

**THESIS OF DOCTORAL (Ph.D.)
DISSERTATION**

**Predictive Modelling Applied to The
Microbial Safety of Plant-based Milk
Products**

by

Maha Rockaya

Supervisors:

Dr. József Baranyi

and

Dr. Mariem Ellouze



**UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF FOOD SCIENCES
DEBRECEN**

2026.

1. INTRODUCTION

Ensuring food safety remains a central challenge in the modern food system. Global supply chains, diverse raw materials, and growing consumer demand for sustainable products all increase the potential for microbial contamination and spoilage. Traditional approaches relied on empirical testing and end-product verification. While useful, these methods are resource-intensive, reactive, and often insufficient to anticipate risks in increasingly complex food environments. This has led to the rise of predictive, model-based strategies, where risks are assessed proactively rather than after problems occur.

Predictive microbiology (PM) has become one of the most widely used tools in this shift toward quantitative risk management. PM describes and forecasts microbial behavior in foods under different conditions by combining experiments with mathematical models. Its applications include designing safe storage and processing conditions, shelf-life guidance, and supporting quantitative microbial risk assessments (QMRA) all of what can be directly translated into food safety decisions.

The usefulness of PM, however, depends critically on the reliability of its model parameters. Growth rates, lag times, and cardinal temperatures are estimated from data and carry uncertainty. Some of this variability comes from biological sources (*wet* variability), while other parts are statistical (*dry* variability). If such uncertainty is poorly quantified, predictions may become misleading, leading to either over-conservative decisions that burden industry or under-protective policies that jeopardize consumer safety.

To address this, both experimental (*in-vitro*) and simulation-based (*in-silico*) approaches are needed. Laboratory experiments provide insight into how

microorganisms behave in real matrices, while computational analyses test how model formulations and estimation methods affect the variability and reliability of parameter estimates. In this dissertation, both approaches are combined: growth of *B. licheniformis* was studied in three common plant-based milk alternatives after chemical characterization, while *in-silico* analyses examined the most used predictive models and strategies to improve the performance of their regression towards more accurate predictions. Plant-based milks were chosen because they are a rapidly expanding sector with highly variable composition, while *B. licheniformis*, a spore-forming bacterium commonly found in dairy and non-dairy matrices, provides a realistic model organism for studying microbial growth in such environments. The research was therefore structured around complementary objectives: experimental studies that quantified growth behavior and nutrient influences in plant-based milks, and computational studies that evaluated estimates' reliability, reparameterization strategies, and common misconceptions in predictive microbiology. These objectives are presented in detail in the following section.

RESEARCH OBJECTIVES

Wet (in-vitro) objectives:

1. Measure the growth of *B. licheniformis* in plant-based milks across the entire temperature range supporting growth. Compare its kinetics in BHI (culture medium) and three commercial plant-based milks to determine its cardinal temperatures in these matrices and evaluate possible matrix effects on its growth.
2. Quantify macronutrient composition of the targeted media and test correlations between macronutrient concentrations and *B. licheniformis* growth kinetics.

Dry (in-silico) objectives:

1. Suggest a database ontology for the microbiology and composition of plant-based milk alternatives to standardize the recording of experimental conditions, measurement methods, and microbial kinetic observations; to facilitate consistent data collection, sharing, and cross-lab pooling.
2. Assess the reliability of parameter estimates for two widely used primary models, by analyzing the distribution of their standard error and confidence intervals. Evaluate how much their histogram deviates from the one generated by linear regression as the well-studied golden standard.
3. Develop and recommend, justified by statistical evidence, rescaling and reparameterization transformations for two of the most used secondary models to improve (i) the distribution of parameter estimates, (ii) the numerical stability of the estimation procedure, and (iii) the regression performance especially on poor data.
4. Develop a model, describing the growth of *B. licheniformis* as a function of temperature, by determining its cardinal temperature parameters and their uncertainty in the studied matrices.
5. Clarify common misleading concepts in predictive microbiology met during the above analyses (e.g., misuse of the log transformation, misinterpretation of standard error and standard deviation) and produce clear, example-driven explanations and guidance for correct practices.

2. MATERIALS AND METHODS

2.1. Development of a Database Ontology for Microbiology and Composition of Plant-Based Alternatives

Transforming raw data into useful models requires passing through a crucial intermediate stage: data cleaning, sorting, and formatting within a predefined database ontology.

To enable this process and facilitate both the analysis of the dataset of this work and to serve as a foundation for future studies in plant-based alternatives compositional and microbial research, we developed a structured database ontology called (*Pbase*). This framework, inspired by and adapted from *ComBase* (Baranyi and Tamplin, 2004) and *MilkyBase* (Pacza et al. 2022) to integrate microbial growth kinetics with chemical composition data for plant-based alternatives.

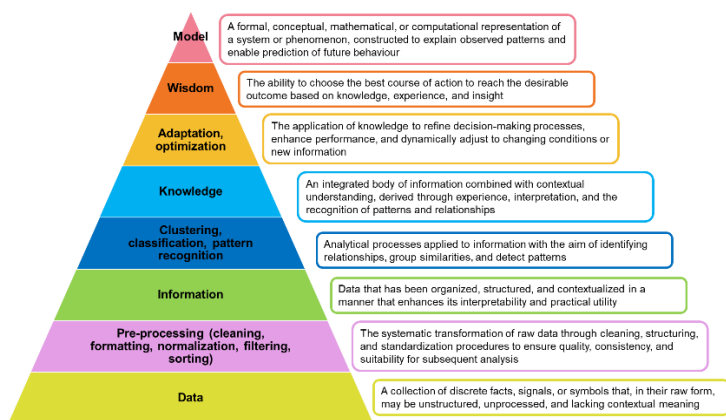


Fig. 2.1 From data to models

Pbase was created in Microsoft Excel, a widely used tool across scientific disciplines, chosen for its familiar interface, ease of use and the future need

of implementing Visual Basic for Applications (VBA), which is supported by Excel, for syntax and semantic check of entries.

Pbase is organized around a *Master* sheet that contains records (Excel rows) each identified by a unique key representing one bacterial growth/death curve with detailed information about the source of the data, the experiment, the log-count and the growth kinetic parameters each in a separate field (Excel columns). The interpretation of each field, along with its allowed values and numerical ranges, is provided in a separate sheet named identically to the corresponding field in the *Master* sheet. These are referred to as *Definition* sheets for obvious reasons. They will serve as key references when applying syntax-checking programs.

2.2. Primary and Secondary Models: Analysis of the Variability of Parameter Estimates and Performance Evaluation After Applying Mathematical Transformation

2.2.1. Analysis of Primary Models

In this section we explain the method used to investigate the regression performance on two widely used sigmoid primary models. Their performance is judged by the reliability of the parameter estimates obtained from fitting simulated log-count data. We compare the regression outputs of the two models with that of the linear regression as a golden standard. As the error-estimation in non-linear regression is carried out by locally-linear approximations, our task is to establish which of the two regression outputs (of the two primary models) generate properties closer to the linear regression.

We focus on *dry* source of error and correlation, by assessing the distributions (histograms) of the parameter estimates and of their relative errors and the

correlation between parameters ($\mu_{\max}$ and h_0). From these, we draw conclusions about the reliability of standard-error estimates and confidence intervals produced by the Least Squares Method (Miller, 2006).

2.2.1.1. Data Simulation

Three primary models have been used: the Linear Model (LM); the Baranyi–Roberts model (BRM); and the reparametrized Gompertz function (GF) to run a *Monte-Carlo* simulation (Mooney, 1997).

Seven sampling times t_i ($i = 1 \dots 7$) were considered: 0, 12, 24, 36, 48, 60 and 72 (h), then 1000 Monte Carlo simulations were performed for each model to generate log-count data.

2.2.1.2. Fitting

The perturbed y_i values were then fitted in two ways: (i) by the same equation that generated them and (ii) by the other sigmoid curve.

