Mühendislik Bilimleri ve Tasarım Dergisi
9(2), 628 – 640, 2021
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.929974

Araştırma Makalesi Research Article

IS IT POSSIBLE TO MAKE FEWER EXPERIMENTS: PREDICTION OF BACTERIAL SURVIVAL/DEATH PROBABILITY FOR HIGH-PRESSURE PROCESSING WITH THE BAYESIAN APPROACH?

Sebahattin Serhat TURGUT*

Suleyman Demirel University, Faculty of Engineering, Department of Food Engineering, Isparta, Turkey

Keywords

High Hydrostatic Pressure, Predictive Microbiology, Logistic Regression, Listeria Monocytogenes, Cronobacter Sakazakii.

Abstract

In the present study, a model based on Bayesian Logistic Regression (*BLR*) was developed to predict the probability of bacterial survival/death treated with high-hydrostatic pressure under different conditions. Previously published data for *Listeria monocytogenes* in phosphate-buffered saline and *Cronobacter sakazakii* in trypticase soy broth and infant formula were used where the process variables were pressure, temperature, medium pH, initial inoculum and processing time. Along with the using possibility of *BLR*, effects of introduced sampling size by changing data split ratio and case prevalence were assessed. The *BLR* model predictions were consistent with both experimental data and the frequentist logistic regression models. Although some overfitting problems arise as the sampling size decrease, *BLR* can produce reliable probability models with a smaller number of experimental data (about 50 experimental samples) than the frequentist approach requires. Moreover, instead of a point estimate, *BLR* offers a posterior distribution for parameters and predictions. So the present study has indicated that *BLR* can be a useful tool to describe the survival/death of microorganisms after high-pressure processes with less experimental data requirement than the frequentist approach and also with the ability to handle missing observation and imbalanced dataset. In the light of these outcomes, the design of new experiments according to *BLR*, save on time and costs for experimental studies and more detailed safety risk assessment may be feasible for the food industry.

DAHA AZ DENEME GERÇEKLEŞTİRMEK MÜMKÜN MÜ: BAYESIAN YAKLAŞIMLA YÜKSEK BASINÇ İŞLEMLERİ İÇİN BAKTERİYEL HAYATTA KALMA/ÖLÜM OLASILIĞININ TAHMİNİ?

Anahtar Kelimeler

Yüksek Hidrostatik Basınç, Öngörücü Mikrobiyoloji, Lojistik Regresyon, Listeria Monocytogenes, Cronobacter Sakazakii.

Öz

Mevcut çalışmada, farklı koşullar altında yüksek hidrostatik basınç işlemine tabi tutulan bakterilerin hayatta kalma/ölüm olasılığını tahmin etmek için Bayesian Logistic Regression'a (*BLR*) dayalı bir model geliştirilmiştir. Bu amaçla *Listeria monocytogenes* (fosfatla tamponlanmış tuzlu su çözeltisi içinde) ve *Cronobacter sakazakii* (triptik soya broth ve bebek maması formülasyonu) bakterileri için daha önce yayımlanmış verilerden faydalanılmış olup, proses değişkenleri basınç, sıcaklık, ortamın pH değeri, ilk aşılama ve işlem süresidir. *BLR* kullanım olasılığının yanı sıra, veri bölme oranları değiştirilerek örneklem büyüklüğünün ve verilerdeki vaka sıklığının etkileri değerlendirilmiştir. Sonuç olarak *BLR* model tahminlerinin hem deneysel verilerle hem de frekansçı lojistik regresyon modelleriyle tutarlı olduğu gözlenmiştir. Örneklem boyutu küçüldükçe bazı aşırı uyum sorunları ortaya çıksa da, *BLR*, frekansçı yaklaşımının gerektirdiğinden daha az sayıda deneysel veriye ile (yaklaşık 50 deneysel örnek) güvenilir olasılık modelleri üretebilmektedir. Dahası *BLR*, nokta tahminleri yerine parametreler ve kestirimler için sonsal dağılımlar sunmaktadır. Bu nedenle mevcut çalışmada, *BLR*'nin frekansçı yaklaşıma göre daha az deneysel veri gereksinimiyle mikroorganizmaların uygulanan yüksek basınç işlemlerinden sonra hayatta kalma/ölme olasılık kestirimleri için yararlı bir

* İlgili yazar / Corresponding author: serhatturgut@sdu.edu.tr, +90-246-211-1668

araç olabileceği, eksik gözlemleri ve dengesiz veri setlerini yönetme kabiliyetine sahip olduğu gösterilmiştir. Bu sonuçların ışığında, *BLR* yaklaşımına uygun yeni deneme tasarımları ile, deneysel çalışmalarda zamandan ve maliyetten tasarruf sağlanması ve gıda endüstrisi için daha ayrıntılı güvenlik riski değerlendirmesi mümkün olabilir.

Alıntı / Cite

Turgut, S. S., (2021). Is It Possible to Make Fewer Experiments: Prediction of Bacterial Survival/Death Probability for High-Pressure Processing with the Bayesian Approach?, Journal of Engineering Sciences and Design, 9(2), 628-640.

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
S. S. Turgut, 0000-0002-9968-4750	**Başvuru Tarihi / Submission Date**	29.04.2021
	Revizyon Tarihi / Revision Date	09.05.2021
	Kabul Tarihi / Accepted Date	12.05.2021
	Yayım Tarihi / Published Date	20.06.2021

1. Introduction

Food safety, especially in terms of microbiological safety, is one of the most important criteria for food producers, regulatory organisations and consumers. The microbiological safety of foods is often ensured by heat-induced methods such as pasteurisation and sterilisation. However, during heating treatments, food products undergo some accelerated biochemical changes resulting in loss of bioactive and beneficial components and/or formation of some colour and flavour compounds some of which may be undesirable (Yamamoto, 2017). But as a result of ongoing changes in consumer preferences tends to fresh-like foods and non-thermal methods under which limited changes occur. Thus, these methods gain increasing popularities. Among the non-thermal methods, high-pressure processing (HPP) takes increasing interest since it was first announced in food science and technology by Hite (1899) due to its superiority over other methods originating from [1] minimal effects on the sensorial, visual and nutritional value of foods; [2] being independent of food geometry and size; [3] utilise ability for packaged or bulk foodstuffs; [4] being environmentally friendly and relatively low energy costs (Stoica et al., 2013; Yamamoto, 2017; Zhang et al., 2019). Today, HPP is actively used by the food industry in the USA, European countries, Korea, Taiwan, Japan etc. (Yamamoto, 2017). It has been studied for several food products such as milk and cheese (Bilbao-Sáinz et al., 2009; Hokmollahi and Ehsani, 2017), fruit and vegetables (Butz et al., 2003), meat and meat products (Campus, 2010; Khan et al., 2014), fish (Gudbjornsdottir et al., 2010), fruit juice (Ferrari et al., 2010) etc.

Microbial inactivation efficacy of the HPP process strictly depends on process variables such as pH, water activity, temperature, applied pressure, process time and initial microbial load (Hajmeer and Basheer, 2002; Koseki et al., 2009; Koseki and Yamamoto, 2007). The changes caused by these variables should be understood, controlled and optimised carefully to ensure the safety of the final product. For that purpose, several statistical learning methods such as kinetic models (Buzrul, 2014; Serment-Moreno et al., 2014), neural networks (Hajmeer and Basheer, 2002), Weibull model (Buzrul et al., 2008), logistic regression (Buzrul, 2019; Koseki et al., 2009; Koseki and Yamamoto, 2007; Wang et al., 2017), piecewise models (Buzrul, 2017) have been utilised. Among these modelling approaches, logistic regression (*LR*), has been widely accepted by researchers due to its ability to handle non-linear kinetics of microbial inactivation and ease of implementation (Koseki et al., 2009). *LR* can be used as a probability estimator of survival or death of microorganisms under given conditions which named as survival/death interface model in the related literature (Koseki and Yamamoto, 2007). However, as a frequentist and discriminative modelling approach, *LR* harbours some disadvantages. The frequentist and discriminative approaches assume the observed data is sampled from a fixed and known distribution. They do not assign probabilities of parameter values. Therefore, frequentist approaches measure model uncertainty relying on the null hypothesis and confidence intervals which does indeed not giving the uncertainty of true estimates. Moreover, they intend to find point estimates of the parameters for the model from given data from maximum likelihood estimation by maximising log-likelihood value for *LR*. Although they can work with the presence of missing data, discriminative models generally require the full set of observations and larger data sets to train the model (O'Brien and Dunson, 2004; Xu, 2020a, b). On the other hand, Bayesian approaches can overcome some negativities of the frequentist approach. For instance, for Bayesian Logistic Regression (*BLR*), the Bayesian version of *LR*, the prior distribution for parameters and data according to the researcher's knowledge is used. So as an output, not a matrix of point estimates but posterior distributions of model parameters and predictions are produced. This will lead to an understanding of credible interval (or probability/plausibility/likelihood) of the data which help to reach the true uncertainty of the model. On the other side, for frequentists, population parameters are assumed to be fixed which requires many experiments to find sampling distribution (Lu, 2019). However, *BLR* requires a comparably smaller number of experimental data and can handle missing observations (O'Brien and Dunson, 2004). However, it is computationally more intensive (Xu, 2020a, b). To the best of our knowledge, there is no study on the use of

BLR for bacterial survival/death after HPP treatments and searching the possibility of developing a probabilistic model conducting less number of experiments.

