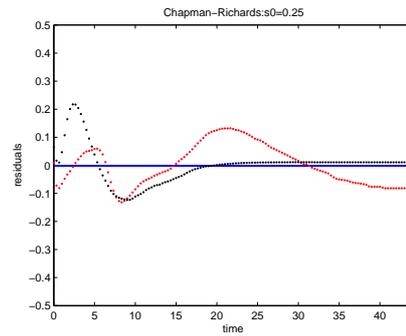(e) The r-square value for the upper curve is 0.9990 and for the lower curve is 0.9973.
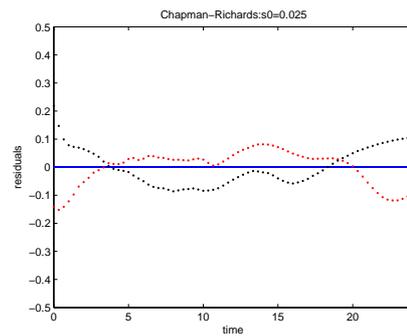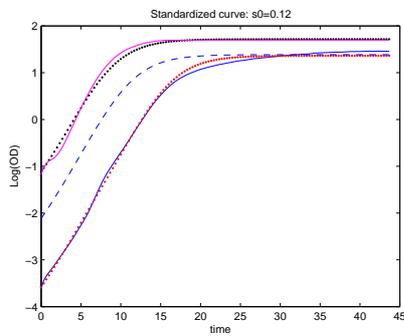
(f) Plot of residuals.

Figure 5.2: The three part model with Chapman-Richards function fitted to the dataset I, well 96 and well 8 (using weights). The observed curve $log(OD)$ (solid), the fitted growth curve (dotted) and the standardized curve (dashed). The corresponding residual plots of the fitted curves are on the right.
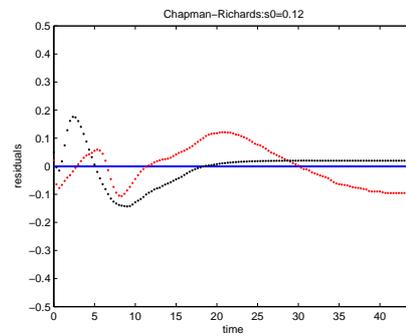
(a) The r-square value for the upper curve is 0.9977 and for the lower curve is 0.9927.
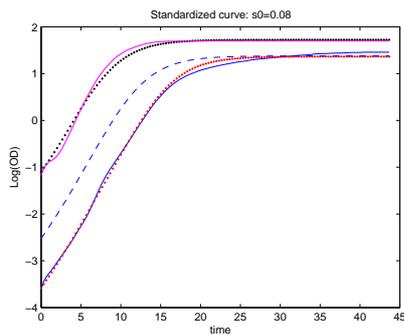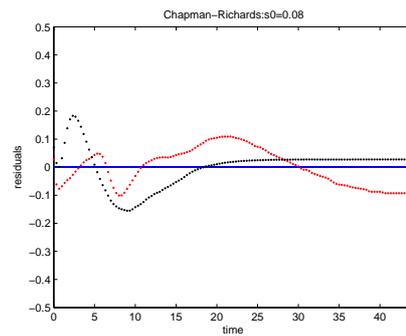
(b) Plot of residuals.

(c) The r-square value for the upper curve is 0.9978 and for the lower curve is 0.9930.
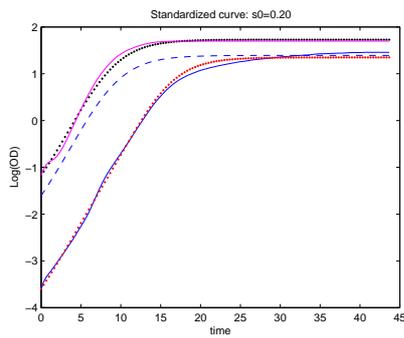
(d) Plot of residuals.

