UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF FOOD SCIENCES

Head of Doctoral School:
Prof. Dr. Kovács Béla Róbert DSc.

Supervisors:
Dr. József Baranyi, PhD.
and
Dr. Mariem Ellouze, PhD.

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

Prepared by:
Maha Rockaya
PhD. candidate

Debrecen
2026.

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

Dissertation submitted in partial fulfilment of the requirements for the doctoral (Ph.D.) degree in food sciences discipline

Written by: **Maha Rockaya**

MSc. in Food and Analytical Chemistry; BSc. in Pharmacy and Pharmaceutical Chemistry

Prepared in the framework of the Food Sciences Doctoral School of the University of Debrecen

(Food Sciences Programme)

Supervisors:

Dr. József Baranyi, PhD.

and

Dr. Mariem Ellouze, PhD.

The official opponents of the dissertation:

Name	Degree	Signature
…….	…….	…….
…….	…….	…….

The evaluation committee:

	Name	Degree	Signature
Chairperson:	…….	…….	…….
Members:	…….	…….	…….
	…….	…….	…….
Secretary:	…….	…….	…….

The date and venue of the dissertation defense:

Debrecen, …….., ……/…… /2026

TABLE OF CONTENT

LIST OF ABBREVIATIONS 6

LIST OF FIGURES 8

LIST OF TABLES 10

1. INTRODUCTION 1

2. LITERATURE REVIEW 3

2.1. PLANT-BASED DAIRY ALTERNATIVES 3

2.2. *BACILLUS LICHENIFORMIS* 9

2.3. PREDICTIVE MICROBIOLOGY 10

2.4. CLASSIFICATION OF MODELS 12

2.5. RESCALING AND REPARAMETERIZATION 13

2.6. BACTERIAL KINETICS 16

2.7. PRIMARY MODELS 18

2.7.1. Exponential Growth Model and Its Linearization 18

2.7.2. Logistic Model 20

2.7.3. Gompertz Model 21

2.7.4. Baranyi and Roberts Model 22

2.8. SECONDARY MODELS 25

2.8.1. Ratkowsky's Square-root Model 26

2.8.2. Cardinal Temperatures Model 27

2.9. TERTIARY MODELS 27

2.9.1. Correction Factor and Bias Factor 28

2.10. ERROR ANALYSIS 28

2.10.1. Standard Error of Parameter 28

2.10.2. Standard Error of Fit 30

2.10.3. Error Sources 31

2.11. LIMITATIONS OF PREDICTIVE MICROBIOLOGY MODELS 33

3. MATERIALS AND METHODS 35

3.1. DEVELOPMENT OF A DATABASE ONTOLOGY FOR MICROBIOLOGY AND COMPOSITION OF PLANT-BASED DAIRY ALTERNATIVES 35

3.2. PRIMARY AND SECONDARY MODELS: ANALYSIS OF THE VARIABILITY OF PARAMETER ESTIMATES AND PERFORMANCE EVALUATION AFTER APPLYING MATHEMATICAL TRANSFORMATION 36

3.2.1. Analysis of Primary Models 36

3.2.1.1. Data Simulation 36

3.2.1.2. Fitting 37
3.2.1.3. Histogram Comparison 37
3.2.1.4. Correlation Comparison 37
3.2.2. Analysis of Secondary Models 37
3.2.2.1. Data Simulation 38
3.2.2.2. Fitting 38
3.2.2.3. Histogram Comparison 39
3.2.2.4. Reparameterization 39
3.2.2.5. Sensitivity to Initial Estimates 39
3.3. Determination of *B. licheniformis* Growth Kinetics and Cardinal Parameters in Plant-based Dairy Alternatives 40
3.3.1. Characterization of Plant-based Dairy Alternatives 40
3.3.2. Inoculum Preparation 40
3.3.3. Growth Experiments 40
3.3.4. Choosing the Temperature Levels 41
3.3.5. Model Fitting and Statistical Analysis 41
4. RESULTS 42
4.1. Final Framework of the *Pbase* Ontology 42
4.2. Primary and Secondary Models: Analysis of the Variability of Parameter Estimates and Performance Evaluation After Applying Mathematical Transformation 47
4.2.1. Analysis of Primary Models 47
4.2.1.1. Histograms Comparison 47
4.2.1.2. Correlation Comparison 49
4.2.1.3. Summary 49
4.2.2. Analysis of Secondary Models 49
4.2.2.1. Histogram Comparison 50
4.2.2.2. Reparameterization 52
4.2.2.3. Sensitivity to Initial Estimates 54
4.2.2.4. Summary 54
4.3. Determination of *B. licheniformis* Growth Kinetics and Cardinal Parameters in Plant-based Dairy Alternatives 56
4.3.1. Matrix Characterization 56
4.3.2. Primary Fitting and Model Performance 58
4.3.3. Secondary Fitting and Model Performance 61
4.3.4. Summary 66
4.4. Clarification of Common Misleading Concepts in Predictive Microbiology Discovered During Analyses 68

4.4.1. What Is the Difference Between Standard Deviation (STD) and Standard Error (SE)? 68
4.4.2. Does Rescaling Affect SE_{fit} and $SE(p)$? 68
4.4.3. Can the Lag Parameter Have a Negative Value? 69
4.4.4. Does It Matter If We Fit Log10Count or LnCount When Using BRM and GF? 70
4.4.5. How Robust are BRM and GF with Poor Datasets Scenarios? 72
4.4.6. What Does *Dry* and *Wet* Correlation Mean and What Is the Difference Between Them? 73
4.4.7. How to Define the '*Happy-growth Region*'? 75
4.4.8. Can We Use Parallel CI When Fitting the Square-root RKM? 76
4.4.9. What are the Shapes of the Most Used Confidence Intervals? 77
5. CONCLUSIONS AND RECOMMENDATIONS 80
6. NEW SCIENTIFIC RESULTS 82
7. PRACTICAL RESULTS 84
8. SUMMARY IN HUNGARIAN 86
9. SUMMARY IN ENGLISH 88
10. BIBLIOGRAPHY 90
11. LIST OF PUBLICATIONS IN THE FIELD OF RESEARCH 100
STATEMENT ..... ERROR! BOOKMARK NOT DEFINED.
ACKNOWLEDGEMENT ..... ERROR! BOOKMARK NOT DEFINED.

LIST OF ABBREVIATIONS

PM	Predictive Microbiology
PbD	Plant-based Dairy (alternatives)
QMRA	Quantitative Microbial Risk Assessment
BHI	Brain Heart Infusion
UHT	Ultra Heat Treatment
CFU	Colony Forming Unit
PCA	Principal Component Analysis
LM	Linear Model
L-Reg	Linear Regression
BRM	Baranyi and Robert Model
BR-Reg	Baranyi and Robert Regression
GF	Gompertz Function
G-Reg	Gompertz Regression
DT	Doubling Time
ODE	Ordinary Differential Equation
E1	pre-inoculation Environment
E2	post-inoculation Environment
CTM	Cardinal Temperature Model
RKM	Ratkowsky Model
PMP	Pathogen Modelling Program
GP	Growth Predictor
MRV	Microbial Responses Viewer
UGPM	Unified Growth Prediction Model
SSP	Seafood Spoilage Predictor
SE	Standard Error
VBA	Visual Basic Application
CI	Confidence Interval
SS	Sum of Squares
AA	Almond Milk from Company ‘A’
CA	Coconut Milk from Company ‘A’

AH	Almond Milk from Company ‘H’
AI	Artificial Intelligence
CV	Coefficient of Variance
ISO	International Organization for Standardization
STD	Standard Deviation
SEM	Standard Error of the Mean
SE_{fit}	Standard Error of the Fit
$SE(\hat{p})$	Standard Error of Parameter

LIST OF FIGURES

2.1. Europe dairy alternatives market estimates and forecasts, by product, 2014-2025

2.2. Quality criteria of PbD alternatives

2.3. Bacterial prevalence in plant-based beverages raw ingredients; A: total viable count, B: total aerobic spore count

2.4. The chemical structure of lichenysin

2.5. Types of equation rearrangements

2.6. Demonstrating reparameterization on the Ratkowsky model. Canonical form: $\sqrt{\mu_{max}} = a_1 \cdot T + a_0$. Reparametrized form: $\sqrt{\mu_{max}} = b \cdot (T - T_{min})$, where $b = a_1$, $T_{min} = - a_0 / a_1$. It is easier to find biological interpretation for T_{min} than for a_0, while it is easier to analyze the statistical properties of a_0 when estimating its value from measured data

2.7. Graphic representation of typical bacterial growth curve

2.8. Demonstration of the confidence interval of the predicted mean (orange band) and the distribution of the slope parameter (inlet), on data from https://combase.errc.ars.usda.gov/.

2.9. Demonstration of the residuals

3.1. From data to models

4.1. Screenshot of *Pbase*

4.2. Milestones when making sense of data

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

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

4.5. The simulated data fitted with RKM (before reparameterization)

4.6. The simulated data fitted with CTM (before reparameterization)

4.7. The simulated data fitted with RKM (after reparameterization)

4.8. The simulated data fitted with CTM (after reparameterization)

4.9. One example of the discontinuous fits of CTM

4.10. Visual example that there is no differences in terms of goodness of fit between the two compared models

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

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

4.13. Specific growth rates (μ_{max}) as a function of temperature fitted by CTM. (a) μ_{max} estimates from the primary model fitting and their fitted CTM curves, without using a link function and (b) the same data but with the natural logarithm link function

4.14. Observed versus Cf-predicted natural logarithm of the maximum specific growth rates in the four studied matrices, the broken lines show the trend in these matrices in the suboptimal region

4.15. Case of negative (geometrical) lag for GF and BRM

4.16. Fitting BRM to the growth curve with ID =ELC0374 in *ComBase* (Blue stars). First, with yi = Ln(x_i) (Red continuous line) then with the $Log_{10}(x_i)$ (Black continuous line), in the latter case, the results were put back to the natural log scale

4.17. An example for a GF-simulated data set (blue diamonds) fitted by GF (blue thick continuous line) and BRM (red thin continuous line)

4.18. Demonstration of the positive correlation between $\sqrt{\mu_{max}}$ and its standard error

4.19. The difference between confidence band and prediction band

LIST OF TABLES

4-1 Description of *Pbase* ontology

4-2 The applied transformation for each parameter

4-3 Summary of the comparison between the two studied secondary models

4-4 Energy, macronutrient, pH and water activity of each matrix

4-5 Estimated parameters and their standard errors for the CTM model in BHI and the three studied Plant-based beverages, resulting from the regression while applying the Ln link function

4-6 Fitting BRM to the growth curve with ID = ELC0374 in *ComBase*. First, with $y_i = \mathrm{Ln}(x_i)$ then with the $\mathrm{Log}_{10}(x_i)$ values, in the latter case, the results were put back to the natural log-scale

4-7 The difference between confidence band and prediction band

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 assessment. 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 uncertainty 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 *Bacillus licheniformis* was studied in three common plant-based dairy 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 beverages 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 beverages, and computational studies that evaluated estimates' reliability, reparameterization strategies, and common misconceptions in predictive microbiology. These objectives are presented in detail in the following paragraph.

RESEARCH OBJECTIVES

Wet (*in-vitro*) objectives:

1. Measure the growth of *B. licheniformis* in plant-based beverages across the entire temperature range supporting growth. Compare its kinetics in BHI (culture medium) and three commercial plant-based beverages 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 parameters.

Dry (*in-silico*) objectives:

1. Suggest a database ontology for the microbiology and composition of plant-based dairy 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.

Based on the research objectives the dissertation goes into three connected dimensions:

1. Development of a database ontology for microbiology and composition of plant-based alternatives.

2. Primary and Secondary models: analysis of the variability of parameter estimates and performance evaluation after applying mathematical transformations.

3. Determination of *B. licheniformis* growth kinetics and cardinal parameters in plant-based dairy alternatives.

2. LITERATURE REVIEW

2.1. Plant-based Dairy Alternatives

Global demand for plant-based beverages has risen sharply since 2018, with steady growth expected through the end of the decade. Asia-Pacific, led by China, dominates the market, while Europe and North America are also expanding (Vegan Society, 2019). In the U.S., plant-based dairy alternatives make up a significant share of plant-based food sales. Almond

and soy milk remain the leading products, with oat and coconut also forming big part of the core segment. Analysts project the sector will continue its strong growth, reaching $32–40 billion by 2030 (Business Research Company, 2024; Business Research insights, 2025).

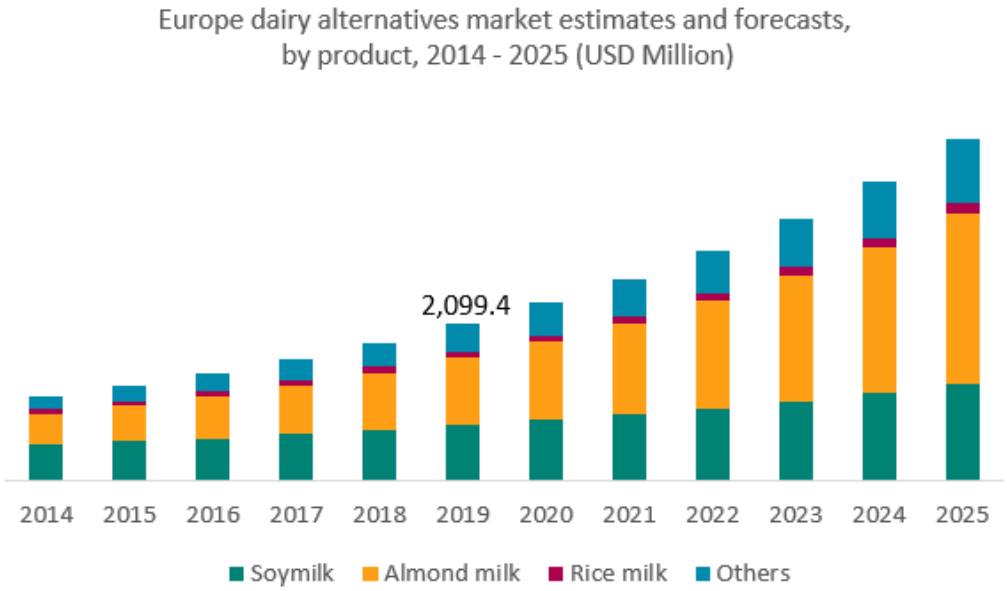

Fig. 2.1 *Europe dairy alternatives market estimates and forecasts, by product, 2014-2025 (USD Million)*

Source: *www.vegansociety.com/news/market-insights/dairyalternative-market/european-plant-milk-market*

The key drivers of using plant-based dairy alternatives can be summarized as: (1) Health consciousness: lactose-free, low-cholesterol, suitable for people allergic to dairy proteins or lactose-intolerant and believed to benefit heart health and weight management; (2) Environmental sustainability: lower carbon, water and land footprints than dairy; (3) Ethical/animal welfare: avoidance of dairy industry practices and animal exploitation (e.g. mass farming of cows); (4) Dietary trends: rise of vegan, vegetarian and flexitarian diets; "*free-from*" and functional food trends (Su et al. 2023; Beacom et al. 2022; McClements and Grossman, 2022).

On the other hand, plant-based beverages differ nutritionally from dairy, often having lower-quality or incomplete proteins, fewer essential amino acids, and anti-nutritional factors like phytates that can reduce mineral absorption; however, these issues can be mitigated by blending sources, fortification, and processing (McClements, 2020). Flavor, texture, and stability challenges, such as off-flavors, separation, or chalkiness, are often addressed with

additives, though these can face consumer criticism as they contrast with the health driver (Lee et al. 2024).

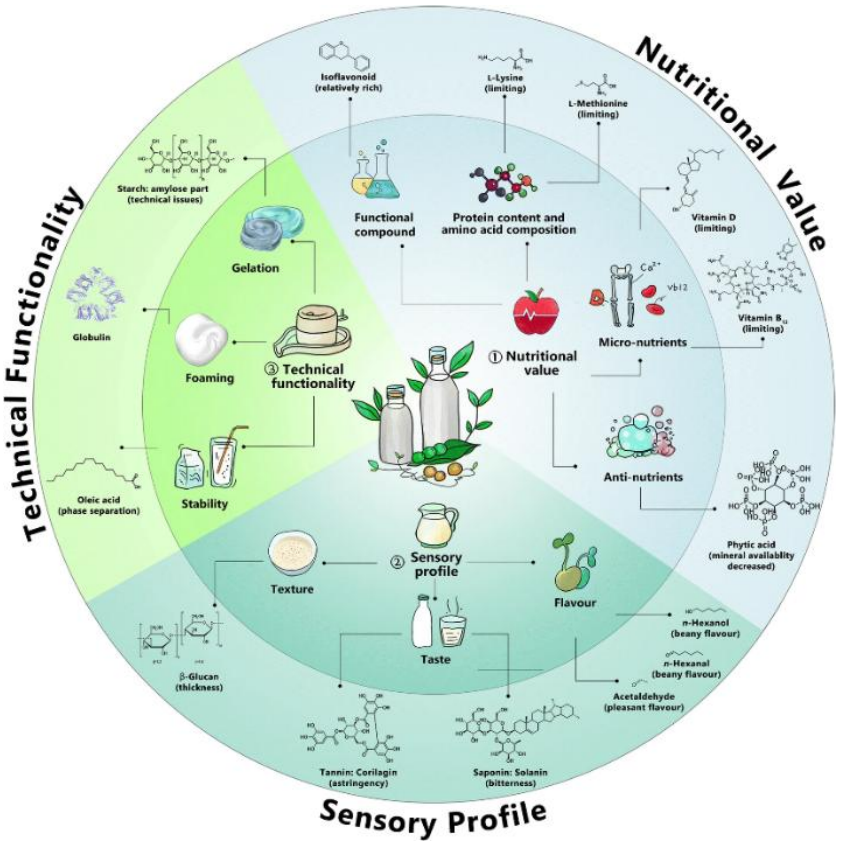

Fig. 2.2 *Quality criteria of PbD alternatives*

Source: *Tangyu et al. 2019*

Plant-based beverages can be broadly categorized by their primary source ingredients, each offering distinct nutritional and environmental profiles. Legume-based varieties, such as soy, pea, fava, and mung bean milk, are valued for their high protein content, though unprocessed forms may contain certain anti-nutritional factors. Nut-based milks, including almond, cashew, macadamia, pistachio, and hazelnut, are often prized for their creamy texture. Grain-based options, such as oat, rice, and quinoa milks, are appreciated for their mild flavor, digestibility, and carbohydrate content, which can provide a quick source of energy. Seed-based alternatives, such as hemp, flax, and sunflower milk, provide functional nutrients, with hemp being rich in omega-3 fatty acids and flaxseed contributing dietary fiber. Emerging and novel products, including pistachio, walnut, buckwheat, watermelon seed, and potato milk, are often positioned in the market for their unique sensory qualities or perceived sustainability advantages (Reyes-Jurado et al. 2023; Karoui and Bouaicha, 2024).

While specific processing and formulation techniques may vary among manufacturers, the production of plant-based beverages typically follows a standardized sequence designed to ensure product quality, safety, and nutritional adequacy. First, *ingredient preparation* involves cleaning and soaking raw plant materials to soften them for further processing. This is followed by *grinding* or *blending*, in which the hydrated material is mixed with water. The suspension then undergoes *filtration* to separate the fibrous pulp (often repurposed for other food applications) from the liquid fraction containing proteins, fats, and soluble nutrients. Next, *heat treatment*, such as pasteurization or ultra-high temperature (UHT) processing, is applied to inactivate microorganisms and extend shelf life. *Homogenization* is subsequently employed to evenly disperse fats and emulsifiers, producing a stable, smooth texture and reducing phase separation. At the *fortification* stage, micronutrients such as vitamins A, D, and B12, along with calcium and other minerals, are incorporated to enhance the nutritional profile and bring it closer to that of regular milk. Finally, *packaging* either as a pasteurized refrigerated product, typically lasting 1–2 weeks, or as a shelf-stable product, in which UHT treatment and aseptic packaging can extend shelf life to 6–12 months (Romulo, 2022; McClements et al. 2019).

Plant-based beverages ingredients frequently harbor high loads of spore-forming bacteria, primarily *Bacillus* spp. (e.g. *B. licheniformis, B. subtilis, B. cereus, B. amyloliquefaciens*) and others (**Fig. 2.3 A-B**), originating from soil and raw plant materials; surveys of oat, almond, pea and rice ingredients report wide variation in viable and spore counts (≈1-8.5 $\log_{10}$ CFU/g) (Gleissle et al. 2025). Although pasteurization processing inactivate yeasts, molds and vegetative cells, bacterial endospores commonly survive this heat treatment and can therefore persist into the final product (Kyrylenko et al. 2023). If they grow, some of these species can produce toxins, of particular concern are thermostable toxins: *B. cereus* can grow in nutrient-rich, refrigerated plant drinks and produce diarrheal enterotoxin as well as the highly heat-stable emetic toxin cereulide, which, once formed in the food cannot be inactivated by further heat treatments (Yang et al. 2023). *B. licheniformis*, often dominant among aerobic spore counts in plant ingredients (**Fig. 2.3 B**) frequently produces lichenysin a surfactant with cytotoxic/emetic potential (Yeak et al. 2022). High initial spore loads therefore enable germination and growth during storage, causing spoilage (acidification, gas formation) and, in some cases, toxin formation.

Given its prevalence in plant-based beverages ingredients, its ability to form heat-resistant spores, and its potential to produce the toxin lichenysin, *B. licheniformis* warrants particular attention. In the next section, we will explore this organism in greater detail, as it is the microbial target of this thesis.

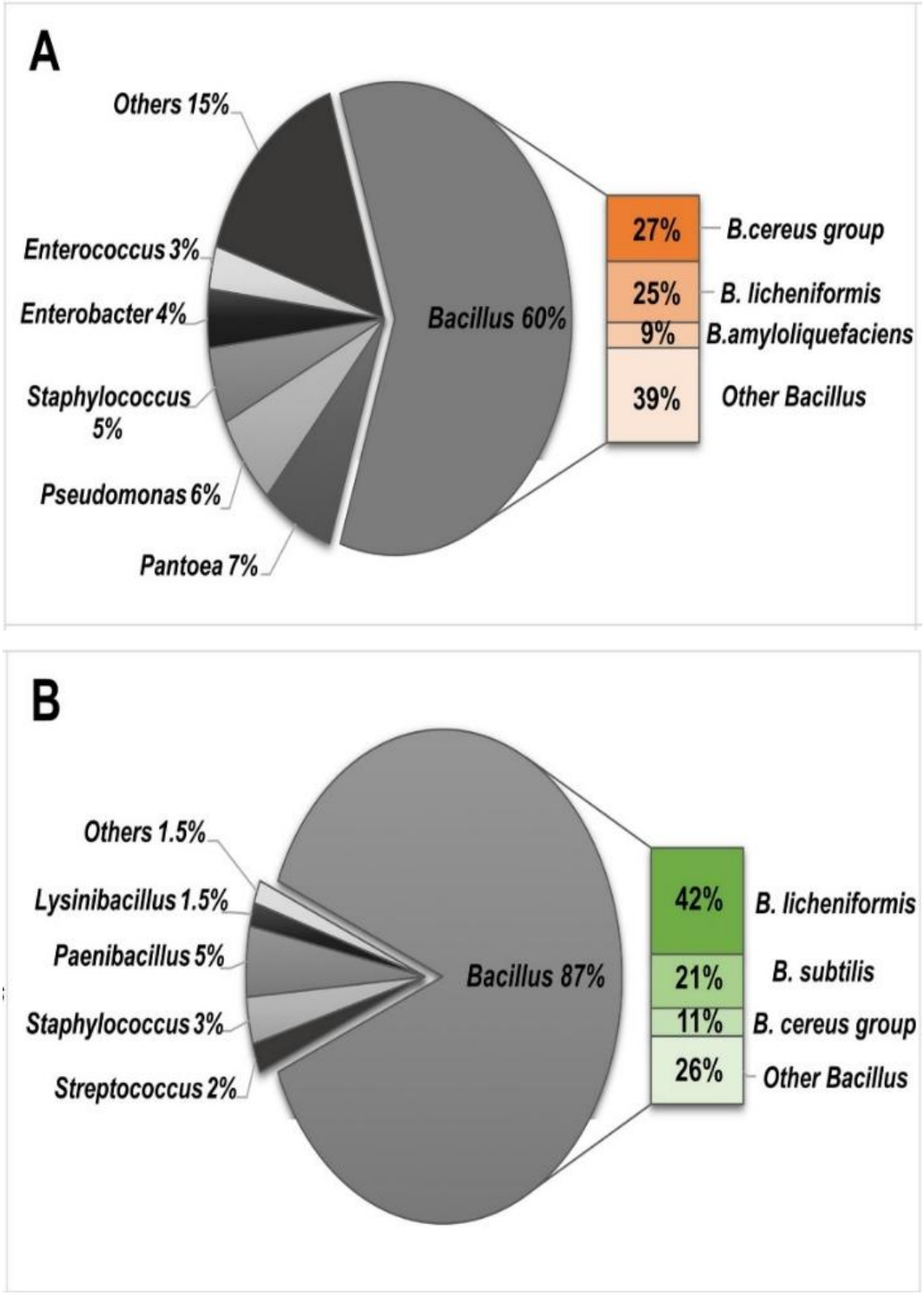

Fig. 2.3 *Bacterial prevalence in plant-based beverages raw ingredients; A: total viable count, B: total aerobic spore count*

Source: *Kyrylenko et al. 2023*

2.2. *Bacillus licheniformis*

Bacillus licheniformis is a Gram-positive, spore-forming, facultatively aerobic bacillus with large genotypic and phenotypic diversity; it is widespread in soil and plant materials and is frequently isolated from plant-based foods, dairy ingredients and processing environments (Dai et al. 2024). *B. licheniformis* is regarded primarily as a spoilage organism in food systems, it produces a range of extracellular enzymes (proteases, lipases) that degrade proteins and lipids, generates off-flavors and sliming, and contributes to the loss of organoleptic quality in dairy and dairy-alternative products (Wang et al. 2021).