So, the objective of the current study is to [1] investigate the possibility of the use of *BLR* as a survival/death interface model for *Listeria monocytogenes* and *Cronobacter sakazakii* which were processed with HPP under varying temperature, pH, inoculum, pressure and time combinations; [2] assess the prediction performance of *BLR* models against the use of reduced numbers and imbalanced case prevalence of experimental data and [3] discuss cons and pros of *BLR* from the perspective of food safety risks.

2. Material and Method

2.1. Data Sets

The datasets used in the present study were taken from previously published papers by Koseki and Yamamoto (2007) and Koseki et al. (2009) with their permission. Brief information about the used dataset including the microorganisms of interest, process conditions and levels, number of observations were given in Table 1. The goal of the given data set is to create a binary classification model which predicts whether the given microorganisms will survive after HPP treatments under known conditions and its probability. The target responses were coded with the value of 1 and 0 for survival and death, respectively. The absence or presence of viable bacteria was determined by plating the samples taken after HPP onto nonselective agar (tryptic soy agar [TSA]). Plates were then incubated at 37°C for 48 h and inspected for either colony formation (survival) or no colony formation (death). For the details of the after HPP procedure, please see Koseki and Yamamoto (2007) and Koseki et al. (2009).

Table 1. Summary of data used for Bayesian Logistic Regression

	Reference	
Parameters	**[1]***	**[2]***
Microorganism	*L. monocytogenes* ATCC 19117	*C. sakazakii* ATCC 29544
Medium	0.01 M phosphate-buffered saline	- Trypticase soy broth (TSB) - Infant formula (IF)
Pressure (MPa)	200, 300, 400, 500	400, 450, 500, 550, 600
Temperature (°C)	22	25 or 40
pH	3, 4, 5, 6, 7	7.2 for TSB
Inoculum (log_{10}CFU/ml)	3, 5, 7	3,5,7
Time (min)	1, 3, 5, 10, 20, 30	1, 3, 5, 10, 20
Number of observations	360 combinations · 3 replicates = 1080	300 combinations · 3 replicates = 900

* [1] Koseki and Yamamoto (2007), [2] Koseki et al. (2009).

2.2. Bayesian regression and model development platform

Since the present study aims to investigate and propound the possibility of the use of the Bayesian approach for probabilistic modelling of bacterial survival/death expectation, the given models' structures were directly taken from the initial models given in the reference papers (Koseki et al., 2009; Koseki and Yamamoto, 2007). Therefore, the models were formed as given below for the data from Koseki and Yamamoto (2007) (Eqs. (1)-(2)) and the data from Koseki et al. (2009) (Eqs. (3)-(5));

$$logit(p) = a_0 + a_1P + a_2log_{10}(t) + a_3pH + a_4IC \quad (1)$$

$$logit(p) = a_0 + a_1P + a_2t + a_3pH + a_4IC + a_5P \cdot t + a_6P \cdot pH + a_7P \cdot IC + a_8t \cdot pH + a_9t \cdot IC + a_{10}pH \cdot IC + a_{11}P^2 + a_{12}t^2 + a_{13}pH^2 + a_{14}IC^2 \quad (2)$$

$$logit(p) = a_0 + a_1P + a_2ln(t) + a_3T + a_4IC + a_5M \quad (3)$$

$$logit(p) = a_0 + a_1P + a_2ln(t) + a_3T + a_4IC + a_5M + a_6P \cdot T \quad (4)$$

$$logit(p) = a_0 + a_1P + a_2t + a_3T + a_4IC + a_5M + a_6P \cdot t + a_7P \cdot T + a_8P \cdot IC + a_9P \cdot M + a_{10}t \cdot T + a_{11}t \cdot IC + a_{12}t \cdot M + a_{13}T \cdot IC + a_{14}T \cdot M + a_{15}IC \cdot M + a_{16}P^2 + a_{17}t^2 + a_{18}T^2 + a_{19}IC^2 \quad (5)$$

where $logit(p)$ is the abbreviation for $ln[p/(1-p)]$, ln is the natural algorithm, p is the probability of survival (in the range from 0 to 1), a_i values are model coefficients, P is the applied pressure (MPa), t is the pressure holding time (min), T is the process temperature (°C), IC is the inoculum concentration of the bacteria of interest in the

test medium ($\log_{10}$CFU/ml) and M is the type of medium for Eqs. (3)-(5) (TSB and IF coded as dummy variables of 0 and 1, respectively). The logarithm was applied for t in Eq. (1), (3) and (4) since the relationship between holding and microbial inactivation is not linear (Buzrul and Alpas, 2004; Buzrul et al., 2005; Koseki et al., 2009; Koseki and Yamamoto, 2007).

BLR models were developed using PyMC3 library which is for probabilistic programming in Python (Salvatier et al., 2016) over Google Colab and the practice given at the PyMC3 package documentation web page (https://docs.pymc.io/notebooks/GLM-logistic.html) for logistic regression was followed. To conduct *BLR*, the first step is to specify a prior distribution to draw samples from the posterior. Especially, when the number of the experimental sampling is very low, prior distribution estimation may have an impact on the outcome (van Boekel, 2020). However, one of the best advantages of the PyMC3 library is its self-tuning strategies for adaptively setting the tunable parameters of MCMC (Markov Chain Monte Carlo) over NUTS (No-U-Turn Sampler). Moreover, PyMC3 provides a very simple, succinct and flexible probabilistic programming platform for quantitative researchers to implement statistical models. Its large library of statistical distributions and pre-defined fitting algorithms help researchers to focus on the scientific problem rather than the implementation details of Bayesian modelling (Salvatier et al., 2016). So, the model settings were left as default in the present study to utilise PyMC3's advantages. As being default, zero-mean normal distribution as a prior with a variance of 10 was applied for parameters, which corresponds to non-informative priors. For the logistics regression model, the error distribution was specified as binomial. Being different from default settings, 2000 tuning samplings were used before the MCMC sampler with 4 chains and "target_accept" parameter was increased to 0.90 to obtain a better convergence with the use of smaller step size.

3. Results and Discussion

3.1. Model Development

Two different data sets consisting of 900 and 1080 observations respectively from Koseki et al. (2009) and Koseki and Yamamoto (2007) were used in the present study. In brief, these data sets contain the observations for survival/death of *L. monocytogenes* and *C. sakazakii* after HPP treatments with varying levels of applied pressure (MPa) for different pH values where temperature (°C) and pressure medium were also changed in Koseki et al. (2009) and Koseki and Yamamoto (2007), respectively. Moreover, initial microbial inoculum ($\log_{10}$CFU/ml) and process time (min) were also other important factors of experiments. Five different models (Eqs. (1)-(2) for the data from Koseki and Yamamoto (2007) and Eqs. (3)-(5) for the data from Koseki et al. (2009)) were developed using the PyMC3 library. Note that although it is not a common practice to separate data as training and test sets for survival/death prediction modelling of bacteria using *LR* (Buzrul, 2019; Koseki et al., 2009; Koseki and Yamamoto, 2007), it has been performed in our study. The purpose of this effort was to figure out whether the models developed by using MCMC *BLR* with smaller number of experiments produce satisfying results as it was previously mentioned as one of the main advantages of Bayesian regression. So, the data sets were randomly separated as training and testing groups using split ratios of 0.10, 0.40, 0.70, 0.90, 0.95 and 0.99 (keeping the ratio of 0/1 (death/survival) which was 55.56 and 33.11% for data from Koseki and Yamamoto (2007) and Koseki et al. (2009), respectively). This means that the given ratio of the entire data was left to a test set for the developed model whereas the rest of the data was used to train it. This was done to observe/compare the prediction performance of the models developed using MCMC *BLR*. The goodness of the models' fit was assessed using simple and classical statistics which are the percentage of model accuracy, sensitivity and specificity with confidence matrix. These statistics were calculated as given by Baratloo et al. (2015) as follows (Eqs. (6)-(8));

$$Accuracy\ (\%) = \frac{TP + FP}{TP + TN + FP + FN} \times 100 \quad (6)$$

$$Sensitivity\ (\%) = \frac{TP}{TP + FN} \times 100 \quad (7)$$

$$Specificity(\%) = \frac{TN}{TN + FP} \times 100 \quad (8)$$

where TP is the number of true-positives (observations correctly identified as death (0)); TN is the number of true-negatives (observations correctly identified as survival (1)); FP is the number of false-positives (observations incorrectly identified as death (0)); FN is the number of false-negatives (observations incorrectly identified as survival (1)). Here, the accuracy of the model indicates its ability to estimate both the survival and death results correctly. Sensitivity is the ability to differentiate the death cases correctly and specificity is the ability to determine the survival cases correctly (Baratloo et al., 2015). FP or fail-safe represents a model that telling the

bacteria survives under the conditions which it should be inactivated. This means that according to the extreme fail-safe model, consumption of the mentioned food is unsafe so it should be discarded (Ratkowsky, 2004). On the other hand, FN or fail-dangerous is also undesired especially for manufacturers. Because in such a case a new treatment with a higher/denser process conditions like higher temperature, longer times would be applied to inactivate the "already inactivated" microorganism which is, of course, waste of time and energy (Buzrul, 2019). Therefore, along with accuracy, sensitivity and specificity, FN and FP are also important to assess the preferability of a model. The model parameters (mean of posterior distributions as points estimates) with their uncertainty (2.5 and 97.5% percentiles of posterior distributions) and goodness of fit statistics were given in Tables 2-6 for Eqs. (1)-(5).

