This document contains the **post-print pdf-version** of the referred paper:

**“Parameter estimations in predictive microbiology:
Statistically sound modelling of the microbial growth rate”**

by **Simen Akkermans, Filip Logist, Jan F. Van Impe**

which has been archived on the university repository Lirias of the KU Leuven
(<https://lirias.kuleuven.be/>).

The content is identical to the content of the published paper, but without the final typesetting by the publisher.

When referring to this work, please cite the full bibliographic info:

Akkermans, S., Logist, F., Van Impe, J.F., 2018. Parameter estimations in predictive microbiology: Statistically sound modelling of the microbial growth rate. *Food Research International*. (In Press)

The journal and the original published paper can be found at:
<https://doi.org/10.1016/j.foodres.2017.11.083>

The corresponding author can be contacted for additional info.

Conditions for open access are available at: <http://www.sherpa.ac.uk/romeo/>

Parameter estimations in predictive microbiology: Statistically sound modelling of the microbial growth rate

Simen Akkermans, Filip Logist, Jan F. Van Impe

BioTeC, Chemical and Biochemical Process Technology and Control,
Department of Chemical Engineering, KU Leuven, Ghent, Belgium,
OPTEC, Optimization in Engineering Center-of-Excellence, KU Leuven,
Belgium,
CPMF², Flemish Cluster Predictive Microbiology in Foods - www.cpmf2.be

[simen.akkermans, jan.vanimpe] @kuleuven.be

Correspondence to:

Prof. J. F. Van Impe

Chemical and Biochemical Process Technology and Control (BioTeC)

Department of Chemical Engineering, KU Leuven

Gebroeders de Smetstraat 1, B-9000 Ghent (Belgium)

jan.vanimpe@kuleuven.be

Tel: +32-16-32.14.66

Fax: +32-9-265.86.24

1 ABSTRACT

2
3 **Keywords:** Secondary model, exponential phase, variance stabilizing
4 transformation, sampling variance, confidence intervals.

5
6 When building models to describe the effect of environmental conditions on the
7 microbial growth rate, parameter estimations can be performed either with a *one-step*
8 *method*, i.e., directly on the cell density measurements, or in a *two-step method*, i.e., via
9 the estimated growth rates. The *two-step method* is often preferred due to its simplicity.
10 The current research demonstrates that the *two-step method* is, however, only valid if
11 the correct data transformation is applied and a strict experimental protocol is followed
12 for all experiments. Based on a simulation study and a mathematical derivation, it was
13 demonstrated that the logarithm of the growth rate should be used as a variance
14 stabilizing transformation. Moreover, the *one-step method* leads to a more accurate
15 estimation of the model parameters and a better approximation of the confidence
16 intervals on the estimated parameters. Therefore, the *one-step method* is preferred and
17 the *two-step method* should be avoided.

1 INTRODUCTION

In predictive microbiology, the effect of environmental conditions on the growth rate of microorganisms is generally modelled in two steps (i.e., via the so-called two-step method). In the first step, the microbial maximum specific growth rate (hereafter referred to as growth rate) is determined at several values of the studied conditions by fitting a primary model to the measured growth curves. In the second step, the relationship between the experimental conditions and the growth rate is modelled by fitting a secondary model to the growth rates determined in the first step (Whiting and Buchanan, 1993). At the end of the previous century, the use of experiments with environmental conditions that change as a function of time was introduced into predictive microbiology as a means of obtaining more information from growth experiments (Cunha et al. 1997; Versyck et al., 1999). A consequence of using these dynamic conditions is that the two-step parameter estimation methods could no longer be applied. Instead, the primary models, which describe the microbial growth as a function of time, were combined with the secondary models, which describe the effect of environmental conditions on microbial growth, and the combined models were identified by minimizing the error on the predicted growth. In this way, the parameter estimation is performed in a single step on the cell density data (leading to the so-called one-step method). Such parameter estimation techniques can also be applied when a set of static experiments is available as done, e.g., in Akkermans et al. (2016). A choice should thus be made between the one-step and two-step method.

A first point of comparison between the one-step and two-step modelling methods is the compatibility of their objective functions with the application of predictive microbiology, i.e., to minimise or prevent the growth of microorganisms in food products for a certain period of time. In the two-step method the error on the predicted

growth rate of a population is minimised whereas the one-step method minimises the error on the predicted growth, i.e., the (logarithm of) the increase of the number of cells. Therefore, the objective of predictive microbiology is more closely linked with the objective function of the one-step method than that of the two-step method. Moreover, the models used in predictive microbiology are overall largely empirical. This means that the models, and the assumptions they represent, are an oversimplification of the true behaviour of microorganisms and can thus never be expected to completely agree with the reality. It should be stressed that this is an entirely different matter than model(parameter) uncertainty. As a consequence, it is worth considering how there should be dealt with the discrepancies that exist between a model for the growth rate and the real microbial response, since this is also determined by the objective function. Comparing the objective functions of the modelling methods considered here, the objective function of the one-step method is preferred, as it minimises the error on the predicted growth instead of the growth rate.

