International Journal of Food Microbiology 282 (2018) 1–8

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

A tutorial on uncertainty propagation techniques for predictive microbiology models: A critical analysis of state-of-the-art techniques☆

Simen Akkermans, Philippe Nimmegeers, 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

ARTICLE INFO

Keywords:
Prediction uncertainty
Parameter estimation
Sigma point method
Linear approximation
Monte Carlo method

ABSTRACT

Building mathematical models in predictive microbiology is a data driven science. As such, the experimental data (and its uncertainty) has an influence on the final predictions and even on the calculation of the model prediction uncertainty. Therefore, the current research studies the influence of both the parameter estimation and uncertainty propagation method on the calculation of the model prediction uncertainty. The study is intended as well as a tutorial to uncertainty propagation techniques for researchers in (predictive) microbiology. To this end, an in silico case study was applied in which the effect of temperature on the microbial growth rate was modelled and used to make simulations for a temperature profile that is characterised by variability. The comparison of the parameter estimation methods demonstrated that the one-step method yields more accurate and precise calculations of the model prediction uncertainty than the two-step method. Four uncertainty propagation methods were assessed. The current work assesses the applicability of these techniques by considering the effect of experimental uncertainty and model input uncertainty. The linear approximation was demonstrated not always to provide reliable results. The Monte Carlo method was computationally very intensive, compared to its competitors. Polynomial chaos expansion was computationally efficient and accurate but is relatively complex to implement. Finally, the sigma point method was preferred as it is (i) computationally efficient, (ii) robust with respect to experimental uncertainty and (iii) easily implemented.

1. Introduction

During the last decades, researchers in the field of predictive microbiology have focused on developing and fine-tuning a wide range of mathematical models that contribute to the assessment and prediction of microbial food safety and quality. Currently, there is a wide interest in moving towards mechanistic modelling methods such as individual based models (e.g., Kreft et al., 1998; Tack et al., 2015) or systems biology approaches (e.g., Brul et al., 2008; Vercammen et al., 2017). In practice, however, the state-of-the-art for real life application will remain for a considerable time the use of grey box models. These grey box models are built to deliver a simplified representation of the relevant microbial response (e.g., growth rate, inactivation rate, probability of growth). Grey box models require experimental data to select mathematical model structures and to estimate the most suitable combination of model parameters. As such, building mathematical models in the field of predictive microbiology will remain, for the time being, a data driven science.

The experimental data used to build a mathematical model will influence the choice of the model structure and the estimated values of the model parameters. As such, the experimental data also influences the model predictions that will be obtained. Knowing this, several publications have focused on assessing the quality and validity of the models that are obtained. For example, Ross (1996) developed indices to evaluate the accuracy and bias of models based on the predicted generation time. Apart from the accuracy, also variation plays an important role when modelling microbial responses. The sources of variation in predictive microbiology were distinguished as follows by Van Impe et al. (2001): (i) the type and quantity of microorganisms in the initial microbial load, (ii) the true intrinsic and extrinsic conditions that characterise a food product, (iii) the lack of observations both in the monitoring points and the number of samples, (iv) random noise which inevitably corrupts measurements. The sources of variation can be categorised as uncertainty or variability. Uncertainty refers to the precision with which a state or parameter is known (e.g., error on an experimental measurement) and variability refers to the natural variation

☆ CPMF^2, Flemish Cluster Predictive Microbiology in Foods - www.cpmf2.be.

* Corresponding author at: Chemical and Biochemical Process Technology and Control (BioTeC), Department of Chemical Engineering, KU Leuven, Gebroeders de Smetstraat 1, B-9000 Ghent, Belgium.

E-mail addresses: simen.akkermans@kuleuven.be (S. Akkermans), philippe.nimmegeers@kuleuven.be (P. Nimmegeers), jan.vanimpe@kuleuven.be (J.F. Van Impe).

https://doi.org/10.1016/j.ijfoodmicro.2018.05.027
Received 22 November 2017; Received in revised form 19 April 2018; Accepted 27 May 2018
Available online 29 May 2018
0168-1605/

of a variable or process (e.g., microbial growth rate).

Due to the inevitable presence of variation in building predictive models, it is generally deemed important to assess the accuracy of the model predictions. This is often simplified to finding the confidence intervals of the parameter estimates. The confidence intervals of the parameter estimates (or simply the variation of the parameter estimates) can lead to the calculation of the uncertainty on the model prediction. As such, the user of a predictive model can be provided with an estimate of, e.g., a 95% confidence interval of the model prediction. The determination of this uncertainty is indispensable when using predictive models for quantitative microbial risk assessments (Zwietering, 2015). As (the uncertainty on) the estimated values of the model parameters are determined by (the uncertainty on) the experimental data, also the calculated uncertainty on the model parameters and model prediction will be determined by the experimental data. Consequently, it is worth wondering how to ensure that the provided uncertainty is actually reliable.

