



Beyond the 4°C Alarm.

01
|
|
|
|
|

*What a temperature log can and cannot
tell us about spoiling milk*

SURYANSH PRATAP SINGH
Shri Mata Vaishno Devi University

Suryansh Pratap Singh

Beyond the 4 °C Alarm: What a Temperature Log Can and Cannot Tell About Milk Spoilage. Published by the author, Katra, Jammu and Kashmir, October 2026. This book has not been peer reviewed.

Evidence status. Chapter 4 and the proofs are mathematical results that hold within the stated models. Chapters 5, 9, 10 and 11, Section 8.2 and the second half of Chapter 12 report simulations: their numbers describe a virtual cold chain built from published parameters, not field measurements. Chapter 7, the first half of Chapter 8 and the first half of Chapter 12 use real published growth data. No new laboratory measurements were made.

Code, data, simulation results and the source of this text: <https://github.com/CopyNinja07-07/dairyguard>. Every number in the text, tables and figures is generated from the result files by script, with fixed random seeds, and can be reproduced with `make all`.

Released under the MIT licence. ComBase records remain subject to the ComBase terms of use.

Preface: how this began

DairyGuard did not start as a research question. It started as a box. For the Smart India Hackathon our team designed a small cold-chain pod for dairy logistics: an ESP32 microcontroller reading a DS18B20 temperature probe, a Peltier cooler, and an app that would warn when milk had been too warm for too long. To turn a temperature reading into “how much life does this milk have left”, the pod needed a shelf-life model, and the usual answer in the literature is a time–temperature integrator with an Arrhenius law.

Building that model raised a question the hardware could not answer. The probe measures the air and the pack. It does not measure the bacteria. Two cartons on the same truck, with the same temperature history, can be in very different states because they started with different numbers and kinds of bacteria. So how much can any temperature log, from any logger, actually tell us about the milk? And if it is wrong, is it wrong because the thermometer is cheap, because it reads too rarely, because the formula is poor, or because of something no thermometer can see?

This book answers that question as carefully as I could. I built a virtual cold chain in which the true answer is always known, tied every number in it to published experiments and to the public ComBase database, and measured how much accuracy each missing piece of information costs. I then asked why the losses fall as they do, and found that the mathematics of non-isothermal kinetics, developed in thermal analysis for chemical reactions, explains them: bacterial growth reads the temperature history only through one accumulated quantity, a biological clock. Finally I tested the parts that can be tested on real data, and set out what the rest would need.

Everything in this book is open. The code, the data, the tests and the scripts that write every number are in the repository named at the front of this book, and anyone can run them. Where a result comes from simulation, the text says so; where it comes from real measurements, it says that too.

Suryansh Pratap Singh

Department of Mathematics and Computing, Shri Mata Vaishno Devi University

October 2026

Abstract

A cold-chain monitor records temperature, but whether milk spoils also depends on which bacteria it carries, how many and how ready they are to grow, none of which a thermometer observes. This book asks how much of the error in predicting spoilage from a temperature record is caused by that unobserved biology, how much by the sensor, how much by the sampling interval and how much by the choice of decision rule.

The question is answered with an open Monte Carlo benchmark of 240 000 simulated deliveries. A Baranyi–Roberts growth model with a square-root temperature law stands in for real milk, and a closed-form Gompertz shelf-life accumulator, cheap enough to run on a microcontroller, tries to predict it from noisy, sampled records. Every rule is judged by its ability to rank spoiled above unspoiled trips, measured by the area under the receiver operating characteristic curve (AUC), with paired bootstrap comparisons on identical trips and an oracle that knows each trip’s biology. Parameters are tied to published experiments and to the ComBase database. The model reproduces a published fluctuating-temperature experiment within 4 %, fits 50 ComBase records of natural spoilage flora, and predicts 53 held-out growth rates of *Listeria monocytogenes* with an accuracy factor of 1.37.

Unobserved biology costs 0.27–0.40 AUC, an order of magnitude more than anything an engineer controls. A platinum resistance sensor instead of a low-cost digital probe gains at most 0.02; hourly sampling is indistinguishable from minute-by-minute sampling in 11 of 12 settings; and a growth-model score improves on a plain degree-hour count by at most 0.04. A Monte Carlo probability is far better calibrated than a single alarm. The same pattern holds in five parameter sets, including a hazard case study for *Listeria*. Treating growth as a non-isothermal kinetic process explains these results: under the square-root law the growth models, and the heat released by a growing culture, depend on the temperature history only through a generalized time, analogous to Ozawa’s generalized time in thermal analysis. Error propagation predicts the size of each loss before any simulation is run; in simulation no Arrhenius or Q_{10} integrator outperforms the square-root integrator in any of 60 settings; and an isothermal calorimetric detection time of the batch, combined with the log, recovers 69–94 % of the loss to biology. Real data supports the central assumption: across 11 independent ComBase studies of *Listeria* in milk, the square-root slope varies with a coefficient of variation of 0.25, the upper end of the range simulated. A monotone neural growth law trained on these records predicts unseen studies no better than the classical laws, and a logistic model trained on 4 000 labelled trips is no better calibrated than the monitor’s own probability, which needs no labels. For this problem, the accuracy of a temperature monitor is limited by biology, not by the thermometer, and effort is better spent on measuring the product than on measuring its temperature more precisely.

The study in numbers

Every cold-chain monitor makes mistakes. This study measured where they come from.

240 000

simulated deliveries

five parameter sets, 12 settings each, every trip read with every sensor, sampling rule and score

≤ 0.02

AUC gained by a lab-grade sensor

a PT1000 platinum probe against a low-cost DS18B20

0.27–0.40

AUC lost to unseen biology

strain, starting load and lag: what no thermometer can measure

≤ 0.04

AUC a growth model adds over degree-hours

a simple degree-hour count captures most of the value

11 of 12

settings where hourly = every minute

reading the temperature once an hour lost nothing measurable

0.03–0.07

calibration error of a probability

against 0.34–0.43 for a single red-or-green alarm

0.25

between-study spread of growth rates

real ComBase data from 11 studies: the hidden biology is not an assumption

69–94 %

of the lost accuracy won back

by one calorimetric test of the batch at dispatch, combined with the log

Five things to take away

1. **The thermometer is not the bottleneck.** The biology inside the carton costs about ten times more accuracy than any choice of sensor or sampling interval.
2. **Cheap hardware is enough.** A ± 0.5 °C digital probe read every 15–60 minutes loses almost nothing.
3. **Simple rules come close, if they are the right ones.** Degree-hours nearly match a growth model; the maximum temperature and time above 8 °C do not.
4. **Say “how likely”, not “spoiled”.** A probability is honest about what the log cannot know and is far better calibrated than an alarm.

5. **Measure the milk, not just the air around it.** One calorimetric test of the batch recovers most of what the log misses.

Contents

Preface: how this began	iii
Abstract	iv
The study in numbers	v
Notation	xiv
1 Introduction	1
1.1 The problem	1
1.2 Why the question matters	1
1.3 Approach	2
1.4 Research questions	2
1.5 Structure of the book	2
2 Background	3
2.1 Spoilage of pasteurised milk	3
2.2 Predictive microbiology	3
2.2.1 Primary models	3
2.2.2 Secondary models	4
2.2.3 Tertiary models and dynamic conditions	4
2.3 Cold-chain monitoring	4
2.4 Uncertainty analyses of shelf life	4
2.5 The gap addressed here	4
3 Models	6
3.1 Temperature and growth rate	6
3.2 The reference model	7
3.3 The edge estimator	7
3.4 Cumulative shelf life	7
3.5 Degree-hours as an approximation	8
3.6 A worked example	8
3.7 An illustrative trip	9
4 Non-isothermal growth kinetics	10
4.1 Relation to Arrhenius kinetics and the apparent activation energy	10
4.2 Reduction to a generalized time	10
4.3 Degree-hours as a first-order approximation	12
4.4 Fixed-law integrators as generalized times	12
4.5 Sampling error	13

4.6	Propagation of sensor error	13
4.7	Propagation of biological variability	14
4.8	From estimation error to ranking accuracy	14
5	The benchmark	16
5.1	Design	16
5.2	Trip generator	17
5.3	Reference biology	17
5.4	Labels	17
5.5	Sensors and sampling	17
5.6	Scores	18
5.7	Parameter sets	18
5.8	Probabilistic output	18
5.9	Protocol	18
6	Statistical methods	19
6.1	The area under the ROC curve	19
6.2	Precision and sample size	19
6.3	Bootstrap intervals	20
6.4	Paired comparisons	20
6.5	Multiplicity	20
6.6	Calibration	20
6.7	Growth-rate fits	20
7	Validation against real data	21
7.1	A fluctuating-temperature experiment	21
7.2	Natural spoilage flora	21
7.3	<i>Listeria monocytogenes</i>	22
7.4	The compiled data	22
8	Temperature laws compared	24
8.1	Four laws fitted to ComBase records	24
8.2	Against Arrhenius and Q_{10} integrators	25
9	Results	27
9.1	Loss budget (RQ1)	27
9.2	Against simple rules (RQ3)	28
9.3	Sensors and sampling (RQ2)	29
9.4	Robustness across parameter sets	30
9.5	Calibration (RQ4)	30
9.6	Mathematics and simulation agree	30
10	<i>Listeria monocytogenes</i>: a hazard case study	32
10.1	Why a pathogen behaves differently	32
10.2	Results	32
10.3	Labels and estimator structure	32
10.4	Scope	33
11	Calorimetry: measuring what the log cannot see	34
11.1	Connection with isothermal calorimetry	34
11.2	Value of a calorimetric measurement of the lot	35
11.3	A role for calorimetry in the cold chain	36

12 Further results on real data and the monitor	37
12.1 How much do real growth rates differ?	37
12.2 Which temperature law predicts unseen studies?	37
12.3 Is a learned probability better?	38
12.4 How many plausible batches are needed?	39
12.5 Correlated biology	39
12.6 How long will this batch last?	39
12.7 What these results ask for	39
13 Monitor design and software	40
13.1 Design rules	40
13.2 Processing on the logger	40
13.3 Software	41
13.4 Firmware	41
13.5 Browser demonstrator	42
14 Discussion	43
14.1 Where effort belongs	43
14.2 Consistency with earlier studies	43
14.3 Hazard and spoilage	43
14.4 Limitations	43
15 Conclusions	45
A Parameters and provenance	46
B Data	48
C Proofs	49
D Reproducing the results	50
References	51

List of Figures

3.1	One excursion through the models	8
3.2	One delivery, three cartons	9
4.1	Temperature dependence of growth. (a) Square-root law for spoilage flora (baseline, solid line) and <i>L. monocytogenes</i> (ComBase fit, dashed line). (b) Shelf life at a steady temperature, $SL = K/\mu$, for the 3×10^4 CFU mL ⁻¹ limit (logarithmic axis). (c) Equivalent Arrhenius activation energy from Eq. (4.1)	11
4.2	The generalized-time reduction (Proposition 1) shown numerically with the Baranyi–Roberts model. (a) Two thermal histories with the same total biological time: a steady 5.5 °C (trip 1, solid line) and 4 °C with a 6-h spell at 13 °C (trip 2, dashed line). (b) Biological time τ . (c) Bacterial counts against elapsed time. (d) Bacterial counts against biological time; the two curves coincide	11
4.3	Why hourly readings are enough (Proposition 3). True growth rate during a 70-min warm spell to 13 °C (solid line) and the rate assumed from hourly readings held until the next reading (dashed line, dots). The hatched area on the rising side is missed and the dotted area on the falling side is over-counted; the two largely cancel	13
4.4	Error budget and its effect on ranking. (a) First-order standard deviation of ln CSL at 4 °C for each source (Table 4.2); the upper four bars are biological, the lower three are instrumental. (b) Binormal relation between noise-to-signal ratio and AUC, Eq. (4.6), with the oracle and a monitor with a perfect thermometer marked	15
5.1	Design of the benchmark	16
7.1	Reproduction of a fluctuating-temperature experiment	21
7.2	Natural spoilage flora in pasteurised milk	22
7.3	<i>Listeria monocytogenes</i> in milk	23
8.1	Secondary models fitted to ComBase records. (a) Spoilage flora in pasteurised milk and (b) <i>L. monocytogenes</i> : growth rates (points) with the square-root (solid), Arrhenius (dashed), constant- Q_{10} (dotted) and Arrhenius–Davey (dash-dotted) fits. (c) Accuracy factor for the records below 8 °C when each law is fitted at or above 8 °C; plain bars spoilage flora, hatched bars <i>Listeria</i> . Data from ComBase [24]	25
8.2	Paired AUC differences between the square-root integrator CSL and (a) an Arrhenius integrator with the activation energy fitted to ComBase records, (b) an Arrhenius integrator with $E_a = 120$ kJ mol ⁻¹ , (c) a $Q_{10} = 3$ rule and (d) degree-hours, on identical trips (DS18B20, 15 min). Bars span the 12 settings of each run and open circles mark the median; the numbers give the settings whose 95 % paired-bootstrap interval excludes zero. Positive values favour CSL	26
9.1	Loss budget	28

9.2	AUC of the five scores	28
9.3	Paired differences in the baseline run	29
9.4	Four effects across the five parameter sets	30
9.5	Calibration of probabilities and alarms	31
11.1	Calorimetry and the generalized time. (a) Simulated heat-flow curves of the same milk biology (baseline parameters, $h_0 = 3.8$, $N_0 = 50$ CFU mL ⁻¹) under isothermal programmes at 20 °C (solid) and 12 °C (dashed) and a step from 6 to 22 °C after 30 h (dotted), with the heat flow taken as proportional to the rate of cell formation. (b) The same curves divided by $\mu(T)$ and plotted against the generalized time τ ; they coincide (Proposition 6). (c) AUC over the 12 baseline settings when the log is combined with information on the biology: none, the true lag and starting load, the true rate law, an isothermal calorimetric detection time at 20 °C with 5 % measurement error, and all of it (oracle). Bars span the settings and circles mark the mean	35
12.1	Growth laws on real data. (a) Held-out accuracy factor of five laws; dark bars <i>Listeria</i> (leave one study out), light bars spoilage flora (five-fold). (b) <i>L. monocytogenes</i> growth rates in milk from ComBase (points) with the square-root law and the monotone neural law; the band is ± 2 ensemble standard deviations. (c) Square-root slope b fitted separately to each study with at least three records, with the median (dashed)	38
12.2	Further tests of the monitor (simulation, baseline world). (a) Calibration error and AUC of a single alarm, a logistic regression trained on 4 000 labelled trips, and the monitor's Monte Carlo probability. (b) Calibration error against the number of plausible batches. (c) AUC lost to biology when $\ln b$ and $-T_{\min}$ are correlated	39
13.1	Processing on the logger	41

List of Tables

2.1	Approaches to predicting spoilage from a temperature history and the question each leaves open for a monitor	5
4.1	Weight of one hour at temperature T relative to one hour at 4°C under each temperature law (spoilage flora; square-root law with $T_{\min} = -7.7^\circ\text{C}$)	12
4.2	First-order error budget for lnCSL at $\bar{T} = 4^\circ\text{C}$ (baseline biology).	14
5.1	Benchmark runs. “Assumed” marks values without a direct source	18
7.1	Composition of the compiled ComBase milk database by organism. n : records; Rate: maximum growth rate stated; Curve: growth curve only; None: no growth reported; N/T: rate not transcribed.	23
8.1	Four secondary models fitted to the ComBase milk records. RMSE and cross-validated (leave one temperature out) RMSE of $\ln \mu$; AICc (best in bold); accuracy factor A_f on the fitting data; bias and accuracy factors when the model is fitted at $T \geq 8^\circ\text{C}$ and used to predict the records below 8°C ; A_f on the 53 <i>Listeria</i> records held out by condition. Data from ComBase [24]	25
8.2	The CSL score against single-parameter integrators on identical trips (DS18B20, 15 min): minimum/maximum paired AUC difference over the 12 settings of each run, with the number of settings whose 95 % paired-bootstrap interval excludes zero in brackets	26
9.1	Loss budget: mean AUC over the 12 settings of each run as information is removed, and the range of the biology gap (oracle minus ideal sensor).	27
9.2	AUC of the five scores from a DS18B20 record sampled every 15 min (baseline run, 4 000 trips per row). The best score in each row is in bold. The last two columns give the same kinetic score with an ideal sensor sampled every minute, and the oracle that knows each trip’s biology.	29
9.3	Paired AUC differences on identical trips: minimum/maximum over the 12 settings of each run, with the number of settings (of 12) whose 95 % paired-bootstrap interval excludes zero in brackets.	29
9.4	Calibration: range of expected calibration error over the 12 settings, and mean Brier score, for each run. Lower is better.	30
9.5	Analytic predictions (Chapter 4) against simulated results (baseline run unless stated).	31
11.1	Value of a measurement of the biology (baseline run, DS18B20 record sampled every 15 min, 12 settings of 4 000 trips). Gap closed: share of the oracle-minus-log-only AUC difference recovered	36

12.1 Accuracy factor A_f on held-out data (1 is perfect). <i>Listeria</i> : leave one study out; spoilage flora: five-fold cross-validation	38
13.1 Design rules for a milk cold-chain monitor	40
13.2 Automated tests	42
A.1 Model parameters and provenance	46
B.1 ComBase records of aerobic spoilage bacteria in pasteurised milk (Tasmanian Institute of Agriculture; 51 records). $\mu_{\max}$: maximum growth rate in decimal-logarithm units per hour	48

Notation

Symbol	Meaning	Unit
$T, T(t)$	product temperature, temperature history	°C
$\mu(T)$	maximum specific growth rate, natural-log basis	h^{-1}
μ_{10}	maximum growth rate, decimal-log basis, $\mu / \ln 10$	h^{-1}
b	slope of the square-root law	$\text{h}^{-1/2} \text{K}^{-1}$
$T_{\min}$	notional minimum growth temperature	°C
$N, N_0, N_{\max}$	bacterial count, starting count, maximum population density	CFU mL^{-1}
N_L	limit against which spoilage is judged	CFU mL^{-1}
y	natural logarithm of the count, $\ln N$	–
q, q_0	physiological state of the Baranyi–Roberts model and its initial value	–
h_0	lag parameter, $\ln(1 + 1/q_0)$; equal to $\lambda\mu$ at constant temperature	–
λ	lag time	h
A, L	$\log_{10}(N_{\max}/N_0)$ and $\log_{10}(N_L/N_0)$	–
$\text{SL}(T)$	shelf life at a constant temperature T	h
K	shelf-life budget, $\text{SL}(T)\mu(T)$	–
CSL	cumulative shelf life used; the limit is reached at $\text{CSL} = 1$	–
Δt	sampling interval	min, h
cv	coefficient of variation of b between strains	–
σ_T	standard deviation of $T_{\min}$ between trips	K
AUC	area under the receiver operating characteristic curve	–
B_f, A_f	bias and accuracy factors of a growth-rate model	–
ECE	expected calibration error	–

Introduction

A carton of milk carries two histories: the temperature it has been through, and the bacteria it started with. A logger records only the first.

1.1 The problem

Milk is perishable in a precise way. The psychrotrophic bacteria that spoil pasteurised milk grow slowly at 4 °C and several times faster at 13 °C, and every hand-off between the dairy, the truck, the depot, the shop and the household refrigerator is an opportunity for the temperature to drift. Predictive microbiology turns such a temperature history into an estimate of growth: a primary model describes the count against time, a secondary model describes the growth rate against temperature, and a tertiary model applies both along a measured history [1–3].

Temperature loggers are now cheap enough to travel with every crate [4–6], and dynamic shelf-life labels, which replace a printed date by a prediction from the log, have been proposed for dairy and meat [7–9]. A monitor of this kind will be trusted to decide which crates are sold first, which are tested and which are discarded. It must therefore come with an honest account of how often it is wrong, and why.

Given only a temperature record, how well can an inexpensive calculation tell whether milk has passed a limit, and what prevents it from doing better?

1.2 Why the question matters

Post-pasteurisation contamination with psychrotrophic bacteria causes roughly half to three-fifths of fluid-milk spoilage in the United States [10], and Gram-negative psychrotrophs that enter after pasteurisation are the main reason high-temperature short-time milk spoils early [11]. Foodborne disease more broadly causes an estimated 600 million illnesses and 420 000 deaths a year, 175 000 of them in the WHO South-East Asia region [12]. These deaths are dominated by agents other than milk-borne bacteria and are quoted only as context; this work models growth, not disease.

The limits a monitor reports against are process-hygiene criteria. The Food Safety and Standards Authority of India (FSSAI) sets an aerobic plate count of 3×10^4 CFU mL⁻¹ for pasteurised milk [13], and quality deteriorates noticeably near 10^6 CFU mL⁻¹ [10, 14]; both are used here. Throughout this report, a trip is called *spoiled* when the modelled count crosses the stated limit; this is a microbial-limit exceedance, not a sensory judgement or a safety verdict. A recent review of fresh dairy concludes that few studies predict the end of shelf life, that endpoint definitions are inconsistent, that few models are

validated beyond their development data, and that probabilities are preferable to single verdicts [15]. Without a quantitative account of where accuracy is lost, two opposite mistakes are made: money is spent on precise thermometers that cannot improve the prediction, and alarms are believed that are wrong a large fraction of the time.