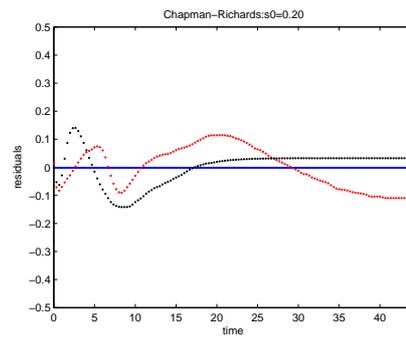(e) The r-square value for the upper curve is 0.9981 and the lower curve is 0.9916.
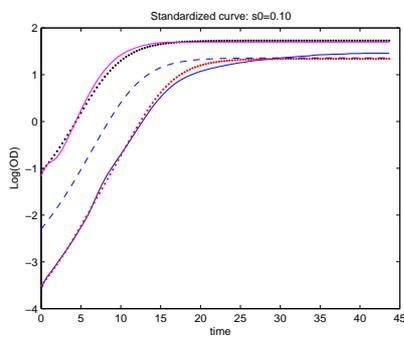
(f) Plot of residuals.

Figure 5.3: The three part model with Chapman-Richards function fitted to the dataset II, well 186 and well 84. The observed curve $log(OD)$ (solid), the fitted growth curve (dotted) and the standardized curve (dashed). The corresponding residual plots of the fitted curves are on the right.
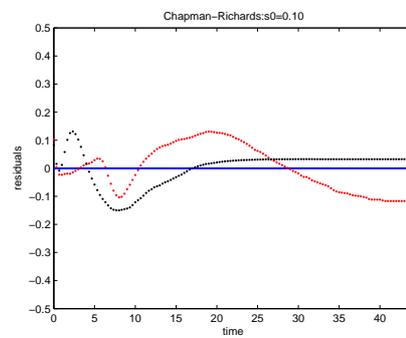
(a) The r-square value for the upper curve is 0.9968 and for the lower curve is 0.9938.

(b) Plot of residuals.

(c) The r-square value for the upper curve is 0.9971 and for the lower curve is 0.9933.

(d) Plot of residuals.

(e) The r-square value for the upper curve is 0.9978 and the lower curve is 0.9902.

(f) Plot of residuals.

Figure 5.4: The three part model with Chapman-Richards function fitted to the dataset II, well 186 and well 84 (using weights). The observed curve $log(OD)$ (solid), the fitted growth curve (dotted) and the standardized curve (dashed). The corresponding residual plots of the fitted curves are on the right.
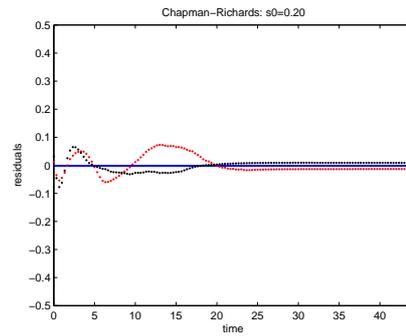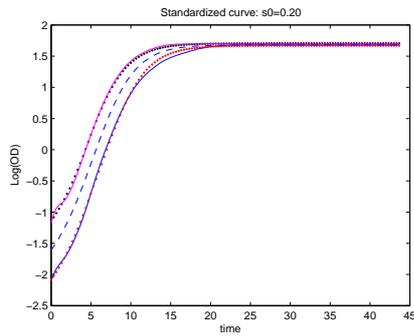
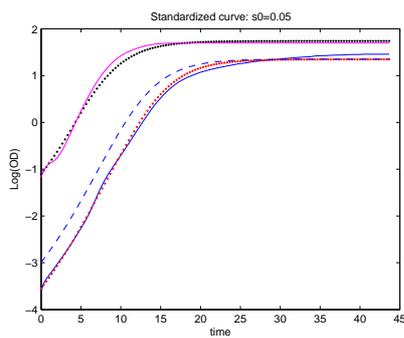(a) The r-square value for the upper curve (well 84) is 0.9991 and for the lower curve (well 26) is 0.9993.

(b) Plot of residuals.

(c) The r-square value for the upper curve (well 77) is 0.9994 and for the lower curve (well 106) is 0.9989.

(d) Plot of residuals.
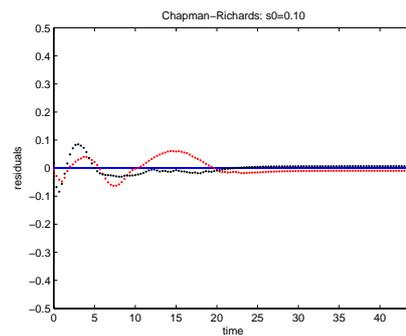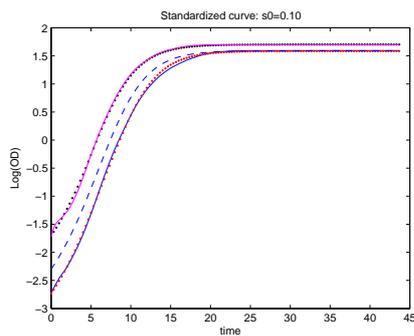
(e) The r-square value for the upper curve (well 16) is 0.9995 and the lower curve (well 95) is 0.9994.

