Food Research International 137 (2020) 109374

Contents lists available at ScienceDirect

Food Research International

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

# Multilevel modelling as a tool to include variability and uncertainty in quantitative microbiology and risk assessment. Thermal inactivation of *Listeria monocytogenes* as proof of concept

Alberto Garre, Marcel H. Zwietering, Heidy M.W. den Besten*

*Food Microbiology, Wageningen University & Research, P.O. Box 17, 6700 AA Wageningen, the Netherlands*

ARTICLE INFO

*Keywords:*
Variance analysis
Hierarchical models
Predictive microbiology
Food safety
Bayesian statistics
Exposure assessment

ABSTRACT

Variability is inherent in biology and also substantial for microbial populations. In the context of food safety risk assessment, it refers to differences in the response of different bacterial strains (between-strain variability) and different cells (within-strain variability) to the same condition (e.g. inactivation treatment). However, its quantification based on empirical observations and its incorporation in predictive models is a challenge for both experimental design and (statistical) analysis.

In this article we propose the use of multilevel models to quantify (different levels of) variability and uncertainty and include them in the predictions. As proof of concept, we analyse the microbial inactivation of *Listeria monocytogenes* to thermal treatments including different levels of variability (between-strain and within-strain) and uncertainty. The relationship between the microbial count and time was expressed using a (non-linear) Weibullian model. Moreover, we defined stochastic hypotheses to describe the different types of variation at the level of the kinetic parameters, as well as in the observations (microbial counts). The model parameters (kinetic parameters and variances) are estimated using Bayesian statistics.

The multilevel approach was compared against an analogous, single-level model. The multilevel methodology shrinks extreme parameter estimates towards the mean according to uncertainty, thus mitigating overfitting. In addition, this approach enables to easily incorporate different levels of variation (between-strain and/or within-strain variability and/or uncertainty) in the predictions. On the other hand, multilevel (Bayesian) models are more complex to define, implement, analyse and communicate than single-level models. Nevertheless, their ability to incorporate different sources of variability in predictions make them very suitable for Quantitative Microbial Risk Assessment.

## 1. Introduction

Mathematical modelling is a basic tool in food science. It is broadly used to predict quality changes of food products (van Boekel, 2009; González-Tejedor et al., 2017) or to describe the microbial response (growth or inactivation) to processing or storage conditions (González-Tejedor, Garre, Esnoz, Artés-Hernández, & Fernández, 2018; Perez-Rodriguez & Valero, 2012). Mathematical models can be applied, for instance, to aid in the establishment of the shelf-life of a food product or in Quantitative Microbial Risk Assessment (QMRA) (Franz et al., 2019; García et al., 2015; Possas, Valdramidis, García-Gimeno, & Pérez-Rodríguez, 2019).

Most of the mathematical models employed in food science are parametric models. Well-known examples of model parameters are the specific growth rate (for microbial growth) or the *D*-value (the heating time required for a ten-fold reduction in microbial concentration). Besides being required for making predictions, many parameters have a biological interpretation (e.g. the *D*-value can be related to stress resistance). This information can be valuable for risk assessment and process design. Examples for microbial inactivation include the comparison of the effectiveness of different treatments or the identification of the most resistant bacterial population to a treatment (Ros-Chumillas, Garre, Maté, Palop, & Periago, 2017; van Asselt & Zwietering, 2006).

Model parameters can depend on a broad range of implicit (e.g. bacterial strain), intrinsic (e.g. type of media) and extrinsic factors (e.g. temperature) in a manner that is not yet entirely understood. Hence, they must be estimated using experimental data and, because experimental error is unavoidable, their values cannot be known with absolute certainty. Moreover, additional variation in the observations is due

* Corresponding author.
*E-mail address:* heidy.denbesten@wur.nl (H.M.W. den Besten).

https://doi.org/10.1016/j.foodres.2020.109374
Received 22 January 2020; Received in revised form 27 May 2020; Accepted 31 May 2020
Available online 02 June 2020

to the pivotal role of biological variability in the microbial response. In the context of microbial food safety, variability includes inherent sources of variation (e.g. differences in the response of single cells or in the composition of the food media). It is, thus, different to uncertainty, which encompasses those sources of variation that are not considered in the system (e.g. measurement errors or model misspecifications). Therefore, whereas uncertainty can be reduced by gathering more experimental data of higher quality, variability is part of the process and cannot be reduced simply by more and better experimentation (Nauta, 2000; Thompson, 2002). Although conceptually there is a clear separation between variability and uncertainty, the quantification of these terms and the incorporation in predictions is a complex problem from the point of view of experimental design and statistical analysis. Consequently, they are not considered in many case studies in food science, where they are poorly quantified and/or have little impact in the outcome. There are other applications, such as Quantitative Microbial Risk Assessment (QMRA), where variability and uncertainty can be equally or more relevant than kinetic and process parameters, so they must be incorporated in the analysis (den Besten, Aryani, Metselaar, & Zwietering, 2017; EFSA Scientific Committee et al., 2018).

The study of biological variability and the quantification of its impact in QMRA has gained interest during the last years (den Besten et al., 2017; Koutsoumanis & Aspridou, 2017). This problem has been tackled mainly from two different directions. Some studies have applied a bottom-up approach, where probabilistic hypothesis (mechanistic or empirical) are proposed to describe individual sources of variability and/or uncertainty. Then, numerical simulations can be applied to quantify the impact of the variability and uncertainty of the parameter estimates in the variation of the variable of interest (usually the microbial count) (Aguirre, Pin, Rodriguez, & Garcia de Fernando, 2009; Aspridou & Koutsoumanis, 2015; Garre, Egea, Esnoz, Palop, & Fernandez, 2019; Koyama et al., 2019a). This approach, however, can only reflect the impact of a limited number of uncertainty and/or variability sources. Furthermore, the definition of realistic hypotheses to describe biological variability (e.g. the distribution of *D*-values) can be challenging.

Other studies have applied a top-down method to analyse variability and uncertainty. This approach applies statistical methods to analyse datasets of bacterial responses and tries to quantify the contribution of different sources of variation (Aryani et al., 2015a; Aryani et al., 2015b; Augustin et al., 2011; den Besten et al., 2017; Guillou & Membré, 2019; Koyama et al., 2019b; Santos, Samapundo, Gülay, Van Impe, & Sant'Ana, Devlieghere, 2018; Wells-Bennik et al., 2019). In this approach, experiments are repeated for the same condition (i.e. bacterial strain, media…). Note that, because it is entirely data-driven, the experimental design defines what sources of variation can be quantified and the precision of the estimates. Once the data has been gathered, the contribution of each source of uncertainty or variability to the total variance is usually analysed by fitting a primary model (growth or inactivation) to each experiment, followed by an analysis of the variance of the parameters of interest. An alternative to gathering experimental data is the use of model parameters already available in the literature to perform a meta-analysis (den Besten and Zwietering, 2012; Liu et al., 2020). The main limitation of a top-down approach is that it requires a significant amount of published experimental work, which rapidly increases with the number of sources of variability/uncertainty considered. Moreover, most previous studies following this approach have applied a post-hoc statistical analysis to quantify variability and/or uncertainty, so errors in model fitting have an impact on the estimation of the different sources of variance. These latter issues could be mitigated following a different modelling approach, where variation is included in the first steps of modelling definition.

Multilevel models are an extension of classical regression where coefficients can vary by group (McElreath, 2016). In the scientific literature, this type of model is also called mixed-effects model (especially in frequentist statistics) or hierarchical models (Gelman & Hill, 2007). We use the term "multilevel" in this article because it is consistently used in scientific works applying a Bayesian approach. This methodology is able to incorporate hypothesis explaining the causes for the variation (variability and uncertainty) of the experimental observations in the model definition. Multilevel models are applied in different scientific fields and, in most cases, provide a better description of the model variance (also variation at different levels), especially for datasets with more than one sampling per individual (technical replicates in this context) or when there are imbalances in the dataset (McElreath, 2016). For these reasons, multilevel models can be of interest for kinetic models relevant for predictive microbiology, where variability is inherent and can be very relevant (van Boekel, 2020). However, their application in this field has been mostly used in meta-analyses or meta-regression models, which aim to estimate model parameter (e.g. the D-value) using published data from studies available in the literature (Cadavez et al., 2017; Gonzales-Barron & Butler, 2011; Guillou & Membré, 2019; Jaloustre, Guillier, Morelli, Noël, & Delignette-Muller, 2012; Liu et al., 2020; Rigaux, Denis, Albert, & Carlin, 2013). Meta-regression models usually apply linear models to describe the relationship between the variable of interest (e.g. the logarithm of the *D*-value) and the explanatory variables (temperature, pH…), simplifying parameter estimation of the multilevel model. However, the relationship between the microbial count and the treatment time is rarely log-linear (Peleg & Cole, 1998). As a result, inactivation models are not linear, so the methods applied for fitting multilevel models linear in the parameters are rarely applicable (Gelman, 2014).

Model fitting of a multilevel model can be done from a frequentist or a Bayesian perspective. Although there are algorithms for model fitting of multilevel models using a frequentist approach (e.g. the *lme4* R package (Bates, Mächler, Bolker, & Walker, 2015)), they are limited to relatively simple models (without strong non-linearity and relatively simple stochastic models for variability). More complex models require the application of a Bayesian approach to model fitting. The Bayesian approach differs from the frequentist approach already in the philosophy of the fitting (Taper and Lele, 2004). The frequentist approach assumes that there is a unique "true" value of the model parameters (e.g. the *D*-value) and that the estimated value may differ from it because of the experimental error. In this context, confidence intervals bound the true value of the model parameter based on asymptotic results (Eichelsbacher et al., 2013). On the other hand, the Bayesian approach does not consider the existence of a unique "true" value for the model parameters. Instead, it considers that they follow an unknown probability distribution, to be approximated based on data. Then, the variable of interest is the conditional probability of the model parameter, $\theta$, given the data available ($P(\theta|data)$). In a Bayesian context, $P(\theta|data)$ is usually called posterior-distribution. According to Bayes rule, ($P(\theta|data)$ is proportional to $P(data|\theta)\cdot P(\theta)$, where $P(data|\theta)$ is the likelihood of the data given the vector of parameters $\theta$ and $P(\theta)$ is the prior distribution for $\theta$. Therefore, Bayesian parameter estimation can be seen as a balance between prior belief and evidence from the data. If there is very strong prior belief before making the analysis, it can only be changed by extraordinary evidence from the data. If prior belief is weak, the data will have a strong contribution to the conditional probability of parameter $\theta$.

In this article, we demonstrate how multilevel models can be used for kinetic modelling in predictive microbiology accounting for different sources of variability and uncertainty. As proof of concept, we use an already published dataset dealing with variability in the inactivation of *Listeria monocytogenes*. The model defines contributors to variation to describe the variability and uncertainty at different levels (the ln *D*-value, the shape parameter, $\beta$, of the Weibull primary model and $\log_{10} N$) and uses a non-linear inactivation model. Therefore, it is an extension with respect to previous models in the field that either used simple error models (e.g. regression models) or were mixed-effect models based on linear relationships.The basic hypotheses of the multilevel modelling approach are presented in Section 2, together with a

detailed description of the model built for this case study. Note that, although the article is limited to one case study, the modelling approach can be used to a broad range of phenomena to be quantified in food science. In Section 3, the parameter estimates of the model are compared against those obtained using an analogous single-level model. This is followed in Section 4 by a discussion of the benefits and limitations of the multilevel modelling approach against a "classical" single-level approach. The article ends in Section 5, which presents the conclusions drawn from this study.