1.3 Approach

A real cold chain never reveals the truth for every carton, so the question cannot be answered from field data alone. The approach taken here is a controlled simulation in which the truth is always known. Thousands of deliveries are generated with realistic temperature excursions and realistic variation in the bacteria; a reference model decides which cartons truly passed the limit; and a monitor that sees only a sensed, sampled temperature record tries to tell them apart. Information is then removed one source at a time, from an oracle that knows everything to a low-cost probe read once an hour, and the accuracy lost at each step is measured on the same trips. The result is a *loss budget*: a single table that states how much each source of error costs. Every parameter is tied to a published study, a public database or an explicit assumption, and the model is checked against real growth data before any result is read.

1.4 Research questions

- RQ1. How much ranking accuracy is lost to biology that a temperature record cannot see?
- RQ2. How much is lost to the sensor and to the sampling interval, and how much does a better sensor or faster sampling recover?
- RQ3. Does a growth-model score rank trips better than simple rules such as degree-hours, time above a threshold or the maximum temperature?
- RQ4. Is a probability better calibrated than a single alarm?
- RQ5. Why do the losses fall as they do, and does the choice of temperature law matter?
- RQ6. How much of the loss to biology can a calorimetric measurement of the batch recover?

1.5 Structure of the book

Chapter 2 reviews the microbiology of milk spoilage, predictive microbiology and existing approaches to cold-chain monitoring. Chapter 3 sets out the growth models and the edge estimator, and Chapter 4 develops their kinetic theory. Chapter 5 describes the benchmark and Chapter 6 its statistical methods. Chapter 7 validates the model against published and ComBase data, and Chapter 8 compares temperature laws. Chapter 9 reports the results, Chapter 10 the *Listeria* case study, Chapter 11 the calorimetric route to the missing biology, Chapter 12 further tests on real data and of the monitor, and Chapter 13 the design of a monitor, the software and the firmware. Chapter 14 discusses the findings and their limits, and Chapter 15 concludes. The appendices list the parameters with their provenance, the data and the procedure for reproducing every result.

Background

Decades of predictive microbiology tell us how fast bacteria grow at a given temperature. Far less is known about how well that knowledge works inside a real monitor.

KEY POINTS

Pasteurised milk spoils mainly through psychrotrophic bacteria introduced after pasteurisation. Their growth is described by primary and secondary models fitted under constant conditions and applied along changing temperature histories. Existing work quantifies how variable spoilage is and which unknown input matters most; none measures what is left to predict once the temperature has been logged.

2.1 Spoilage of pasteurised milk

Pasteurisation destroys vegetative pathogens and most spoilage organisms, so the bacteria that spoil pasteurised fluid milk are either those that survive heating, chiefly spore-formers, or those that enter afterwards from fillers, pipework and the plant environment [10, 11]. Post-pasteurisation contaminants are dominated by Gram-negative psychrotrophs, above all *Pseudomonas*, which grow at refrigeration temperature, produce heat-stable proteases and lipases, and cause bitterness, off-flavours and coagulation once counts approach 10^6 – 10^7 CFU mL⁻¹ [10]. The starting load is usually very low, often between one and fifty cells per 100 mL [10], so the time to spoilage is long at a well-controlled 4 °C and highly sensitive to warm intervals.

2.2 Predictive microbiology

2.2.1 Primary models

At constant temperature a bacterial population passes through a lag phase, an exponential phase and a stationary phase. The modified Gompertz model [16, 17] describes this sigmoid curve with parameters that have a direct biological meaning (lag, maximum rate and asymptote) and has a closed-form inverse, which makes it convenient for computing a time to a limit. The Baranyi–Roberts model [18, 19] describes the same curve through a differential equation with a physiological state variable that measures how ready the cells are to grow; because it is a differential equation, it handles changing temperature naturally, and it has become the reference model for dynamic conditions.

2.2.2 Secondary models

The secondary model gives the growth rate as a function of temperature. Below the optimum the square-root law of Ratkowsky [20] is the most widely used for psychrotrophs; the cardinal-temperature model [21] describes the whole biokinetic range, and Arrhenius-type laws with a fixed activation energy are used in time–temperature integrators [22]. ComBase gathers tens of thousands of growth records from which such models can be fitted [23, 24]. For psychrotrophic *Pseudomonas* in milk, the square-root law was fitted to 68 growth curves and tested in industry [25, 26].

2.2.3 Tertiary models and dynamic conditions

Models fitted at constant temperature are applied along changing histories by integrating the growth rate [27]. Growth under fluctuating temperature has been measured directly for *P. fluorescens* in milk [14], and shifts of temperature during the lag phase can add an extra lag [28]. Whether the extra lag matters for a monitor is one of the structural questions tested here.

2.3 Cold-chain monitoring

Time–temperature indicators and mean kinetic temperature integrate an Arrhenius rate with a fixed activation energy [9, 22]. The Safety Monitoring and Assurance System predicts remaining shelf life from logged temperature with a square-root model [7, 29]. Field studies of Swedish chains show large gaps between dynamically predicted and printed shelf life [8]; reviews describe how often cold chains break [30]; neural networks estimate product temperature from vehicle and door data [31]; heat-transfer simulation quantifies the cost of failures [32]; and recent work couples growth models with learned components [33].

2.4 Uncertainty analyses of shelf life

The studies closest to this work are Monte Carlo analyses of shelf life [34]. For fluid milk, Lau et al. modelled spoilage from Gram-negative post-pasteurisation contamination and found the maximum growth rate and the storage temperature most influential [35]. Buehler et al. and Qian et al. modelled psychrotolerant spore-formers and deployed the model as a supply-chain tool, with home storage temperature, transport temperature and initial spore load ranked highest [36, 37]. Giannakourou and Taoukis propagated kinetic uncertainty and cold-chain variability through nested Monte Carlo loops [38]. In all of these, temperature enters as an uncertain input drawn from a distribution.

2.5 The gap addressed here

Table 2.1 sets the approaches side by side. They ask how variable the outcome is, which input moves it most, or how to build a decision system. None asks the question a monitor faces: starting from a *measured* record and judged against a known truth, how much of the ability to rank trips by spoilage is lost to unobserved biology, to sensor error, to the sampling interval and to the choice of rule. The two views are complementary. When temperature is unknown it dominates the spread of outcomes; once it is logged, what remains is mostly biology, and that is what a loss budget measures.

Table 2.1 Approaches to predicting spoilage from a temperature history and the question each leaves open for a monitor

Approach	Input	Temperature law	Biological variability	Open question for a monitor
Threshold alarm, time above a limit	logged temperature	step	ignored	how often the alarm is wrong
Degree-hours	logged temperature	linear	ignored	how close it comes to a growth model
TTI, mean kinetic temperature [9, 22]	logged or physical integral	Arrhenius	ignored	accuracy lost to biology
Dynamic shelf life, SMAS [7, 29]	logged temperature	square-root	nominal	loss to biology, sensor and sampling
Monte Carlo spoilage models [35–37]	simulated temperature	square-root	sampled	what remains once temperature is logged
Nested uncertainty [38]	simulated temperature	Arrhenius	sampled	accuracy of a monitor
Product-temperature estimation [31]	vehicle data	Q_{10}	ignored	effect on ranking of spoiled units
This work	sensed, sampled log	square-root	sampled, with oracle	–

Models

Three equations carry the whole study: one for how temperature sets the growth rate, one for real milk, and one cheap enough to run on a low-cost microcontroller.

KEY POINTS

A square-root law gives the growth rate; a Baranyi–Roberts model stands in for real milk; and a Gompertz accumulator, deliberately a different model, is what the monitor computes. The accumulator reduces a temperature record to the integral of the growth rate divided by a fixed budget, which is why a log can be read cheaply and why the budget, which carries the biology, limits accuracy.

3.1 Temperature and growth rate

The square-root law relates the maximum specific growth rate μ (h^{-1} , natural-log basis) to temperature below the optimum [20],

$$\sqrt{\mu(T)} = b(T - T_{\min}), \quad T > T_{\min}, \quad (3.1)$$

and $\mu = 0$ at or below $T_{\min}$. Neumeier fitted Eq. (3.1) to 68 growth curves of a fast-growing *Pseudomonas* in broth and milk, giving $T_{\min} = -7.7^\circ\text{C}$ and, on the natural-log basis used here, $b = 0.030 \text{ h}^{-1/2} \text{ K}^{-1}$ [25]. Rates of *P. fluorescens* in UHT milk at 4, 15 and 29°C [14] give $b = 0.027$ and $T_{\min} = -9.0^\circ\text{C}$. The two independent sources agree within 15 % on the generation time at 4°C , 5.6 and 6.4 h.

The law implies that the rate grows with the square of the distance from $T_{\min}$:

$$\frac{\mu(T_2)}{\mu(T_1)} = \left(\frac{T_2 - T_{\min}}{T_1 - T_{\min}} \right)^2. \quad (3.2)$$

With $T_{\min} = -7.7^\circ\text{C}$, one hour at 13°C is worth $(20.7/11.7)^2 = 3.1$ hours at 4°C , and one hour at 17.5°C is worth 4.6 hours. Short warm spells are therefore expensive, and the cost rises faster than the temperature.

3.2 The reference model

The Baranyi–Roberts model [18, 19] stands in for real milk. For $y = \ln N$ and the physiological state q ,

$$\frac{dy}{dt} = \mu(T) \frac{q}{1+q} (1 - e^{y-y_{\max}}), \quad (3.3)$$

$$\frac{dq}{dt} = \mu(T) q, \quad h_0 = \ln(1 + 1/q_0). \quad (3.4)$$

The factor $q/(1+q)$ starts small and rises towards one as the cells adapt, which produces the lag; the factor $1 - e^{y-y_{\max}}$ stops growth at the maximum density. At constant temperature the lag is $\lambda = h_0/\mu$, so the product $\lambda\mu = h_0$ is constant for cells with the same history: cells that are slow to start are slow at every temperature in proportion. Because μ is evaluated along the trace, Eqs. (3.3)–(3.4) apply to any temperature history. They are integrated on a one-minute grid.

3.3 The edge estimator

A logger cannot integrate differential equations for many parameter sets cheaply, and the estimator should in any case differ from the reference so that the comparison is not circular. The monitor therefore uses the modified Gompertz curve [17],

$$\log_{10} \frac{N(t)}{N_0} = A \exp \left\{ -\exp \left[\frac{\mu_{10} e}{A} (\lambda - t) + 1 \right] \right\}, \quad (3.5)$$

with $\mu_{10} = \mu / \ln 10$ and $A = \log_{10}(N_{\max}/N_0)$. Setting $\log_{10}(N/N_0) = L = \log_{10}(N_L/N_0)$ and solving for t gives the shelf life at a constant temperature,

$$\text{SL}(T) = \lambda + \frac{A}{\mu_{10}(T)e} \left[1 - \ln \left(-\ln \frac{L}{A} \right) \right], \quad \lambda = \frac{h_0}{\mu(T)}. \quad (3.6)$$

Both terms are proportional to $1/\mu(T)$, so

$$\text{SL}(T) = \frac{K}{\mu(T)}, \quad K = h_0 + \frac{A \ln 10}{e} \left[1 - \ln \left(-\ln \frac{L}{A} \right) \right], \quad (3.7)$$

where the budget K depends on the lag and on the starting, limiting and maximum counts, but not on temperature.

3.4 Cumulative shelf life

A reading T_i held over an interval Δt_i consumes a fraction $\Delta t_i / \text{SL}(T_i)$ of the shelf life. The cumulative shelf life used is

$$\text{CSL} = \sum_i \frac{\Delta t_i}{\text{SL}(T_i)}, \quad \text{limit reached when } \text{CSL} \geq 1, \quad (3.8)$$

and, by Eq. (3.7),

$$\text{CSL} = \frac{1}{K} \sum_i \mu(T_i) \Delta t_i \approx \frac{1}{K} \int_0^t \mu(T(s)) ds. \quad (3.9)$$

The monitor integrates the growth rate along the log and compares the result with a fixed budget. In practice $\text{SL}(T)$ is stored as a table on a 0.01 °C grid, and each reading costs one table look-up and one addition.

The log measures how much growth the temperature allowed. It cannot measure how much growth the milk could afford.

Equation (3.9) separates what the record can supply from what it cannot. The integral is a property of the temperature history and is measured by the logger. The budget K and the rate scale b^2 are properties of the bacteria in the carton: their starting load, their physiological state and their strain. No temperature record contains them.

3.5 Degree-hours as an approximation

Writing $T - T_{\min} = (T_r - T_{\min}) + (T - T_r)$ in Eq. (3.1) and expanding the square shows that the integral in Eq. (3.9) consists of a term common to all trips of equal length, a term proportional to the signed degree-hours about T_r , and a quadratic term that gives extra weight to large excursions. Near refrigeration temperature the linear term dominates: for $T_r = 4\text{ °C}$, an excursion to 13 °C receives 211 units from the linear term and 81 from the quadratic term, so degree-hours capture about 72 % of its cost, and about 47 % for an excursion to 30.7 °C . A degree-hour count should therefore rank trips almost as well as the growth-model score, losing a little on hot excursions. Chapter 9 tests this.

3.6 A worked example

With the baseline parameters, $\mu(4) = (0.030 \times 11.7)^2 = 0.123\text{ h}^{-1}$ and $\mu(13) = 0.386\text{ h}^{-1}$. For the $3 \times 10^4\text{ CFU mL}^{-1}$ limit the shelf life is about 83 h at a steady 4 °C and about 26 h at 13 °C . One hour at 4 °C consumes 1.2 % of the shelf life and one hour at 13 °C 3.8 %, so a 90-min spell at 13 °C costs about 4 % of shelf life beyond what the same time at 4 °C would have cost. Figure 3.1 follows a single excursion through the reference model and the accumulator.

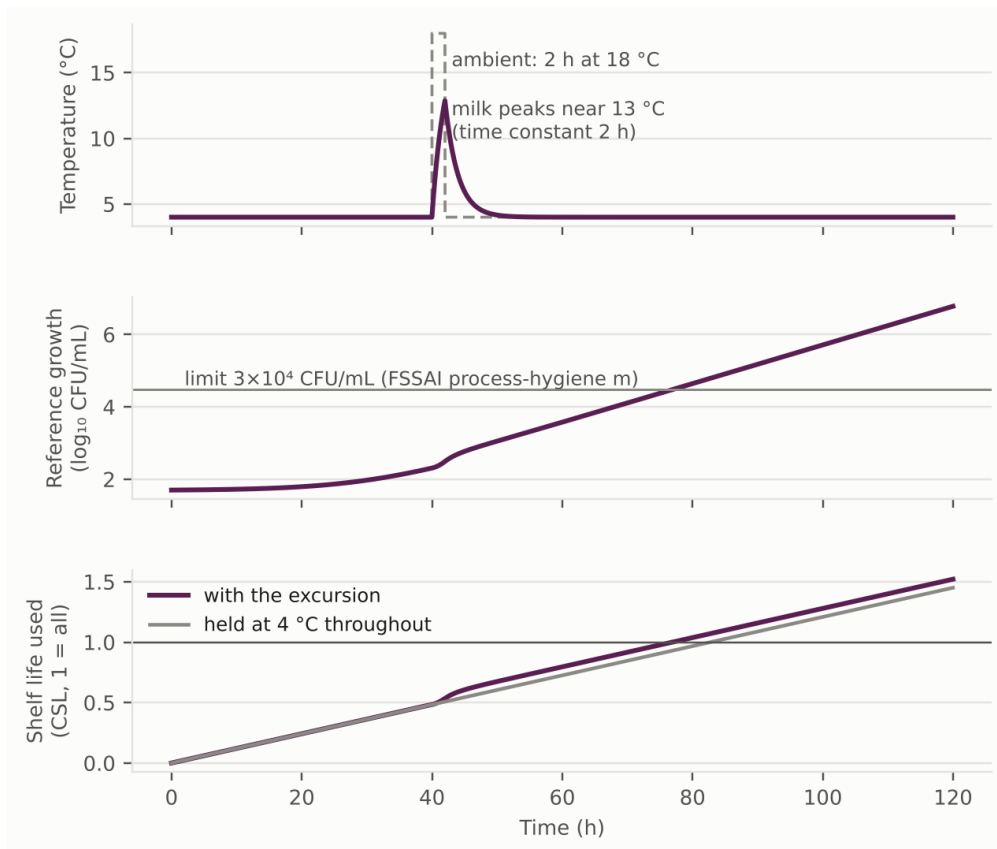


Figure 3.1 One excursion through the models. Top: ambient temperature (2 h at 18 °C) and product temperature of an insulated crate. Middle: reference growth. Bottom: cumulative shelf life with and without the excursion

3.7 An illustrative trip

Figure 3.2 shows the central difficulty on one 72-h delivery. Three cartons travel together and therefore share a single temperature history; they differ only in their starting load and the physiological state of their bacteria. The monitor, reading the log, returns one prediction for all three. It is late for one carton and early for the other two, and a more accurate thermometer cannot change that, because the information it lacks is not in the temperature record.

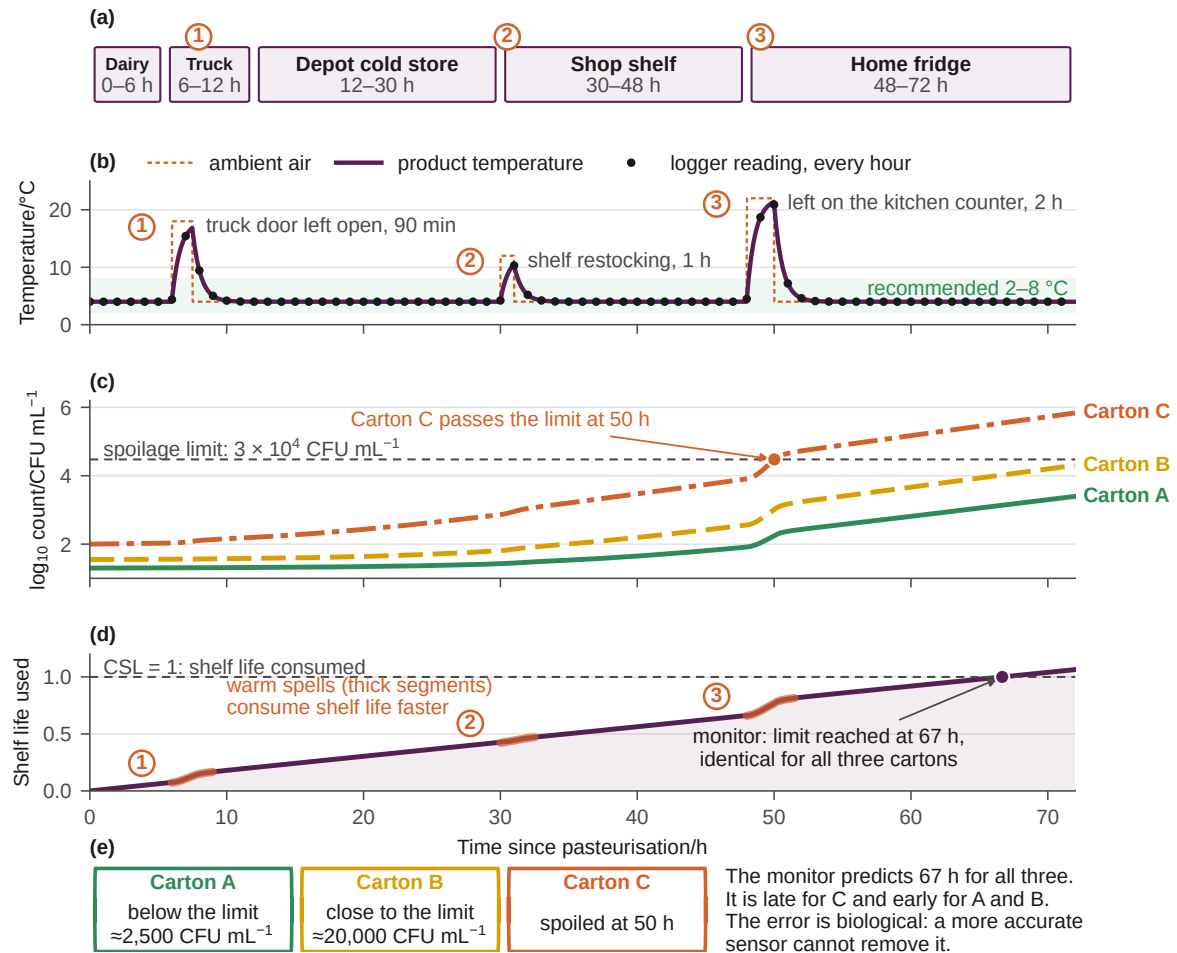


Figure 3.2 One delivery, three cartons. (a) Stages of the trip. (b) Ambient air and product temperature with three excursions and hourly logger readings. (c) Simulated counts in three cartons with the same history but different starting load and lag. (d) Cumulative shelf life computed by the monitor; warm spells consume shelf life faster. (e) One prediction for three cartons whose true fates differ

Non-isothermal growth kinetics

Why does a cheap degree-hour count come so close to a growth model, and why can no thermometer fix the rest? The answer is one accumulated number.

KEY POINTS

Growth in a changing temperature is a non-isothermal kinetic process, the same mathematical problem that thermal analysis has studied for sixty years. Under the square-root law, both growth models depend on the temperature history only through a generalized time, the integral of the growth rate. Every simple integrator is an approximation of that integral, and error propagation shows that biology outweighs sensor error about sevenfold.