(f) Plot of residuals.

Figure 5.5: The three part model with Chapman-Richards function fitted to the dataset II. The observed curve $log(OD)$ (solid), the fitted growth curve (dotted) and the standardized curve (dashed). The corresponding residual plots of the fitted curves are on the right.
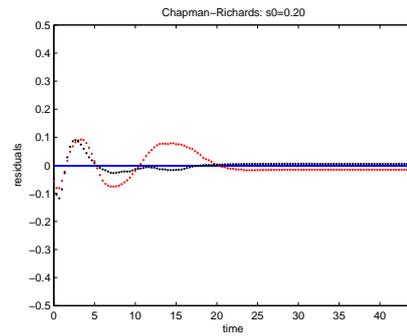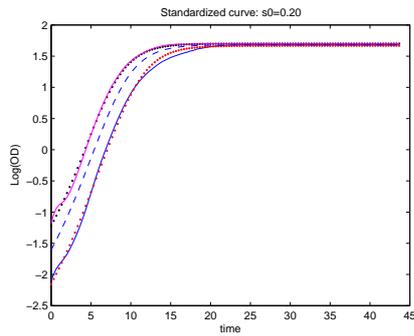
(a) The r-square value for the upper curve (well 84) is 0.9985 and for the lower curve (well 26) is 0.9988.
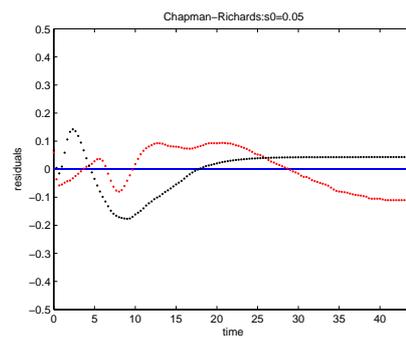
(b) Plot of residuals.

(c) The r-square value for the upper curve (well 77) is 0.9985 and for the lower curve (well 106) is 0.9981.

(d) Plot of residuals.

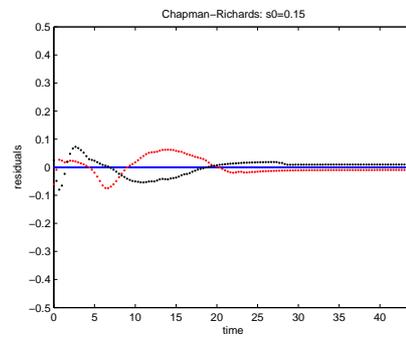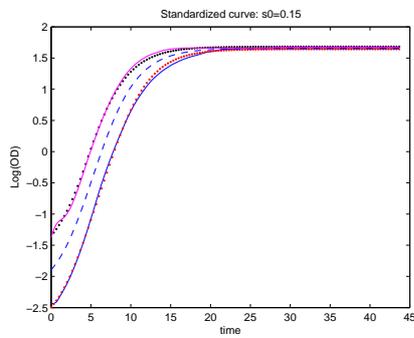(e) The r-square value for the upper curve (well 16) is 0.9993 and the lower curve (well 95) is 0.9993.

(f) Plot of residuals.