## 2. Materials and methods

### 2.1. Case study analysed: Variability in heat resistance of Listeria monocytogenes

The dataset on microbial inactivation of *Listeria monocytogenes* published by Aryani et al. (2015a) has been used as a case study. The goal of that research was to quantify the impact of different sources of variability on the thermal resistance of *L. monocytogenes* cells. Three different sources were described in their study: between-strain variability, within-strain variability (i.e. biological variability) and the variation among experiments (the uncertainty related to experimental variability). An experimental protocol was devised in their study to quantify each variability source. Briefly, isothermal treatments were performed for twenty different *L. monocytogenes* strains using a water bath set at 55, 60 and 65 °C. Within-strain variability was analysed by doing the experiment for each strain on three different days with freshly prepared cultures. To quantify variation among experiments, each experiment was performed twice in parallel using the same culture. Culture preparation and enumeration of survivors was done using standard microbiological methods. For details on the methodology, the reader is referred to the original article by Aryani et al. (2015a). The microbial counts obtained in that study are plotted in Supp. Fig. 1.

### 2.2. Primary inactivation model

The inactivation model used by Aryani et al. (2015a) and initially proposed by Metselaar, den Besten, Abee, Moezelaar, and Zwietering (2013) was used in this study. This model is a reparameterization of the Mafart isothermal inactivation model (Mafart, Couvert, Gaillard, & Leguerinel, 2002), where it is considered that the microbial resistance of individual cells to a thermal stress follows a Weibull probability distribution. Consequently, the log-microbial count ($\log_{10}N$) at time, $t$, can be calculated as shown in Equation (1), where $N_0$ is the initial microbial count. The inactivation kinetics are described by the parameters $\beta$ and $D_\Delta$. Parameter $\beta$ introduces the non-linearity in the survivor curve ($\beta < 1$ results in curves with upward concavity whereas $\beta > 1$in downward concavity). The inactivation rate is described by $D_\Delta$, which equals the decimal reduction time of an ideal log-linear inactivation curve that intersects the survivor curve at $\Delta$ log-cycles of inactivation. The approach followed here is slightly different to the one applied by Aryani et al. (2015a), who estimated the time to reach six log-reductions ($t_{6\Delta} = 6\hat{A}\cdot D_\Delta$). We have decided to fit $D_\Delta$ instead, to estimate the variability in the parameter. The parameter $\Delta$ was set to 6, as in the study of Aryani et al. (2015a).

$$\log_{10}N(t) = \log_{10}N_0 - \Delta\cdot\left(\frac{t}{\Delta\cdot D_\Delta(T)}\right)^\beta \quad (1)$$

### 2.3. Single-level modelling of microbial inactivation

For the construction of models according to a single-level approach, the model parameters of the reparameterized Weibull model (Equation (1)) were fitted independently to each biological replicate. The data set consists of isothermal treatments at three different temperatures, applied to three biological replicates of each of the 20 strains analysed. Hence, 180 parameters (20 strains × 3 replicates × 3 parameters [$\log N_0$, $D_\Delta$, $\beta$]) were estimated per temperature. Note that this approach differs slightly from the one applied by Aryani et al. (2015a), who fitted the model individually for the two technical replicates (doubling the number of curves). We have not disaggregated by technical replicate to ease comparison with our multilevel model, where we assign a vector of model parameters [$D_\Delta$, $\beta$] for each biological replicate and include a variable for the remaining variation (see Section 2.4).

The single-level model was fitted by non-linear regression with the *nls* function included in the *stats* package of R (R Core Team, 2016) using the Gauss-Newton algorithm (Bates & Watts, 2007). This algorithm minimizes the sum squared error of the fittings with respect to the data. As it is usual in quantitative microbiology, we have used the residuals of the log-microbial count for model fitting (Eq. (1)). In non-linear regression, this is due to the hypothesis that the error of the log-microbial count follows a normal distribution with constant variance $\sigma_e^2$. The expected value of the normal distribution is the one calculated by the primary model (Eq. (1)). These hypotheses can be written as a stochastic model as shown in Eq. (2), where the symbol "~" can be read as "[…] follows a […] distribution" and *Norm* refers to the normal distribution. Note that, because in the single-level approach the model is fitted independently for each biological replicate, a parameter vector ([$\log N_0$, $D_\Delta$, $\beta$]) is obtained by non-linear regression for each experiment (strain/biological replicate/temperature), as well as the variance, $\sigma_e^2$.

$$\log_{10}N \sim Norm\left(\log_{10}N_0 - \Delta\cdot\left(\frac{t}{\Delta\cdot D_\Delta(T)}\right)^\beta, \sigma_e\right) \quad (2)$$

### 2.4. Hypotheses of multilevel models; description of the notation used in the manuscript

The cornerstone of the multilevel modelling approach is the definition of an underlying statistical model to describe the variability of some model parameters. Before defining the multilevel model used for the case study, we will use a simpler example to describe the basic ideas of this approach, as well as to describe the notation used throughout this article. Let $\mathbf{y}$ be an observable quantity, such as the $D$-value in an inactivation experiment (or any transformation, like its natural logarithm in our example). We will write $y_i$ as the value of $\mathbf{y}$ observed for observation $i$ ($i \in 1, 2, \cdots, n$). If we define $\mu$ as the expected value of $\mathbf{y}$ for the complete data set, the difference between observation $y_i$ and $\mu$ can be attributed to several sources of variation, as illustrated in Fig. 1. In this example (and also in the case study), these will be limited to three (Eq. (3)): between-strain variability, within-strain variability and uncertainty.

$$y_i = \mu + \varepsilon_{sj[i]} + \varepsilon_{bioj,k[i]} + \varepsilon_i;$$
$$i \in 1, 2, \cdots, n;\ \ k \in 1, 2, \cdots, n_b;\ \ j \in 1, 2, \cdots, n_s \quad (3)$$

The between-strain variability describes the fact that some strains are more resistant to heat stresses than others. This variation is described by the term $\varepsilon_{\mathbf{sj}}$, which is a vector of length $n_s$ (the number of strains). Each observation, $i$, corresponds unmistakable to one of the strains included in the experimental design. Hence, $\varepsilon_{sj[i]}$ can be read as "of the $n_s$ elements of $\varepsilon_{\mathbf{sj}}$, take the one corresponding to the strain used for observation i". Therefore, this term is constant for every observation for the same strain and the differences between observations of the same strain are explained by the remaining error terms.

In this context, the within-strain variability describes differences between bacterial populations of the same strain. It is described by the coefficient $\varepsilon_{\mathbf{bio_{j,k}}}$, which is a vector of length $n_s{\cdot}n_b$ ($n_b$ stands for the number of biological replicates per strain). Similarly as for $\varepsilon_{sj[i]}$, this parameter can be read as "of the $n_s{\cdot}n_b$ elements of $\varepsilon_{\mathbf{bio_{j,k}}}$, take the one corresponding to the strain and the biological replicate for observation

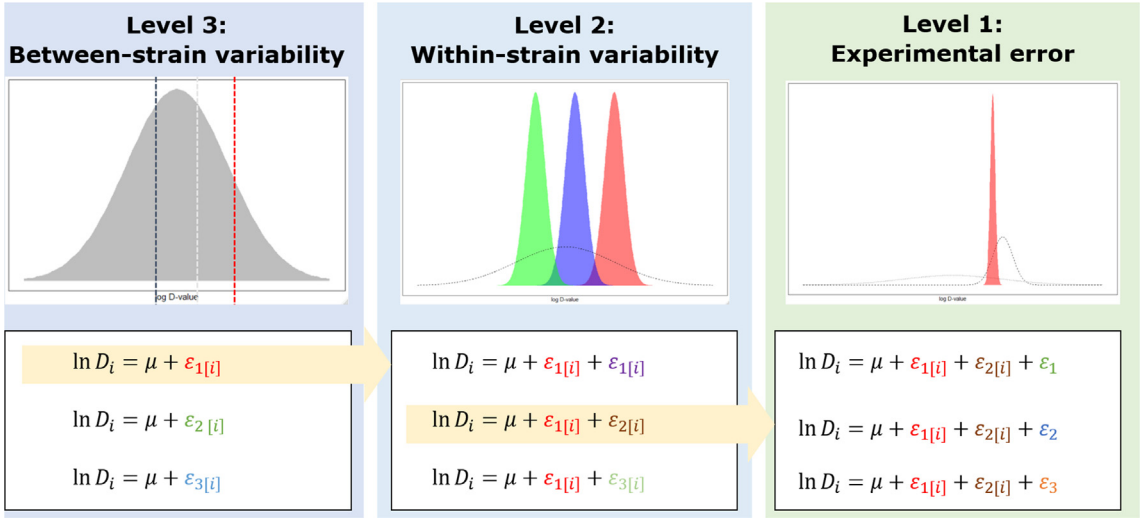


**Fig. 1.** Illustration of the underlying concepts for the multilevel modelling approach using the *D*-value. Briefly, we assume the existence of a theoretical hyper-population of *Listeria* strains with unknown mean and variance (Level 3; between-strain variability). Then, the expected *D*-value for each strain is a random draw from this distribution. There is some variability between the biological replicates quantified by an unknown correlation between the log *D*-values observed for each biological replicate (within-strain variability; Level 2). Finally, the difference between the observed ln *D*-value and the expected one is modelled using another normal distribution with unknown variance (experimental uncertainty; Level 1). Note that this Figure refers to the example described in section 2.4. The model actually used to analyse the data defines the experimental error (Level 1) in the microbial count, not the natural logarithm of the *D*-value (Eq. (4)).

*i*". These two terms describe the natural variation (variability) of the observed variable, *y* (in this example, the natural logarithm of the *D*-value). However, there are other factors that can cause variation in the experimental observations (e.g. experimental uncertainty), so $\varepsilon_{\mathbf{s_j}}$ and$\varepsilon_{\mathbf{bio_{j,k}}}$ are unlikely to describe the variation of observable, *y*. This remaining variation is described by the term $\varepsilon_i$. Note that this term does not include any bracket, indicating that it is not "grouped"; a different value corresponds to each observation. Therefore, $\varepsilon$is a vector of size *n* (the total number of data points).

Several studies have divided the total variation of the observed variables in a similar fashion, although they were not quantified using multilevel models (Aryani, den Besten, Hazeleger, & Zwietering, 2015a; Aryani, den Besten, Hazeleger, & Zwietering, 2015b; den Besten et al., 2017). Instead, they applied a post-hoc variance analysis. The multilevel modelling approach, on the other hand, defines a stochastic model for the distribution of the different error terms ($\varepsilon_{s_j[i]}$, $\varepsilon_{bio_{j,k}[i]}$ and $\varepsilon_i$). Then, the model is fitted to the complete dataset considering these probabilistic hypotheses.