4.1 Relation to Arrhenius kinetics and the apparent activation energy

In thermal analysis the rate constant usually follows Arrhenius, $k(T) = A \exp(-E/RT)$ [39]. Ratkowsky et al. found that bacterial growth rates bend away from an Arrhenius line and are described better by Eq. (3.1) [20]. The two laws can be compared through the apparent activation energy, defined as for any rate,

$$E_{\text{app}}(T) = RT_K^2 \frac{d \ln \mu}{dT} = \frac{2RT_K^2}{T - T_{\min}}, \quad (4.1)$$

with T_K the absolute temperature. For spoilage flora with $T_{\min} = -7.7^\circ\text{C}$ it falls from 109 kJ mol^{-1} at 4°C to 84 at 8°C and 66 at 13°C ; for *Listeria* ($T_{\min} = -2.3^\circ\text{C}$) from 203 to 89 kJ mol^{-1} (Fig. 4.1c). An isoconversional analysis of growth data would therefore report an activation energy that varies with temperature, which is the usual signature of a process that is not a single Arrhenius step [40]. For the monitor this matters because a time–temperature indicator or a mean-kinetic-temperature rule built on one fixed E [9, 22] treats warm spells differently from the bacteria.

4.2 Reduction to a generalized time

The central observation of this book is that the growth equations care about temperature only through one accumulated quantity.

Definition 4.1 (Biological time). For a temperature history $T(\cdot)$, the biological time elapsed by t is $\tau(t) = \int_0^t \mu(T(s)) ds$.

Proposition 1 (Generalized-time reduction). Let $\mu(T)$ drive both Eqs. (3.3)–(3.4). Then y and q depend on the history only through $\tau(t)$: there are functions Y, Q with $y(t) = Y(\tau(t))$, $q(t) = Q(\tau(t))$, where Y, Q solve the same equations with $\mu \equiv 1$. In particular $q(t) = q_0 e^{\tau(t)}$, the time at which N reaches N_L is the first t with $\tau(t) \geq \tau^*$, and τ^* depends only on $(h_0, N_0, N_{\max}, N_L)$, not on temperature.

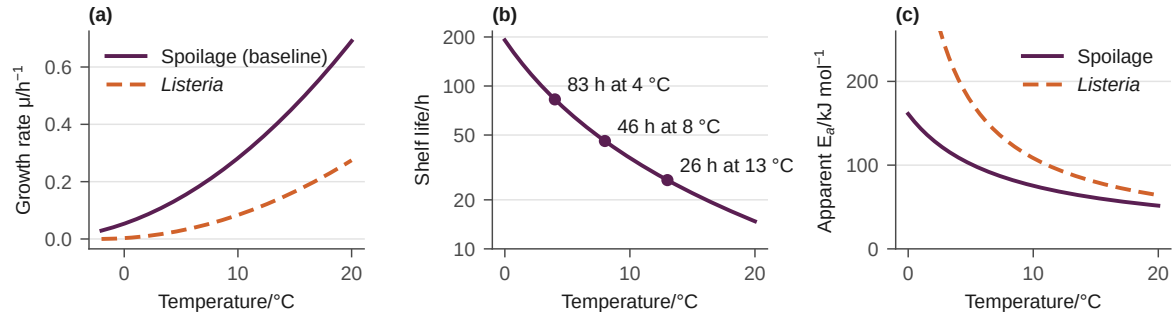


Figure 4.1 Temperature dependence of growth. (a) Square-root law for spoilage flora (baseline, solid line) and *L. monocytogenes* (ComBase fit, dashed line). (b) Shelf life at a steady temperature, $SL = K/\mu$, for the 3×10^4 CFU mL⁻¹ limit (logarithmic axis). (c) Equivalent Arrhenius activation energy from Eq. (4.1)

The construction is the same as Ozawa's generalized time in thermal analysis, $\theta = \int_0^t \exp(-E_a/RT) ds$, which condenses any thermal history into the time an isothermal experiment at infinite temperature would need to reach the same conversion [41–43]. In that notation the kinetic equation $d\alpha/dt = A \exp(-E_a/RT) f(\alpha)$ becomes $d\alpha/d\theta = A f(\alpha)$, and plots of the rate against conversion in generalized time are master plots that do not depend on the temperature programme [44, 45]. Proposition 1 is the same statement for bacterial growth: the right-hand side of Eqs. (3.3)–(3.4) divided by μ is the kinetic model function, τ is the generalized time with the square-root law in place of the Arrhenius factor, and the count plotted against τ is a master curve. Figure 4.2 checks the proposition numerically: two very different histories with the same biological time give different growth paths in elapsed time but exactly the same curve in biological time.

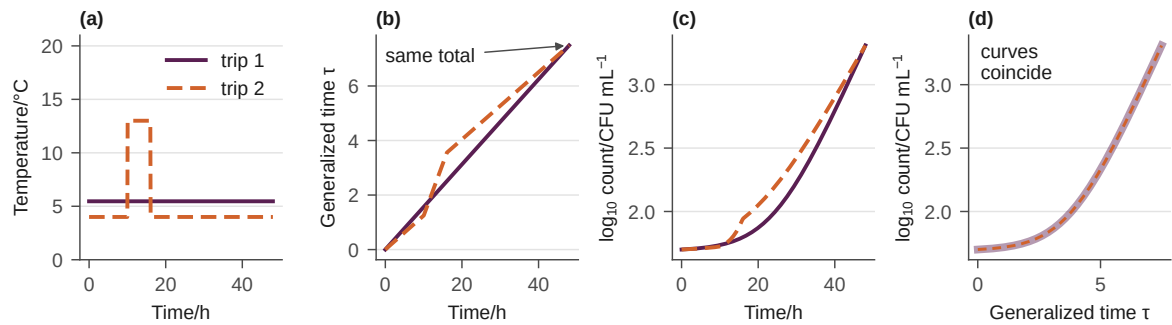


Figure 4.2 The generalized-time reduction (Proposition 1) shown numerically with the Baranyi–Roberts model. (a) Two thermal histories with the same total biological time: a steady 5.5 °C (trip 1, solid line) and 4 °C with a 6-h spell at 13 °C (trip 2, dashed line). (b) Biological time τ . (c) Bacterial counts against elapsed time. (d) Bacterial counts against biological time; the two curves coincide

Proposition 2 (The accumulator is a function of generalized time). *With continuous readings, $CSL(t) = \tau(t)/K$. Hence the Gompertz accumulator and the Baranyi–Roberts reference are both monotone functions of the same biological time, and they cross their limits at $\tau = K$ and $\tau = \tau^*$ respectively.*

A proof of each proposition is given in Appendix C.

4.3 Degree-hours as a first-order approximation

Corollary 1 (Degree-hours as a first-order approximation). *Expand μ about a reference T_r . Under Eq. (3.1),*

$$\text{CSL} = \frac{b^2}{K} \left[(T_r - T_{\min})^2 t + 2(T_r - T_{\min}) \int_0^t (T - T_r) ds + \int_0^t (T - T_r)^2 ds \right]. \quad (4.2)$$

For trips of equal length the first term is common to all trips, the second is a signed degree-hour count about T_r , and the third gives extra emphasis to large excursions. With $T_r = 4$ °C, an excursion to 13 °C gets $(9)^2 = 81$ from the quadratic term against $2 \times 11.7 \times 9 = 211$ from the linear term, so degree-hours capture about 72 % of its cost; for an excursion to 30.7 °C they capture only 47 %. A degree-hour count should therefore rank trips almost as well as CSL, losing a little on hot excursions, which is what the simulation finds (Chapter 9).

4.4 Fixed-law integrators as generalized times

Every single-parameter temperature integrator is a generalized time with some rate law. A time–temperature indicator or mean kinetic temperature (MKT) uses the Arrhenius factor with a fixed E_a [9, 22],

$$\theta_E = \int_0^t \exp \left[-\frac{E_a}{R} \left(\frac{1}{T_K} - \frac{1}{T_{\text{ref}}} \right) \right] ds, \quad (4.3)$$

$$T_{\text{MKT}} = \frac{E_a/R}{E_a/(RT_{\text{ref}}) - \ln(\theta_E/t)},$$

a Q_{10} rule uses $Q_{10}^{(T-T_{\text{ref}})/10}$, and degree-hours use $(T - T_{\text{ref}})^+$. For trips of equal duration T_{MKT} is a monotone function of θ_E , so it ranks trips in the same order. Any of these integrators therefore differs from CSL only in the weight it gives an hour at temperature T relative to an hour at the reference. Table 4.1 lists these weights. Around a warm spell to 13 °C the laws differ by less than a factor of two, but for a 30 °C excursion the Arrhenius weights are three to nine times the square-root weight. An Arrhenius integrator therefore over-penalises hot, short spells relative to the growth they cause, and the larger its E_a the more it does so; this is the mechanism behind the benchmark result in Chapter 9.

Table 4.1 Weight of one hour at temperature T relative to one hour at 4 °C under each temperature law (spoilage flora; square-root law with $T_{\min} = -7.7$ °C)

$T/^\circ\text{C}$	Square-root	Arrhenius 92 kJ mol ⁻¹	Arrhenius 120 kJ mol ⁻¹	$Q_{10} = 3$
8	1.80	1.77	2.10	1.55
13	3.13	3.52	5.14	2.69
17.5	4.64	6.43	11.2	4.41
30.7	10.8	33.8	97.2	18.8

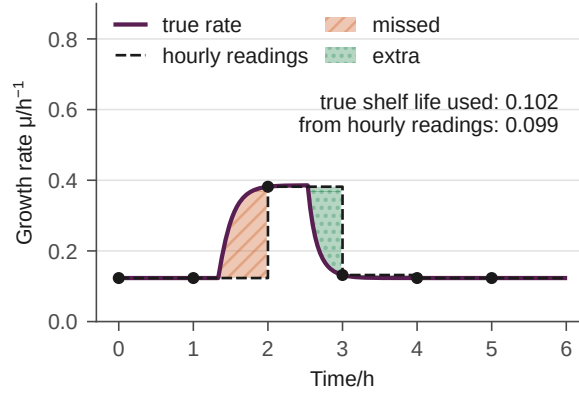


Figure 4.3 Why hourly readings are enough (Proposition 3). True growth rate during a 70-min warm spell to 13 °C (solid line) and the rate assumed from hourly readings held until the next reading (dashed line, dots). The hatched area on the rising side is missed and the dotted area on the falling side is over-counted; the two largely cancel

4.5 Sampling error

A sampled record replaces $\int \mu dt$ by a left Riemann sum. The error has a clean bound.

Proposition 3 (Sampling bound). *Let readings be taken every Δt and held until the next reading. (a) For any history, $|\text{CSL}_{\Delta t} - \text{CSL}| \leq \Delta t V/K$, where V is the total variation of $\mu(T(t))$ over the trip. (b) For a single rectangular excursion from T_0 to T_1 , $|\text{CSL}_{\Delta t} - \text{CSL}| < \Delta t [\mu(T_1) - \mu(T_0)]/K$, whatever its timing. (c) If the excursion's timing is uniform relative to the sampling grid, the error has mean zero and standard deviation $\Delta t [\mu(T_1) - \mu(T_0)]/(K\sqrt{6})$.*

Figure 4.3 shows the mechanism on a single warm spell.

Numerical example. Hourly sampling of a 1-h excursion from 4 to 13 °C changes CSL by less than $(0.386 - 0.123)/10.2 = 0.026$, that is under 3 % of the shelf life, and on average by zero (standard deviation 0.011). Product thermal mass smooths the excursion further and makes the error smaller still.

4.6 Propagation of sensor error

Proposition 4 (Sensor propagation). *Let the sensor read $\tilde{T} = T + \beta + \varepsilon + e_q$ with a per-trip offset β , zero-mean noise ε of variance σ_n^2 and quantisation error $e_q \sim \mathcal{U}(-\delta_q/2, \delta_q/2)$. To first order, for a trip near a steady temperature $\bar{T}$,*

$$\Delta \ln \text{CSL} \approx \frac{2\beta}{\bar{T} - T_{\min}}, \quad \text{E} \Delta \text{CSL}_{\varepsilon} = \frac{b^2}{K} \sigma_n^2 t, \quad (4.4)$$

and the noise and quantisation terms average out over n readings as $1/\sqrt{n}$.

For a DS18B20 with offset uniform within ± 0.5 °C [46], $\text{sd}(\beta) = 0.5/\sqrt{3} = 0.29$ °C, so $\text{sd}(\ln \text{CSL}) = 0.049$; for a Class A PT1000 within ± 0.15 °C [47], 0.015. Noise of 0.05 °C adds a bias of $\sigma_n^2/(\bar{T} - T_{\min})^2 \approx 2 \times 10^{-5}$, which is negligible.

4.7 Propagation of biological variability

Taking logarithms of $\text{CSL} = (b^2/K) \int (T - T_{\min})^2 dt$,

$$\ln \text{CSL} = 2 \ln b + \ln \int (T - T_{\min})^2 dt - \ln K. \quad (4.5)$$

Each unseen biological quantity enters additively on the log scale, and the delta method gives the contributions in Table 4.2: a strain-to-strain cv of 0.10 in b gives $2 \times 0.10 = 0.20$; a spread $\sigma_T = 1.5$ K in $T_{\min}$ gives $2\sigma_T/(\bar{T} - T_{\min}) = 0.26$; and the spread of h_0 and N_0 gives 0.08 through K . In quadrature the biology contributes 0.33, about 7 times the DS18B20 offset.

Table 4.2 First-order error budget for $\ln \text{CSL}$ at $\bar{T} = 4$ °C (baseline biology).

Source	Formula	sd of $\ln \text{CSL}$
Growth-rate slope b (cv 0.10)	$2cv$	0.20
Minimum temperature (σ_T 1.5 K)	$2\sigma_T/(\bar{T} - T_{\min})$	0.26
Lag and starting load (h_0, N_0)	sd of $\ln K$	0.08
Biology, combined	quadrature	0.33
DS18B20 offset (± 0.5 °C)	$2sd(\beta)/(\bar{T} - T_{\min})$	0.049
PT1000 offset (± 0.15 °C)	same	0.015
Hourly sampling, 1-h excursion	Prop. 3(c)	0.011

4.8 From estimation error to ranking accuracy

Figure 4.4a shows the budget of Table 4.2 graphically. Error in $\ln \text{CSL}$ matters only through how often it swaps the order of a spoiled and an unspoiled trip. A simple model makes this precise.

Proposition 5 (Binormal AUC). *Let a trip's true log-margin be $z \sim \mathcal{N}(0, \sigma_z^2)$, spoiled iff $z > 0$, and let the score be $s = z + e$ with independent $e \sim \mathcal{N}(0, \sigma_e^2)$. Then, with $r = \sigma_e/\sigma_z$,*

$$\text{AUC}(r) = \frac{1}{2} + \frac{2}{\pi} \arcsin \frac{1}{\sqrt{2(1+r^2)}}. \quad (4.6)$$

AUC falls from 1 at $r = 0$ to $\frac{1}{2}$ as $r \rightarrow \infty$, and its slope flattens as r grows. Inverting Eq. (4.6) at the simulated AUC of an ideal sensor gives a noise-to-signal ratio of about 2.4 for biology alone. Adding the DS18B20 offset of Table 4.2 then moves r only slightly, and Eq. (4.6) predicts a loss of 0.002 AUC; the simulation finds 0.010. The model is cruder than the simulation, treating every error as Gaussian on one scale, but both put the sensor loss more than an order of magnitude below the biology loss.

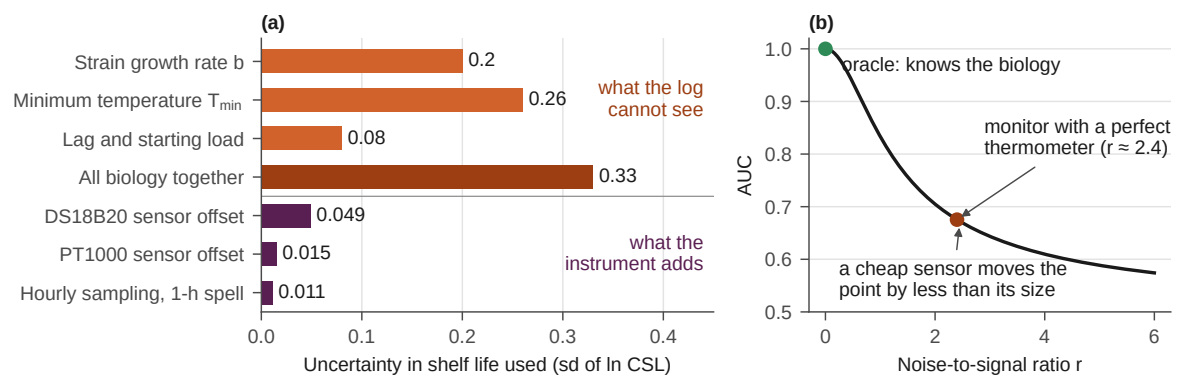


Figure 4.4 Error budget and its effect on ranking. (a) First-order standard deviation of $\ln \text{CSL}$ at 4 °C for each source (Table 4.2); the upper four bars are biological, the lower three are instrumental. (b) Binormal relation between noise-to-signal ratio and AUC, Eq. (4.6), with the oracle and a monitor with a perfect thermometer marked

The benchmark

In a real cold chain nobody knows the truth for every carton. In a simulated one, we always do.

KEY POINTS

Each simulated trip has a true temperature history and a true biology. The reference model labels it spoiled or not; the monitor sees only a sensed, sampled record. Five parameter sets, 12 settings each and 4 000 trips per setting give 240 000 trips, each read with every sensor, sampling rule and score.

5.1 Design

Figure 5.1 summarises the design. A simulated world supplies the truth: a temperature history, a biology and, through the reference model, a label. The monitor sees only a record that has passed through a sensor and a sampling rule, and turns it into a score. Scores are compared with labels by AUC. An oracle with the true biology and the true temperature sets the ceiling, and removing information one step at a time yields the loss budget.

Simulated truth, hidden from the monitor

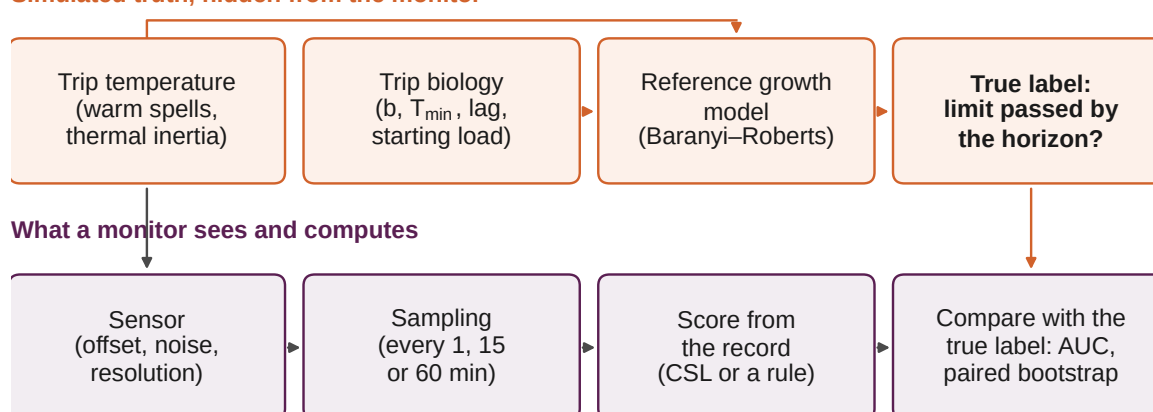


Figure 5.1 Design of the benchmark. Top: simulated truth, hidden from the monitor. Bottom: what the monitor sees and computes. Scores are judged against the true labels by AUC

5.2 Trip generator

Each trip has a product-temperature trace on a one-minute grid. The base temperature of the chain is drawn uniformly from 2.5–5 °C. Excursions arrive as a Poisson process with a mean of one per 144 h of horizon (at most eight per trip); each peaks at 12, 14.3, 17.5 or 30.7 °C with weights 0.30, 0.25, 0.30 and 0.15, representing transport and retail handling, and lasts a lognormal time with median 60 min and log-standard deviation 0.8, clipped to 10–480 min. Product temperature follows the ambient temperature with first-order dynamics,

$$\frac{dT}{dt} = \frac{T_a(t) - T}{\tau_p}, \quad (5.1)$$

with the product time constant τ_p drawn from 5–20 min for a low-mass pouch and 1–4 h for an insulated crate. The two products form one axis of the design: in a pouch the milk follows the air closely, while in a crate a short excursion is strongly damped.

5.3 Reference biology

For each trip the reference population starts from a random load and grows by Eqs. (3.3)–(3.4) along the trace, with

- $T_{\min} \sim \mathcal{N}(\bar{T}_{\min}, \sigma_T^2)$, the spread between strains and trips;
- $b = \bar{b}\varepsilon$ with ε lognormal with coefficient of variation $cv \in \{0, 0.10, 0.25\}$;
- h_0 uniform over a range derived from measured lag times;
- $\log_{10} N_0$ uniform over an assumed range.

In the structure-risk run, a quarter of the lag state is re-imposed whenever the product temperature departs from, or returns to, its two-hour running mean by more than 4 °C, following the extra lag observed after temperature shifts [28]. The estimator is not told about this mechanism.