Figure 5.6: The three part model with Chapman-Richards function fitted to the dataset II using weights. The observed curve $log(OD)$ (solid), the fitted growth curve (dotted) and the standardized curve (dashed). The corresponding residual plots of the fitted curves are on the right.
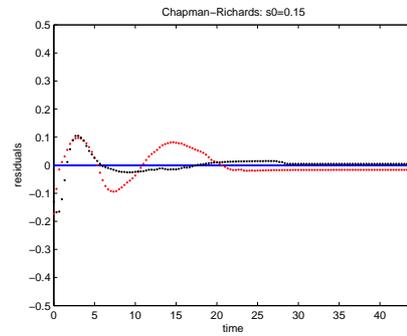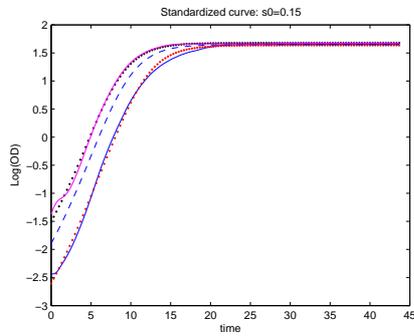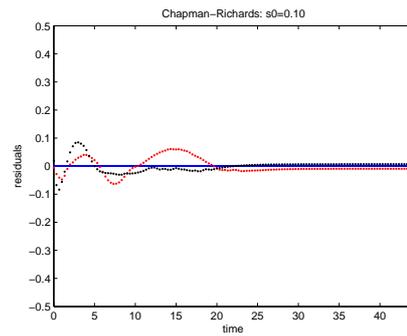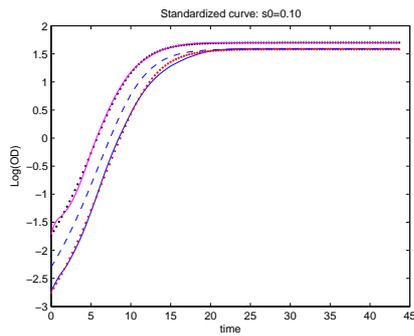
between the two observed curves. The fitting error becomes very small when the two observed curves are close and the weighting does not have substantial effect.

Table 5.2: The summary of results derived by the proposed method and Pylvänäinen methods from datasets I.

Parameter	Method	Mean	St. dev.	CV(%)
λ	Pylvänäinen method	2.1765	0.1466	6.74
	Proposed method (weighted)	2.3996	0.2373	9.89
μ	Pylvänäinen method	0.3561	0.0064	1.81
	Proposed method (weighted)	0.3533	0.0050	1.41
Y	Pylvänäinen method	5.3711	0.1802	3.35
	Proposed method (weighted)	5.2950	0.3355	6.10

Table 5.3: The summary of results derived by the proposed method and Pylvänäinen methods from datasets II.

Parameter	Method	Mean	St. dev.	CV(%)
λ	Pylvänäinen method	0.8645	0.2279	26.37
	Proposed method	0.7668	0.1581	20.62
	Proposed method (weighted)	0.7753	0.1351	17.42
μ	Pylvänäinen method	0.3634	0.0100	2.74
	Proposed method	0.3458	0.0115	3.32
	Proposed method (weighted)	0.3459	0.0110	3.17
Y	Pylvänäinen method	5.4976	0.1638	2.98
	Proposed method	4.6607	0.2806	6.02
	Proposed method (weighted)	4.6564	0.2902	6.23

From Table 5.2, we see that the estimate of the lag time λ by proposed method has slightly higher coefficient of variation than Pylvänäinen methods for dataset I. One reason for this is could be– we choose very close observed curves to standardize by Pylvänäinen methods and distant curves for the proposed method. But coefficient of variation of the relative maximum growth rate μ, is still lower for the proposed method than Pylvänäinen methods.

From Table 5.3, the lag time λ estimated by proposed method, has less coefficient of variation than estimated by Pylvänäinen methods, even when the observations are not weighted for dataset II. The coefficient of variation is much

less when the weighting is used. But the coefficient of variation is slightly higher for the proposed method for both cases when the weighting is used and also not used. The coefficient of variation of the stationary phase OD increment by Pylvänäinen methods is less for both datasets.

Chapter 6

Discussion and Conclusion

Quantitative phenomics is a fundamental approach for better understanding the cellular process, to determine the cellular role of genes through screening the quantitative changes of growth phenotypes in a wide variety of growth conditions for yeast. We mainly focus one problem related to the microbial growth– how to standardize the growth curves with respect to an initial OD, having observed the growth curves. We use the Chapman-Richards growth function as our basic tool.