Ratkowsky et al. (1991) was probably the first publication in predictive microbiology to draw attention to one of the main difficulties when working with the two-step method. They pointed out that the variance of bacterial growth responses such as the growth rate or lag time are dependent on the magnitude of the mean, i.e., that they are heteroscedastic. However, when a least squares regression is used to fit a model to data, this data should be homoscedastic, i.e., independent of its value. When heteroscedastic data is used in an unweighted least squares regression, too much confidence is placed in data with a high variance compared to data with a low variance. Such errors in the application of least squares regression techniques can lead to significant deviations between the model and the microbial system, which in their turn lead to poor predictions and too much confidence being placed in these predictions. A

possible solution to this problem is making a variance stabilizing transformation of the data before commencing the fitting procedure. Several authors proposed the use of either a square root transformation (Augustin and Carlier, 2000; Coroller, 2005; Zwietering, 1994) or logarithmic transformation (Alber and Schaffner, 1992; Masana, 2000; te Giffel and Zwietering, 1999) to obtain homoscedastic data. No consensus has been reached on the required transformation because (i) the probability distribution of measurements depends on the experimental methods, (ii) experimental studies of the variance were based on too limited data sets and (iii) simulation studies are highly dependent on their assumptions (Schaffner, 1998). If a transformation of the growth rates is made, the objective of minimizing the error on the transformed growth rate should be compared with the objective of minimizing the error on the microbial growth. This question of heteroscedasticity of the growth rates is of no importance to the one-step modelling method, since the model is fitted to the data of the population density and not to the estimates of the growth rate.

The one-step and two-step modelling methods also differ in the calculation of the confidence intervals of the model parameters. Since these confidence intervals are essential for modellers to know how much trust can be placed in the estimated parameters (and consequently in the model predictions), it is worth considering the effect of the chosen modelling method on the calculation of the confidence intervals. Even though the importance of determining the confidence intervals is generally accepted, this matter has not been discussed in previous publications in the field of predictive microbiology. The use of the one-step method in relation to confidence intervals has only been considered for case studies with time-varying environmental conditions, where the use of the two-step method is not possible (e.g., Longhi et al. 2017).

Jewell (2012) previously presented a comparative study of the application of the one-step and two-step procedure in predictive microbiology. For this purpose, parameter estimations were performed on three experimental datasets. The main conclusions of this research were that the 1-step method results in more precise parameter estimates and that it is more data efficient. It is however difficult to know the validity of these conclusions because the correct transformations of the growth rate were not considered when using the two-step procedure, even though this directly impacts the parameter estimation results.

To select a modelling method that is compatible with the estimation of parameters for secondary models in predictive microbiology, this research studies the variance of the growth rate and the calculation of confidence intervals in the one-step and two-step modelling method. Two options present themselves to study the variance of the maximum specific growth rate. The first being an experimental study and the second a simulation study. In case of an experimental study of the variance of the growth rate, both the experimental measurement error and the microbial variability would be included. The problem with an experimental determination of the variance of the growth rate (or in fact any variance) is that a large amount of data is needed to obtain an accurate estimate of the variance. Even though it is a common practice to determine a variance, standard deviation or standard error on as little as three data points, many more are needed to obtain an accurate estimate of the population variance. This point is proven by means of a simple simulation study in this research.

Based on a simulation study, this article investigates (i) the relevance of microbial variability and (ii) the sampling protocol when modelling the maximum specific growth rate, (iii) the required variance stabilizing transformations for the one- and two-step

118 method and (iv) the effect of the parameter estimation method on the calculation of
119 confidence intervals.

2 MATERIALS AND METHODS

2.1 Monte Carlo simulations

Random values of the normal, lognormal and Poisson distribution are generated during the simulations in this research by using the function *random* from the Statistical Toolbox of Matlab version 9.0 (The MathWorks Inc.).

2.2 Mathematical models

This article focuses on the estimation and description of the microbial growth rate. As such, the following simple primary model is applied in Sections 3.4 to estimate the microbial growth rate:

$$\frac{dN(t)}{dt} = \mu_{\max} \cdot N(t) \quad (1)$$

with $N(t)$ [CFU/mL] the cell density as a function of time and $\mu_{\max}$ [h^{-1}] the maximum specific growth rate. This mathematical model only considers the exponential phase of growth. The logarithmic transformation of this equation is often preferred:

$$\frac{dn(t)}{dt} = \mu_{\max} \quad (2)$$

where $n(t) \left[\ln \left(\frac{CFU}{mL} \right) \right]$ is the natural logarithm of the cell density. In Section 3.7, the more complex model of Baranyi and Roberts (1994) is used to study the calculation of confidence intervals of the model parameters:

$$\frac{dn(t)}{dt} = \frac{1}{1 + \exp(-q(t))} \cdot \mu_{\max}(T) \cdot [1 - \exp(n(t) - n_{\max})] \quad (3)$$

$$\frac{dq(t)}{dt} = \mu_{\max}(T) \quad (4)$$

with $\mu_{\max}$ [h^{-1}] the maximum specific growth rate and $n_{\max}$ [$\ln(CFU/mL)$] the maximum cell density. $q(t)$ [—] is a measure for the physiological state of the cells and

serves to describe the lag phase of the growth curve. The initial values of $n(t)$ and $q(t)$ are respectively $7 \ln(\text{CFU/mL})$ and -1 for all simulations. $n_{\max}$ was chosen at a value of $22.55 \ln(\text{CFU/mL})$. All growth curves were simulated until the population reached a level of just over $22.53 \ln(\text{CFU/mL})$ to ensure that all parameters of the primary model were identifiable from the simulation data.