5.4 Labels

A trip is *spoiled* if the reference population exceeds the limit within the label horizon. The horizon is the time the estimator takes to reach the limit at a constant 4 °C with nominal parameters, rounded up to a whole hour, as a printed shelf life would be. It was fixed from label balance before any rule was compared. Prevalence is therefore a property of the design (about 0.5 for spoilage and 0.08–0.38 for *Listeria*), and absolute predictive values must not be read as field rates. Only ranking (AUC) and calibration against the constructed labels are interpreted.

5.5 Sensors and sampling

Each sensor adds a per-trip offset drawn uniformly within its accuracy band, white noise and quantisation:

- DS18B20 digital sensor: offset within ± 0.5 °C [46], noise 0.05 °C, resolution 0.0625 °C;
- PT1000 platinum resistance sensor read by a MAX31865: offset within ± 0.15 °C (IEC 60751 Class A [47]), noise 0.03 °C, resolution 0.03125 °C;
- an ideal sensor with no error.

Sampling intervals are 1, 15 and 60 min, plus an adaptive rule that samples every minute while the temperature is changing and every 15 min otherwise. Every trip is read with every sensor and sampling rule, so that differences between them are measured on identical histories.

5.6 Scores

Five scores are computed from the measured record: the cumulative shelf life CSL of Eq. (3.8) with nominal parameters; degree-hours above 4 °C; time above 5 °C; time above 8 °C; and the maximum temperature. The oracle computes CSL with each trip's true parameters and noise-free temperature. A sixth output, a probability, is described in Sect. 5.8.

5.7 Parameter sets

Table 5.1 lists the five runs. The baseline uses the Neumeyer parameters and a measured lag range. The extra-lag run adds the structural risk described above. The low-load run uses a starting load close to real post-pasteurisation levels. Because parameters from inoculated isolates need not describe natural flora, the milk-flora run uses parameters fitted to ComBase records of spoilage bacteria in pasteurised milk (Chapter 7). The *Listeria* run models a pathogen with labels of +0.5 and +1 log₁₀ growth above the starting load; these are round-number labels, not regulatory limits. Each run has 12 settings (two limits, two products, three values of cv) of 4 000 trips. The full list of parameters, with the source of every value, is given in Appendix A.

Table 5.1 Benchmark runs. “Assumed” marks values without a direct source

Run	Biology and labels
Baseline	$b = 0.030$, $T_{\min} = -7.7$ °C [25]; $\sigma_T = 1.5$ K (assumed); $h_0 \in [2.5, 5]$ [14]; N_0 30–100 CFU mL ⁻¹ (assumed); limits 3×10^4 and 10^6 CFU mL ⁻¹ ; horizons 83 and 113 h
Extra lag	as baseline, plus extra lag after temperature shifts [28]
Low load	as baseline, N_0 0.1–3 CFU mL ⁻¹ [10]; horizons 120 and 154 h
Milk flora	$b = 0.0256$, $T_{\min} = -6.6$ °C (ComBase, $n = 50$), low load; horizons 201 and 258 h
<i>Listeria</i>	$b = 0.0235$, $T_{\min} = -2.3$ °C (ComBase, $n = 46$); $h_0 \in [2, 4.5]$, $\sigma_T = 1.0$ K, N_0 0.06–0.18 CFU mL ⁻¹ (assumed); labels +0.5 and +1 log; horizons 100 and 202 h

5.8 Probabilistic output

Instead of a single verdict, the monitor can draw $M = 80$ parameter sets $\theta_k = (b_k, T_{\min,k}, h_{0,k}, N_{0,k})$ from a prior and report

$$\hat{p} = \frac{1}{M} \sum_{k=1}^M \mathbf{1}\{\text{CSL}(\theta_k) \geq 1\}, \quad (5.2)$$

the fraction of plausible biologies under which the limit has been reached. Its Monte Carlo standard error is at most $\sqrt{0.25/80} = 0.056$. Two priors are compared with the single alarm that uses nominal parameters: a prior that matches the simulated variation, and an overconfident prior that ignores the strain variation in b .

5.9 Protocol

For every run and setting: (1) generate 4 000 temperature histories; (2) simulate the reference population and label each trip; (3) read each history with every sensor and sampling rule; (4) compute all scores and the oracle; (5) estimate every AUC with a bootstrap interval and every paired contrast with a paired-bootstrap interval; (6) compute the probability and its calibration. Random seeds are fixed, and the whole benchmark runs in about three hours on two processor cores.

Statistical methods

A small difference between two monitors is real only if it survives on the same trips, again and again.

KEY POINTS

Rules are compared by AUC, which measures ranking and does not depend on the constructed prevalence. Differences between rules are tested on identical trips with a paired bootstrap, and calibration is measured by the Brier score and the expected calibration error.

6.1 The area under the ROC curve

Let S_1 be the score of a randomly chosen spoiled trip and S_0 that of a randomly chosen unspoiled trip. The primary metric is

$$\text{AUC} = \Pr(S_1 > S_0) + \frac{1}{2} \Pr(S_1 = S_0), \quad (6.1)$$

the probability that the monitor ranks a spoiled trip above an unspoiled one [48]. A value of 0.5 is a coin toss and 1 is perfect ranking. AUC does not depend on a threshold or on the prevalence, which in this benchmark is constructed, so it measures how much information a score carries. For n_1 spoiled and n_0 unspoiled trips it is estimated by the Mann–Whitney statistic [49],

$$\widehat{\text{AUC}} = \frac{1}{n_1 n_0} \left(\sum_{i: y_i=1} r_i - \frac{n_1(n_1 + 1)}{2} \right), \quad (6.2)$$

where r_i are the ranks of the pooled scores.

6.2 Precision and sample size

Hanley and McNeil approximated the standard error of $\widehat{\text{AUC}}$ by

$$\text{SE}^2 = \frac{A(1 - A) + (n_1 - 1)(Q_1 - A^2) + (n_0 - 1)(Q_2 - A^2)}{n_1 n_0}, \quad (6.3)$$

with $Q_1 = A/(2 - A)$ and $Q_2 = 2A^2/(1 + A)$ [48]. For $A = 0.67$ and $n_1 = n_0 = 2000$, a typical setting, $\text{SE} = 0.0085$. This fixed the number of trips: the loss to biology is more than fifteen standard errors, while sensor and sampling effects are of the order of one standard error and can be resolved only by paired comparison.

6.3 Bootstrap intervals

Intervals for single AUCs are percentile intervals from $B = 200$ bootstrap resamples of trips [50]. The trip is the resampling unit because trips are independent by construction, while the many readings within a trip are not.

6.4 Paired comparisons

Two scores computed on the same trips are correlated, because both respond to the same temperature histories. For a difference $D = \text{AUC}_1 - \text{AUC}_2$,

$$\text{Var}(\hat{D}) = \text{Var}(\widehat{\text{AUC}}_1) + \text{Var}(\widehat{\text{AUC}}_2) - 2\text{Cov}(\widehat{\text{AUC}}_1, \widehat{\text{AUC}}_2), \quad (6.4)$$

and the covariance term is large and positive for similar scores. Testing the two AUCs separately would ignore it and lose most of the power; this is why every trip is read with every sensor, sampling rule and score. The interval for D is obtained by resampling the same trips for both scores $B = 400$ times, the resampling counterpart of the method of DeLong et al. [51].

6.5 Multiplicity

Each run has 12 settings and each contrast is examined in all of them. No conclusion is drawn from a single setting: for each contrast the range of the difference over the settings is reported, together with the number of settings whose 95 % interval excludes zero. Under no real effect about 0.3 such exclusions in a favourable direction would be expected among 12 settings by chance.

6.6 Calibration

A probability $\hat{p}_i$ is calibrated if, among trips given probability p , a fraction p is spoiled. Calibration is measured by the Brier score and by the expected calibration error over ten equal-width bins [52, 53],

$$\text{Brier} = \frac{1}{n} \sum_i (\hat{p}_i - y_i)^2, \quad (6.5)$$

$$\text{ECE} = \sum_{m=1}^{10} \frac{|B_m|}{n} |\bar{y}_{B_m} - \bar{p}_{B_m}|. \quad (6.6)$$

6.7 Growth-rate fits

Secondary models are fitted by ordinary least squares on $\sqrt{\mu}$, with 95 % intervals from bootstrap resamples of records. Agreement between predicted and observed rates is summarised by the bias and accuracy factors of Ross [54],

$$B_f = 10^{\frac{1}{n} \sum \log_{10}(\hat{\mu}_i / \mu_i)}, \quad A_f = 10^{\frac{1}{n} \sum |\log_{10}(\hat{\mu}_i / \mu_i)|}, \quad (6.7)$$

where $B_f = 1$ means no bias and $A_f = 1$ is the typical relative error. Competing laws are compared by the small-sample Akaike criterion [55, 56],

$$\text{AICc} = n \ln \frac{\text{RSS}}{n} + 2k + \frac{2k(k+1)}{n-k-1}, \quad (6.8)$$

where RSS is the residual sum of squares of $\ln \mu$ and k the number of parameters, and by cross-validation in which all records at one temperature are left out in turn.

Validation against real data

A simulation earns trust only when it meets real data and survives.

KEY POINTS

The model reproduces a published fluctuating-temperature experiment within 4%, fits natural spoilage flora with the same temperature shape as inoculated isolates but a lower level, and predicts held-out *Listeria* growth rates without bias.

7.1 A fluctuating-temperature experiment

Lin et al. held UHT milk inoculated with *P. fluorescens* at 4 °C, with and without two four-hour excursions to 29 °C, and report that the time to 10^6 CFU mL⁻¹ fell from about 90 h to about 55 h [14]. With parameters fitted only to their growth rates and their 4 °C lag, the reference model gives 90 h and 53 h (Fig. 7.1), an error of -4 % on the excursion case, which was not used for fitting. This is one experiment with one organism and an inoculated load far above real post-pasteurisation levels, and the published times are approximate values from the text.

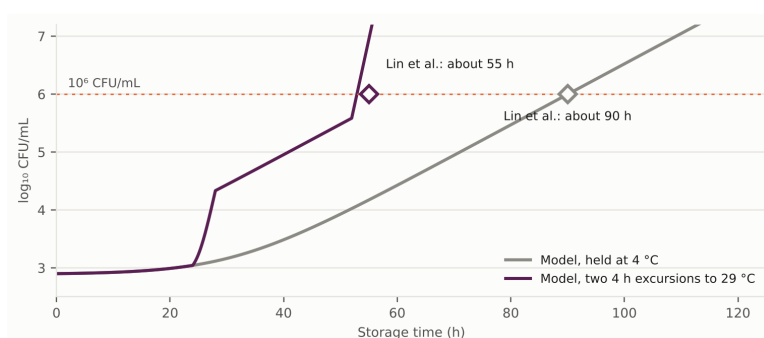


Figure 7.1 Reproduction of the experiment of Lin et al. [14]: model growth at a constant 4 °C and with two four-hour excursions to 29 °C, with the published times to 10^6 CFU mL⁻¹ (diamonds)

7.2 Natural spoilage flora

ComBase holds 51 records of the maximum growth rate of aerobic spoilage bacteria in pasteurised milk at 2.7–14.9 °C from the Tasmanian Institute of Agriculture [24]; they are listed in Appendix B. One record at 6.8 °C lies far below its neighbours. Without it, the square-root law fits the remaining 50 records with $b = 0.0256$, $T_{\min} = -6.6$ °C and a residual standard deviation of 0.16 in natural-log units; with it, $b = 0.0260$ and $T_{\min} = -6.3$ °C, so no conclusion depends on it. Parameters from inoculated

isolates overpredict these records with a bias factor of 1.62 that is nearly constant across temperature (Fig. 7.2): the temperature *shape* transfers from isolates to natural flora, the *level* does not. This is consistent with studies that choose a fast isolate as a worst case, and it is the reason for the milk-flora run.

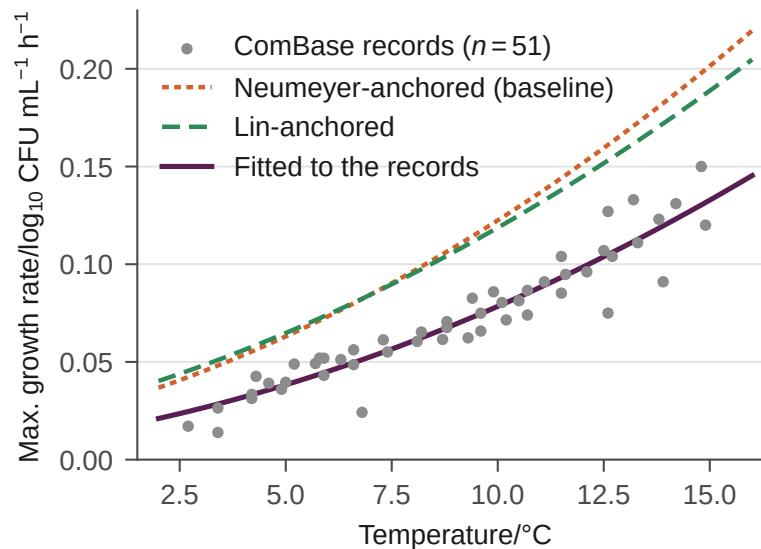


Figure 7.2 Maximum growth rates of natural spoilage flora in pasteurised milk (points; ComBase [24]) with the Neumeyer-anchored model (dotted), the Lin-anchored model (dashed) and the square-root fit to the records (solid)

7.3 *Listeria monocytogenes*

ComBase milk records with a stated growth rate for *L. monocytogenes* at 2–16 °C were split by condition: 46 records without added factors were used for fitting and 53 records with other stated conditions (sterilised or previously heated milk, added salts or acids) were held out. The fit gives $b = 0.0235$ (95 % bootstrap interval 0.0203–0.0265) and $T_{\min} = -2.3$ °C (–4.2 to –0.9 °C), a generation time of 32 h at 4 °C, and on the held-out records $B_f = 1.02$ and $A_f = 1.37$ ($A_f = 1.33$ on the fitted records). The split is by condition rather than by study, so some held-out records share a source with fitted ones and the test is weaker than a split by source. Because $T_{\min}$ lies closer to refrigeration temperature, the rate at 13 °C is 5.9 times that at 4 °C for *Listeria*, against about 3.4 times for spoilage flora (Fig. 7.3).

7.4 The compiled data

A ComBase search with milk as the food matrix returned about 480 records, of which 373 were compiled (Table 7.1), together with the 51 records of spoilage flora. Records without a stated rate were not used for fitting. All compiled records are released with the code.

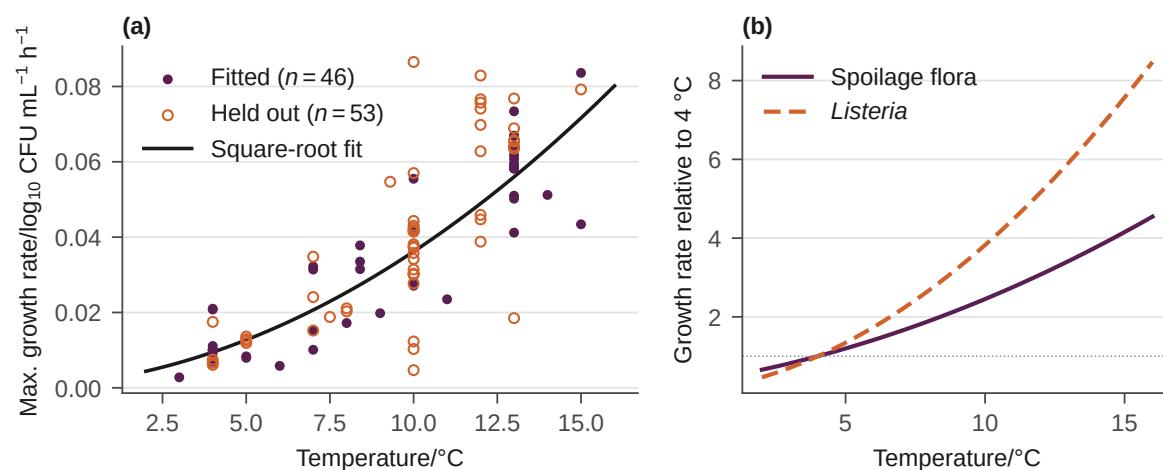


Figure 7.3 *Listeria monocytogenes* in milk. (a) Maximum growth rates used for fitting (filled) and held out by condition (open), with the square-root fit (ComBase [24]). (b) Growth rate relative to 4 °C for spoilage flora (solid) and *Listeria* (dashed)

Table 7.1 Composition of the compiled ComBase milk database by organism. *n*: records; Rate: maximum growth rate stated; Curve: growth curve only; None: no growth reported; N/T: rate not transcribed.

Organism	<i>n</i>	Rate	Curve	None	N/T
<i>L. monocytogenes/innocua</i>	149	102	43	0	4
<i>Bacillus cereus</i>	94	47	36	3	8
<i>Escherichia coli</i>	66	22	15	4	25
<i>Salmonella</i> spp.	25	4	7	13	1
Campylobacteraceae	16	0	13	0	3
Shigellae	11	0	10	0	1
<i>Cronobacter</i> spp.	6	0	5	0	1
Aerobic spoilage bacteria	4	4	0	0	0
Lactic acid bacteria	1	0	1	0	0
Other	1	0	1	0	0
Total	373	179	131	20	43

Temperature laws compared

Shelf-life labels and indicators often assume an Arrhenius law. Does the choice of law matter once the biology is unknown?

KEY POINTS

Four temperature laws fit ComBase growth records about equally well inside the range of the data, but only the square-root law predicts growth in the cold without bias. In the simulated world, whose reference biology follows the square-root law, no Arrhenius or Q_{10} integrator outperforms the square-root integrator in any of 60 settings, although the margins are small. Tested study by study on real *Listeria* data, no law is clearly best (Chapter 12).

8.1 Four laws fitted to ComBase records

Five secondary models are common in food microbiology and thermal analysis. The square-root law, Eq. (3.1), is used throughout; for the whole biokinetic range it is extended to $\sqrt{\mu} = b(T - T_{\min})\{1 - \exp[c(T - T_{\max})]\}$ [20], which reduces to Eq. (3.1) well below the maximum temperature $T_{\max}$. The Arrhenius law and its three-parameter extension by Davey are

$$\ln \mu = \ln A_0 - \frac{E_a}{RT_K}, \quad (8.1)$$

$$\ln \mu = C_0 + \frac{C_1}{T_K} + \frac{C_2}{T_K^2}, \quad (8.2)$$

with T_K the absolute temperature. A constant temperature coefficient gives the exponential law

$$\ln \mu = a + cT, \quad Q_{10} = e^{10c}, \quad (8.3)$$

and the cardinal-temperature model of Rosso et al. describes the rate through $T_{\min}$, the optimum and $T_{\max}$ [21]. Below the optimum, which is the only range relevant to a cold chain, the cardinal model and the extended square-root law reduce to curves of the same shape as Eq. (3.1), so we compare the four laws of Eqs. (3.1) and (8.1)–(8.3).

Each law was fitted by ordinary least squares, the square-root law on $\sqrt{\mu}$ and the others on $\ln \mu$, to the 51 ComBase records of spoilage flora in pasteurised milk and to the 46 *L. monocytogenes* fitting records (Table 8.1, Fig. 8.1). The statistics used are defined in Chapter 6. Within the range of the data all four laws fit about equally well, and the square-root law has the lowest AICc for both organisms. The decisive difference appears in the cold. When each law is fitted to the records at or above 8 °C and used to predict the records below 8 °C, the square-root law is almost unbiased (bias factor 1.01 for spoilage flora and 1.15 for *Listeria*), while the Arrhenius and constant- Q_{10} laws overpredict by

Table 8.1 Four secondary models fitted to the ComBase milk records. RMSE and cross-validated (leave one temperature out) RMSE of $\ln \mu$; AICc (best in bold); accuracy factor A_f on the fitting data; bias and accuracy factors when the model is fitted at $T \geq 8^\circ\text{C}$ and used to predict the records below 8°C ; A_f on the 53 *Listeria* records held out by condition. Data from ComBase [24]

Data	Model	RMSE	AICc	CV RMSE	A_f	Cold B_f	Cold A_f	Held-out A_f
Spoilage flora	Square-root	0.188	-166.2	0.197	1.13	1.01	1.19	–
	Arrhenius	0.202	-159.0	0.214	1.15	1.16	1.23	–
	Constant Q_{10}	0.204	-158.0	0.216	1.15	1.18	1.24	–
	Arrhenius–Davey	0.187	-164.3	0.203	1.13	1.13	1.22	–
<i>L. monocytogenes</i>	Square-root	0.375	-85.9	0.448	1.33	1.15	1.51	1.37
	Arrhenius	0.381	-84.4	0.431	1.33	1.46	1.74	1.39
	Constant Q_{10}	0.383	-84.1	0.430	1.33	1.49	1.76	1.39
	Arrhenius–Davey	0.373	-84.3	0.549	1.33	0.84	1.61	1.38

16–18 % and 46–49 %. The curvature of Eq. (3.1) on an Arrhenius plot is exactly what Ratkowsky et al. observed [20]. Fitted over the whole range, Eq. (8.1) gives $E_a = 92 \text{ kJ mol}^{-1}$ for spoilage flora and 130 kJ mol^{-1} for *Listeria*, of the same order as the 120 kJ mol^{-1} measured calorimetrically for carrot-juice spoilage [57]; the corresponding constant Q_{10} values are 4.0 and 7.1. On held-out *Listeria* records all four laws perform alike, so the advantage of the square-root law is specific to extrapolation into the cold, which is where a cold chain operates.