### 2.5. A multilevel model for microbial inactivation including variability and uncertainty

A multilevel model provides a framework to include in the model stochastic hypotheses regarding the sources of variability. Therefore, the hypotheses of the multilevel model are not limited to how the inactivation rate varies (i.e. the Weibullian model in the single-level approach), but also include hypotheses regarding how different sources of variation affect the variation of the model variables. As a basic block for the multilevel model, we have used the primary inactivation model proposed by Metselaar et al. (2013). Then, we have defined stochastic hypotheses for the two sources of variability and the experimental uncertainty considered relevant for this case study. We introduce the first level of variability (the uncertainty, or experimental error) as a normal distribution that describes the difference between the expected microbial count for condition *i* ($\mu_i$) at the specific time point $t_i$ and the observed microbial count $\log_{10}N$ as a normal distribution with mean $\mu_i$ and unknown standard deviation, $\sigma_e$. This hypothesis is equivalent to the one commonly used in (non–)linear regression (Eq. (2)). Note that this model differs slightly from the one used in the example in section 2.4, where the observed variable was the logarithm of the *D*-value (to simplify the notation, the subindex Δ is omitted in the remaining of the manuscript). These hypotheses are written in mathematical form in Eq. (4), where a difference is made between stochastic (symbol "~") and deterministic (symbol "←") relationships. In this equation, the parameters *D* and *β* have been ln-transformed to improve the convergence of the Monte Carlo sampler (described in section 2.6) (Gelman, 2014). Note that, because of the use of a Monte Carlo sampler, this transformation only improves the convergence of the algorithm and does not influence inference. This equation uses the same notation as the one used in the previous equations. The subindex *i* refers to a single observation, *j* refers to the strain and *k* to the biological replicate. When a coefficient does not vary between replicates, the subindex *i* is written within brackets (e.g. $\beta_{j,k[i]}$ varies between strain and biological replicate and $N_{0,j[i]}$ only between strains).

$$\log_{10}N \sim Norm(\mu, \sigma_e)$$

$$\mu_i \leftarrow \log_{10}N_{0,j[i]} - \Delta\left(\frac{t_i}{\Delta e^{\ln D_{j,k[i]}}}\right)^{e^{\ln\beta_{j,k[i]}}} \quad (4)$$

In the dataset analysed, there are three biological replicates per strain. In the multilevel approach, we assign one *D*-value and one value of *β* to each of the biological replicates. In this sense, it does not differ from the single-level approach. However, whereas in the single-level approach the model parameters calculated for different conditions are independent, the multilevel approach enables the definition of stochastic relationships (e.g. an error model; Fig. 1) between them.

The natural logarithm of the *D*-value (ln*D*) calculated for each biological replicate can be written as a 3x20 matrix (3 biological replicates and 20 strains), $\boldsymbol{A_{k,j}}$, where component $a_{k,j}$ corresponds to the ln*D* calculated for biological replicate *k* and strain *j*. In order to illustrate the assumptions behind the modelling of the within-strain variability, let us assume that the ln*D* for the first biological replicate for every strain is known, i.e. the first row of $\boldsymbol{A_{k,j}}$. If we repeated every inactivation experiment, we would expect that the ln *D*-values calculated for a second biological replicate would be "similar" to the ones obtained for the first replicate. For instance, the strain with the highest ln*D* should be among the most resistant ones. This can be expressed as the existence of a correlation ($\rho_{a,12}$) between rows $\mathbf{a}_{1,j}$ and $\mathbf{a}_{2,j}$. The same logic can be applied between any arbitrary pair of rows of$\mathbf{A_{k,j}}$. The

correlation $\boldsymbol{\rho_a} = \begin{pmatrix} 1 & \rho_{12}\rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13}\rho_{23} & 1 \end{pmatrix}$ is, then, a symmetric, square matrix with the same number of rows as the matrix $\mathbf{A_{k,j}}$ (i.e. the number of biological replicates).

The between-strain variability has been modelled considering that the natural logarithm of the *D*-values of every *Listeria* strain follows a normal distribution with mean ($\mu_a$) and standard deviation ($\sigma_a$) (i.e. between-strain variability), both unknown. Then, the *D*-values of the observed strains are random draws from that distribution (Fig. 1). The selection of a normal distribution can be justified using several arguments. First, as an empirical argument, we inspected density plots of the *D*-values estimated using the single-level model, identifying that the ln-transformation resulted in a distribution resembling a normal distribution. As a more theoretical argument, the normal distribution has maximum entropy among all real-valued distributions supported on $(-\infty, +\infty)$ with a specified variance. Therefore, from the point of view of information theory, it is the modelling approach with the smallest influence on the predicted posterior distributions.

The hypotheses defined to describe the within-strain and between-strain variability can be modelled using a multivariate normal distribution (*MVN*), as shown in Eq. (5). Note that the correlation matrix $\boldsymbol{\rho_a} = \begin{pmatrix} 1 & \rho_{12}\rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13}\rho_{23} & 1 \end{pmatrix}$ has three unknown parameters ($\rho_{12}$, $\rho_{13}$, $\rho_{23}$). A similar modelling approach was followed for $\beta$, introducing the vector of standard deviations $\sigma_b$ and the correlation matrix $\boldsymbol{\rho_b}$ also with three unknowns. Hence, the variances $\sigma_a^2$ and $\sigma_b^2$ are unknown parameters to be estimated from the data that quantify the between-strain variability. In a similar way, the correlation matrixes $\boldsymbol{\rho_a}$ and $\boldsymbol{\rho_b}$ are fitted to the data to describe the within-strain variability.

$$\begin{pmatrix} a_{1,1} & \cdots & a_{20,1} \\ a_{1,2} & \cdots & a_{20,2} \\ a_{1,3} & \cdots & a_{20,3} \end{pmatrix} = \begin{pmatrix} \mathbf{a_{s,1}} \\ \mathbf{a_{s,2}} \\ \mathbf{a_{s,3}} \end{pmatrix} \sim MVN\left( \begin{pmatrix} \mu_{a1} \\ \mu_{a2} \\ \mu_{a3} \end{pmatrix}, \begin{pmatrix} \sigma_{a1} \\ \sigma_{a2} \\ \sigma_{a3} \end{pmatrix}, \boldsymbol{\rho_a} \right)$$

$$\begin{pmatrix} b_{1,1} & \cdots & b_{20,1} \\ b_{1,2} & \cdots & b_{20,2} \\ b_{1,3} & \cdots & b_{20,3} \end{pmatrix} = \begin{pmatrix} \mathbf{b_{s,1}} \\ \mathbf{b_{s,2}} \\ \mathbf{b_{s,3}} \end{pmatrix} \sim MVN\left( \begin{pmatrix} \mu_{b1} \\ \mu_{b2} \\ \mu_{b3} \end{pmatrix}, \begin{pmatrix} \sigma_{b1} \\ \sigma_{b2} \\ \sigma_{b3} \end{pmatrix}, \boldsymbol{\rho_b} \right) \quad (5)$$

The deterministic primary model (Eq. (4)) and the stochastic model describing the variation in ln *D* and $\beta$ (Eq. (5)) have been combined using dummy variables. The value of $\ln D_i$ and $\ln \beta_i$ in Eq. (4) can be calculated as shown in Eq. (6), where $x_{bk[i]}$ are dummy variables that take the value 1 if the sample corresponds to biological replicate *k* and 0 otherwise. The parameters $a_{j,k}$ and $b_{j,k}$ describe the values of ln *D* and ln $\beta$ estimated for biological replicate *k*, as defined in Eq. (5). For instance, if observation *i* corresponds to biological replicate 1, variables $x_{b2}$ and $x_{b3}$ both equal zero, and $x_{b1} = 1$. Hence, the first line of Eq. (6) reduces to $\ln D_i \leftarrow a_{j,1}$.

$$\ln D_i \leftarrow a_{j,1}x_{b1[i]} + a_{j,2}x_{b2[i]} + a_{j,3}x_{b3[i]}$$

$$\ln \beta_i \leftarrow b_{j,1[i]}x_{b1[i]} + b_{j,2[i]}x_{b2[i]} + b_{3,1[i]}x_{b3[i]} \quad (6)$$

### 2.6. Bayesian parameter estimation

Although, some multilevel models can be fitted using a frequentist approach as a mixed-effects model (Gelman & Hill, 2007), a Bayesian scheme has been used due to the model complexity. This approach requires the definition of priors for every unknown model parameter. Then, the posterior distribution of the model parameters is approximated balancing evidence (likelihood) and prior distribution. For every standard deviation, an exponential prior with expected value $\lambda = 1$ has been used. We have chosen a value larger than expected based on available experimental data to ease the convergence of the Monte Carlo sampler (McElreath, 2016). The analysis was repeated using different values for $\lambda$ and also a uniform prior, without observing any impact on the parameter estimates (results not shown). The mean of the multivariate normal prior for the logarithm of the *D*-value was selected based on the scatter plot in Supp. Fig. 1 and set to 2 ln-min (7.4 min) for the data at 55 °C, −1 ln-min (0.37 min) at 60 °C and −4 ln-min (1.1 s) at 65 °C and to 0 for the logarithm of $\beta$ for every temperature (i.e. $\beta = 1$). For the correlation matrixes, an LKJ prior (the name of the distribution is an acronym for the authors of the papers where it was first defined (Lewandowski, Kurowicka, & Joe, 2009)) with parameter $\tau = 2$ has been selected. The family of LKJ distributions are an alternative to the Inverse-Wishart distribution with better properties for multilevel models. Note that covariance matrices must fulfil several requirements (e.g. positive definite and symmetrical) that restrict the matrices that can be a correlation matrix, so generating valid priors for them is not a trivial problem. A complete discussion of this family of distributions is out of the scope of this article, so we refer the interested reader to the original article by Lewandowski et al. (2009). We have repeated the analysis for different parameters for the prior distributions (within a reasonable range), without observing a significant impact on the results.

The parameters of the multilevel model have been estimated using *Stan* (Carpenter et al., 2017), through the interface provided in the *rethinking* R package (McElreath, 2016). *Stan* is a programming language specially developed for statistical inference. For model fitting, it uses the no-U-turn sampler (NUTS), a type of Hamiltonian Monte Carlo (HMC) sampler (Hoffman & Gelman, 2011). This sampling scheme that is usually more efficient than the one used by Metropolis or Gibbs samplers (Duane, Kennedy, Pendleton, & Roweth, 1987).

The convergence of the Markov chain was assessed according to typical guidelines (Brooks, 2011; McElreath, 2016). Trace and pair plots were visually inspected to ensure appropriate mixing and convergence of the chain. Furthermore, the number of iterations was increased until the parameter $\widehat{R}$ was lower than 1.1, following usual guidelines for Bayesian modelling (McElreath, 2016). Accordingly, 4000 iterations after 1000 warmup iterations were required to fulfil these requirements.

Note that the model has been constructed with a similar scope as the study of Aryani et al. (2015a). Different multilevel models could be defined to analyse the same dataset. The R code implemented to define, fit and post-process the models is openly available in the GitHub page of one of the co-authors (https://github.com/albgarre/multilevel_inactivation).

### 2.7. Modelling of the initial microbial count

Regarding the initial microbial count, $\log_{10}N_0$, two competing hypothesis can be made, both of them reasonable from a biological point of view. On the one hand, considering that the method to prepare the strains is the same for every experiment, it seems reasonable that the initial count of the 20 strains is the same in every experiment. On the other hand, due to between-strain variability, it is possible that the procedure resulted in insignificant differences in the initial count compared to the experimental error ($\sigma_e$). Consequently, both models have been constructed and fitted to the data. Then, their goodness of fit has been compared as described in section 2.8 below. This exercise illustrates the possibility to compare different modelling approaches using multilevel models.

The first modelling approach does not consider strain variability in the initial microbial count, assuming that it is the same for all strains ($\log_{10}N_0$). Hence, in this model, $\log_{10}N_0$ is a scalar (i.e. a vector of length one). A normal prior with mean eight and standard deviation one has been used as shown in Eq. (7). A relatively large value for the standard deviation has been chosen to reduce the impact of the prior distribution on the posterior and to improve the convergence of the HMC sampler.

$$\log_{10}N_0 \sim Norm(8, 1) \quad (7)$$

The second modelling approach considers that the initial count is

the same for every experiment of each individual strain; i.e. there is between-strain variability, but no within-strain variability. Therefore, the parameter $\log_{10}N_{0,s[i]}$ is a vector of length $n_s$ (20 in the case study). We have hypothesized that each element of this vector is a random draw from a normal distribution with unknown mean and variance (Eq. (8)). For the prior distribution, we have used a normal prior with mean 8 log CFU/ml (taken by visually inspecting Supp. Fig. 1) and unknown standard deviation $\sigma_{\log N_0}$, for which an exponential prior with expected value one ($Exponential(1)$) has been used.

$$\log_{10}N_{0,s} \sim Norm(8, \sigma_{logN_0})$$

$$\sigma_{\log N_0} \sim Exponential(1) \quad (8)$$

### 2.8. Model comparison

The Akaike Information Criterion (Akaike, 1974) is commonly used for model selection and comparison. However, several hypotheses that lead to the definition of the *AIC* are violated for multilevel Bayesian models, so the use of this index should be avoided in this context (Gelman, Hwang, & Vehtari, 2014). Consequently, models were compared according to the Widely Applicable Information Criterion (*WAIC*). This index, which has a similar interpretation as the Akaike Information Criterion (models with lower *WAIC* are favoured), is defined as $WAIC = -2(lppd - pWAIC)$. The term *lppd* is the log-pointwise predictive density (with a similar interpretation to the log-likelihood). The coefficient *pWAIC* is the effective number of parameters. This parameter is introduced because for non-linear models the number of degrees of freedom should not be calculated as *n-p*, where *n* is the number of data points and *p* the number of model parameters. A typical example of this is lasso regression (James et al., 2013). For further details regarding the calculation of these indexes, please see the article by Gelman et al. (2014).