B. licheniformis can produce a heat-stable cyclic lipopeptide, lichenysin (**Fig. 2.4**), which is cytotoxic to intestinal cells and has emetic activity, therefore it can act as an intoxication risk under permissive conditions and as a biosurfactant, which is known to destabilize emulsions and to chelate calcium ions (Yeak et al. 2022)

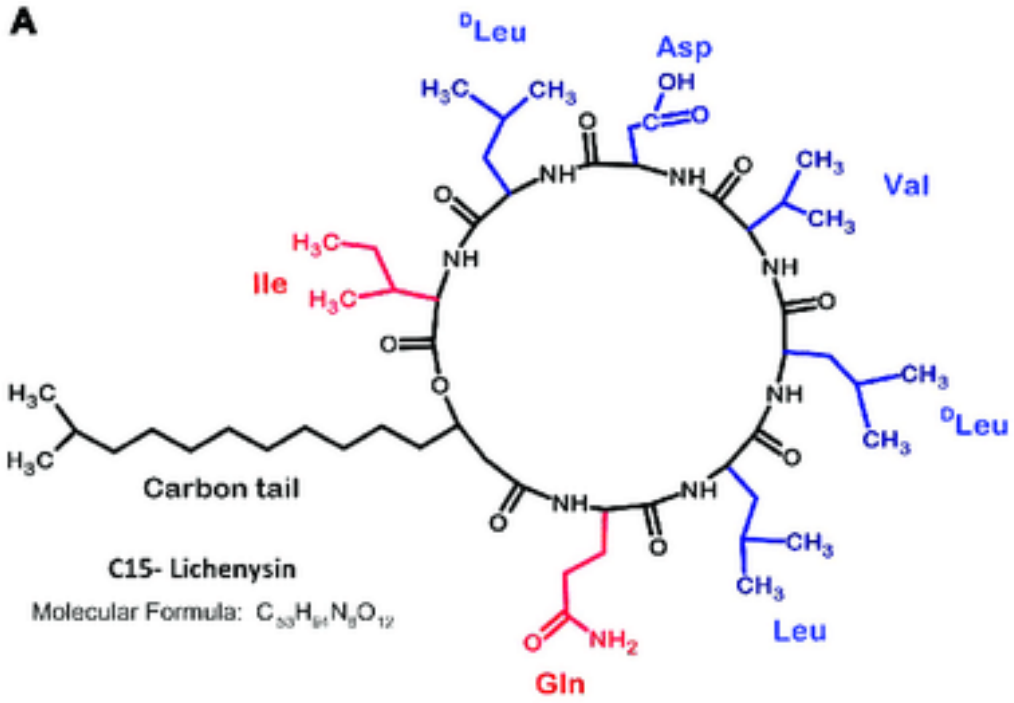

Fig. 2.4 *The chemical structure of lichenysin*

Source: *Yeak et al. 2022*

A critical food-safety concern is the ability of *B. licheniformis* to form heat-resistant spores that survive common thermal treatments (pasteurization, evaporation, spray-drying and in some matrices even UHT), enabling persistence through processing and subsequent

germination after the treatment and finally growth during storage; Its inactivation kinetics were found to be matrix-dependent (Champidou et al. 2024). Recent experimental work demonstrates that spores of *B. licheniformis* exhibit noticeably different thermal inactivation in plant-based matrices (e.g., pea-based drinks) versus bovine milk or laboratory medium, implying that conventional thermal regimes validated on bovine milk may under-perform for plant-based beverages (Scheldeman et al. 2005; Champidou et al. 2024). When raw plant ingredients contain high spore loads, the probability that spores will germinate, multiply during storage, and generate spoilage metabolites or thermostable toxins increases markedly, especially under warm, nutrient-rich, or prolonged-storage conditions (Gleissle et al. 2025).

Persistence in the processing environment is compounded by biofilm formation, in which *B. licheniformis* cells and spores embed in extracellular matrix and resist cleaning and sanitizing procedures; biofilms act as reservoirs of recurrent contamination in production lines; evaporation, spray-drying and filling equipment are critical control points (Diarra et al. 2023; Wang et al. 2021).

Control is therefore multi-pronged: rigorous raw-material screening and supplier controls to limit spore loads, validated sanitation and biofilm mitigation strategies, careful design of thermal processes with attention to matrix effects, strict cold chain and storage-time limits, and consideration of targeted interventions (e.g., germination-followed-by-inactivation strategies or non-thermal technologies such as high-pressure processing) to reduce viable spore and toxin risk in finished products (Maojin et al. 2025).

Accurate risk assessment and effective management of *B. licheniformis* spoilage require detailed characterization of its growth response to critical variables (temperature, pH, water activity and food composition) across relevant food matrices. This is where predictive microbiology, implemented via challenge tests and kinetic modelling, provides the essential tools to quantify growth parameters and define growth boundaries.

2.3. Predictive Microbiology

Predictive microbiology is a scientific discipline that aims to quantify and predict the behavior of microorganisms in food systems using mathematical models (Zwietering et al. 1990). Understanding this behavior helps us address different types of problems such as toxin production, shelf-life/spoilage, development of pathogenic bacteria, microbial inactivation,

etc. This field is sometimes referred to as *quantitative microbial ecology of food* – a term, some researchers argue, may more accurately reflect the field's scope (McMeekin et al. 2002; Baranyi, 2010). It has various applications to ensure food safety and quality such as shelf-life assessment, process parameters settings by modelling and simulating microbial responses to various environmental factors encountered during food processing, transportation, distribution, and storage (Huang, 2014). As a multidisciplinary field, predictive microbiology synthesizes knowledge from microbiology, mathematics, engineering, chemistry, and statistics (Schaffner and Labuza, 1997), allowing researchers and practitioners to move beyond traditional microbial testing toward a proactive, data- and model-driven understanding of microbial dynamics.

The origins of predictive microbiology can be traced back to the 1920s, when mathematical tools were first applied to calculate thermal inactivation times of microorganisms, a breakthrough that revolutionized the canning industry (Esty and Meyer, 1922; Bigelow, 1921). However, widespread academic and industrial interest in microbial modeling did not fully materialize until the 1980s. This delay is often attributed to three converging developments: the sharp rise in foodborne illness outbreaks during the 1980s and the growing public and regulatory demand for safer food products (Ross & McMeekin, 1994), as well as the arrival of personal computers, which enabled more accessible and practical applications of microbial modeling techniques (Tijskens et al. 2001; Whiting, 1995). Since the 1990s it has become a distinct discipline with conferences (e.g. *ICPMF;* https://icpmf.github.io/website/), and databases (e.g. *ComBase;* www.combase.cc).

Mathematical models are commonly employed to predict microbial growth, inactivation, or survival in response to various storage conditions or food processing techniques. These predictions are typically based on specifically designed experiments or derived from existing data, including published studies and observations from real-world situations (González-Tejedor et al. 2017, 2018; van Boekel, 2009). Such models are also used to estimate the shelf-life of food products (García et al. 2015; Possas et al. 2019). One of the key advantages of predictive models is their ability, within certain boundaries, to infer and provide insights for conditions that were not directly assessed during the original experiments (Walls et al. 1996; Walls and Scott, 1997, 1996).

While predictive microbiology is a powerful tool, it is not without limitations. Its reliability is high under well-controlled conditions, and applications to complex food matrices demand rigorous calibration and validation. Even when imperfect, such models remain valuable as decision-support tools, guiding experimental design, informing policy, and providing a transparent rationale for food safety interventions. Rather than substituting empirical evidence, they complement laboratory and field data by generating hypotheses that can be tested and refined (Zwietering et al. 1996; Wijtzes et al. 1998).

2.4. Classification of Models

A common distinction of predictive microbiology models is made between kinetic (dynamic) and probabilistic (event-based) models. *Kinetic* models describe the temporal evolution of microbial populations, typically using differential equations to predict changes in concentration over time. In contrast, *probabilistic* models focus on the likelihood of specific microbial events, such as growth versus no-growth or the exceedance of a critical threshold (e.g., a detection limit or infective dose). However, this distinction is not strict, as probabilistic elements can be incorporated into kinetic models through stochastic formulations (McMeekin et al. 2002).

From another perspective, models can be categorized as deterministic or stochastic, depending on how variability and uncertainty are handled. *Deterministic* models assume fixed input parameters and therefore produce a single outcome, whereas *stochastic* models account for variability and/or uncertainty by representing inputs and/or parameters as probability distributions. In such cases, techniques such as Monte Carlo simulations are used to propagate variability and/or uncertainty through the model and generate a distribution of possible outcomes (Poschet et al. 2003).

In the context of risk assessment, it is important to distinguish between variability and uncertainty (Nauta, 2000). *Variability* reflects the natural heterogeneity of biological systems and environmental conditions and can be described using probability distributions. *Uncertainty*, on the other hand, arises from limited knowledge about model structure, parameters, or assumptions, and is more difficult to quantify (WHO, 2008). While both aspects can be addressed within stochastic modeling frameworks, this study focuses exclusively on variability. Accordingly, one-dimensional Monte Carlo (1D MC) simulations

were applied to propagate variability in model inputs and parameters, without explicitly separating uncertainty.

Models can also be described as *empirical* or *mechanistic*. Empirical models are often built by regressing data with mathematical functions, without any links to the mechanism. These models (e.g polynomial models) may offer high accuracy in specific settings while not suitable for extrapolation, i.e. for predictions outside the range of observations. They also often lack generalizability and interpretability. In contrast, mechanistic models are built on idealistic, simplifying assumptions preferably stemming from fundamental principles of science, such as mass balance, enzyme kinetics, or thermodynamics.

As Baranyi (2010) emphasized, a mathematical model is more than just a set of equations; it includes the definitions of its variables and parameters, possibly constraints on their values, and domain-specific interpretations. Thus, calling an equation a "*model*" assumes that its interpretation and conditions are already known and understood. Referring to a mathematical model as an "*equation*" without its contextual grounding reduces it to a purely mathematical artifact, stripping away its explanatory power.

In this thesis, the focus will be exclusively on kinetic-type models, whether empirical or mechanistic, following the hierarchical classification proposed by Whiting and Buchanan (1993).

According to this classification, microbiological models are categorized into three levels: *primary*, *secondary*, and *tertiary*.

To help develop and implement kinetic models, two key concepts will be defined, as they will be used frequently in what follows.

2.5. Rescaling and Reparameterization

Rescaling and reparameterization are mathematical rearrangements which are classified under a subgroup called mathematical transformations (**Fig. 2.5**). They are frequently used to enhance the implementation, numerical stability, and interpretability of predictive models. When fitting models, either in simple environments like that of Microsoft Excel spreadsheets, or using more complex tools, one can improve the computational performance of the regression by transforming the variables and/or the parameters.

Rescaling refers to monotonic, such as logarithmic or square root, transformations of a variable, either to stabilize the variance of its measured values, or to linearize relationships, or prepare the data for regression (e.g. using the logarithm of the bacterial count before fitting) (Baranyi, 1992).

Rescaling is widely used in microbiology and biochemistry; A classic example is the logarithmic definition of pH, where hydrogen ion concentration is expressed on a log scale to compress a wide numerical range into a manageable one. In bacterial growth modelling, the cell-concentration is often log-transformed for two reasons: (i) its exponential proliferation becomes linear on the logarithmic scale that simplifies the numerical analyses; (ii) since cell concentrations are estimated via dilution and colony counting, the log transformation helps maintain a constant relative error regardless of the cell concentration level.

*Reparameterization*s are carried out to make parameters estimates uncorrelated or more biologically interpretable, so it is easier to provide initial estimates for them (which is a must for non-linear regression), or to verify if the results are reasonable. Even more importantly, fitting the new parameters may result in a procedure closer to linear regression, with all its advantages. Reparameterizations do not alter the core assumptions. Rather, they serve to optimize the model's usability, facilitating more robust and accessible applications.

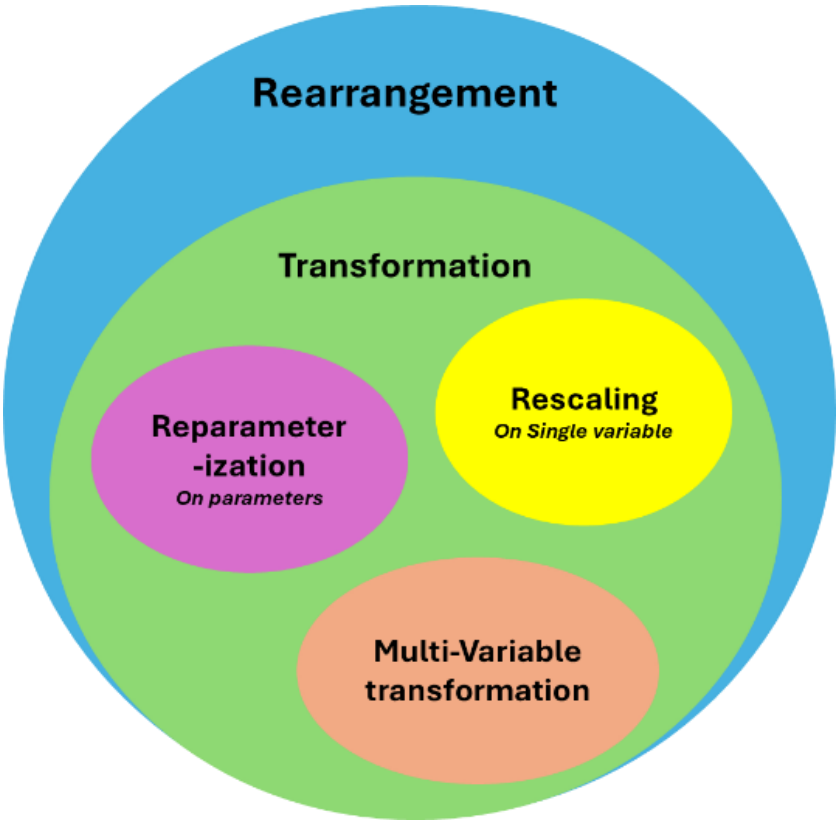

Fig. 2.5 *Types of equation rearrangements*

Consider the linear model $y = a_1 \cdot x + a_0$, which is the starting point of any regression analysis. Not only that y is a linear function of x, but the model is also linear in terms of its parameters (a_1 and a_0), which means they can be estimated by linear regression resulting in explicit formulae for the parameter estimates, depending only on the x_i, y_i data points to be fitted. Apply a reparameterization to this model to be $y = a_1 \cdot (x - x_0)$, where the new parameter $x_0 = a_0/a_1$ is the intercept with the horizontal x axis. The model itself does not change, only its parameters are recombined. Fitting this reparametrized model (which is still linear in terms of the relationship between x and y) results in a non-linear regression, which is more complicated than the linear regression. However, the transformation can still be useful if the new parameter x_0 has easier biological interpretation than a_0 had. Such an interpretability makes it also easier to provide initial estimates needed for non-linear fitting.

An example for this is the model of Ratkowsky (1982) where the square root of the maximum specific growth rate is a linear function of temperature in the sub-optimal region: $\sqrt{\mu_{max}} = b \cdot (T - T_{min}) = (b \cdot T) - (b \cdot T_{min}) = b \cdot T + a_0$ where T_{min} is a parameter close to the minimum temperature of growth (**Fig. 2.6**). The $a_0 = -b \cdot T_{min}$ reparameterization brings the

model to the canonical form $a_1 \cdot T + a_0$, which is linear in the new parameters, therefore their estimates can be obtained by direct formulae and the distribution of the estimates of a_1 and a_0 is well-known and analyzed. The original parameterization lends itself to nonlinear regression, for which the distribution of the parameter estimates can be obtained by simulation (so-called boot-strap method) only.

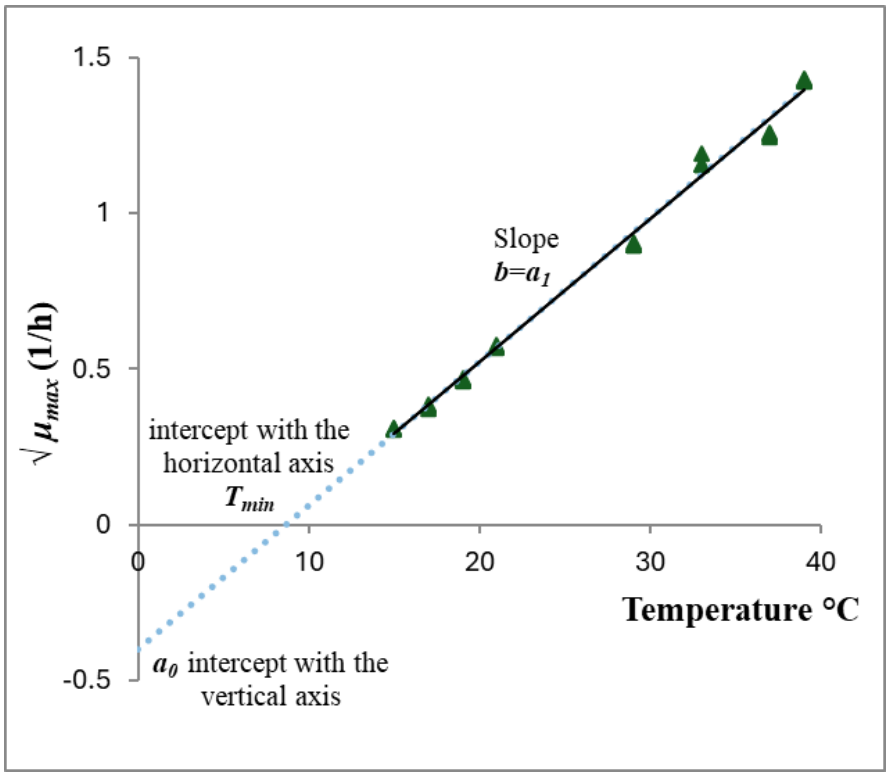

Fig. 2.6 *Demonstrating reparameterization on the Ratkowsky model. Canonical form:* $\sqrt{\mu_{max}} = a_1 \cdot T + a_0$ *. Reparametrized form:* $\sqrt{\mu_{max}} = b \cdot (T - T_{min})$*, where* $b = a_1$*,* $T_{min} = -a_0 / a_1$*. It is easier to find biological interpretation for* T_{min} *than for* a_0*, while it is easier to analyze the statistical properties of* a_0 *when estimating its value from measured data*

2.6. Bacterial Kinetics

Bacterial kinetics is a term used to describe the rate of change of microbial populations with time, including growth and death (**Fig. 2.7**). Kinetic curves usually represent the $\log_{10}$ of colony-forming units *per* milliliter or gram of the bacterial concentration in a sample (CFU/ml or CFU/g), as a function of time. In this work, we will focus on growth kinetics.

The four phases of the normal bacterial growth curve are lag, exponential (or log), stationary, and death phase (Monod, 1949).

During the lag phase, cell division is delayed for some time while they adjust to the new environment, resulting in non-quantifiable change in population size (Rolfe et al. 2012). During this phase cells synthesize the proteins required to metabolize the available substrates and adjust to the conditions surrounding them (Madigan et al. 2000). The lag phase is influenced by several factors, such as the pre-inoculation history of the cells, temperature variation (static / dynamic), differences in nutrient concentration between the original and new media or the inoculum size (Augustin et al. 2000). Thus, unsurprisingly, researchers have shown high variation in the length of the lag phase across different strains of bacteria, even for the same strain under different pre-inoculation stress conditions, and even between different cells within a homogeneous inoculum population (Hamill et al. 2020).

Once adaptation is complete, the bacteria reach the exponential (log) phase, in which cells reproduce rapidly using the available nutrients (Monod, 1949). With increasing bacterial numbers, depletion of nutrients and potential accumulation of toxins, growth decelerates. Subsequently, the rate of cell death becomes about the same as the rate of cell division, and the population size remains constant. This phase is referred to as the stationary phase (Maier and Pepper, 2015). The subsequent death phase is typically triggered by changes in pH due to metabolic end products and the accumulation of toxic wastes (Monod, 1949).

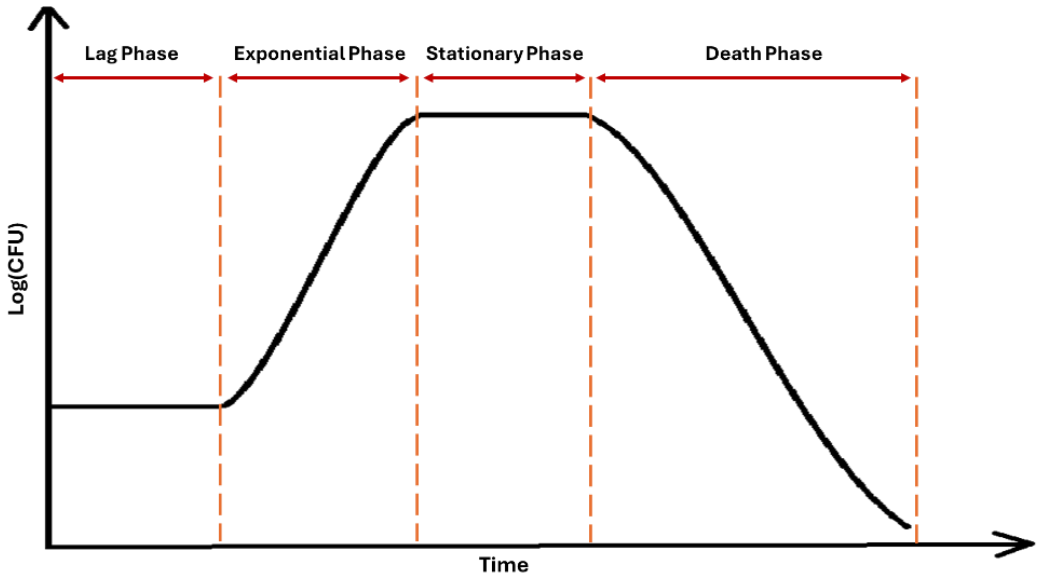

Fig. 2.7 *Graphic representation of typical bacterial growth curve*

2.7. Primary Models

Primary models in predictive microbiology describe the temporal response of a microbial population under constant environmental conditions. They are used to quantify key growth parameters such as lag phase duration, maximum specific growth rate, and final population density. These models are based on experimental data that track changes in microbial population over time, typically represented as a microbial growth or death curve.