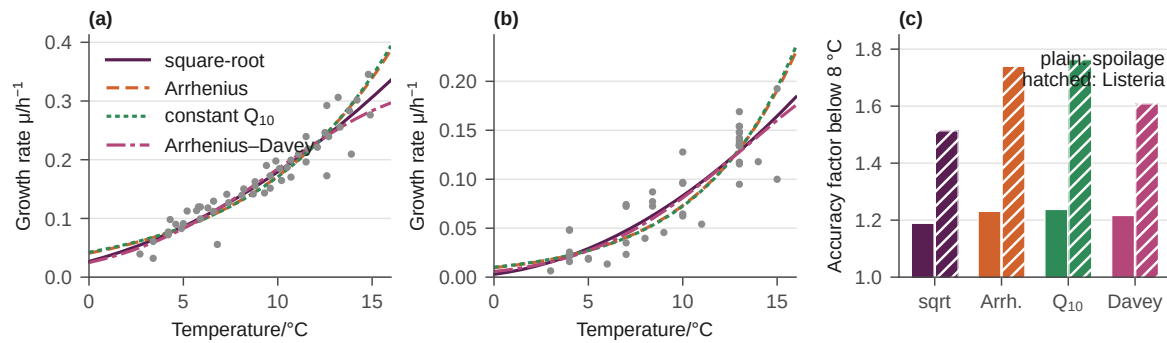


Figure 8.1 Secondary models fitted to ComBase records. (a) Spoilage flora in pasteurised milk and (b) *L. monocytogenes*: growth rates (points) with the square-root (solid), Arrhenius (dashed), constant- Q_{10} (dotted) and Arrhenius–Davey (dash-dotted) fits. (c) Accuracy factor for the records below 8°C when each law is fitted at or above 8°C ; plain bars spoilage flora, hatched bars *Listeria*. Data from ComBase [24]

8.2 Against Arrhenius and Q_{10} integrators

The rules most often used in practice to summarise a temperature log are a mean kinetic temperature, that is an Arrhenius integrator with a fixed activation energy, and a Q_{10} rule. We compared the square-root integrator CSL with three of them on identical trips: Arrhenius with the activation energy fitted to the ComBase records of the run's organism (92 kJ mol^{-1} for spoilage flora, 130 kJ mol^{-1} for *Listeria*), Arrhenius with the 120 kJ mol^{-1} measured calorimetrically for a food-spoilage process [57], and $Q_{10} = 3$ [31] (Table 8.2, Fig. 8.2). No integrator was significantly ahead of CSL in any of the 60 settings. CSL was ahead of the fitted Arrhenius integrator in 52 of 60 settings, significantly so in 28; ahead of the 120 kJ mol^{-1} integrator in 57 (35 significant); and ahead of the Q_{10} rule in 48 (15 significant). By the sign test these proportions are far from what chance would give ($p < 10^{-5}$ for each), with the caveat on dependence noted in Chapter 6. The advantage is largest when extra lag follows temperature shifts, where CSL leads in every setting by up to 0.038 AUC, and it grows with the activation energy, as Table 4.1 predicts. The margins are nevertheless small, of the order of 0.01 AUC: every integrator is a generalized time, and all are limited far more by unobserved biology than by their temperature law.

Table 8.2 The CSL score against single-parameter integrators on identical trips (DS18B20, 15 min): minimum/maximum paired AUC difference over the 12 settings of each run, with the number of settings whose 95 % paired-bootstrap interval excludes zero in brackets

Contrast	Baseline	Extra lag	Low load	Milk flora	<i>Listeria</i>
CSL – Arrhenius, fitted E_a	–0.002/0.008 (4)	0.001/0.021 (11)	–0.001/0.008 (2)	–0.002/0.010 (5)	–0.000/0.028 (6)
CSL – Arrhenius, 120 kJ mol ^{–1}	–0.002/0.019 (5)	0.002/0.038 (11)	0.000/0.016 (5)	0.001/0.023 (8)	–0.001/0.022 (6)
CSL – $Q_{10} = 3$	–0.002/0.005 (3)	0.000/0.014 (9)	–0.001/0.005 (1)	–0.002/0.005 (2)	–0.000/0.003 (0)
Arrhenius, fitted – degree-h	0.003/0.015 (10)	0.003/0.017 (11)	0.001/0.015 (8)	0.002/0.020 (6)	–0.011/0.006 (2)

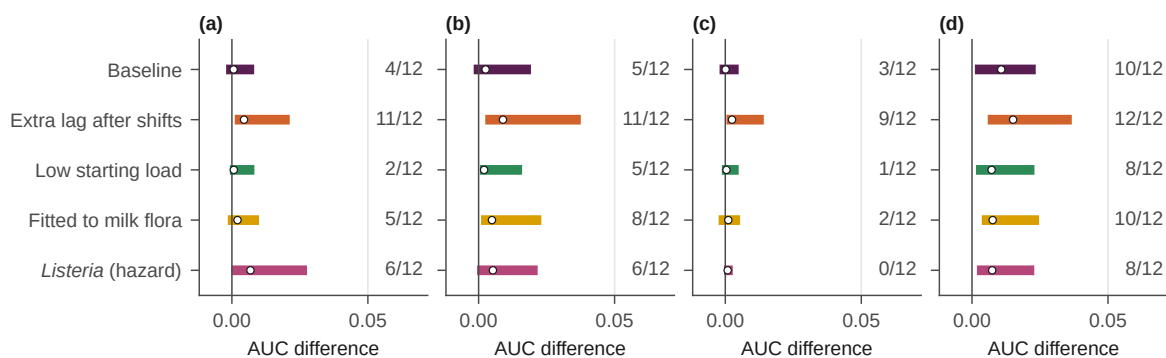


Figure 8.2 Paired AUC differences between the square-root integrator CSL and (a) an Arrhenius integrator with the activation energy fitted to ComBase records, (b) an Arrhenius integrator with $E_a = 120 \text{ kJ mol}^{-1}$, (c) a $Q_{10} = 3$ rule and (d) degree-hours, on identical trips (DS18B20, 15 min). Bars span the 12 settings of each run and open circles mark the median; the numbers give the settings whose 95 % paired-bootstrap interval excludes zero. Positive values favour CSL

Results

Remove one piece of information at a time, and watch where the accuracy goes.

PRINCIPAL FINDINGS

- Unobserved biology costs 0.27–0.40 AUC in the baseline run, an order of magnitude more than any sensor or sampling choice.
- A PT1000 instead of a DS18B20 gains at most 0.02 AUC; hourly sampling matches one-minute sampling in 11 of 12 settings.
- The growth-model score beats time above 5 °C and the maximum temperature in every setting, but improves on degree-hours by at most 0.04.
- A Monte Carlo probability reaches ECE 0.03–0.07; a single alarm 0.34–0.43.
- The pattern is the same in all five parameter sets.

9.1 Loss budget (RQ1)

An oracle that knows each trip's biology and temperature separates spoiled from unspoiled trips almost perfectly, by construction. The same CSL with nominal biology and an ideal sensor loses 0.27–0.40 AUC across the 12 baseline settings (Table 9.1, Fig. 9.1). The loss is large even when strains do not differ in growth rate ($cv = 0$), because the minimum growth temperature, the lag state and the starting load still differ between trips. In terms of Eq. (3.9), these quantities set the rate scale and the budget K , which the temperature record does not contain; no sensor or sampling rate can recover them. Replacing the ideal sensor by a PT1000 or a DS18B20 sampled every 15 or 60 min removes only a further 0.00–0.02.

Table 9.1 Loss budget: mean AUC over the 12 settings of each run as information is removed, and the range of the biology gap (oracle minus ideal sensor).

Run	Oracle	Ideal	PT1000	DS18B20	DS18B20	Biology gap
		1 min	15 min	15 min	60 min	
Baseline	1.000	0.671	0.670	0.661	0.662	0.27–0.40
Extra lag	0.988	0.652	0.652	0.644	0.645	0.27–0.40
Low load	1.000	0.667	0.666	0.656	0.657	0.28–0.40
Milk flora	1.000	0.664	0.663	0.654	0.653	0.28–0.40
Listeria	0.950	0.738	0.737	0.725	0.726	0.09–0.34

A better thermometer is a poor way to improve a spoilage estimate.

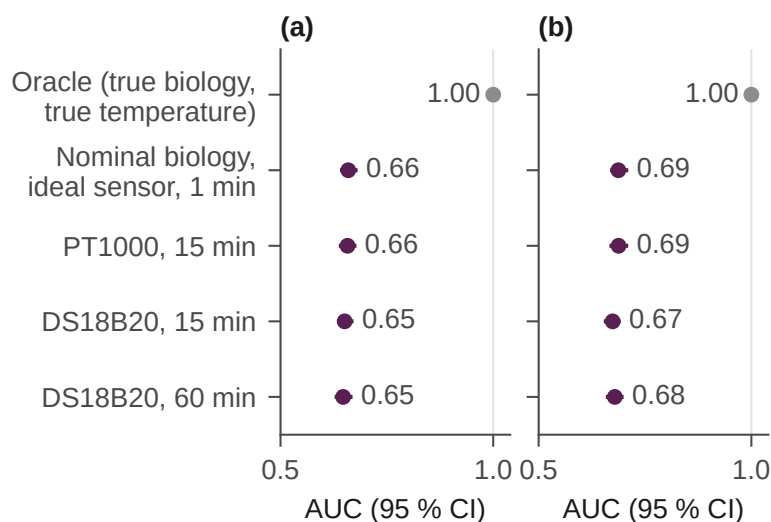


Figure 9.1 Loss budget for (a) a low-mass pouch and (b) an insulated crate (limit 3×10^4 CFU mL⁻¹, $cv = 0.10$). Grey: oracle with true biology and temperature. Purple: the growth-model score as information is removed step by step. Bars: 95% bootstrap intervals

9.2 Against simple rules (RQ3)

On identical trips, the growth-model score beats time above 5 °C in 12 of 12 settings and the maximum temperature in 12 (Table 9.2, Fig. 9.2). Against degree-hours above 4 °C, a rule that needs no growth model, the gain is positive with an interval excluding zero in 10 of 12 settings, but small: 0.00–0.02 AUC in the baseline run and up to 0.04 when extra lag is present. This is the behaviour predicted in Chapter 3: degree-hours are the linear part of the same integral, and the growth model gains only on the large excursions where the quadratic part matters. Time above 8 °C and the maximum temperature discard the duration or the depth of an excursion and rank trips worst.

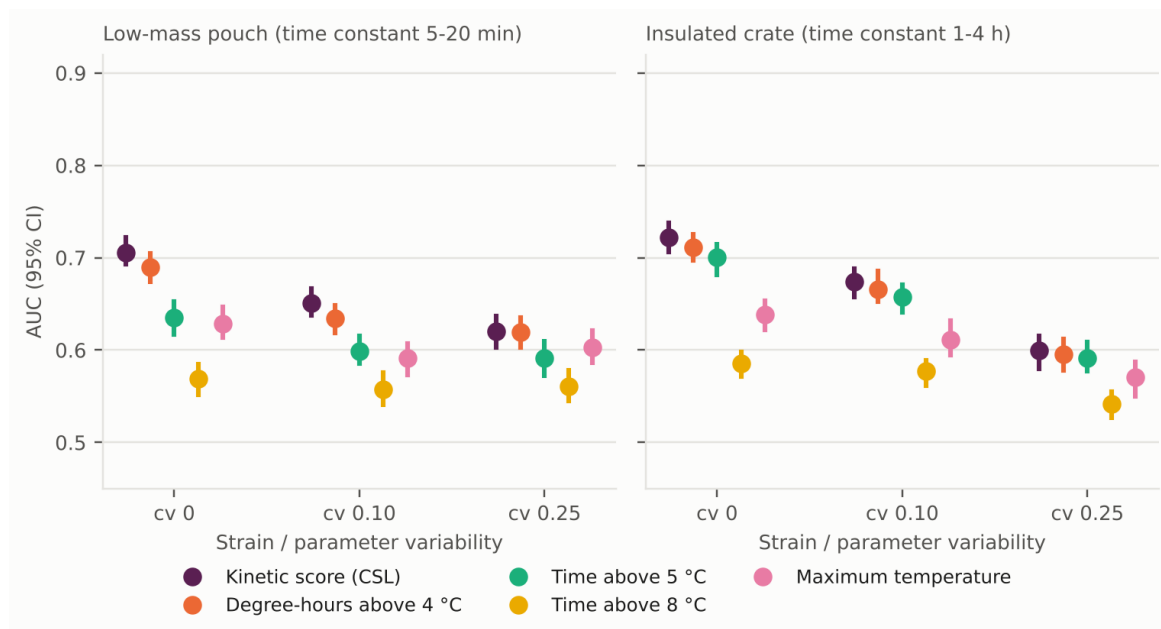


Figure 9.2 AUC of the five scores from a DS18B20 record sampled every 15 min in the baseline run, by product and strain variability (limit 3×10^4 CFU mL⁻¹). Bars: 95% bootstrap intervals

Table 9.2 AUC of the five scores from a DS18B20 record sampled every 15 min (baseline run, 4 000 trips per row). The best score in each row is in bold. The last two columns give the same kinetic score with an ideal sensor sampled every minute, and the oracle that knows each trip's biology.

Limit	Product	c_v	CSL	Degree-h > 4 °C	Time > 5 °C	Time > 8 °C	$T_{\max}$	Ideal, 1 min	Oracle
A	Pouch	0.00	0.705	0.690	0.635	0.569	0.629	0.718	1.000
A	Pouch	0.10	0.651	0.634	0.599	0.557	0.591	0.659	1.000
A	Pouch	0.25	0.620	0.619	0.591	0.561	0.602	0.627	1.000
A	Crate	0.00	0.722	0.711	0.701	0.586	0.638	0.732	1.000
A	Crate	0.10	0.674	0.666	0.657	0.577	0.611	0.687	1.000
A	Crate	0.25	0.599	0.596	0.592	0.541	0.571	0.603	1.000
B	Pouch	0.00	0.710	0.686	0.636	0.564	0.587	0.723	1.000
B	Pouch	0.10	0.656	0.634	0.606	0.539	0.564	0.670	1.000
B	Pouch	0.25	0.604	0.594	0.565	0.531	0.549	0.608	1.000
B	Crate	0.00	0.708	0.696	0.687	0.600	0.629	0.724	1.000
B	Crate	0.10	0.676	0.670	0.663	0.585	0.611	0.686	1.000
B	Crate	0.25	0.603	0.595	0.591	0.538	0.552	0.608	1.000

Table 9.3 Paired AUC differences on identical trips: minimum/maximum over the 12 settings of each run, with the number of settings (of 12) whose 95 % paired-bootstrap interval excludes zero in brackets.

Contrast	Baseline	Extra lag	Low load	Milk flora	Listeria
CSL – degree-h	0.001/0.023 (10)	0.006/0.037 (12)	0.001/0.023 (8)	0.004/0.025 (10)	0.002/0.023 (8)
CSL – time > 5 °C	0.008/0.074 (12)	0.001/0.104 (11)	0.005/0.066 (11)	0.006/0.075 (11)	0.010/0.077 (12)
PT1000 – DS18B20	0.003/0.015 (9)	0.003/0.013 (7)	0.005/0.018 (10)	0.002/0.015 (8)	0.005/0.020 (8)
60 min – 1 min	-0.008/0.004 (1)	-0.008/0.004 (1)	-0.011/0.008 (2)	-0.009/0.006 (0)	-0.008/0.009 (0)
Adaptive – 1 min	-0.006/0.005 (0)	-0.006/0.006 (0)	-0.008/0.009 (1)	-0.007/0.006 (0)	-0.006/0.009 (0)

9.3 Sensors and sampling (RQ2)

Replacing a DS18B20 with a PT1000 gains 0.00–0.02 AUC at 15-min sampling (Table 9.3, Fig. 9.3). The reason is visible in Eq. (3.1): a sensor offset of ± 0.5 °C shifts $\sqrt{\mu}$ by $0.5b$, about 4 % of its value at 4 °C, whereas the trip-to-trip spread of $T_{\min}$, h_0 and N_0 moves the rate scale and the budget by tens of per cent. Sampling hourly instead of every minute shows no detectable difference in 11 of 12 settings and at most about 0.01 where it does, and the adaptive rule is indistinguishable from fixed one-minute sampling. Excursions in the generator last at least 10 min and typically an hour; the conclusion on sampling would change if short, sharp excursions were common.

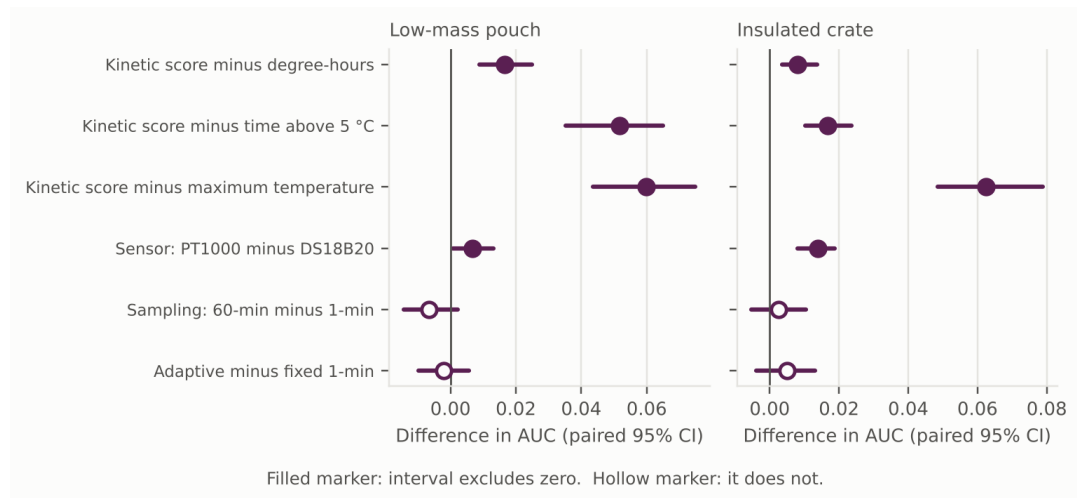


Figure 9.3 Paired AUC differences on identical trips in one baseline setting per product ($c_v = 0.10$, limit 3×10^4 CFU mL⁻¹), with 95 % paired-bootstrap intervals. Filled markers: interval excludes zero

Table 9.4 Calibration: range of expected calibration error over the 12 settings, and mean Brier score, for each run. Lower is better.

Variant	Baseline	Extra lag	Low load	Milk flora	<i>Listeria</i>
Probability, matched prior	0.03–0.07 / 0.231	0.01–0.07 / 0.236	0.02–0.11 / 0.233	0.02–0.09 / 0.234	0.08–0.41 / 0.216
Probability, overconfident	0.04–0.13 / 0.237	0.05–0.13 / 0.244	0.02–0.14 / 0.238	0.02–0.13 / 0.239	0.11–0.44 / 0.218
Single alarm	0.33–0.43 / 0.380	0.34–0.44 / 0.391	0.34–0.44 / 0.386	0.34–0.43 / 0.385	0.34–0.45 / 0.400

9.4 Robustness across parameter sets

Figure 9.4 summarises the four effects across all five runs. The pattern holds whether the biology is anchored to inoculated isolates, includes extra lag, starts from a realistic low load, is fitted to natural flora or describes *Listeria*: the biology gap is 0.28–0.40 (low load), 0.28–0.40 (milk flora) and 0.09–0.34 (*Listeria*); the gain over degree-hours stays below 0.04 in every run; and the sensor gain stays small.

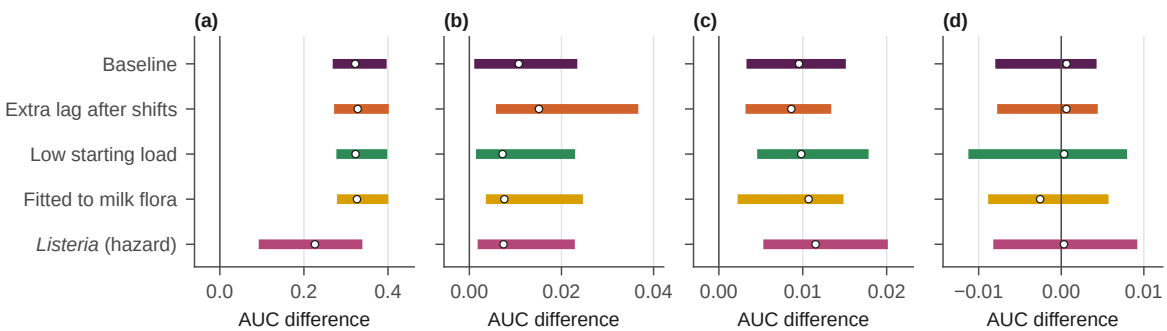


Figure 9.4 Four effects across the five parameter sets, as AUC differences on identical trips: (a) loss to unobserved biology, (b) growth-model score over degree-hours, (c) PT1000 over DS18B20, (d) hourly over one-minute sampling. Bars span the 12 settings of each run; dots mark the median. Horizontal scales differ

9.5 Calibration (RQ4)

A single alarm with nominal parameters is badly calibrated against the constructed labels (ECE 0.34–0.43; Table 9.4, Fig. 9.5). A Monte Carlo probability with a prior that matches the truth reaches ECE 0.03–0.07, and one with an overconfident prior on the growth rate 0.04–0.13. Absolute values reflect the constructed prevalence; the ordering of the three variants is the finding. A probability that admits what the log cannot know is better calibrated than a verdict that ignores it, and even a prior that understates the strain variation retains most of the benefit.