Simulation, histograms plotting and linear regression were performed using Microsoft Excel, its statistical functions and Add-Ins. Nonlinear regression was performed with an in-house VBA package implementing the Levenberg–Marquardt algorithm (Press et al. 2007).

2.2.1.3. Histogram Comparison

From each simulated fit, the growth parameter and its standard error were estimated (total = 1000 $\mu_{\max}$ estimates and 1000 $RE_{\mu_{\max}}$). These were used to build histograms to be compared.

Each histogram was compared to the shape of: (i) the same histogram generated by LM, and (ii) the normal distribution curve that has the same mean and standard deviation as the respected histogram does.

2.2.1.4. Correlation Comparison

One of the properties of L-Reg ($y=a \cdot x+b$) is that the intercept b and the slope a are positively correlated (Nau, 2016).

To compare it with G-Reg and BR-Reg, we reparametrized BRM and GF using the intercept parameter (h_0) introduced by Baranyi and Robert (1994). As can be readily seen, h_0 is the equivalent of the intercept b , and $\mu_{\max}$ is analogous to the slope a of linear model.

We then examined the correlation between h_0 and $\mu_{\max}$ for the two tested regression. Comparing this correlation to the strong positive correlation between a and b of LM indicates fitting of which sigmoid model is closer to linear regression.

2.2.2. Analysis of Secondary Models

In publications, parameter estimates are often reported as $\hat{p} \pm 2 \cdot SE(\hat{p})$. This practice relies on two strong assumptions: (i) the original data are normally distributed, and (ii) estimation is performed by linear regression. In predictive microbiology, these assumptions rarely hold. Data are not always normally distributed, and almost all models' regressions are nonlinear. Therefore, it is necessary to evaluate how well this conventional approach applies to secondary models' parameters. Here, we investigate the robustness of two widely used models by examining: (i) the reliability of their standard errors and confidence intervals, (ii) their sensitivity to initial estimates, and (iii) whether reparameterization can bring their regression closer to linear, especially under poor-data scenarios.

2.2.2.1. Data Simulation

Two secondary models: Ratkowsky's Square-root Model (RKM) and the Cardinal Temperature Model (CTM) were used to run a *Monte-Carlo* simulation.

Eight sampling temperatures T_i ($i = 1 \dots 8$) were considered: 6, 12, 18, 24, 30, 36, 39, 41°C and 1000 *Monte Carlo* simulations were performed using each model to generate specific growth rate data μ_{max_i} . The parameter values for RKM: $b = 0.03$, $T_{min} = -3^\circ\text{C}$, $c = 0.5$, $T_{max} = 43^\circ\text{C}$; for CTM: $T_{min} = -3^\circ\text{C}$, $\mu_{opt} = 1.17 \text{ h}^{-1}$, $T_{opt} = 36^\circ\text{C}$, $T_{max} = 43^\circ\text{C}$.

For each T_i the response μ_{max_i} , obtained by each model, was perturbed with normally distributed random errors $\varepsilon(0, \sigma)$ where $\sigma/\mu_{max_i} = 0.1$, mimicking a typical 10% biological variability, caused by (i) unbiased measurement noise; (ii) error in the primary model used to estimate μ_{max_i} .

2.2.2.2. Fitting

The perturbed μ_{max_i} values were fitted in two ways: (i) by the same equation that generated them and (ii) by the other sigmoid curve. Only those estimates were retained for further analysis which met the following criteria : (i) $R^2 > 60\%$; (ii) standard error of c , b , or $\mu_{opt} < 50\%$; (iii) standard error of T_{min} , T_{opt} , or $T_{max} < 10^\circ\text{C}$.

2.2.2.3. Histogram Comparison

From each simulated fit, the parameters of each model and their standard errors were estimated (1000 of each parameter and its corresponding SE). These were used to build the histograms and calculate the 95% confidence intervals using the conventional formula: $95\% \text{ CI} = \hat{p} \pm 2 \cdot SE(\hat{p})$.

Then the number of outliers (i.e. number of estimates that fall outside of the 95% CI) was counted, and the resulting histograms were compared to the theoretical bell-shaped t -distribution.

2.2.2.4. Reparameterization

Comparing the resultant histograms emerged two cases: (i) parameters with histogram shapes sufficiently close to bell-shape and symmetric confidence intervals were considered reliable and no reparameterization was applied; and (ii) parameters with skewed, stretched, or flat histograms were reparametrized to improve symmetry and bring their confidence intervals closer to those expected under linear regression. After applying the suggested reparameterization (see **Table 2-2**), the previous steps were repeated using the reparametrized version of the models (data-simulation → fitting → plotting histograms → calculating 95% CI → counting outliers). This will end up with before/after reparameterization to be compared.

2.2.2.5. Sensitivity to Initial Estimates

To investigate the sensitivity of the model to initial estimates, the previous analysis was repeated but this time with semi-automated (rough) initial values for RKM: $b = 0.01$, $T_{min} = 0^{\circ}\text{C}$, $c = 0.6$, $T_{max} = 51^{\circ}\text{C}$; for CTM: $T_{min} = 0^{\circ}\text{C}$, $\mu_{opt} = 1.2 \text{ h}^{-1}$, $T_{opt} = 30^{\circ}\text{C}$, $T_{max} = 51^{\circ}\text{C}$.

The sensitivity was evaluated by two criteria: (i) the possibility to result in biologically unacceptable shapes, (ii) how many out of 1000 fits met the criteria set before (see **Section 2.2.2.2**).

Strange or nonsense fitting shapes result from discontinuity during the optimization (i.e. an iteration to look for a minimum SS “Sum of Squares”)

this discontinuity will cause the iteration to stop and the program to give an overflow error message.

2.3. Determination of *B. licheniformis* Growth Kinetics and Cardinal Parameters in Plant-based Milk Alternatives

2.3.1. Characterization of Plant-based Milk Alternatives

To capture the variability caused by the food matrix, we used three plant-based milk alternatives available in the Hungarian market: almond milk (A) and a coconut milk (C) from the same company (A) (so the products are named 'AA' and 'CA' respectively), as well as another almond milk from a different company (H) named 'AH'. Brain Heart Infusion 'BHI' (VWR Chemicals) was used for culture medium, as a base for comparison.

For each matrix, macronutrient, pH and water activity measurements were performed using Kjeldahl method for protein, phenol-sulfuric acid titration for carbohydrates, Soxtec extraction for fat, a pH meter with a glass electrode (SevenEasy, Mettler Toledo) for pH and a water-activity meter (rotronic HygroLab C1) for aw.

2.3.2. Inoculum Preparation

B. licheniformis DSM 13 reference strain (Leibniz Institute DSMZ) was used in this study. A working culture of the strain was produced and maintained at -20°C on Microbank™ cryobeads (Pro-Lab Diagnostics) prior to use. The culture of the strain was prepared by inoculating 10 mL of BHI broth with a bead from our working culture and incubating it at 37°C (ESCO) while shaking under aerobic conditions for 8 hours. After the incubation period, 100 µL of the subculture was transferred to a tube containing 9.9 mL of BHI broth and incubated for another 8 hours under the same conditions. For all

experiments the strain was harvested at the end of the exponential phase and at the beginning of its stationary phase.

2.3.3. Growth Experiments

For each experiment, a new inoculum was prepared as described above. It was then decimally diluted four times in Tryptone salt diluent (NutriSelect® Basic, Millipore), and 100 µL of the diluted subculture was then inoculated into a 100 mL Erlenmeyer flask containing 70 mL of sample medium (AH, AA, CA, BHI) to reach a target initial concentration around 102 CFU/mL. For the 'AH' and 'BHI' media, triplicates were prepared at all temperatures. For the 'AA' and 'CA' media, triplicates were prepared at cold temperatures but only duplicates at high temperatures.

For each experiment, the inoculated samples were incubated at the selected temperatures (ranging from 15 to 55°C) while shaking under aerobic conditions.

To enumerate the variation of *B. licheniformis* concentration over time, plate counting method was used by collecting 1 mL of each matrix, adopting the adequate dilutions before pour-plating on BHI Agar (Liofilchem).