3.2. Model Test and Prediction

As known, the Bayesian approach does not give only one best fitting line or point estimates but it provides an approximation for whole posterior distribution for model parameters (Salvatier et al., 2016). For that purpose, PyMC3 uses Markov Chain Monte Carlo to draw samples from the priors to approximate the posteriors. Monte Carlo is a popular technique of drawing random samples from a distribution, and Markov Chain refers to that the next sample will be drawn according to only previous sample value. The combination of these two methods reveals Markov Chain Monte Carlo (MCMC). As more samples are drawn, the approximation of the posterior will converge on the true posterior distribution for the model parameters at the end (Koehrsen, 2018). To assess the goodness of convergence, trace plots are commonly utilised. In Fig. 1, the trace plot for Eq. (1) was given (for clarity the trace plots for other models were not presented). This plot visualises all the samples (except the discarded tuning samples) drawn for all the variables as divided into two columns. On the left-hand side of Fig. 1, the final approximate posterior distribution for model parameters (in fact they are our model) and on the right-hand side the individual sample values at each step of samplings during MCMC were given. The four different colours in the figures represent the four different sampling chains. It is seen that the model parameters were not point estimates. They were given as posterior distributions displaying how we were uncertain about the true values. From the right-hand side of Fig. 1, the sampling chains for the individual parameters seem well converged and stationary (there are no large drifts or other odd patterns). So, we can use the mean values of distributions as most likely estimates which are given also in Table 2 for Eq. (1) and in Tables 5-6 for other models (PyMC3, 2018). Note that, since MCMC *BLR* is an iterative method producing a distribution, it is possible to face very small and ignorable differences in most likely estimates of model parameters if one reruns the code using even the same data. Moreover, MCMC *BLR* approximated a similar conclusion by almost figuring out a normal distribution for model parameters and found almost identical parameter' estimates (mean of posterior distribution) compared to the parameters' point estimates given in Koseki and Yamamoto (2007) for the same model structure. However, if a model is still really unsure about the parameters since they distribute in a wide interval (Koehrsen, 2018), it may be narrowed further by changing model hyperparameters.

The information about the posterior distributions for all models was given in Table 2-6. Regarding Eq. (1), all the accuracy, sensitivity and specificity statistics both for training and test data were found greater than 96% at the split ratio of 0.10. It shows that irrespective of the experimental results (being 0 or 1), the model developed using MCMC *BLR* correctly classified the cases with more than 96% accuracy. Moreover, very close goodness of fit values among test and train samples indicated that there was not either an over or under-fitting problem. When the test split ratio was increased from 0.1 to 0.4, any apparent change in the goodness of statistics did not come through. This behaviour went on up to the test split ratio of 0.90. Even though there were very indistinct variations among changing test split ratios, models' prediction performances were almost identical between 0.10-0.90 level for Eq. (1). However, when the split ratio went beyond 0.90, the goodness of fit statistics started to indicate an increasing overfitting problem. This was previously noted in the literature that a not proper prior distribution estimation and a low number of sampling (high number of unknowns) may be the cause of overfitting issues in Bayesian-based regression methods (Pacifico, 2021; van Boekel, 2020). Parallel with decreasing prediction certainty, parameters uncertainties given with 95% credible interval (2.5 and 97.5% percentiles of highest density interval) of posterior distributions were getting wider as the sample size decreased. Although the philosophy behind them is very different, credible interval gives an idea about how confidence interval gives in frequentists approach. Therefore, as it is valid for confidence interval when the sample size is small, the posterior distributions of the parameters (also credible interval) and the predictions become wider. If there is a wider posterior for parameters, it means that one has a less certain conclusion/estimate from the model (du Prel et al., 2009). Thus, in brief, as the test split ratio increased, we got more brittle models from MCMC *BRL*. Despite all these negativities, it can be concluded that MCMC *BLR* correctly classified more than 86% of the TP and FP values of testing set using a non-informative prior distribution and only 10 training data that randomly picked among 1080 observation.

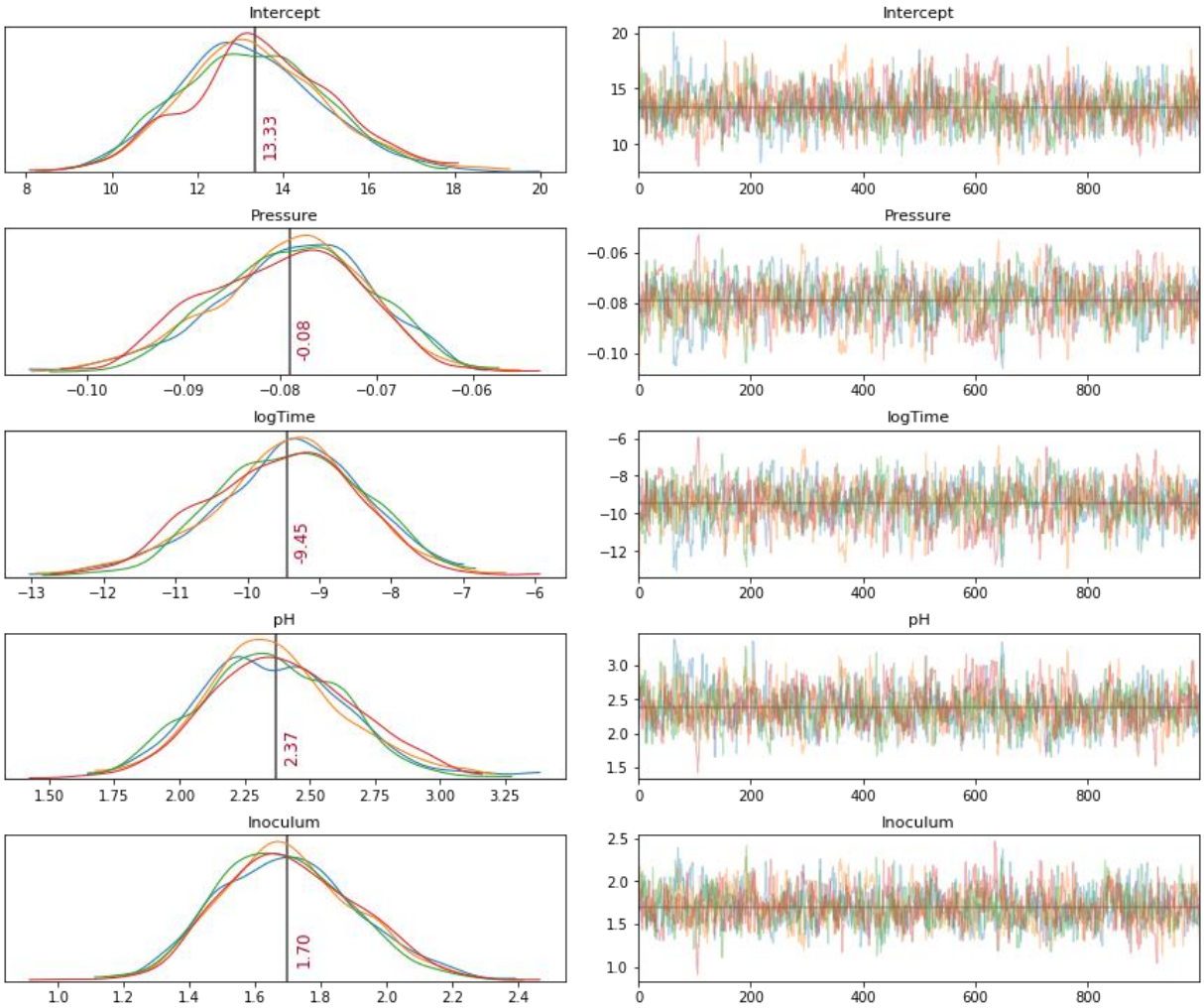

Figure 1. Trace plot and posterior distribution of parameters for Eq. (1) (The final approximate of the posterior distributions for model parameters are on the left-hand side and the individual sample values at each step of samplings are on the right-hand side for the data from Koseki and Yamamoto (2007))

Table 2. Posterior summaries of parameter estimates and model agreement statistics for Eq. (1)