The three part model is fitted for a wide range of wild-type yeast growth curves and the standardized growth curves are derived with a specified OD. The model works well when the observed curves are close. For the distant observed curves, some fitting errors arise. However, the fitting errors can be reduced to a certain extent by considering lower tolerance levels during each iteration while running `lsqnonlin` function in `Matlab` and also using very small termination criterion. Here it is to be mentioned that to reduce the running time of the `Matlab` codes, we consider the termination criterion `c` (the difference between the stationary phase OD increment of the three part model curve and standardize curve) equals `0.01` and keep `TolX=0.001` and `TolFun=0.001` for the parameter `options` in `lsqnonlin` function for all runs, while the default values are `TolX=1e-6` and `TolFun=1e-6`. We also report a systematic error in the very beginning of all observed curves. In most of the curves we see a weird behavior as– it suddenly goes up and then it comes down and follows the typical shape of the yeast growth curve for each of the observed curves. But we do not know why this happens; there might be some biological reasons. When we fit the upper curve, we perform

the *'golden section search'* to find out the optimum linear part for the upper curve for which the fitting error of the both observed curves is minimum. Also we adjust the optimum linear piece for the lower curve so that the stationary phase OD increment of the lower curve becomes equal to the stationary phase OD increment of the standardized curve.

Another important objective of this thesis, is to compare the efficiency of the proposed method of standardization with the existing methods, in particular, the Pylvänäinen methods. We compare our proposed method of standardization with the Pylvänäinen methods by deriving standardized curves and comparing the estimates of the growth parameters. The proposed method shows better results as it has lower coefficient of variation for the growth parameter: the relative maximum growth rate for dataset I and the lag time for dataset II. But for lag phase in dataset I and maximum relative growth rate for dataset II, Pylvänäinen methods show slightly better results. The stationary phase OD increment is better by Pylvänäinen methods for both datasets. Since Pylvänäinen methods are sensitive as those overestimate or underestimate the growth parameters for distant observed curves, we choose to use the proposed method for the distant observed curves and Pylvänäinen methods for the close observed curves for standardization. The result would possibly be more in favor to the proposed method if we choose the same observed curves for the proposed method.

The proposed method can easily be generalized to obtain a single standard curve from n pairs of observed curves where n observed curves have initial OD greater than the specified initial OD for standardization and the other n observed curves have initial OD less than the specified initial OD for standardization. The three part models are to be the fitted to the observed curves. The upper n curves can be fitted by choosing the optimum linear pieces by the *'golden section search'*. The linear parts for the lower n curves can be adjusted so that the stationary phase OD increment of the standardized curve becomes equal to the average of the stationary phase OD increment of the three part model curves, having initial OD less than the standard initial OD. We assume that the lag time and the growth rate are the same for all the three part model curves and the standardized curve.

6.1 Future work

Further study can be conducted to standardize more accurately and to get the robust estimate of growth parameters by a number of ways. Some of the techniques, could then be–

- Improving algorithms: We use higher tolerance during each iteration while fitting and standardizing growth curves to reduce the computational time. One can try to improve the algorithm to increase the computational speed and use lower tolerance to get more precise fitting of the observed curves. For instance, one can use high tolerance in the beginning of the fitting and lower tolerance when fine-tuning the fit. At the same time, different weights can also be used.

- Using more curves: One can use another observed curve in between the two other curves to integrate some more information as well. The add in information from the middle curve can strengthen the estimates of the growth parameters.

- Exploring the bulb shape: During the fitting of the observed curves, we always notice a bulb shape in the very beginning for which a systematic error always occurs in the fitting. We believe there might be some biological reasons, but we do not integrate this bulb shape in our models. It would be interesting to explore this further.

- Applying more advanced statistical methods: Some more advanced statistical tools like time series analysis using the state space models or exponential smoothing, Bayesian approach can also be applied to get more precise fitting and estimate of the growth parameters of yeast.

Bibliography

Botstein, D., Chervitz, S. A. and Cherry, J. M. (1997). Yeast as a model organism, *Science* **277**(5330): 1259–1260.

Ericson, E. (2004). *Large-scale phenotypic analysis of Saccharomyces cerevisiae deletion mutants*, Licentiate Thesis, Department of Cell and Molecular Biology, Microbiology, Gotebörg University.

Ericson, E., Pylväniänen, I., Fernandez-Ricaud, L., Nerman, O., Warringer, J. and Blomberg, A. (2006). Genetic pleiotropy in saccharomyces cerevisiae quantified by high-resolution phenotypic profiling, *Molecular Genetics and Genomics* **275**: 605–614.

Fernandez-Ricaud, L., Warringer, J., Ericson, E., Pylvänäinen, I., Kemp, G. J. L., Nerman, O. and Blomberg, A. (2005). PROPHECY– a database for high-resolution phenomics, *Nucleic Acids Research* **33**(Database issue): D369–D373.

Gani, J. and Saunders, I. W. (1977). Fitting a model to the growth of yeast colonies, *Biometrics* **33**: 113–120.

