A New Probabilistic Approach in Predictive Microbiology

Inauguraldissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik und Informatik der Freien
Universität Berlin

vorgelegt von

Nadine Schöne

Berlin 2011

Erstgutachter: Prof. Dr. Alexander Bockmayr
Freie Universität Berlin

Zweitgutachter: Prof. Dr. Bernd Appel
Bundesinstitut für Risikobewertung

Tag der Disputation: 29.06.2011

Hiermit bestätige ich, dass diese Dissertation von mir selbstständig angefertigt wurde und alle von mir genutzten Hilfsmittel und Quellen angegeben wurden.

Alle wörtlich oder dem Sinne nach anderen Veröffentlichungen entnommenen Stellen wurden von mir kenntlich gemacht.

Ich versichere, daß die Dissertation weder an einer anderen wissenschaftlichen Einrichtung eingereicht noch abgelehnt worden ist.

Die Promotionsordnung des Fachbereichs Mathematik und Informatik der Freien Universität Berlin habe ich zur Kenntnis genommen.

Acknowledgments

It is an honour for me to thank my supervisors Prof. Dr. Alexander Bockmayr and Prof. Dr. Bernd Appel for their support and guidance. I would like to show my gratitude to my advisor Dr. Annemarie Käsbohrer, whose encouragement, guidance and support from the initial to the final level enabled me to finish this work.

This thesis would not have been possible without the encouragement of my parents and the rest of our family.

I want to thank my proofreaders and friends for their structural and moral support. I offer my regards to all of those who supported me in any respect during the completion of this thesis.

I want to thank the library staff in the Bundesinstitut für Risikobewertung (Federal Institute for Risk Assessment). I also thank the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) for financing the project PathoSafe (13N9550) this thesis derived from.

´Essentially, all models are wrong, but some are useful.´

[Box and Draper, 1987]

´Behind every simple is a huge tail of complicated.´

[Pratchett, 2010]

Contents

CHAPTER 1

Introduction

In this thesis a new probabilistic model for prediction of microbial growth (NPMPM) is presented. Today, almost all models in predictive microbiology are deterministic. A deterministic approach, like a combination of best case, worst case, and average case analysis, gives equal probabilities to all possible outcomes. Other probabilistic models are derived from deterministic models by substitution of deterministic variables with probability distributions.

The NPMPM is based on a new approach. It is empirical and not predicated on scientific regularities. It is a generic method. Therefore, the NPMPM can be adapted to other problems than bacterial growth, e.g. it can be used for prediction of metabolism from genome sequence data. In this thesis the NPMPM will predict the contamination of milk products with *Listeria* spp.

The NPMPM is fitted with experimental data. These data were taken from the ComBase, a database on microbial responses to the food environment [**Anonymous, a**]. The implementation was done in the R programming language [**R Development Core Team, 2010**] yielding a downloadable R package. The decision against a spreadsheet-based program like @Risk, Crystal Ball or Model Risk was based on the fact that these programs can not easily cope with large amounts of data and big numbers.

Predictive modelling is used to forecast the development of microorganisms in food or feed, using mathematical formulas. There is no detection strategy that ensures the absence of contamination during food processing [**Kennedy, 2008**], but it is possible to develop protection strategies. For this task the prognosis of growth and decay of pathogens or spoilage microorganisms is essential. In food processing, food safety and food characteristics have to be balanced. Predictive models can facilitate this optimisation process.

In case of an intentional or unintentional contamination, the consequence can be simulated depending on the foodstuff (matrix), microorganism, quantity of contamination (inoculum) and step in the supply chain. Getting the desired information solely by means of laboratory experiments is time and cost intensive. For instance if the goal is to investigate the behaviour of microorganisms depending on three factors, e.g., temperature, time, and pH value, for ten different values for each factor, $10 \cdot 10 \cdot 10 = 1000$ bacterial counts have to be measured. The default method is to redo the same 1000 experiments two more times to quantify the variability of the bacterial counts. Hence, 3000 experiments have to be done to get the essential

information.

In most manufacturing processes, parameters like time and temperature are not fixed values but vary over certain ranges; deterministic models are not able to take into account this variability. Probabilistic models estimate the probability distribution of the output values calculated. They are able to take into account the variability of all parameters. Several authors confirm that both variability and uncertainty are very important in predictive microbiology and that it is essential to use probabilistic models [**Delignette-Muller and Rosso, 2000**], [**Membre and Lambert, 2008**], [**Standaert et al., 2007**], [**Standaert et al., 2007**], [**Van Boekel, 2002**], [**Koutsoumanis, 2008**]. As a probabilistic approach provides far more realistic results, the use of probabilistic models in predictive microbiology supports a more profound risk assessment of the food or feed supply chain than deterministic models.

In our approach, variability and uncertainty present in the experimental data are directly taken into account. For every single experiment (one series of measured values) a primary model is fitted to the data points. The idea is to calculate a set of values. These are considered to be a sample from a probability distribution that reflects the distribution of bacterial counts under the given conditions. In this thesis a log-normal distribution of bacterial counts is assumed.

The model can be employed when little information as input or only few experimental data points are available, as well as when there is large amounts of high-quality data. When little information is available the uncertainty is high, and this is reflected in an expanded distribution of the output.

1.1. Organisation

This thesis is a multidisciplinary one. It provides the biological background needed to understand the complexity of bacterial growth and decay mechanisms. The mathematical knowledge essential to develop predictive models is also supplied.

The subject of foodborne diseases is covered in chapter 2. This chapter emphasises the importance of the efforts to achieve food safety. An overview over possible symptoms, causing agents, and associated foodstuff is given. Risk factors and the infectious dose are discussed. The chapter ends with examples of foodborne illnesses and verified outbreaks.

In order to model bacterial growth, one must gain a fundamental understanding of the underlying mechanisms. In chapter 3, the causing agents, mechanisms and kinetics of bacterial growth and decay are covered. First, the differences between bacteria that are related to growth and decay are introduced. Afterwards, the molecular bases of growth and decay are examined and compared. Then the influence of the environment is outlined.

Finally, the kinetics of growth and decay are discussed.

Chapter 4 provides an overview of the milk supply chain. After describing the composition of cows milk, the milk flora is described. Then the main process steps of dairy processing are outlined. Finally, the properties and production of certain milk products are highlighted. This knowledge about the milk supply chain is a prerequisite for simplification of production processes with the goal of modelling them.

Chapter 5 follows up the kinetics discussed in chapter 3. Different model types are outlined. Then existing approaches to model bacterial growth and decay are discussed. Afterwards, the difficulties in communication of probabilistic information are laid out.

The Monte Carlo method is introduced in chapter 6. After that, the model assumptions are outlined. Finally, the methods used for validation and sensitivity analysis are described.

Chapter 7 presents the `NPMPM`. An overview of the model structure is given. After discussing the theoretical background and a description of the data used in the simulations, the algorithm is introduced.

In chapter 8, the results are presented and discussed in form of validation and sensitivity analysis.

Finally, contribution and future work are introduced in chapter 9.

In the appendix, the terminology used, a brief description of the downloadable R package containing the model, and results of statistical tests can be found.

CHAPTER 2

Foodborne Diseases

Human illness that results from consumption of contaminated food or water is called a foodborne disease or foodborne illness. Only an occurrence of at least two cases of the same disease that are related to each other, is called an outbreak (*´Ausbruch´*). If the disease can be allocated to a special foodstuff or a food producing facility, the term foodborne disease outbreak is used. Some foodstuff-pathogen combinations are found more frequently than others, for example *Listeria monocytogenes* in milk products, or *Salmonella* spp. in chicken. Contamination may occur during production, but also during processing and preparation.

Foodborne disease is caused by bacteria and their toxins, or by viruses, parasites and chemicals. An infectious disease that can be transmitted from infected animals to humans is called zoonosis. In this thesis only zoonotic bacteria are taken into account, with a special focus on *Listeria* spp..