The Cardinal Temperature Model with Inflection (CTMI, Rosso et al. 1993) was used as a secondary model to describe the maximum specific growth rate $\mu_{\max} [\text{h}^{-1}]$ as a function of temperature $T [^{\circ}\text{C}]$:

$$\mu_{\max}(T) = \mu_{\text{opt}} \cdot \frac{(T-T_{\min})^2 \cdot (T-T_{\max})}{(T_{\text{opt}}-T_{\min}) \cdot [(T_{\text{opt}}-T_{\min}) \cdot (T-T_{\text{opt}}) - (T_{\text{opt}}-T_{\max}) \cdot (T_{\text{opt}}+T_{\min}-2T)]} \quad (5)$$

with $T_{\min} [^{\circ}\text{C}]$ and $T_{\max} [^{\circ}\text{C}]$ the minimum and maximum temperature that define the range of environmental conditions where growth is possible. $T_{\text{opt}} [^{\circ}\text{C}]$ is the temperature at which the optimum growth rate $\mu_{\text{opt}} [\text{h}^{-1}]$ is reached. The nominal parameter values were selected based on the results of Van Derlinden et al. (2008) and are specified in Table 1.

2.3 Parameter estimations

The objective function of the parameter estimations considered in this research is the minimization of the sum of squared errors (SSE) for $M_{\#}$ measurements of the output variable X , i.e., the cell density or the growth rate :

$$\text{SSE} = \sum_{i=1}^{M_{\#}} \left(X_{m,i} - X_{p,i}(p) \right)^2 \quad (6)$$

with $X_{m,i}$ the measured value and $X_{p,i}(p)$ the predicted value for a set of parameters p . The $(100 - \alpha)\%$ confidence interval of every parameter p_i is calculated based on the Student's t -distribution:

$$\left[p_i \pm t_{(100-\frac{\alpha}{2}), M_{\#}-P_{\#}} \cdot \sqrt{\sigma_{p_i}^2} \right] \quad (7)$$

where $P_{\#}$ is the number of parameters and consequently $M_{\#} - P_{\#}$ is the number of degrees of freedom. $\sigma_{p_i}^2$ is the variance on the model parameter p_i and is found on the main diagonal of the variance covariance matrix V , which is approximated as the inverse of the Fisher Information Matrix (FIM) F (Walter and Pronzato, 1997):

$$\sigma_{p_i}^2 = V(i, i) \quad (8)$$

$$V = F^{-1} \quad (9)$$

$$F = \frac{1}{MSE} \cdot J^T \cdot J \quad (10)$$

$$MSE = \frac{SSE}{M_{\#}-P_{\#}} \quad (11)$$

with J the Jacobian matrix and MSE the mean sum of squared errors. The optimal set of parameters are calculated using the *lsqnonlin* routine of the Optimization Toolbox of Matlab version 9.0.

2.4 Experimental protocol and variability

The variability on the measurements is of course dependent on the experimental protocol. The simulations are performed assuming a widely adopted protocol with viable plate counts. For this protocol, dilutions of a sample are made and the cells in this diluted sample are grown on a Petri dish containing solid medium. In this research, the amount of cells is expressed as concentration but the same conclusions can be reached when working with a total number of cells. The concentration of cells in the sample is calculated after counting the number of colonies that originate from the individual cells. In this protocol, a relatively high cell density is needed (e.g., more than 1,000 CFU/mL). For the simulations in this research, 50 μ L is transferred from a 10-fold dilution to a Petri dish to obtain between 40 and 400 colonies.

190 To obtain the experimental uncertainties, the necessary probability distributions
191 are included in the calculation. Pipetted volumes of 100 and 900 μL used to make the
192 dilution series are assigned a normal distribution with a standard deviation of
193 respectively 0.3 and 1.8 μL , based on the manufacturer's specifications (Biohit,
194 Sartorius AG). The amount of cells contained in a sample when making dilutions and
195 when plating are determined by a Poisson distribution.

3 RESULTS AND DISCUSSION

In a first step, the difficulty of an experimental study of the variance of, e.g., the maximum specific growth rate is discussed. Then, the variance of the growth rate that is characteristic to the variability of the individual cells in a population is studied. Continuing on these results, the effect of the experimental variability of the cell density measurements on the calculated growth rates is determined. Then, the effect of the experimental protocol on the variance of the growth rate is examined. Finally, the calculation of confidence intervals based on the two parameter estimation methods is compared.

3.1 Experimental study of the variance

Say that for a certain experiment the growth rate is normally distributed with a variance of 0.10 h^{-2} , then a number of growth rates can be randomly sampled from this distribution. The variance of the growth rate is calculated with the general formula for the sample variance. This calculation is represented by Eq. (11), with the number of parameters equal to one and the predicted value equal to the mean. This same calculation is performed 50 times for situations with 5, 20, 50 and 100 samples. The calculated values of the variance of the growth rate are plotted in Fig. 1. These simulation results illustrate that a large number of samples is needed to obtain an accurate estimate of the variance, i.e., with a high certainty of being close to the true population variance. This could also be concluded by calculating the confidence intervals of the variance for different sample sizes (for further information, see, e.g., Kutner et al., 2004).

Moreover, when experimentally determining the microbial growth rate, its variability will in part be due to the uncertainty on the environmental conditions. Small

differences in, e.g., the temperature, pH and media composition will lead to differences in the growth rate of a population. When performing experiments to build secondary models, such errors have to be minimised so that they do not influence the parameter estimation, since a simple least squares regression assumes that there is no uncertainty on the independent variables. When studying the variance of the growth rate it would of course be impossible to separate this uncertainty from the experimental variability so that they would be analysed together.

For these reasons, it is in practice infeasible to obtain information on the homo- or heteroscedasticity of the growth rate measurements based on an experimental method. Consequently, this study of the variance of the growth rate is performed by means of simulations.

3.2 Variance due to the variability between individual cells

It is well known that individual cells have varying division times under the exact same environmental conditions. The combination of the cell division times of all the individual cells results in the growth rate of a population and therefore also in the variability on the growth rate that is caused by the behaviour of individual cells. The variance of the growth rate that originates from this variability between individual cells is discussed in this section based on simulations of the growth of a population as a group of individually considered cells. For the first simulation, the growth of a population is simulated starting from a single cell.