Parameters of Eq. (1)		**Test Split Ratio**											
		0.10		**0.40**		**0.70**		**0.90**		**0.95**		**0.99**	
Intercept (a_0)		13.33 [10.14|16.75]		13.94 [10.07 | 18.11]		16.64 [9.36 | 24.94]		24.99 [6.59 | 46.39]		3501.54 [1416.53 | 5538.06]		11797.12 [-3885.00 | 31847.39]	
P (a_1)		-0.08 [-0.10 | -0.06]		-0.08 [-0.10 | -0.06]		-0.12 [-0.16 | -0.08]		-0.20 [-0.34 | -0.08]		-22.57 [-35.78 | -8.96]		-62.28 [-143.84 | -12.50]	
$log_{10}(t)$ (a_2)		-9.45 [-11.45 | -7.41]		-9.64 [-12.17 | -7.30]		-12.2 [-17.26 | -7.59]		-20.42 [-36.46 | -8.05]		-2216.81 [-3485.78 | -884.41]		-1143.81 [-2352.38 | -85.49]	
pH (a_3)		2.37 [1.82 | 2.90]		2.48 [1.84 | 3.14]		3.53 [2.21 | 5.04]		6.16 [2.30 | 10.79]		552.40 [218.29 | 906.46]		622.50 [12.23 | 1639.54]	
IC (a_4)		1.70 [1.32 | 2.10]		1.82 [1.34 | 2.33]		2.7 [1.71 | 3.92]		4.91 [1.62 | 8.44]		541.41 [237.20 | 876.25]		933.69 [-134.12 | 2010.31]	
		S	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***
CM of training sets	***S***	524	16	346	14	177	3	58	2	30	0	6	0
	D	17	415	13	275	7	137	2	46	0	24	0	4
CM of test sets	***S***	58	2	232	8	396	24	524	16	558	12	558	36
	D	1	47	9	183	14	322	31	401	42	414	111	365
For training sets	**Ac**	96.61		95.83		96.91		96.3		100		100	
	Se	97.04		96.11		98.33		96.67		100		100	
	Sp	96.07		95.49		95.14		95.73		100		100	
For test sets	**Ac**	97.22		96.06		94.97		95.16		94.74		86.26	
	Se	96.67		96.67		94.29		97.04		97.89		93.94	
	Sp	97.92		95.31		95.83		92.82		90.79		76.68	

- a_i values are model coefficients (results are given as most likely estimates [95% credible interval]), P is the applied pressure (MPa), t is the pressure holding time (min), T is the process temperature (°C), IC is the inoculum concentration of the bacteria of interest (log_{10}CFU/ml)
- ***S***:Survival, ***D***:Death, ***Ac***: Accuracy, ***Se***: Sensitivity, ***Sp***: Specificity, ***CM***: Confusion matrix.
- The given colours indicate 50 60 70 80 90 100 levels as percentage.

Regarding Eq. (2), the model developed using the same data with Eq. (1), showed slightly better agreement with training and testing sets for 0.1 and 0.4 test split ratios. Accuracy, sensitivity and specificity statistics of these models were almost within the range of 97-98%. For the 0.70 test split ratio, the goodness of fit statistics for the training data set maintained close to 98% where a 3% reduction was observed for agreement between model predictions and the test set. However, since the difference between model agreement statistics of training and testing data was so small and ignorable, it is worth saying that using only 30% of the experimental data set (324 observation instead of 1080) can give a good estimate of the posterior probability of bacterial survival/death after HPP treatment. However, for greater split ratios, the models started to overestimate the results and reached remarkably wider posterior distributions faster (at higher sampling sizes) than Eq. (1) did. As the test split ratio increased, specificity and accuracy percentages decreased, as well. Not only for Eq. (2) but also for Eq. (1), the most affected goodness of fit criteria from the decrement of the number observation introduced to the model were specificity and accuracy in the decreasing order. On the other hand, sensitivity somehow was not affected by the change of sample size. It means that the models, developed using a very limited number of samples, were highly capable of foreseeing if HPP treatment is sufficient and the bacterial death takes place. However, they have a higher tendency to label survival cases as death (or fail-safe) for high split ratios. These may lead to food safety problems as it may cause the acceptance of unsafe foods as if they were safe. But overall, even the 0.90 split ratios led to acceptable prediction accuracy from the point of tests' prediction perspective. On the other hand, parameter removal can improve the performance of Eq. (2) and it may also be beneficial to solve the overfitting problem at high test split ratios. As it is well-known, model complexity is the main source of overestimating problems (James et al., 2013).

Table 3. Posterior summaries of parameter estimates and model agreement statistics for Eq. (2)

Parameters of Eq. (2)	**Test Split Ratio**					
	0.10	**0.40**	**0.70**	**0.90**	**0.95**	**0.99**
Intercept (a_0)	50.99 [28.99 | 71.19]	40.84 [17.25 | 66.66]	106.64 [39.69 | 182.99]	9551.77 [3742.87 | 14664.53]	14589.76 [572.90 | 117.97]	20567.35 [-31738.04 | 123821.62]
P (a_1)	-0.27 [-0.40 | -0.14]	-0.1 [-0.37 | -0.06]	-0.39 [-0.82 | 0.02]	-52.02 [-93.28 | 0.002]	1.78 [-105.23 | 2115.42]	-93.59 [-275.17 | 21.94]
t (a_2)	-1.61 [-2.74 | -0.81]	-1.14 [-2.14 | -0.10]	-1.78 [-4.94 | 1.73]	711.50 [-64.56 | 1614.56]	265.40 [-540.40 | 2115.42]	130.54 [-767.91 | 1223.20]
pH (a_3)	-1.90 [-5.83 | 1.46]	-3.43 [-8.13 | 1.26]	-12.35 [-23.87 | -1.96]	685.11 [-1459.07 | 2322.94]	328.14 [-1190.71 | 1577.80]	-2.39 [-1588.52 | 2084.44]
IC (a_4)	4.77 [1.54 | 8.14]	4.86 [0.64 | 8.99]	4.33 [-7.80 | 16.85]	-413.88 [-1978.28 | 1184.28]	-40.06 [-965.90 | 783.25]	436.39 [-1497.10 | 2474.87]
$P \cdot t$ (a_5)	0.00 [-0.002 | 0.003]	-0.001 [-0.01 | 0.002]	-0.01 [-0.02 | 0.004]	-6.54 [-11.16 | -2.44]	-8.05 [-13.5 | -2.51]	-10.39 [-32.94 | -1.19]
$P \cdot pH$ (a_6)	-0.01 [-0.02 | 0.01]	-0.00 [-0.02 | 0.02]	0.01 [-0.04 | 0.05]	0.23 [-2.71 | 5.25]	3.43 [-5.86 | 13.52]	1.13 [-7.49 | 15.22]
$P \cdot IC$ (a_7)	-0.02 [-0.03 | -0.004]	-0.02 [-0.03 | -0.003]	0.00 [-0.05 | 0.05]	4.45 [0.44 | 8.53]	-0.47 [-5.17 | 5.88]	16.79 [3.10 | 39.52]
$t \cdot pH$ (a_8)	0.13 [0.02 | 0.25]	0.14 [-0.003 | 0.31]	0.58 [0.05 | 1.13]	40.00 [-77.35 | 136.98]	36.22 [-181.20 | 233.94]	333.07 [-380.41 | 1711.29]
$t \cdot IC$ (a_9)	-0.09 [-0.16 | -0.02]	-0.10 [-0.21 | -0.002]	0.103 [-0.25 | 0.44]	23.53 [-33.34 | 92.76]	-12.59 [339.00 | 465.95]	-271.10 [-1118.02 | 720.22]
$pH \cdot IC$ (a_{10})	0.09 [-0.38 | 0.49]	0.22 [-0.39 | 0.86]	-0.23 [-1.75 | 1.22]	119.57 [-146.31 | 476.125]	120.92 [-562.26 | 542.77]	605.82 [-216.03 | 2516.81]
P^2 (a_{11})	0.00 [0.00 | 0.001]	0.00 [-0.00 | 0.001]	0.00 [-0.00 | 0.001]	-0.05 [-0.11 | 0.04]	-0.25 [-0.43 | -0.01]	-0.25 [-0.54 | -0.03]
t^2 (a_{12})	0.02 [0.01 | 0.03]	0.01 [-0.002 | 0.03]	-0.01 [-0.07 | 0.05]	17.42 [4.07 | 28.71]	25.84 [-20.87 | 60.58]	-103.95 [-605.40 | 274.51]
pH^2 (a_{13})	0.57 [0.15 | 1.01]	0.49 [-0.07 | 1.04]	1.42 [0.21 | 2.81]	200.05 [46.06 | 4995.45]	123.41 [-81.13 | 385.90]	442.05 [-338.22 | 1332.83]
IC^2 (a_{14})	0.35 [0.04 | 0.69]	0.42 [-0.04 | 0.87]	0.27 [-0.86 | 1.44]	58.61 [-66.65 | 182.98]	435.68 [172.72 | 983.18]	-67.18 [-899.08 | 985.58]

		S	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***
CM of training sets	***S***	529	11	554	6	176	4	60	0	30	0	6	0
	D	14	418	10	278	3	141	0	48	0	24	0	4
CM of test sets	***S***	59	1	236	4	400	20	513	27	555	15	579	15
	D	1	47	5	187	15	321	33	399	48	408	159	317
For training sets	**Ac**	97.43		98.11		97.84		100		100		100	
	Se	97.96		98.93		97.78		100		100		100	
	Sp	96.76		96.53		97.92		100		100		100	
For test sets	**Ac**	98.15		97.92		95.37		93.83		93.86		83.74	
	Se	98.33		98.33		95.24		95		97.37		97.47	
	Sp	97.92		97.4		95.54		92.36		89.47		66.6	

- a_i values are model coefficients (results are given as "most likely estimates [95% credible interval]"), P is the applied pressure (MPa), t is the pressure holding time (min), T is the process temperature (°C), IC is the inoculum concentration of the bacteria of interest ($\log_{10}$CFU/ml)
- ***S***:Survival, ***D***:Death, ***Ac***: Accuracy, ***Se***: Sensitivity, ***Sp***: Specificity, ***CM***: Confusion matrix.
- The given colours indicate 50 60 70 80 90 100 levels as percentage.