Today a contaminated food batch may reach a vast number of consumers, because of big production facilities and long-range distribution. The trend for ready-to-eat food which is consumed without heating increases the probability of foodborne illness, too. Therefore, numerous people may be involved in an outbreak. An example is the large outbreak caused by *Escherichia coli* O157:H7 that occurred in Japan in 1996 [**Mermin and Griffin, 1999**]. More than 7000 persons were involved. Foodborne illness was caused by contaminated white radish sprouts served through a centralised lunch program. The origin of contamination could not be determined.

Foodborne diseases can have an enormous economic impact due to loss of working hours, and costs of a recall of affected foodstuff. The consumer may avoid one or all products of a special manufacturer, an industry or even a country. In 1989 Chilean grapes were found to be contaminated with cyanide. Consequence was a recall of Chilean fruit in Canada and the United States of America. In succession more than 100 growers and shippers went bankrupt. The damage added up to several hundred million dollars [**Organization, 2008**].

Predictive microbiology can help to prevent foodborne disease outbreaks, because it helps to detect probable contamination sites, and to define monitoring points. It is a basic component of microbial risk assessment. In this chapter an overview of foodborne illnesses caused by bacteria is given. First, symptoms related to foodborne illness are described. Then the focus is laid on different causing agents, especially bacteria. In the next

section, foodstuff often associated with foodborne illness, and contamination routes are discussed. This is followed by an illumination of individual risk factors for infection and complications. After an overview of influences on the infectious dose of several agents, some examples for foodborne disease outbreaks reported in the literature are listed.

2.1. Symptoms and mortality

After an incubation period ranging from hours to days, first symptoms show up. The timespan depends on the agent. For toxins it is usually shorter than for bacteria. For some toxins it may be less than one hour. Bacterial infections have longer incubation periods; for example after ingestion of *Brucella* spp., first symptoms may show after one week up to several months. The incubation period of *Listeria monocytogenes* has a mean of approximately three weeks. The timespan differs in wide ranges, a timespan of three to 70 days was reported during outbreaks [**Gerner-Smidt and Whichard, 2007**].

Symptoms and mortality rate differ in wide ranges; dose-response relationship is complex and depends on several factors. The type of agent and foodstuff, and the quantity of bacteria ingested may influence the progression of disease. In unborn children, infants, the elderly and those with a compromised immune system (´YOPI´ - young, old, pregnant, immunocompromised) often progression is more severe.

Common symptoms of foodborne diseases are nausea, vomiting, abdominal cramps and diarrhea - the body tries to get rid of the agent. It is vital to prevent dehydration by oral replacement of water and electrolytes. Depending on their mechanism of action, ingestion of toxins can lead to impaired vision, numbness of parts of the body or paralysis. Bacteria may be transported with the bloodstream, this is referred to as septicaemia. A colonisation of other parts of the body than the intestine causes a variety of symptoms. The inflammatory response of the immune system to microbes in body tissues is called sepsis.

Mortality rate is measured as number of deaths in a population in a certain time span, scaled to the size of the population. To refer to the prevalence of a disease, or to the incidence rate, the term morbidity rate is used. Mortality rate of foodborne illnesses differs in wide ranges, depending on the agent, the amount ingested and the immune status of the host. A poor immune status increases the risk of severe illness and complications, and therewith the mortality rate, too. Medical treatment increases the probability of survival. Bacteria resistant to antibiotics are of special concern, because there may be no efficient therapy left.

2.2. Causing agents

Generally, there are three mechanisms leading to bacterial foodborne illnesses. Bacterial toxins ingested with foodstuff cause intoxication. Bacteria may colonise the intestine where they produce toxins leading to toxiinfection [**Weber, 2006**]. After conquering defence mechanisms of the intestine, bacteria may infect other parts of the body.

Microorganisms or toxins may cause foodborne diseases. Some toxins are organic, e.g. tetrodotoxin found in pufferfish. Others are inorganic, e.g. arsenic. Pathogenic species are found among viruses, bacteria, protists and metazoa (trematoda). Until now no pathogenic archaea were found [**Cavicchioli et al., 2003**]. [**Conway de Macario and Macarioa, 2009**] found hints that methane producing archaea aid pathogenic bacteria in colonisation of the host. In this thesis we will concentrate on vegetative cells of bacteria as causing agents.

Symptoms of an intoxication range from mild to severe, depending on toxin and relative amount swallowed. Botulinum toxin is a powerful neurotoxin produced by *Clostridium botulinum*. It is destroyed by heat, e.g. cooking [**Wein and Liu, 2005**]. *Staphylococcus aureus* produces, amongst others, an enterotoxin that causes gastrointestinal symptoms. This toxin is relatively resistant to heat. Aflatoxins are produced by certain fungi. They are carcinogenic and cause liver necrosis.

Colonisation of the intestine with toxin-producing bacteria not crossing the mucosa is called toxiinfection [**Weber, 2006**]. *Clostridium botulinum* is a fastidious anaerobe bacterium leading to spoilage and contamination with botulinum toxin of not proper sterilised canned food. Its spores are ubiquitous. Ingestion of spores of *Clostridium botulinum* by infants can cause a toxiinfection called visceral botulism. Unlike in adults, the intestine of infants provides favourable conditions for spore germination; bacteria can multiply and produce toxins. Food-borne botulism is caused by ingestion of even small amounts of botulinum toxin. The main symptom of botulism is paralysis, firstly affecting the eyes. Food-borne botulism without intensive medical treatment almost sure leads to respiratory failure.

Numerous species of bacteria can cause infections. Some bacteria remain in the intestine, others are more invasive and try to colonise different parts of the human body. Intracellular pathogens are shielded from the immune system, and from antibiotic treatment; they may cause repeated relapses. Vaccines for humans exist only for few pathogens that cause foodborne illness, e.g. for *Vibrio cholerae*.

Listeria monocytogenes is an ubiquitous zoonotic bacterium. It has been found in several kinds of raw foods, including meat, milk products, fruit and vegetables. Unlike most other foodborne pathogens, it is able to grow at refrigeration temperatures [**Kemmeren et al., 2006**]. Despite

being relatively rare [**Kemmeren et al., 2006**], Listeriosis is one of the most important causes of death from foodborne illness in industrialised countries [**Anonymous, 2008**]. *Listeria monocytogenes* causes severe illness in YOPI with a high mortality rate of 20-30% [**Anonymous, 2004**]. Infection of pregnant women may lead to miscarriages. But most individuals infected with *Listeria monocytogenes* do not develop obvious symptoms [**Kemmeren et al., 2006**]. In 2008 from 653 confirmed human cases of listeriosis in the EU where the outcome was known, 134 cases died, from which 87 cases were aged over 65 [**Anonymous, 2008**]. Overall notification rate was highest for elderly aged over 65 followed by children younger than five years old [**Anonymous, 2008**]. In laboratory experiments *Listeria monocytogenes* is often substituted by *Listeria innocua* which has similar behaviour but is non-pathogenic for humans.

2.3. Associated foodstuff

Almost all sorts of food can be linked to foodborne illness. A contamination of food with bacteria can occur at almost every point of the production chain. Unfortunately, pathogenic bacteria often do not cause spoilage in concentrations already sufficient for an infection. Hence, contaminated food usually can not be recognised by the consumer. The Centers for Disease Control and Prevention assigned one-third of foodborne disease outbreaks in the United States of America between 1998 and 2002 to production and processing or cross-contamination in the kitchen [**Gerner-Smidt and Whichard, 2007**]. This means that two-third of the outbreaks could not be clarified or were associated with other routes of contamination.

Pathogenic bacteria may be already present in raw food. Vegetables can be contaminated with faecal bacteria by means of organic fertilisers. Dust or water may transmit ubiquitous bacteria and spores [**Beuchat and Ryu, 1997**]. Infected animals may show no symptoms of disease, leading to contamination of meat, eggs and dairy products. For example *Brucella* spp. are excreted from the udder of infected cows together with milk.