Microbial responses used to build primary models can be measured directly, via colony counts or microscopy, or indirectly through indicators such as optical density, electrical conductivity, or the detection of metabolic byproducts like organic acids and toxins (Gracias and McKillip, 2004).

2.7.1. Exponential Growth Model and Its Linearization

A standard introduction to primary modeling begins with the simplest case, where a healthy bacterial population is transferred to a nutritive environment and is assumed to grow immediately without restrictions. Under constant environmental conditions, this process can be described using a first-order growth kinetics model with a constant specific growth rate, as shown in **Eq(2-1)**. This is known as the pure exponential growth model or Malthus Model (Malthus, 1798):

$$\frac{dx(t)}{dt} = \mu_{max} \cdot x(t) \qquad \textbf{(2-1)}$$

where $x(t)$ is the population size at the time t, and μ_{max} is the maximum specific growth rate (hour^{-1}).

The solution to this differential equation, assuming an initial population size x_0, is:

$$x(t) = x_0 \cdot e^{\mu_{max} \cdot t} \qquad \textbf{(2-2)}$$

Linearization of this model is done by rescaling the response variable taking the natural logarithm of both sides to give this linear equation which will be referred to in this thesis as LM and L-Reg for its regression:

$$y(t) = Ln\big(x(t)\big) = y_0 + \mu_{max} \cdot t = Ln(x_0) + \mu_{max} \cdot t \qquad \textbf{(2-3)}$$

Where μ_{max} is the slope and $\mathrm{Ln}(x_0) = y_0$ is the intercept. $x(0) = x_0$ represents an initial value for the differential equation.

Not only that the natural logarithm of the cell concentrations is a linear function of time, but it is linear in terms of the μ_{max} and y_0 parameters, too, so they can be estimated by linear regression.

The main properties of linear regression (Nau, 2016) are as follows:

(1) the parameter estimators $\widehat{y_0}$ and $\widehat{\mu_{max}}$ are explicit expressions of the t_i, y_i values (unlike the implicit way in non-linear regression which is carried out by locally linear approximations)

(2) if RE($\widehat{\mu_{max}}$)=SE($\widehat{\mu_{max}}$)/$\widehat{\mu_{max}}$ denotes the relative error of the $\widehat{\mu_{max}}$ estimate, $SE(\widehat{\mu_{max}})$ being its standard error, then the reciprocal of the relative error follows Student's t-distribution which allows the accurate calculation of confidence intervals for the estimates

(3) y_0 estimate positively correlates with μ_{max} estimate.

These properties of the linear model are known and well analyzed, making it a reference point for any curve fitting procedure.

Since food microbiologists often work with $\log_{10}$ of cell concentrations, the conversion from natural logarithm to common logarithm is applied using the factor Ln(10) ≈ 2.3 :

$$Log(x(t)) = Log(x_0) + \left(\frac{\mu_{max}}{Ln(10)}\right) \cdot t \qquad \textbf{(2-4)}$$

Here, μ_{max} is the maximum specific growth rate expressed in natural logarithm units Ln(CFU)/hour. The term [μ_{max} / Ln(10)] is the slope of the growth curve of $\log_{10}$(CFU)/hour which is "*the maximum growth rate*" and usually referred to as v_{max} so the equation can be written:

$$Log(x(t)) = Log(x_0) + v_{max} \cdot t \qquad \textbf{(2-5)}$$

Since μ_{max} and $v_{max} \approx \mu_{max}/2.3$ can both be used via this transformation, care must be taken not to confuse them.

On the other hand, the Doubling time Dt (i.e., the time required for the population to double) can be derived as follows:

$x(t+Dt) = 2 \cdot x(t)$

$Ln(x(t+Dt)) = Ln(2) + Ln(x(t))$

$Ln(x_0) + \mu_{max} \cdot (t + Dt) = Ln(2) + Ln(x_0) + \mu_{max} \cdot t$

$\mu_{max} \cdot Dt = Ln(2)$

$$Dt = \frac{Ln(2)}{\mu_{max}} \quad \textbf{(2-6)}$$

The limitation of this model is that it provides a good description of bacterial growth only under optimal, non-limiting conditions and assumes that no lag phase occurs. When a lag phase is present in the observed data, this simple exponential model becomes less accurate. When the environment supports growth but under limiting conditions the resulting growth curve of plotting the logarithm of cell concentration against time, is of sigmoid shape. (Baranyi et al. 1993).

Several widely used primary models capture the sigmoid pattern of growth curves, including that of Gompertz (1825), Verhulst, (1845), Richards (1959). The most popular primary models, the reparametrized Gompertz function (Zwietering et al. 1991) and the model of Baranyi and Roberts (1994), are built on those classical growth models.

It is important to emphasize that models such as that of Gompertz and Verhulst (the latter one commonly called logistic model) were developed for the population size and not for its logarithm. They should not be called Gompertz or logistic model once they are applied for log-counts data. Their true sigmoid shape only holds on the arithmetic scale; on the log scale, both the Gompertz and logistic model would have a monotonically decreasing slope. The observed sigmoid behavior in microbial growth results from biological adaptation processes, particularly during the lag phase, necessitating the use of models that are appropriate both for the biological context and the scale of analysis (Baranyi, 2010).

The logistic model (Verhulst, 1845), the reparametrized Gompertz function (Zwietering et al. 1991) and the model of Baranyi and Roberts (1994) will be summarized in the next section.

2.7.2. Logistic Model

In a finite space, the closer a growing population to the (x_{max}) carrying capacity of the environment, the more its μ_{max} maximum specific growth rate is reduced. The logistic model (Verhulst, 1845) was the first to use the factor $u(x) = (1 - x/x_{max})$ to attenuate the maximum specific growth rate.

$$\frac{dx(t)}{dt} = \mu_{max} \cdot x(t) \cdot (1 - \frac{x}{x_{max}}) \quad \textbf{(2-7)}$$

This $u(x)$ attenuator function is between 0 and 1, and monotone converges to 1 as $x(t)$ increases.

As mentioned, the logistic growth model was originally developed to model how a population grows in an environment with limited resources. Using it for the logarithm of a population would lose its background, and the correct term in predictive microbiology should be "*logistic function (not model) used for the log-counts data*".

2.7.3. Gompertz Model

The original Gompertz model (Gompertz, 1825) was initially developed to model the time-variation of a population size. Later it was useful to describe the growth of a tumor.

Its ODE-form is:

$$\frac{dx(t)}{dt} = c \cdot x(t) \cdot Ln\left(\frac{x_{max}}{x}\right) \quad \textbf{(2-8)}$$

where $x(t)$: population size (or other quantity), x_{max}: asymptotic maximum and c: rate constant.

Zwietering et al. (1991) reparametrized this function to obtain parameters with the biological meaning customary in predictive microbiology:

$$Log(x(t)) = Log(x_0) + \left[A \cdot e^{-e^{\left(\frac{\mu max}{A} \cdot e \cdot (\lambda - t) + 1\right)}}\right] \quad \textbf{(2-9)}$$

Where x(t): population size at time t;

x_0: initial population size;

x_{max}: asymptotic maximum population size;

μ_{max}: maximum specific growth rate;

λ: lag time;

t: time;

A: difference between the logarithm of the maximum and initial population size (growth amplitude), defined as:

$A = Log(x_{max}) - Log(x_0)$

$\mu_{max} = \frac{c \cdot Log(x_{max})}{e}$ and $\lambda = t_r - \frac{1}{r}$

where

t_r: time at which the absolute growth rate is maximal (inflection point of the curve)

r: relative growth rate at the inflection point

This equation will be referred to as Gompertz function, GF and G-Reg for its regression.

Note that reparameterization does not change the model, only the parameters' meaning.

2.7.4. Baranyi and Roberts Model

As mentioned, predictive microbiology focuses on the *logarithm* of the bacterial population that, in a constant and finite environment typically exhibits a sigmoidal shape over time. However, the first (lag) phase is "*optional*" as it is due to the change in the environment at the inoculation; the smaller this change, the shorter the length of the lag period, that therefore can be zero, too. Baranyi and Roberts (1994) proposed a model that expresses this (so-called non-autonomous) property. In what follows, the model will be referred to as BRM and BR-Reg for its regression.

BRM distinguishes the pre-inoculation environment (E1) from the post-inoculation environment (E2). During the ideal exponential phase in E2, the logarithm of the cell concentration increases linearly, along a slope μ_{max} which is the maximum specific growth rate the cells can achieve in E2. This rate is attenuated preceding and following the exponential phase based on established models of population-dynamics and biochemistry.

Using the logarithm link function for the cell concentration is particularly relevant as bacterial counts are traditionally measured by the plate-count method resulting in the *lognormal* distribution of the measurement error, with a constant relative error. This is independent of time and bacterial concentration since the latter one is obtained after serial dilution so that colony numbers fall in a practical range (≈150–300; see Corry et al. 2007).

Consequently, when BRM is fitted to y_i = Ln(cell concentration at t_i), the residuals will be independent, identically and normally distributed, just as needed for data fitting.

During the transition phase preceding the exponential phase, an attenuation (or adjustment) function

$$\alpha(t) = \frac{q(t)}{1+q(t)} \quad \textbf{(2-10)}$$

is applied to the maximum specific growth rate (i.e. the slope of the linear function in the exponential phase). Its $\alpha(0)$ initial value depends on E1 (history-effect, just as the inoculum level). The $\alpha(t)$ function monotone converges to 1 as *t* increases.

The model is completed by an attenuation function mentioned at the logistic model, expressing the limited carrying capacity of E2:

$$\frac{dy(t)}{dt} = \frac{q(t)}{1+q(t)} \cdot \mu_{max} \cdot (1 - e^{y-y_{max}}) \quad \textbf{(2-11)}$$

Here *q(t)* describes the accumulation of a "*critical, bottle-neck substance*", necessary to induce the exponential growth. Biologically, it represents a "*critical substance*" required for growth. It is not a specific molecule identifiable in all systems, but an abstract representation of internal factors such as enzymes, ribosomes, or metabolic intermediates that must accumulate to a threshold level before maximum growth can occur. The *q(t)* process is described assumed to obey a simple linear kinetics as shown by **Eq**(**2-12**) and **Eq**(**2-13**).

A key advantage of the model is that it offers an algebraic solution under constant environmental conditions. In a dynamic environment, however, it should be solved by numerical methods.

$$\frac{dq(t)}{dt} = \nu \cdot q \quad \textbf{(2-12)}$$

$$q(0) = q_0 \quad \textbf{(2-13)}$$

where *v* is the accumulation rate of the aforementioned "*critical substance*".

The simplification $v = \mu_{max}$ makes the model more suitable for practical curve fitting procedures.

The following reparameterizations of q_0 into a_0 (which is a measure of the initial physiological state of the cells) have biological interpretations (that will be discussed later in this section) and advantageous numerical/statistical properties that are useful when using the model for curve fitting:

$$\alpha_0 = \frac{q_0}{1+q_0} \quad \textbf{(2-14)}$$

$$h_0 = -Ln\,(\alpha_0) \quad \textbf{(2-15)}$$

The solution of the differential **Eq(2-11)** can be expressed in terms of $y(t) = \mathrm{Ln}(x(t))$ the natural logarithm of the cell concentration:

$$y(t) = y_0 + \mu_{max} \cdot A(t) - \frac{1}{m} \cdot \mathrm{Ln}\left(1 + \frac{e^{m \cdot \mu_{max} \cdot A(t)} - 1}{e^{m \cdot (y_{max} - y_0)}}\right) \quad \textbf{(2-16)}$$

where $A(t)$, the integral of $\alpha(t)$, can be considered as a gradually delayed measure of time:

$$A(t) = t - \lambda + \frac{1}{n \cdot \nu} \cdot \mathrm{Ln}\left(1 - e^{-n \cdot \nu \cdot t} + e^{-n \cdot \nu \cdot (t-\lambda)}\right) \quad \textbf{(2-17)}$$

The $A(t)$ function together with the post-inoculation conditions, describes the lag phase. Assuming that the specific growth rate responds instantaneously to environmental changes, this model can describe microbial growth kinetics under conditions where parameters such as pH, water activity, and temperature vary over time. A key advantage of the model is its ability to simulate microbial behavior in dynamic environments. m and n are tuning (curvature) parameters.

The four parameters of this model (y_0, y_{max}, μ_{max}, and λ) can be interpreted as follows:

μ_{max}: the maximum specific growth rate is an autonomous parameter, characterizing purely the intrinsic ability of the bacteria to grow under the E2 environment, independently of their prior history. This reflects the expectation that, after any necessary adaptation, cells will ultimately grow at a rate determined entirely by the current environment.

y_{max}: the final cell concentration is also an autonomous (history-independent) parameter but is of considerably less practical importance than the maximum specific growth rate, as food microbiology primarily focuses on low cell concentration levels.

y_0: the initial cell concentration is purely history dependent. In experiments, it is set up by the experimenter and can be relatively easily estimated. In real food, however, its estimation

can be complicated, which can cause difficulties when estimating the error of predictions of bacterial concentration in the actual environment.

λ: this is the time when, for the adjustment function, $\alpha(\lambda) = 1/2$. It is affected by both the history (E1) and the growth environment (E2). Baranyi and Roberts (1995) *reparametrized* the system and introduced the $h_0 = \mu_{max} \cdot \lambda$ quantity that can quantify the work to be done during the lag phase. This parameter has been shown to have no correlation with temperature (and therefore neither with μ_{max}) in the normal physiological growth region. It reflects certain aspects of the inoculation procedure (for example, inoculating from early or late stationary phase?), just as the inoculum level does. Taking it as a temperature-independent constant represents the assumption that, at both higher and lower temperatures, the same physiological processes "*work to be done*" occur during the lag phase, only at higher or lower rates.

A rescaling of the h_0 parameter, $\alpha_0 = \exp(-h_0)$, is a sort of "*suitability*" parameter, between 0 and 1, quantifying how much the history of the cells suitable to the actual environment. $\alpha_0 = 1$ means optimum history, when there is no lag at all ($\lambda = 0$); and $\alpha_0 = 0$ marks the infinitely long lag situation (no growth). Therefore, the system has two initial values: y_0 and α_0 (or h_0). With this concept, the lag obviously depends on both history and the actual environment shown by the simple **Eq(2-18)**

$$\lambda = \frac{h_0}{\mu_{max}} = \frac{-Ln(\alpha_0)}{\mu_{max}} \qquad \textbf{(2-18)}$$

It can be also shown that the $\alpha_0 = \exp(-\mu_{max} \cdot \lambda)$ quantity expresses the fraction of cells that would have been able to grow, without lag, into the same exponential phase . Therefore, for example, $\alpha_0 = 0.04$ means that if only 4% of the cells grow, they would reach a certain (high) concentration level at the same time as the actual growth curve, if those 4% can grow without lag.

Altogether, this model provides a biologically interpretable and mathematically robust framework for describing microbial growth under both static and dynamic conditions, making it a widely used primary model in predictive microbiology.

2.8. Secondary Models

Secondary models describe how the parameters of a primary model (such as maximum specific growth rate or lag time) change in response to changes in the environmental factors

like temperature, pH, water activity (aw), packaging atmosphere, or the presence of preservative (e.g. organic acids). In predictive microbiology, secondary models provide the essential link between microbial kinetics and the conditions under which growth occurs. While often applied to growth parameters, secondary models are also used for inactivation kinetics.

Various secondary functions have been proposed in literature. For the temperature effect, the model of Ratkowsky et al. (1983), and the Cardinal Temperatures Model 'CTM' of Rosso et al. (1993) are the most widely used. For multivariate cases, when including the impact of different factors to predict the maximum specific growth rate, second order polynomial surface models can be used (e.g. *ComBase*) or the "*gamma*" concept of Zwietering et al. (1996) including different gamma terms reflecting the impact of each factor. These gamma terms can be expressed through different equations including the Ratkowsky Square Root Model, the Cardinal model and others..

2.8.1. Ratkowsky's Square-root Model

Ratkowsky et al. (1982) suggested a linear relationship between the square root of growth rate and temperature covering the sub-optimal growth area **Eq(2-19)**

$$\sqrt{\mu_{max}} = b \cdot (T - T_{min}) \quad \textbf{(2-19)}$$

$T < T_{min} \Rightarrow \mu_{max} = 0$

where T is the temperature explanatory variable; b is the slope of the linear model; $T_{\min}$ is a hypothetical temperature considered as the theoretical minimum temperature for growth

In 1983, Ratkowsky *et al*. expanded this model to include the entire biokinetic range of temperatures.

$$\sqrt{\mu_{max}} = b \cdot (T - T_{min}) \cdot \left(1 - e^{(-c \cdot (T_{max} - T))}\right) \quad \textbf{(2-20)}$$

$T < T_{min} \Rightarrow \mu_{max} = 0$

$T > T_{max} \Rightarrow \mu_{max} = 0$

where T is the temperature explanatory variable; b is the slope of the linear model; $T_{\min}$ is a hypothetical temperature considered as the theoretical minimum temperature for growth; T_{max} is the maximum temperature for growth. The turn of trend around the optimum temperature

is quantified by an empirical curvature parameter, *c,* that controls how sharply this trend changes around zero. This model will be referred to from now on as RKM.

2.8.2. Cardinal Temperatures Model

The Cardinal Temperatures Model (CTM) reads as:

$$\mu_{max} = \frac{\mu_{opt} \cdot (T-T_{max}).(T-T_{minC})^2}{(T_{opt}-T_{minC}).[(T_{opt}-T_{minC}).(T-T_{opt})-(T_{opt}-T_{max}).(T_{opt}+T_{minC}-2T)]} \quad \textbf{(2-21)}$$

$T_{minC} < T < T_{max} \Rightarrow \quad \mu_{max} > 0$

$T < T_{minC} \Rightarrow \ \mu_{max} = 0$

$T > T_{max} \Rightarrow \mu_{max} = 0$

where the four parameters are the minimum T_{minC}, optimum T_{opt} and maximum T_{max} growth temperatures, and μ_{opt} (the specific growth rate obtained at T_{opt}).

The shape of RKM and CTM is like that of an asymmetric triangle.

2.9. Tertiary Models

Early definitions of tertiary models described them as user-friendly computational tools that combined primary and secondary models into interactive software systems (Whiting, 1995). In this sense, tertiary modelling was seen primarily as an application layer, enabling non-specialists to access predictive microbiology data and model outputs. Examples of such tools include the Pathogen Modelling Program *PMP* (USA), Growth Predictor *GP* (UK), *ComBase* (UK/USA), *Sym'Previus* (France), the Microbial Responses Viewer *MRV* (Koseki, 2009), the Unified Growth Prediction Model *UGPM* (Psomas et al. 2011), and the Seafood Spoilage Predictor *SSP* (Dalgaard et al. 2002).

Baranyi et al. (2017) later argued that the term tertiary model should not refer to software interfaces and implementations, or a combination of primary and secondary models. Instead, logically, it should describe how the parameters of a secondary model depend on category variables, for example, the species or strain of the microorganism, or the type of food matrix.

One early example consistent with this more mathematical definition was presented by Rosso et al. (1993) who identified strong correlations between the cardinal values of multiple

species. Similarly, Baranyi et al. (2017) reported correlations between two parameters of Ratkowsky model (T_{min}) and (b), for different strains of *Bacillus cereus*.

2.9.1. Correction Factor and Bias Factor

Based on the assumption that the cardinal temperatures are matrix-independent, predictions on the maximum specific growth rate in food can then be generated via a correction factor typical of that food (Buss da Silva et al. 2017). Because the cardinal temperatures are assumed to be matrix-independent, the correction factor c_f then becomes the ratio between the optimum maximum specific growth rate in the food in question and that in culture medium or broth:

$$c_f = \frac{\mu_{opt}{}^{(food)}}{\mu_{opt}{}^{(broth)}} \quad \textbf{(2-22)}$$

On the other hand, the bias factor, introduced by Ross et al. (1996), is a widely used metric to evaluate the performance of predictive models. It quantifies the systematic deviation (over- or underprediction) between predicted and observed values and is particularly useful when assessing model performance in environments different from those used for model development (e.g., broth versus food systems).

$$Log(b_f) = \frac{1}{n}\sum_{i=1}^{n} Log\left(\mu_i{}^{(observed)}\right) - Log\left(\mu_i{}^{(predicted)}\right) \quad \textbf{(2-23)}$$

2.10. Error Analysis

A regression output typically includes parameter estimates ($\hat{p}$) and their standard errors $SE(\hat{p})$ as well as the standard error of fit SE_{fit}. These are two important but different terms to evaluate the reliability of a fitting. In this section both these terms will be clarified as they are the center of our model performance analysis.

2.10.1. Standard Error of Parameter

When regressing measured data, the parameters of the fitted model are estimated by some numerical method. Part of this method is a quantification of the confidence in the parameter estimates, which is expressed by their standard error and confidence interval. This is, however, another estimation from the data, which depends on experimental and sampling randomness, as well as the robustness of the regression algorithm.

In predictive microbiology, as well as, generally, in biological studies, the 95% confidence interval for a $\hat{p}$ estimate is often approximated by $\hat{p} \pm 2 \cdot SE(\hat{p})$. This approach finds its root in the Student's theorem, which says that, ***if*** the raw data were distributed around the assumed model according to the normal distribution, with constant variance and zero mean, then the *t*-score of the estimate, $\hat{p}/SE(\hat{p})$, follows the Student distribution, which, for large degree of freedom, is close to normal. However, normality assumption often does not hold for the raw data, and the regression is not always linear. Despite this, confidence intervals are sometimes preferred to standard errors because of their visual representation, even though it would be more robust to evaluate the reliability of the fitting estimates based on the standard error.

It is a common scenario that a model accurately describes the biological system but if the test points to be fitted (observations) are poorly chosen or/and the regression algorithm is not robust enough, then the standard errors of the parameter estimates become unrealistic. The opposite of this is, when the chosen regression is robust, such as the linear regression, but it does not reflect reality with sufficient accuracy; for example, when inherently sigmoid growth curves are fitted by a linear model.

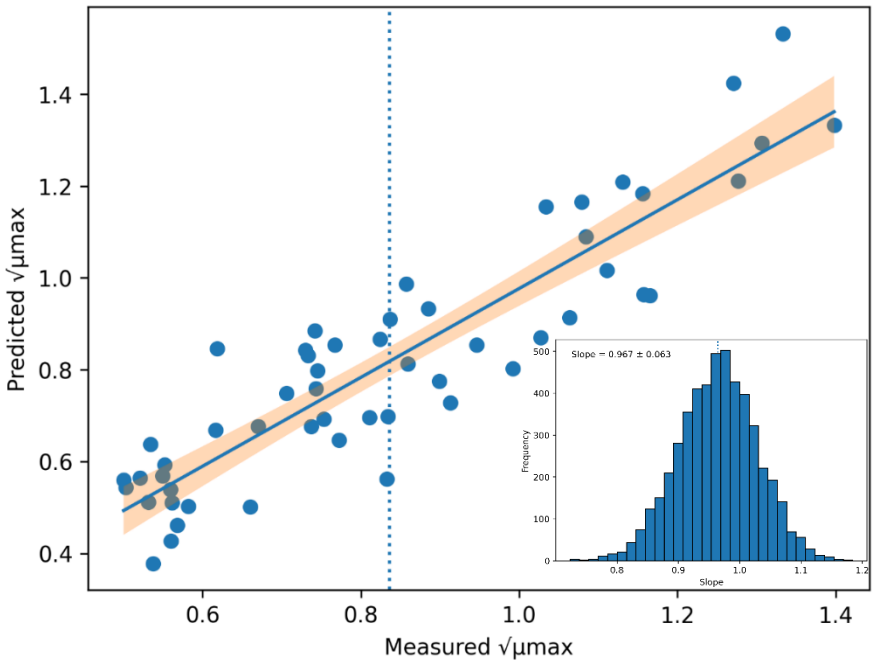

Fig. 2.8 *Demonstration of the confidence interval of the predicted mean (orange band) and the distribution of the slope parameter (inlet), on data from https://combase.errc.ars.usda.gov/.*

Figure (**2.8**) illustrates the concept of the standard error of a model parameter. The dataset (blue markers) is fitted with a linear model (blue solid line). The shaded orange region

represents the 95% confidence band of the fitted line, indicating the range within which the true mean response is expected to lie. This uncertainty arises because the fitted line is only one of many possible fits that could result from repeated experiments under the same conditions. Each of these fits would yield slightly different parameter estimates (e.g., slopes), forming a sampling distribution of that parameter. The inserted blue histogram inlet visualizes the distribution in case of the slope: its spread (quantified by its standard deviation) corresponds to the standard error of the estimated slope.