9.6 Mathematics and simulation agree

Table 9.5 sets each analytic prediction of Chapter 4 beside the corresponding simulated result. They were obtained independently: the propositions use only the model equations and parameter ranges, and the simulation uses none of the propositions.

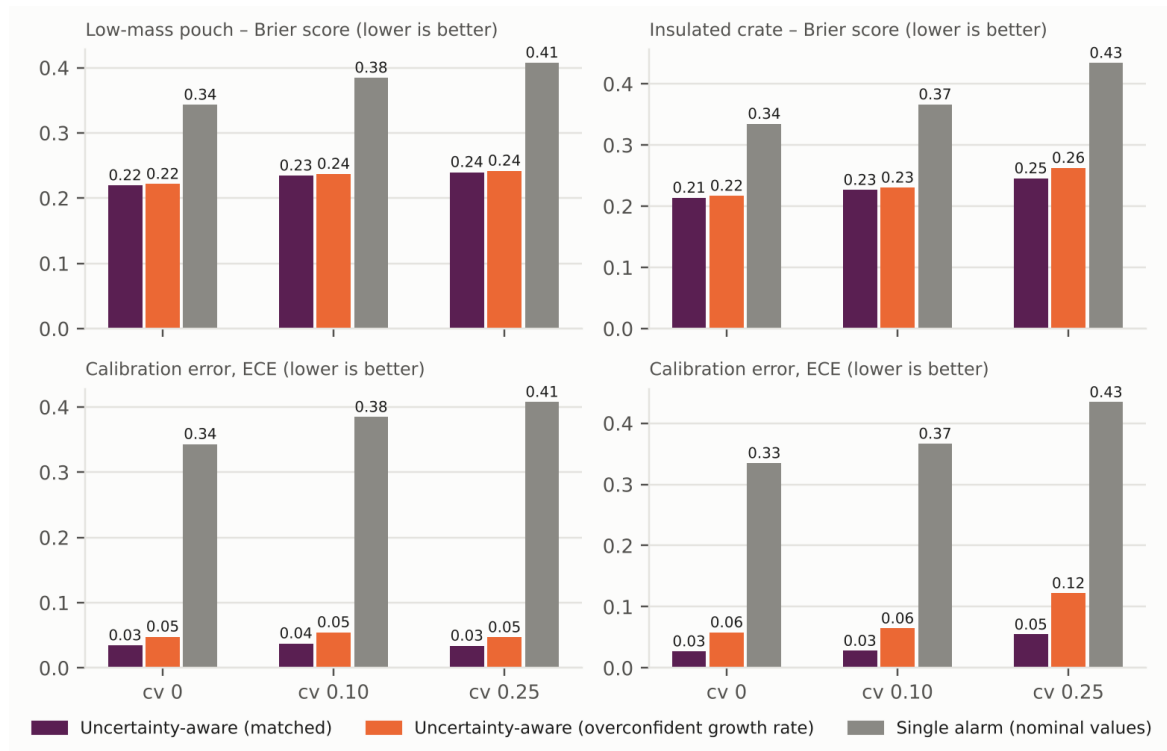


Figure 9.5 Brier score and expected calibration error of the Monte Carlo probability with a matched prior, with an overconfident prior, and of a single alarm with nominal parameters (baseline run, limit 3×10^4 CFU mL⁻¹). Lower is better

Table 9.5 Analytic predictions (Chapter 4) against simulated results (baseline run unless stated).

Question	Mathematics predicts	Simulation finds
Does an oracle rank perfectly?	Yes, both models read τ (Props. 1, 2)	Oracle AUC 1.000 (baseline); below 1 only where structure differs (extra lag 0.988, <i>Listeria</i> half-log)
Biology versus sensor	sd of ln CSL: 0.33 against 0.049, ratio about 7	Biology gap 0.27–0.40 AUC against DS18B20 loss of about 0.01
Effect of sensor on AUC	Binormal model: 0.002 (Prop. 5)	0.010 (ideal → DS18B20, 1 min)
Degree-hours versus CSL	First-order term of the same generalized time (Cor. 1)	CSL ahead by at most 0.04 AUC
Hourly versus every-minute	Error < 0.026 per 1-h excursion, mean zero (Prop. 3)	–0.01 to 0.00 AUC; not detectable in 11 of 12
Which biology matters most	Rate parameters: sd 0.20 (b) and 0.26 ($T_{\min}$) against 0.08 (lag, load)	True rate law adds 0.26–0.40 AUC; true lag and load 0.01–0.04
Calorimetric measurement	Detection time measures rate and budget together (Prop. 6)	Closes 69–94 % of the biology gap
Value of a probability	sd of $\hat{p} \leq 0.056$ with $M = 80$	ECE 0.03–0.07 against 0.34–0.43 for an alarm

Listeria monocytogenes: a hazard case study

For a pathogen that still grows in the refrigerator, a warm hour is even more expensive. The lesson stays the same.

KEY POINTS

For a pathogen with a minimum growth temperature close to refrigeration, a warm hour costs more, but the loss budget has the same shape: biology first, instruments a distant second.

10.1 Why a pathogen behaves differently

The *Listeria* fit places $T_{\min}$ at -2.3°C , much closer to refrigeration than the -7.7°C of spoilage *Pseudomonas*. By Eq. (3.2), the same warm hour is therefore relatively more expensive: the rate at 13°C is 5.9 times the rate at 4°C , against about three times for spoilage flora. At 4°C itself growth is slow, with a generation time of 32 h. Temperature excursions thus matter even more for the hazard than for spoilage.

10.2 Results

With labels of $+0.5$ and $+1 \log_{10}$ growth, unobserved biology costs 0.09–0.34 AUC, a PT1000 gains 0.01–0.02, hourly sampling matches minute sampling in 12 of 12 settings, and the gain of the growth-model score over degree-hours is at most 0.02 (Tables 9.1–9.4). The lower end of the biology gap is smaller than for spoilage because the narrow starting-load range and the slow growth at 4°C leave less room for biological differences to reorder trips.

10.3 Labels and estimator structure

The half-log label requires a caveat. The limit then lies inside the lag-to-growth transition, where the Gompertz estimator approximates the Baranyi–Roberts reference poorly; even the oracle reaches only 0.88–0.92 AUC (1.00–1.00 for the $+1 \log$ label). With the $+1 \log$ label the probability reaches ECE 0.08–0.20 against 0.34–0.41 for the alarm; with the half-log label it remains better than the alarm (0.28–0.41 against 0.42–0.45) but is poorly calibrated in absolute terms. For very small growth targets, the structure of the estimator, and not biology alone, limits accuracy.

10.4 Scope

This case study concerns growth, not illness. Risk of illness depends on dose, virulence, host susceptibility and the contamination event, none of which are modelled. The FSSAI criterion for *L. monocytogenes* in milk products is absence in 25 mL [13], which no growth threshold represents.

Calorimetry: measuring what the log cannot see

A thermometer cannot see the bacteria. A calorimeter can feel their heat.

KEY POINTS

The heat released by a growing culture depends on the temperature programme only through the same generalized time, so heat-flow curves collapse onto one master curve. An isothermal calorimetric detection time of the batch measures the biology that dominates the loss budget; combined with the log, it recovers 69–94 % of the gap to the oracle in the simulated benchmark.

11.1 Connection with isothermal calorimetry

Isothermal calorimetry follows microbial growth through the heat flow P of a sample [57–59]. During growth the heat flow is taken as proportional to the rate at which cells are formed, $P = Y_Q V \, dN/dt$, where Y_Q is the heat released per cell formed and V the sample volume [58, 60]. Under the Baranyi–Roberts model, $dN/dt = N \, dy/dt = \mu(T) N \frac{q}{1+q} (1 - N/N_{\max})$, and Proposition 1 turns this into a statement about calorimetric curves.

Proposition 6 (Calorimetric master curve). *Let Y_Q be constant and let the growth follow Eqs. (3.3)–(3.4). Then, for any temperature programme,*

$$\frac{P(t)}{\mu(T(t))} = Y_Q V G(\tau(t)), \quad Q(t) = \int_0^t P \, ds = Y_Q V [N(\tau(t)) - N_0], \quad (11.1)$$

where $G(\tau) = N \frac{q}{1+q} (1 - N/N_{\max})$ is evaluated on the solution with $\mu \equiv 1$. The calorimetric conversion $\alpha = Q/Q_\infty = (N - N_0)/(N_{\max} - N_0)$ therefore obeys $d\alpha/d\tau = f(\alpha)$ with a model function f that does not depend on the temperature programme.

Equation (11.1) is the biological counterpart of the master plots of thermal analysis, $d\alpha/d\theta = Af(\alpha)$ [44, 45]: thermograms recorded under different programmes differ in elapsed time, but divided by $\mu(T)$ and plotted against τ they coincide (Fig. 11.1a,b). Two uses follow. First, a single non-isothermal calorimetric run, for example a stepwise programme, can be used to test the square-root generalized time and to estimate b and $T_{\min}$ for a given product, in the same way as Ozawa’s method uses runs at several heating rates [41, 45]; a calorimeter that measures at several temperatures at once [57] gives $\mu(T)$ and hence E_{app} , Eq. (4.1), directly.

Second, calorimetry measures the biology that a temperature log cannot. Consider a sample of the lot incubated isothermally at T_{ref} until the heat flow exceeds a detection threshold, the detection-time

method used to screen for contamination [60]. If detection corresponds to a count N_{det} , Proposition 1 gives

$$\mu_{\text{lot}}(T_{\text{ref}}) t_{\text{det}} = \tau^*(h_0, N_0, N_{\text{det}}), \quad (11.2)$$

so the detection time combines the lot's own growth rate with its lag and starting load, which are precisely the quantities that dominate the error budget of Table 4.2. Because the limit is reached at $\tau = K$, and $K - \tau^*(N_{\text{det}})$ depends only weakly on h_0 and N_0 once the lag is over, a monitor that divides the logged integral by the measured detection time,

$$S_{\text{cal}} = \frac{1}{t_{\text{det}}} \int_0^t \mu_{\text{nom}}(T(s)) ds, \quad (11.3)$$

ranks trips by an estimate of τ/K in which the lot's rate scale and budget have been measured rather than assumed. As written, S_{cal} has units of h^{-1} ; multiplying by the constant $1/\mu_{\text{nom}}(T_{\text{ref}})$ gives the dimensionless ratio $\tau_{\text{nom}}(t)/[\mu_{\text{nom}}(T_{\text{ref}}) t_{\text{det}}]$. If the lot's rate law has the nominal shape, $\mu_{\text{lot}} = c \mu_{\text{nom}}$, Eq. (11.2) turns this ratio into $\tau_{\text{lot}}(t)/\tau^*$, the fraction of the lot's own budget up to detection that the trip has used. A constant factor does not change a ranking, so the AUC is the same for either form. What remains unmeasured is only the change of shape of $\mu(T)$ between T_{ref} and the storage temperature, which comes from the spread of T_{min} . Sect. 11.2 quantifies how much of the biological loss this recovers.

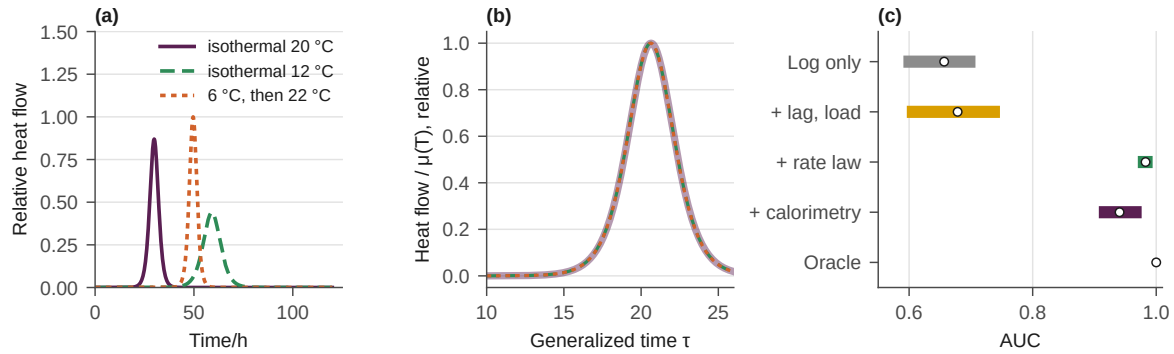


Figure 11.1 Calorimetry and the generalized time. (a) Simulated heat-flow curves of the same milk biology (baseline parameters, $h_0 = 3.8$, $N_0 = 50 \text{ CFU mL}^{-1}$) under isothermal programmes at 20 °C (solid) and 12 °C (dashed) and a step from 6 to 22 °C after 30 h (dotted), with the heat flow taken as proportional to the rate of cell formation. (b) The same curves divided by $\mu(T)$ and plotted against the generalized time τ ; they coincide (Proposition 6). (c) AUC over the 12 baseline settings when the log is combined with information on the biology: none, the true lag and starting load, the true rate law, an isothermal calorimetric detection time at 20 °C with 5 % measurement error, and all of it (oracle). Bars span the settings and circles mark the mean

11.2 Value of a calorimetric measurement of the lot

The loss budget says what is lost; Proposition 6 suggests how much of it a laboratory measurement could buy back. We regenerated the 12 baseline settings with a separate random stream (4 000 trips each) and scored every trip from the same DS18B20 record, sampled every 15 min, with four kinds of added information (Table 11.1, Fig. 11.1c). Knowing each trip's true lag and starting load raises the AUC by only 0.01–0.04. Knowing its true rate law (b and T_{min}) raises it by 0.26–0.40, to 0.97–0.99. This ordering is the one predicted by the first-order budget of Table 4.2, in which the rate parameters contribute 0.20 and 0.26 to the spread of $\ln \text{CSL}$ and lag and load only 0.08.

A calorimetric detection time combines rate, lag and load in one number, Eq. (11.2). We simulated an isothermal measurement of each trip's biology at $T_{\text{ref}} = 20 \text{ °C}$ to a detection level of 10^5 CFU mL^{-1} , with a lognormal measurement error of 5 %, and scored the trip by Eq. (11.3). The detection times have a median of 16 h (5th–95th percentile 11–24 h), short enough for a sample taken at dispatch. The

score reaches AUC 0.91–0.98, against 0.59–0.71 for the log alone; the gain exceeds zero in all 12 settings, and the measurement closes 69–94 % (mean 82 %) of the gap to the oracle. With a 15 % measurement error the AUC is 0.82–0.96 (mean 66 % of the gap), and with $T_{\text{ref}} = 10\text{ °C}$ it is 0.96–0.99 (mean 92 %). What remains, 0.02–0.09 AUC, is the change of shape of $\mu(T)$ between the reference and storage temperatures, which a single isothermal measurement cannot resolve.

These numbers are an upper bound in one respect: each simulated trip carries the biology of its own sample, whereas in practice a sample represents a lot whose cartons differ among themselves. They also assume a constant heat yield per cell and a fixed detection count. Within these assumptions the result is clear: one isothermal calorimetric measurement is worth an order of magnitude more ranking accuracy than any improvement of the thermometer.

Table 11.1 Value of a measurement of the biology (baseline run, DS18B20 record sampled every 15 min, 12 settings of 4 000 trips). Gap closed: share of the oracle-minus-log-only AUC difference recovered

Information added to the log	AUC range	Mean	Gap closed/%
None (log only)	0.591–0.707	0.657	0
True lag and starting load	0.596–0.747	0.679	1–14
True rate law (b , T_{min})	0.970–0.994	0.983	90–99
Calorimetric detection time, $T_{\text{ref}} = 20\text{ °C}$, 5 % error	0.907–0.976	0.941	69–94
Calorimetric detection time, $T_{\text{ref}} = 20\text{ °C}$, 15 % error	0.821–0.958	0.891	41–90
Calorimetric detection time, $T_{\text{ref}} = 10\text{ °C}$, 5 % error	0.957–0.989	0.973	86–97
All biology (oracle)	1.000	1.000	100

11.3 A role for calorimetry in the cold chain

Calorimetry cannot travel with a crate, and a temperature log cannot see the biology; the two are complementary. Proposition 6 shows that they measure the same generalized time from opposite sides: the log measures how much of it has elapsed, and an isothermal detection time measures, in the same units, how much of it the lot can afford. In the benchmark a single detection time at dispatch recovers most of the loss that no sensor can touch (Sect. 11.2). Isothermal microcalorimeters already detect contamination in hours [60] and follow the shelf life of fresh produce [61], and multi-temperature instruments give the temperature dependence of growth in one experiment [57]. Three questions must be answered before the idea is used: how well one sample represents a lot, how stable the heat yield per cell is in milk, and how the detection count depends on the instrument and the sample volume. Each can be answered by a calorimetric experiment, and the generalized-time framework says which quantities such an experiment should report.

Further results on real data and the monitor

The simulation answers what a log can tell. Real data and further tests say how far the answer can be trusted.

KEY POINTS

On real ComBase records, growth rates differ between independent studies with a coefficient of variation of 0.25, so the biology a log cannot see is real and of the size assumed. Tested on studies they have not seen, the square-root, Arrhenius, Q_{10} and Davey laws and a monotone neural network perform within a narrow band, with no clear winner. The monitor's Monte Carlo probability is as well calibrated as a logistic model trained on 4 000 labelled trips, it is stable from 80 plausible batches, and the loss to biology survives correlated biological parameters. The monitor's estimate of hours left is cautious but imprecise, for the same reason that the log cannot see the biology.

12.1 How much do real growth rates differ?

The benchmark assumes that the slope b of the square-root law varies between batches with a coefficient of variation of 0, 10 or 25 %. ComBase allows a direct check. Its *L. monocytogenes* records in milk with a stated growth rate between 0 and 15 °C come from many independent studies, each made with its own strains, media and methods. Fitting the square-root law with one shared $T_{\min}$ and a separate slope for each of the 11 studies with at least three records (87 records) gives a spread of $\ln b$ between studies of 0.25, while the scatter within a study is 0.26 on $\ln \mu$ (Fig. 12.1c). Real growth rates therefore differ between studies by about as much as the largest spread simulated. The biological variation that dominates the loss budget is not a modelling convenience; it is visible in the published data.

12.2 Which temperature law predicts unseen studies?

Chapter 8 compared four laws within the range of the data and when extrapolating into the cold. A stricter test is to predict a study that the law has never seen. Each law was fitted to all *Listeria* studies but one and used to predict the growth rates of the study left out, for each of the 20 studies in turn (101 records). For the spoilage flora, which come from a single source, five-fold cross-validation over records was used instead (51 records). Alongside the four classical laws we trained a monotone neural network, $\sqrt{\mu} = \sum_k e^{u_k} \text{softplus}(e^{v_k}(T - c_k)/10)$ with four units, which can take any increasing shape but cannot decrease with temperature; five independently initialised networks were averaged, and their spread estimates the network's own uncertainty (Fig. 12.1b).

Table 12.1 Accuracy factor A_f on held-out data (1 is perfect). *Listeria*: leave one study out; spoilage flora: five-fold cross-validation

Law	<i>Listeria</i> (20 studies)	Spoilage flora
Square-root	1.45	1.13
Arrhenius	1.37	1.15
Q_{10} (exponential)	1.37	1.15
Davey	1.39	1.14
Monotone neural network	1.42	1.14

All five laws fall within a narrow band (Table 12.1, Fig. 12.1a). For *Listeria* the Arrhenius and Q_{10} laws generalise slightly better across studies, for the spoilage flora the square-root law does, and the neural network sits in between. The differences between laws, about 0.1 in the accuracy factor, are small next to the differences between studies. Two conclusions follow. First, the choice of temperature law is a second-order question for a monitor; the first-order question is the biology of the batch, as the loss budget says. Second, a more flexible model of the rate, here a neural network, gains nothing when all it can learn from is maximum rates at a handful of temperatures. Learning more requires richer data: complete growth curves, including the lag, measured under changing temperatures and paired with information about each batch.