A contamination can occur during processing, often due to poor hygiene. Insects like flies may transport infectious material on their bodies and transmit it onto food [**Buchanan, 1907**]. An infected food worker can disperse enteric pathogens directly to other persons, or indirectly from contact surfaces or food to person [**Todd et al., 2009**]. Healthy workers may transmit infectious material from other people, e.g. from ill children in their family, to surfaces or directly to food or to other people [**Todd et al., 2009**]. An overview on transmission of pathogens by food workers is given by Todd *et al.* in [**Todd et al., 2009**].

Fruits can be contaminated by insects, soil, dust or bird droppings. Persons who pick and handle them are a source of contamination, too.

Freezing kills some microorganisms on fruits [**Marth and Steele, 2001**]. Sugar concentrations in candied or glacéed fruits lower the a_w value below the level that permits microbial growth. In March 1997 frozen strawberries were associated with an outbreak of hepatitis A in Michigan [**Marth and Steele, 2001**].

Food that contains contaminated raw eggs can lead to salmonellosis when it is not proper handled and cooled. Eggs contaminated with *Salmonella* spp. normally do not show a significant increase in bacterial number in the first 10-20 days [**Marth and Steele, 2001**]. One reason is that there is only little iron available in the egg, and most bacteria require iron for growth [**Marth and Steele, 2001**]. During preparation raw eggs are mixed with other ingredients containing additional nutrients. This changes the conditions for bacterial growth. When the foodstuff is not kept at refrigeration temperatures, *Salmonella* may start to grow. Most foodborne pathogens can not grow at refrigeration temperatures [**Kemmeren et al., 2006**]. *Listeria monocytogenes* is an exception, it also tolerates relatively low pH and a_w values [**Kemmeren et al., 2006**].

There is an increasing trend for ready-to-eat food [**Havelaar et al., 2008**]. This type of food is usually not heated before consumption. So pathogens already present are not killed of. Improper cooling allows bacteria to grow. During longer storage the number of *Listeria monocytogenes* will increase even at refrigeration temperatures. Additionally, ready-to-eat food often has a large surface, because of being cut in convenient pieces. The big surface facilitates growth of bacteria, e.g. in ready-to-eat salad. Proper cooking and storage reduces the risk for foodborne illness.

2.4. Risk factors

The human gastro-intestinal tract has several defence mechanisms against virulent bacteria. Secretion of gastric acid and bile causes changes to a very low and a high pH, respectively. Digestive enzymes kill microorganisms by digestion, and mucosa and epithelium serve as barriers. Commensal microflora suppresses colonisation by other microorganisms. These defence mechanisms are weakened by certain medication that influences the amount of digestive secretion, e.g. H_2 receptor antagonists that raise the pH value in the stomach. Ingestion of foodstuff raising the pH together with pathogenic bacteria may lead to a higher risk of infection, too. Milk and alcohol can speed up the intake of certain toxins.

Innate and adaptive immune system provide a defence against non-symbiotic microorganisms. Hence, YOPI often have a higher risk of infection. Rate of complication and severity of disease is usually higher than in healthy people.

2.5. Oral infectious dose

The number of microorganisms that has to be ingested to cause a foodborne disease mainly depends on the immune status of the host, composition of the foodstuff, and the microorganism itself. It is called oral infectious dose. Acute toxicity of a toxin is usually specified by the median lethal dose (LD_{50}). The LD_{50} is the amount of toxin that kills 50% of individuals of a tested population after a certain time.

Dose-response relationships for human pathogens can be investigated by feeding studies or by analysis of foodborne outbreaks. Most experiments to determine the oral infectious dose in humans were done with healthy young males [**Kothary and Babu, 2001**]. Transfer of the results to high-risk groups is difficult. Additionally the pathogens are given in a non-food matrix like sodium bicarbonate or in milk [**Kothary and Babu, 2001**]. This may lead to a better survival of microorganisms in the stomach. The infectious dose of *Listeria monocytogenes* can only be estimated from foodborne outbreaks; febrile gastroenteritis in healthy hosts is apparently caused by a high dose of several million bacteria [**Gerner-Smidt and Whichard, 2007**]. For ethical reasons, experimental data from infection of humans are lacking [**Kothary and Babu, 2001**].

The infectious dose may vary for different strains of pathogens. Kothary and Babu reported an infectious dose of 10^4 cells of two particular strains of *Vibrio cholerae*, but no infection after ingestion of 10^6 cells of another strain [**Kothary and Babu, 2001**].

Generally there are other routes of infection than the oral route, with different infectious doses. Spores, viruses, or aerosols containing bacteria can be inhaled. Microorganisms may enter the body via lesions of skin or mucosa. Brucellosis is gained usually by the oral route but may be acquired from aerosols via the inhalatory route or via the conjunctiva.

2.6. Examples for foodborne outbreaks

In 2008 in the European Union 5332 foodborne outbreaks (both possible and verified) were reported to the EFSA [**Anonymous, 2008**]. In 74.2% of these outbreaks the causative agent is known. 14,001 human cases were verified, from which 20 died. Germany reported 20% of these outbreaks. The number of cases and hospitalisations varied considerably, depending on the pathogen causing the outbreak, as well as on the setting.

In 2009, an outbreak of listeriosis was caused by 'Quargel' cheese produced by an Austrian manufacturer. Nine people died and another 26 people became ill [**Fretz et al., 2010b**], [**Fretz et al., 2010a**]. The outbreak was caused by two different strains of *Listeria monocytogenes* due to consumption of the same white cheese from the same manufacturer [**Fretz et al., 2010a**]. The product was withdrawn from the market in

January 2010.

An example for cross-contamination as contributing factor is the national outbreak of *Salmonella* Enteritidis infection in the United States in 1994 [**Hennessy et al., 1996**]. A number of 224,000 persons developed gastroenteritis after consumption of ice cream. *Salmonella* Enteritidis was identified as causing agent, and was isolated from ice cream samples. Ice cream was manufactured from a pasteurised ice cream premix. This premix was transported to the producer in tanker trailers that had previously transported non-pasteurised liquid egg. Cleaning routines of the tankers were irregular, so most likely the foodborne outbreak was due to cross-contamination of ice cream premix with remaining egg containing *Salmonella* Enteritidis.

In 1985 a *Salmonella* Typhimurium outbreak in the United States was associated with consumption of properly pasteurised milk. About 23,000 people became ill. Despite the detection of *S.* Typhimurium in milk samples, the contamination site in the dairy plant could not be determined [**Weber, 2006**].

CHAPTER 3

Bacterial Growth and Decay

In order to model bacterial growth and decay it is useful to understand underlying structures and mechanisms. The border between growth and decay is not a sharp one. Vermeulen *et al.* found no straight cut-off, but a narrow transition zone both for monocultures and mixed cultures of strains of *Listeria monocytogenes* [**Vermeulen et al., 2007**]. At environmental conditions near the transition zone, growth was significantly slowed down.

In a medium containing unlimited nutrients, energy sources and room for growth, and in absence of predators, some bacteria can double in less than 20min at optimal temperatures. The medium in which bacteria are grown is also called matrix. Microbes grow until the supply of nutrient ceases or toxic by-products accumulate. The defensive response to such inhospitable conditions are cannibalism, synthesising of stress proteins, expression of starvation genes, and formation of dormant cells or spores [**Price and Sowers, 2004**]. The persistence of bacteria under different environmental conditions is called tenacity.

'All living matter is endowed with the capacity for multiplication and for coming to terms with its environment, and these fundamental properties are exhibited even by the simplest type of organism, the single cell, of which the bacterium is the commonest example.'