This research studies how a reliable determination of the prediction uncertainty can be obtained. The focus lies on modelling and predicting the growth of microorganisms as a function of temperature, but the results should be transferable to other conditions and to modelling of microbial inactivation as well. However, further research should be performed to confirm the conclusions of this research for other applications. It is worth noting that an accurate determination of the model prediction uncertainty will become more difficult for more complex models (e.g., in case of multiple influencing variables and interactions). Two steps in the modelling procedure are investigated with respect to their influence on determining the model prediction uncertainty: (i) the parameter estimation method and (ii) the uncertainty propagation method. These are deemed most influential on the calculation of the prediction uncertainty. For this purpose, a case study was applied in which a mathematical model was built for the effect of temperature on the microbial growth rate and used to predict microbial growth for a temperature profile that is characterised by variability. This research also is meant to serve as a tutorial to uncertainty propagation techniques for scientists working in the field of (predictive) microbiology.

2. Materials and methods

For the current research, data is simulated according to the protocol explained in Section 2.1. The parameters of the predictive model will be estimated according to the methods explained in Section 2.2. This section also explains the method generally used to determine the model parameter accuracy. Finally, Section 2.3 elaborates on the different methods for uncertainty propagation that are tested in this publication to calculate the model prediction uncertainty.

2.1. Simulation protocol

Experiments are always simulated at the same 8 temperatures (10, 15, 20, 25, 30, 35, 40, 45 °C). At each temperature, the maximum specific microbial growth rate μ_{max} [h^{-1}], which is reached during the exponential phase of growth, is calculated according to the Cardinal Temperature Model with Inflection (CTMI) of Rosso et al. (1993):

$$\mu_{max}(T) = \mu_{opt} \cdot \frac{(T - T_{min})^2 \cdot (T - T_{max})}{(T_{opt} - T_{min}) \cdot [(T_{opt} - T_{min}) \cdot (T - T_{opt}) - (T_{opt} - T_{max}) \cdot (T_{opt} + T_{min} - 2T)]} \quad (1)$$

In this equation, T_{min} [°C] and T_{max} [°C] represent the minimum and maximum temperature that allow microbial growth. T_{opt} [°C] is the optimum temperature at which the optimum growth rate μ_{opt} [h^{-1}] is reached, as such $\mu_{opt} = \mu_{max}(T_{opt})$. The value of μ_{max} [h^{-1}] (at any temperature) is then used to simulate a growth curve using the model of Baranyi and Roberts (1994):

$$\frac{dn(t)}{dt} = \frac{1}{1 + \exp(-q(t))} \cdot \mu_{max}(T) \cdot [1 - \exp(n(t) - n_{max})] \quad (2)$$

$$\frac{dq(t)}{dt} = \mu_{max}(T) \quad (3)$$

with n(t) [ln(CFU/mL)] the natural logarithm of the population density at a time point t [h], n_{max} [ln(CFU/mL)] the natural logarithm of the maximum population density and q(t)[−] the natural logarithm of the physiological state of the cell. The initial values of n(t) and q(t) are respectively n_0 and q_0. Nominal values for T_{min}, T_{opt}, T_{max}, μ_{opt}, n_0, q_0 and n_{max} were chosen arbitrarily for a hypothetical microorganism and are listed in Table 1. Growth curves were simulated until the population density reached a value approximating the nominal n_{max}. In these growth curves, 8 samples were taken at equidistant time points. Gaussian noise with zero mean was added to these samples to simulate the variation of the experimental data. The standard deviation of the Gaussian noise was taken equal to 0.28 ln(CFU/mL) based on the mean squared error of previous (unpublished) parameter estimation results with secondary models for growth. Discrepancy between the model structure and the microbial system under study is not considered in this research. Also the effect of the experimental design was not considered in this research.

Table 1
Nominal parameter values of the CTMI and the model of Baranyi and Roberts (1994).

Parameters	Values
T_{min} [°C]	2.3
T_{opt} [°C]	40.6
T_{max} [°C]	45.5
μ_{opt} [h^{-1}]	0.623
n_0 [ln(CFU/mL)]	7.00
q_0 [−]	−1.00
n_{max} [ln(CFU/mL)]	22.55

The simulations used to compare different methods for assessing the propagation of uncertainty from experimental data to model predictions (Section 3.2) are based on a temperature profile that is characterised by variability as well. An arbitrary temperature profile was selected for these simulations to mimic the food chain of a product that is kept at refrigeration temperatures. The different steps of the temperature profile are listed in Table 2. Fig. 1 illustrates the temperature profile with all parameters at their mean value. The durations of each step was considered to have a uniform distribution. Both the linear approximation and the sigma point method (described in Section 2.3) rely on the mean value and variance for their computations. As such,

Table 2
Five different steps of the temperature profile used to simulate microbial growth as a function of time with prediction uncertainty. Normal distributions are marked with their mean and variance and uniform distributions with their lower and upper bound.

Description	Temperature [°C]	Time [h], uniform distribution	Time [h], approximate normal distribution
Storage after production	$N(6.0, 1.5)$	$U(10.00, 22.00)$	$N(16.00, 3.46^2)$
Transportation to shops	$N(10.0, 1.0)$	$U(0.50, 4.00)$	$N(2.25, 1.01^2)$
Storage in shops	$N(6.0, 1.0)$	$U(1.00, 168.00)$	$N(84.50, 48.21^2)$
Transport to customer's home	$N(20.0, 2.0)$	$U(0.08, 1.00)$	$N(0.54, 0.26^2)$
Storage at home	$N(7.0, 1.0)$	Remaining time of total 240 h	