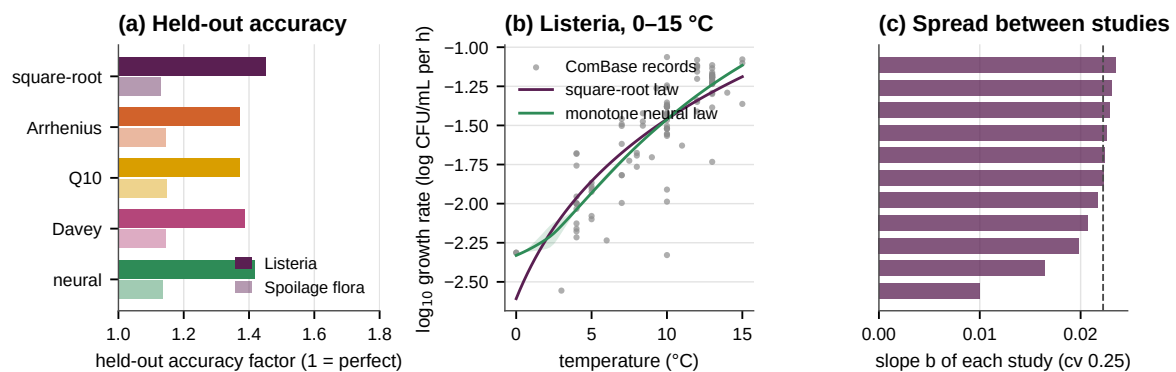


Figure 12.1 Growth laws on real data. (a) Held-out accuracy factor of five laws; dark bars *Listeria* (leave one study out), light bars spoilage flora (five-fold). (b) *L. monocytogenes* growth rates in milk from ComBase (points) with the square-root law and the monotone neural law; the band is ± 2 ensemble standard deviations. (c) Square-root slope b fitted separately to each study with at least three records, with the median (dashed)

12.3 Is a learned probability better?

Comparing the Monte Carlo probability with a single alarm is an easy test, because an alarm can only say 0 or 1. A fairer comparison is with a probability learned from data. We trained a logistic regression on six features of the log (degree-hours above 4 °C, hours above 5 and 8 °C, maximum and mean temperature, and the logarithm of CSL) on 4 000 simulated trips with known outcomes, and tested it on 4 000 independent trips, against the monitor's probability and the single alarm (insulated crate, DS18B20 every 15 min). The trained model reached an AUC of 0.669 and a calibration error of 0.026; the monitor's probability 0.670 and 0.026; the alarm 0.618 and 0.382 (Fig. 12.2a). The low-mass pouch gave the same picture (0.673 and 0.675). A probability built from the biology is therefore as good as one learned from thousands of labelled trips, and it needs no labels, which a real cold chain does not have.

12.4 How many plausible batches are needed?

The monitor's probability is the share of K plausible batches whose consumed shelf life reaches 1. With $K = 10$ the calibration error is 0.160, with 20 it is 0.131, with 40 0.054, and from 80 it stays between 0.014 and 0.028 (Fig. 12.2b). The 80 batches used in the benchmark are therefore enough, and the firmware's 64 are close to that.

12.5 Correlated biology

The benchmark draws the slope b and the minimum temperature $T_{\min}$ independently, but strains that grow fast in the cold may also have a low $T_{\min}$. With a correlation of -0.6 , 0 and $+0.6$ between $\ln b$ and $-T_{\min}$ (strain spread 25 %), the loss to biology was 0.35, 0.39 and 0.42 AUC (Fig. 12.2c). Correlation shifts the loss by a few hundredths but leaves it the dominant term.

12.6 How long will this batch last?

A dairy wants more than a probability: it wants to know how many hours of sale are left. After each simulated trip that stayed under the limit (1945 trips), the monitor estimated the hours the batch would last at 4 °C, typically (the median over plausible batches) and cautiously (the value nine batches in ten exceed). The true time was computed by continuing the reference model at 4 °C. The cautious figure was safe in 99.9 % of trips, more often than the nominal 90 %, so it is conservative. The typical figure was within 16 h of the truth in half the trips, against a median true time of 22 h, and its rank correlation with the truth was only 0.11. This is the loss budget seen from another side: the log fixes how much biological time has passed, but the time a batch can afford depends on biology the log cannot see. A measurement of the batch at loading is what would make this estimate precise.

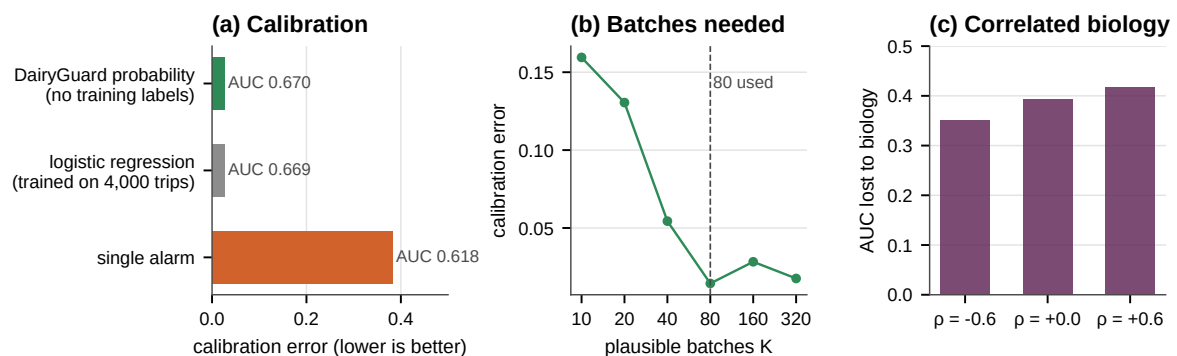


Figure 12.2 Further tests of the monitor (simulation, baseline world). (a) Calibration error and AUC of a single alarm, a logistic regression trained on 4 000 labelled trips, and the monitor's Monte Carlo probability. (b) Calibration error against the number of plausible batches. (c) AUC lost to biology when $\ln b$ and $-T_{\min}$ are correlated

12.7 What these results ask for

Each result in this chapter points the same way. Real growth rates vary between studies as much as the benchmark assumed; a flexible model gains nothing from rates alone; a learned probability is no better than one built from biology; and the hours left are uncertain because the batch is unknown. The next step is therefore not a cleverer formula for the temperature log but new information about the batch: complete growth curves of real milk under changing temperatures, measured together with simple tests of each batch at loading. A model that can use such data, keeping the generalized-time structure that this book establishes and learning only what it leaves unknown, is the natural successor to DairyGuard Kinetic.

Monitor design and software

What the results mean for anyone who builds, buys or trusts a cold-chain monitor.

KEY POINTS

The results translate into a short list of design rules. The estimator runs on a microcontroller, a browser demonstrator runs the same kernel, and the whole study is reproducible from one command.

13.1 Design rules

Table 13.1 converts the results into rules for a milk cold-chain monitor.

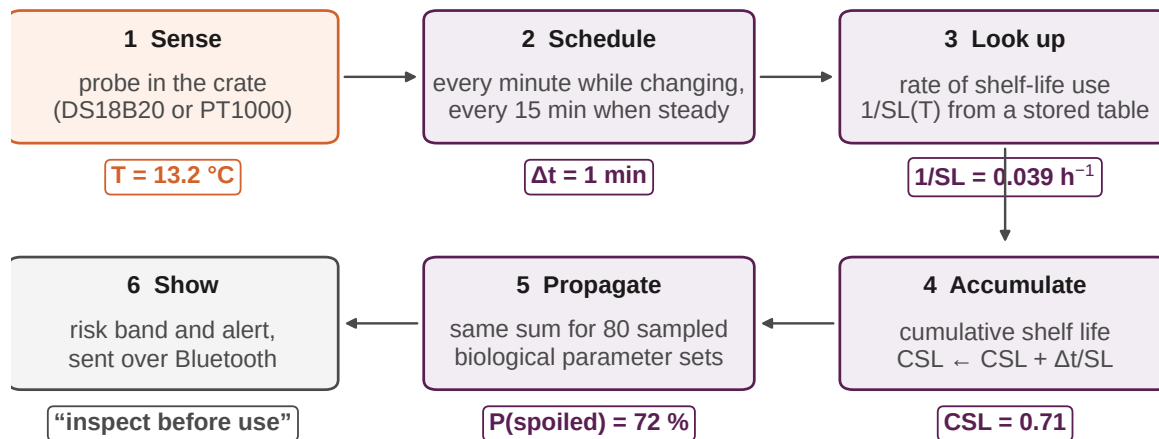
Table 13.1 Design rules for a milk cold-chain monitor

Decision	Rule and evidence
Sensor	A ± 0.5 °C digital probe suffices; a PT1000 gains at most 0.02 AUC
Sampling	15–60 min is enough unless excursions shorter than about 10 min are common
Score	Cumulative shelf life; degree-hours above 4 °C are an acceptable fallback (at most 0.04 AUC lower)
Rules to avoid	Maximum temperature and time above 8 °C; both rank trips clearly worse
Output	A probability, not an alarm (ECE 0.03–0.07 against 0.34–0.43)
Parameters	Fit to natural flora of the product; inoculated isolates overpredict by a factor of 1.62
Improvement	Measure the biology: starting load, lag and strain spread

13.2 Processing on the logger

Figure 13.1 shows how the estimator runs on a logger. A probe reading is taken on an adaptive schedule, the shelf-life rate $1/\text{SL}(T)$ is read from a stored table, Eq. (3.8) is updated, and the same sum is carried for 80 sampled biological parameter sets to produce a probability. Parameter fitting and the shelf-life table are prepared once, offline. The memory required is a table of a few thousand entries and 80 running sums, well within a low-cost microcontroller.

Example values under each block: 40 min after a truck door was left open. Steps 2–5 run on the logger.



Done once, offline: fit b , $T_{\min}$, lag and starting load to ComBase and published data; store the table (step 3) and the 80 parameter sets (step 5) on the logger

Figure 13.1 Processing on the logger, with example values 40 min after a truck door was left open. Steps 2–5 run on the device; parameter fitting and the shelf-life table are prepared offline

13.3 Software

The code is an installable Python package, `dairyguard`, with the research scripts kept separately in `experiments/`. The package is organised in layers: `kinetics` implements Eqs. (3.1)–(3.8); `simulate` and `benchmark` implement Chapters 5 and 6; `growthlaws` holds the secondary models of Chapter 8, the monotone neural law and the validation tools of Chapter 12; `products` holds the parameter sets of each product; and `monitor` is the estimator that a logger or a server runs as readings arrive. A command-line tool judges any logged trip,

```
dairyguard assess trip.csv -product milk
```

and prints the shelf life used, the probability that the limit has been crossed, the hours left at 4 °C (typical and cautious) and a plain verdict. Scripts in `generated/` write every number, table and figure of this book from the result files. Numerical work uses NumPy, SciPy and Matplotlib [62–64]. Twenty-seven automated tests (Table 13.2) run on every change. Parts of the code were developed with an AI coding assistant; all of it is covered by the tests and was checked by the author.

13.4 Firmware

The monitor also exists as portable C for microcontrollers (`firmware/`). The core keeps the consumed shelf life of the nominal batch and of 64 plausible batches drawn at start-up with a small deterministic random-number generator, using under 2 kB of memory and no dynamic allocation; an ESP32 sketch reads a DS18B20 probe every 15 min, updates the monitor, drives a green, amber or red light and reports over serial. A test compiles the core on a desktop and checks that it reproduces the Python monitor’s shelf life at every reading of a trip to a relative error below 2×10^{-4} .

Table 13.2 Automated tests

Group	What is checked
Kinetics (4)	Gompertz agrees with exponential growth; reference and estimator agree at constant temperature with equal parameters; the look-up table matches the closed form; shelf life falls with temperature
Validation (3)	Fitted parameters are plausible; the constant-4 °C and two-excursion times of Lin et al. are reproduced within 10 %
Statistics (6)	Fast AUC matches a reference implementation; known AUC values are recovered; the paired difference of identical scores is zero; a real improvement is detected; the probability rises with temperature; calibration of perfect and poor forecasts
Units and headlines (5)	ComBase rates are converted from decimal to natural logarithms before fitting; shelf life scales correctly with the rate; the headline numbers follow from the result files
Monitor and laws (6)	The monitor matches the closed form at constant temperature; a warmer trip is riskier; log reading and the command line work; the square-root fit recovers known parameters; the neural law can only rise with temperature
Firmware (2)	The C code for microcontrollers gives the same shelf life as Python at every reading; its probability and hours-left behave correctly
Parity (1)	The browser kernel matches the Python kernel

13.5 Browser demonstrator

A public demonstrator runs the same kernel in the browser (<https://copyninja07-07.github.io/dairyguard/explorer.html>); a parity test ensures that it reproduces the Python implementation. A user chooses or draws a trip, such as a shop door left open or a failed truck refrigerator, and sees the probability that the milk has passed the limit, the simulated growth in 60 cartons that share the same log, the fraction of shelf life used with its biological spread, simulated heat-flow curves of three batches with their detection times, and the comparison of the log alone with the log plus a calorimetric test. Every model in the demonstrator answers the same request, a temperature log and any tests at loading, so a future model can be added beside the present one and compared on the same trip. The demonstrator is software only.

Discussion

Where should the next rupee of effort go?

14.1 Where effort belongs

In every run, what the thermometer cannot observe costs about an order of magnitude more accuracy than any choice of sensor or sampling interval. Equation (3.9) explains why: a log measures the integral of the growth rate well, but the rate scale and the budget against which the integral is compared vary from carton to carton. A more accurate thermometer is therefore a poor way to improve a spoilage estimate, and a low-cost probe sampled every 15–60 min loses little under the assumptions made here. Effort is better spent on the biology: measuring the starting load or a proxy of handling history, learning the lag and strain spread of the product at hand, and reporting a probability instead of a verdict.

14.2 Consistency with earlier studies

Sensitivity analyses of milk spoilage rank storage temperature among the most influential inputs [35–38]; this study finds that biology dominates. The two answer different questions. When temperature is drawn from a distribution, it drives the spread of outcomes; once it is measured, the residual uncertainty is biological. The small gain of the growth-model score over degree-hours agrees with the long practical use of degree-hour and time–temperature integrators [9, 22], and the finding that parameters from inoculated isolates overpredict natural flora agrees with the practice of choosing a fast isolate as a worst case.

14.3 Hazard and spoilage

For spoilage flora a warm hour is costly but the baseline rate is already high; for *Listeria*, growth at 4 °C is slow but rises about 5.9-fold by 13 °C. The same benchmark and the same design rules hold for both, which supports treating temperature excursions as the common risk factor while recognising that a growth model says nothing about illness.

14.4 Limitations

- (i) The reference and the estimator both reduce the temperature history to an integral of the growth rate, so the benchmark measures how much accuracy is lost to the listed sources, not whether a growth-model score describes growth better than another model; the structure-risk run relaxes one structural assumption only.

- (ii) The strain spread, the starting load, the excursion statistics and the *Listeria* lag are assumed (Appendix A).
- (iii) Labels are constructed, with prevalence near 0.5 for spoilage and 0.08–0.38 for *Listeria*.
- (iv) ComBase records come from heterogeneous studies and report maximum rates, not curves, so lag and responses to temperature shifts are untested on them.
- (v) A single experiment validates growth under fluctuating temperature, and no new plate counts were made.
- (vi) Neither organism class represents spore-formers, which account for a large share of milk spoilage [10].
- (vii) For very small growth targets, such as the half-log *Listeria* label, the estimator's structure rather than biology limits accuracy.

Conclusions

Measure the milk, not just the air around it.

/ A temperature log can say much about how much growth the milk could have had, and little about how much it did.

In an open benchmark of 240 000 trips anchored to published and ComBase data, unobserved biology costs 0.27–0.40 AUC; a degree-hour count comes within 0.04 of a growth model; sensor quality and sampling interval matter little; and a Monte Carlo probability is far better calibrated than an alarm. The answers to the six research questions are therefore:

- RQ1. Biology the log cannot see is the dominant loss in every run.
- RQ2. A better sensor or faster sampling recovers at most about 0.02 AUC.
- RQ3. A growth-model score clearly beats threshold and maximum rules but only narrowly beats degree-hours.
- RQ4. A probability is far better calibrated than a single alarm.
- RQ5. Growth reads the temperature history only through a generalized time, which explains every loss. In simulation the square-root law is never outperformed by Arrhenius or Q_{10} rules; on real data from studies a law has not seen, the four laws are close, and the square-root law is the safest in the cold.
- RQ6. A single calorimetric detection time of the batch recovers 69–94 % of the loss to biology.

The results of Chapter 12 show what the next stage needs: richer data about each batch rather than a cleverer formula for the log. Four directions follow: measurements of real logger data from Indian dairy routes to replace the assumed excursion model; challenge tests with natural flora under fluctuating temperature, with new plate counts; calorimetric experiments on milk under stepwise temperature programmes, to test the master curve and measure detection times of real batches; and a hardware logger running the estimator.

Acknowledgements

The author thanks the ComBase consortium and the laboratories whose published growth records made this study possible.

Parameters and provenance

Every parameter of the simulation is listed in Table A.1 with its value and its source. Values marked as assumed are the main candidates for measurement.

Table A.1 Model parameters, their values and their provenance, as recorded in `data/params.csv`. SOURCED: taken from a cited study; FITTED: estimated from data in this work; ASSUMED: no direct source

Parameter	Value	Unit	Provenance
Tmin_mean	-7.7	degC	SOURCED: Neumeyer 1995 thesis Table 4.9 (-7.7 C, 68 growth curves, broth and milk). Literature mean 265.4 K = -7.75 C. Lin 2016 own data give -9.15 C
Tmin_sd	1.5	degC	ASSUMED. Neumeyer Table 5.1: ± 0.7 K is the 95% CI of the MEAN over 64 sources; their SD is 2.83 K but includes method noise. 1.5 K lies between the two and should be replaced by a measured strain SD
b_nominal	0.030	$\sqrt{1/h}/\text{degC}$ (ln basis)	SOURCED, two independent sources agree: Neumeyer 1995 Table 4.9 ($b=0.00465$ on $\sqrt{1/\text{generation time in min}}$) converts to 0.030; Lin et al. 2016 (growth rates 0.056/0.17/0.46 per h at 4/15/29 C, log10 basis) gives 0.027
b_cv	0 / 0.10 / 0.25	fraction	ASSUMED sensitivity axis, not measured
h0_truth	U(2.5,5.0)	dimensionless	SOURCED RANGE from Lin 2016 Table 2: lag x growth rate (ln basis) = 3.8, 5.0 and 3.0 at 4, 15, 29 C. Inoculated UHT milk, so natural contamination may differ
h0_estimator	3.8	dimensionless	SOURCED: Lin 2016 at 4 C (lag 29.5 h)
log10N0_truth	U(1.5,2.0)	log10 CFU/mL	ASSUMED. Martin et al. 2021 notes post-pasteurisation contamination is often very low (1 to 50 cells per 100 mL), so this is probably high and needs measuring
log10N0_estimator	1.7	log10 CFU/mL	ASSUMED
log10Nmax	9.0	log10 CFU/mL	ASSUMED (typical order for milk psychrotrophs)
Nlimit_A	3e4	CFU/mL	SOURCED as a number: FSSAI Appendix B v2 (04.11.2024) Table 2A, pasteurized milk APC $m=3 \times 10^4/\text{mL}$. A process-hygiene criterion used here as a quality marker. US PMO limit 20,000 CFU/mL (Martin 2021)
Nlimit_B	1e6	CFU/mL	SUPPORTED: about 10^6 CFU/mL is where quality deteriorates (Martin et al. 2021, citing Ranieri & Boor 2009; Lin 2016 citing Momani & Natsheh 2012)
label_shelf_life	83 (A) / 113 (B)	h	DESIGN: time for the estimator at nominal parameters held at 4 C to reach the limit; set for label balance before any method comparison
events_per_trip	Poisson, 1 per 144 h of horizon, max 8	count	ASSUMED
event_peak_ambient	12 / 14.3 / 17.5 / 30.7 (w .30/.25/.30/.15)	degC	ASSUMED: representative excursion peaks for transport and retail handling
event_duration	lognormal median 60 min, sigma 0.8, clip 10-480	min	ASSUMED

Parameter	Value	Unit	Provenance
product_time_constant	low-mass U(5,20) min; insulated U(1,4) h	h	ASSUMED; to be measured on real packs
base_temp	U(2.5,5.0)	degC	ASSUMED
shift_lag_fraction	0 (baseline) / 0.25 (structure-risk run)	fraction of h0	SOURCED VALUE, ASSUMED IMPLEMENTATION: Zwietering et al. 1994 (additional lag is one quarter of the lag at the new temperature, <i>Lactobacillus plantarum</i>); Lin 2016 saw extra lag after shifts in milk
sensor_DS18B20	bias U(-0.5,0.5), noise sd 0.05, res 0.0625	degC	accuracy ± 0.5 C from the manufacturer datasheet; noise ASSUMED
sensor_PT1000_ MAX31865	bias U(-0.15,0.15), noise sd 0.03, res 0.03125	degC	accuracy ± 0.15 C (IEC 60751 Class A tolerance at 0 C); noise ASSUMED
combase_fit	b 0.0256, Tmin -6.6 C (n=50 robust); log resid sd 0.16		EMPIRICAL: 51 ComBase records (Tasmanian Institute of Agriculture, pasteurised milk, 2.7-14.9 C); one record with an implausible temperature removed; Neumeier/Lin parameters overpredict these by 1.6x
runs_reported	baseline; extra lag; low starting load; ComBase-fitted milk flora; <i>Listeria</i>		all use seeds fixed in src/dairyguard/benchmark.py; 4000 trips per setting
lis_b	0.0235	ln basis	FITTED: 46 ComBase records of <i>Listeria monocytogenes</i> in milk, conditions not specified, 2-16 C; bootstrap 95% 0.0203-0.0265. Held-out check on 53 records: bias factor 1.02, accuracy factor 1.37 (split by condition, not by study)
lis_Tmin	-2.3	degC	FITTED: same fit; bootstrap 95% -4.2 to -0.9 C
lis_Tmin_sd	1.0	K	ASSUMED: of the order of the bootstrap standard error (about 0.8 K)
lis_h0	3.0 (estimator); U(2.0,4.5) (truth)	dimensionless	ASSUMED: no <i>Listeria</i> lag data were used; range borrowed from <i>Pseudomonas</i>
lis_N0	U(-1.25,-0.75) narrow; estimator -1.0	log10 CFU/mL	ASSUMED: about 0.06-0.18 cells per mL. Kept narrow on purpose: the +0.5 log label sits at -0.5, so a wider start range would put some trips above the label at time zero
lis_labels	+0.5 log10 (A) and +1 log10 (B) growth over starting load; horizons 100 h and 202 h		ASSUMED round-number labels, NOT regulatory limits (FSSAI Table 2B requires absence in 25 mL)

Data

Table B.1 lists the ComBase records of spoilage flora used in Chapter 7. The complete compilation of 373 milk records, including the 99 *Listeria* records with their conditions and sources, is in the repository (data/combase_milk_records.csv).