Pour-plating was chosen because, under aerobic conditions, *B. licheniformis* rapidly spreads across the surface of the Petri dish, making colony counting difficult. To mitigate this, plates were incubated at 37°C for 20 hours (shorter than the standard 24 hours) for easier counting, and colonies were then counted manually.

For each experiment, we aimed to take regular samples for plate counting, acquiring a minimum of 3 points in the lag phase, 6 points in the exponential phase, and 3 points in the stationary phase.

2.3.4. Choosing the Temperature Levels

Growth experiments were carried out at 13 different temperatures (15, 17, 19, 21, 29, 33, 37, 39, 45, 49, 51, 53, 55°C) in BHI and 12 temperatures (15, 17, 19, 21, 29, 35, 39, 45, 49, 51, 53, 55°C) in almond milk (AH). For white almond milk (AA), and coconut milk (CA) the experiment was carried out in the same temperature range, at 9 temperature points (15, 17, 19, 21, 33, 37, 45, 51, 55°C).

2.3.5. Model Fitting and Statistical Analysis

The measured Ln-count data was fitted with BRM as a function of time (primary modelling), and the estimated μ_{max} for every matrix in every studied temperature were fitted using CTM as a function of temperature (secondary modelling). Both fittings were performed using an in-house Visual Basic for Applications (VBA) Excel Add-In that implemented the Levenberg–Marquardt algorithm for non-linear regression.

To evaluate whether the cardinal parameters (μ_{opt} , T_{min} , T_{opt} , T_{max}) differed significantly across matrices, pairwise Welch’s test **Eq(2-1)** were performed using the parameter estimates and their standard errors (Welch, 1974):

$$t_{welch} = \frac{Estimate_A - Estimate_B}{\sqrt{SE^2_A + SE^2_B}} \quad (2-1)$$

where:

- Estimate_A, Estimate_B: Estimates for the two matrices A and B;
- SE_A, SE_B: respective standard errors of the estimates.

The degree of freedom was calculated using the Welch–Satterthwaite equation **Eq(2-2)** to account for unequal variances.

$$df = \frac{\frac{SE^2_A}{n_A} + \frac{SE^2_B}{n_B}}{\frac{\left(\frac{SE^2_A}{n_A}\right)^2}{n_A-1} + \frac{\left(\frac{SE^2_B}{n_B}\right)^2}{n_B-1}} \quad (2-2)$$

The significance of the difference was assessed via the calculated t_{welch} -values and their corresponding p -values. Differences with $p < 0.05$ were considered statistically significant at the 95% confidence level.

3. RESULTS AND DISCUSSION

3.1. Final Framework of the Phase Ontology

The final framework of *Pbase* (**Fig. 3.1**) is organized around a *Master* sheet that contains records (Excel rows) each identified by a unique key representing one bacterial growth/death curve. The fields (Excel columns) can be categorized into three groups:

- (i) admin fields with details on how and where the data were collected, (ii) explanatory fields related to the experimental details under which the data were produced (e.g., temperature, strain, measurement method)
- (iii) response fields contain the measured/calculated data (e.g., initial counts, maximum counts, specific growth rate).

Admin Fields			Explanatory Fields							Response Fields						
key	org	source	details	mf	measMeth	propOrg	propMF	Comp	temp	pH	Aw	tObs	logSpan	logD	logC	specRate
Bl_AH_19_1	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	19	7.6	0.958	77.00	4.28	2.33	Bl_AH_19_1	0.18
Bl_AH_19_2	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	19	7.6	0.958	77.00	4.25	2.18	Bl_AH_19_2	0.18
Bl_AH_19_3	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	19	7.6	0.958	77.00	4.46	2.15	Bl_AH_19_3	0.19
Bl_AH_21_1	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	21	7.6	0.958	73.17	7.42	1.62	Bl_AH_21_1	0.22
Bl_AH_21_2	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	21	7.6	0.958	96.67	7.24	1.60	Bl_AH_21_2	0.19
Bl_AH_21_3	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	21	7.6	0.958	96.67	7.38	1.60	Bl_AH_21_3	0.21
Bl_AH_29_1	Bl	InstNutr_UniDeb	M01	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	29	7.6	0.958	46.75	7.72	1.86	Bl_AH_29_1	0.57
Bl_AH_29_2	Bl	InstNutr_UniDeb	M01	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	29	7.6	0.958	46.75	7.89	1.84	Bl_AH_29_2	0.52
Bl_AH_29_3	Bl	InstNutr_UniDeb	M01	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	29	7.6	0.958	46.75	7.88	1.83	Bl_AH_29_3	0.59
Bl_AH_33_1	Bl	InstNutr_UniDeb	M01	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	33	7.6	0.958	20.00	6.81	1.99	Bl_AH_33_1	0.88
Bl_AH_33_2	Bl	InstNutr_UniDeb	M01	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	33	7.6	0.958	20.00	6.82	2.03	Bl_AH_33_2	0.87
Bl_AH_33_3	Bl	InstNutr_UniDeb	M02	ABM	pourPlating	provider=DSM2 ; strain=DSM14	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	33	7.6	0.958	20.00	6.88	1.99	Bl_AH_33_3	0.96
Bl_AH_35_1	Bl	InstNutr_UniDeb	M01	ABM	pourPlating	provider=DSM2 ; strain=DSM13	brand=Happy ; atmosphere=aerobic	Comp=IAH_Happy	35	7.6	0.958	16.50	6.42	2.01	Bl_AH_35_1	1.27
Field	Master	logC	org	source	mf	propOrg	propMF	Comp							measMeth	

Fig. 3.1 Screenshot of *Pbase*

The *Master* sheet is linked to auxiliary sheets like “source”, “comp”, “mf”...etc., defining all the fields in the *Master* sheet such as food composition data, organism properties, matrix properties, and temporal

growth measurements (logCount vs time), these sheets are demonstrated in (Fig. 3.1) by unique colors representing each column.

While our current version does not include custom programming, Excel's built-in support Visual Basic for Applications (VBA) offers the potential to enhance functionality in the future by implementing VBA macros to validate the syntax and semantics of entries, making users aware of errors while data is being entered and thereby improving overall data quality.

Although it still requires further data and programming before it can be considered a full database, our ontology hopefully contributes to food sciences, particularly for emerging sectors like plant-based foods.

In conclusion, combining numerical data from publications into a standardized framework facilitates the identification of trends, correlations, and potential causal links that may otherwise go unnoticed.

This work is wished to be a small initiative toward a Big Data and Artificial Intelligence (AI) applications, with ontologies serving as a vital step toward both, ultimately aimed at advancing research on plant-based foods.

3.2. Primary and Secondary Models: Analysis of the Variability of Parameter Estimates and Performance Evaluation After Applying Mathematical Transformation

3.2.1. Analysis of Primary Models

3.2.1.1. Histograms Comparison

The symmetry of the parameter histograms, particularly for μ_{max} and its relative error $RE(\mu_{max})=SE(\mu_{max})/\mu_{max}$, provides a direct indication of how closely non-linear regressions approximate linear regression. As shown in (Fig. 3.2), BR-Reg produced distributions are much closer in shape to those

from the linear regression, whereas G-Reg produced wider, skewed histograms. Importantly, the horizontal axes in (Fig. 3.2) are not on the same scale; the purpose is not to compare absolute errors, but rather the shape of the distributions and their alignment with similar outputs from linear regression. This analysis highlights regression of which model yields more reliable error estimates and confidence intervals, as these are typically derived under the assumption of t-distribution. In all cases, BR-Reg produced error distributions closer to those when of L-Reg.

The robustness of this conclusion were also tested by cross-fitting, which means: BRM was applied to data simulated by GF and vice versa. Regardless of the generating function, fitting BRM consistently provided more reliable error estimation than fitting GF.

Remember, our comparison does not focus on which model better describes the biological process, but on which regression algorithm produces more statistically sound estimates.