Goffeau, A. (1997). The yeast genome directory, *Nature* **387**(Suppl.): 5–6.

Hietala, K. A., Lynch, M. L., Allshouse, J. C., Johns, C. J. and Roane, T. M. (2006). A mathematical model of saccharomyces cerevisiae growth in response to cadmium toxicity, *Journal of Basic Microbiology* **46**: 196–202.

Medawar, W., Strehaiano, P. and Dèlia, M.-L. (2003). Yeast growth: lag phase modelling in alcoholic media, *Food Microbiology* **20**: 527–532.

37

Pienaar, L. V. and Turnbull, K. J. (1973). The chapman-richards generalization of von bertalanffy's growth model for basal area growth and yield in even - aged stands, *Forest Science* **19**(21): 2–22.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press.

Pylvänäinen, I. (2005). *A Parametric Approach to Yeast Growth Curve Estimation and Standardization*, Doctoral Thesis, Division of Mathematical Statistics, Department of Mathematical Sciences, Chalmers University of Technology.

Richards, O. W. (1928). The growth of the yeast *Saccharomyces cerevisiae:* the growth curve, its mathematical analysis and the effects of temparature on the yeast growth, *Annals of Botany* **os–42**(1): 271–283.

Scherens, B. and Goffeau, A. (2000). The uses of genome-wide yeast mutant collections, *Genome Biology* **229**: 1–8.

Sherman, F. (1997). Yeast genetics, *The Encyclopedia of Molecular Biology and Molecular Medicine* **6**: 302–325, (edited by R. A. Meyers).

Szewczyk, K. W. (1989). A model for backer's yeast growth, *Bioprocess Engineering* **4**: 261–264.

Warringer, J., Ericson, E., Fernandez, L., Nerman, O. and Blomberg, A. (2003). High-resolution yeast phenomics resolves different physiological features in the saline response, *PNAS* **100**(26): 15724–15729.

Zwietering, M. H., Jongenburger, Rombouts, F. M. and Riet, K. V. (1990). Modeling of the bacterial growth curve, *Applied And Environmental Microbiology* **56**(1): 1875–1881.

Appendix A

Pylvänäinen methods

A.1 Standardizing upwards

This method was developed to predict what would happen if the initial OD is fixed and larger than the observed initial OD. The three part model is used to fit the observed curve so that a standardize curve can be obtained by 'lifting' the fitted curve to start from $log(s_0)$ and removing the linear piece from the middle, Figure A.1. The growth parameters lag time, maximum relative growth rate and stationary phase OD increment, are to be the same for the three part model curve and the standardize curve.

The three part model given by

$$
g_t^* = \begin{cases} g_t; & t \leq t_I \\ g_{t_I} + \mu(t - t_I); & t_I \leq t \leq t_I + \Delta \\ g_{t-\Delta} + \mu\Delta; & t \geq t_I + \Delta \end{cases} \tag{A.1.1}
$$

where

$$
g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{\frac{1}{(1-\beta_3)}} + D \tag{A.1.2}
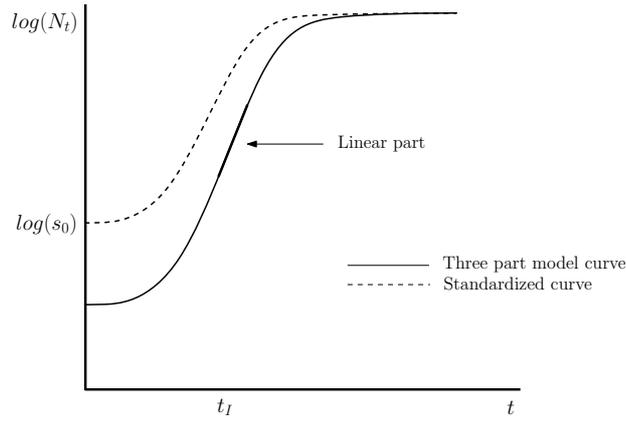$$

Figure A.1: Illustration of standardizing curve when the standard initial OD is higher than the observed initial OD. Here N_t is the population size at time t.

is the Chapman-Richards function.