Examples of studies of single cell behaviour are found, e.g., in Francois et al. (2005), Metris et al. (2005), Pin and Baranyi (2006) and Koutsoumanis and Lianou (2013). In these publications, the probability distributions of the division times of several generations are reported. The results indicated that the mean and variance of

these distributions are approximately constant after the lag phase. Since only the growth rate is considered in this research, and not the lag phase duration, the parameters of the probability distributions are assumed to be constant. The probability distribution is assumed to be lognormal, since the division time can never have a negative value. It is important to stress that the conclusions of the following simulation study would remain unchanged if, e.g., a Weibull or gamma probability distribution is used. Based on the results of Pin and Baranyi (2006) for *Escherichia coli* K12 at 32°C, the mean and variance of the division time are set equal to 0.50 h and 0.04 h², respectively. This probability distribution is used to simulate 500 growth curves, starting from a single cell. The simulations are performed in time steps of 0.1 min. Each time a cell reaches the division age, the mother cell is replaced by two daughter cells and they are each assigned a random division time according to the previously described probability distribution. The resulting growth curves are shown in Fig. 2. Fig. 2 (a) illustrates that a large variation in cell densities can occur, due to the variability of the individual cells. However, when looking at the logarithm of the cell densities in Fig. 2 (b), it seems that the main divergence between the growth curves of the different simulations is confined to the first 2-ln increase of the cell density. After this point, the evolution of the logarithmic cell densities continues more or less parallel to one another. This can easily be explained by the law of large numbers or mathematically by this simple formula that expresses the variance of the mean σ_m^2 of a number of uncorrelated samples $S_{\#}$, each of which has a known variance σ^2 :

$$\sigma_m^2 = \frac{\sigma^2}{S_{\#}} \quad (12)$$

Applied to the current case study, this means that the probability distribution of the growth rate of a population becomes narrower with an increasing number of cells. Since growth experiments in liquid media are traditionally performed with

concentrations between 10^3 and 10^9 CFU/mL and in volumes of at least several millilitres, it is worth wondering if this variability is still relevant if such a high number of cells is considered.

To investigate this question, 500 Monte Carlo simulations of the growth of a population are performed starting from a single cell to about $9 \cdot 10^6$ cells. All these simulations are plotted in Fig. 3, starting from the point where they reach 1,000 cells. In this way, the figure illustrates the possible differences in the growth of a population starting from 1,000 cells. Where the growth of cultures starting from 1 cell (Fig. 2) showed clearly that many different scenarios are possible, there is almost no distinction visible in the growth curves starting from 1,000 cells (Fig. 3 (a) and (b)). Only when looking very closely at the first minutes after 1,000 cells are reached in Fig. 3 (c), the slight differences between the simulations are clear.

When studying the growth rate of a population considering the variability in single cell behaviour, it is worth making a distinction between the *instantaneous* specific growth rate and the *time-independent* specific growth rate. In this distinction the instantaneous growth rate is valid at a specific point in time. This growth rate is dependent on the division times of all the cells in the population and it is ever changing with time. The time-independent specific growth rate is calculated from the growth of a population over a certain time period by considering that the logarithm of the cell density versus time is linear. For this calculation the specific growth rate is assumed to be independent of time.

The simulations of Fig. 2 and Fig. 3 demonstrate that the variance of the instantaneous specific growth rate decreases with time, due to the increase in the number of cells, as formulated by Eq. (12). The variance of the time-independent specific growth rate on the other hand depends on the range of cell densities considered

during the growth of the population. It is the instantaneous specific growth rate that represents the variability of the growth of a population and it is the effect of the variance of this growth rate that should be studied by researchers who are interested in providing probability distributions for the number of cells in a population as a function of time (contrary to what has been done by, e.g., Koutsoumanis and Lianou (2013)). The growth rate that is studied in this research, on the other hand, is the time-independent growth rate. It can be determined for each simulation by estimating the slope of the curves in Fig. 3 (b). The mean of these growth rates is 1.37 h^{-1} with a variance of just $9.41 \cdot 10^{-7} \text{ h}^{-2}$. This very low variance demonstrates that the variability of the growth rate of the population itself can be neglected during conventional growth experiments. The variance on the growth rate that is found is thus due to the experimental variability on the measurements of the cell density. This experimental variability is studied in the next section.

3.3 Variability of the cell density measurements

Based on the experimental protocol (explained in the Materials and Methods section), the cell density measurements with experimental variability are simulated. For this simulation, an experiment is considered in which cells are grown from a concentration of 7 to $19 \ln(\text{CFU/mL})$ at a growth rate of 1 h^{-1} and concentration measurements are determined according to the protocol in Section 2.4. To determine the variance of the measurements of the cell density, a Monte Carlo simulation with 50,000 iterations is performed at every cell density. The resulting variance is plotted as a function of the logarithm of the cell density in Fig. 4 (a). This figure illustrates how the variance increases linearly with the amount of cells plated, due to the variance of the Poisson distribution, which increases with the mean. It also demonstrates that the

error increases stepwise with every additional dilution, since the error is multiplied by 10 for each dilution when calculating the cell density in the undiluted sample. If the square root of the cell density is used, the variance is stabilised within the same dilution but still increases with every dilution (Fig. 4 (b)). When taking the logarithm of the cell density, the variance decreases with an increasing number of cells within the same dilution (Fig. 4 (c)). In this way, the increase of the variance with every additional dilution is compensated. Thus, taking the logarithm of the cell density doesn't completely stabilise the variance but it prevents making a systematic error during fitting. Fig. 4 (c) also indicates the experimentally determined variance of the logarithm of the cell density that was reported by Van Derlinden et al. (2008) for the same experimental procedure. Naturally, this experimentally determined value is higher than the ideal values of the current study.