[**Dean and Sir Hinshelwood, 1966**], page 1

Despite the fact that bacteria are genetically and structurally simple, their division depends on a large number of events [**Slater and Schaechter, 1974**]. Hence, a lot of factors may influence growth and death kinetics. Due to evolutionary adaption, different genera and species, and also stems and biovars of the same species, show variant behaviour. The speed of growth and tenacity also depends on the history of the cells and current environmental conditions. Temperature, pH value, water content, availability of nutrients and other factors have dynamic and interactive effects. The presence of competitive flora, bactericides or preservatives may inhibit growth or enhance decay.

In chapter 2 the impact of foodborne diseases caused by bacteria is described. To prevent foodborne illness it is essential to understand the underlying mechanisms of bacterial growth and decay. In this chapter bacterial growth and death kinetics and the influencing factors are described. This provides knowledge for predictive modelling introduced in chapter 5. First, the kinetics are characterised. Then, cellular structures and their connection

to kinetics are discussed. After that, basic mechanisms of growth and death of bacterial cells are described. Then, the influence of environmental conditions on growth and death are considered. Finally, experimental data used in the `NPMPM` are characterised.

3.1. Counting bacteria

Usually, to obtain bacterial growth or death data, aliquots of a stock culture are used to inoculate a medium, which then is exposed to some treatment, e.g. heating. At different times samples are withdrawn from the treated culture, and the bacteria are counted. Various methods are available for quantification, some only count viable cells, others also count dead bacteria, yielding different quantities. Not all counting methods are able to distinguish different species in a mixed culture. In predictive microbiology mathematical models are fitted to experimental data. Databases with experimental data often do not contain information about the counting method.

The gold standard for counting bacteria is plating of defined volumes of a serial dilution on agar plates, and incubating these plates (plate count). Incubation temperature and sort of growth medium used for dilution and cultivation are chosen to provide optimal conditions for the bacteria. A defined volume of the matrix containing the bacteria is sampled and diluted with certain volumes of growth medium. Defined volumes of this serial dilution are evenly spread on agar plates. During incubation the bacteria able to grow will form colonies that (after some time) can be seen with the naked eye. At least for some dilutions the number of colonies is small enough to distinguish and count them. For high initial concentrations of bacteria, colonies form a bacterial lawn and can not be distinguished. For higher dilutions, there may be no colonies. The number of colonies is called the number of colony forming units, or short, cfu. Finally, the concentration of bacteria in the sample is calculated from cfu, dilution rate and sample volume.

For slow growing bacteria a primary culture may take a long time. For example *Mycobacterium avium* ssp. *paratuberculosis* has a generation time of 20h, under optimal conditions. A primary cultivation may take six months [**Rowe and Grant, 2006**]. Dormant cells (see section 3.3) or damaged bacteria that are still alive, but do not grow in the timespan of incubation, can not be detected. Other bacteria present may overgrow the species of interest. This can be prevented by addition of certain antibiotics.

A clear medium is clouded by increasing numbers of bacteria. This can be used for deduction of the number of bacteria from turbidity measurement in a photometer. A pure culture (a culture that contains only the bacteria of interest) in a clear liquid medium is required. Dormant, damaged, and dead cells enhance cloudiness, too. To calculate numbers of bacteria from turbidity measurements, the measured values must be compared with

turbidity measurements of defined counts of the bacteria, e.g. determined by plate count.

Bacterial cells can be counted directly. A visual counting of a pure culture can be done microscopically. If the matrix is not a clear liquid, a defined volume is sampled and diluted in a defined volume of growth medium or isotonic saline solution. Bacteria can be stained to distinguish living and dead cells. Then a defined volume of this suspension is filled into a counting chamber, and the bacteria are counted under a transmitted light microscope. This method is very time - and labour-intensive. Flow cytometry and coulter counter automatise the counting process. The bacteria have to be suspended in a special fluid. This passes by an electronic detector counting cells. A pure culture or a specific dye for the regarded bacterium is necessary.

A recently developed Raman spectroscope combines fluorescence microscopy and Raman spectroscopy for counting. First, the sample containing the bacteria is fixed on a membrane. Living cells are dyed with a fluorescent dye for distinction from other particles. Then the combined microscope-spectroscope focuses on fluorescing particles. Now a Raman spectrum is measured and aligned with a database to identify genus, species and stem of the bacterium. This method can simultaneously count cfus of different species.

The DNA content of a bacterial population is roughly proportional to its number of cells. This fact can be used for indirect counting of bacteria. From a defined sample volume bacterial DNA is extracted. A simple method is to perform an agarose gel electrophoresis of sample DNA together with a standard DNA probe. In the gel the DNA is separated into different bands. A comparison of the width of the bands of both probes is used to roughly estimate the number of bacteria present in the sample volume. Another possibility is to make a quantitative polymerase chain reaction (qPCR) of the sample DNA. A DNA standard from a dilution series of the same bacterium must be amplified, too. A linear regression of the DNA standard data yields a straight line from which the number of bacteria in the sample can be read off.

In this thesis the number of viable cells is referred to as cfu. The terms *cfu/ml* and *logcfu/ml* denominate the number of bacteria in one ml that was measured in an experiment, and the decadic logarithm of *cfu/ml*, respectively.

3.2. Kinetics

The term *kinetics* describes dynamics of bacterial growth and death, and velocity of changes in population size. For visualisation a curve can be drawn by plotting the decadic logarithm of the cfu against time. A general growth curve, beginning with inoculation, consists of four phases: lag phase with no changes in population size, log phase with exponential growth,

stationary phase with no changes, and death phase. An idealised growth curve is shown in figure 3.1.

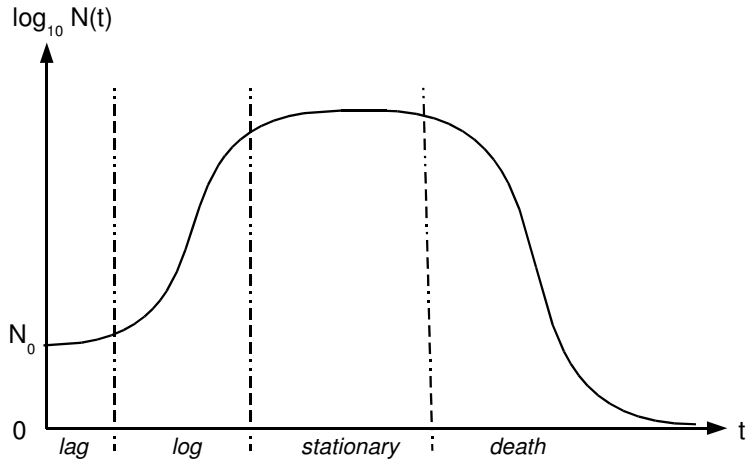


FIGURE 3.1. The four (possible) phases of a growth curve

Cultivation in a closed system with a fixed amount of nutrient culture medium is called a batch culture. The composition of the growth medium changes with time due to metabolic activity and increasing population size. Hence, the growth curve consists of all four phases, including the death phase when the nutrients are consumed or toxic metabolic substances reach a certain threshold. Cultivation in an open system with a permanent substitution of old culture medium with fresh one is called a continuous culture. The rate with which new medium is added controls the growth rate of the microorganisms. Hence, in a continuous culture cells permanently grow exponentially, but the population size is constant, because bacteria are constantly removed, and the death phase is missing.

During the lag phase bacteria adapt to the new environment. Proteins appropriate for the substrate and the environmental conditions are synthesised, e.g. enzymes. During the lag phase the RNA content of cells is six to twelve times higher than in other growth phases [**Fuchs, 2007**]. Cells in lag phase do not divide, and population size does not change. Fast changing environmental conditions inhibit the adaptation of cells and yield a longer lag phase. The age of the cells has also an influence, i.e., the time from inoculation of the culture on [**Stern and Frazier, 1941**] [**Dean and Sir Hinshelwood, 1966**]. Muñoz-Cuevas *et al.* report a dependency of the length of the lag phase on the magnitude of change between previous and actual conditions, and on current growth conditions [**Muñoz Cuevas et al., 2010**]. In a minimal medium it will be longer than in a full medium [**Dean and Sir Hinshelwood, 1966**]. It was shown in numerous studies that predictions of the length of the lag phase as a function of current growth conditions are usually inaccurate

[**Muñoz Cuevas et al., 2010**].