### 2.9. Analysis of the posterior distributions and calculation of predictions

The HMC algorithm provides samples of the posterior distributions of the model parameters conditional to the data. In the case studied, (non-linear) operations on the posterior distributions are required for making predictions. Similar transformations are also required for comparison with the parameters estimated using the single-level approach (the *D*-value was ln-transformed in the multilevel, but not on the single-level approach). In the frequentist approach, the calculation of standard errors for the model parameters after the application of non-linear transformations can be quite complex. That is not the case for the HMC algorithm, where a (non-)linear transformations of the samples from the posterior distribution follow the distribution of the transformed variable (Gamerman & Lopes, 2006). This means that, for instance, to calculate the standard deviation of the logarithm of a parameter, we can calculate the logarithm of every element of the sample. Then, the standard deviation of the transformed values equals the one of the logarithm of the random variable. This has been applied to compare the *D*-values and $\beta$s estimated using the multilevel approach against those estimated using the single-level approach. Then, credible intervals have been calculated using the appropriate quantiles of the (transformed) random samples.

A similar approach can be applied to predict the microbial count considering different levels of variability. According to Eq. (4), $\log_{10}N$ follows a normal distribution with expected value $\mu$ and variance $\sigma_e^2$, both random variables estimated from the data. This error term (estimated from the data) describes the variation in this variable that is not explained by the primary model and the sources of variability considered in upper levels of the model (within-strain and between-strain variability). The variation due to these sources of variability is reflected in the posterior distribution of $\mu$, which is a non-linear combination (the Weibull model in Eq. (4)) of the posterior distributions of the matrices with elements $a_{j,k}$ and $b_{j,k}$ (Eq. (6)) considering the link function (Eq. (6)). The HMC algorithm provides a sample of the elements $a_{j,k}$ and $b_{j,k}$ that can be used to obtain a vector of values that simulates the posterior distribution of $\mu$. This vector includes the contribution of the within-strain and between-strain variabilities. This vector of values, together with the estimates of $\sigma_e$ (uncertainty), define a normal distribution that can be used to predict the microbial count accounting for the three sources of variation. Because the parameter $\sigma_e$ only accounts for uncertainty, the posterior of the expected value of $\log_{10}N$, $\mu$, can be used to predict the microbial count omitting the variation due to the uncertainty. This is equivalent to setting $\sigma_e = 0$. Furthermore, simulations can be made for a particular strain, *S*, (disregarding between-strain variability) by taking only the elements $a_{S,k}$ and $b_{S,k}$ corresponding to that strain.

## 3. Results

### 3.1. Comparison of model hypotheses regarding the initial microbial count

The mixed-effects model was fitted independently for the data obtained at each treatment temperature (55, 60 and 65 °C). The fitting was performed twice, once using random initial counts ($N_0$ varies between strains) and using fixed initial count ($N_0$ is constant) as described in the materials and methods section. The complete table of parameter estimates is provided as supplementary material. Supp. Table 1 reports the values of the *WAIC* calculated for each model and the *pWAIC*. Statistical indices used to evaluate the quality of the fit are also random variables, which implies that they have an associated uncertainty (Murdoch, Tsai, & Adcock, 2008). This includes indexes based on information theory, such as the *AIC* or the *WAIC*. However, for most of these indexes the calculation of their uncertainty has high complexity, requiring the application of computationally expensive resampling techniques. The *WAIC* has the advantage with respect to other indexes that it is calculated based on the posterior distribution of the likelihood of the observations, enabling the estimation of a standard error for its estimate. This value can be used to make inference between *WAIC* calculated for different conditions under the assumption that the *WAIC* follows a normal distribution. Column *se(WAIC)* of Supp. Table 1 reports the standard error of the *WAIC* estimated for each model. The results show that the hypothesis used to model the distribution of $N_0$ had little impact on the *WAIC* for the three temperatures tested. For the data at 55 °C the *WAIC* was lower for the model with random initial count, whereas the opposite happened for the other two temperatures. Nevertheless, in every case analysed, the differences between the *WAIC*s are lower than their standard errors. Consequently, there is high uncertainty on what model has better predictive capabilities. In this case, the only hypothesis that varies between models is the variability of the initial count. The posterior distributions of the remaining model parameters (variances, *D*-values and $\beta$s) obtained for each modelling approach were compared, without observing major differences (data not shown). Moreover, very low values have been estimated for the variability of the microbial log count ($\sigma_{\log N_0}$; defined in Eq. (8)) (mean of the posterior of 0.19 log CFU/ml at 55 °C, 0.06 at 60 °C and 0.10 at 65 °C). Hence, for the case studied, the variability of the initial count has little practical impact on the results. For the sake of simplicity, the initial count will be considered constant among all the strains for the remaining of the analysis (but having an uncertainty).

Another point worth mentioning is related to the number of effective parameters of each model. The single-level modelling approach fits one *D*-value, one $\beta$ and one initial count for each biological replicate. This results in 180 model parameters in total (20 strains × 3 replicates × 3 parameters). Then, for making post-hoc inference on the model, we would calculate the number of degrees of freedom of the model as $n - 180$, where $n$ is the number of observations (or some variation thereof depending on the type of inference); i.e. 789, 594 and 694 degrees of freedom for the models fitted at 55, 60 and 65 °C,

respectively. In the multilevel model we have defined additional model parameters to describe variability. That is, parameter $\sigma_a$, for the between-strain variability of ln $D$; $\sigma_b$, for the between-strain variability of ln $\beta$; $\rho_a$, for the within-strain variability of ln $D$; and $\rho_b$ for the within-strain variability of ln $\beta$. Then, the uncertainty is described by $\sigma_e$, that describes the experimental error (error in log $N$). Moreover, we estimate the uncertainty associated with these model parameters. Hence, in the multilevel model we have used more model parameters than in the single-level approach. However, the number of degrees of freedom is not given by $n$ minus the number of parameters. The multilevel model introduces a constraint between the parameter estimates (namely, the multi-variate normal distribution for the parameters) that may increase the degrees of freedom of the model (McElreath, 2016). For instance, the $D$-value estimated for some experiment is expected to be "similar" to the one obtained for the other biological replicates of the same strain (Fig. 1). As a result, the models have 115.5, 107.3 and 109.3 effective parameters at 55, 60 and 65 °C, respectively (Supp. Table 1). This results in 853.5, 666.7 and 764.7 degrees of freedom, respectively; i.e. the multilevel models has more degrees of freedom than the single level model (789, 594, 694) despite having more parameters. Consequently, it is important that for this type of models the calculation of degrees of freedom is based on effective number of parameters, not their raw count. This fact emphasizes that model comparison should be based on appropriate statistical indexes. For multilevel (Bayesian) models, model comparison should be performed only using statistical indexes able to estimate the number of effective parameters (e.g. the *WAIC*).

### 3.2. Comparison of parameter estimates between the single-level and multilevel models

The values and standard deviations of the posteriors of $D_\Delta$ and $\beta$ are provided as supplementary material to this article. As expected, the $D$-value averaged over all the strains tested was strongly affected by the temperature of the treatment (15.4 min at 55 °C, 0.17 min at 60 °C, 0.021 min at 65 °C; Supp. Table 2). The standard deviation of the log $D$-values at the between-strain level (0.13 log min at 55 °C, 0.21 log min at 60 °C, 0.21 log min at 65 °C) was much higher than the average standard deviation at the within-strain level (0.02 log min at 55 °C, 0.04 log min at 60 °C, 0.04 log min at 65 °C; Supp. Table 2). The between-strain variability is also much higher than the uncertainty in the estimation of the $D$-value (average standard deviation of 0.01 log min at 55 °C, 0.02 log min at 60 °C, 0.02 log min at 65 °C; Supp. Table 2). Regarding $\beta$, at 55 °C almost log-linear curves are observed (averaged $\beta$ of 1.02; Supp. Table 3). Increasing the treatment temperature results in a reduction of $\beta$ (0.85 at 60 °C, 0.78 at 65 °C; Supp. Table 3), indicating the presence of tail effects. As well as for the $D$-value, between-strain variability is the most important contributor to the total variation of $\beta$ (0.18 at 55 °C, 0.28 at 60 °C, 0.18 at 65 °C; Supp. Table 3), higher than the one of within-strain variability (0.01 at 55 °C, 0.01 at 60 °C, 0.01 at 65 °C; Supp. Table 3) and uncertainty (0.07 at 55 °C, 0.08 at 60 °C, 0.08 at 65 °C; Supp. Table 3). These results agree with those reported by Aryani et al. (2015a) for the same dataset. Therefore, between-strain variability is the most important contributor to variation for the response of *Listeria monocytogenes* to thermal treatments. As an illustration of the between strain variability, strain LO28 has the lowest $D$-value at 65 °C (0.60 s), more than six times smaller than the one of the most resistant strain (L6, 3.89 s). Therefore, according to the model predictions, a treatment at 65 °C able to reduce the microbial count of a population of strain LO28 in 6 log-cycles, would not even cause a ten-fold reduction of a population of strain L6. This result shows the importance of quantifying the impact of different sources of variability and uncertainty and to include them in risk assessments (den Besten, Wells-Bennik, & Zwietering, 2018).

Figs. 2 and 3 compare the model parameters estimated for each biological replicate using the multilevel and single-level models, as well as their associated uncertainties. Note that both modelling approaches make different hypotheses regarding the distribution of the model parameters. Consequently, confidence intervals/regions are depicted at the 68% confidence level ( ± 1 standard error for the single-level) to show the central part of the distribution, avoiding the complexity introduced by differences in the tail of the distributions. The results of both approaches are comparable and the overall conclusions of the analysis do not vary between methods. For instance, strain L6 is the most heat resistant at the temperatures tested and between-strain variability is higher than within-strain. Also, both models estimate the value of $\beta$ with more uncertainty than the value of $D_\Delta$. However, there are significant differences in the parameter estimated for some of the biological replicates. The most distinctive feature of the multilevel model is that extreme values are rarer than for the single-level model. This effect can be readily visualized in Fig. 4, where the solid points illustrate the parameters estimated by the single-level model and the tip of the arrow the estimates of the multilevel model. Hence, the longer the arrow, the bigger the difference between both estimates. Most of the arrows point towards the mean of all the observations (confidence ellipsoids illustrated with dashed lines). This effect is a well-known, positive feature of Bayesian models with regularizing priors, usually denominated "shrinkage" (McElreath, 2016). Due to the multilevel nature of the model, shrinkage also takes place at a within-strain level; parameter estimates for each biological repetition are "shrunk" towards the mean of the three repetitions. This "secondary" shrinkage can be visualized, for instance, for the results obtained for strain L6 at 65 °C (highlighted using a red square with dotted lines). The amount of shrinkage depends on several factors, including the relative uncertainty of the parameter estimate with respect to the other biological replicates, its relative uncertainty with respect to the other strains, the magnitude of the differences between the parameter estimates, and the impact that changes in the parameter values have on the likelihood of the observations. For a linear model, the shrinkage can be calculated (under some simplifying assumptions) as a weighted average according to the standard deviations of the parameter estimates and the number of data points (Gelman & Hill, 2007). However, the calculations for non-linear models (such as the one used for this case study) are more complex and can rarely be calculated analytically. Nevertheless, the lack of an analytical solution for more complex model does not affect the basis concepts that define the direction and the magnitude of the shrinkage.