Eqs. (4)-(6) were developed for the data taken from Koseki et al. (2009). The reason for applying the same modelling procedure for similar datasets was the difference between the distribution balance of them. As previously noted, the ratio of 0/1 (death/survival) is 55.56% for the data from Koseki and Yamamoto (2007) which is well balanced whereas it is only 33.11% for the data from Koseki et al. (2009). As known, an imbalanced dataset may lead to biased parameter estimates and classification performance of a logistic regression model (Rahman et al., 2020). So in the present study, the performance of Bayesian-based logistic regression was also conducted once more to observe whether it is capable of handling moderately imbalanced bacterial survival data at various sampling sizes. At this point, it should also be noted that the performance of a model is not only dependent on the methods but also affected by various factors which of one is the data itself. But in our case, since the models in the original papers predicted quite well for both data, it was thought that performing a comparison using different data sets with changing case prevalence was appropriate.

Table 4. Posterior summaries of parameter estimates and model agreement statistics for Eq. (3)

Parameters of Eq. (3)		**Test Split Ratio**											
		0.10		**0.40**		**0.70**		**0.90**		**0.95**		**0.99**	
Intercept (a_0)		19.14 [14.99 | 23.72]		16.68 [12.13 | 21.62]		20.18 [12.10 | 28.27]		2675.88 [528.57 | 4324.96]		3725.6 [665.98 | 7608.41]		-7325.47 [-15996.09 | 1915.64]	
P (a_1)		-0.05 [-0.05 | -0.04]		-0.4 [-0.05 | -0.03]		-0.05 [-0.07 | -0.03]		-6.33 [-10.55 | -1.47]		-12.96 [-23.71 | -3.52]		5.46 [-11.51 | 26.43]	
$ln(t)$ (a_2)		-3.53 [-4.21 | -2.87]		-3.45 [-4.25 | -2.59]		-3.75 [-4.92 | -2.60]		-499.42 [-864.84 | -154.08]		-1068.63 [-1890.85 | -389.22]		-1493.99 [-2445.49| -421.73]	
T (a_3)		-0.28 [-0.34 | -0.21]		-0.26 [-0.34 | -0.19]		-0.27 [-0.38 | -0.15]		-32.49 [-54.87 | -7.75]		-29.67 [-92.13 | 21.58]		48.71 [-242.68 | 266.66]	
IC (a_4)		2.03 [1.61 | 2.42]		1.91 [1.48 | 2.39]		2.10 [1.44 | 2.87]		243.68 [57.84 | 410.62]		647.07 [206.43 | 1213.23]		657.36 [-442.08 | 1794.71]	
M (a_5)		8.29 [6.80 | 9.80]		7.69 [5.99 | 9.47]		-9.97 [6.90 | 13.12]		1021.38 [371.79 | 2064.88]		2091.38 [842.96 | 3281.05]		1335.80 [138.40 | 2784.19]	
		S	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***
CM of training sets	***S***	513	29	337	24	175	6	60	0	30	0	6	0
	D	29	239	18	161	10	79	2	28	0	15	0	3
CM of test sets	***S***	58	2	226	15	405	16	525	17	544	28	516	80
	D	1	29	5	114	29	180	40	228	53	230	150	145
For training sets	**Ac**	92.84		92.22		94.07		97.78		100		100	
	Se	94.65		93.35		96.69		100		100		100	
	Sp	89.18		89.94		88.76		93.33		100		100	
For test sets	**Ac**	96.67		94.44		92.86		92.96		90.53		74.19	
	Se	96.67		93.78		96.2		96.86		95.1		86.58	
	Sp	96.67		95.8		86.12		85.07		81.27		49.15	

- a_i values are model coefficients (results are given as "most likely estimates [95% credible interval]"), P is the applied pressure (MPa), t is the pressure holding time (min), T is the process temperature (°C), IC is the inoculum concentration of the bacteria of interest (log_{10}CFU/ml)
- ***S***: Survival, ***D***:Death, ***Ac***: Accuracy, ***Se***: Sensitivity, ***Sp***: Specificity, ***CM***: Confusion matrix.
- The given colours indicate 50 60 70 80 90 100 levels as percentage.

Table 5. Posterior summaries of parameter estimates and model agreement statistics for Eq. (4)

Parameters of Eq. (4)		**Test Split Ratio**											
		0.10		**0.40**		**0.70**		**0.90**		**0.95**		**0.99**	
Intercept (a_0)		12.30 [1.92 | 23.49]		10.32 [-2.433 | 23.94]		20.58 [-1.80 | 42.24]		-806.19 [-969.87 | 2577.42]		6276.49 [-2120.08 | 13597.42]		-9516.54 [-25288.56 | 6433.26]	
P (a_1)		-0.03 [-0.05 | -0.01]		-0.03 [-0.05 | -0.00]		-0.05 [-0.10 | -0.01]		-0.86 [-7.51 | 4.38]		-17.36 [-31.58 | -0.09]		15.18 [-20.32 | 54.00]	
$ln(t)$ (a_2)		-3.55 [-4.29 | -2.91]		-3.50 [-4.32 | -2.72]		-3.80 [-5.11 | -2.63]		-735.51 [-971.41 | -433.91]		-1004.82 [-1559.24 | -520.316]		-1434.74 [-2511.71 | -307.97]	
T (a_3)		-0.05 [-0.38 | 0.29]		-0.05 [-0.46 | 0.35]		-0.27 [-0.94 | 0.35]		92.17 [-15.29 | 192.82]		-114.38 [-334.26 | 111.26]		139.47 [-224.77 | 713.98]	
IC (a_4)		2.04 [1.61 | 2.43]		1.94 [1.47 | 2.42]		2.13 [1.36 | 2.89]		375.07 [227.27 | 502.84]		622.27 [260.53 | 951.71]		519.30 [-506.30 | 2053.55]	
M (a_5)		8.31 [6.78 | 9.99]		7.77 [6.00 | 9.58]		10.12 [6.80 | 13.60]		1625.54 [1001.57 | 2220.34]		2057.71 [1071.13 | 3100.93]		1253.74 [93.89 | 2766.49]	
$P \cdot T$ (a_7)		-0.00 [-0.001 | 0.00]		-0.00 [-0.001 | 0.00]		0.00 [-0.001 | 0.001]		-0.25 [-0.44 | -0.06]		0.15 [-0.26 | 0.59]		-0.27 [-1.41 | 0.58]	
		S	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***
CM of training sets	***S***	516	26	338	23	175	6	59	1	30	0	6	0
	D	25	243	19	160	9	80	2	28	0	15	0	3
CM of test sets	***S***	56	4	230	11	405	16	522	20	546	26	488	108
	D	1	29	10	109	32	177	49	219	42	241	115	180
For training sets	**Ac**	93.7		92.22		94.44		96.67		100		100	
	Se	95.2		93.63		96.69		98.33		100		100	
	Sp	90.67		89.29		89.89		93.33		100		100	
For test sets	**Ac**	94.44		94.17		92.38		91.48		92.05		74.97	
	Se	93.33		95.44		96.2		96.31		95.45		81.88	
	Sp	96.67		91.6		84.69		81.72		85.16		61.02	

- a_i values are model coefficients (results are given as "most likely estimates [95% credible interval]"), P is the applied pressure (MPa), t is the pressure holding time (min), T is the process temperature (°C), IC is the inoculum concentration of the bacteria of interest (log_{10}CFU/ml)
- ***S***: Survival, ***D***:Death, ***Ac***: Accuracy, ***Se***: Sensitivity, ***Sp***: Specificity, ***CM***: Confusion matrix.
- The given colours indicate 50 60 70 80 90 100 levels as percentage.

Table 6. Posterior summaries of model parameter estimates and model agreement statistics for Eq. (5)