After adaption to a new environment, bacteria start to grow exponentially. This phase is called logarithmic phase (log phase), because a semi-logarithmic cfu vs. time plot often yields a straight line. Most times there is a more or less slow gradation between lag and log phase. The log phase is characterised by maximal autosynthetic activity [**Dean and Sir Hinshelwood, 1966**]. The time needed for doubling the number of cells is called (mean) generation time. During log phase the generation time is minimal. The frequency of procreation, i.e., the number of cleavages per time, is called growth rate. It depends on generation time, on the proliferative fraction present in the population, and on death rates [**Knolle, 1988**]. There are also correlations between individual generation times of sister cells, and between mother and daughter cells [**Takahashi, 1968**]. When the increase of cell mass is only influenced by growth rate and cell density, the population grows exponentially. Exponential growth may be followed by most living cells, but not by the total number of cells in a population. Additionally not all daughter cells have offspring themselves; some of them die. Growth may become linear due to influences of limiting factors [**Fuchs, 2007**].

The population size is a function of time, growth rate, and initial number of bacteria. The simplest assumption is exponential growth of all bacteria present in the population. With the number of cfu at time t, $N(t)$, initial bacterial count $N(0) = N_0$ and frequency β of growing by a factor of e, i.e.,

$$N(t) = N_0 \exp(\beta t). \tag{3.2.1}$$

Often the decadic logarithm of the number of bacteria is of interest, e.g. in the concept of D-values and z-values described in section 5.4. Formula 3.2.1 can be rewritten to

$$N(t) = N_0 10^{kt}, \tag{3.2.2}$$

with a growth constant k that describes the frequency of growing by factor 10. It holds true that $k = \frac{\ln 10}{\beta}$. Both equations 3.2.1 and 3.2.2 do only describe the log phase. β and k depend on the method of counting bacteria [**Baranyi et al., 1993a**].

There are many reasons for deviations from exponential growth. One single cell doubles its mass and divides into two identical daughter cells. In a population cells usually are at random points of the growth cycle. Cells can be synchronised in their generation periods. In a synchronised population a plot of logcfu/ml over time yields a series of distinct steps instead of a smooth curve.

Bacteria growing in a matrix with two different sources of carbohydrates may show similar kinetics to that shown in figure 3.2. First the cells consume the preferred carbon source. Use of the first nutrient prevents use of the second one. After exhaustion of the preferred carbon source, an adaption for metabolisation of the second carbohydrate is needed, e.g. a

second lag phase occurs. This phenomenon is called diauxie.

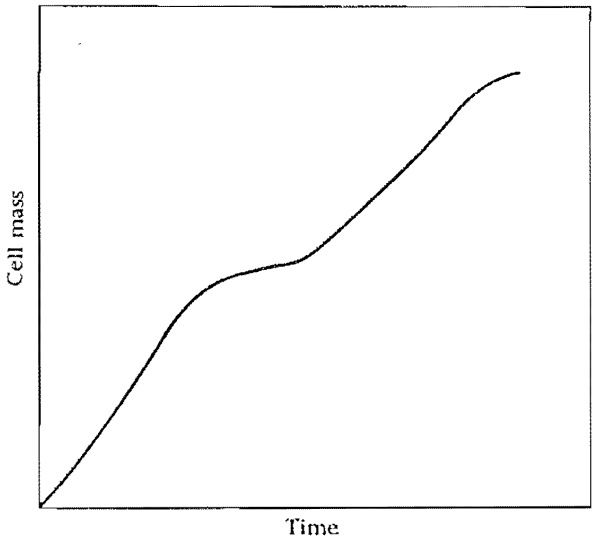


FIGURE 3.2. Diauxie (diphasic growth), from [**Dean and Sir Hinshelwood, 1966**]

During growth bacteria consume nutrients and produce metabolic substances that may be toxic. Additionally, the space between bacteria, e.g. the volume of matrix per bacterium, decreases. Growth slows down and part of the bacteria die. When the concentration of bacteria reaches a threshold, rate of growth and rate of death are equal. The population size then remains constant for some time. This phase of the growth curve is called stationary phase.

There may be up to 10^9 cfu per ml in a bacterial culture, depending on cell size and environmental conditions [**Anonymous, 2010b**]. Spheroidal bacteria, also called cocci, have diameters of about 1μm. Rhod shaped bacteria usually have a width of 0.2-1μm and a length of 1-10 μm. The volumes calculated from cell measurements are approximately $0.52\mu\text{m}^3$ for cocci and $0.03 - 7,85\mu\text{m}^3$ for rhods. Hence, in one ml there is space for about 1.9×10^9 cocci, 3.3×10^{10} of small rhods or 1.3×10^8 of big rhods - and no matrix providing nutrients in between them. There are bacterial species with smaller volumes, yielding slightly higher maximal quantities per ml. Nevertheless, the experimental upper bound is reasonable.

An enrichment of bacteria is possible, e.g. by centrifugation. The pellet obtained by centrifugation must be resuspended in a small amount of fresh medium, otherwise the cells are killed of after a short period. The concentration of the suspension should be at most 10^{12} cfu/ml (for small bacteria), higher concentrations lead to rapid death of the cells [**Anonymous, 2010b**]. A dilution by factor 1000, e.g. a contamination of a 1000litre-raw-milk tank with one litre of a concentrated suspension, yields

10^9 cfu/ml. Hence, the upper bound of 10^9 cfu/ml is also reasonable in case of an intentional contamination.

The stationary phase is followed by the death phase. When conditions worsen, death rate increases whilst growth rate decreases, and the population declines. The death phase is not really investigated [**Fuchs, 2007**]. To some extend autolysis may occur [**Fuchs, 2007**].

The simplest death curve consists of only one phase of exponential death, e.g. a straight line in a semi-logarithmic cfu vs. time plot. Exponential decay of a population can be described by equation 3.2.1 with negative β. Under the assumption that not all cells in the population are equally susceptible to a harmful factor and that the distribution of this factor can be uneven in the population, other kinetics are possible. For *Listeria monocytogenes* it was shown by Pagán [**Págan et al., 1997**] and Rowan [**Rowan and Anderson, 1994**]) that the shape of the survival curve could drastically change depending on pretreatment (history) of the cells.

In figure 3.3 the number of bacteria is plotted against time. For exponential decline we get curve I. If the cells may stand a damaging influence up to a certain degree, e.g. heat, and then immediately die off, we get a curve similar to curve II. Curves III and IV indicate a distribution of times of tolerance in the population, e.g. the presence of sub-populations.

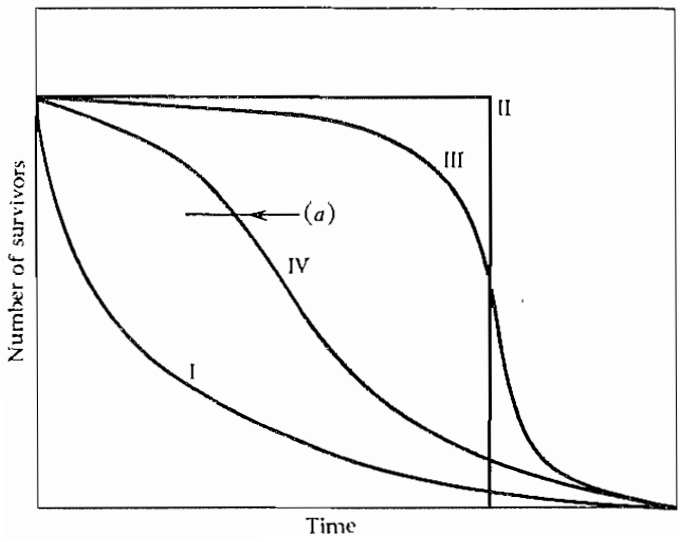


FIGURE 3.3. Death-rate of bacteria according to various hypotheses, from [**Dean and Sir Hinshelwood, 1966**]

A description of the modelling of these kinetics is given in chapter 5.