Figure (3.2) shows the histograms (blue columns) for the estimates of the maximum specific growth rates (μ_{max}) and their relative errors (green columns) in case of LM, BRM and GF functions. The red continuous lines represent the normal distribution that has the same mean and standard deviation that the respective histogram does. The fitted raw data were 1000 growth curves sets of $N=7$ simulated logCount generated by the same model that was used for fitting, having identical parameters for the three simulation studies including those of the Gaussian noise. The closer the skewness of the histogram to zero, the more symmetric the distribution.

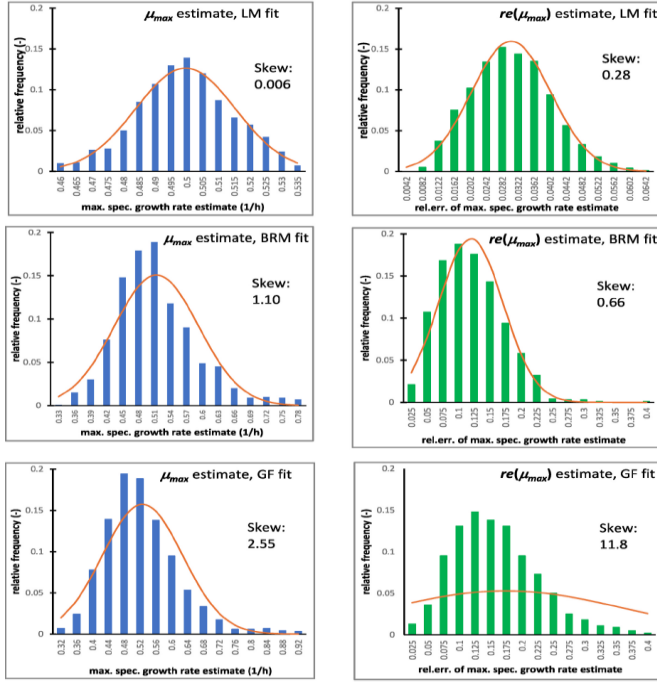


Fig. 3.2 Histograms for the estimates of the maximum specific growth rates and their relative errors in case of LM, BRM and GF functions

3.2.1.2. Correlation Comparison

Linear regression is known to exhibit a strong correlation between slope and intercept estimates as mentioned before. When reparametrized as ($h_0 = \mu_{max} \cdot \lambda$) both BR-Reg and G-Reg displayed a similar positive correlation between μ_{max} and h_0 . The strength of this correlation was higher for BR-Reg ($R^2 > 0.7$) than for G-Reg ($R^2 = 0.61$), again confirming that BR-Reg inherits more of the statistical properties of linear regression (**Fig. 3.3**).

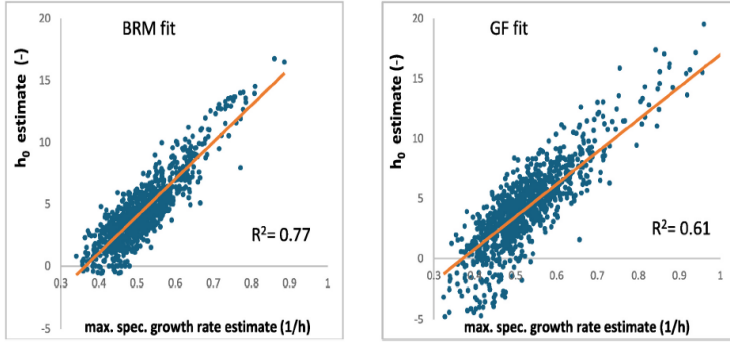


Fig. 3.3 Correlation between the $\widehat{\mu_{\max}}$ and $\widehat{h_0}$ estimates when fitting the simulated data in case of BRM and GF

This is unrelated to their biological (*wet*) relationship where we would expect the opposite trend.

3.2.2. Analysis of Secondary Models

3.2.2.1. Histogram Comparison

The results revealed that the parameters b and μ_{opt} showed distributions that were close to the t -distribution, indicating reliable estimation without the need for reparameterization (**Fig. 3.4**). For T_{min} , the situation needed a compromise: the histogram was not as close to the ideal distribution, and reparametrizing it improved the histogram shape considerably, but this came at the cost of losing biological interpretability thus losing and the practical advantage of easily defining initial estimates for fitting. Therefore, T_{min} was kept in its original form.

In contrast, parameters T_{max} and c in RKM (**Fig. 3.4**) and T_{max} and T_{opt} in CTM (**Fig. 3.5**) displayed histograms far from the t -distribution (skewed, asymmetric, or stretched). Meaning, these parameters required reparameterization to achieve more symmetric histograms and reliable

confidence intervals. Importantly, the choice of the model (RKM or CTM) was not decisive than the choice of reparameterization: both of them give very similar and very satisfactory goodness of fit and once reparametrized, both models provided well-shaped histograms and robust statistical behavior.

Figures (3.4 and 3.5) show the histograms (blue and green columns) for the estimates of RKM and CTM parameters. The red continuous lines represent the normal distribution that has the same mean and standard deviation that the respective histogram does.

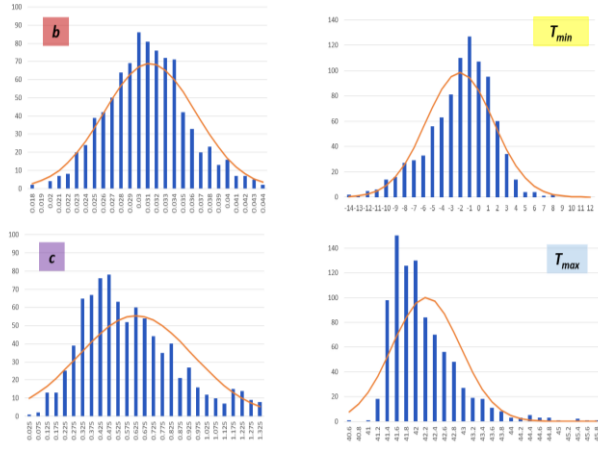


Fig. 3.4 The simulated data fitted with RKM (before reparameterization)

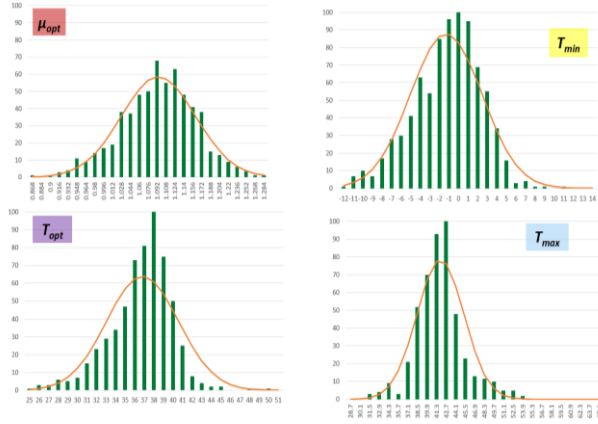


Fig. 3.5 The simulated data fitted with CTM (before reparameterization)

3.2.2.2. Reparameterization

Following the histograms comparison, we suggested a reparameterization method that targets the most problematic parameters. The reparametrized equations were re-fitted to the simulated data, and the results were satisfactory as both models gave symmetric and reliable new histograms (**Fig. 3.6**) and (**Fig. 3.7**). The transformations were defined in (**Table 3-1**)

Table 3-1: The applied transformation on each parameter

	RKM <i>before</i> reparameteriz ation	RKM <i>after</i> reparameteriz ation	CTM <i>before</i> reparameteriz ation	CTM <i>after</i> reparameteriz ation
<i>a</i>₁	T_{min}	T_{min}	T_{min}	T_{min}
<i>a</i>₂	b	b	μ_{opt}	μ_{opt}
<i>a</i>₃	c	$c \cdot T_{span}$	T_{opt}	q_c
<i>a</i>₄	T_{max}	T_{span}	T_{max}	T_{span}

where $T_{span} = T_{max} - T_{min}$ and $q_c = (T_{max} - T_{opt}) / (T_{opt} - T_{min})$

Notice that the new $a_3 = c \cdot T_{span}$ not only gave a smoother histogram but also introduced a dimensionless curvature parameter that is independent of whether the temperature is measured in say Celsius or Kelvin.