Parameters of Eq. (5)	Test Split Ratio					
	0.10	**0.40**	**0.70**	**0.90**	**0.95**	**0.99**
Intercept (a_0)	19.04 [-1701.74 | 1842.36]	516.46 [-1358.73 | 231667]	-1204.25 [-3460.17 | 567.69]	1546.76 [-1468.81 | 9007.78]	3260.69 [-4571.08 | 10062.48]	124904.24 [-514247.5 | 2297439.9]
P (a_1)	0.00 [-0.13 | 0.11]	0.06 [-0.04 | 0.15]	-0.24 [-0.53 | 0.01]	2.09 [-4.88 | 11.75]	14.29 [-47.51 | 61.14]	-127.80 [-1768.90 | 1506.74]
t (a_2)	-0.39 [-1.97 | 1.65]	0.69 [-0.29 | 1.53]	-1.26 [-4.60 | 1.57]	-629.61 [-1031.17 | -287.73]	211.38 [1659.27 | 2171.30]	-5.45 [-1405.36 | 830.20]
T (a_3)	-1.63 [-119.60 | 110.12]	-34.67 [-151.52 | 86.89]	82.47 [-27.96 | 236.60]	157.20 [-46.56 | 474.95]	166.01 [-68.75 | 820.31]	-89.92 [-1658.70 | 1377.97]
IC (a_4)	4.16 [-1.21 | 10.02]	2.44 [-0.13 | 7.89]	0.39 [-9.39 | 11.61]	725.35 [-201.82 | 1672.02]	377.11 [-550.76 | 1451.28]	-500.44 [-2793.53 | 1361.02]
M (a_5)	19.19 [-5.61 | 37.50]	9.95 [-5.18 | 23.02]	39.21 [0.81 | 72.89]	87.18 [-1149.62 | 1309.29]	683.99 [-195.29 | 2444.31]	6.94 [-1480.20 | 1784.24]
$P \cdot t$ (a_6)	-0.002 [-0.01 | 0.001]	-0.004 [-0.01 | -0.003]	-0.01 [-0.01 | 0.001]	-0.53 [-1.47 | 0.08]	-12.81 [-20.38 | -7.49]	-776.88 [-2566.53 | 5.83]
$P \cdot T$ (a_7)	0.00 [-0.001 | 0.002]	-0.00 [-0.002 | 0.001]	0.002 [-0.001 | 0.01]	-0.65 [-1.04 | -0.35]	-1.03 [-3.23 | 0.88]	86.87 [-95.14 | 396.96]
$P \cdot IC$ (a_8)	-0.01 [-0.02 | 0.01]	-0.001 [-0.01 | 0.01]	0.01 [-0.01 | 0.02]	-0.85 [-3.40 | 1.23]	9.41 [0.48 | 27.73]	-81.64 [-1288.78 | 263.91]
$P \cdot M$ (a_9)	-001 [-0.04 | 0.04]	0.005 [-0.03 | 0.03]	-0.01 [-0.08 | 0.06]	1.51 [-3.15 | 7.13]	11.65 [-30.08 | 48.99]	660.90 [-65.92 | 2062.86]
$t \cdot T$ (a_{10})	0.00 [-0.02 | 0.02]	-0.01 [-0.03 | +0.01]	-0.2 [-0.01 | 0.05]	-6.87 [-19.55 | 6.63]	-1.83 [-17.22 | 5.75]	-529.77 [-2677.45 | 556.56]
$t \cdot IC$ (a_{11})	-0.36 [-0.14 | 0.09]	0.01 [-0.11 | 0.12]	0.15 [-0.05 | 0.37]	42.49 [-10.72 | 111.82]	341.69 [-74.98 | 1320.75]	-210.41 [-1593.44 | 1212.26]
$t \cdot M$ (a_{12})	0.17 [-0.24 | 0.77]	0.36 [-0.03 | 0.88]	0.85 [-0.19 | 1.93]	134.17 [-85.67 | 398.57]	238.14 [-919.56 | 15995.39]	-215.62 [-2062.35 | 1207.76]
$T \cdot IC$ (a_{13})	-0.05 [-0.11 | 0.01]	-0.05 [-0.12 | 0.002]	0.02 [-0.10 | 0.12]	-4.08 [-16.87 | 15.37]	-8.07 [-117.59 | 62.42]	57.74 [-1525.81 | 2164.48]
$T \cdot M$ (a_{14})	0.06 [-0.15 | 0.31]	0.04 [-0.18 | 0.30]	-0.02 [-0.42 | 0.44]	105.57 [17.93 | 224.93]	512.82 [-109.06 | 1316.83]	238.67 [-1165.81 | 2407.44]
$M \cdot IC$ (a_{15})	-1.18 [-2.91 | 0.39]	-1.15 [-2.68 | 0.64]	-3.30 [-7.22 | 0.94]	19.39 [-187.82 | 331.72]	437.09 [-1404.80 | 2912.93]	112.89 [-1193.40 | 967.14]
P^2 (a_{16})	-0.00 [0.00 | 0.00]	-0.00 [-0.00 | 0.00]	0.00 [-0.00 | 0.00]	-0.003 [-0.02 | 0.01]	-0.09 [-0.25 | 0.02]	3.29 [-17.68 | 35.31]
T^2 (a_{17})	0.03 [0.01 | 0.06]	0.02 [0.01 | 0.05]	0.07 [0.02 | 0.12]	16.38 [5.02 | 26.51]	22.54 [-239.30 | 327.03]	-236.37 [-2530.95 | 1994.51]
t^2 (a_{18})	0.02 [-1.69 | 1.83]	0.53 [-1.34 | 2.32]	-1.29 [-3.57 | 0.51]	0.39 [-2.35 | 4.59]	-2.31 [-10.47 | 4.26]	-298.12 [-2304.11 | 994.74]
IC^2 (a_{19})	0.35 [0.05 | 0.54]	0.28 [0.05 | 0.5]	0.12 [-0.42 | 0.70]	60.78 [-2.59 | 210.25]	-11.07 [-69.85 | 3.33]	324.25 [-1446.21 | 2578.79]

		S	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***	***S***	***D***
CM of training sets	***S***	524	18	346	15	176	5	60	0	30	0	6	0
	D	22	246	12	167	8	81	2	28	0	15	0	3
CM of test sets	***S***	57	3	227	14	401	20	524	18	540	32	371	225
	D	2	28	3	116	28	181	44	224	57	226	97	198
For training sets	**Ac**	95.06		95		95.19		97.78		100		100	
	Se	96.68		95.84		97.24		100		100		100	
	Sp	91.79		93.3		91.01		93.33		100		100	
For test sets	**Ac**	94.44		95.28		92.38		92.35		89.59		63.86	
	Se	95		94.19		95.25		96.68		94.41		62.25	
	Sp	93.33		97.48		86.6		83.58		79.86		67.12	

- a_i values are model coefficients (results are given as "most likely estimates [95% credible interval]"), P is the applied pressure (MPa), t is the pressure holding time (min), T is the process temperature (°C), IC is the inoculum concentration of the bacteria of interest ($\log_{10}$CFU/ml)
- ***S***: Survival, ***D***:Death, ***Ac***: Accuracy, ***Se*****:** Sensitivity**,** ***Sp*****:** Specificity**,** ***CM*****:** Confusion matrix**.**
- The given colours indicate 50 60 70 80 90 100 levels as percentage.

The posterior information of the simplest model (Eq. (3)) for this moderately imbalanced data set was given in Table 4. When the test split ratio was set as 0.10, the goodness of fit statistics for the test set was calculated as a little higher than their training counterparts. This may be attributed to the imbalance of case prevalence of the used data set. Because it brings to mind that the test data set might consist of easy to predict (more compatible with the model) examples than the training set. For such a case, model parameters can be changed, test and training data can be redistributed or other validation methods such as cross-validation might be employed to use all possible subsets of the data for both training and testing. Similar to the previous model, as the test splitting ratios increased the model started to overestimate the results and the model uncertainty increased gradually. But still relatively acceptable prediction accuracy was obtained up to 0.9-0.95 split level. The addition of $P \cdot T$ parameter to Eq. (3) gave Eq. (4). When the results were analysed, this change did not result in a big difference in the final model accuracy except a slightly better-balanced goodness of fit statistics appearing between training and test. However, the mentioned differences were small and the models could still be considered as acceptable. Since the $P \cdot T$ parameter in Eq. (4) had so small values of coefficients for different splitting ratios, it had a very negligible contribution to the model prediction comparing to the other model parameters. A similar situation present for Eq. (5), as well and some of the model parameters were frivolous. So using model simplification methods such as comparison of WAIC (Widely-applicable Information Criterion) values or LOO (Leave-one-out cross-validation) of

different models, the best possible parameter subset of the models can be chosen (Vehtari et al., 2017). A similar procedure was carried out by Koseki et al. (2009) and some of the parameters were removed from Eq. (5) to obtain minimum AIC (Akaike's Information Criterion). However, since the main purpose of the present paper is to assess the potential use of MCMC *BLR* and its reaction against low sampling sizes and changing case prevalence, a parameter elimination practice was not followed. Regarding the effect of increased test split ratios, model behaviour is almost identical for Eqs. (3)-(5) with Eqs. (1)-(2).

To compare the effects of different model parameters on model output, there are several alternatives. The most popular and the easiest one is the comparison of the "Odds Ratio" of the model parameters and their uncertainties. The odds ratio is a popular statistic for likelihood approximation and it simply tells us that one unit of increase in an explanatory variable changes the expected likelihood results at the level of one odds ratio times. Therefore, if the odds ratio of a parameter is equal to one, this means that two outputs (death or survival) are independent of that parameter (Bland and Altman, 2000; UCLA, 2016). On the other hand, when the odds ratio is between 0-1, this means that there is a negative correlation between the parameter and expected case compared to the reference, and vice versa for an odds ratio that is greater than 1 (Szumilas, 2010). For instance, the odds ratio of pressure term for Eq. (1) was calculated as 0.96 to expect survival of bacteria after HPP treatment. It means that every unit of increase of the pressure is likely to decrease the survival probability of the bacteria 0.96 times.