3.3. Cellular structures

There is a basic cell configuration found in all bacteria. Due to evolutionary adaptation to a specific ecological niche, there are alterations or additional structures. Hence, genera, species and even stems and biovars differ from each other. These differences are due to mutations, i.e., changes in the DNA sequence. Temporary alterations are done by up- and downregulation of genes already present in the cell. Unlike mutations this is fully reversible. Another type of adaption is the exchange of plasmids containing genes coding for resistance against antibiotics or for virulence factors. In the presence of the antibiotics, possession of the associated resistance genes is vital. In an environment free of antibiotics expression of resistance mechanisms and needed replication of the additional plasmid DNA is an evolutionary disadvantage. Therefore, after a certain time period without antibiotics, a bacterial population usually (and fortunately) is free of the resistance mechanisms again.

Bacterial cells demarcate themselves from the environment by use of a cell wall. Its main component is peptidoglycan (murein). It acts as a filter mechanism and provides a barrier against external noxes. Additionally, it gives structural support, and protects the cell from deformation due to osmotic pressure. There are two main types of murein cell wall that respond differently to a special type of staining, the so-called Gram stain. Bacteria with a high amount (up to 90%) of murein are stained by crystal violet. They are named Gram-positive. Gram-negative bacteria are not stained by crystal violet. Their cell wall has less murein (about 10%) and an additional outer membrane. There are some bacteria that yield a Gram-variable pattern. Tenacity of Gram-positive bacteria may differ from that of Gram-negative bacteria.

Pathogenic bacteria often synthesise a capsule that protects them from phagocytosis in the host. Usually, capsules consist of polysaccharides, whereas for example the capsule of *Bacillus anthracis* contains only glutamic acid. Encapsulated bacteria are comparatively resistant to thermal and physical shock [**Marth and Steele, 2001**]. Cells that lost the ability to express capsules are more likely damaged by freezing than encapsulated cells of the same strain [**Marth and Steele, 2001**]. The famous experiment of Griffith [**Griffith, 1928**] with strains of *Streptococcus pneumoniae* showed the dependency of its pathogenicity on a capsule.
Griffith took two different strains of *S. pneumoniae*. One of which was pathogenic, had a capsule and formed smooth colonies on an agar plate. The other one was not pathogenic, had no capsule and formed rough colonies. Mice infected with the smooth (S) strain developed pneumonia and died, mice infected with the rough (R) strain did not show symptoms at all. Injection of cells of the S strain killed by heat did not lead to illness, either. But when a mixture of killed S strain and alive R strain bacteria were injected the mice died, and viable cells of the S strain could be cultivated. This is the first experiment that was noticed to show transformation of bacteria.

A specific surface adhesion to host cells is guided by fimbria. These structures are also involved in biofilm formation. A biofilm is a community of bacterial cells embedded in a polymer matrix that sticks to an inert surface or a living tissue. Bacteria located in biofilms are protected from environmental conditions leading to a higher tenacity. In general in a biofilm different species of bacteria are present. They supplement each others abilities of gathering, processing and removing nutrients from their environment. Biofilm formation in the facility sometimes is a problem in food processing. It can cause high bacterial counts and spoilage in the end product [**Burgess et al., 2010**], [**Mariani et al., 2007**].

The genera *Clostridium* and *Bacillus* form spores as dormant bodies. Spores have a high tenacity and are very heat resistant. This is due to the thick spore wall, a low water content, production of calcium dipicolinate, and an almost stopped metabolism. Spores germinate under convenient environmental conditions. Sometimes a trigger is needed for germination, e.g. a high temperature for a short time. Spores can be stocked for many years. There are also super-dormant spores that require an increased signal for triggering spore germination compared to most spores in populations [**Ghosh and Setlow, 2009**].

There are non-spore forming bacteria that constitute dormant cells that are characterised by a drastically decreased metabolic activity, enhanced resistance to harmful factors and absence of cell division. Unlike formation of spores the dormant state is reversible, but dormant cells are often non-culturable. Transition from growth to non-culturability and vice versa are both poorly understood. An example is the latent form of *Mycobacterium tuberculosis* [**Shleeca et al., 2010**]. Dormant cells have a higher tenacity than viable cells.

Ubiquitous bacteria are adapted to a variety of changing conditions and can grow under a wide range of environmental conditions. Colonisation of certain ecological niches require a specialisation. Parasitic bacteria are adapted to the conditions inside their hosts, e.g. human pathogenic bacteria are all able to grow at 37°C. Facultative parasitic bacteria generally have a larger tenacity outside their hosts than obligate parasitic bacteria. In general all organisms specialised to live under certain environmental conditions have a very low tenacity under differing conditions. For example enzymes of bacteria able to grow at low temperatures have a relatively high turnover number and catalytic efficiency. Therefore, they are more susceptible to heat denaturation. Psychrophilic bacteria show growth at low temperatures between -5 and 20°C. *Listeria monocytogenes* is an example for a psychrophilic bacterium. The cells contain an unusually high proportion of branched chain fatty acids. Furthermore, cells grown at 6°C contain about one-third more total lipid than those grown at 30°C [**Mastronicolis et al., 1998**]. Thermophilic bacteria may grow at temperatures between 45°C and 80°C. Whilst the latter are able to form biofilms in processing facilities, e.g. grow on the surface of plate heat exchangers,

psychrophilic bacteria may lead to spoilage at refrigerator temperatures.

Speed of adaptation to new environmental conditions, depends on the physiological condition of the cells [**Vadasz and Vadasz, 2007**]. Hence, the history of the cells is of importance. If shortly after adaption to a new medium bacteria are retransferred to the old medium, a rapid reversion occurs and the adaption gets lost. Only after a longer period of growth in the new medium the adaption becomes stable and a retransfer to the old medium leads to necessity of new adaption [**Dean and Sir Hinshelwood, 1966**]. This kind of adaption is due to up- and downregulation of different genes.

An example for changing kinetics depending on the history of the cells is pictured in figure 3.4. *Aerobacter aerogenes* was grown with various concentrations of proflavine (2,8-di-amino-acridine). After an adaption period the cells were transferred to media with different concentrations of proflavine. The general shape of the growth curves in the new medium is similar, but the lag phase gets longer for certain concentrations, depending on training concentration.

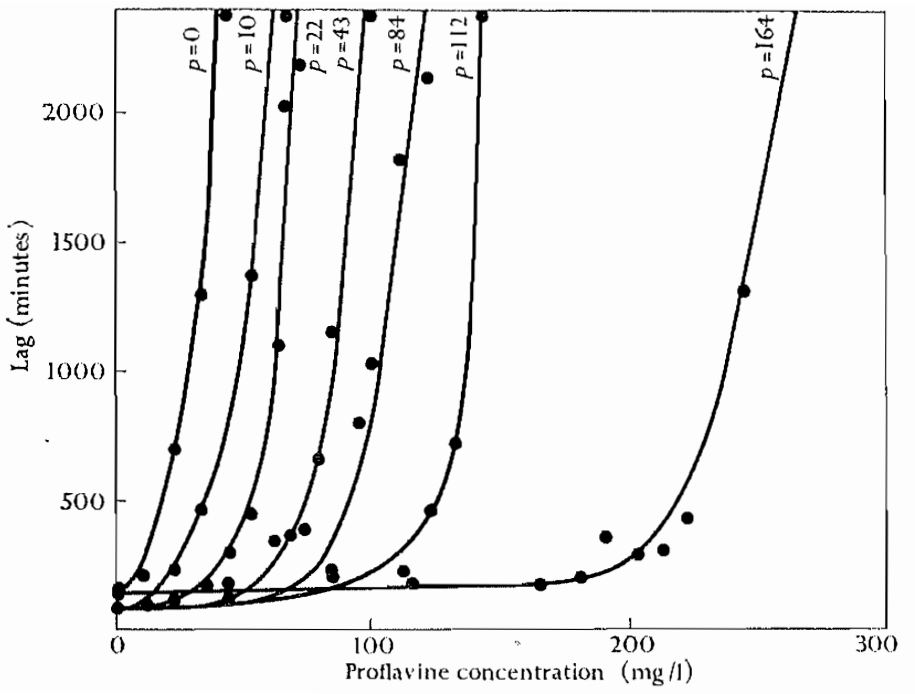


FIGURE 3.4. Family of lag-concentration curves for trained strains of *Aerobacter aerogenes* in proflavin - the value p indicates the concentration at which the cells were trained, from [**Dean and Sir Hinshelwood, 1966**]