The crucial feature of these transformations is that they are reversible, with clear inverse functions that allow the original parameters to be recovered. This ensures that the models remain the same rather than becoming entirely different equations. It is also important to note that the newly estimated 95% confidence intervals (CIs) will not necessarily be symmetric in the reparametrized form. However, they will more accurately reflect the true variability of the estimates, unlike the old symmetric CIs that were roughly and approximately imposed/presumed. This leads to a more reliable quantification of uncertainty and improves the accuracy of decision-making based on these predictions.

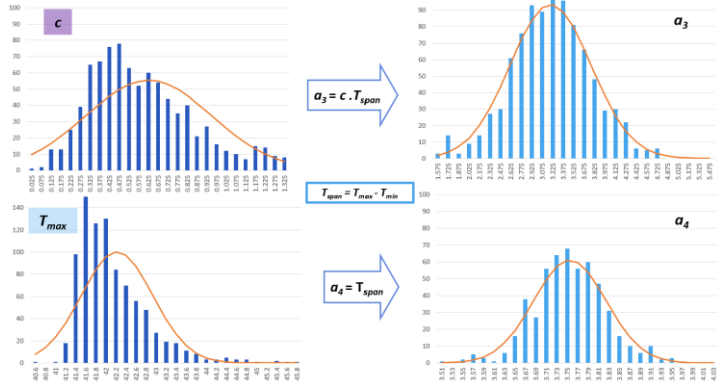


Fig. 3.6 The simulated data fitted with RKM (after reparameterization)

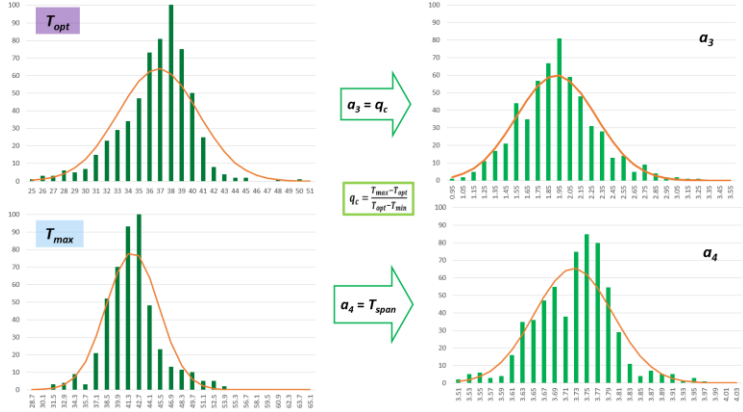


Fig. 3.7 The simulated data fitted with CTM (after reparameterization)

3.2.2.3. Sensitivity to Initial Estimates

RKM consistently provided robust iteration convergence for a wide range of initial estimates. CTM showed convergence for a smaller range of initial estimates which means that occasionally produced discontinuous or nonsensical fits, this caused the program to stop due to an over-flow error, its

polynomial structure is likely the reason. This explains why with CTM we had to exclude ~ 200 out of 1000 simulations, compared to ~ 100 for RKM when assessed the fits against our criteria (see **Section 2.2.2.2**). Finally, because more extreme fits were filtered out, CTM histograms sometimes appeared closer to the t -distribution, but this was a consequence of excluding poor fits rather than inherently better statistical properties.

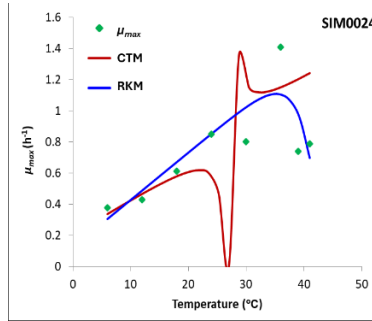


Fig. 3.8 One example of the discontinuous fits of CTM

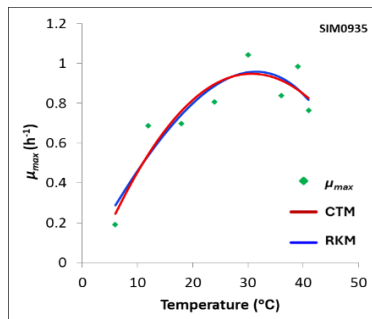


Fig. 3.9 Visual example that there is no significant difference in terms of goodness of fit between the two compared models.

3.3. Determination of *B. licheniformis* Growth Kinetics and Cardinal Parameters in Plant-based Milk Alternatives

3.3.1. Matrix Characterization

Among the tested matrices, BHI had the highest protein content (2.42 g/100 mL) and the lowest fat, while AH was the only matrix with clearly measurable carbohydrates and sugars (2.05 and 1.98 g/100 mL, respectively). AA exhibited the highest total fat, and CA stood out for its high saturated fat content, reflecting the typical fatty acid profile of coconut-based products. pH and water activity values were similar across matrices, with all falling within optimal ranges for microbial growth; therefore, their effect was not modeled.

Nevertheless, additional matrix-specific effects were observed. For example, protein precipitation at elevated temperatures occurred in AH during the stationary phase, which may be related to protein denaturation or the production of the biosurfactant lichenysin. Lichenysin is known to destabilize emulsions and chelate calcium ions (Yeak et al. 2022), both of which could explain the phenomenon. Similarly, fat solidification (e.g. lauric acid) was detected at low temperatures in CA. While these changes were visible, they did not appear to significantly impact microbial growth kinetics. Moreover, pH measurements taken at all temperatures and in all matrices, both before and after incubation, showed no significant changes, confirming that pH was not driving growth differences.

Table 3-2: Energy, macronutrient, pH* and water activity of each matrix

in 100 ml product	Almond Milk (AH)	Almond Milk (AA)	Coconut Milk (CA)	BHI
Energy (kcal)	20.20	16.80	9.98	12.00
Fat (g)	0.569	1.140	0.415	0.020
of which saturated fatty acids (g)	0.069	0.133	0.385	0.012
Carbohydrate (g)	2.05	<0.10	<0.10	0.20
of which sugars (g)	1.98	<0.10	<0.10	0.20
Protein (g)	0.993	0.747	0.496	2.42
Salt (g)	<0.1	<0.1	<0.1	<0.1
pH	7.62 (0.06)	7.38 (0.07)	7.32 (0.05)	7.4 (0.2)
aw	0.958	0.959	0.956	0.992

*The numbers between parentheses are the standard deviations

3.3.2. Primary Fitting and Model Performance

Plate count data (CFU/mL) over time, collected from experiments in BHI and the three plant-based milk alternatives (AH, AA, CA) at all testes temperature (Section 2.3.4) were log-

transformed Ln (CFU/mL) and fitted using the model of Baranyi and Roberts (1994). Two representative one-replicate fittings are shown in (Fig. 3.11 a-b)

Although the primary focus was on the maximum specific growth rate (μ_{max}), it is still worth noting that the maximum cell density (y_{max}) also varied across

both media and temperatures, with white almond milk consistently showing the highest values.

At 45°C growth was rapid and similar among the media, (**Fig. 3.11 b**), whereas at 15°C, growth was markedly slower, with average lag times of 68.8 h in BHI and 69.9 h in AA (**Fig. 3.11 a**).