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

is the inflection time point 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. First the three part model is fitted to the observed data points with the constraints that removing the linear piece from the middle and shifting the curve by $\tau \geq 0$. The standardized curve which is denoted by $g_t^{\bullet}$, is obtained:

$$g_t^{\bullet} = \beta_0\left[1 - \beta_1 e^{-\beta_2 t}\right]^{\frac{1}{(1-\beta_3)}} + D + \tau \tag{A.1.3}$$

Let s_0 be the standard initial OD, the fixed value. Equating (4.3) at $t = 0$, we get $\beta_0[1-\beta_1]^{\frac{1}{(1-\beta_3)}} + D + \tau$ which is equal to $log(s_0)$, i.e.

$$\beta_0[1-\beta_1]^{\frac{1}{(1-\beta_3)}} + D + \tau = log(s_0) \tag{A.1.4}$$

Solving (A.1.4) for τ gives

$$\tau = log(s_0) - \beta_0[1 - \beta_1]^{\frac{1}{(1-\beta_3)}} \tag{A.1.5}$$

Then the time span of the linear part Δ, is so adjusted 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 to 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

$$\Delta = \frac{-\beta_0 - D + log[e^{\beta_0 + D + \tau} - e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D + \tau} + e^{\beta_0(1-\beta_1)^{\frac{1}{1-\beta_3}} + D}]}{\mu} \tag{A.1.6}$$

A.2   Standardizing downwards

When the standard initial OD is less than the observed initial OD, then standard-izing downward method can be applied. A Chapman-Richard model curve that lacks a part in the middle is fitted to the observed data. The standardized curve is then the Chapman-Richard model curve, including the part in the middle. The stationary phase OD increment of the standardized curve is to be same as of the observed curve, but the lag time and the growth rate do not need to be the same.

The model of the observed curve is written as

$$g_t^* = \begin{cases} g_t; & t \leq t_L \\ g_{t+\Delta} - (g_{t_U} - g_{t_L}); & t \geq t_L \end{cases} \tag{A.2.7}$$

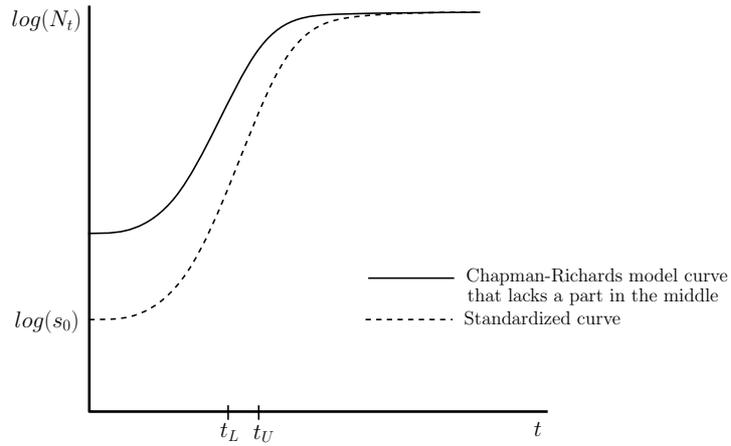The following figure shows the standardizing downwards by Pylvänäinen method

Figure A.2: Illustration of standardizing curve when the standard initial OD is lower than the observed initial OD. Here N_t is the population size at time t.

where

$$g_t = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{\frac{1}{(1-\beta_3)}} + D \tag{A.2.8}$$

is the Chapman-Richards function and

$$\Delta = t_U - t_L.$$

The model of the standardized curve can be written as

$$g_t^{\bullet} = \beta_0 \left[1 - \beta_1 e^{-\beta_2 t} \right]^{\frac{1}{(1-\beta_3)}} + D + \tau \tag{A.2.9}$$

where $\tau (\leq 0)$ is obtained by setting the initial $log(OD)$ value of the standardized curve to $log(s_0)$

$$\beta_0 [1 - \beta_1]^{\frac{1}{(1-\beta_3)}} + D + \tau = log(s_0) \tag{A.2.10}$$

Solving (A.2.10) for τ yields

$$\tau = log(s_0) - \beta_0 [1 - \beta_1]^{\frac{1}{(1-\beta_3)}} \tag{A.2.11}$$

The stationary phase OD increment of the standardized curve

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

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

$$\Delta = -t_L + \frac{1}{\beta_2} log \left[\frac{\beta_1}{1 - \left((1 - \beta_1 e^{-\beta_2 t_L})^{\frac{1}{1-\beta_3}} + [\varphi] \right)^{1-\beta_3}} \right] \tag{A.2.12}$$

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