The size of the initial number of bacteria may have an effect on duration of the lag phase, e.g. time needed for adaption to a new medium. Augustin *et*

al. showed that lag time of *Listeria monocytogenes* grown under suboptimal conditions is extended, when a small (and severely stressed) inoculum is used [**Augustin et al., 2000**]. Davis *et al.* observed an effect of inoculum size on culturability and isolation of soil bacteria [**Davis et al., 2005**].

Bacterial cells compete for available nutrients. In a mixed culture, some bacteria may produce metabolites that are toxic for other species. Additionally, the change in environmental conditions induced by one species can be fatal for other species. For example *Lactobacilli* metabolise lactose to lactic acid that lowers the pH value. Most bacteria are adapted to a neutral pH value of 7.0 and cease to grow or die off under acidic or alkaline conditions. Production of lactic acid is the main antimicrobial mechanism in *Lactobacilli* [**Zdolec et al., 2009**].

Some bacteria show an antimicrobial activity by production of bacteriocins. Bacteriocins have a protein or peptide component essential for their bactericidal function. They may inhibit DNA synthesis, promote DNA degradation or cause changes in membrane permeability [**Farkas-Himsley, 1980**]. Nisin is the only bacteriocin commercially available [**Tomé et al., 2005**].

3.4. Basic mechanisms of growth

Bacteria have offspring by binary fission. Unlike for eukaryotic cells, growth does not follow a cell cycle. Division begins with start of replication of DNA at the origin of replication (ori). Then the two DNA strands attach to opposite sites of the cell membrane. Now the cell starts to elongate, leading to separation of the DNA strands. After invagination of the cell membrane in between the DNA strands, the cell is divided into two almost identical daughter cells. A new daughter cell is approximately half the size of the mother cell.

The offspring of one cell is only almost identical, and growth of the two daughter cells is in most cases not synchronous. This shift in time will add up over the next generations. Therefore, a population of bacteria usually is not synchronised in growth. Hence, a simple up-scaling from kinetics of one bacterium to a population may not be sufficient. Synchronisation of generation times in a population of bacterial cells can be induced by shock treatment. After temporary interruption of growth, the cells may almost all be in the same phase. Synchronous cultures soon revert to a population with random phases of individual growth [**Dean and Sir Hinshelwood, 1966**].

3.5. Basic mechanisms of decay

Bacteria die off due to external factors, and there is no apoptosis like in *eukaryotes*. A direct damage of cellular structures leads to necessity of repair and therefore to delayed growth, but can also cause instantaneous death. Some antibiotics inhibit polymerases or build complexes with the DNA, e.g. actinomycin, or induce strand breaks, e.g. mitomycin C [**Fuchs, 2007**].

This leads to adverse effects on the synthesis of cellular components.

Every cell dies if it rans out of nutrients. This can be due to interruption of external supply, e.g. when there are no nutrients left in the matrix, or due to a breakdown of transport mechanisms into the cell. Blocking of metabolic pathways also leads to death. An example for competitive inhibition is the inhibition of cellular respiration by carbon monoxide that competes with oxygen for cytochrome oxidase [**Fuchs, 2007**]. Vital enzymes can be denaturated by heat, ethanol or high salt concentrations, and they may be damaged by heavy metals.

Bacteria die off when the border between the cell and its environment is destroyed. Detergents damage the cell membrane. Intracellular or extracellular ice crystals may cause destruction by punctuating it. Antibiotic substances may block synthesis of the cell wall, e.g. cephalosporine or vancomycin [**Fuchs, 2007**].

3.6. Environmental conditions

Environmental conditions have a big influence on the behaviour of microorganisms. The effect depends on the microorganism concerned, e.g. spores are much more resistant than vital cells. Sensitivity for harsh environmental conditions is not constant in time, even for one single bacterium [**Weber, 2006**]. Guillier *et al.* found that stress induced a big variability in lag times of *Listeria monocytogenes* [**Guillier et al., 2005**]. Stress in the meaning of unfavourable conditions like heat leads to alterations of gene expression patterns. Stress conditions imply damages of cell components that have to be repaired. When the DNA is affected the mutation rate may increase. Proteins with increased expression in stress conditions are called heat shock proteins (HSP). Many HSPs are chaperones.

Combination of different preservation factors (´hurdles´) in production processes is used to prepare safe, but minimal processed food. For example a microorganism that is relatively resistant to heat under otherwise optimal conditions may be killed off at lower temperatures when the pH value is changed. The concept of hurdle technology for food preservation was introduced by Leistner and Gorris in 1995 [**Leistner and Gorris, 1995**]. It is widely used to improve safety, stability, quality, and economic aspects of a variety of different foods [**Barbosa-Cánovas et al., 2005**].

All organotrophic bacteria need an organic carbon source for growth, which may be a sugar like glucose. Additionally essential elements such as magnesium, nitrogen, phosphorus, and sulfur must be available to allow the synthetisation of proteins and nucleic acids. If one nutritional component is missing, bacteria stop growing and eventually die. Oxygen is needed to metabolize the carbon source, and without carbon dioxide no gluconeogenesis can be done. The amount of CO_2 can affect both length of lag phase and growth rate [**Dean and Sir Hinshelwood, 1966**]. Parasitic bacteria

living in blood, tissue or intestines are adapted to a higher concentration of CO_2, and an absence of CO_2 may inhibit growth.

Several substances may slow down or even stop growth of microorganisms; the effect often depends on concentration. Many substances that stop growth at high concentrations will prevent fission of the cells in lower concentrations, yielding non-septate filaments in rhods [**Slater and Schaechter, 1974**]. Bacteriocins kill bacteria. Bacteriostatic agents interrupt growth until they are removed. Ethanol is a bacteriocidic substance that leads to coagulation of proteins. All detergents damage the cytoplasmatic membrane. There are some antibiotics that accumulate in the cytoplasmatic membrane leading to increased permeability. Enzymes are damaged by some heavy metals. Carbon monoxide competitively inhibits cellular respiration. Other substances inhibit the synthetisation of cellular components.

One bottleneck factor for growth is the availability of water. Water molecules have a three-dimensional tetrahedral structure, they are electric dipoles and form intermolecular hydrogen bonds with other water molecules. Because of its extraordinary physical and chemical properties water plays a crucial role in almost all cellular processes. Water content of a matrix comprises of water bound to charged molecules and free water. The content of free water is called a_w value or water activity. Only the free water is available for microbial, chemical, and enzymatic reactions [**Lewis and Heppell, 2000**]. Water can be bound by different hydrophilic molecules. Molecules with a high water binding capacity are called humectants. Examples are saccharose in jam or salt in ham; smoking decreases the water content, too.

A high a_w value lowers the heat resistance, probably because water increases the rate of heat-induced protein coagulation [**Marth and Steele, 2001**]. A low water activity inhibits bacterial growth. An a_w value between 0.6 and 0.9 largely protects food from microbial spoilage [**Belitz et al., 2009**]. *Listeria monocytogenes* is able grow from a water activity of at least 0.91 [**Anonymous, 2004**]. Usually the a_w value is calculated as $\frac{p}{p_s}$ with partial vapor pressure of food moisture p, and saturation vapor pressure of pure water p_s [**Lewis and Heppell, 2000**]. Calculation of the a_w value in intermediate moisture food can be done with Ross´ equation. It is computed as product of each components water activity, assuming that each a_w lowering component behaves independently [**Chirife, 1978**].