Rather than odds ratio, posterior predictive regression lines allow us to visually inspect the most likely estimate of model predictions and its distribution versus varying levels of models' explanatory variables (Fig. 2). In these figures, 1000 different curves for varying levels of the inspected parameters were given where the mean values of the other parameters were used. For example, in Fig. 2(a), pressure values were changed between 200-500 MPa but the rest of the model parameters were kept at their means which were $pH = 5$, $IC = 5$ $\log_{10}$CFU/ml and $log_{10}(t) = 0.83$ min. Moreover, where it was necessary, some plots were drawn by extrapolating the model to an acceptable level out of the experimental range to provide a clearer view of the trend for posterior predictive. In those figures, the black dashed lines display the most likely estimate of the prediction which were drawn using coefficients of parameters given in Table 2. Since the red curves correspond to 1000 different curves for each level of a given parameter from the posterior distribution, they seem blurry and the saturation of the colour gives an idea about the frequency and distribution of the predictions. Moreover, 0.1, 0.5 and 0.9 probability (or 10, 50 and 90% probability) for achieving bacterial inactivation were marked. The region of probability <0.5 can be assigned as "likely to achieve the target log reduction" (Buzrul, 2019). From the figures, it can be concluded that there was a negative correlation between P, $log_{10}(t)$ and survival probability of *L. monocytogenes*. On the other hand, lower values of pH and IC lead to a higher probability of bacterial log reduction. For the values approximately $P > 325$ MPa, $log_{10}(t) > 0.65$ min, $pH < 0.57$ and $IC < 6$ $\log_{10}$CFU/ml, inactivation probability of bacteria was greater than 0.5. Moreover, for the values $P > 360$ MPa, $log_{10}(t) > 0.85$ min, $pH < 4.75$ and $IC < 4.70$ $\log_{10}$CFU/ml, achieving desired bacterial log reduction probability was greater than 0.9 for the given levels of the other parameters. As an advantage of the probabilistic modelling, given probabilities in Fig. 2, are also valid for the levels of the parameters higher/lower than the studied range (Buzrul, 2019). For instance, the effects of $P > 360$ MPa were also distinguishable. As an advantage of the *BLR*, the distribution of the output uncertainty around the most likelihood curve was also visible. As it was apparent from the trace plot of parameters (Fig. 1), posterior distributions were unimodal and quite symmetric. As a result, the predictive posterior distribution of survival probability from Eq. (1) also showed a similar distribution as seen from Fig. 2. However, for the given parameter levels, posterior predictive distributions were a little wider for higher values of pH, IC and lower values of $log_{10}(t)$. For all the mentioned levels, bacteria was more prone to survive. So, the producers need to pay more attention for these more uncertain regions in their processes to avoid from possible food safety problems.

3.3. Perspective

Probabilistic modelling is an emerging technique in statistical learning of bacterial survival/death or growth/no growth. Among them, *LR* is very popular and almost the standard method to analyse binary and ordered categorical outcome data. However, the classical frequentist approach (the maximum likelihood estimation) to determine point estimates of *LR* requires large sample arguments and their performance is often problematic for small or moderate volume size of data sets (O'Brien and Dunson, 2004). Bayesian modelling, which is an important sub-discipline of the probabilistic approach is suitable for such cases. It has several important advantages over likelihood-based frequentist methods for analysing multivariate categorical data. Not all but the most important advantages of *BLR*, especially for MCMC algorithm, are: [1] there is no need for large sample data sets, [2] it can handle the existence of missing data, [3] it is easy to implement with the use of available libraries like PyMC3 and [4] it gathers realistic and not fixed information about the distribution and uncertainty of predictive outputs and model parameters (O'Brien and Dunson, 2004; Salvatier et al., 2016; van Boekel, 2020).

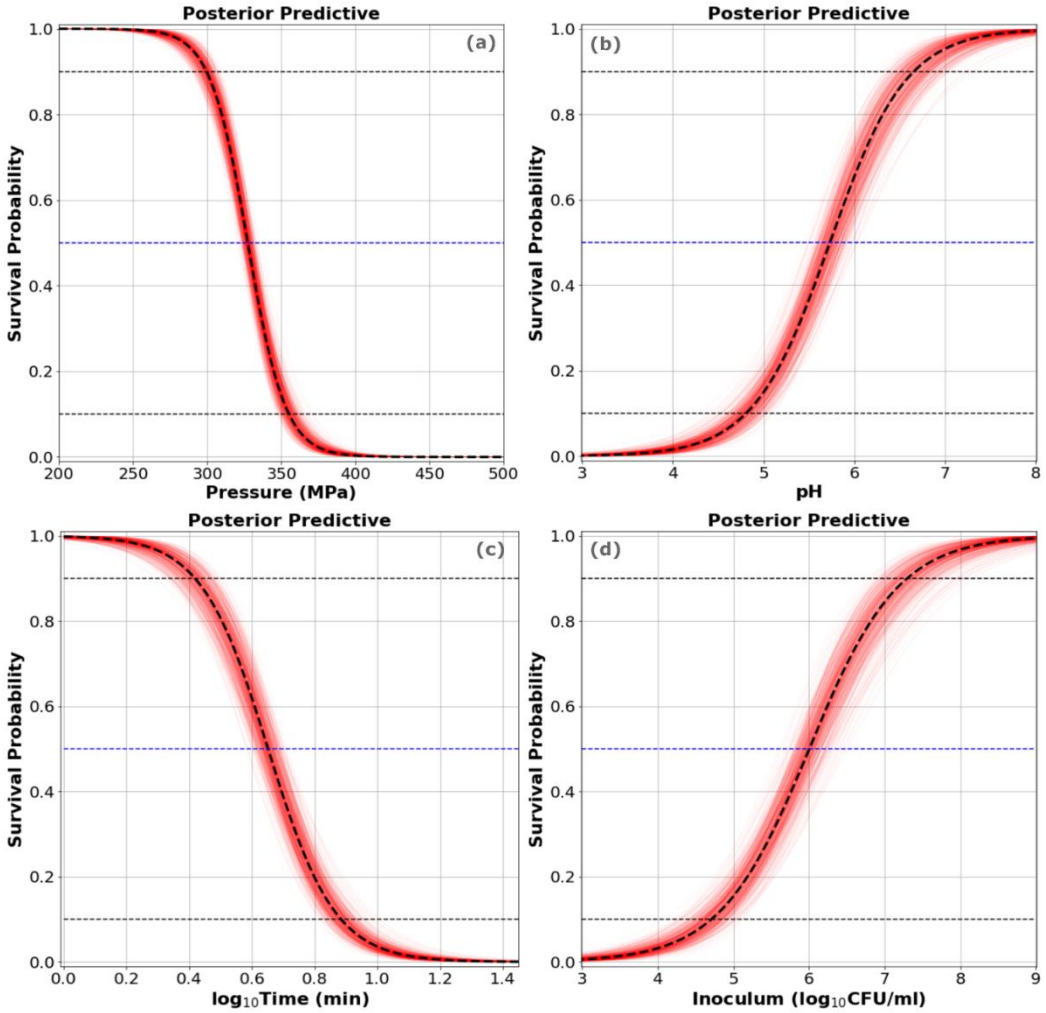

Fig. 2. Posterior predictive distributions for *L. monocytogenes* (for the data from Koseki and Yamamoto (2007)) using Eq. (1) for varying levels of (a) pressure (P), (b) pH, (c) $\log_{10}$Time ($log_{10}(t)$) and (d) inoculum (IC) (for all plots, the constant process variables are $P = 350$ MPa, $pH = 5$, $IC = 5$ $\log_{10}$CFU/ml and $log_{10}(t) = 0.83$ min except the one given for each where survival and death cases were coded as 1 and 0, respectively)

Compared to reference studies (Koseki et al. (2009) and Koseki and Yamamoto (2007)), MCMC *BLR* can produce satisfying results even for test data sets using 90 and 108 number of experimental data (test split ratio of 0.90) instead of 900 and 1080, respectively, for Koseki et al. (2009) and Koseki and Yamamoto (2007). Compared to the per cent concordance statistics (the equivalent of per cent accuracy calculated in the present study) given in those references, MCMC *BLR* can produce similar model agreements for both training and test comparison (note that the given concordance results in the reference studies were calculated using all data set as training and leave one out cross-validation was also applied only by Koseki et al. (2009)). Furthermore, the use of 45 and 54 data points (test split ratio of 0.95) still provides a fair amount of knowledge to evaluate the processes although the misclassification rates slightly increased. When the time and cost savings with the reduction of experimental works are considered, a small increase in the misclassification ratios might be acceptable. However, when the data size reduced to 9 and 10, an obvious overfitting problem arose which may lead producers to give wrong decisions. So, it can be concluded that MCMC *BLR* can be used with a smaller number of experimental data (about 50 experimental sample) to produce a relatively reliable probability models. Although it did not conduct in the present study, elimination of weakly or not contributing parameters may further improve the models' prediction accuracy. On the other hand, although a good convergence was obtained for given models during MCMC sampling, using a more informative prior may improve model performance instead of a non-informative one. If someone already has an idea about the parameter distributions, posteriors can take shape and credible interval can narrow faster. And the use of a weakly informative prior is always suggested since it prevents Bayesian regressions from overfitting (McElreath, 2016), and has a stabilising effect on MCMC simulation (Korner-Nievergelt et al., 2015). A detailed guideline about covering the tips about prior selection was given by van Boekel (2020). Being like it was done in the current study, for scientists non-experienced in *BLR*, a non-informative prior may work when the number of tuning iterations were increased which is very easy with a cost of more computation time. Using tuning samples, a stationary prior is automatically supplied for a speed-up calculation of expectations. In the PyMC3

library, the tuning samples were discarded since it would be theoretically shaky to suggest that they came from the stationary distribution at all (Carroll, 2019).