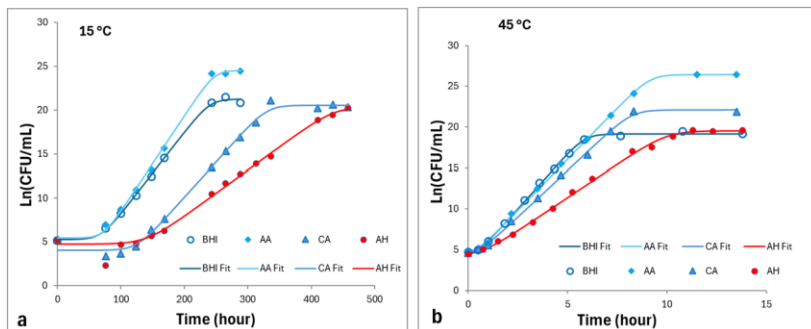


Fig. 3.11 a-b LogCount data as a function of time in various plant-based matrices at 15°C (**a**) and 45°C (**b**), fitted by BRM

Interestingly, both almond and coconut milk, displayed a sudden population decline immediately after inoculation, a phenomenon termed the “*phoenix phenomenon*” (Aspridou et al. 2018). This caused prolonged lag phases of 141.2 h in AH and 129.2 h in CA before exponential growth resumed. A similar decline-recovery behavior was also observed at 55°C but this time in all the tested media, including the BHI broth, suggesting that both 15°C and 55°C represent conditions close to the biological limits of growth. The primary fits (for the 4 matrices and at all temperatures) yielded $SE < 0.2$ log(CFU) standard-error-of-fit and $R^2 > 99\%$ for the goodness of fit, indicating a good accuracy for the original log-count measurements.

More importantly, the data showed that the relative error of the maximum specific growth rate estimates **Eq(3-1)** did not show correlation with the temperature ($p=0.576$; see **Fig. 3.12**):

$$RE(\mu_{max}) = SE(\mu_{max}) / \mu_{max} \quad (3-1)$$

This confirms earlier suggestions by Rockaya and Baranyi (2025) that the logarithmic transformation stabilizes variance when data are obtained via the dilution-based counting method. Indeed, plotting $SE(\mu_{max})$ against μ_{max} reveals a clear linear relationship **Eq(3-2)**, with slope c corresponding to the mid-value of the relative errors (**Fig. 3.12**). This implies that the error of $\text{Ln}(\mu_{max})$ is constant, a desirable statistical property for secondary modelling later on:

$$SE(\mu_{max}) = c \cdot \mu_{max} \quad (3-2)$$

The estimated c slope corresponds to a certain mid-value of the relative errors of the μ_{max} estimates. Remember, the approximation that the errors of the $\text{Ln}(\mu_{max})$ estimates are close to the relative errors of the μ_{max} estimates comes from the **Eq(3-3)** and **Eq(3-4)**:

$$\frac{\mu_{obs} - \mu}{\mu} = \varepsilon \implies \mu_{obs} = \mu (1 + \varepsilon) \quad (3-3)$$

$$\implies \text{Ln}(\mu_{obs}) = \text{Ln}(\mu) + \text{Ln}(1 + \varepsilon) \approx \text{Ln}(\mu) + \varepsilon \quad (3-4)$$

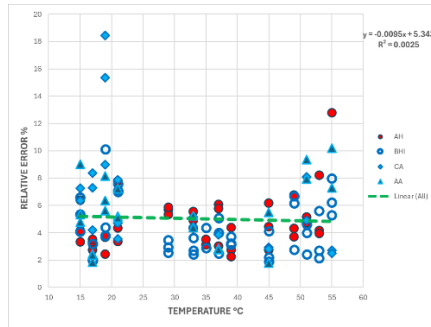


Fig. 3.12 Relative errors of the maximum specific growth rate estimates, generated by the primary model fitting, against temperature, through all the four studied media

Figure (3.12) shows that the relative error (RE) does not show a trend with the temperature ($p \gg 0.05$). Notably, between 25°C and 45 °C, the RE values are rather small (2–6%), corresponding to the “happy-growth region” described by Baranyi et al. (2024). However, RE increases below 20°C and above 50°C as the wet/measurement uncertainty increases close to the growth/no-growth boundaries. Such pattern was previously reported in other food systems, too (Le Marc et al. 2005).

The resultant ca. 5% RE for the specific growth rate estimates in our experiments represents high accuracy and significant improvement compared to its typical values (~10%) generated by other food-based growth experiments (e.g. Pin et al. 2004).

Examples for culture-based growth experiments for *B. licheniformis* can be found in *ComBase* (Baranyi and Tamplin, 2004). There, 48 growth curves (see the records with *ComBase* ID starting with B227, B266, B267) provided the base for its predictive model. These were generated at the same laboratory, under standardized protocol and in culture medium broth; what is more, in a

smaller temperature range than the one we use here. Still the relative error of their maximum specific growth rate estimates is around 20%. The high accuracy of our maximum specific growth rate estimates plays a critical role in the claim that the cardinal temperatures are not the same for the almond-based milk and for the culture medium.

The performance of our model exceeded the ISO guidelines (ISO 23691, 2025), which recommend a Relative Error (RE) below 10% for maximum specific growth rate estimates and a Coefficient of Variation (CV) below 15% when three replicates are used. In our study, the RE was below 15% for all data points, except for two values in the coconut matrix at 19°C, which showed slightly elevated RE-s (15.3% and 18.4%; see **Fig. 3.12**). Across all conditions where replicates were available, the CV of μ_{max} remained below 15%, with an overall average of 4.02%, indicating a high precision of the growth rate estimations.

3.3.3. Secondary Fitting and Model Performance

The CTM model was fitted to our specific growth rate estimates, on the $\text{Ln}(\mu_{max})$ scale, the standard error-of-fit was 0.10. This value shows higher accuracy than those characterizing compatible secondary models on the entire (T_{min}, T_{max}) interval.

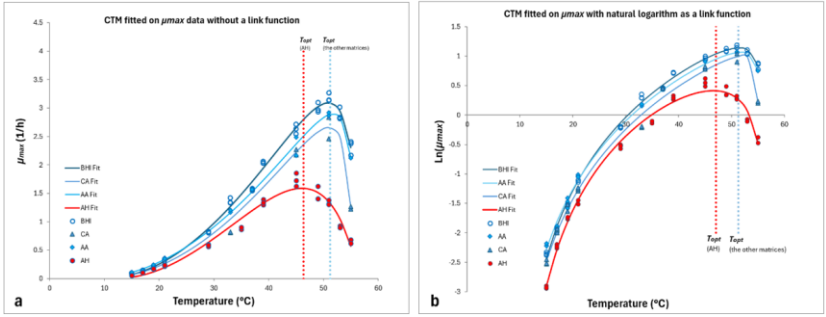


Fig. 3.13 Specific growth rates (μ_{max}) as a function of temperature fitted by CTM.

The reason why the logarithm link function should be used for the μ_{max} response variable was analyzed by Akkermans et al. 2018, also confirmed by Rockaya and Baranyi (2025).

The temperature dependence of $\ln(\mu_{max})$ was modelled separately for each matrix using the Cardinal Temperature model (CTM). Figure (3.13 a-b) demonstrates that the model describes the μ_{max} vs. temperature relationship satisfactorily.

However, beyond the goodness of fit, it is equally important to consider the standard errors of the model parameters, which play a crucial role in comparing the performance of models.

Table (3-3) lists the estimated cardinal parameters, i.e. μ_{opt} , T_{min} , T_{opt} and T_{max} , along with their respective standard errors.

The Cardinal Temperatures of *B. licheniformis* growing in the white almond milk (AA) and the coconut milk (CA) were similar to those in BHI, though the μ_{opt} estimates showed the trend: $\mu_{opt}(BHI) > \mu_{opt}(AA) > \mu_{opt}(CA)$.

An interesting and unexpected difference was observed in the parameters of almond milk (AH) compared to those of the other plant-based matrices. The well-known delta-shaped curve shifted on the x-axis resulting in higher T_{min} and T_{max} and a very different T_{opt} compared to its values in the other media (**Fig. 3.13 a-b**). Besides, almond milk (AH) had the smallest μ_{opt} value.

To support the results statistically, pairwise Welch's test was carried out to compare the cardinal parameters across all the media. It showed that the cardinal parameters for almond milk (AH) differed significantly from the parameters of the other matrices (culture medium 'BHI'; coconut 'CA', and white almond 'AA') in all parameters ($p < 0.05$) except T_{max} . In contrast, no significant differences were observed between the cardinal values in BHI, coconut, and white almond matrices ($p \gg 0.05$). The performance of the model was evaluated according to ISO guidelines (FDIS-ISO 23691, 2025), which recommend a Relative Error (RE) below 10% for (μ_{opt}) and 1°C SE for T_{min} . In our study, the RE for (μ_{opt}) estimate was below 10%, and the SE of T_{min} was even less than 0.5°C for all matrices.