2.10.2. Standard Error of Fit

The standard error of the fit, SE_{fit}, also called the root mean square error (RMSE) of the residuals, measures how far, the observed responses are from the model's predicted values (**Fig. 2.9**). In fact, it is an estimate of the σ standard deviation of the normal distribution of raw data assuming it spreads around the correctly chosen model according to the normal distribution with a constant variance. A small SE_{fit} means the fitted values are close to the observed ones. Unlike $SE(\hat{p})$, which measures uncertainty in a single parameter, SE_{fit} is an overall performance measure affected by the suitability of the model structure to describe the dataset, the data variability and the experimental noise. Note that a model can achieve a small SE_{fit} while having large uncertainty in one or more parameters, or conversely, have precise parameter estimates but still fit the data poorly, for example because of a lot but noisy measurements (large σ).

In the case of linear regression, the standard error of the slope-estimate is proportional to the standard error of the fit: $SE(slope) = SE_{fit}/C$ (for detailed proof see Nau, 2016), where C is a constant, depending solely on the sampling times.

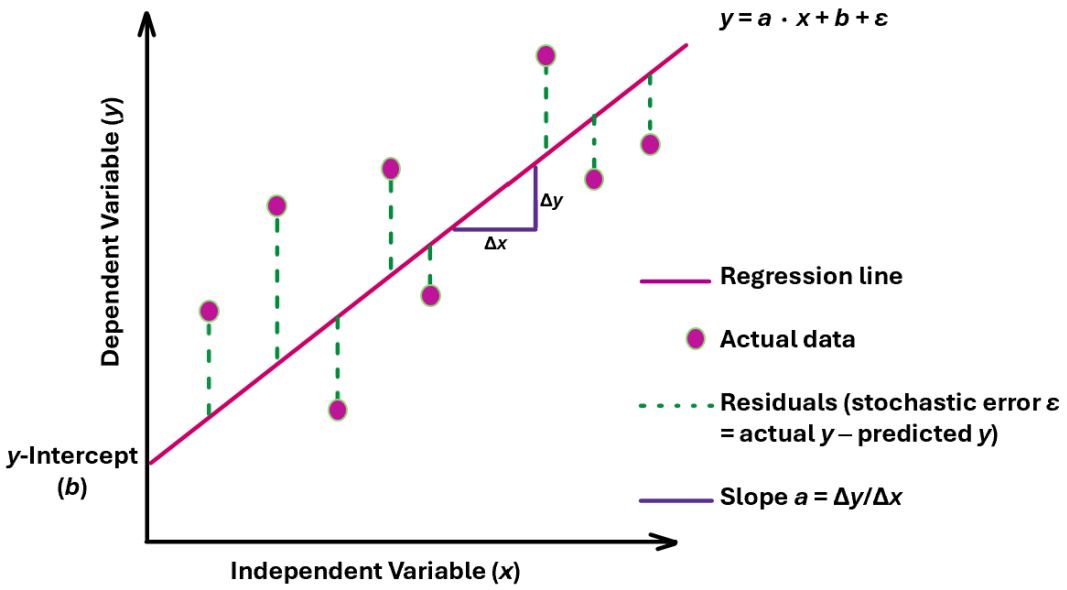

Fig. 2.9 *Demonstration of the residuals*

2.10.3. Error Sources

When assessing confidence in estimates and predictions, the total error arises from various sources, form them we mention:

(1) experimental factors include sampling-induced variability (e.g. pipetting imprecision, inconsistent sample size or handling), (2) bacterial-culture-induced variability (e.g. subculture history, inoculum heterogeneity, cell-to-cell differences in division time), (3) bench- and method-factors (e.g., laboratory environment, operator technique, equipment calibration), (4) numerical sources arising during data handling, model selection and parameter estimation (choice of model, regression algorithm).

Significant effort has been made (Corry et al. 2007; Akkermans et al. 2018; Nauta, 2000) to define and categorize these sources.

In predictive microbiology and quantitative microbial risk assessment, error separation is most commonly framed through the distinction between *variability* and *uncertainty*, as formalized by Nauta (2000). *Variability* refers to the inherent heterogeneity of biological systems, including differences between strains, individuals, environments, and time points, and is therefore considered irreducible. In contrast, *uncertainty* represents the quantitative expression of lack of knowledge, arising from limitations in data, measurement, model structure, or parameter estimation, and can in principle be reduced by additional information or improved methods

In this work, we chose to divide them according to Rockaya and Baranyi (2025) into two broad categories:

- *Wet* Source: analogously to the term "*wet science*," this category encompasses all factors contributing to variability and uncertainty in observations related to biology and/or the conditions under which the observations were made. Examples include strain- or cell-to-cell variability, environmental randomness, the accuracy of laboratory methods or equipment, and even the expertise of the laboratory operator.
- *Dry* Source: this category covers all factors that introduce errors or uncertainties in the algorithms used for data cleaning, inference and estimations, i.e. the entire process from data recording to processing and numerical calculations/approximations.

The *wet* and *dry* classification proposed in this work is conceptually different to Nauta's (2000) framework but together they provide complementary perspectives: the *variability*/*uncertainty* framework characterizes the epistemic nature of the error (irreducible vs. reducible), whereas the *wet*/*dry* framework identifies its source (physical vs. computational).

According to these definitions, both variability and uncertainty may originate from either wet or dry sources. For example, variability may arise from biological heterogeneity (wet source) or from stochastic elements introduced during numerical procedures, such as Monte Carlo sampling (dry source), while uncertainty may stem from experimental imprecision (wet source) or from model structure and parameter estimation (dry source).

This orthogonality motivated the use of the wet/dry classification in this work, as it provides a more direct and clear-cut identification of the origin of error, avoiding conceptual overlap between categories.

In practice, error often reflects a combination of both origins, arising from intertwined experimental and computational contributions. Conceptualizing error within the wet/dry framework and considering its propagation through *successive stages* of data generation and analysis, facilitates clearer interpretation and more effective identification of potential mitigation strategies. *Wet* sources can be addressed through improved experimental design and measurement protocols, while *dry* sources can be reduced through improved statistical methods, model formulations, and numerical implementations.

2.11. Limitations of Predictive Microbiology Models

Predictive models are valuable decision-support tools for optimizing experiments, informing policy, and guiding food-safety and quality interventions. Nevertheless, their practical utility is constrained by the following limitations:

- Models are often developed from optimal laboratory media (broth) and then extrapolated to complex, real-food matrices, reducing predictive accuracy in practice.

- Datasets of varying quality: not every existing growth and inactivation dataset adequately covers the diversity of environmental factors such as food types and temperatures, and processing encountered in real systems.

- Lack of standardized protocols and databases. Variation in treatment procedures and measurement methods as well as lack of databases with compatible ontology where related data can be documented, create systematic inconsistencies that hinder data pooling and model comparison.

- Bio-composition and microstructure of food can modulate microbial behavior in ways that current models which are trained mainly on broth cannot capture yet.

- Industrial processing and storage commonly involve fluctuating temperatures, treatments, and time-varying stresses which need to be addressed using dynamic models.

- Poor uncertainty quantification and reporting. Many papers report on point estimates without robust confidence or prediction intervals or sensitivity analyses and are still used in risk-informed decision making, increasing the risk for poor decisions.

- Empirical models may fit data well but lack biological parameters or mechanistic bases, reducing interpretability and generalizability across matrices and conditions.

Addressing these limitations is important to improve the reliability and applicability of predictive models but it is challenging to address them all. This research aims to target some of these gaps to enhance model performance in real-world food safety applications.

3. MATERIALS AND METHODS

3.1. Development of a Database Ontology for Microbiology and Composition of Plant-Based Dairy 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 food alternatives.

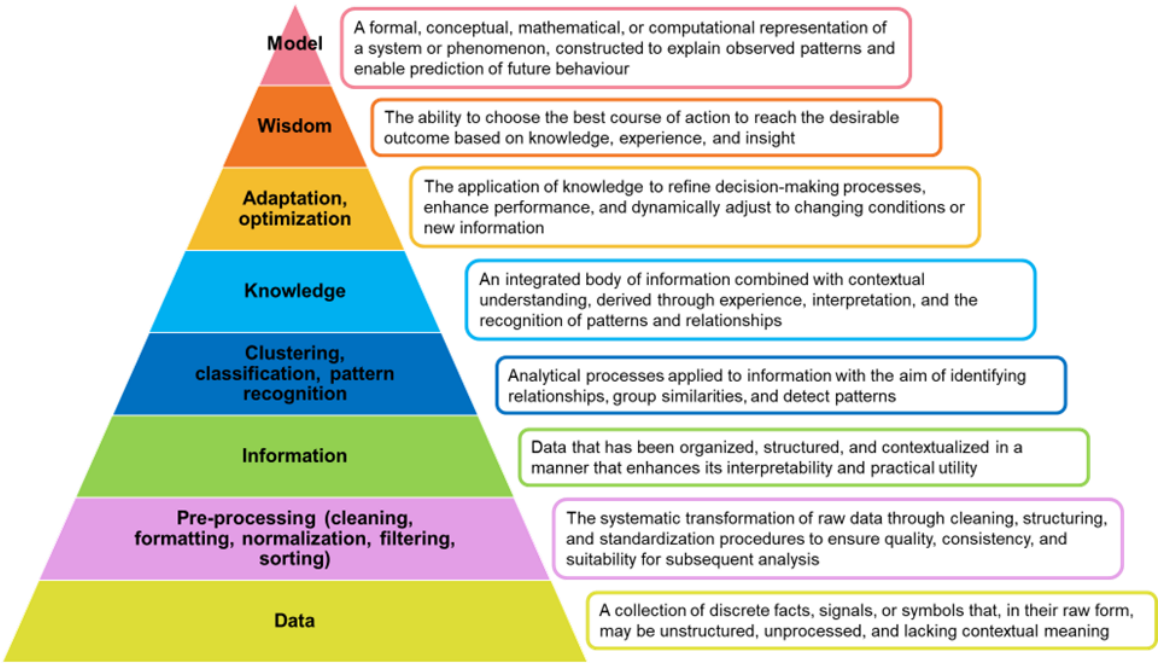

Fig. 3.1 *From data to models*

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.

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

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).

3.2.1.1. Data Simulation

Three primary models have been described in (**Section 2.7**): the Linear Model (LM, **Eq. 2-3**); the Baranyi-Roberts model (BRM, **Eq. 2-11**); and the reparametrized Gompertz function (GF, **Eq. 2-9**). We used them to run a *Monte-Carlo* simulation (Mooney, 1997).

Seven sampling times t_i (i = 1…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. The parameter values were: $\mu_{\max}$ = 0.5 (h^{-1}), y_0 = 5 (Ln cell/ml). For BRM and GF also λ = 8 (h) and $y_{\max}$ = 23 (Ln cell/ml). For each time t_i the response y_i , obtained by each model, was perturbed by normally distributed random errors (Gaussian noise) with mean zero and standard deviation σ = 1 to mimic the random measurements error.

3.2.1.2. 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).

3.2.1.3. Histogram Comparison

From each simulated fit, the growth parameter and its standard error were estimated (total = 1000 μ_{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. This comparison was feasible by calculating the skewness (i.e. the closer the skewness of a histogram is to zero, the more this histogram is symmetric and close to the t-distribution of a linear regression).

3.2.1.4. 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 (see **Section 2.7.1** and Nau, 2016).

To compare with G-Reg and BR-Reg, we reparametrized BRM and GF using the intercept parameter (h_0) introduced by Baranyi and Robert (1994) and explained in detail in **Section 2.7.4**). As can be readily seen, h_0 is the equivalent of the intercept b, and μ_{max} is analogous to the slope a of linear model.

We then examined the correlation between h_0 and μ_{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.

3.2.2. 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.

3.2.2.1. Data Simulation

The two secondary models described in (**Section 2.8**): Ratkowsky's Square-root Model (RKM, **Eq. 2-20**) and the Cardinal Temperature Model (CTM, **Eq. 2-21**) were used to run a *Monte-Carlo* simulation.

Eight sampling temperatures T_i (i = 1….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°C, c = 0.5, T_{max} = 43°C ; for CTM: T_{min} = −3°C, μ_{opt} = 1.17 h^{-1}, T_{opt} = 36°C, T_{max} = 43°C.

We used the same values of T_{min} and T_{max} for both models aiming to have similar μ_{max}.

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

As a simplification, it was assumed that the temperature would not impact the μ_{max_i} variability observed for each temperature level.

3.2.2.2. 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 μ_{opt} < 50% ; (iii) standard error of T_{min} , T_{opt} , or T_{max} < 10°C.

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

3.2.2.3. 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: $95\%\ CI = \hat{p} \pm 2 \cdot SE(\hat{p})$.

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.

3.2.2.4. 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 4-2**), 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.

3.2.2.5. 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: $b = 0.01$, $T_{min} = 0°C$, $c = 0.6$, $T_{max} = 51°C$; for CTM: $T_{min} = 0°C$, $\mu_{opt} = 1.2\ h^{-1}$, $T_{opt} = 30°C$, $T_{max} = 51°C$.

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 3.2.2.2.**).

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.

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

3.3.1. Characterization of Plant-based Dairy Alternatives

To capture the variability originating from the food matrix, we used three plant-based dairy 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 International; Debrecen, Hungary) 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; Budapest, Hungary) for pH and a water-activity meter (rotronic HygroLab C1; Bassersdorf, Switzerland) for water activity (aw) measurement.

3.3.2. Inoculum Preparation

B. licheniformis DSM 13 reference strain (Leibniz Institute DSMZ; Braunschweig, Germany) was used in this study. A working culture of the strain was produced and maintained at −20°C on Microbank™ cryobeads (Pro-Lab Diagnostics; Ontario, Canada) 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 in a water bath at 37°C (ESCO; Barnsley, UK) while shaking under aerobic conditions for 8 hours. After the incubation period, 100 µL of the culture 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.

3.3.3. 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 10^2 CFU/mL. For the 'AH' and 'BHI' media, triplicates were prepared at all temperatures

ranging from 15 to 55°C. For the 'AA' and 'CA' media, triplicates were prepared at cold temperatures (15 to 21°C) but only duplicates at high temperatures (29 to 55°C).

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

To enumerate the changes in *B. licheniformis* concentration over time, plate counting method was used by collecting 1 mL of each inoculated flask and adopting the adequate dilutions before pour-plating on BHI Agar (Liofilchem; Roseto degli Abruzzi, Italy) to obtain between 30 and 300 colonies on the enumeration plates..

Pour-plating was preferred to surface plating because, under aerobic conditions, *B. licheniformis* rapidly spread 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, regular samples were taken for plate counting, aiming at a minimum of 3 points in the lag phase, 6 points in the exponential phase, and 3 points in the stationary phase.

3.3.4. 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 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).

The temperature selection was optimized to test a broad range of temperatures directly in food matrices not only in broth, allowing us to verify whether growth patterns observed in broth are consistent across different plant-based dairy alternatives and to improve the accuracy of growth models for *B. licheniformis* in both broth and plant-based dairy alternatives.

3.3.5. Model Fitting and Statistical Analysis

The measured Ln-count data was fitted with BRM **Eq(2-11**) as a function of time (primary modelling), and the estimated μ_{max} for every matrix in every studied temperature were fitted

using CTM **Eq(2-21)** 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(3-1)** between two matrices at a time, were performed using the parameter estimates and their standard errors (Welch, 1974):

$$t_{welch} = \frac{Estimate_A - Estimate_B}{\sqrt{SE^2{}_A + SE^2{}_B}} \quad \textbf{(3-1)}$$

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(3-2)** to account for unequal variances.

$$df = \frac{\frac{SE^2{}_A}{n_A} + \frac{SE^2{}_B}{n_B}}{\frac{\left(\frac{SE^2{}_A}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{SE^2{}_B}{n_B}\right)^2}{n_B - 1}} \quad \textbf{(3-2)}$$

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.

4. RESULTS

4.1. Final Framework of the *Pbase* Ontology

The final framework of *Pbase* (**Fig. 4.1**) 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).

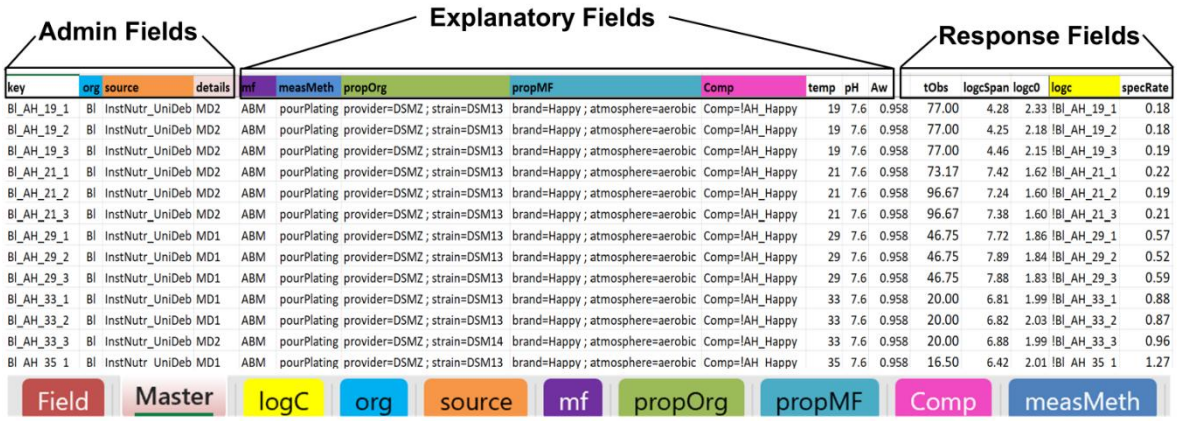

Fig. 4.1 *Screenshot of Pbase*

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. 4.1**) by unique colors representing each column and its components are described in (**Table 4-1**).

As of the time of writing this dissertation, *Pbase* contained 415 records, each corresponding to a distinct *B. licheniformis* growth curve. Of these, 133 records originate from our previously published experiments (Rockaya et al., 2025). Each record represents the growth of a bacterial population at a specific temperature, along with associated pH and aw values. These records are supported by two types of tables: a dynamic one containing the raw logarithmic cell counts (LogCount) and their sampling time and another detailing the chemical composition of the growth matrix. The remaining records were obtained from an external source (Misiou et al., 2023), which provided quantitative growth parameters such as maximum specific growth rates (μ_{max}) and environmental factors such as pH, and aw. However, these external records do not include the raw Log C data nor detailed compositional information of the matrices used.

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.

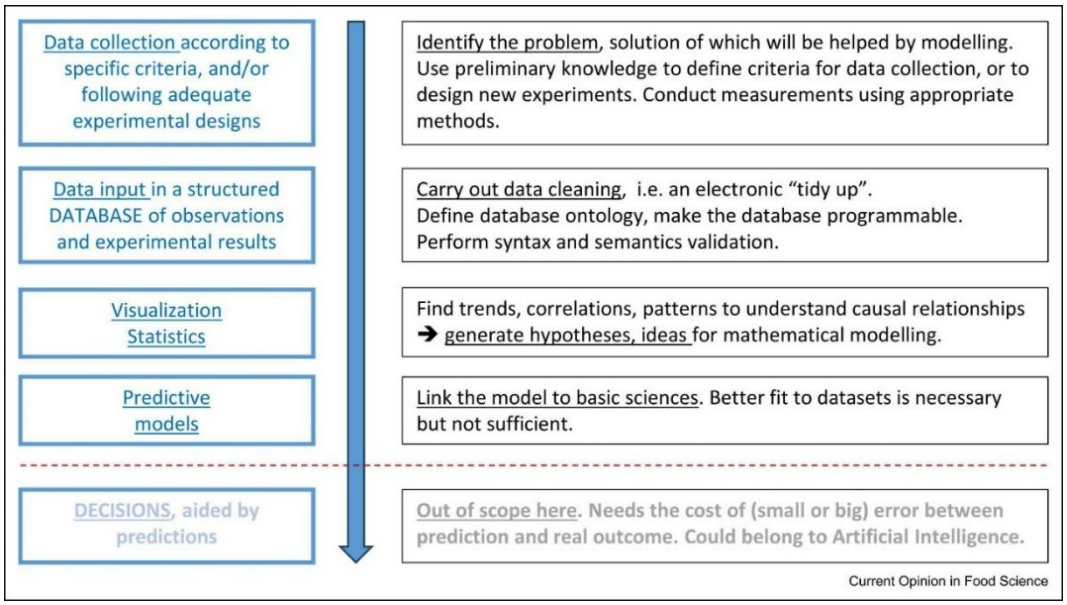

Fig. 4.2 *Milestones when making sense of data*

Source: *Baranyi et al. 2024*

Figure 4.2 conceptualizes the proposed framework within a broader data-to-knowledge pipeline, illustrating the sequential transformation from raw data acquisition to model-informed insights. The process begins with data collection under well-defined experimental designs and criteria, ensuring reproducibility and relevance. These data are subsequently structured and stored within a database environment, where systematic data cleaning and validation procedures (including syntactic and semantic checks) are essential to ensure consistency and reliability. Once curated, the dataset enables exploratory analysis through visualization and statistical methods, facilitating the identification of trends, correlations, and potential causal relationships. These insights form the basis for hypothesis generation and the development of predictive models, which are further grounded in fundamental scientific principles rather than relying solely on empirical fitting. The framework emphasizes that, while model development is a critical step, decision-making based on model predictions requires additional considerations, such as uncertainty quantification, the evaluation of discrepancies between predicted and observed outcomes and the potential of utilizing AI to help in this step.

Although *Pbase* 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 dairy or animal-protein alternatives.

This work is wished to be a small initiative toward 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.

Table 4-1: Description of *Pbase* ontology

Field	Description	Example
key	identification unique code of the record, built by concatenating key information to summarize each record	Bl_CA_37_2 which stands for Bl: *Bacillus licheniformis* ; CA: abbreviation of a commercial coconut milk ; 37: temperature, in C, at which the bacteria grew ; 2: the number of biological replicates. Any alphanumeric string is allowed for the key, as long as it is unique.
Org	name of the organism	Bl: *Bacillus licheniformis*
source	source of the data	InstNutr_UniDeb : Institute of Nutrition, University of Debrecen
details	preparation of the experiment	70mL of media/milks were inoculated with 100microliter of *Bacillus licheniformis* suspended in then transferred into….etc,.
mf	medium/food where the experiment was done	TSB: Tryptone Soy Broth ABM: Almond Milk
measMeth	method to measure response	plate count, pour plating.
propOrg	properties of the organism	strain, provider, history.
propMF	properties of the matrix	brand, treatment, preservatives…
Comp	chemical composition of the medium	carbohydrates, fatty acids, minerals.
Temp	temperature	45°C
pH	pH of the medium	7.5
Aw	Water activity of the medium	0.998
tObs	interval of observation (duration of the experiment)	60 hours
logcSpan	span of log10(orgConcentration) during tObs	9.04
logc0	log10 of initial orgConcentration	2.23
logc	log10(orgConcentration)	!Bl_CA_37_2 is a pointer referring to a table of temporal growth measurements (logCount vs time curve) in a complementary logC sheet
specRate	specific growth/death rate of the organism	1.4 h^{-1}
nVal	sample size, number of observations; number of points per curve	13
inputBy	the responsible person who inputted the data	MR: Maha Rockaya
year	year of data submission to the database	2025

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

Parameter estimation in predictive models should always be accompanied by quantifying uncertainty (e.g., confidence intervals, standard errors), keeping in mind that these are also estimates from the raw data. The more robust and reliable the error estimation, the stronger its contribution to decision-making in the food sector. In this section, we compared the statistical properties of error estimates obtained by the most used primary and secondary models to evaluate how closely they align with the benchmark properties of linear regression and its t-distribution and how rescaling and reparameterization enhance it.