In order to compare in further detail the single-level and multilevel approaches, we have analysed the results obtained for strain FBR12 at 55 °C, because of the large differences between the parameter estimates obtained using either approach. The single-level model estimates for parameter $D_\Delta$ (at 55 °C), 12.5, 13.0 and 13.5 min for each biological replicate, and 0.95, 0.87 and 2.12 for parameter $\beta$. The multilevel model, on the other hand, estimates parameter values with lower variation. For parameter $D_\Delta$ it estimates 11.7, 11.9 and 12.1 min for each biological replicate, whereas 0.83, 0.80 and 1.22 are estimated for parameter $\beta$. Therefore, the one-level model predicts that the survivor curve has a slight upwards curvature for two of the biological replicates and a strong downwards curvature for the other one. Such a significant qualitative change in the microbial response is hard to justify from a biological point of view, as the curvature of the survivor curve is usually explained by cell-to-cell variations in the stress resistance (Mafart et al., 2002). The differences between models are further illustrated in Fig. 5, where the survivor curves fitted by the two models (Fig. 5A for the one-level, 5B for the multilevel) are depicted together with the microbial counts. The plot shows that for one of the biological replicates (the one depicted with solid squares), a higher count was observed at 30 min than at 25 min. An increase in the microbial count during an inactivation treatment like the one shown in Fig. 5 is likely associated with an experimental error. The single-level model only uses the information available for one biological replicate and finds the curve that better fits the data. Consequently, it is unable to consider the data about the other repetitions and fits a curve with a strong downwards curvature, very different to the one fitted for the other two

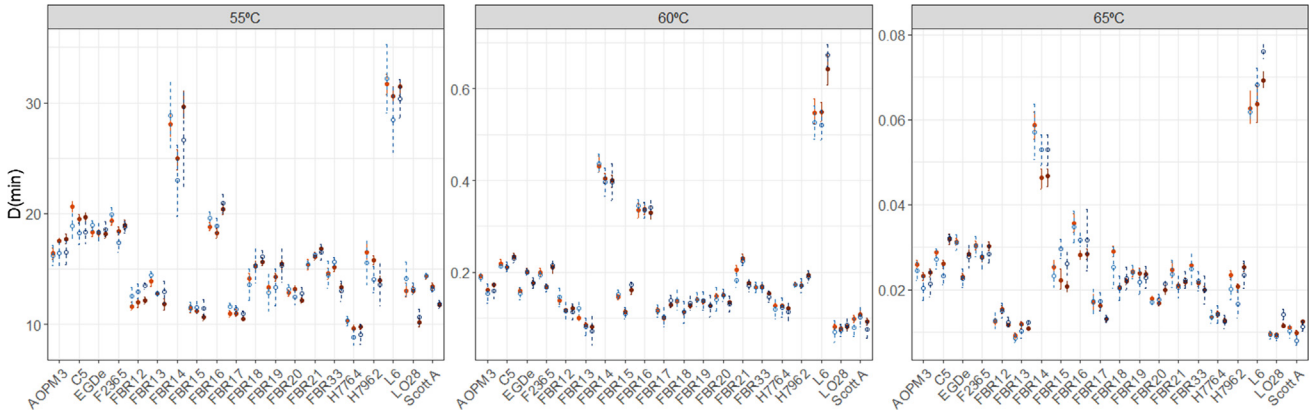


**Fig. 2.** Comparison between the values of $D_\Delta$ estimated using the single-level (blue, open circles) and multilevel (orange, closed circles) at three different temperatures. The error bars (dashed for the multilevel, solid for the single-level) represent the standard error for the frequentist approach and 68% credible intervals for the Bayesian approach. Each point correspond to a biological replicate (two technical replicates) as indicated by the different shades of orange/blue.

repetitions. On the other hand, the multilevel model uses all the data available to fit the model (the one corresponding to other biological replicates of the same strain, as well as the data for other strains). The model, then, "distrusts" extreme values, such as the value of $\beta$ that differs from the one of the other two replicates by one unit. It introduces a correction in the model parameters, pushing them towards the mean (Fig. 4). In the case analysed here, two regularizing hypotheses have been defined, one for the between-strain variability and another one for the within-strain variability. The parameters are, thus, "shrunk" towards the mean of the three biological replicates and the grand mean of all the strains with a magnitude that depends on the uncertainty of the parameter estimates. Therefore, the multilevel modelling approach mitigates the risk of generating spurious parameter estimates due to model overfitting.

### 3.3. Quantification of variability and uncertainty: Impact on predictions

Single-level models associate a standard deviation to the model parameters that quantifies their total uncertainty. However, in order to separate the different sources of variation (biological variability, experimental uncertainty…), post-hoc statistical methods are required (Aryani et al., 2015a; den Besten et al., 2017). The multilevel model developed in this study explicitly models the between-strains and within-strain variability of the stress response, defining specific parameters for their variance (standard errors and correlations; Eqs (4)–(8)). The estimates of these parameters (or the distributions they define) can be used to quantify and compare different sources of variability. For instance, the parameter $\sigma_e$ explains variations between the observed microbial count and the one predicted that cannot be explained based on the within-strain and between-strain variability (Eq. (4)). It includes errors in the sampling (Poisson/Binomial error due to dilution and plating), errors taking the time points and other sources of experimental error, as well as model misspecifications. In our experiments, we observed that the estimate of this error depended on the treatment temperature. At 55 ˚C, it took a value of 0.45 ± 0.01 log CFU/ml, slightly lower than the one observed at 60 ˚C (0.59 ± 0.02 log CFU/ml) and at 65 ˚C (0.61 ± 0.02 log CFU/ml). This can be justified because at higher temperatures inactivation is faster and errors in the sampling time (and other experimental setting) have a higher influence on the observed microbial count. Nevertheless, this conclusion is based on a single dataset (the one used in this investigation) and should be verified using additional data.

In the multilevel model, we have explicitly defined a stochastic model to describe the within-strain and between-strain variability. Their contribution to the total variation is quantified by the parameters $\sigma_a$, $\sigma_b$, $\rho_a$ and $\rho_b$, as described in Eq. (5). However, it is hard to draw any conclusion by simply analysing the values of these parameters (posterior means: $\sigma_a = 0.34$lnmin, $\sigma_b = 0.26$, $\rho_a = 0.97$, $\rho_b = 0.67$ at 55 ˚C, $\sigma_a = 0.77$lnmin, $\sigma_b = 0.43$, $\rho_a = 0.97$, $\rho_b = 0.76$ at 60 ˚C, $\sigma_a = 0.49$lnmin, $\sigma_b = 0.43$, $\rho_a = 0.90$, $\rho_b = 0.70$ at 65 ˚C). Instead, Eqs. (5) and (6) define a probability distribution for the $D$-value and $\beta$. In most cases, an analysis of these distributions is more insightful than an

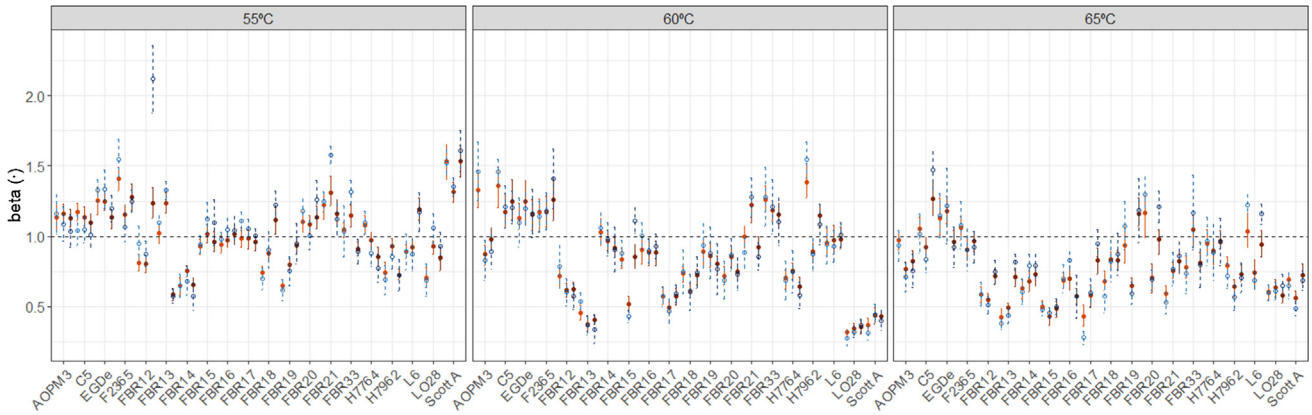


**Fig. 3.** Comparison between the values of $\beta$ estimated using the single-level (blue, open circles) and multilevel (orange, closed circles) at three different temperatures. The error bars (dashed for the multilevel, solid for the single-level)) represent the standard error for the frequentist approach and 68% credible intervals for the Bayesian approach. The horizontal, dashed line marks the point where $\beta = 1$. Each point correspond to a biological replicate (two technical replicates) as indicated by the different shades of orange/blue.

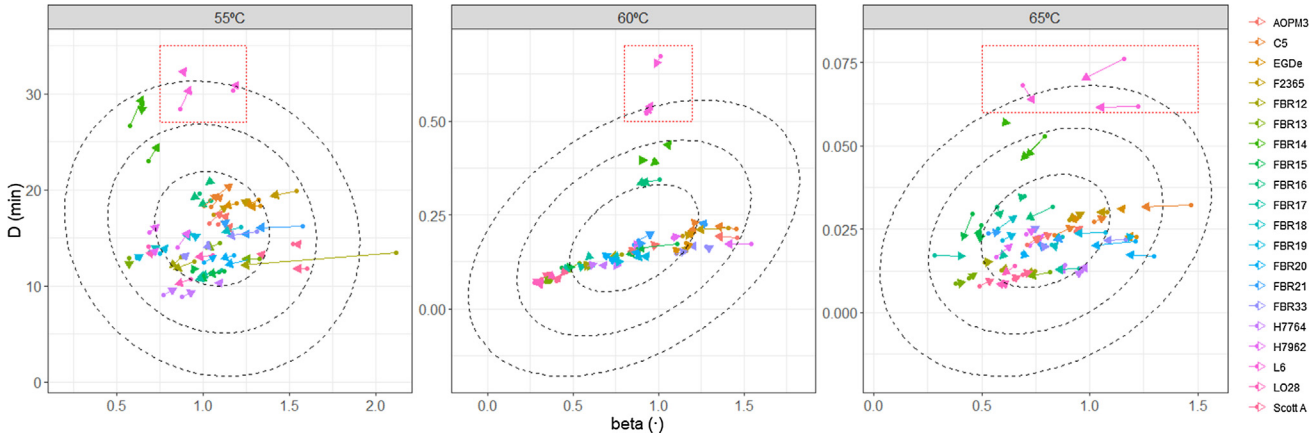


**Fig. 4.** Shrinkage of the parameter values due to the application of a Bayesian model with regularizing priors at three different temperatures. The solid dots represent the estimates of the single-level model and the tip of the arrows the estimates of the Bayesian model. The dashed line depict the (multi-variate normal) confidence ellipsoids according to the single-level model with 50, 90 and 99% confidence level. The position of the parameters for strain L6 are highlighted using a red square with dotted lines.

analysis of the model parameters that generate them. Fig. 6A plots the probability density function of the decimal logarithm of the *D*-value of the hyper-population at 55 °C. The mode of these distributions represents the expected value for each parameter, whereas its "width" describe the within-strain variability. This distribution is drawn based on the values of $\sigma_a$ that have been estimated from the experimental data. Hence, as every model parameter, $\sigma_a$ cannot be known exactly and has an uncertainty associated with its estimated value. As a result, the probability density function has an uncertainty associated. It is shown in Fig. 6A as a shaded area, which represents a credible interval (90% confidence) for this probability density function. Note that the uncertainty illustrated in this plot is the one associated with the estimation of the *D*-value, so it is different to the experimental uncertainty ($\sigma_e$). A similar approach has been followed to estimate the between-strain variability in $\beta$, reported in Fig. 6B. It is evident from these two plots that the uncertainty associated with $\beta$ is higher than the one associated with the *D*-value. This is reasonable for the dataset analysed, as the parameter $\beta$ is estimated with higher uncertainty than the *D*-value (Figs. 2 and 3).