Since the uncertainty distribution for posterior predictive is available with *BLR*, manufacturers can consider and pay more attention to the levels of process parameters separately or in combination for which high deviation exists. In addition to high uncertainty, the posterior distribution of parameter estimates and the model prediction may have a non-symmetric distribution skewing left or right. Having a full perspective about posterior distributions of parameters and estimates strengthen manufacturers hands against food safety risks allowing them to foresee unusual deviations of the system.

Contrary to its ease of implementation, calculation of expectations by MCMC *BLR* may take hours especially if the dataset is big and the computational sources are limited. However, it can be tolerable when the amount of time possibly saved from the experimental workload is considered. Moreover, with the use of more powerful sources like Google Colab or high-performance computing services like TRUBA which is provided by TUBITAK ULAKBIM, calculation times can be reduced to minutes. No doubt, this makes the use of MCMC *BLR* more attractive.

4. Conclusion

This study indicates that Bayesian Logistic Regression can be a useful tool to describe the survival/death probability of microorganisms after high-pressure processes with lower experimental data requirement (with about 50 experimental samples) than the frequentist approach and also with the ability to handle missing observation and imbalanced dataset. Although weakly informative priors of parameters are generally required for a better and faster convergence, with the help of well developed and documented libraries like PyMC3 for Python, both coding and implementation of Markov Chain Monte Carlo Bayesian Logistic Regression is easily applicable and can produce satisfactory predictions even with the use of non-informative priors. Since Bayesian Logistic Regression produces posterior distributions for parameters and predictions, more realistic information about the model uncertainty is possible to obtain instead of the presumed and fixed distributions followed in the frequentist approach. Moreover, with the use of free high computational sources, the iterations of the Markov Chain Monte Carlo can be completed only in minutes and Bayesian approaches become more attractive. Therefore, not only for high-pressure processes but also for other food operations, the design of new studies requiring a smaller number of experiments to save time and costs may be possible for researchers. Besides, a more detailed risk assessment paying attention to changing and possibly non-symmetric distributions for varying levels of process parameters alone or in combination may be feasible for the food industry.

Acknowledgement

I would like to thank Prof. Shige Koseki (Food and Agricultural Process Engineering, Hokkaido University, Japan) for sharing the experimental data and giving permission to reuse it.

Conflict of Interest

No conflict of interest was declared by the author.

References

Baratloo, A., Hosseini, M., Negida, A., El Ashal, G., (2015). Part 1: Simple definition and calculation of accuracy, sensitivity and specificity. Emergency 3(2), 48-49.

Bilbao-Sáinz, C., Younce, F.L., Rasco, B., Clark, S., (2009). Protease stability in bovine milk under combined thermal-high hydrostatic pressure treatment. Innovative Food Science & Emerging Technologies 10(3), 314-320.

Bland, J.M., Altman, D.G., (2000). The odds ratio. BMJ 320(7247), 1468.

Butz, P., Fernández García, A., Lindauer, R., Dieterich, S., Bognár, A., Tauscher, B., (2003). Influence of ultra high pressure processing on fruit and vegetable products. Journal of Food Engineering 56(2), 233-236.

Buzrul, S., (2014). Multi-pulsed high hydrostatic pressure inactivation of microorganisms: A review. Innovative Food Science & Emerging Technologies 26, 1-11.

Buzrul, S., (2017). Evaluation of different dose-response models for high hydrostatic pressure inactivation of microorganisms. Foods 6(9), 79.

Buzrul, S., (2019). High hydrostatic pressure inactivation of microorganisms: A probabilistic model for target log-reductions. International Journal of Food Microbiology 309, 108330.

Buzrul, S., Alpas, H., (2004). Modeling the synergistic effect of high pressure and heat on inactivation kinetics of Listeria innocua: a preliminary study. FEMS Microbiology Letters 238(1), 29-36.

Buzrul, S., Alpas, H., Bozoglu, F., (2005). Use of Weibull frequency distribution model to describe the inactivation of Alicyclobacillus acidoterrestris by high pressure at different temperatures. Food Research International 38(2), 151-157.

Buzrul, S., Alpas, H., Largeteau, A., Demazeau, G., (2008). Modeling high pressure inactivation of Escherichia coli and Listeria innocua in whole milk. European Food Research and Technology 227(2), 443-448.
Campus, M., (2010). High pressure processing of meat, meat products and seafood. Food Engineering Reviews 2(4), 256-273.
Carroll, C., (2019). Pragmatic probabilistic programming: Parameter adaptation in PyMC3.
du Prel, J.-B., Hommel, G., Röhrig, B., Blettner, M., (2009). Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. Deutsches Arzteblatt international 106(19), 335-339.
Ferrari, G., Maresca, P., Ciccarone, R., (2010). The application of high hydrostatic pressure for the stabilization of functional foods: Pomegranate juice. Journal of Food Engineering 100(2), 245-253.
Gudbjornsdottir, B., Jonsson, A., Hafsteinsson, H., Heinz, V., (2010). Effect of high-pressure processing on Listeria spp. and on the textural and microstructural properties of cold smoked salmon. LWT - Food Science and Technology 43(2), 366-374.
Hajmeer, M., Basheer, I., (2002). A probabilistic neural network approach for modeling and classification of bacterial growth/no-growth data. Journal of Microbiological Methods 51(2), 217-226.
Hite, B.H., (1899). The effects of pressure in the preservation of milk. West Virginia Agricultural and Forestry Experiment Station Bulletins 58, 15-35.
Hokmollahi, F., Ehsani, M., (2017). High pressure processing and its application in cheese manufacturing: a review. Journal of Food Biosciences and Technology 7, 57-66.
James, G., Witten, D., Hastie, T., Tibshirani, R., (2013). An Introduction to statistical learning with applications in R. Springer-Verlag New York.
Khan, M.A., Ali, S., Abid, M., Ahmad, H., Zhang, L., Tume, R.K., Zhou, G., (2014). Enhanced texture, yield and safety of a ready-to-eat salted duck meat product using a high pressure-heat process. Innovative Food Science & Emerging Technologies 21, 50-57.
Koehrsen, W., (2018). Bayesian linear regression in python: using machine learning to predict student grades part 2. Medium.
Korner-Nievergelt, F., Roth, T., Von Felten, S., Guélat, J., Almasi, B., Korner-Nievergelt, P., (2015). Bayesian data analysis in ecology using linear models with R, BUGS, and Stan. Academic Press.
Koseki, S., Matsubara, M., Yamamoto, K., (2009). Prediction of a required log reduction with probability for Enterobacter sakazakii during high-pressure processing, using a survival/death interface model. Applied and Environmental Microbiology 75(7), 1885.
Koseki, S., Yamamoto, K., (2007). Modelling the bacterial survival/death interface induced by high pressure processing. International Journal of Food Microbiology 116(1), 136-143.
Lu, S., (2019). Do you know credible interval? Medium.
McElreath, R., (2016). Statistical rethinking. A bayesian course with examples in R and stan. CRC Press, Boca Raton.
O'Brien, S.M., Dunson, D.B., (2004). Bayesian Multivariate Logistic Regression. Biometrics 60(3), 739-746.
Pacifico, A., (2021). Robust open Bayesian analysis: Overfitting, model uncertainty, and endogeneity issues in multiple regression models. Econometric Reviews 40(2), 148-176.
PyMC3, (2018). GLM: Linear regression. The PyMC Development Team.
Rahman, H.A., Wah, Y.B., Huat, O.S., (2020). Predictive performance of logistic regression for imbalanced data with categorical covariate. pertanika journal of science and technology 28.
Ratkowsky, D.A., (2004). Model fitting and uncertainty, in: McKellar, R.C., Lu, X. (Ed.), Modeling microbial responses in food. CRC Press, Boca Raton, FL.
Salvatier, J., Wiecki, T.V., Fonnesbeck, C., (2016). Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2, e55.
Serment-Moreno, V., Barbosa-Cánovas, G., Torres, J.A., Welti-Chanes, J., (2014). High-pressure processing: kinetic models for microbial and enzyme inactivation. Food Engineering Reviews 6(3), 56-88.
Stoica, M., Mihalcea, L., Borda, D., Alexe, P., (2013). Non-thermal novel food processing technologies. An overview. Journal of Agroalimentary Processes and Technologies 19(2), 212-217.
Szumilas, M., (2010). Explaining odds ratios. Journal of the Canadian Academy of Child and Adolescent Psychiatry 19(3), 227-229.
UCLA, (2016). How do i interpret odds ratios in logistic regression? UCLA: Statistical Consulting Group.
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.
Vehtari, A., Gelman, A., Gabry, J., (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27(5), 1413-1432.
Wang, J., Koseki, S., Chung, M.-J., Oh, D.-H., (2017). A novel approach to predict the growth of Staphylococcus aureus on rice cake. Frontiers in Microbiology 8(1140).
Xu, S., (2020a). Frequentist vs. Bayesian approaches in machine learning: A comparison of linear regression and bayesian linear regression. Medium.
Xu, S., (2020b). Generative vs. Discriminative probabilistic graphical models. A comparison of naive bayes and logistic regression. Medium.
Yamamoto, K., (2017). Food processing by high hydrostatic pressure. Bioscience, Biotechnology, and Biochemistry 81(4), 672-679.
Zhang, Z.-H., Wang, L.-H., Zeng, X.-A., Han, Z., Brennan, C.S., (2019). Non-thermal technologies and its current and future application in the food industry: a review. International Journal of Food Science & Technology 54(1), 1-13.