4.2.1. Analysis of Primary Models

4.2.1.1. 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. 4.3**), 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. 4.3**) 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 used to fit 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 (**4.3**) 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.

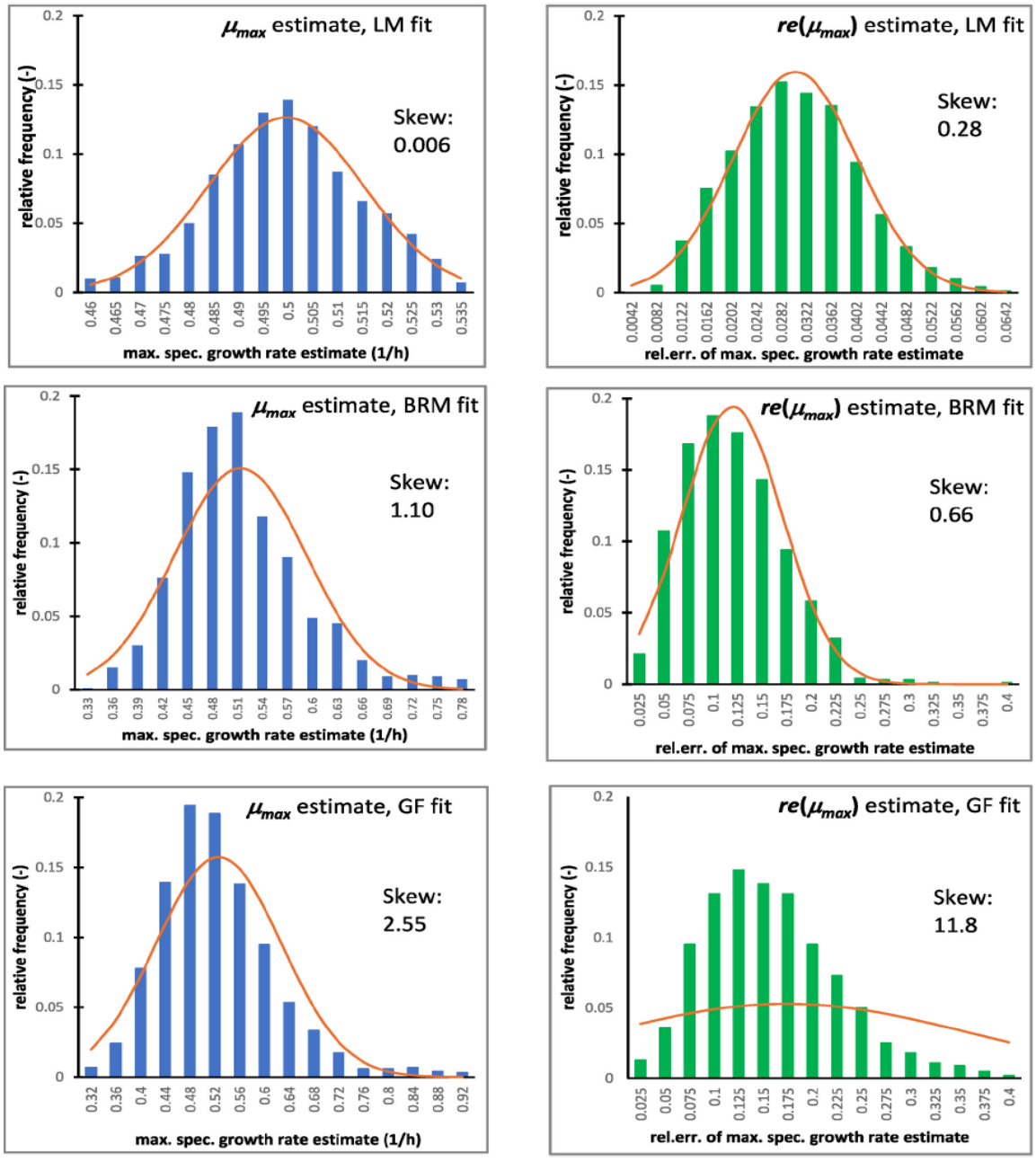

Fig. 4.3 *Histograms for the estimates of the maximum specific growth rates and their relative errors in case of LM, BRM and GF functions*

4.2.1.2. 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. 4.4**).

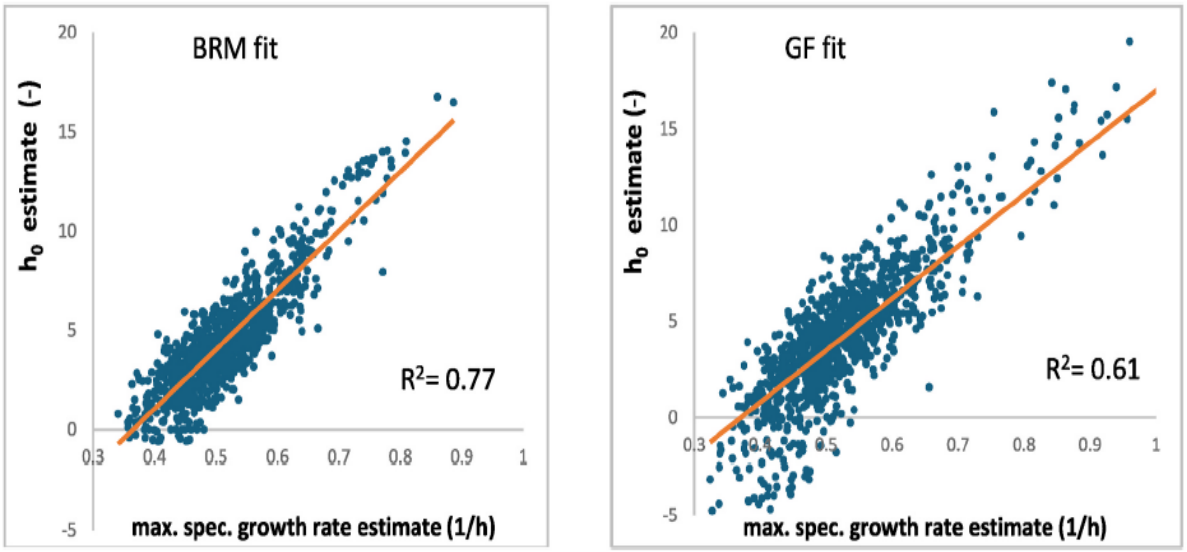

Fig. 4.4 *Correlation between the $\widehat{\mu_{max}}$ and $\widehat{h_0}$ estimates when fitting the simulated data in case of BRM and GF*

This is unrelated to their biological (*wet*) relationship where we would expect the opposite trend, the difference between the *dry* and *wet* correlations will be detailed later (see section **4.4.6.**)

4.2.1.3. Summary

The histogram analysis and the correlation patterns demonstrate that BR-Reg (BRM fitting) consistently performs more similarly to linear regression than G-Reg (GF fitting), regardless of which model generated the data. This is because BRM is itself a refinement of the LM and therefore inherits its more favorable statistical properties.

4.2.2. Analysis of Secondary Models

To evaluate the robustness of parameter estimation in secondary models and the impact of reparameterization on it, Monte Carlo simulations were performed by generating datasets of

μ_{max} versus temperature, as outlined in the *Materials and Methods* section. Each dataset was fitted using the Ratkowsky model (RKM) and the Cardinal Temperature Model (CTM). The resulting parameter estimates from 1000 fits were then used to construct histograms for each parameter. These were compared with the theoretical bell-shaped t-distribution, constructed from the same mean and standard deviation of the estimated sets.

4.2.2.1. 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. 4.5**). 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. 4.5**) and T_{max} and T_{opt} in CTM (**Fig. 4.6**) 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 (**4.5** and **4.6**) 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.

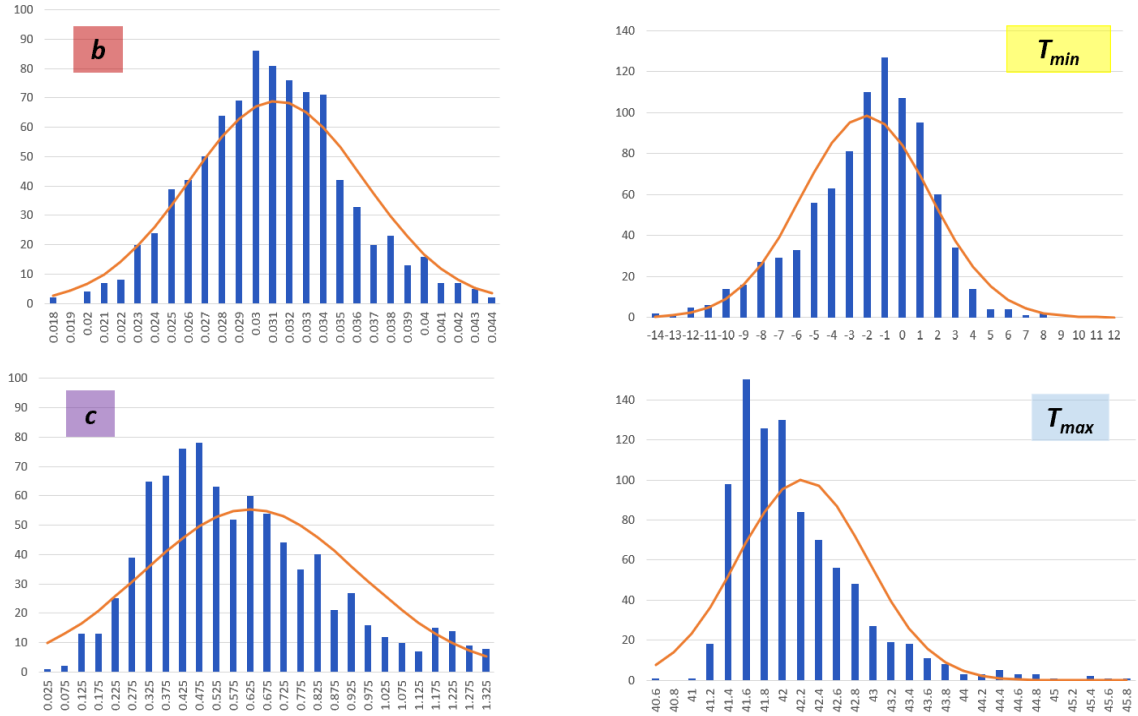

Fig. 4.5 *The simulated data fitted with RKM (before reparameterization)*

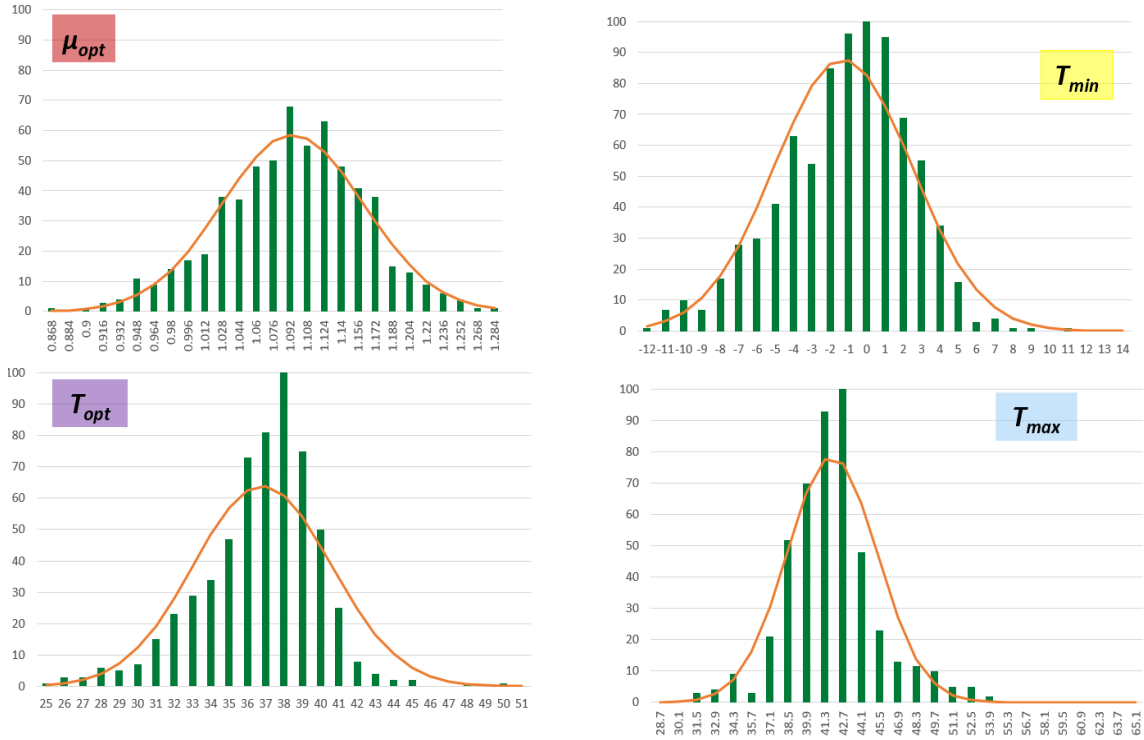

Fig. 4.6 *The simulated data fitted with CTM (before reparameterization)*

4.2.2.2. 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. 4.7**) and (**Fig. 4.8**).

The transformations were defined in (**Table 4-2**):

Table 4-2: The applied transformation on each parameter

	RKM *before* reparameterization	**RKM *after* reparameterization**	**CTM *before* reparameterization**	**CTM *after* reparameterization**
a_1	T_{min}	T_{min}	T_{min}	T_{min}
a_2	b	b	μ_{opt}	μ_{opt}
a_3	c	$c.T_{span}$	T_{opt}	q_c
a_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.

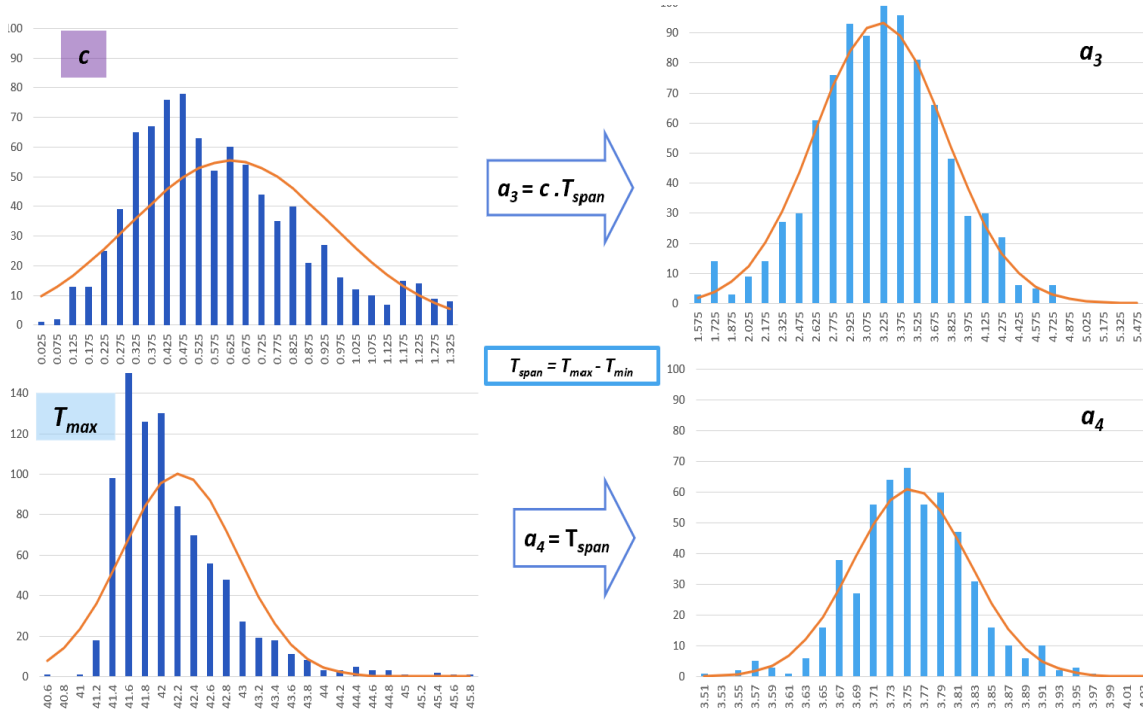

Fig. 4.7 *The simulated data fitted with RKM (after reparameterization)*

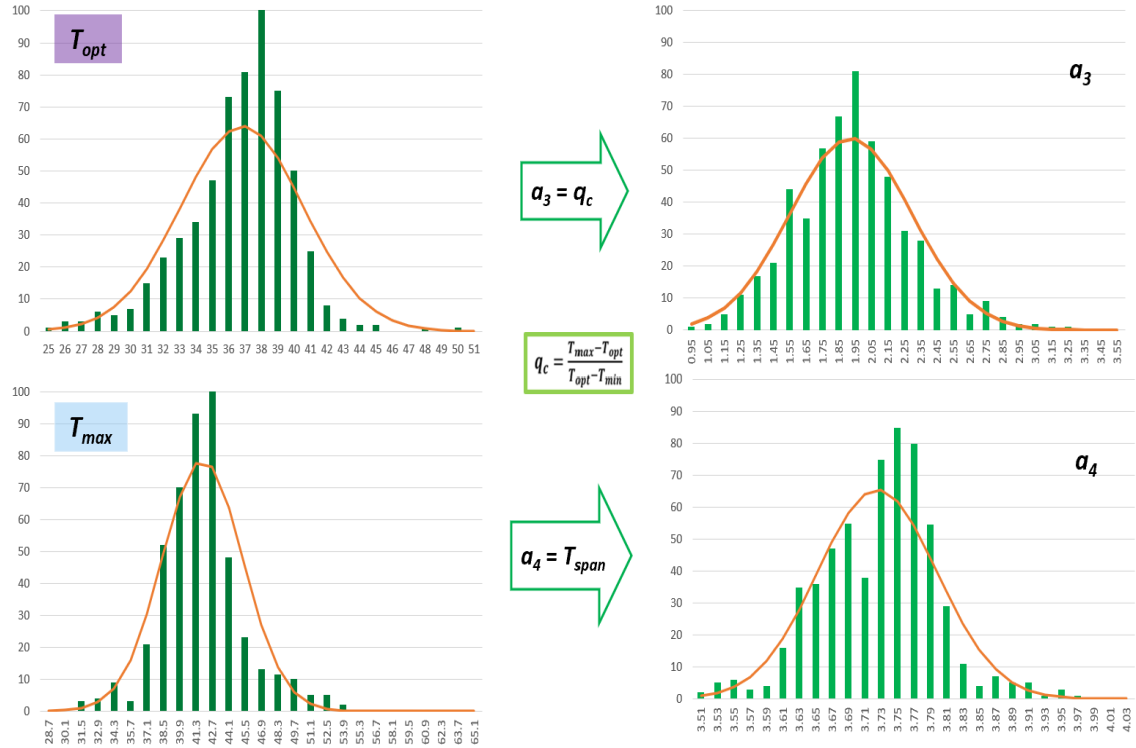

Fig. 4.8 *The simulated data fitted with CTM (after reparameterization)*

4.2.2.3. 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 3.2.2.2**). 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.

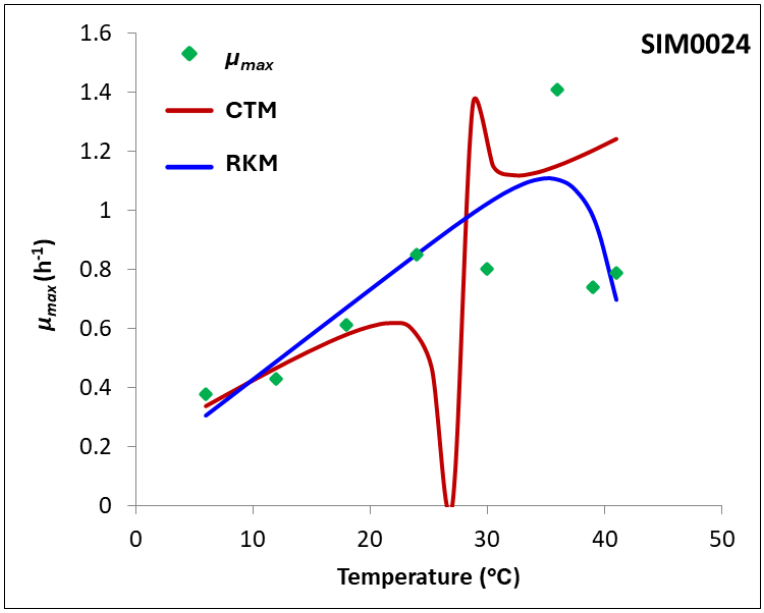

Fig. 4.9 *One example of the discontinuous fits of CTM*

4.2.2.4. Summary

Once reparameterized, both CTM and RKM provided symmetric, reliable histograms and accurate confidence intervals. RKM produced fewer failed fits overall, but CTM was similar only once suitable initial estimates were provided. Eventually, the results confirm that it is not the choice of model, but the choice of reparameterization that determines reliable error estimation and practical applicability. **Table (4-3)** summarizes the results of this section.

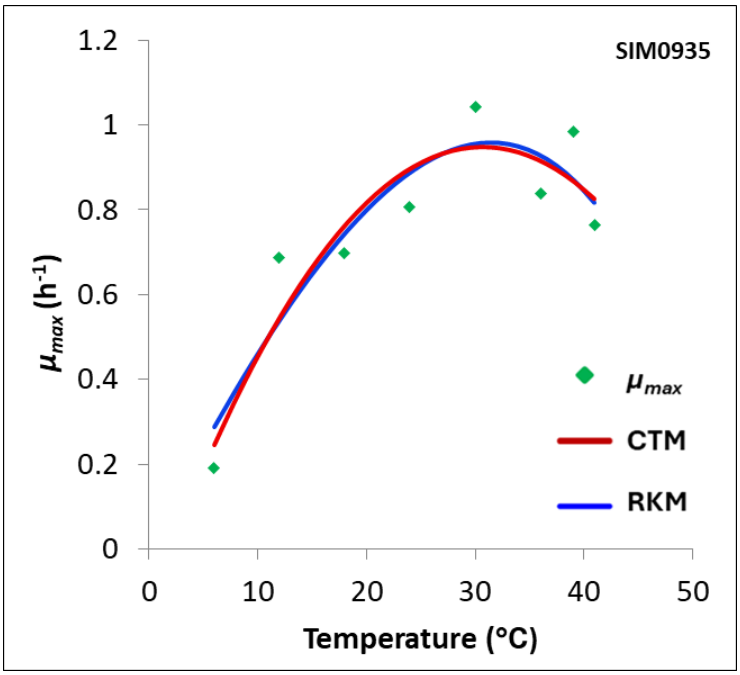

Fig. 4.10 *Visual example that there is no significant difference in terms of goodness of fit between the two compared models*

Table 4-3: Summary of the comparison between the two studied secondary models

POV	CTM	RKM
Goodness of fit	No difference	
Parameters Interpretation (How easy to find initial values)	All easy to interpret biologically	Exponent "*c*" is difficult
Iteration Convergence (Sensitivity to initial estimates)	Discontinuity → strange fits "*smaller range of initial estimates*"	No discontinuity → normal fits "*wider range of initial estimates*"
Excluded by the criteria	~ 200/1000	~ 100/1000
Need to be reparameterized	T_{max}, T_{opt}	T_{max}, c
Symmetry of Histograms with the original parameters	Closer to t-distribution "*more bad fits were excluded*"	Less close
Symmetry of Histograms (after reparameterization)	Equally good histograms	

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

4.3.1. Matrix Characterization

Table 4-4: 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 next to the pH values are the standard deviations

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 the inoculated flasks of AH during the stationary phase, which may be related to protein denaturation or to the production of the toxin and 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 the inoculated flasks of 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 in this study.