The model definition in the single-level approach is only based on hypotheses regarding the model variables (a kinetic model in the case studied). The multilevel approach also includes hypotheses for the variation (error model), enabling to distinguish between different sources of variability. Hence, it provides a good framework to make predictions including different sources of variation. This is especially interesting for QMRA, where variability is inherent to the system studied (e.g. variability in stress response) and has to be carefully accounted for in the analysis (Nauta, 2000). Fig. 6C depicts prediction bands (90% confidence) for the inactivation of three specific strains at 55 °C. For each of the strains, these intervals are based on Monte Carlo simulations calculated using the posterior distribution of the *D*-value and $\beta$ corresponding to the selected strains only (as described in Section 2.9). Hence, these predictions include the contribution of the within-strain variability (quantified by $\rho_a$ and $\rho_b$). These predictions also include the uncertainty associated with $\rho_a$ and $\rho_b$. Therefore, the predictions include the uncertainty associated with the within-strain variability, but not the experimental error. The reason for that is that the focus of QMRA is generally the response of the microbial population to the processing or storage conditions. In that case, the uncertainty associated with the estimates of the *D*-value and $\beta$ are of interest, but not the uncertainty associated with the experimental error that largely depends on the laboratory conditions. The multilevel model structure proposed in this article defines a variance term for the error of the microbial count ($\sigma_e$) different from the one used to describe the variation of the kinetic parameters of the model. Therefore, it is relatively simple to make predictions of the microbial count considering different sources of variability by selecting what variance components to include in the Monte Carlo simulations.

The posterior distributions of the model parameters of the hyper-population can also be used to construct prediction bands. Fig. 6D plots the prediction interval (90% confidence) for the response of the theoretical hyper-population to a treatment at 55 °C. Note that this figure

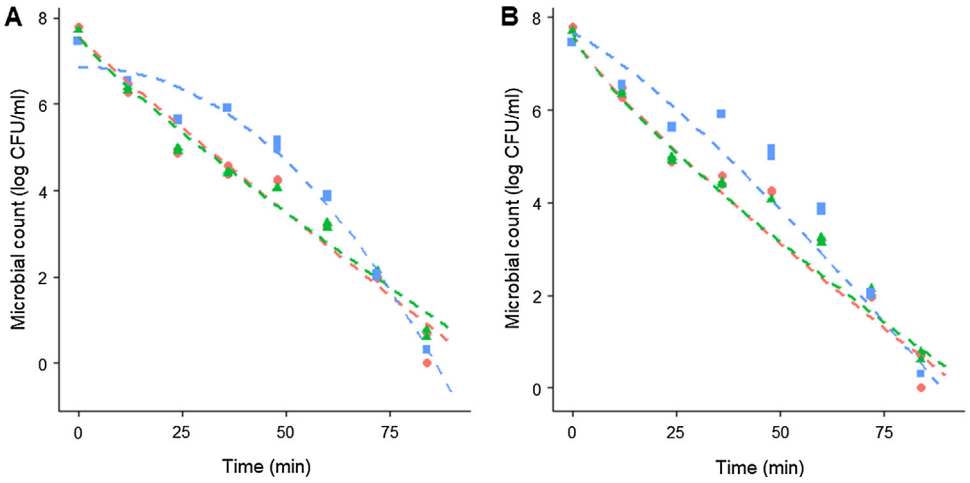


**Fig. 5.** Comparison of the survivor curves fitted for strain FBR12 at 55 °C using the single-level (A) and multilevel (B) modelling approaches. The symbols (dots, triangles and squares) indicate the observations of three different biological replicates (different shapes) and the dashed lines the fitted curves for each biological replicate.

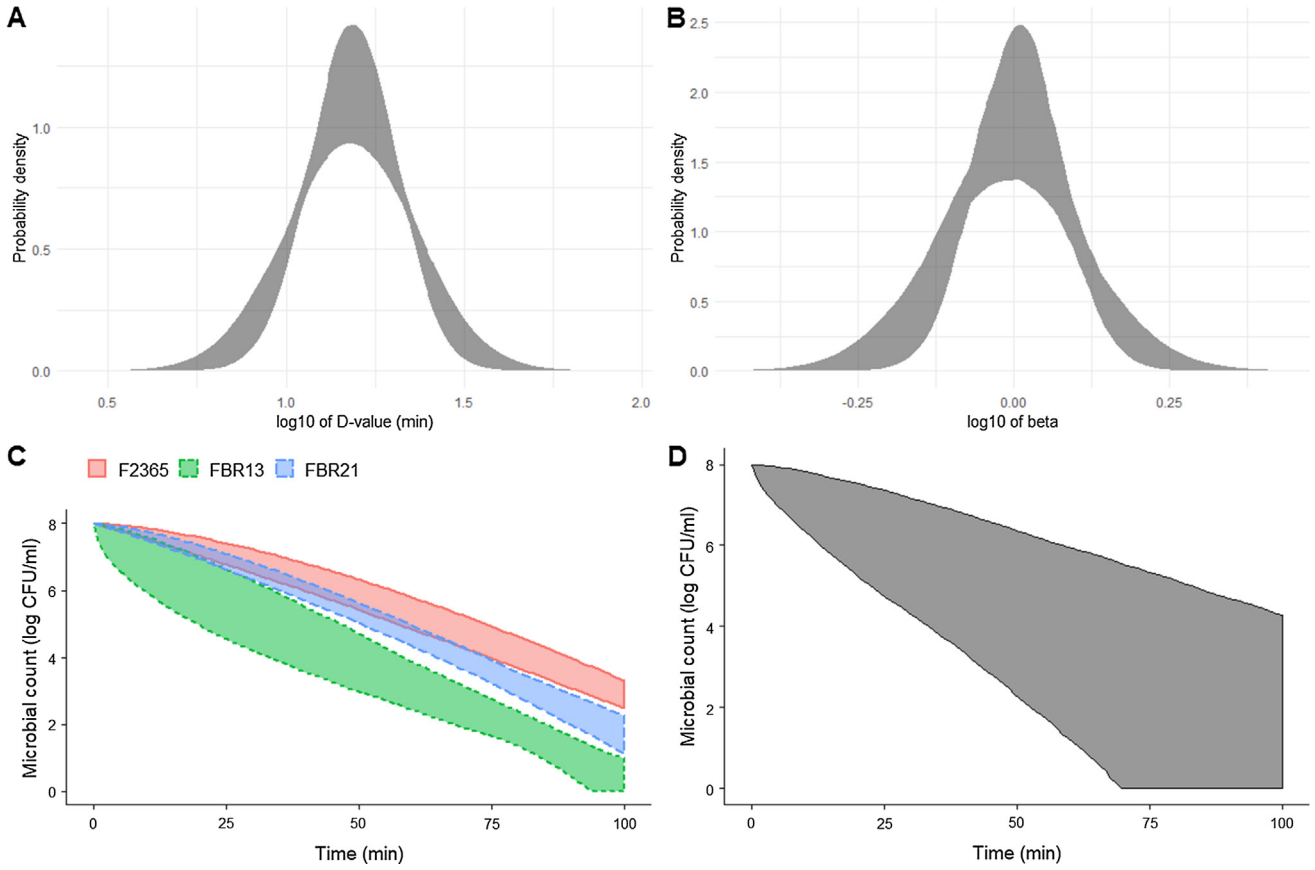


**Fig. 6.** Illustration of the application of a multilevel Bayesian modelling approach to include variability in model predictions and quantify different sources of variation. Distribution of the decimal logarithm of the *D*-value (A) and $\beta$ (B) for the theoretical hyper-population. Shaded area represents the uncertainty in strain variability (90% confidence interval. (C) Prediction intervals (90% confidence) considering within-strain variability, but not experimental uncertainty for three different strains (FBR12, red; FBR14, green; F2365, blue). (D) Prediction interval (90% confidence) of the inactivation of the hyper-population at 55 °C where the only source of variation considered is the between-strain variability.

reflects the between-strain variability and its uncertainty, but not the within-strain variability or the experimental uncertainty. A comparison between Fig. 6C and 6D clearly illustrates that predictions associated with a single strain have less uncertainty than those associated with the theoretical hyper-population. This is a reflection of the fact that, for this case, between-strain variability is much higher than within-strain variability, as already reported by Aryani et al. (2015a). The higher uncertainty affects both the inactivation rate and the curvature of the survivor curve. Due to the differences in the curvature direction of the strains analysed, the prediction for the theoretical hyper-population does not have a clear curvature direction. However, when the simulations are restricted to a single strain, the uncertainty on the curvature direction is largely reduced. Based on the data, it can be ensured that strain F2365 has a downwards curvature and strain FBR1 an upwards curvature. Regarding strain FBR12, there is some uncertainty regarding its curvature direction (as already discussed in Fig. 5). Nevertheless, for the dataset analysed, differences in the *D*-value of the strains have more practical relevance than slight, though significant, differences in $\beta$. Note that a correlation exists between the value of $\beta$ and $D_{\Delta}$ that can be attributed to parameter identifiability issues (Dolan & Mishra, 2013) or biological effects (Baranyi, Buss da Silva, & Ellouze, 2017). The parameter values used in the Monte Carlo simulations based on the multilevel model are drawn from the posterior distribution of the model parameters, which implicitly consider that correlation.

## 4. Discussion

Strain variability and population heterogeneities can have a significant impact on risk assessment (den Besten et al., 2017). As illustrated in Supp. Fig. 1, the effectiveness of a treatment can largely vary depending on the bacterial strain. For instance, there is a difference of approximately 6 log-reductions after 1 min of treatment at 60 °C between the most sensitive and the most resistant *Listeria* strains included in the case study. Because of that, variance is not simply a nuisance parameter in predictive microbiology. It is a parameter associated with a biological characteristic of the microorganism (variability) that can be as relevant as other parameters, such as growth or inactivation rates, for the case studied. Multilevel models are a good framework for the analysis of variation, because they incorporate variability and uncertainty in the first steps of the modelling process. In this approach, model hypotheses are not limited to model kinetics; an error model has to be defined to describe variability and uncertainty using statistical hypothesis that can later be contested by other experts or compared using model selection techniques. Besides, multilevel models can define parameters (e.g. correlations and standard deviations) to quantify the impact of each source of variation, enabling a quantification and comparison of different sources of variation (variability and uncertainty). Conversely, single-level models fit model parameter independently for each one of the different repetitions (biological replicates or at any relevant level). Hence, the contribution of each source of variation can only be done post-hoc, by an analysis of the variance of the parameter estimates. It can, then, be concluded that multilevel models are more flexible than single-level models and enable the application of some hypothesis that cannot be easily incorporated using other approaches. One clear example is the hypothesis that the technical error ($\sigma_e$ in this work) should be the same in every experiment.