These results demonstrate that the necessary variance stabilizing transformation is dependent on the experimental protocol used. Care should be taken not to make a logarithmic transformation of (a measure of) the cell density without considering if this is a valid procedure for the given experimental protocol. The same type of simulations is used in the next section to generate in silico results at different growth rates.

3.4 Experimental uncertainty of the growth rate

Since it has been established in the previous section that logarithmic transformations of the cell density measurements are acceptable for the considered experimental protocol, parameter estimations are performed on such data to estimate the growth rate. Measurements are simulated for experiments with growth rates between 0.1 and 2.0 h⁻¹. For each experiment, 10 samples are taken and these samples are taken at the same concentrations for all experiments, irrespective of the growth rate.

The distribution of the cell density measurements is identical to the distribution described in the previous section. The logarithmic form of the simple primary model described in Eq. (2) is fitted to these cell densities to find the growth rate. This procedure is repeated 50,000 times at every growth rate, to obtain an accurate estimate of the variance.

Fig. 5 (a) represents the relationship between the growth rate and its variance. These results demonstrate that the variance on the experimentally determined growth rate is much higher at high growth rates than at low growth rates. Therefore, when fitting a secondary model on untransformed growth rate data with an unweighted least squares method, too much confidence is placed in the results at high growth rates in comparison with the results at low growth rates. This leads to inaccurate predictions of the microbial growth at low growth rates. Performing a weighted least squares regression is often not a good solution because too much experimental data is required to obtain an accurate estimation of the variance on the growth rate (as demonstrated previously). The square root transformation of the growth rate is often suggested as a variance stabilizing transformation but the results of Fig. 5 (b) illustrate that this is not the case in the current study. This transformation only transforms the quadratic relationship into a linear relationship. On the other hand, the logarithmic transformation results in homoscedastic variance (Fig. 5 (c)). This behaviour of the variance leads to the conclusion that the parameter estimates of the growth rate follow a Weibull distribution. The logarithmic transformation is therefore most suitable for researchers who wish to apply the two-step modelling method for the considered experimental protocol. It can even be demonstrated that for the very strict experimental protocol used in these simulations, the two-step method with logarithmically transformed growth rates has an objective function that is equivalent to that of the one-step method and

therefore leads to the same combination of parameter estimates. The experimental protocol assumed for the simulations is however unrealistically strict. Therefore, Section 3.6 the effect of changes in the experimental protocol on the variance of the growth rate.

3.5 Mathematic derivation of the growth rate transformation

Based on a simulation study, the previous section demonstrates that the logarithmic transformation of the growth rate should be used to stabilise its variance. A specific experimental protocol was taken in mind to perform these simulations. Consequently, it is interesting to further generalise this analysis of the variance to eliminate as many assumptions as possible. This is done here by mathematically deriving the relationship between the variance of an estimated growth rate and its mean. By mathematically proving that the variance of the estimated growth rate is proportional to the square of the mean, it is also proven that the logarithmic transformation has to be used to stabilise the variance (Box et al., 2005).

A simple exponential growth process is described by the following equation:

$$n(t) = \mu \cdot t + n_0 \quad (13)$$

with μ the microbial growth rate and n_0 the initial cell density. An identical number of samples is taken in every experiment, regardless of the growth rate and the experiment stops once a maximum population level $n_{\max}$ has been reached. The duration t_E of such an experiment is calculated as follows:

$$t_E = \frac{n_{\max} - n_0}{\mu} \quad (14)$$

As the parameter estimation of μ and n_0 is a simple linear regression, the estimated growth rate $\hat{\mu}$ of an experiments is (Kenney and Keeping, 1966):

$$\hat{\mu} = \frac{\bar{t} \circ \bar{n} - \bar{t} \cdot \bar{n}}{\bar{t} \circ \bar{t} - \bar{t}^2} \quad (15)$$

with $\mathbf{n}$ a vector of measurements of the population level at times $\mathbf{t}$ (with equidistant time points). The ‘ $\circ$ ’ notation refers to the Hadamard product in which every element of the first vector is multiplied with the corresponding element of the second vector. The sample times $\mathbf{t}$ can be replaced with the multiplication of the duration of the experiment with the relative sample times $\mathbf{t}_r$. These relative sample times are ideally identical for every experiment, independent of the growth rate. This leads to the following equation:

$$\hat{\mu} = \frac{t_E}{t_E^2} \cdot \frac{\overline{\mathbf{t}_r \circ \mathbf{n}} - \overline{\mathbf{t}_r} \cdot \overline{\mathbf{n}}}{\overline{\mathbf{t}_r \circ \mathbf{t}_r} - \overline{\mathbf{t}_r}^2} = \frac{\mu}{n_{\max} - n_0} \cdot \frac{\overline{\mathbf{t}_r \circ \mathbf{n}} - \overline{\mathbf{t}_r} \cdot \overline{\mathbf{n}}}{\overline{\mathbf{t}_r \circ \mathbf{t}_r} - \overline{\mathbf{t}_r}^2} \quad (16)$$