4.3.2. Primary Fitting and Model Performance

Plate count data (CFU/mL) over time, collected from experiments in BHI and the three plant-based dairy alternatives (AH, AA, CA) at all testes temperature (**Section 3.3.4**) 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. 4.11 a-b**)

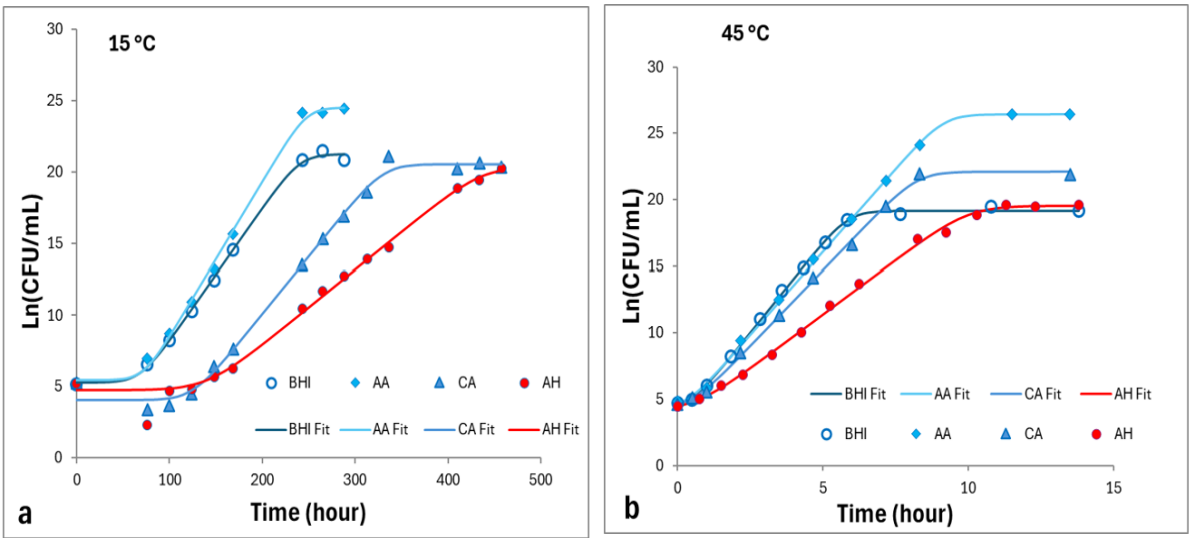

Fig. 4.11 a-b *LogCount data as a function of time in various plant-based matrices at 15°C (**a**) and 45°C (**b**), fitted by BRM*

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 AA consistently showing the highest values.

At 45°C growth was rapid and similar among the media, (**Fig. 4.11 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. 4.11 a**).

Interestingly, both AH and CA milks, 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**(**4-1**) did not show correlation with the temperature (*p = 0.576*; see **Fig. 4.12**).

$$RE(\mu_{max}) = SE(\mu_{max}) / \mu_{max} \quad (\textbf{4-1})$$

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**(**4-2**), with slope *c* corresponding to the mid-value of the relative errors (**Fig. 4.12**). This implies that the error of $Ln(\mu_{max})$ is constant, a desirable statistical property for secondary modelling later on.

$$SE(\mu_{max}) = c \cdot \mu_{max} \quad (\textbf{4-2})$$

The estimated *c* slope corresponds to a certain mid-value of the relative errors of the μ_{max} estimates.

Remember that the approximation, that the errors of the $Ln(\mu_{max})$ estimates are close to the relative errors of the μ_{max} estimates comes from the **Eq**(**4-3**) and **Eq**(**4-4**):

$$\frac{\mu_{obs}-\mu}{\mu} = \varepsilon \quad \Longrightarrow \quad \mu_{obs} = \mu\,(1+\varepsilon\,) \quad (\textbf{4-3})$$

$$\Longrightarrow Ln(\mu_{obs}) = Ln(\mu) + Ln(1+\varepsilon\,) \approx Ln(\mu) + \varepsilon \quad (\textbf{4-4})$$

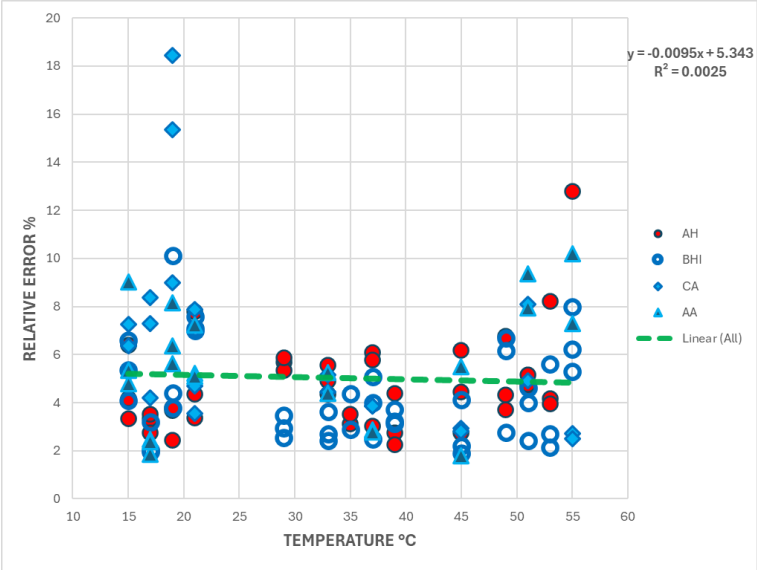

Fig. 4.12 *Relative errors of the maximum specific growth rate estimates, generated by the primary model fitting, against temperature, through all the four studied media*

Figure (**4.12**) shows that the relative error (RE) does not show a trend with the temperature ($p >> 0.05$). 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 were above the requirements of ISO guidelines for primary model parameters (FDIS-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. 4.12**). Across all conditions where replicates were available, the CV of μ_{max} remained below 15%, with an overall average of 4.02%, indicating good reproducibility of the experiments.

4.3.3. Secondary Fitting and Model Performance

The CTM model was fitted to our maximum specific growth rate estimates, on the Ln(μ_{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.

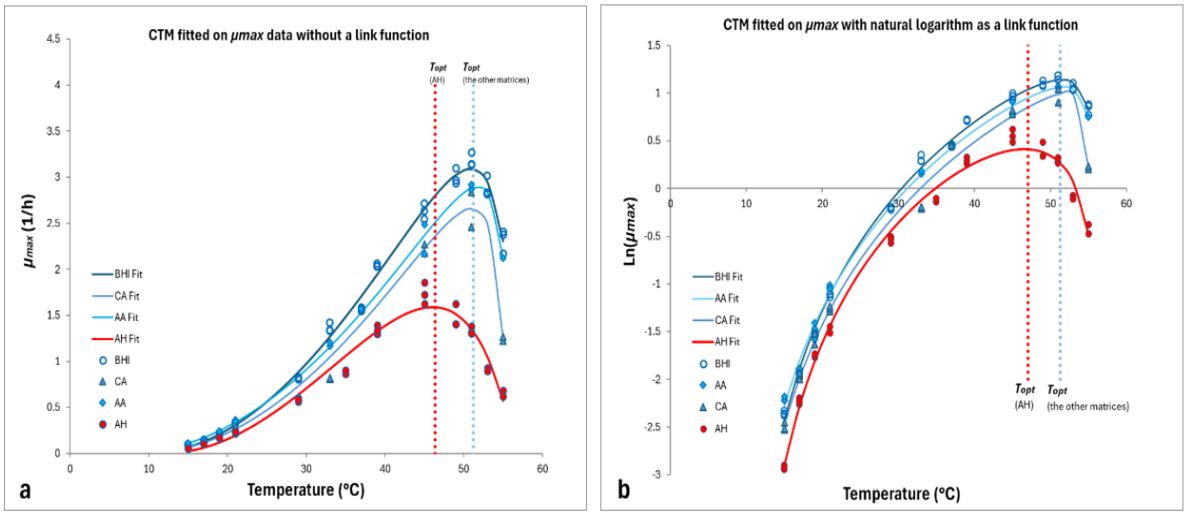

Fig. 4.13 *Specific growth rates (μ_{max}) as a function of temperature fitted by CTM. (**a**) μ_{max} estimates from the primary model fitting and their fitted CTM curves, without using a link function and (**b**) the same data but with the natural logarithm link function*

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(μ_{max}) was modelled separately for each matrix using the Cardinal Temperature model (CTM). Figure **(4.13 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 (**4-5**) lists the estimated cardinal parameters, i.e. μ_{opt}, T_{min}, T_{opt} and T_{max}, along with their respective standard errors.

The performance of the model was evaluated according to ISO requirements (FDIS-ISO 23691, 2025) for secondary model parameters, 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 4-5: Estimated parameters and their standard errors for the CTM model in BHI and the three studied Plant-based beverages, resulting from the regression while applying the Ln link function

	T_{min}	SE_{Tmin}	μ_{opt}	$SE_{\mu opt}$	T_{opt}	SE_{Topt}	T_{max}	SE_{Tmax}	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*
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 Cardinal Temperatures (T_{min}, T_{opt}, T_{max}) of *B. licheniformis* growing in the almond milk (AA) and the coconut milk (CA) were similar to those in BHI. The μ_{opt} estimates showed the trend: μ_{opt}(BHI) > μ_{opt}(AA) > μ_{opt}(CA) reflecting the expected matrix effect and growth being faster in culture medium compared to other food matrices.

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. 4.13 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 almond 'AA') in all parameters ($p < 0.05$) except T_{max}. In contrast, no significant differences were observed between the cardinal values in BHI, CA, and AA matrices ($p >> 0.05$).

Further evidence comes from food correction factor analyses (**Fig. 4.14**). It compares observed and predicted values of the natural logarithm of the maximum specific growth rate, Ln(μ_{max}), across the three food matrices (AA, CA, AH) using food correction factors with the predictions in BHI as a baseline. The equality line represents perfect agreement between predicted and observed values. While predictions in AA and CA show good alignment with the equality line, AH-predictions exhibit notable deviations, particularly at lower growth rates. This indicates that the correction factor, i.e. the ratio μ_{food}/μ_{BHI} is not necessarily temperature-independent for every food matrix, as AH is a counterexample.

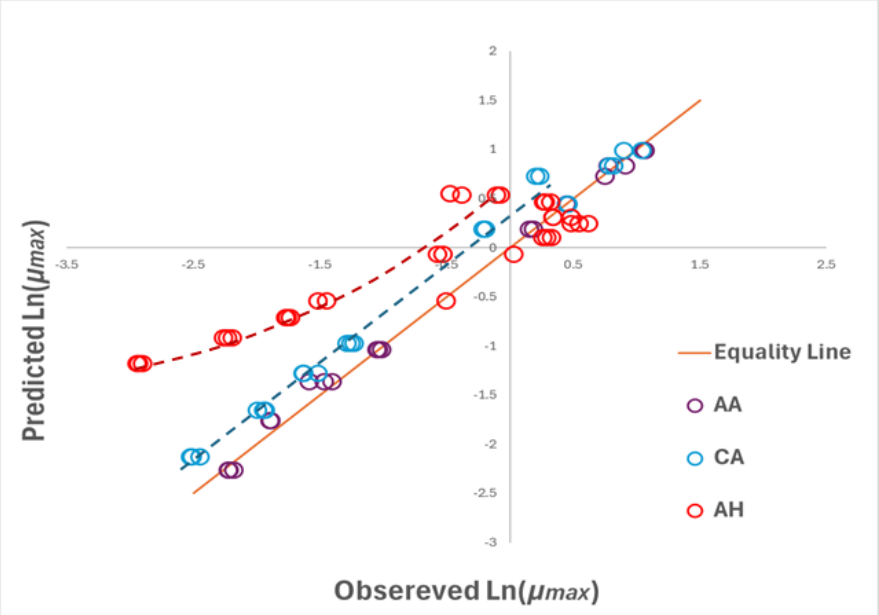

Fig. 4.14 *Observed versus C_f-predicted natural logarithm of the maximum specific growth rates in the four studied matrices, the broken lines show the trend in these matrices in the suboptimal region*

Notably, the purple AA points closely follow the equality line, indicating that growth kinetics in AA is similar to that in BHI. Meanwhile, for lower values of Ln(μ_{max}) (below 0.2 h^{-1}), the blue CA points run roughly parallel to the equality line. This shows that a correction factor can reliably adjust BHI-based predictions for CA in this lower range of Ln(μ_{max}). However,

the AH points (red) increasingly deviate from the equality line as Ln(μ_{max}) decreases. This indicates that a constant correction factor is inadequate for predictions in AH. These patterns suggest three distinct behaviors in the sub-optimal region: (1) AA behaves like BHI almost without adjustment; (2) CA aligns with BHI after applying the food correction factor C_f; and (3) AH exhibits more complex deviation, where the constant correction factor method is not valid for this matrix.

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. From **Table** (**4-4**) we observe that the fat content is twice as low in Almond milk AH (fat content = 0.57 g/100 ml) compared to Almond Milk AA (fat content = 1.14g/100 ml). In addition, the carbohydrates concentrations are quite different (2.05 g/100 ml for AH compared to <0.1 g/100 ml for AA).

Based on these observations, it can be hypothesized that the material source (e.g., almond, soy, oat) is not necessarily a good predictor of microbial behavior and kinetics parameters. An example of the compositional difference impact 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 difference found. 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.

4.3.4. Summary

By testing a broad range of temperatures directly in food matrices, this study provides high-resolution data for evaluating microbial growth variability under real-world conditions. Such datasets can enhance *in-silico* simulations and help quantify uncertainty in food safety modeling. The findings emphasize that, when collected transparently and reproducibly, microbial growth data should be considered, even when they depart from convenient but simplified assumptions.

For reliable use, particularly in ambient-stored products, we recommend applying the developed models only within the tested temperature range of 15°C to 55°C. Although the

model estimates T_{min} and T_{max}, these represent extrapolations and may not reflect true biological limits.

This work draws the attention that the effect of some matrix components could be significant, beyond broad food categorizations (e.g., "*almond milk*"). Converting them into continuous variables would mean more advanced "*tertiary modelling*".

4.4. Clarification of Common Misleading Concepts in Predictive Microbiology Discovered During Analyses

4.4.1. What Is the Difference Between Standard Deviation (STD) and Standard Error (SE)?

Standard Deviation (STD) tells, on average, how far each value lies from the mean, it represents the spread/scatter of the data.

$$\mathrm{STD} = \sqrt{\frac{(\bar{x} - x_i)^2}{n-1}} \quad \textbf{(4-5)}$$

Where $\bar{x}$ is the sample mean (average of the data), x_i ($i = 1 \ldots n$) datapoints and n is the number of datapoints in the used dataset (sample size).

Meanwhile, the Standard Error (SE) or sometimes refer to as the Standard Error of the Mean (SEM) tells, how far the sample mean (average) deviates from the actual population mean, it represents the uncertainty of the estimation. It can be calculated from the data using:

$$\mathrm{SE} = \frac{\mathrm{STD}}{\sqrt{n}} \quad \textbf{(4-6)}$$

4.4.2. Does Rescaling Affect SE_{fit} and $SE(\hat{p})$?

Rescaling of variables has straightforward consequences when the variables are rescaled by a proportionality constant (linear rescaling). This is demonstrated below by changing the units of the axes of primary models:

1. Rescale the time unit from hours to, say, days. The sampling points then become $t_i' = t_i /24$. The standard error of fit, SE_{fit}, remains unchanged since it is measured on the vertical axis. However, the estimate for the slope and its standard error are both altered by the factor 24. Therefore, the relative error *RE* of the slope estimate (the reciprocal of the t-score) does not change.
2. Rescale the log-count observations from decimal to, say, natural logarithm. This will change both the standard error of fit and the standard error of the slope by the factor $\mathrm{Ln}(10) \approx 2.3$. The relative error of the slope, again, does not change.

It can be readily seen that, if the response variable depends on the parameter in a non-linear manner, then the relative error of the slope estimate does change.

4.4.3. Can the Lag Parameter Have a Negative Value?

In this section we examine why the estimated lag parameter may take negative values and how to interpret them. We illustrate the point with two common approaches: the mechanistic Baranyi–Roberts model (BRM) and the geometric Gompertz function (GF).

BRM relies on a mechanistic assumption: during the lag phase cells must produce a key metabolic product P(t) (for example, an enzyme) that is required for growth in the new environment. The effect of this product on the maximum growth rate is written according to Michaelis–Menten kinetics, a concept well-known in enzyme kinetics (Srinivasan, 2022):

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{P(t)}{K_P + P(t)} \cdot \mu_{max} \qquad \textbf{(4-7)}$$

Since K_p is a constant that sets the scale for how much P(t) product is "*enough*", then the ratio $P(t)/K_p$ is what matter the most to express how ready are the cells to initiate the exponential growth. This naturally motivates rescaling the variable $P(t)$ via dividing it by K_P, defining a unitless quantity that simplifies the equation and the interpretation:

$$q(t) = \frac{P(t)}{K_P} \qquad \textbf{(4-8)}$$

Now the growth inhibition factor simplifies to:

$$\frac{P(t)}{K_P + P(t)} = \frac{q(t)}{1 + q(t)} \qquad \textbf{(4-9)}$$

As mentioned before, the reparameterizations $\alpha_0 = \frac{q(0)}{1+q(0)}$ and $h_0 = -\mathrm{Ln}(\alpha_0)$ both reduce the number of free parameters and supply biologically meaningful descriptors of the initial physiological state: α_0 (or h_0) quantifies how "*ready*" the cells are to grow after inoculation. BRM also assumes that $P(t)$ accumulates approximately at the same exponential rate (μ_{max}). This assumption further reduces complexity while remaining biologically credible.

A direct consequence of this mechanistic formulation is that the estimated lag time can be negative. In BRM the lag is defined as the time until the physiological state reaches the level required for rapid growth (e.g., when $q(t) = 1$ or $\alpha(t) = 0.5$). If the cells are already quite ready at inoculation (i.e., $P(0) > K_P(0)$ so $q(0) > 1$), the time when $q(t)$ would have crossed the threshold actually lies (in the past) before time zero. In other words, the cell's readiness

reached the required level earlier than the start of observation. This makes the computed lag negative.

GF also can give a negative lag, but the interpretation is different. The GF lag is a geometric feature of the fitted sigmoid (defined from the tangent at the inflexion point) and not derived from a mechanistic internal variable, this is a property of the curve fit rather than a direct physiological statement. Relatedly, GF's parameter y_0 is the lower geometric asymptote of the fitted sigmoid, It is usually close to the real initial measurement $y(0)$ but not guaranteed, a distinction that becomes important when the fitted lag is negative (**Fig. 4.15**).

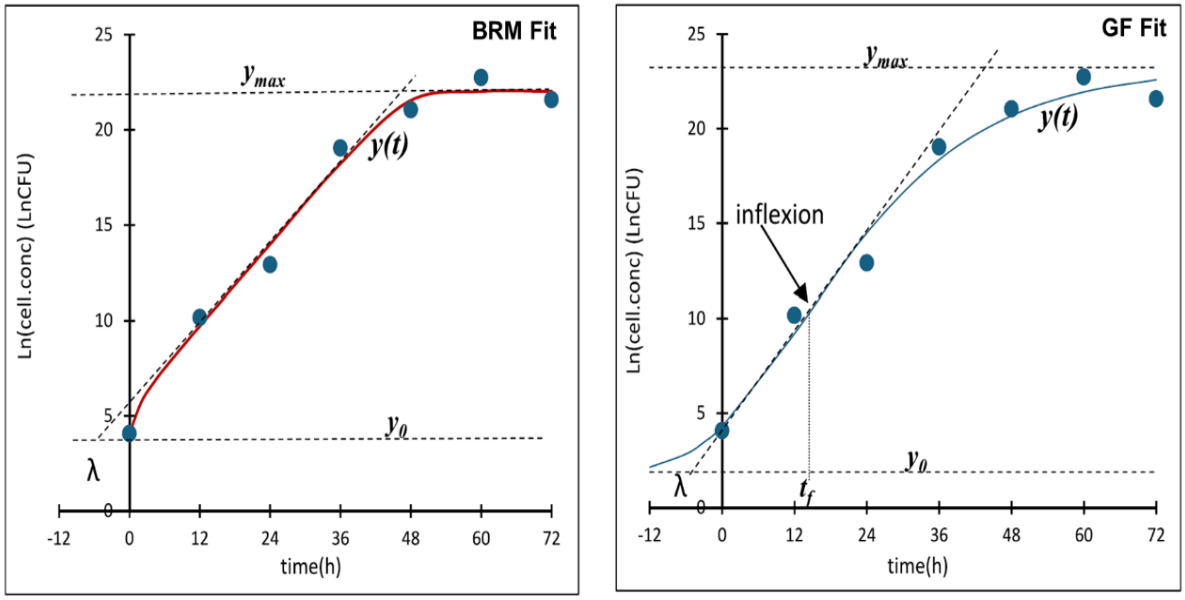

Fig. 4.15 *Case of negative lag for GF and BRM. Blue circles: GF-simulated data set. Blue and red continuous lines: fitted by GF and BRM.*

4.4.4. Does It Matter If We Fit Log10Count or LnCount When Using BRM and GF?

The answer differs between the two studied sigmoid equations. For Gompertz Function (GF) the slope parameter is geometrically linked to time (the tangent at the inflection), so a linear rescaling of the log-axis simply rescales the estimated rate and its standard error by the constant factor Ln(10), (see section **4.4.1.**) when converting between Log_{10} and Ln. In practice this makes GF indifferent to whether the fit is performed on $Log_{10}(x)$ or $Ln(x)$: the numerical rate on the natural-log scale (μ_{max}) can be recovered by multiplying the Log_{10}–based rate estimate by Ln(10).

However, BRM behaves differently because its parameter μ_{max} is a *potential* maximum specific growth rate, not simply the observed tangent slope at the inflection time. BRM's inhibition function commonly appears as $u(y) = 1 - \exp(y - y_{max})$ with $y = \mathrm{Ln}(x)$. If y is taken as $\mathrm{Log}_{10}(x)$, then the back-transformation (inverse function) of $\exp(y)$ is no longer valid: one must use $10^y = \exp[y \cdot \mathrm{Ln}(10)]$. Failing to replace $\exp(y)$ by 10^y produces a structural mismatch in the inhibition term and leads to biased parameter estimates. **Table** (**4-6**) illustrates this. The discrepancy is often small (<5%) when the exponential phase is well sampled, because the dominant information about growth rate comes from that region. However, when only a few exponential-phase points are available or when the inoculum level is not very low, the misuse of the scaling function can inflate μ_{max} by 20–50% and increase its estimated uncertainty. Since standard errors are affected too, the confidence intervals based on the incorrect formulation will also be misleading.

Practical recommendation is to fit BRM to natural-log counts $\mathrm{Ln}(x)$ or, if you have to fit $\mathrm{Log}_{10}(x)$, explicitly replace every occurrence of $\exp(y)$ by 10^y in the model structure before back-transforming parameter estimates. For GF this step is unnecessary because the exponential appears in a time argument, not in a log-count back-transform argument. Always report which log base was used and how any conversions were performed so readers can interpret the rates and their uncertainties correctly.

Table 4-6: Fitting BRM to the growth curve with ID = ELC0374 in *ComBase*. First, with $y_i = \mathrm{Ln}(x_i)$ then with the $\mathrm{Log}_{10}(x_i)$ values, in the latter case, the results were put back to the natural log-scale (see **Fig. 4.16**).