Another example is the initial microbial count. As we have illustrated in this article, multilevel models can be used to assess the need for different hypotheses regarding these parameters. This added flexibility is especially interesting for analyses where the variance is a variable of interest, such as studies aiming to quantify the different sources of variability and uncertainty (den Besten et al., 2017; Koutsoumanis & Aspridou, 2017).

Besides their advantages for describing variability, multilevel models can increase the robustness of the model from a statistical point of view, reducing overfitting. Current microbiological procedures are affected by experimental error (Duarte, Stockmarr, & Nauta, 2015; Garre et al., 2019; Jarvis, 2008; Jongenburger, Reij, Boer, Gorris, & Zwietering, 2010) that, together with other sources of uncertainty, can introduce a bias in the estimated values of kinetic parameters. In some situations, such as the one illustrated in Fig. 5, the existence of an experimental outlier can be easily spotted. However, in many situations the impact of the experimental outlier in the results is more subtle, being hard to distinguish from a biological effect (Garcés-Vega & Marks, 2014; Garre et al., 2019). Although the uncertainty of the parameter estimates can be minimized by the application of optimum sampling schemes (Balsa-Canto, Alonso, & Banga, 2008; Garre, González-Tejedor, Peñalver-Soto, Fernández, & Egea, 2018), they cannot completely eliminate the risk of spurious parameter estimates due to an experimental error. The multilevel (Bayesian) approach suggested in this article uses all the information available in the dataset to estimate each model parameter. As a result, the parameter estimates obtained for one condition do not depend only on the data obtained for that condition; they are influenced by their uncertainty, the values estimated for other conditions and their uncertainties, and the total variation of the dataset. Therefore, multilevel models put some "distrust" in extreme values, which results in a "shrinkage" towards the mean (McElreath, 2016). This philosophy can be summarized via the statement popularized by Carl Sagan that "extraordinary claims require extraordinary evidence" (Gillispie, Fox, & Grattan-Guiness, 1997). The multilevel Bayesian approach will certainly not fix every issue related to experimental error. However, it provides a tool to avoid overfitting based on solid, statistical arguments.

As shown in this research, the multilevel model developed allows the analyst to select what sources of uncertainty and/or variability are included in model predictions (Fig. 6). It also defines probability distributions to describe the thermal resistance of the theoretical hyperpopulation of microbial strains, which can be used for making predictions considering the between-strain variability. However, it does not consider variability in prevalence. It is to be explored whether stress resistance is an influencing factor for the prevalence of a particular strain in the environment. The multilevel approach presented here is able to estimate prediction bands for each strain individually, accounting for the within-strain variability. These simulations could potentially be weighted, accounting for differences in prevalence between strains. This analysis is, nonetheless, out of the scope of this article and is left for future works.

On the other hand, the multilevel model has several disadvantages with respect to a single-level modelling approach. First of all, multilevel models are more complex than single-level models. This added complexity involves every step of the modelling process: model definition, implementation, post-processing and communication. The definition of a multilevel model requires that every underlying model assumption must be explicitly formulated as a stochastic or deterministic relationship. The complexity of this step should not be overlooked; myriad stochastic models can be defined for complex case studies (Gelman & Hill, 2007). Therefore, the application of model comparison and selection techniques may be required to select the best modelling approach. Due to the complex non-linear relationships between model parameters, model comparison can also be challenging in multilevel models. Hence, it is the task of the analyst to assess whether this increased complexity is required to solve the problem or a more simple model is enough, in other words, "simple is not stupid and complex is not always more correct" (Zwietering, 2009). Once the equation describing the stochastic and deterministic relationships of the models have been defined, the model must be fitted to the data. Multilevel models can be fitted applying a frequentist approach (usually called mixed-effects) or a Bayesian approach. The frequentist approach is more simple than the Bayesian one, but it is limited to models of low complexity (Gelman & Hill, 2007). Although Bayesian model fitting, is based in the "simple" Bayes rule, the conditional probability distributions can rarely be integrated analytically in most complex problems. This requires the application of so-called Approximate Bayesian Computations (ABC), which apply numerical integration (usually, some sort of Monte Carlo integration) to estimate the posterior distributions of the model parameters (Sunnåker et al., 2013). Although several methods are already available for this, most of them require advanced programming skills for their application. As an example, the one used in this study (*Stan*) is a programming language on itself. Furthermore, the results of an ABC are a sample of the posterior-distribution of the model parameters, whose analysis and interpretation is different to the output of frequentist analysis. Consequently, analysts able to, first, assess the validity of an ABC and, then, interpret the posterior distributions are required to apply this kind of models. Otherwise, the application of a Bayesian approach for parameter estimation makes more harm than good.

Last but not least, the results of the analysis must be communicated in a comprehensible way to other scientists who may not be experts in statistical modelling. As illustrated in this article, multilevel models can be of great interest for quantitative microbial risk assessment; a very sensitive topic where communication is an important step (ICF, Etienne, Chirico, Gunabalasingham, Jarvis, 2018; Thompson, 2002). This modelling approach applies advanced concepts (Bayesian statistics, multilevel modelling…) that may not be easy to understand by, for instance, managers who need to take discrete decisions (Zwietering, 2015). It is, therefore, crucial to develop effective ways of communicating the results of the study comprehensible for scientists and non-scientist with different backgrounds. It is, therefore, crucial the development effective ways of communicating the results of the study comprehensible for scientists and non-scientist with different backgrounds need to be developed.

## 5. Conclusions

This article has illustrated the construction of a multilevel model accounting for different sources of variability and uncertainty in predictive microbiology. This model includes variation at different levels (i.e. the $D$-value, the parameter $\beta$ and $\log_{10} N$) and is based on a non-linear model. Therefore, it presents a novel approach for kinetic modelling of microbial responses. This approach has several advantages with respect to single-level models. Firstly, it is based on stochastic hypotheses to describe different types of variation. Consequently, the model parameter quantify each contributor to variation and one can choose what sources of variation to include in model predictions. Another advantage is related to the "distrust" in extreme values through shrinkage that can mitigate the occurrence of spurious model parameters. On the other hand, multilevel models are harder to implement (advanced programming languages), analyse (Markov chains and posterior distributions) and communicate than single-level models. Hence, they are unlikely to replace single-level models for every case study. Nevertheless, considering current efforts towards quantifying variability and uncertainty of the microbial responses, multilevel models are likely to gain popularity during the next years. Although this work has been limited to a case study of microbial inactivation, this methodology can potentially be applied to a broad range of problems in predictive microbiology.

## CRediT authorship contribution statement

**Alberto Garre:** Conceptualization, Methodology, Software, Visualization. **Marcel H. Zwietering:** Conceptualization, Visualization. **Heidy M.W. den Besten:** Conceptualization, Visualization.

## Acknowledgments

Alberto Garre was supported by a postdoctoral granted from the Fundación Seneca (20900/PD/18) and by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Individual Fellowship grant No 844423 (FANTASTICAL).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2020.109374.

## References

Aguirre, J. S., Pin, C., Rodriguez, M. R., & Garcia de Fernando, G. D. (2009). Analysis of the variability in the number of viable bacteria after mild heat treatment of food. *Applied and Environmental Microbiology, 75*, 6992–6997. https://doi.org/10.1128/AEM.00452-09.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Aryani, D. C., den Besten, H. M. W., Hazeleger, W. C., & Zwietering, M. H. (2015a). Quantifying variability on thermal resistance of *Listeria monocytogenes*. *International Journal of Food Microbiology, 193*, 130–138. https://doi.org/10.1016/j.ijfoodmicro.2014.10.021.

Aryani, D. C., den Besten, H. M. W., Hazeleger, W. C., & Zwietering, M. H. (2015b). Quantifying strain variability in modeling growth of *Listeria monocytogenes*. *International Journal of Food Microbiology, 208*, 19–29. https://doi.org/10.1016/j.ijfoodmicro.2015.05.006.

Aspridou, Z., Koutsoumanis, K.P., 2015. Individual cell heterogeneity as variability source in population dynamics of microbial inactivation. Food Microbiology, Special Issue on Predictive modelling in food 45, Part B, 216–221. https://doi.org/10.1016/j.fm.2014.04.008.

Augustin, J.-C., Bergis, H., Midelet-Bourdin, G., Cornu, M., Couvert, O., Denis, C., ... Stahl, V. (2011). Design of challenge testing experiments to assess the variability of *Listeria monocytogenes* growth in foods. *Food Microbiology, Predictive Modeling in Foods, 28*, 746–754. https://doi.org/10.1016/j.fm.2010.05.028.

Balsa-Canto, E., Alonso, A. A., & Banga, J. R. (2008). Computing optimal dynamic experiments for model calibration in predictive microbiology. *Journal of Food Process Engineering, 31*, 186–206. https://doi.org/10.1111/j.1745-4530.2007.00147.x.

Baranyi, J., Buss da Silva, N., & Ellouze, M. (2017). Rethinking tertiary models: Relationships between growth parameters of *Bacillus cereus* strains. *Frontiers in Microbiology, 8*, 1890. https://doi.org/10.3389/fmicb.2017.01890.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Soft. 67*, pp.

Bates, D.M., Watts, D.G., 2007. Nonlinear regression analysis and its applications, 1 edition. ed. Wiley-Interscience, New York, NY.

Boekel, M.A.J.S. van, 2009. Kinetic modeling of reactions in foods, Food science and technology. CRC Press, Boca Raton.

den Besten, H. M. W., & Zwietering, M. H. (2012). Meta-analysis for quantitative microbiological risk assessments and benchmarking data. *Trends in Food Science and Technology, 25*(1), 34–39. https://doi.org/10.1016/j.tifs.2011.12.004.

Brooks, S. (2011). *Handbook of Markov chain.* Monte Carlo.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: a probabilistic programming language. Journal of Statistical Software 76, 1–32. https://doi.org/10.18637/jss.v076.i01.

Cadavez, V. A. P., Rodrigues, V., & Gonzales-Barron, U. A. (2017). Microbiological Safety of Goat Milk and Cheese: Evidences from a Meta-Analysis. In J. Simões, & C. Gutiérrez (Vol. Eds.), *Sustainable Goat Production in Adverse Environments: Vol. I*, (pp. 379–390). Springer International Publishing. https://doi.org/10.1007/978-3-319-71855-2_21.

den Besten, H. M. W., Aryani, D. C., Metselaar, K. I., & Zwietering, M. H. (2017). Microbial variability in growth and heat resistance of a pathogen and a spoiler: All variabilities are equal but some are more equal than others. *International Journal of Food Microbiology, 240*, 24–31. https://doi.org/10.1016/j.ijfoodmicro.2016.04.025.

den Besten, H. M. W., Wells-Bennik, M. H. J., & Zwietering, M. H. (2018). Natural diversity in heat resistance of bacteria and bacterial spores: Impact on food safety and quality. *Annual Review of Food Science and Technology, 9*, 383–410. https://doi.org/10.1146/annurev-food-030117-012808.

Dolan, K. D., & Mishra, D. K. (2013). Parameter estimation in food science. *Annual Review of Food Science and Technology, 4*, 401–422. https://doi.org/10.1146/annurev-food-022811-101247.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B, 195*, 216–222. https://doi.org/10.1016/0370-2693(87)91197-X.

Duarte, A. S. R., Stockmarr, A., & Nauta, M. J. (2015). Fitting a distribution to microbial counts: Making sense of zeroes. *International Journal of Food Microbiology, 196*, 40–50. https://doi.org/10.1016/j.ijfoodmicro.2014.11.023.