The variance of an estimated growth rate can thus be written as:

$$\sigma^2(\hat{\mu}) = \sigma^2 \left(\frac{\mu}{n_{\max} - n_0} \cdot \frac{\overline{\mathbf{t}_r \circ \mathbf{n}} - \overline{\mathbf{t}_r} \cdot \overline{\mathbf{n}}}{\overline{\mathbf{t}_r \circ \mathbf{t}_r} - \overline{\mathbf{t}_r}^2} \right) = \mu^2 \cdot \sigma^2 \left(\frac{1}{n_{\max} - n_0} \cdot \frac{\overline{\mathbf{t}_r \circ \mathbf{n}} - \overline{\mathbf{t}_r} \cdot \overline{\mathbf{n}}}{\overline{\mathbf{t}_r \circ \mathbf{t}_r} - \overline{\mathbf{t}_r}^2} \right) \quad (17)$$

In the theoretical situation used in this example, $n_{\max}$, n_0 and $\mathbf{t}_r$ are identical for every experiment. The variance due to the values of $\mathbf{n}$ will not change between experiments as long as a sufficiently large number of data points is used. As such, the variance of the growth rate is clearly proportional to the square of its mean and can therefore be stabilised with a logarithmic transformation.

In the current derivation, there were no assumptions on a specific experimental protocol. As such, the decreasing variance with a decreasing growth rate is only due to the fact that the duration of the experiment increases. The main assumptions that were still made in this derivation were that (i) the number of samples is constant, (ii) the samples are equally distributed over the duration of the experiment and (iii) the errors on the cell density measurements are independent and identically distributed normal random variables. The current derivation was also made for a simple log-linear model, instead of a more complex model that includes the lag and stationary phases of growth.

The conclusion holds for any model that describes the exponential phase as a log-linear relationship as long as the model parameters are uncorrelated.

3.6 Applicability of the two-step method

The simulations made in the previous section demonstrated that it is possible to stabilise the variance on the experimentally determined growth rates by making a logarithmic transformation. However, a constant variance is only obtained if the measurements are made according to a very strict experimental protocol. The analysis was made for a situation in which only experimental results of the exponential phase of growth are considered. In practice this would mean that the lag and stationary phase have to be omitted during sampling or that the measurements of these phases of growth have to be removed from the results. However, excluding these phases of growth requires the modeller to make decisions on when the lag phase ends and when the stationary phase begins. This exclusion of data also results in a loss of information on the growth rate.

The logarithmic transformations of the growth rate can only be homoscedastic if the same amount of samples are taken, at the same concentrations and with the same number of Petri dishes for each sample in every experiment. This is, however, a very unlikely scenario and experimenters will often choose to change the sample frequency, the amount of Petri dishes made per sample, the amount of experiments at some conditions and the range of concentrations included in an experiment. All these changes in the experimental protocol have a significant effect on the variance of the growth rate, as demonstrated in Fig. 6. The variance of the logarithm of the cell density decays exponentially with an increase of (i) the number of samples, (ii) the concentration range and (iii) the number of Petri dishes. This means that, in practice, the variance of the

growth rates still varies, even if the correct variance stabilizing transformation is applied. It should be noted that if experimenters choose to use a fixed sampling rate instead of a fixed total number of samples for each experiment, the variance on the estimated growth rate will decrease even more with decreasing growth rates. This means that the square root transformation will definitely not be suitable in such a case either. But also the logarithmic transformation would assign too little value to measurements at low growth rates as opposed to those at high growth rates.

3.7 Calculation of confidence intervals

Apart from estimating the parameters of a secondary model, also the confidence intervals of these parameters should be provided. These intervals give an indication of the amount of confidence that can be placed in the parameter estimates for a given set of experimental data and can be used to produce confidence intervals on the model predictions. However, the confidence intervals are estimates themselves. They are commonly calculated via the FIM as explained in Section 2.3. The accuracy of these confidence intervals is determined by (i) the estimated parameter values, (ii) the approximation of the variance on the experimental measurements by the MSE and (iii) the Cramér-Rao lower bound. The Cramér-Rao bound states that the inverse of F for the nominal model parameters provides the lower bound of the variance on the estimated model parameters $\bar{p}$ (Walter and Pronzato, 1997):

$$V(\bar{p}) \geq F(p)^{-1} \quad (18)$$

All three of these approximations improve when using larger amounts of data. To study the accuracy of the confidence intervals, a Monte Carlo simulation with 1,000 iterations was performed in which the parameters of a secondary model were estimated on a randomised dataset and the confidence intervals were calculated. Both the one-

step and two-step parameter estimation procedure were applied. The secondary model used to describe the effect of temperature on the microbial growth rate and is given in Section 2.2. Growth rates were calculated at equidistant temperatures in the range of 10 to 45°C using the CTMI. Growth curves were simulated using the model of Baranyi and Roberts (1994) as explained in Section 2.2. Based on the set of parameter estimates that resulted from the Monte Carlo method, an approximation of the 95% confidence bounds was made. This is seen as an approximation of the real variability of the parameter estimates for the given experimental data. The 95% confidence bounds were also calculated in each iteration via the FIM as explained in Section 2.3 (Eq. (8) – Eq. (11)).

Initially, 6 experiments were generated and each experiment contained 6 data points. The experimental variability characterised in Section 3.3 was added to the simulation data during the Monte Carlo method. The number of experiments and the number of data points were simultaneously increased to 10 in steps of 1. The comparison of the total width of the 95% confidence bounds from the Monte Carlo method and the FIM approximation is presented in Fig. 7 for different quantities of data. Only the results for the estimation of $T_{\min}$ are presented as the results for the other parameter estimates led to the same conclusions. First of all, the observed trend is not fluent. This is a consequence of the fact that the model parameter accuracy is not only affected by the amount of data but also by the sampling conditions (specific temperatures of the experiments and sampling times). As such, it is possible that the amount of information decreases with an increasing amount of data. With respect to the variability of the parameter estimates, the results demonstrate that there was hardly any difference between the one-step and two-step method. However, the calculation of the 95% confidence bounds with the FIM through the two-step method is significantly less

accurate and more variable than for the one-step method. Increasing the amount of experimental data improves the accuracy of the parameter estimates and variability of confidence bounds for both methods but the advantage of the one-step method over the two-step method remains. Even though the overestimation of the inaccuracy by the two-step method can be seen as fail-safe, it is problematic that the uncertainty cannot be estimated accurately without using excessive amounts of data.