Method to fit BRM	μ_{max} (h^{-1})	$SE(\mu_{max})$	λ(h)	SE(λ)	$\hat{y}_{max}$	$SE(\hat{y}_{max})$
Eq(**2-11**) fitted to $y_i = \mathrm{Ln}(x_i)$	0.442	0.0703	21.9	4.70	22.9	4.79
Eq(**2-11**) fitted to $y_i = \mathrm{Log}_{10}(x_i)$, then the appropriate parameter estimates are multiplied by $\mathrm{Ln}(10) \approx 2.3$	0.542	0.0942	26.5	4.16	22.9	4.98

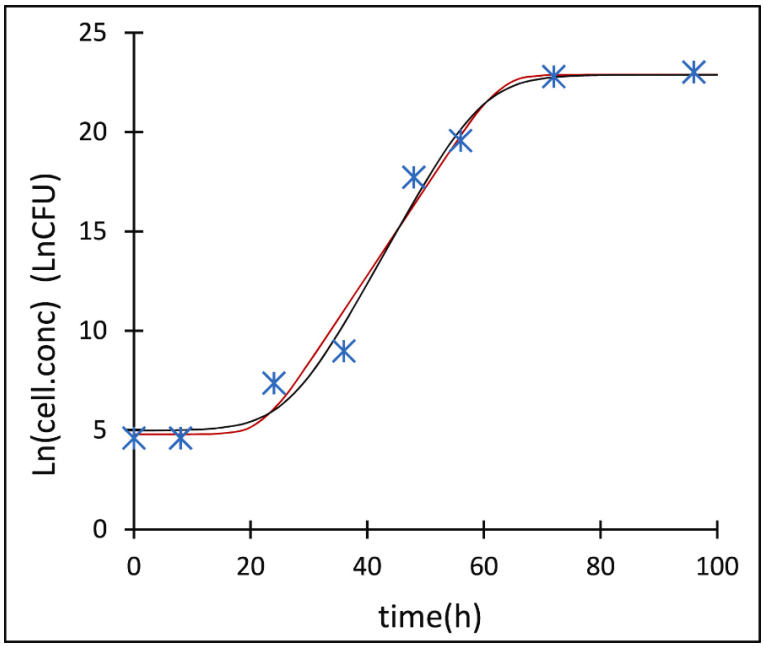

Fig. 4.16 *Fitting BRM to the growth curve with ID =ELC0374 in ComBase (Blue stars). First, with $y_i = Ln(x_i)$ (Red continuous line) then with the $Log_{10}(x_i)$ (Black continuous line), in the latter case, the results were put back to the natural log scale*

4.4.5. How Robust are BRM and GF with Poor Datasets Scenarios?

When fitting poor datasets, GF was much less stable than BRM. In 10 out of 1000 simulations, GF produced extreme overestimates of μ_{max}, with above 100% relative errors (the true value being μ_{max} = 0.5 h^{-1}). These overestimations were driven by the limited number of data points during the exponential phase. A single outlier in this critical region could distort GF fits dramatically. Figure (**4-17**) illustrates one example of such sensitivity where the GF-estimate of the maximum specific growth rate is more than double of that produced by BRM regression, and more than three times greater than the true value used to generate the simulated dataset. In contrast, BRM never produced such outliers: even when the simulated data were generated from GF, BRM produced stable, unique estimates for any reasonable initial values.

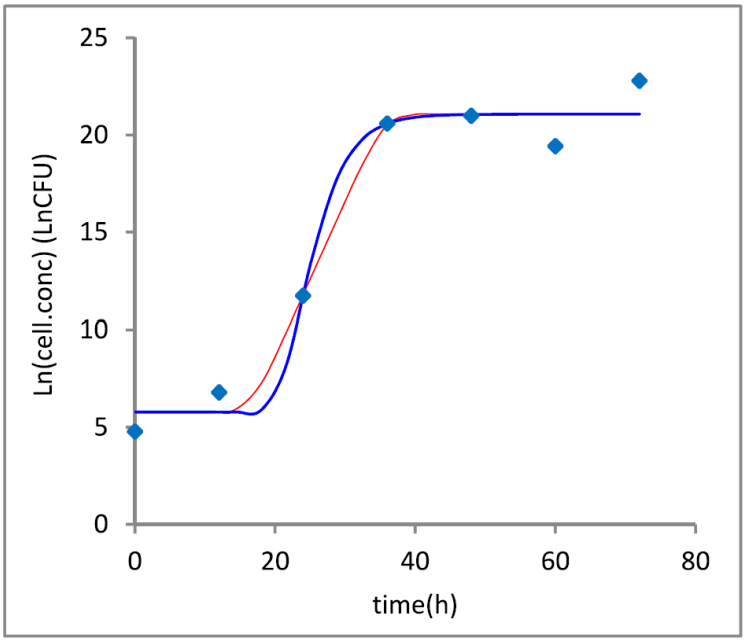

Fig. 4.17 *An example for a GF-simulated data set (blue diamonds) fitted by GF (blue thick continuous line) and BRM (red thin continuous line)*

Opposite example can be tested that, if accurately measured bacterial concentrations are all in the exponential phase, except the inoculum and one stationary phase point, still GF would way overestimate the maximum specific growth rate unlike BRM would.

4.4.6. What Does *Dry* and *Wet* Correlation Mean and What Is the Difference Between Them?

In brief, *dry* correlation is a statistical artefact of the estimation process; *wet* correlation is the biological (or causal) association between parameters across experimental conditions.

More precisely, every regression produces a covariance matrix for the parameter estimates; the variance of an estimate is the square of its standard error and the covariance between two estimates gives their *dry* correlation via:

$$\rho_{1,2} = \frac{\mathrm{Cov}(\widehat{p_1},\widehat{p_2})}{SE_{\widehat{p_1}} \cdot SE_{\widehat{p_2}}} \qquad \textbf{(4-10)}$$

This quantity depends only on the model, the data sampling method and the noise model (shape of the residuals distribution), not on biology *per se*, and so it is called *dry* (statistical) correlation. It can be large when (1) the model ties those parameters together in a nonlinear

way, (2) the collected data give little independent information about each parameter, or (3) when the numerical optimization find many parameter combinations that give nearly identical sums-of-squares (flat "*valleys*" in the multivariate objective function to be minimized).

Practical recommendations that reduce *dry* correlation include (i) centering the explanatory and the response variables (shifting the data so the mean is zero, or in another words, the center of gravity of their measured values coincides with the origin of the coordinate system), (ii) a proper reparameterization, if found.

Wet correlation, by contrast, refers to the true relationship between the underlying biological parameters across conditions (for example, how the lag λ and maximum growth rate μ_{max} covary when temperature changes). This correlation is driven by physiology and causal effects (e.g., as temperature rises within the sub-optimal range, biology predicts $\lambda\downarrow$ while $\mu_{max}\uparrow$, producing a negative *wet* correlation). *Wet* correlation must be estimated by comparing independently obtained parameter values across controlled conditions (different temperatures, media, etc.), and it is not eliminable by algebraic reparameterization.

A useful example: reparametrizing BRM and GF by $h_0 = \mu_{max} \cdot \lambda$ revealed a strong positive *dry* correlation between μ_{max} and h_0 (**Fig. 4.4**) and between μ_{max} and λ as well, even though the *wet* correlation between the μ_{max} and h_0 across temperatures is essentially zero in the normal physiological range and the *wet* correlation between the μ_{max} and λ is strongly negative as temperature increases (but remains sub-optimal), the lag phase would shorten, and the maximum specific growth rate μ_{max} would increase. Thus, one observes (1) a positive *dry* correlation among the estimates produced by joint fitting and (2) a zero (or negative) *wet* correlation reflecting biology.

Another example of purely *wet* correlation between parameters is the one reported by Baranyi et al. (2017) between two parameters of Ratkowsky model (T_{min}) and (b), for different strains of *Bacillus cereus*.

It is important to emphasize here that a frequent source of confusion in applied modeling is the distinction between model parameters and their estimates. Model parameters (e.g., μ_{max}, λ) are, in a strict statistical sense, fixed but unknown constants that characterize the

underlying biological system. As such, they do not possess variability and *dry* correlation. In contrast, parameter estimates are obtained from experimental data and therefore depend on the specific dataset, experimental design, and estimation procedure. Consequently, they behave as random variables but with a fixed probability distribution (which is, however, difficult to determine exactly, except in ideal cases like that of linear regression) and may exhibit both variance and correlation. Thus, whenever correlations are discussed in the context of parameter estimation or model fitting, they should be understood as correlations between parameter estimates rather than between the parameters themselves.

4.4.7. How to Define the '*Happy-growth Region*'?

We have seen that, if the dilution-based sampling method is the main source of the error, then the relative error for the bacterial concentration measurement is constant, i.e. independent of the magnitude of μ_{max}. Partly because of this, partly to obtain a linear model in the exponential phase, the natural logarithm of the cell concentration should be fitted by the model of Baranyi and Roberts (1994).

This has an important consequence for secondary models. If the environment does not affect the error of the (y_i, t_i) measurements, then the logarithm link function should be used for the secondary model, too (Akkermans et al, 2018; Rockaya and Baranyi, 2025). The constant relative error for μmax can be conceived this way: if the conditions change, all physiological processes happen in the same way, just proportionally faster or slower, as if we had changed the time unit. We name the region of the environment, where this is true, the "*happy-growth region*"; i.e. where the environment does not cause extra error of wet origin, only proportional change in the cellular kinetics, as if using a different unit for time. As the various biochemical rates change, also the lag phase is proportionally shorter or longer as well as the transition to the stationary phase. Consequently, the product of the lag and the growth rate (which is analogous to the $h_0 = \lambda \cdot \mu_{max}$ "*relative lag*"; see McMeekin et al. 2002) is constant. In other words, we can define the "*happy-growth region*" also by this: it is the region of the growth environment where, assuming the same pre-inoculation history, the h_0 product of the lag and the maximum specific growth rate is constant. Since h_0 has also been used to quantify the "*work-to-be-done*" during the lag; (Bovill et al. 2001), another definition for the happy growth region is this: where, assuming the same history, the work to be done is the same to

adjust to the new environment (though this work could be carried out slower or faster). This region coincides with that of a constant relative change in maximum specific growth rate.

4.4.8. Can We Use Parallel CI When Fitting the Square-root RKM?

In principle, parallel (constant-width) confidence bands for the square-root Ratkowsky model (sq-RKM) can only be justified if the residual variance of $\sqrt{\mu_{max}}$ is approximately constant across the fitted temperature (or any other environmental factor) range. Empirically and biologically, this is seldom true: the square-root link linearizes the relationship between $\sqrt{\mu_{max}}$ and temperature in the sub-optimal region but does not stabilize variance (it rather increases with the rate **Fig. 4.18**) and variance typically increases even more as bacteria approaches the growth boundary because the *wet* source of the error becomes increasingly significant (for example, it is more difficult to maintain a constant environment for longer times, or the cells' physiological variability increases). Therefore, assuming homoscedastic errors and reporting parallel CIs will understate uncertainty near the boundaries and overstate it in the central region, producing misleading inference about temperature effects and parameter precision. A more defensible workflow is (i) check homoscedasticity manually (e.g. residual plots), (ii) fit the linear $\sqrt{\mu_{max}}$ vs. T relationship only in the well-behaved sub-optimal window to obtain robust starting values (initial estimates), and then (iii) fit the full model by weighted regression or by using the previous obtained initial estimates to fit (Ln-RKM) i.e. Ln(μ_{max}) vs. T.

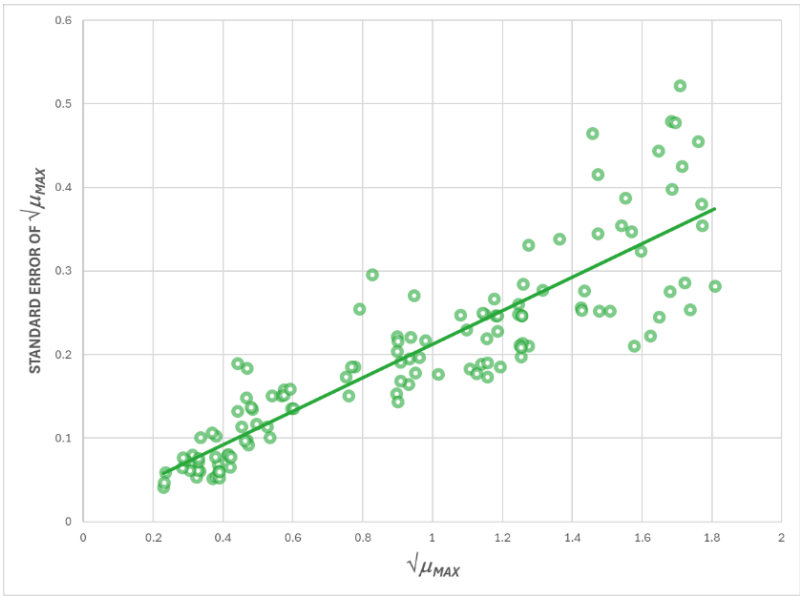

Fig. 4.18 *Demonstration of the positive correlation between* $\sqrt{\mu_{max}}$ *and its standard error*

4.4.9. What are the Shapes of the Most Used Confidence Intervals?

We should have a direct and clear understanding of confidence intervals, their application and related terminology, to avoid confusion. The most used ones in predictive microbiology papers are the confidence intervals of predicting a new single observation and the confidence intervals of predicting the mean of a mean of a new dataset.

Figure (**4.19**) demonstrates the linear model (blue, thick continuous line). The confidence band around it, with a radius estimable by the standard error of fit (paralell pink shaded area) demonstrates that *ca* 2/3 of the possible further *y* measurements will be expected to fall within the band of the parallels. This is called the prediction band for the individual-observation. However, the curved confidence band (purple, shaded area) indicates that the same proportion of the means generated from those further points will fall within this curved band which is called the prediction band of the means. With increasing number of data, the confidence band for the measurements will not change significantly, only the radius of the band will be more certain; however, the curved band will be thinner and thinner (roughly proportionally with the square-root of the number of measurements). The comparison is analogous with the mentioned difference between the standard deviation of the data and the

standard error of the average of the data, as an estimation of the population mean (see **Section 4.4.1**).

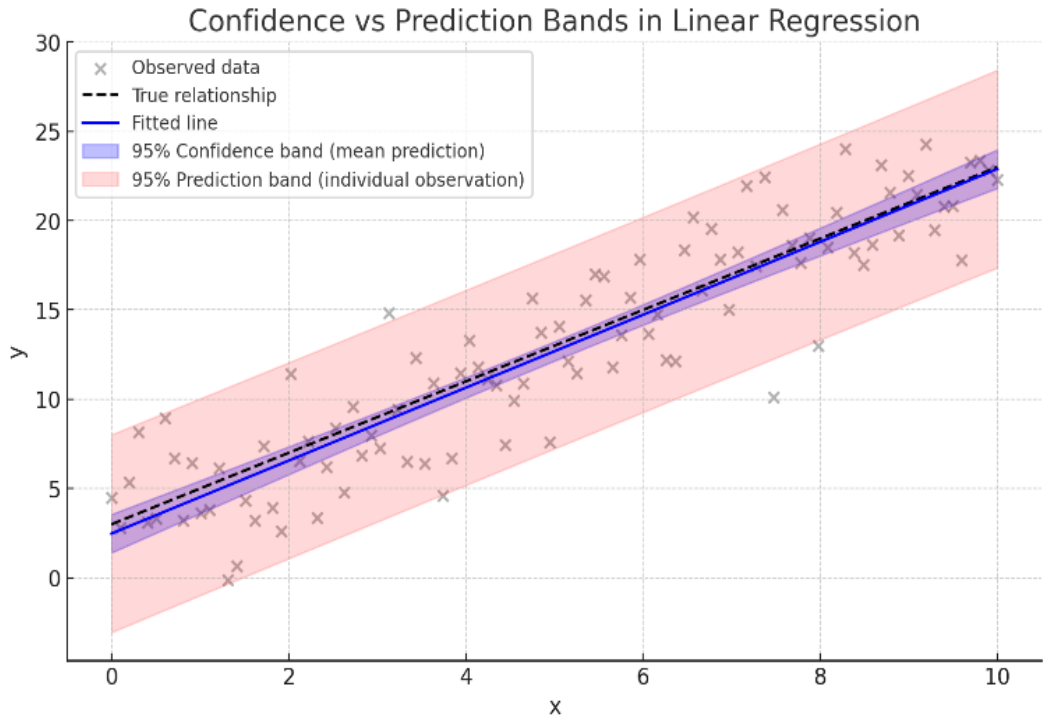

Fig. 4.19 *The difference between confidence band and prediction band*

Table 4-7: The difference between confidence band and prediction band

Feature	Confidence Band (Mean Response)	Prediction Band (Individual Observation)
Definition	Interval estimating where the *mean* of future observations will fall.	Interval estimating where a *single new observation* will fall.
Purpose	Quantifies uncertainty in estimating the *average response*.	Quantifies uncertainty in predicting an *individual future value*.
Width	Narrower	Wider
Includes	Only uncertainty of the estimated mean (model structure + parameter error).	Both uncertainty of the mean (parameter error) and the natural variability of data (random error) BUT not the model structure!
Dependence on sample size (n)	Width decreases with √n (more data → narrower band).	Largely independent of n (band remains about the same width).
Interpretation	"*If we repeated the experiment many times, the average of the responses would lie within this band 95% of the time.*"	"*If we repeated the experiment many times, 95% of individual responses would fall within this band.*"
Common use	To show model precision and trend confidence.	To predict future measurements and variability.

5. 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 that of *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 beverages 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 aw 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.

6. 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 with 415 records so far. 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 dairy alternatives.

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 beverages 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 beverages 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 the other 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°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.

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.

The study introduces a conceptual distinction between two sources of error: *wet* (biological/experimental) and *dry* (statistical/model-based). This 2-category framework parallels the traditional classification of uncertainty and variability but offers a clearer, more practical way to interpret error in predictive microbiology. By separating errors arising from experimental conditions and biological systems from those introduced by modeling and data analysis.

7. PRACTICAL 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 beverages 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.

8. SUMMARY IN HUNGARIAN

Ez a disszertáció *in silico* (szimuláción alapuló) elemzéseket és *in vitro* (laboratóriumi) kísérleteket ötvözött, tükrözve ezzel két egymást kiegészítő megközelítést.

A dolgozat első része a *Pbase* adatbázis bemutatására összpontosított. Ez egy olyan strukturált ontológián alapul, amelyet a mikrobiológiai és összetételi adatok standardizálására fejlesztettek ki a növényi alapú alternatívák területén. Bár ez a fejlesztés még korai szakaszban van, a munka rávilágít arra, hogy a kísérleti adatok strukturálása és digitalizálása elengedhetetlen előfeltétele annak, hogy a prediktív és kockázatalapú megközelítések (például a Big Data és a mesterséges intelligencia alkalmazásai) előreléphessenek az élelmiszertudományokban.

A második rész a paraméterbecslések statisztikai megbízhatóságát vizsgálta prediktív modellekben. Az elsődleges modellek esetében a hisztogram- és korrelációelemzések azt mutatták, hogy a Baranyi–Roberts-regresszió (BR-Reg) statisztikus tulajdonságai jobban egyeznek a lineáris regresszió tulajdonságaival, mint a Gompertz-regresszió (G-Reg) becslései. Ez annak köszönhető, hogy a BRM a lineáris alapmodellt finomítja. A másodlagos modellek esetében kulcsfontosságúnak bizonyult az átparaméterezés: az átalakítás után mind a Cardinal Temperature Model (CTM), mind a Ratkowsky Model (RKM) megbízható konfidenciaintervallumokat és szimmetrikus statiszikai eloszlást eredményezett. Az RKM robusztusabbnak bizonyult, míg a CTM csak akkor teljesített hasonlóan jól, ha megfelelő kezdőértékekkel támasztották alá. Ezek az eredmények hangsúlyozzák, hogy a prediktív modellezés megbízhatóságát nem csak a modell matematikai kifejezése, hanem annak paraméterezése is jelentősen befolyásolja.

A harmadik rész a *B. licheniformis* növekedési kinetikáját vizsgálta növényi alapú tejekben. Széles hőmérsékleti tartományban végzett kísérletek révén nagy felbontású adatokat gyűjtöttünk élelmiszermátrixokban, amelyek megkérdőjelezték a prediktív mikrobiológiában gyakran alkalmazott egyszerűsítő feltételezéseket. A vizsgálat különösen azt mutatta meg, hogy a kardinális hőmérsékleti paraméterek (minimum, optimum és maximum) nem mindig függetlenek az élelmiszer-mátrixtól. Ez a kivétel kiemeli a pontos adatgyűjtés és a mikrobiális növekedési adatok világos közlésének fontosságát. Emellett a kutatás rávilágított arra, hogy az élelmiszer-összetevők a tág termékkategóriákon (pl.

„*mandulatej*”) túl jelentősen befolyásolhatják a mikroorganizmusok viselkedését. Ez alátámasztja annak lehetőségét, hogy a mátrix összetételét a jövőbeli harmadlagos modellekben, folytonos változóként kezeljük. A gyakorlati alkalmazás szempontjából a fejlesztett modellek 15–55 °C hőmérsékleti tartományban érvényesek. Az extrapolált T_{min} és T_{max} értékek nem feltétlenül tükrözik a biológiailag értelmezett értékeket.

Az adatbázis-ontológia fejlesztésének, a modellek statisztikai értékelésének és az alkalmazott mikrobiológiai vizsgálatoknak a kombinálásával ez a disszertáció hozzájárul a prediktív mikrobiológia módszertani szigorának növeléséhez, valamint annak gyakorlati alkalmazásához, például a növény-alapú alternatívák területén.

9. SUMMARY IN ENGLISH

The dissertation combined *in-silico* (simulation-based) analyses and *in-vitro* (laboratory) experiments, leveraging these two complementary perspectives to utilize predictive microbiology methods in practical solutions.

The first part of the thesis introduced the *Pbase* database, constructed on a well-defined ontology that had been developed to standardize microbiological and food-composition data for plant-based alternatives. Although still at a preliminary stage, this contribution highlights the value of structuring and digitizing experimental data as a prerequisite for advancing computational and predictive approaches, such as Big Data and AI applications, in food sciences. The second part evaluated the statistical reliability of parameter estimates when regressing predictive models. For primary models, histogram and correlation analyses revealed that the Baranyi-Roberts regression (BR-Reg) produced estimates more consistent with those generated by linear regression than the Gompertz regression (G-Reg) did, owing this to the construction of BRM which was in fact a refinement of a linear model. For secondary models, reparameterization proved to be a critical method: once transformed, both the Cardinal Temperature Model (CTM) and the Ratkowsky Model (RKM) produced reliable confidence intervals and symmetric error distributions. RKM displayed greater robustness, while CTM performed similarly only when supported by suitable initial estimates. These results emphasize that reliability in predictive modeling is not dictated by model choice only, but also the arrangement in which the model is presented, line its parameterization.

The third part focused on *B. licheniformis* growth kinetics in plant-based beverages. By testing a wide range of temperatures in real food matrices, this study provided high-resolution data that questioned one of the simplifying assumptions often made in predictive microbiology. In particular, the findings demonstrated that cardinal temperatures (minimum, optimum, maximum) are not always independent of the food matrix. This exception underscores the importance of collecting high accuracy data and reporting microbial growth data transparently, even when they depart from convenient generalizations. Moreover, the study showed that food components beyond broad product categories (e.g., "*almond milk*") can significantly influence microbial behavior, highlighting the potential of treating matrix composition as a continuous variable in future tertiary models. For practical use, models

developed here should be applied within the experimental range of 15–55°C, as extrapolated T_{min} and T_{max} values may not reflect biological reality. By combining database ontology development, statistical evaluation of models, and applied microbiological studies, it contributes both to the methodological rigor of predictive microbiology and to its practical application in emerging food sectors such as plant-based alternatives.

10. BIBLIOGRAPHY

1. Akkermans, S., Logist, F., & Van Impe, J. F. (2018). Parameter estimations in predictive microbiology: Statistically sound modelling of the microbial growth rate. *Food Research International*, *106*, 1105–1113.
2. 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.
3. Augustin, J. C., Brouillaud-Delattre, A., Rosso, L., & Carlier, V. (2000). Significance of inoculum size in the lag time of *Listeria monocytogenes*. *Applied and Environmental Microbiology*, *66*(4), 1706–1710.
4. 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.
5. Baranyi, J. (1992). Notes on reparameterization of bacterial growth curves. *Food Microbiology*, *9*(2), 169–171.
6. Baranyi, J. (2010). *Modelling and parameter estimation of bacterial growth with distributed lag time* [Doctoral dissertation, University of Szeged (Hungary)].
7. Baranyi, J., Buss da Silva, N., & Ellouze, M. (2017). Rethinking tertiary models: relationships between growth parameters of *Bacillus cereus* strains. *Frontiers in Microbiology*, *8*, 1890.
8. Baranyi, J., Rockaya, M., & Ellouze, M. (2024). From data to models and predictions in food microbiology. *Current Opinion in Food Science*, *57*, 101177.
9. Baranyi, J., & Roberts, T. A. (1994). A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, *23*(3–4), 277–294.
10. Baranyi, J., Roberts, T. A., & Mcclure, P. (1993). Some properties of a nonautonomous deterministic growth model describing the adjustment of the bacterial population to a new environment. *Mathematical Medicine and Biology: A Journal of the IMA*, *10*(4), 293–299.
11. Baranyi, J., & Tamplin, M. L. (2004). ComBase: A common database on microbial responses to food environments. *Journal of Food Protection*, *67*(9), 1967–1971.
12. Beacom, E., Repar, L., & Bogue, J. (2022). Consumer motivations and desired product attributes for 2.0 plant-based products: a conceptual model of consumer insight for

market-oriented product development and marketing. *SN Business & Economics*, *2*(8), 115.