Scientific Committee, E. F. S. A., Benford, D., Halldorsson, T., Jeger, M. J., Knutsen, H. K., More, S., ... Hardy, A. (2018). Guidance on uncertainty analysis in scientific assessments. *EFSA Journal, 16*, 5123. https://doi.org/10.2903/j.efsa.2018.5123.

Eichelsbacher, P., Elsner, G., Kösters, H., Löwe, M., Merkl, F., Rolles, S. (Eds.), 2013. Limit theorems in probability, statistics and number theory: in honor of Friedrich Götze, Springer Proceedings in Mathematics & Statistics. Springer-Verlag, Berlin Heidelberg.

Franz, C. M. A. P., den Besten, H. M. W., Böhnlein, C., Gareis, M., Zwietering, M. H., & Fusco, V. (2019). Reprint of: Microbial food safety in the 21st century: Emerging challenges and foodborne pathogenic bacteria. Trends in Food Science & Technology, Keeping Food Safety on the Agenda for 15 years. *The SAFE consortium, 84*, 34–37. https://doi.org/10.1016/j.tifs.2019.01.009.

Gamerman, D., Lopes, H.F., 2006. Markov chain Monte Carlo: stochastic simulation for Bayesian inference, 2nd ed. ed, Texts in statistical science series. Taylor & Francis, Boca Raton.

Garcés-Vega, F., & Marks, B. P. (2014). Use of simulation tools to illustrate the effect of data management practices for low and negative plate counts on the estimated parameters of microbial reduction models. *Journal of Food Protection, 77*, 1372–1379. https://doi.org/10.4315/0362-028X.JFP-13-462.

García, M. R., Vilas, C., Herrera, J. R., Bernárdez, M., Balsa-Canto, E., & Alonso, A. A. (2015). Quality and shelf-life prediction for retail fresh hake (*Merluccius merluccius*). *International Journal of Food Microbiology, 208*, 65–74. https://doi.org/10.1016/j.ijfoodmicro.2015.05.012.

Garre, A., Egea, J. A., Esnoz, A., Palop, A., & Fernandez, P. S. (2019). Tail or artefact? Illustration of the impact that uncertainty of the serial dilution and cell enumeration methods has on microbial inactivation. *Food Research International.*. https://doi.org/10.1016/j.foodres.2019.01.059.

Garre, A., González-Tejedor, G., Peñalver-Soto, J. L., Fernández, P. S., & Egea, J. A. (2018). Optimal characterization of thermal microbial inactivation simulating non-isothermal processes. *Food Research International, 107*, 267–274. https://doi.org/10.1016/j.foodres.2018.02.040.

Gelman, A., 2014. Bayesian data analysis, Third edition. ed, Chapman & Hall/CRC texts in statistical science. CRC Press, Boca Raton.

Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models, Analytical methods for social research. Cambridge University Press, Cambridge ; New York.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24*, 997–1016. https://doi.org/10.1007/s11222-013-9416-2.

Gillispie, C. C., Fox, R., & Grattan-Guiness, I. (1997). Pierre-Simon Laplace, 1749–1827: A Life in Exact Science. *Princeton University Press*. https://doi.org/10.2307/j.ctv39x7f8.

Gonzales-Barron, U., & Butler, F. (2011). The use of meta-analytical tools in risk assessment for food safety. *Food Microbiology, Predictive Modeling in Foods, 28*, 823–827. https://doi.org/10.1016/j.fm.2010.04.007.

González-Tejedor, G. A., Garre, A., Esnoz, A., Artés-Hernández, F., & Fernández, P. S. (2018). Effect of storage conditions in the response of *Listeria monocytogenes* in a fresh purple vegetable smoothie compared with an acidified TSB medium. *Food Microbiology, 72*, 98–105. https://doi.org/10.1016/j.fm.2017.11.005.

González-Tejedor, G. A., Martínez-Hernández, G. B., Garre, A., Egea, J. A., Fernández, P. S., & Artés-Hernández, F. (2017). Quality changes and shelf-life prediction of a fresh fruit and vegetable purple smoothie. *Food Bioprocess Technology, 1–13*. https://doi.org/10.1007/s11947-017-1965-5.

Guillou, S., & Membré, J.-M. (2019). Inactivation of *Listeria monocytogenes*, *Staphylococcus aureus*, and *Salmonella enterica* under high hydrostatic pressure: A quantitative analysis of existing literature data. *Journal of Food Protection, 82*, 1802–1814. https://doi.org/10.4315/0362-028X.JFP-19-132.

Hoffman, M.D., Gelman, A., 2011. The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. arXiv:1111.4246 [cs, stat].

ICF, Etienne, J., Chirico, S., Gunabalasingham, T., Jarvis, A., 2018. Final report: Clear communications and uncertainty: external scientific report. RC/EFSA/COMM/2015/01. EFSA Supporting Publications 15. https://doi.org/10.2903/sp.efsa.2018.EN-1412.

Jaloustre, S., Guillier, L., Morelli, E., Noël, V., & Delignette-Muller, M. L. (2012). Modeling of *Clostridium perfringens* vegetative cell inactivation in beef-in-sauce products: A meta-analysis using mixed linear models. *International Journal of Food Microbiology, 154*, 44–51. https://doi.org/10.1016/j.ijfoodmicro.2011.12.013.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning: with applications in R, Edición: 1st ed. 2013, Corr. 6th printing 2016. ed. Springer.

Jarvis, B. (2008). *Statistical aspects of the microbiological examination of foods.* S.l.: Elsevier Academic Press.

Jongenburger, I., Reij, M. W., Boer, E. P. J., Gorris, L. G. M., & Zwietering, M. H. (2010). Factors influencing the accuracy of the plating method used to enumerate low numbers of viable micro-organisms in food. *International Journal of Food Microbiology, 143*, 32–40. https://doi.org/10.1016/j.ijfoodmicro.2010.07.025.

Koutsoumanis, K. P., & Aspridou, Z. (2017). Individual cell heterogeneity in Predictive Food Microbiology: Challenges in predicting a "noisy" world. *International Journal of Food Microbiology, 240*, 3–10. https://doi.org/10.1016/j.ijfoodmicro.2016.06.021.

Koyama, K., Abe, H., Kawamura, S., & Koseki, S. (2019). Stochastic simulation for death probability of bacterial population considering variability in individual cell inactivation time and initial number of cells. *International Journal of Food Microbiology, 290*, 125–131. https://doi.org/10.1016/j.ijfoodmicro.2018.10.009.

Koyama, K., Aspridou, Z., Koseki, S., & Koutsoumanis, K. (2019). Describing uncertainty in *Salmonella* thermal inactivation using Bayesian statistical modeling. *Frontiers in Microbiology, 10*. https://doi.org/10.3389/fmicb.2019.02239.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*, 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008.

Liu, Y., Sun, W., Sun, T., Gorris, L. G. M., Wang, X., Liu, B., & Dong, Q. (2020). The prevalence of *Listeria monocytogenes* in meat products in China: A systematic literature review and novel meta-analysis approach. *International Journal of Food Microbiology, 312*, 108358. https://doi.org/10.1016/j.ijfoodmicro.2019.108358.

Mafart, P., Couvert, O., Gaillard, S., & Leguerinel, I. (2002). On calculating sterility in thermal preservation methods: Application of the Weibull frequency distribution model. *International Journal of Food Microbiology, 72*, 107–113. https://doi.org/10.1016/S0168-1605(01)00624-9.

McElreath, R., 2016. Statistical rethinking: a Bayesian course with examples in R and Stan, Chapman & Hall/CRC texts in statistical science series. CRC Press/Taylor & Francis Group, Boca Raton.

Metselaar, K. I., den Besten, H. M. W., Abee, T., Moezelaar, R., & Zwietering, M. H. (2013). Isolation and quantification of highly acid resistant variants of *Listeria monocytogenes*. *International Journal of Food Microbiology, 166*, 508–514. https://doi.org/10.1016/j.ijfoodmicro.2013.08.011.

Murdoch, D. J., Tsai, Y.-L., & Adcock, J. (2008). P-values are random variables. *The American Statistician, 62*, 242–245. https://doi.org/10.1198/000313008X332421.

Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology, 57*, 9–18.

Peleg, M., & Cole, M. B. (1998). Reinterpretation of Microbial Survival Curves. *Critical Reviews in Food Science and Nutrition, 38*, 353–380. https://doi.org/10.1080/10408699891274246.

Perez-Rodriguez, F., & Valero, A. (2012). *Predictive Microbiology in Foods* (first ed.). New York: Springer.

Possas, A., Valdramidis, V., García-Gimeno, R. M., & Pérez-Rodríguez, F. (2019). High hydrostatic pressure processing of sliced fermented sausages: A quantitative exposure assessment for *Listeria monocytogenes*. *Innovative Food Science & Emerging Technologies, 52*, 406–419. https://doi.org/10.1016/j.ifset.2019.01.017.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rigaux, C., Denis, J.-B., Albert, I., & Carlin, F. (2013). A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: Estimation of D at 121.1°C and pH 7, zT and zpH of *Geobacillus stearothermophilus*. *International Journal of Food Microbiology, 161*, 112–120. https://doi.org/10.1016/j.ijfoodmicro.2012.12.001.

Ros-Chumillas, M., Garre, A., Maté, J., Palop, A., & Periago, P. M. (2017). Nanoemulsified D-limonene reduces the heat resistance of *Salmonella* Senftenberg over 50 times. *Nanomaterials, 7*, 65. https://doi.org/10.3390/nano7030065.

Santos, J.L.P., Samapundo, S., Gülay, S.M., Van Impe, J., Sant'Ana, A.S., Devlieghere, F., 2018. Inter- and intra-species variability in heat resistance and the effect of heat treatment intensity on subsequent growth of Byssochlamys fulva and Byssochlamys nivea. International Journal of Food Microbiology. https://doi.org/10.1016/j.ijfoodmicro.2018.04.035.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology, 9*, e1002803. https://doi.org/10.1371/journal.pcbi.1002803.

Taper, M. L., & Lele, S. (Eds.). (2004). *The nature of scientific evidence: statistical, philosophical, and empirical considerations*. Chicago: University of Chicago Press.

Thompson, K. M. (2002). Variability and uncertainty meet risk management and risk communication. *Risk Analysis, 22*, 647–654. https://doi.org/10.1111/0272-4332.00044.

van Asselt, E. D., & Zwietering, M. H. (2006). A systematic approach to determine global thermal inactivation parameters for various food pathogens. *International Journal of Food Microbiology, 107*, 73–82. https://doi.org/10.1016/j.ijfoodmicro.2005.08.014.

van Boekel, M. A. J. S. (2020). On the pros and cons of Bayesian kinetic modeling in food science. *Trends in Food Science & Technology, 99*, 181–193. https://doi.org/10.1016/j.tifs.2020.02.027.

Wells-Bennik, M. H. J., Janssen, P. W. M., Klaus, V., Yang, C., Zwietering, M. H., & Den Besten, H. M. W. (2019). Heat resistance of spores of 18 strains of *Geobacillus stearothermophilus* and impact of culturing conditions. *International Journal of Food Microbiology, 291*, 161–172. https://doi.org/10.1016/j.ijfoodmicro.2018.11.005.

Zwietering, M. (2015). Risk assessment and risk management for safe foods: Assessment needs inclusion of variability and uncertainty, management needs discrete decisions. *International Journal of Food Microbiology, 213*, 118–123. https://doi.org/10.1016/j.ijfoodmicro.2015.03.032.

Zwietering, M. H. (2009). Quantitative risk assessment: Is more complex always better? Simple is not stupid and complex is not always more correct. *International Journal of Food Microbiology, 134*, 57–62. https://doi.org/10.1016/j.ijfoodmicro.2008.12.025.