Apart from the effect of the quantity of experimental data, also the effect of its quality was tested. This was tested by adding Gaussian noise with a mean of zero and variance of up to $0.5 \ln(\text{CFU/mL})^2$ to the cell density measurements on top of the variability that was already included. These simulations were all performed with 10 experiments in each dataset and 10 data points in each experiment. The 95% confidence bounds as determined by the Monte Carlo method and FIM approximation are presented in Fig. 8. The two-step method was influenced stronger by a decrease in the quality of the experimental measurements. This is seen from the fact that there was a higher increase in the variability on the confidence bounds derived from the FIM.

In the results of Fig. 7 and 8, the real variability of the parameter estimates was almost identical between the one-step and two-step method. It is worth noting that this is a consequence of the correct weighing of the growth rates by using the logarithmic transformation. As stated in the previous section, this transformation only provides a full stabilization of the variance under ideal conditions, which were met during these simulations. As such, differences are expected when applying the two-step method on real experimental data. Finally, Fig. 7 and Fig. 8 also reveal that the quality and quantity of experimental data has a larger effect on the accuracy of confidence intervals than on the accuracy of the model parameters. In other words, more or better quality experimental data may be required to accurately determine the uncertainty of the

519 parameters than to accurately determine the model parameters themselves. The one-
520 step parameter estimation method will always provide more accurate estimates of the
521 model parameter uncertainty as this method obtains information from a much larger
522 number of measurements (cell densities instead of growth rates).

4 CONCLUSIONS

The first important conclusion of this work is that the variability between individual cells can generally be neglected when studying the variability of the growth rate, since the large amount of cells in the considered populations causes there to be very little variability in the growth rate of the entire population. Therefore, the variability that is found on the growth rate is a consequence of errors on the experimental measurements. By simulating this variability, it is confirmed that the logarithmic transformation of the cell densities should be used when working with dilutions and a plate count method. But it is also important to remember that the necessary transformation is clearly dependent on the chosen experimental protocol. Based on a further simulation study, it was demonstrated that the logarithmic transformation is most suitable to stabilise the variance on the estimated growth rate. The same conclusion was reached by making a mathematical derivation of the relationship between the variance of the estimated growth rate and its mean. Based on these results, it seems unlikely that any experimental protocol exists in which the square root transformation of the growth rate is preferable to the logarithmic transformation. Moreover, this variance is proven to be highly dependent on the specifications of the applied experimental protocol. This means that the individual growth rate estimates will in practice not be weighted correctly, even if the appropriate transformation is used. Finally, it is concluded that the one-step method will result in a more accurate determination of the confidence intervals of the parameter estimates than the two-step method. Based on all these findings it can be concluded that the one-step method is more suitable for the estimation of parameters of secondary models.

546 **5 ACKNOWLEDGEMENTS**

547 This work was supported by project PFV/10/002 (Center of Excellence OPTEC-
548 Optimization in Engineering) of the KU Leuven Research Fund, projects G.0930.13
549 and KAN.15189.13 of the Fund for Scientific Research-Flanders, and the Belgian
550 Program on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science
551 Policy Office (IAP Phase VII/19 DYSCO).

6 REFERENCES

- Akkermans, S., Noriega Fernandez, E., Logist, E., Van Impe, J.F., 2016. Introducing a novel interaction model structure for the combined effect of temperature and pH on the microbial growth rate. *International Journal of Food Microbiology*.
- Alber, S.A., Schaffner, D.W., 1992. Evaluation of data transformations used with the square root and Schoolfield models for predicting bacterial growth rate. *Applied and Environmental Microbiology*, 58(10): 3337-3342.
- Augustin, J.-C., Carlier, V., 2000. Modelling the growth rate of *Listeria monocytogenes* with a multiplicative type model including interactions between environmental factors. *International Journal of Food Microbiology* 56: 53-70.
- Baranyi, J., Roberts, T.A., 1994. A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology* 23, 277-294.
- Box, G.E.P., Hunter, J.S., Hunter, W.G., 2005. *Statistics for Experimenters*. 2nd ed. John Wiley & Sons, Hoboken, USA.
- Coroller, L., Guerrot, V., Huchet, V., Le Marc, Y., Mafart P., Sohier, D., Thuault, D., 2005. Modelling the influence of single acid and mixture on bacterial growth. *International Journal of Food Microbiology*, 100: 167-178.
- Cunha, L.M., Oliveira, F.A.R., Brandão, T.R.S., Oliveira, J.C., 2013. Optimal experiment design for estimating the kinetic parameters of the Bigelow Model. *Journal of Food Engineering* 33(1-2): 111-128.
- Francois, K., Devlieghere, F., Smet, K., Standaert, A.R., Geeraerd, A.H., Van Impe, J.F., Debevere, J., 2005. Modelling the individual cell lag phase: Effect of temperature and pH on the individual cell lag distribution of *Listeria monocytogenes*. *International Journal of Food Microbiology*, 100: 41-53.