Table 3-3: Estimated parameters and their standard errors for the CTM model in BHI and the three studied Plant-based Milks, resulting from the regression while applying the Ln link function

	T_{min}	SE_{Tmin}	μ_{opt}	$SE_{\mu_{opt}}$	T_{opt}	$SE_{T_{opt}}$	T_{max}	$SE_{T_{max}}$	SE_{fit}
BHI (n=39) (L=13)	8.59	0.16	3.10	0.06	51.36	0.25	56.47	0.33	0.08
Almond Milk (AH) (n=32) (L=12)	9.60	0.27	1.51	0.03	46.55	0.38	56.74	0.39	0.09
White Almond Milk (AA) (n=21) (L=9)	7.47	0.31	2.89	0.15	51.63	0.62	56.20	0.65	0.06
Coconut Milk (CA) (n=22) (L=9)	7.76	0.37	2.71	0.24	52.01	1.22	55.29	0.49	0.11

n = number of fitted μ_{max} ; L = number of temperature levels

The results raise key questions about why AH behaves differently. One reason could be the difference between the beverages in terms of their compositional profile.

AH and AA were derived from the same raw material (almond), but they came from different manufacturers so their formulation and processing were not the same, which might have caused the difference in their microbial kinetics. It is, importantly, noted that unlike bovine milk, which has a relatively standardized composition, plant-based alternatives are human-formulated products with highly variable ingredient profiles and processing methods. Therefore, their source material (e.g., almond, soy, oat) is not

necessarily a significant factor to create groups based on their microbial kinetics. An example for their compositional difference is shown by their total carbohydrate content. There is a significant negative correlation (slope = -1.73 , $p < 0.05$, $R^2 = 0.996$) between the carbohydrate content in AH and T_{opt} , hinting that the elevated sugar concentrations may have been one of the reasons for the found difference. This is consistent with the observations of Santos and Martins (2003) and Yu et al. (2017) that high carbohydrate levels can exert osmotic stress, repress enzyme synthesis, or alter metabolic pathways, thereby might impact temperature-dependent intracellular processes. Previous studies on thermophilic bacteria have reported that their maximum growth temperature (T_{max}) depends on the growth substrate and its energy yield (Heinen, 1970; Merkel and Perry, 1977; Ramaley et al. 1975). Although *B. licheniformis* is typically considered thermotolerant, our observation of a significantly lower T_{opt} in a high-carbohydrate matrix supports the idea that growth temperature parameters can be modulated by matrix composition. Ouhib et al. (2006) highlighted sugar-induced repression or stimulation of toxin and metabolite production for a closely related species, *Bacillus cereus*. Similarly, Balay et al. (2018) demonstrated that carbohydrate type and concentration significantly influence microbial growth and resistance development against antimicrobial agents, highlighting the complex interplay between matrix composition and bacterial adaptation. This further supports the need to consider carbohydrate effects when modeling microbial behavior in diverse food systems. Efforts in this direction have already begun, as shown by Nev et al. (2021), who developed growth models that incorporate variable nutrient conditions and demonstrated that key microbial growth parameters, are influenced by initial nutrient concentrations. Fat and protein content, though less variable among the tested matrices, are also known modulators of microbial thermal kinetics (Finn et

al. 2013; Geeraerd et al. 2000). These point toward the potential value of incorporating continuous variables (e.g., macronutrient composition) instead of food classifications (e.g. almond milk alternative, coconut milk alternatives) into modeling frameworks, encouraging future research beyond the use of categorical food classifications. Incorporating such variables could bridge tertiary and secondary modelling, improving the predictive accuracy and generalizability of microbial growth models across diverse food matrices. As with the successful integration of continuous factors like pH and water activity in secondary models, this will require rigorous experimental designs to isolate the effects of individual nutrients while accounting for their interactions with environmental factors.

4. CONCLUSIONS AND RECOMMENDATIONS

This research confirms that predictive microbiology requires equal attention to both biological variability addressed by rigorous experiments and careful experimental design and statistical variability addressed by using the proper rescaling and reparameterization to reach a reliable estimates confidence intervals. Ontologies like *Pbase* demonstrate the feasibility of structuring experimental data for large-scale analysis, paving the way for AI-driven risk assessment tools. Future work should focus on expanding such databases and linking them with openly available microbial growth observations. The analyses of primary and secondary models showed that rescaling and reparameterization are powerful tools for improving parameter estimation, reducing estimator bias, and producing robust confidence intervals. These results recommend that practitioners adopt reparameterized versions of CTM and RKM in their modeling workflows, especially in cases where data are poor or unevenly distributed. The experimental findings with *B. licheniformis* demonstrated that plant-based milks cannot be treated as homogeneous categories for microbial modeling. Matrix-specific effects on cardinal parameters were observed, underlining the need for food safety assessments that take into consideration the compositional data. These findings are particularly relevant for the rapidly expanding plant-based food sector, where product variability is high and shelf-life decisions carry important safety and economic implications.

In practice, nutrient composition might be an interesting research area where incorporating their effects in secondary modelling similar to the successful one of pH and *a_w* as continuous explanatory factors, could enhance their performance and redefine the link between secondary and tertiary modelling.

Finally, clarified common misconceptions, such as confusion between statistical error quantification, misuse of transformations or parameters interpretation (e.g. lag time), can seriously undermine reliability. Providing clearer explanations, alongside structured data and robust modeling practices, is essential for ensuring that PM continues to serve as a trustworthy foundation for modern, science-based food safety management.

5. NEW SCIENTIFIC RESULTS

1- Development of a novel ontology for plant-based foods.

A new database ontology (*Pbase*) was designed to organize microbial and compositional data of plant-based alternatives. This framework enables pooling, structuring, and analyzing data, and provides a foundation for future data-driven applications in predictive microbiology and risk assessment.

2- Characterization of *B. licheniformis* growth parameters in plant-based milks.

The growth of *B. licheniformis* was described not only in standard culture medium (BHI), as commonly done, but also in three widely consumed plant-based milks available on the Hungarian market.

Cardinal growth parameters (T_{min} , T_{opt} , T_{max} , μ_{opt}) of *B. licheniformis* were experimentally determined in three plant-based milk matrices across a broad temperature range, yielding high-resolution in-vitro data for predictive modeling. The parameter values were as follows: in almond milk (AH) 9.60°C, 46.55°C, 56.74°C, 1.51 h⁻¹; in white almond milk (AA) 7.47°C, 51.63°C, 56.20°C, 2.89 h⁻¹; and in coconut milk (CA) 7.76°C, 52.01°C, 55.29°C, 2.71 h⁻¹, respectively.

3- Matrix-Dependent Shift of Cardinal Growth Temperatures.

This research demonstrates that cardinal growth temperatures of *B. licheniformis* are matrix-dependent and can significantly differ from values determined in standard culture media.

In the almond milk matrix (AH), the optimal temperature ($T_{opt} = 46.55^{\circ}\text{C}$) was significantly lower than the 51.36°C measured in culture medium ($p < 0.05$). These findings challenge the common assumption that cardinal temperatures are intrinsic species characteristics independent of the surrounding matrix. The results show that food composition can measurably alter thermal growth responses. From a food-safety perspective, this implies that predictive models based solely on culture-medium parameters may misestimate microbial behavior in real food systems, thereby influencing risk assessment and temperature-control strategies.

4- Comparative evaluation of primary models.

Two widely used primary models, Baranyi–Roberts (BRM) and Gompertz (GF), were systematically compared in terms of regression properties. BR-Reg (skew = 1.10) was shown to perform more consistently with linear regression (skew = 0.006) compared to the GF-Reg (skew = 2.55), providing clearer guidance for practitioners seeking reliable parameter estimates.