Table B.1 ComBase records of aerobic spoilage bacteria in pasteurised milk (Tasmanian Institute of Agriculture; 51 records). $\mu_{\max}$: maximum growth rate in decimal-logarithm units per hour

$T/^{\circ}\text{C}$	$\mu_{\max}/\text{h}^{-1}$	$T/^{\circ}\text{C}$	$\mu_{\max}/\text{h}^{-1}$	$T/^{\circ}\text{C}$	$\mu_{\max}/\text{h}^{-1}$
2.7	0.0171	6.8	0.0242	10.7	0.0866
3.4	0.0139	7.3	0.0613	11.1	0.0910
3.4	0.0264	7.4	0.0551	11.5	0.0852
4.2	0.0334	8.1	0.0605	11.5	0.1040
4.2	0.0313	8.2	0.0653	11.6	0.0948
4.3	0.0426	8.7	0.0615	12.1	0.0961
4.6	0.0391	8.8	0.0706	12.5	0.1070
4.9	0.0360	8.8	0.0676	12.6	0.0750
5.0	0.0396	9.3	0.0623	12.6	0.1270
5.2	0.0489	9.4	0.0826	12.7	0.1040
5.7	0.0492	9.6	0.0749	13.2	0.1330
5.8	0.0520	9.6	0.0658	13.3	0.1110
5.9	0.0431	9.9	0.0859	13.8	0.1230
5.9	0.0519	10.1	0.0804	13.9	0.0910
6.3	0.0512	10.2	0.0715	14.2	0.1310
6.6	0.0486	10.5	0.0813	14.8	0.1500
6.6	0.0562	10.7	0.0740	14.9	0.1200

Proofs

Proposition 1. Where $\mu > 0$, $\tau(t)$ is strictly increasing and $d\tau = \mu dt$. Dividing Eqs. (3.3)–(3.4) by μ gives $dy/d\tau = \frac{q}{1+q}(1 - e^{y-y_{\max}})$ and $dq/d\tau = q$, an autonomous system with no reference to T . Its solution (Y, Q) from (y_0, q_0) is unique (the right-hand side is Lipschitz), so $y(t) = Y(\tau(t))$ and $q(t) = q_0 e^{\tau(t)}$. Since Y is increasing while $y < y_{\max}$, N reaches N_L exactly when τ reaches $\tau^* = Y^{-1}(\ln N_L)$, which depends on $(h_0, y_0, y_{\max}, N_L)$ only. Intervals with $\mu = 0$ ($T \leq T_{\min}$) leave τ , y and q unchanged. $\square$

Proposition 2. From Eq. (3.6), $1/\text{SL}(T) = \mu(T)/K$, and the continuous limit of Eq. (3.8) is $\int_0^t \mu(T(s)) ds/K = \tau(t)/K$. $\square$

Corollary 1. Write $T - T_{\min} = (T_r - T_{\min}) + (T - T_r)$, square, multiply by b^2/K and integrate. $\square$

Proposition 6. With $P = Y_Q V dN/dt$ and $dN/dt = N dy/dt$, Eq. (3.3) gives $P = Y_Q V \mu(T) N \frac{q}{1+q} (1 - N/N_{\max})$. By Proposition 1, N and q are functions of τ alone, so $P/\mu = Y_Q V G(\tau)$. Integrating $P dt = Y_Q V dN$ gives Q , and $d\alpha/d\tau = G(\tau)/(N_{\max} - N_0)$, which is a function of α because α is strictly increasing in τ while $N < N_{\max}$. $\square$

Proposition 3. Let $g(t) = \mu(T(t))$. The error is $\frac{1}{K} \sum_i \int_{t_i}^{t_{i+1}} [g(t_i) - g(t)] dt$. (a) Each integrand is bounded by the oscillation of g on $[t_i, t_{i+1}]$, which is at most the variation V_i there, so the error is at most $\Delta t \sum_i V_i/K = \Delta t V/K$. (b) For a pulse of height $\Delta g = \mu(T_1) - \mu(T_0)$ starting at $s \in [t_i, t_{i+1}]$ and ending at $e \in [t_j, t_{j+1}]$, the start is undercounted by $(t_{i+1} - s)\Delta g$ and the end overcounted by $(t_{j+1} - e)\Delta g$; each lies in $[0, \Delta t \Delta g]$, so their difference lies in $(-\Delta t \Delta g, \Delta t \Delta g)$. (c) If the phases are independent and uniform, the two terms are independent $\mathcal{U}(0, \Delta t \Delta g)$ variables; their difference has mean 0 and variance $2(\Delta t \Delta g)^2/12$. $\square$

Proposition 4. With $\mu = b^2(T - T_{\min})^2$, $\partial \ln \text{CSL}/\partial \beta = 2 \int (T - T_{\min}) dt / \int (T - T_{\min})^2 dt$, which equals $2/(\bar{T} - T_{\min})$ for a steady trace. For noise, $E(T + \varepsilon - T_{\min})^2 = (T - T_{\min})^2 + \sigma_n^2$; the cross term $2(T - T_{\min})\varepsilon$ has mean zero and its sum over n independent readings has standard deviation proportional to $\sqrt{n}$, so its relative contribution falls as $1/\sqrt{n}$. Quantisation is handled in the same way with variance $\delta_q^2/12$. $\square$

Proposition 5. Take a spoiled trip (z_1, s_1) and an unspoiled trip (z_0, s_0) . With $u = z_1$, $v = -z_0$ and $w = s_1 - s_0 = u + v + e_1 - e_0$, $\text{AUC} = \Pr(w > 0 \mid u > 0, v > 0) = 4\Pr(u > 0, v > 0, w > 0)$. The vector (u, v, w) is zero-mean normal with $\text{corr}(u, v) = 0$ and $\text{corr}(u, w) = \text{corr}(v, w) = \sigma_z / \sqrt{2(\sigma_z^2 + \sigma_e^2)} = \rho$. The trivariate orthant probability is $\frac{1}{8} + \frac{1}{4\pi}(\arcsin 0 + 2\arcsin \rho)$, so $\text{AUC} = \frac{1}{2} + \frac{2}{\pi} \arcsin \rho$, which is Eq. (4.6). $\square$

Reproducing the results

All results are generated from fixed random seeds. With Python 3.10 or later, the packages in `requirements.txt` and a LaTeX installation:

Command	Effect
<code>make test</code>	run the 19 automated tests
<code>make runs</code>	run the five benchmark runs (about three hours on two cores)
<code>make validate</code>	reproduce the validation of Chapter 7
<code>make extensions</code>	run the growth-law study and the monitor studies of Chapter 12
<code>make generated</code>	write all numbers, tables and figures from the result files
<code>make report</code>	build this book
<code>make whitepaper</code>	build White Paper WP-01
<code>make all</code>	everything above

Every number quoted in the text is read from the result files by a script and written into the document as a macro; none is typed by hand.

References

- [1] Pérez-Rodríguez F, Valero A. Predictive microbiology in foods. New York: Springer; 2013. <https://doi.org/10.1007/978-1-4614-5520-2>
- [2] Tarlak F. The use of predictive microbiology for the prediction of the shelf life of food products. *Foods*. 2023;12:4461. <https://doi.org/10.3390/foods12244461>
- [3] Stavropoulou E, Bezirtzoglou E. Predictive modeling of microbial behavior in food. *Foods*. 2019;8:654. <https://doi.org/10.3390/foods8120654>
- [4] Badia-Melis R, Mc Carthy U, Ruiz-Garcia L, Garcia-Hierro J, Robla Villalba JI. New trends in cold chain monitoring applications – a review. *Food Control*. 2018;86:170–82. <https://doi.org/10.1016/j.foodcont.2017.11.022>
- [5] Mercier S, Villeneuve S, Mondor M, Uysal I. Time–temperature management along the food cold chain: a review of recent developments. *Compr Rev Food Sci Food Saf*. 2017;16:647–67. <https://doi.org/10.1111/1541-4337.12269>
- [6] Jedermann R, Nicometo M, Uysal I, et al. Reducing food losses by intelligent food logistics. *Philos Trans R Soc A*. 2014;372:20130302. <https://doi.org/10.1098/rsta.2013.0302>
- [7] Koutsoumanis K, Taoukis PS, Nychas G-JE. Development of a Safety Monitoring and Assurance System for chilled food products. *Int J Food Microbiol*. 2005;100:253–60. <https://doi.org/10.1016/j.ijfoodmicro.2004.10.024>
- [8] Göransson M, Nilsson F, Jevinger Å. Temperature performance and food shelf-life accuracy in cold food supply chains: insights from multiple field studies. *Food Control*. 2018;86:332–41. <https://www.sciencedirect.com/science/article/pii/S0956713517305145>
- [9] Kim J, Santos CA, Kim B-S, Kim J, Koo J. Estimation of real-time remaining shelf life using mean kinetic temperature. *LWT*. 2020;134:109968. <https://doi.org/10.1016/j.lwt.2020.109968>
- [10] Martin NH, Torres-Frenzel P, Wiedmann M. Invited review: controlling dairy product spoilage to reduce food loss and waste. *J Dairy Sci*. 2021;104:1251–61. <https://doi.org/10.3168/jds.2020-19130>
- [11] Ranieri ML, Boor KJ. Short communication: bacterial ecology of high-temperature, short-time pasteurized milk processed in the United States. *J Dairy Sci*. 2009;92:4833–40. <https://doi.org/10.3168/jds.2009-2181>
- [12] World Health Organization. WHO’s first ever global estimates of foodborne diseases find children under 5 account for almost one third of deaths. News release, 3 December 2015. <https://www.who.int/news/item/03-12-2015-who-s-first-ever-global-estimates-of-foodborne-diseases-find-children-under-5-account-for-almost-one-third-of-deaths>
- [13] Food Safety and Standards Authority of India. Appendix B: microbiological requirements, version 2 (4 November 2024), Tables 2A and 2B. [https://fssai.gov.in/upload/uploadfiles/files/Appendix%20B\(1\).pdf](https://fssai.gov.in/upload/uploadfiles/files/Appendix%20B(1).pdf)
- [14] Lin H, Shavezipur M, Yousef A, Maleky F. Prediction of growth of *Pseudomonas fluorescens* in milk during storage under fluctuating temperature. *J Dairy Sci*. 2016;99:1822–30. <https://doi.org/10.3168/jds.2015-10179>
- [15] Lima L, Pereira AI, Pintado M, Carochio M. Fresh dairy quality in the era of smart food systems: deterioration pathways, analytical technologies, and predictive modeling. *Foods*. 2026;15:3439. <https://doi.org/10.3390/foods15193439>
- [16] Gompertz B. 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. *Philos Trans R Soc Lond*. 1825;115:513–83. <https://doi.org/10.1098/rstl.1825.0026>

- [17] Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K. Modeling of the bacterial growth curve. *Appl Environ Microbiol.* 1990;56:1875–81. <https://doi.org/10.1128/aem.56.6.1875-1881.1990>
- [18] Baranyi J, Roberts TA, McClure P. A non-autonomous differential equation to model bacterial growth. *Food Microbiol.* 1993;10:43–59. <https://doi.org/10.1006/fmic.1993.1005>
- [19] Baranyi J, Roberts TA. A dynamic approach to predicting bacterial growth in food. *Int J Food Microbiol.* 1994;23:277–94. [https://doi.org/10.1016/0168-1605\(94\)90157-0](https://doi.org/10.1016/0168-1605(94)90157-0)
- [20] Ratkowsky DA, Olley J, McMeekin TA, Ball A. Relationship between temperature and growth rate of bacterial cultures. *J Bacteriol.* 1982;149:1–5. <https://doi.org/10.1128/jb.149.1.1-5.1982>
- [21] Rosso L, Lobry JR, Flandrois JP. An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. *J Theor Biol.* 1993;162:447–63. <https://doi.org/10.1006/jtbi.1993.1099>
- [22] Taoukis PS, Labuza TP. Applicability of time-temperature indicators as shelf life monitors of food products. *J Food Sci.* 1989;54:783–8. <https://doi.org/10.1111/j.1365-2621.1989.tb07882.x>
- [23] Baranyi J, Tamplin ML. ComBase: a common database on microbial responses to food environments. *J Food Prot.* 2004;67:1967–71. <https://doi.org/10.4315/0362-028X-67.9.1967>
- [24] ComBase. Database of microbial responses to food environments: growth records with milk as the food matrix. Accessed 19 Oct 2026. <https://www.combase.cc>
- [25] Neumeyer K. Modelling pseudomonad growth in milk and milk-based products. Research Master thesis. Hobart: University of Tasmania; 1995. <https://eprints.utas.edu.au/21013/>
- [26] Neumeyer K, Ross T, McMeekin TA. Development of a predictive model to describe the effects of temperature and water activity on the growth of spoilage pseudomonads. *Int J Food Microbiol.* 1997;38:45–54. [https://doi.org/10.1016/S0168-1605\(97\)00089-5](https://doi.org/10.1016/S0168-1605(97)00089-5)
- [27] Corradini MG, Peleg M. Estimating non-isothermal bacterial growth in foods from isothermal experimental data. *J Appl Microbiol.* 2005;99:187–200. <https://doi.org/10.1111/j.1365-2672.2005.02570.x>
- [28] Zwietering MH, de Wit JC, Cuppers HGAM, van 't Riet K. Modeling of bacterial growth with shifts in temperature. *Appl Environ Microbiol.* 1994;60:204–13. <https://doi.org/10.1128/aem.60.1.204-213.1994>
- [29] Koutsoumanis K. Predictive modeling of the shelf life of fish under nonisothermal conditions. *Appl Environ Microbiol.* 2001;67:1821–9. <https://doi.org/10.1128/AEM.67.4.1821-1829.2001>
- [30] Ndraha N, Hsiao H-I, Vlajic J, Yang M-F, Lin H-TV. Time-temperature abuse in the food cold chain: review of issues, challenges, and recommendations. *Food Control.* 2018;89:12–21. <https://doi.org/10.1016/j.foodcont.2018.01.027>
- [31] Zou Y, Wu J, Wang X, Morales K, Liu G, Manzardo A. An improved artificial neural network using multi-source data to estimate food temperature during multi-temperature delivery. *J Food Eng.* 2023;351:111518. <https://doi.org/10.1016/j.jfoodeng.2023.111518>
- [32] Müller WA, Kronbauer M, Ferreira SB, da Silva SB. Computational assessment of temperature failures in the cold supply chain: implications for beef shelf life. *Food Control.* 2026;187:112100. <https://doi.org/10.1016/j.foodcont.2026.112100>
- [33] Sofat A, Sodhi B. Beyond thresholds: a quality-aware decision intelligence framework for cold chain IoT systems. *arXiv preprint arXiv:2608.15082.* 2026. <https://arxiv.org/abs/2608.15082>
- [34] Nauta MJ. Separation of uncertainty and variability in quantitative microbial risk assessment models. *Int J Food Microbiol.* 2000;57:9–18. [https://doi.org/10.1016/S0168-1605\(00\)00225-7](https://doi.org/10.1016/S0168-1605(00)00225-7)
- [35] Lau S, Trmčić A, Martin NH, Wiedmann M, Murphy SI. Development of a Monte Carlo simulation model to predict pasteurized fluid milk spoilage due to post-pasteurization contamination with gram-negative bacteria. *J Dairy Sci.* 2022;105:1978–98. <https://doi.org/10.3168/jds.2021-21316>
- [36] Buehler AJ, Martin NH, Boor KJ, Wiedmann M. Psychrotolerant spore-former growth characterization for the development of a dairy spoilage predictive model. *J Dairy Sci.* 2018;101:6964–81. <https://doi.org/10.3168/jds.2018-14501>
- [37] Qian C, Murphy SI, Lott TT, Martin NH, Wiedmann M. Development and deployment of a supply-chain digital tool to predict fluid-milk spoilage due to psychrotolerant sporeformers. *J Dairy Sci.* 2023;106:8414–33. <https://doi.org/10.3168/jds.2023-23673>

- [38] Giannakourou MC, Taoukis PS. Holistic approach to the uncertainty in shelf life prediction of frozen foods at dynamic cold chain conditions. *Foods*. 2020;9:714. <https://doi.org/10.3390/foods9060714>
- [39] Arrhenius S. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Z Phys Chem*. 1889;4:226–48. <https://doi.org/10.1515/zpch-1889-0416>
- [40] Vyazovkin S, Burnham AK, Criado JM, Pérez-Maqueda LA, Popescu C, Sbirrazzuoli N. ICTAC Kinetics Committee recommendations for performing kinetic computations on thermal analysis data. *Thermochim Acta*. 2011;520:1–19. <https://doi.org/10.1016/j.tca.2011.03.034>
- [41] Ozawa T. A new method of analyzing thermogravimetric data. *Bull Chem Soc Jpn*. 1965;38:1881–6. <https://doi.org/10.1246/bcsj.38.1881>
- [42] Ozawa T. Kinetic analysis of derivative curves in thermal analysis. *J Therm Anal*. 1970;2:301–24. <https://doi.org/10.1007/BF01911411>
- [43] Ozawa T. Non-isothermal kinetics and generalized time. *Thermochim Acta*. 1986;100:109–18. [https://doi.org/10.1016/0040-6031\(86\)87053-8](https://doi.org/10.1016/0040-6031(86)87053-8)
- [44] Málek J, Koga N, Pérez-Maqueda LA, Criado JM. The Ozawa's generalized time concept and YZ-master plots as a convenient tool for kinetic analysis of complex processes. *J Therm Anal Calorim*. 2013;113:1437–46. <https://doi.org/10.1007/s10973-013-2939-0>
- [45] Koga N. Ozawa's kinetic method for analyzing thermoanalytical curves: history and theoretical fundamentals. *J Therm Anal Calorim*. 2013;113:1527–41. <https://doi.org/10.1007/s10973-012-2882-5>
- [46] Analog Devices (Maxim Integrated). DS18B20 programmable resolution 1-Wire digital thermometer. Product page and datasheet. <https://www.analog.com/en/DS18B20>
- [47] International Electrotechnical Commission. IEC 60751:2022 Industrial platinum resistance thermometers and platinum temperature sensors. Geneva: IEC; 2022. <https://webstore.iec.ch/en/publication/63753>
- [48] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- [49] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18:50–60. <https://doi.org/10.1214/aoms/1177730491>
- [50] Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993. <https://doi.org/10.1201/9780429246593>
- [51] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–45. <https://doi.org/10.2307/2531595>
- [52] Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78:1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2)
- [53] Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using Bayesian binning. *Proc AAAI Conf Artif Intell*. 2015;29:2901–7. <https://doi.org/10.1609/aaai.v29i1.9602>
- [54] Ross T. Indices for performance evaluation of predictive models in food microbiology. *J Appl Bacteriol*. 1996;81:501–8. <https://doi.org/10.1111/j.1365-2672.1996.tb03539.x>
- [55] Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19:716–23. <https://doi.org/10.1109/TAC.1974.1100705>
- [56] Hurvich CM, Tsai C-L. Regression and time series model selection in small samples. *Biometrika*. 1989;76:297–307. <https://doi.org/10.1093/biomet/76.2.297>
- [57] Wadsö L, Salamanca Y, Johansson S. Biological applications of a new isothermal calorimeter that simultaneously measures at four temperatures. *J Therm Anal Calorim*. 2011;104:119–26. <https://doi.org/10.1007/s10973-010-1140-y>
- [58] Fessas D, Schiraldi A. Isothermal calorimetry and microbial growth: beyond modeling. *J Therm Anal Calorim*. 2017;130:567–72. <https://doi.org/10.1007/s10973-017-6515-x>
- [59] Schiraldi A, Fessas D. Calorimetry and thermal analysis in food science. *J Therm Anal Calorim*. 2019;138:2721–32. <https://doi.org/10.1007/s10973-019-08166-z>
- [60] Fricke C, Harms H, Maskow T. How to speed up the detection of aerobic microbial contaminations by using isothermal microcalorimetry. *J Therm Anal Calorim*. 2020;142:1933–49. <https://doi.org/10.1007/s10973-020-09986-0>

- [61] Haman N, Signorelli M, Duce C, Franzetti L, Fessas D. Isothermal calorimetry protocols to monitor the shelf life and aftermarket follow-up of fresh cut vegetables. *J Therm Anal Calorim.* 2019;137:1673–80. <https://doi.org/10.1007/s10973-019-08064-4>
- [62] Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature.* 2020;585:357–62. <https://doi.org/10.1038/s41586-020-2649-2>
- [63] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>
- [64] Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng.* 2007;9:90–5. <https://doi.org/10.1109/MCSE.2007.55>

A temperature log can say much about how much growth the milk could have had, and little about how much it did.

240,000 simulated deliveries. One question: when a milk monitor is wrong, what is it wrong about? The answer is not the thermometer. It is the biology inside the carton, which no sensor can see, and which one calorimetric test of the batch could largely recover, as the simulation shows. Real milk is the next test.

Open code, open data, every number reproducible from one command.