- 576 Jewell, K., 2012. Comparison of 1-step and 2-step methods of fitting microbiological
577 models. *International Journal of Food Microbiology*, 160: 145-161.
- 578 Kenney, J. F., Keeping, E.S., 1966. *Mathematics of statistics*. 3rd ed. D. Van Nostrand
579 Company, Princeton, USA.
- 580 Koutsoumanis, K.P., Lianou, A., 2013. Stochasticity in colonial growth dynamics of
581 individual bacterial cells. *Applied and Environmental Microbiology*, 79 (9): 2294-
582 2301.
- 583 Kutner, M.H., Neter, J., Nachtsheim, C.J., Wasseman, W., 2004. *Applied linear*
584 *statistical models*. McGraw-Hill, New York, USA.
- 585 Longhi, D.A., Martins, W.F., da Silva, N.B., Carciofi, B.A.M., de Aragão, G.M.F.,
586 Laurindo, J.B., 2017. Optimal experiment design for improving the estimation of
587 growth parameters of *Lactobacillus viridescens* from under non-isothermal
588 conditions. *International Journal of Food Microbiology*, 240: 57-62.
- 589 Masana, M.O., Baranyi, J., 2000. Adding new factors to predictive models: the effect
590 on the risk of extrapolation. *Food Microbiology*, 17: 367-374.
- 591 Metris, A., Le Marc, Y., Elfving, A., Ballagi, A., Baranyi, J., 2005. Modelling the
592 variability of lag times and the first generation times of single cells of *E. coli*.
593 *International Journal of Food Microbiology*, 100: 13-19.
- 594 Pin, C., Baranyi, J., 2006. Kinetics of single cells: Observations and modelling of a
595 stochastic process. *Applied and Environmental Microbiology*, 72 (3): 2163-2169.
- 596 Ratkowsky, D.A., Ross, T., McMeekin, T.A., Olley, J., 1991. Comparison of
597 Arrhenius-type and Belehradek-type models for prediction of bacterial growth in
598 foods. *Journal of Applied Bacteriology*, 71: 452-459.

- 599 Rosso, L., Lobry, J. R., Flandrois, J. P., 1993. An unexpected correlation between
600 cardinal temperatures of microbial growth highlighted by a new model. *Journal of*
601 *theoretical Biology* 162, 447-463.
- 602 Schaffner, D.W., 1998. Predictive food microbiology Gedanke experiment: why do
603 microbial growth data require a transformation? *Food Microbiology*, 15: 185-189.
- 604 te Giffel, M.C., Zwietering, M.H., 1999. Validation of predictive models describing the
605 growth of *Listeria monocytogenes*. *International Journal of Food Microbiology*, 46:
606 135-149.
- 607 Van Derlinden, E., Bernaerts, K., Van Impe, J.F., 2008. Accurate estimation of cardinal
608 growth temperatures of *Escherichia coli* from optimal dynamic experiments.
609 *International Journal of Food Microbiology*. 128: 89-100.
- 610 Versyck, K.J., Bernaerts, K., Geeraerd, A.H., Van Impe, J.F., 1999. Introducing optimal
611 experimental design in predictive modeling: A motivating example. *International*
612 *Journal of Food Microbiology*, 51: 39-51.
- 613 Walter, E., Pronzato, L., 1997. *Identification of parametric models from experimental*
614 *data*. Springer, Germany, Berlin.
- 615 Whiting, R.C., Buchanan, R.L., 1993. "A classification of models for predictive
616 microbiology." *Food Microbiology*, 10: 175-177.
- 617 Zwietering, M.H., Cuppens, H.G.A.M., De Wit, J.C., Van 't Riet, K., 1992. Evaluation
618 of data transformations and validation of a model for the effect of temperature on
619 bacterial growth. *Applied and Environmental Microbiology*, 60(1): 195-203.

620

FIGURES

621

622 **Fig. 1:** 50 Simulations of calculated variances based on (a) 5, (b) 20, (c) 50 and (d)

623 100 samples generated with a normal distribution and variance of 0.1 h^{-1} .

Fig. 2: Growth of a population during 5 h, starting from a single cell. The population size is presented both in (a) untransformed and (b) logarithmically transformed units.

Fig. 3: 500 Monte Carlo simulations of the growth of a population starting from the point it reaches 1,000 cells. The population size is presented both in (a) untransformed and (b) logarithmically transformed units. A detailed view of the divergence between the simulations is represented in (c).

Fig. 4: Variance of (a) the cell density and its (b) square root and (c) logarithmic transformation with the experimentally determined variance of Van Derlinden et al. (2008) (---).

Fig. 5: The variance of (a) the growth rate and its (b) square root and (c) logarithmic transformation versus the growth rate.

Fig. 6: Dependency of the variance of the logarithm of the growth rate on (a) the number of samples, (b) the concentration range and (c) the number of Petri dishes for each sample.

Fig. 7: The effect of the number of experiments and samples on the determination of the total width of the 95% confidence intervals by using a Monte Carlo simulation (x) and the approximations with the Fisher information matrix (x) with error bars (95% confidence). Results are presented for both the (a) one-step and (b) two-step parameter estimation procedure.

Fig. 8: The effect of the variance added to the experimental measurements on the determination of the total width of the 95% confidence intervals by using a Monte Carlo simulation (x) and the approximations with the Fisher information matrix (x) with error bars (95% confidence). Results are presented for both the (a) one-step and (b) two-step parameter estimation procedure.

TABLES

Parameters	Values
$T_{\min}$ [°C]	2.3
T_{opt} [°C]	40.6
$T_{\max}$ [°C]	45.5
μ_{opt} [h ⁻¹]	2.49

Table 1: Nominal parameter values of the CTMI.