13. Bovill, R. A., Bew, J., & Baranyi, J. (2001). Measurements and predictions of growth for *Listeria monocytogenes* and *Salmonella* during fluctuating temperature: II. Rapidly changing temperatures. *International Journal of Food Microbiology*, *67*(1-2), 131–137.

14. Business Research Company. (2025). *Plant milk global market report 2025*. https://www.thebusinessresearchcompany.com/report/plant-milk-global-market-reports

15. Business Research Insights. (2025). *Plant-Based Milk Market Trends, Size, Share & Forecast 2035*. https://www.businessresearchinsights.com/market-reports/plant-hebased-milk-market-109354

16. Buss da Silva, N., Baranyi, J., & Ellouze, M. (2017). Relationships between the parameters of temperature-dependent growth models for *Bacillus cereus* strains. *Frontiers in Microbiology*, *8*, 1890.

17. Cardilin, T., Almquist, J., Jirstrand, M., Sostelly, A., Amendt, C., El Bawab, S., & Gabrielsson, J. (2017). Tumor static concentration curves in combination therapy. *The AAPS Journal*, *19*(2), 456–467.

18. Champidou, C., Ellouze, M., Campagnoli, M., Robin, O., Haddad, N., & Membré, J. M. (2024). Unveiling the matrix effect on *Bacillus licheniformis* and *Bacillus subtilis* spores heat inactivation between plant-based milk alternatives, bovine milk and culture medium. *International Journal of Food Microbiology*, *422*, 110807.

19. Corry, J. E., Jarvis, B., Passmore, S., & Hedges, A. (2007). A critical review of measurement uncertainty in the enumeration of food micro-organisms. *Food Microbiology*, *24*(3), 230–253.

20. Dai, H., Ma, L., Xu, Z., Soteyome, T., Yuan, L., Yang, Z., & Jiao, X. A. (2024). Invited review: Role of *Bacillus licheniformis* in the dairy industry—Friends or foes? *Journal of Dairy Science*, *107*(10), 7520–7532.

21. Dalgaard P., Buch P., & Silberg S. (2002). Seafood Spoilage Predictor—development and distribution of a product specific application software. *International Journal of Food Microbiology*, *73*(2–3), 343–349.

22. Diarra, C., Goetz, C., Gagnon, M., Roy, D., & Jean, J. (2023). Biofilm formation by heat-resistant dairy bacteria: multispecies biofilm model under static and dynamic conditions. *Applied and Environmental Microbiology*, *89*(10), e00713-23.

23. Erjavec, N., Cvijovic, M., Klipp, E., & Nyström, T. (2008). Selective benefits of damage partitioning in unicellular systems and its effects on aging. *Proceedings of the National Academy of Sciences*, *105*(48), 18764–18769.

24. Esty, J. R., & Meyer, K. F. (1922). The heat resistance of the spores of *B. botulinus* and allied anaerobes. XI. *The Journal of Infectious Diseases*, 650–664.

25. 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.

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

27. 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.

28. Gleissle, A., Schmidt, H., & Hinrichs, J. (2025). Prevalence of spore-forming bacteria in plant-based raw materials used for plant-based milk alternatives. *International Journal of Food Microbiology*, *111255*.

29. Goldblith, S. A., Joslyn, M. A., & Nickerson, J. T. (1961). *Introduction to thermal processing of foods* (Vol. 1). AVI Publishing Company.

30. Gompertz, B. (1825). XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philosophical Transactions of the Royal Society of London*, *(115)*, 513–583.

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

32. González-Tejedor, G. A., Martínez-Hernández, G. B., Garre, A., Egea, J. A., Fernández, P. S., & Artes-Hernandez, F. (2017). Quality changes and shelf-life prediction of a fresh

fruit and vegetable purple smoothie. *Food and Bioprocess Technology*, *10*(10), 1892–1904.

33. Gracias, K. S., & McKillip, J. L. (2004). A review of conventional detection and enumeration methods for pathogenic bacteria in food. *Canadian Journal of Microbiology*, *50*(11), 883–890.

34. Hamill, P. G., Stevenson, A., McMullan, P. E., Williams, J. P., Lewis, A. D., S, S., ... & Hallsworth, J. E. (2020). Microbial lag phase can be indicative of, or independent from, cellular stress. *Scientific Reports*, *10*(1), 5948.

35. Heinen, W. (1970). Growth conditions and temperature-dependent substrate specificity of two extremely thermophilic bacteria. *Archiv für Mikrobiologie*, *76*(1), 2–17.

36. Huang, L. (2014). IPMP 2013—a comprehensive data analysis tool for predictive microbiology. *International Journal of Food Microbiology*, *171*, 100–107.

37. ISO International Organization for Standardization. (2025). *Microbiology of the food chain — Determination and use of cardinal values*. ISO/DIS 23691.

38. Karoui, R., & Bouaicha, I. (2024). A review on nutritional quality of animal and plant-based milk alternatives: a focus on protein. *Frontiers in Nutrition*, *11*, 1378556.

39. Koseki S. (2009). Microbial Responses Viewer (MRV): A new ComBase-derived database of microbial responses to food environments. *International Journal of Food Microbiology*, *134*(1–2), 75–82.

40. Kotte, O., Zaugg, J. B., & Heinemann, M. (2010). Bacterial adaptation through distributed sensing of metabolic fluxes. *Molecular Systems Biology*, *6*(1), 355.

41. Kyrylenko, A., Eijlander, R. T., Alliney, G., Lucas-van de Bos, E., & Wells-Bennik, M. H. (2023). Levels and types of microbial contaminants in different plant-based ingredients used in dairy alternatives. *International Journal of Food Microbiology*, *407*, 110392.

42. 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.

43. Lee, P. Y., Leong, S. Y., & Oey, I. (2024). The role of protein blends in plant-based milk alternative: a review through the consumer lens. *Trends in Food Science & Technology*, *143*, 104268.

44. Madigan, M. T. (2000). Bacterial habitats in extreme environments. In *Journey to Diverse Microbial Worlds: Adaptation to Exotic Environments* (pp. 61–72). Springer Netherlands.

45. Maier, R. M., & Pepper, I. L. (2015). Bacterial growth. In *Environmental microbiology* (pp. 37–56). Academic Press.

46. Malthus, T. R. (1798). *An essay on the principle of population, as it affects the future improvement of society, with remarks on the speculations of Mr. Godwin, M. Condorcet, and other writers*. Johnson.

47. Maojin, T., Zheng, Z., Ying, H., Yanyan, H., & Liang, Z. (2025). Bacterial Spore Inactivation Technology in Solid Foods: A Review. *Journal of Food Protection*, *100479*.

48. McClements, D. J. (2020). Development of next-generation nutritionally fortified plant-based milk substitutes: Structural design principles. *Foods*, *9*(4), 421.

49. McClements, D. J., & Grossmann, L. (2022). Next-generation plant-based foods. *Next-Generation Plant-Based Foods*, *57*(177), 198–207.

50. McClements, D. J., Newman, E., & McClements, I. F. (2019). Plant-based milks: A review of the science underpinning their design, fabrication, and performance. *Comprehensive Reviews in Food Science and Food Safety*, *18*(6), 2047–2067.

51. McMeekin, T. A., Olley, J., Ratkowsky, D. A., & Ross, T. (2002). Predictive microbiology: towards the interface and beyond. *International Journal of Food Microbiology*, *73*(2-3), 395–407.

52. 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.

53. Miller, S. J. (2006). The method of least squares. Mathematics Department Brown University, 8(1), 5-11.

54. Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, *3*, 371–394.

55. Mooney, C. Z. (1997). *Monte carlo simulation* (No. 116). Sage.

56. Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology*, *57*(1-2), 9–18.

57. Nau, R. (2016). *Statistical forecasting: notes on regression and time series analysis. Stepwise and All Possible Regressions*.

58. 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.

59. 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.

60. 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.

61. 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.

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

63. Press, W. H. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press.

64. Psomas A.N., Nychas G-N., Haroutounian S.A., & Skandamis P.N. (2011). Development and validation of a tertiary simulation model for predicting the growth of the food microorganisms under dynamic and static temperature conditions. *Computers and Electronics in Agriculture*, *76*(1), 119–129.

65. 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.

66. Ratkowsky, D. A., Lowry, R. K., McMeekin, T. A., Stokes, A. N., & Chandler, R. (1983). Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *Journal of Bacteriology*, *154*(3), 1222–1226.

67. Ratkowsky, D. A., Olley, J., McMeekin, T. A., & Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. *Journal of Bacteriology*, *149*(1), 1–5.

68. Reyes-Jurado, F., Soto-Reyes, N., Dávila-Rodríguez, M., Lorenzo-Leal, A. C., Jiménez-Munguía, M. T., Mani-López, E., & López-Malo, A. (2023). Plant-based milk alternatives: Types, processes, benefits, and characteristics. *Food Reviews International*, *39*(4), 2320–2351.

69. Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, *10*(2), 290–301.

70. Rockaya, M., & Baranyi, J. (2025). Variability of growth parameter estimates-The role of rescaling and reparametrization. *Food Microbiology*, *128*, 104726.

71. Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D., ... & Hinton, J. C. (2012). Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of Bacteriology*, *194*(3), 686–701.

72. Romulo, A. (2022). Food processing technologies aspects on plant-based milk manufacturing. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1059, No. 1, p. 012064). IOP Publishing.

73. Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. *Journal of Applied Bacteriology*, *81*(6), 501–508.

74. Ross, T., & McMeekin, T. A. (1994). Predictive microbiology. *International Journal of Food Microbiology*, *23*(3-4), 241–264.

75. Rosso, L., Lobry, J. R., & Flandrois, J. P. (1993). An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. *Journal of Theoretical Biology*, *162*(4), 447–463.

76. 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.

77. Schaffner, D. W., & Labuza, T. P. (1997). Predictive microbiology: where are we, and where are we going?

78. Scheldeman, P., Pil, A., Herman, L., De Vos, P., & Heyndrickx, M. (2005). Incidence and diversity of potentially highly heat-resistant spores isolated at dairy farms. *Applied and Environmental Microbiology*, *71*(3), 1480–1494.

79. Shulgin, B., Stone, L., & Agur, Z. (1998). Pulse vaccination strategy in the SIR epidemic model. *Bulletin of Mathematical Biology*, *60*(6), 1123–1148.

80. Srinivasan, B. (2022). A guide to the Michaelis–Menten equation: steady state and beyond. *The FEBS Journal*, *289*(20), 6086–6098.

81. Su, W., Zhang, Y. Y., Li, S., & Sheng, J. (2023). Consumers' preferences and attitudes towards plant-based milk. *Foods*, *13*(1), 2.

82. Tangyu, M., Muller, J., Bolten, C. J., & Wittmann, C. (2019). Fermentation of plant-based milk alternatives for improved flavor and nutritional value. *Applied Microbiology and Biotechnology*, *103*(23), 9263–9275.

83. Tijskens, L. M. M., Hertog, M. L. A. T. M., & Nicolaï, B. M. (Eds.). (2001). *Food process modelling* (Vol. 59). Woodhead Publishing.

84. Topp, B., Promislow, K., Devries, G., Miura, R. M., & T Finegood, D. I. A. N. E. (2000). A model of *β*-cell mass, insulin, and glucose kinetics: pathways to diabetes. *Journal of Theoretical Biology*, *206*(4), 605–619.

85. Van Boekel, M. A. (2008). *Kinetic modeling of reactions in foods*. CRC press.

86. Vegan Society. (2019, September). *Plant milk market*. https://www.vegansociety.com/news/market-insights/plant-milk-market

87. Verhulst, P. F. (1845). Recherches mathématiques sur la loi d'accroissement de la population [Mathematical Researches into the Law of Population Growth Increase]. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, *18*, 1–42.

88. Walls, I., & Scott, V. N. (1996). Validation of predictive mathematical models describing the growth of *Escherichia coli* O157: H7 in raw ground beef. *Journal of Food Protection*, *59*(12), 1331–1335.

89. Walls, I., & Scott, V. N. (1997). Validation of predictive mathematical models describing the growth of *Listeria monocytogenes*. *Journal of Food Protection*, *60*(9), 1142–1145.

90. Walls, I., Scott, V. N., & Bernard, D. T. (1996). Validation of predictive mathematical models describing growth of *Staphylococcus aureus*. *Journal of Food Protection*, *59*(1), 11–15.

91. Wang, N., Gao, J., Yuan, L., Jin, Y., & He, G. (2021). Metabolomics profiling during biofilm development of *Bacillus licheniformis* isolated from milk powder. *International Journal of Food Microbiology*, *337*, 108939.

92. Welch, B. L. (1947). The generalization of Student's' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35.

93. Whiting, R. C. (1993). A classification of models in predictive microbiology-a reply to KR Davey. *Food Microbiol.*, *10*, 175–177.

94. Whiting, R. C. (1995). Microbial modeling in foods. *Critical Reviews in Food Science & Nutrition*, *35*(6), 467–494.

95. Wijtzes, T., Van't Riet, K., in't Veld, J. H., & Zwietering, M. H. (1998). A decision support system for the prediction of microbial food safety and food quality. *International Journal of Food Microbiology*, *42*(1-2), 79–90.

96. Yang, S., Wang, Y., Liu, Y., Jia, K., Zhang, Z., & Dong, Q. (2023). Cereulide and emetic *Bacillus cereus*: Characterizations, impacts and public precautions. *Foods*, *12*(4), 833.

97. 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.

98. Yu, W., Chen, Z., Ye, H., Liu, P., Li, Z., Wang, Y., ... & He, N. (2017). Effect of glucose on poly-*y*-glutamic acid metabolism in *Bacillus licheniformis*. *Microbial Cell Factories*, *16*, 1–10.

99. Zwietering, M. H., De Koos, J. T., Hasenack, B. E., De Witt, J. C., & Van't Riet, K. (1991). Modeling of bacterial growth as a function of temperature. *Applied and Environmental Microbiology*, *57*(4), 1094–1101.

100. Zwietering, M. H., De Wit, J. C., & Notermans, S. (1996). Application of predictive microbiology to estimate the number of *Bacillus cereus* in pasteurised milk at the time of consumption. *International Journal of Food Microbiology*, *30*, 55–70.

101. Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & Van't Riet, K. J. A. E. M. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, *56*(6), 1875–1881.

102. Bigelow, W. (1921). The logarithmic nature of thermal death time curves. *The Journal of Infectious Diseases*, 528-536.

103. World Health Organization. (2008). *Exposure assessment of microbiological hazards in food: guidelines* (Vol. 7). World Health Organization.

104. Poschet, F., Geeraerd, A. H., Scheerlinck, N., Nicola1, B. M., & Van Impe, J. F. (2003). Monte Carlo analysis as a tool to incorporate variation on experimental data in predictive microbiology. *Food microbiology*, *20*(3), 285-295.

11. LIST OF PUBLICATIONS IN THE FIELD OF RESEARCH

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/197/2026.PL
Subject: PhD Publication List

Candidate: Maha Rockaya
Doctoral School: Doctoral School of Food Sciences
MTMT ID: 10099165

List of publications related to the dissertation

Foreign language scientific articles in international journals (3)

1. **Rockaya, M.**, Pecsenye, B., Ellouze, M., Máthé, E., Baranyi, J.: Cardinal Temperatures of Bacillus licheniformis Growing in Various Plant-based Milk Alternatives.
   Int. J. Food Microbiol. 445, 1-8, 2025. ISSN: 0168-1605.
   DOI: https://doi.org/10.1016/j.ijfoodmicro.2025.111497
   IF: 5.2 (2024)

2. **Rockaya, M.**, Baranyi, J.: Variability of growth parameter estimates - The role of rescaling and reparametrization.
   Food Microbiol. 128, 1-8, 2025. ISSN: 0740-0020.
   DOI: http://dx.doi.org/10.1016/j.fm.2025.104726
   IF: 4.6 (2024)

3. Baranyi, J., **Rockaya, M.**, Ellouze, M.: From data to models and predictions in food microbiology.
   Current Opinion in Food Science. 57, 1-6, 2024. ISSN: 2214-7993.
   DOI: http://dx.doi.org/10.1016/j.cofs.2024.101177
   IF: 9.1

Foreign language abstracts (8)

4. **Rockaya, M.**, Baranyi, J.: Incorporating biochemical composition into predictive growth models for plant-based milk products.
   In: ICPMF - Athens Abstract Book, International Committee on Predictive Modelling in Food, Athens, 7, 2025.

5. **Rockaya, M.**, Pecsenye, B., Baranyi, J., Máthé, E.: Modeling microbial growth in plant-based milks: the role of nutrients and experimental design.
   In: A Magyar Táplálkozástudományi Társaság XLVIII. Vándorgyűlésének Programja és az Előadások Összefoglalói. Szerk.: Biró Lajos, Gelencsér Éva, Lugasi Andrea, Rurik Imre, Magyar Táplálkozástudományi Társaság, Budapest, 46, 2025. ISBN: 9786155606175

6. **Rockaya, M.**, Baranyi, J.: Refining Predictive Models for Microbial Growth in Plant-Based Milks through Biochemical Composition and Experimental Design.
   In: IAFP's European Symposium - Madrid Programme Application, IAFP, Madrid, 1, 2025.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

7. **Rockaya, M.**, Baranyi, J.: Exploring the Microbial Dynamics of Plant-Based Milk through its Biochemical Composition.
   In: ICOSTEE 2024 - Book of Abstracts. Szerk.: Bíró István et al, Szegedi Egyetem, Szeged, 86, 2024. ISBN: 9789633069868
8. **Rockaya, M.**, Ellouze, M., Baranyi, J.: Linking the Error of Parameter Estimates for Primary and Secondary Models in Predictive Microbiology.
   In: IAFP's European Symposium - Geneva Programme Application, IAFP, Geneva, 1, 2024.
9. **Rockaya, M.**, Ellouze, M., Baranyi, J.: Comparing the Performance of Two Predictive Models When Fitting Noisy Data.
   In: IAFP's European Symposium - Aberdeen Programme Book, IAFP, Aberdeen, 38-39, 2023.
10. **Rockaya, M.**, Ellouze, M., Baranyi, J.: Database building to study the microbiology of plant-based milk.
    In: FoodMicro2022 - Athens Abstract Book, International Committee on Food Microbiology and Hygiene, Athens, 211, 2022.
11. **Rockaya, M.**, Ellouze, M., Baranyi, J.: Safety and stability of plant-based milk - can predictive microbiology help?
    In: 19th Wellmann International Scientific Conference : Book of abstract. Ed.: Kiss Orsolya, University of Szeged Faculty of Agriculture, Hódmezővásárhely, 73, 2022. ISBN: 9789633068601

List of other publications

Foreign language scientific articles in international journals (2)

12. Pecsenye, B., **Rockaya, M.**, Pacza, T., Mposula, Z., Máthé, E.: Making Sense of Developmental Kinetics Under High-Sugar Stress: Mathematical Modeling of Phenotypic Plasticity in Drosophila melanogaster.
    Nutrients. 18 (8), 1-17, 2026. EISSN: 2072-6643.
    IF: 5 (2024)

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

13. Pacza, T., Martins, M. L., **Rockaya, M.**, Müller, K. E., Chatterjee, A., Barabási, A. L., Baranyi, J.: MilkyBase, a database of human milk composition as a function of maternal-, infant- and measurement conditions.
Sci Data. *9* (1), 1-7, 2022. EISSN: 2052-4463.
DOI: http://dx.doi.org/10.1038/s41597-022-01663-1
IF: 9.8

Total IF of journals (all publications): 33,7
Total IF of journals (publications related to the dissertation): 18,9

The Candidate's publication data submitted to the Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

20 April, 2026

STATEMENT

I wrote this dissertation in the framework of the **Food Sciences** Doctoral School of the University of Debrecen for the purpose of obtaining a doctoral degree (Ph.D.) from the University of Debrecen.

Debrecen, 30./.March../2026

.....

signature of the candidate

STATEMENT

I hereby certify that the doctoral candidate **Maha Rockaya** has carried out her work under my supervision within the framework of the above-mentioned Doctoral School between **2021-2026**. The candidate has made a decisive contribution to the results of the dissertation through her independent creative work, and the dissertation is the candidate's independent work. I recommend that the dissertation be accepted.

Debrecen, 30.../March../2026

.....

signature of the supervisor

ACKNOWLEDGEMENT

To Hungary, a country that offered me not only the opportunity to pursue my studies, but also a safe home where I could grow, learn, and rebuild a sense of stability. I promise to always be good to you, and to ensure that my work reflects what you have given me. You will forever hold a place in my heart as my second home.

To my homeland Syria. Despite the pain, distance, and the many hardships it has endured and caused, it remains deeply rooted in who I am. I carry it with me in everything I do. My hope is to one day see it peaceful and prosperous again.

To the Tempus Foundation and the Stipendium Hungaricum Scholarship, without which I would not have been able to pursue this journey independently.

To the University of Debrecen for providing an environment where I could grow both academically and personally.

To my supervisor, Dr. József Baranyi, for believing in my potential and supporting me unconditionally throughout these years. Every opportunity and experience I have had or will have and everything I have learned I owe it all to you.

To the head of our institute, Dr. Endre Máthé, for his trust, guidance, and kindness. He has been a constant source of support, a mentor and a father figure we could always rely on in times of uncertainty.

To my co-supervisor, Dr. Mariem Ellouze, for being a role model I deeply admire. Her professionalism, strength, and dedication have inspired me and shaped my understanding of what it means to be a true scientist.

To Miss. Judit Szepesi, the angel of our institute. Thank you for all your help, and for the warm afternoon coffees we shared, which always brought lightness to busy days.

To my parents, who raised me to be hardworking and ambitious, and who always respected my choices and encouraged me to be unapologetically myself. They silently endured my absence, my frustrations, and my endless preoccupation with research, your sacrifices are woven into every page of this dissertation.

To my father, this achievement is a tribute to you. In many ways, I feel that I have carried forward a dream that was once taken from you, and I hope this work brings you pride.

To my mother, without your prayers, love, and patience, this work would not exist, and it is as much yours as it is mine.

To my brother, my best friend for life, you are one of the greatest blessings I have. Thank you for being my backbone. I wish you a smooth, fulfilling, and memorable PhD journey of your own.

To my extended family in Syria, your love and care have been my constant source of strength, keeping me warm through the cold days of diaspora.

To my life partner, Bence Pecsenye, thank you for bringing joy, calm, and light into my life. With you, I have found happiness and a sense of home. You have also been my best academic collaborator, patiently turning ideas, experiments, and deadlines into something we somehow survived (and even published!).

To my Hungarian family (Apa, Anya, Nagy mama, Zsuzsi, Samu and Matyi) thank you for making me feel like your real daughter with your love, support and kindness. I will always be grateful to have found a second family in you.

To my friends (near and far). Your presence, your laughter, and your support have made even the hardest days lighter. You may not always realize it, but you have made this tough journey not only bearable, but deeply meaningful.

I also wish to honor the memory of those I have lost. To my beloved uncle Hussam and my master's supervisor Dr. Moufid Yassin (both of whom I lost to COVID), your impact continues to live within me, and I hope I made you proud.

To my city, Baniyas, and its martyrs, especially my chemistry teacher Mr. Ahmed Moussa, your memory remains deeply cherished. I also remember all the innocent souls lost to war, the pandemic, the earthquake, and extremism. This work stands in remembrance of them, until we meet again.

Finally, I dedicate this thesis to all PhD students who are silently battling exhaustion, loneliness, and self-doubt. May you find comfort in knowing that you are not alone.