5- Proposal and application of new reparameterizations for RKM and CTM.

New reparameterizations were developed for the Ratkowsky model (RKM) and the Cardinal Temperature Model (CTM), improving parameter estimate distributions, standard errors, and confidence intervals, particularly in poor-data conditions. After reparameterization, both models produced symmetric, reliable histograms and accurate confidence intervals. RKM showed fewer failed fits (100/1000) than CTM (200/1000), although CTM performed similarly when appropriate initial estimates were provided, demonstrating

that reparameterization (not model choice) determines reliable error estimation and practical applicability.

6- Introduction of a new classification of error sources in predictive microbiology.

A conceptual distinction between *wet* (biological/experimental) and *dry* (statistical/model-based) sources of variability was introduced, providing a new way to categorize and interpret uncertainty in predictive microbiology.

6. PRACTICAL UTILIZATION OF THE RESULTS

1- The developed ontology provides a standardized way to store and organize microbial and compositional data of plant-based alternatives. This facilitates cross-study comparability, supports meta-analyses, and prepares the ground for large-scale data pooling. In practice, Pbase can serve as a basis for future AI-driven predictive models and for harmonized risk assessment approaches in the plant-based food sector.

2- The experimental determination of growth parameters in real plant-based milks offers directly usable datasets for predictive microbiology and risk assessment. Instead of relying on broth-based values, food safety models can now incorporate more realistic, matrix-specific estimates to improve shelf-life predictions, product formulation safety testing.

3- By demonstrating that cardinal temperatures may depend on food composition, this work highlights the importance of avoiding blind parameter transfer across different products. Practically, risk assessors and product developers should take into consideration compositional analysis and industry guidelines can be adapted to include matrix-adjusted predictive parameters.

4- Beside its mechanistic foundation, the comparative analysis provides clear evidence that BRM yields more reliable regression outputs than GF. For practical applications, this means food safety practitioners, regulators, and model developers should prioritize BRM when estimating bacterial growth parameters, especially when reliable confidence intervals and biological interpretability are essential.

5- The newly proposed reparameterizations for secondary models provide better numerical stability, improved confidence interval accuracy, and robustness under poor-data scenarios. These modifications can be implemented in modeling software (R, MATLAB, Excel add-ins) and used in standard operating procedures for food safety laboratories, enhancing the reliability of predictive outcomes.

6- Introducing the novel *dry/wet* error classification allows practitioners to better interpret uncertainty. In practice, separating biological variability from statistical estimation error enables more transparent reporting, supports better understanding of safety margins, and strengthens regulatory dossiers by clarifying the origin of uncertainty in model outputs.

7- The thesis provides practical guidance on commonly misused concepts. These clarifications can be converted into training materials, best-practice checklists, and decision-support tools for both industry and regulatory authorities, ultimately improving the quality and trustworthiness of predictive microbiology applications

7. REFERENCES

- ✓ **Aspridou, Z.**, Balomenos, A., Tsakanikas, P., Manolakos, E., & Koutsoumanis, K. P. (2019). Heterogeneity of single cell inactivation: Assessment of the individual cell time to death and implications in population behavior. *Food Microbiology*, 80, 85–92.
- ✓ **Balay, D. R.**, Gänzle, M. G., & McMullen, L. M. (2018). The effect of carbohydrates and bacteriocins on the growth kinetics and resistance of *Listeria monocytogenes*. *Frontiers in Microbiology*, 9, 347.
- ✓ **Baranyi, J.**, Rockaya, M., & Ellouze, M. (2024). From data to models and predictions in food microbiology. *Current Opinion in Food Science*, 57, 101177.
- ✓ **Baranyi, J.**, & Tamplin, M. L. (2004). ComBase: A common database on microbial responses to food environments. *Journal of Food Protection*, 67(9), 1967–1971.
- ✓ **Finn, S.**, Condell, O., McClure, P., Amézquita, A., & Fanning, S. (2013). Mechanisms of survival, responses and sources of *Salmonella* in low-moisture environments. *Frontiers in Microbiology*, 4, 331.
- ✓ **Geeraerd, A. H.**, Herremans, C. H., & Van Impe, J. F. (2000). Structural model requirements to describe microbial inactivation during a mild heat treatment. *International Journal of Food Microbiology*, 59(3), 185–209.
- ✓ **Heinen, W.** (1970). Growth conditions and temperature-dependent substrate specificity of two extremely thermophilic bacteria. *Archiv für Mikrobiologie*, 76(1), 2–17.
- ✓ **ISO International Organization for Standardization.** (2025). Microbiology of the food chain — Determination and use of cardinal values. ISO/DIS 23691.

- ✓ **Le Marc, Y.,** Pin, C., & Baranyi, J. (2005). Methods to determine the growth domain in a multidimensional environmental space. *International Journal of Food Microbiology*, 100(1–3), 3–12.
- ✓ **Merkel, G. J.,** & Perry, J. J. (1977). Effect of growth substrate on thermal death of thermophilic bacteria. *Applied and Environmental Microbiology*, 34(6), 626–629.
- ✓ **Miller, S. J.** (2006). The method of least squares. *Mathematics Department Brown University*, 8(1), 5-11.
- ✓ **Mooney, C. Z.** (1997). *Monte carlo simulation* (No. 116). Sage.
- ✓ **Nau, R.** (2016). Statistical forecasting: notes on regression and time series analysis. Stepwise and All Possible Regressions.
- ✓ **Nev, O. A.,** Lindsay, R. J., Jepson, A., Butt, L., Beardmore, R. E., & Gudelj, I. (2021). Predicting microbial growth dynamics in response to nutrient availability. *PLoS Computational Biology*, 17(3), e1008817.
- ✓ **Ouhib, O.,** Clavel, T., & Schmitt, P. (2006). The production of *Bacillus cereus* enterotoxins is influenced by carbohydrate and growth rate. *Current Microbiology*, 53(3), 222–226.
- ✓ **Pacza, T.,** Martins, M. L., Rockaya, M., Müller, K., Chatterjee, A., Barabási, A. L., & Baranyi, J. (2022). MilkyBase, a database of human milk composition as a function of maternal-, infant-and measurement conditions. *Scientific Data*, 9(1), 557.
- ✓ **Pin, C.,** Velasco de Diego, R., George, S., García de Fernando, G. D., & Baranyi, J. (2004). Analysis and validation of a predictive model for growth and death of *Aeromonas hydrophila* under modified atmospheres at refrigeration temperatures. *Applied and Environmental Microbiology*, 70(7), 3925–3932.
- ✓ **Press, W. H.** (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press.

- ✓ **Ramaley, R. F.**, Bitzinger, K., Carroll, R. M., & Wilson, R. B. (1975). Isolation of a new, pink, obligately thermophilic, gram-negative bacterium (K-2 isolate). *International Journal of Systematic Bacteriology*, 25(4), 357–364.
- ✓ **Rockaya, M.**, & Baranyi, J. (2025). Variability of growth parameter estimates-The role of rescaling and reparametrization. *Food Microbiology*, 128, 104726.
- ✓ **Santos, E. D. O.**, & Martins, M. L. L. (2003). Effect of the medium composition on formation of amylase by *Bacillus* sp. *Brazilian Archives of Biology and Technology*, 46, 129–134.
- ✓ **Welch, B. L.** (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28–35.
- ✓ **Yeak, K. Y. C.**, Perko, M., Staring, G., Fernandez-Ciruelos, B. M., Wells, J. M., Abee, T., & Wells-Bennik, M. H. (2022). Lichenysin production by *Bacillus licheniformis* food isolates and toxicity to human cells. *Frontiers in Microbiology*, 13, 831033.
- ✓ **Yu, W.**, Chen, Z., Ye, H., Liu, P., Li, Z., Wang, Y., ... & He, N. (2017). Effect of glucose on poly-γ-glutamic acid metabolism in *Bacillus licheniformis*. *Microbial Cell Factories*, 16, 1–10.