Growth media used in the laboratory are either liquid or solid. On a solid medium cells remain more or less in the same location. They use up the nutrients in the local area. Metabolites accumulate, and are only slowly distributed by diffusion. In liquid media bacteria get new nutrients due to facilitated distribution by convection, diffusion or agitation of the cultivation container. But metabolites that may be toxic for the cells are easily dispensed, too. Hence, rheological properties of the matrix may have an influence on kinetics.

Some bacteria are able to grow at a wide temperature range, e.g. *Listeria monocytogenes* is able to grow at 37°C, but can also multiply in food from a temperature of at least 0°C [**Anonymous, 2004**]. Other bacteria are very sensitive to deviations from their temperature optimum. At the beginning of the lag phase the cells are the most sensitive to heat; in the stationary phase they are the most insensitive to heat [**Weber, 2006**], [**Marth and Steele, 2001**]. Killing off by heat is based on coagulation of cellular proteins. Heat resistance of bacteria is influenced by several factors. Protein and colloidal particles in the matrix serve as a thermal buffer, like high concentrations of bacteria [**Marth and Steele, 2001**]. If bacteria are clumped together the resistance to heat also depends on clump size. Other factors are a_w value and pH value. In a solution of reducing sugars, e.g. lactose, the thermal resistance of at least some types of bacteria is increased [**Baumgartner, 1938**]. Marth and Steele [**Marth and Steele, 2001**] report a connection between the acidity of fruit and the heat treatment needed for preservation. Antibiotics and other inhibitory compounds often reduce the heat resistance [**Marth and Steele, 2001**]. For *Lactobacillus bulgaricus* Stern and Frazier [**Stern and Frazier, 1941**] found varying dependencies between inoculum size, growth rate, and maximum population size with changing temperature. *Listeria monocytogenes* may show a largely increased heat resistance after heat shock, depending on temperature, heat shock duration and heating menstruum [**Sergelidis and Abrahim, 2009**].

Low temperatures slow down the metabolisms, but there is no evidence of a minimum temperature [**Price and Sowers, 2004**] - apart from absolute zero, of course. Death due to cold is a result of the formation of ice crystals that destroy cellular structures and of denaturation of proteins and lipoproteins [**Marth and Steele, 2001**]. Formation of intracellular ice crystals can be reduced by increasing the osmolarity or reducing the water content of the cell plasma [**Fuchs, 2007**]. The amount of extracellular ice depends on the a_w value of the matrix [**Marth and Steele, 2001**]. Robertson [**Robinson, 1982**] states that intracellular ice is more harmful to microorganisms than extracellular ice. This is contradictory to findings of Ray and Speck [**Ray and Speck, 1973**] that during freezing, formation of extracellular ice was the principal cause of bacterial death. Some bacteria are able to produce HSPs that have a protective effect against freezing [**Marth and Steele, 2001**]. Stresses of freezing are more damaging to large than to small cells [**Marth and Steele, 2001**]. Hence, adaption of bacteria to low temperatures is done by reduction of cell size and capsular polysaccharide coat thickness, alterations of their fatty acid and phospholipid composition, or a decrease of the amount of cellular water. Therefore, their physiological condition affects the survival of microorganisms during freezing and frozen storage [**Marth and Steele, 2001**]. Other factors are the rate and method of freezing, and temperature, time, and condition of storage. Ray and Speck [**Ray and Speck, 1973**] report that cells in the stationary phase of growth resist freezing better than those in the logarithmic phase. Marth and Steele report the following order of survival of bacteria in frozen desserts, ranked from highest to

lowest survivability: bacterial spores, gram-positive bacteria, gram-negative bacteria [**Marth and Steele, 2001**]. Microbial toxins are resistant to freezing [**Marth and Steele, 2001**].

The lower bound for bacterial growth (including *Clostridium botulinum*) is a pH value of about 4.5 [**Lewis and Heppell, 2000**]. *Listeria monocytogenes* is able to grow in food from a pH value of at least 4.3 [**Anonymous, 2004**]. The pH value also has an influence on the effect of other environmental conditions. There are drugs that are much more effective in acidic solutions than in neutral ones. This is due to the positive ion of the drug as active species and the influence of the hydrogen ion on partial electric charges of the drug molecule [**Dean and Sir Hinshelwood, 1966**]. The pH value of a solution may depend on temperature [**Marth and Steele, 2001**]. Delignette-Muller *et al.* describe in [**Delignette-Muller, 1998**] a negative linear correlation between pH value and natural logarithm of the ratio of lag time and generation time. They suggests that this effect is due to physiological stress expanding the lag phase.

Other environmental conditions like pressure, radiation or shear forces affect viable cells and spores. Bacteria are killed by high pressures. A combination of increased pressure of at least 500 MPa and a temperature of 60°C kills even spores of the genus *Bacillus* [**Margosch et al., 2004**]. Radiation may kill off bacteria. UV-radiation is rich in wave length of 260nm that is absorbed by nucleic acids. Ionising rays lead to formation of hydroxyle radicals that damage macromolecules. There are some bacteria that are relatively resistant to radiation, e.g. *Deinococcus radiodurans* [**Battista et al., 1999**]. The effect of moderate shear forces on microorganisms is not investigated yet. Shear forces occur for example in the food chain when liquid foodstuff is pumped through pipes.

CHAPTER 4

The Milk Supply Chain

Milk products are an important protein source in many parts of the world. In some countries milk is produced and processed in small-scale farming. Other countries have dairy farms with up to a couple of thousand cows, and processing is done in dairy plants. Some countries export a large proportion of production, but in many countries with a large production the bigger part is consumed internally. In 2007 India was the major producer of milk with 114.4 million tons ECM ('*energy corrected milk*' with 4% fat and 3.3% protein content). In the same year Germany produced 29.4 million tons ECM [**et al., 2008**].

Dairy farming started almost 6000 years ago [**Bylund, 2003**]. Today the major part of dairy milk is produced from cattle (*Bos primigenius taurus*). Goats, sheep, buffaloes and camels play a major role in small-scale dairy farming. Other dairy animals are horses, donkeys, llamas, yaks, reindeer, water buffalos and moose. In this thesis the term milk without specification of origin refers to cows milk.

Only a cow that recently gave birth to a calf produces milk. In dairy farming the lactation period is kept about 305 days long. Six to eight weeks after calving the cow is mated again. Approximately eight weeks before the next calf is born the cow is caused to cease secreting. A calf needs approximately 1,000 litres of milk to grow up. The average cow milks more than 6,000 litres per lactation, but there are cows that reach more than 14,000 litres per lactation [**Bylund, 2003**]. Daily produce of milk depends on the point of time in lactation cycle, breed, feed and physical comfort. Lactating dairy cattle at present constitutes about 20% of the worlds domestic cattle population [**Marth and Steele, 2001**].

In chapter 2 we have a look at impact and the causes of foodborne diseases. Chapter 3 discusses the dynamics of bacterial growth and decay, showing that the modelling of these kinetics would be very complex without intense simplifications. In this chapter the diversity of milk products and their production processes is shown from different views. First, safety concepts in dairy manufacture are presented. Then the composition and properties of raw milk and dairy products are described. Milk flora, including the natural flora and contaminants, is discussed. After a description of the different production processes, the variety of milk products produced from raw milk by such processes is characterised.

4.1. Safety concepts

Milk and dairy products may be contaminated with pathogenic bacteria. Contamination can occur on the farm, at transport, and during processing or packaging. Infected cows may shed bacteria like *Brucella* spp. with milk. Bacteria are transferred from teats and milking equipment to raw milk. Ubiquitous microorganisms like *Listeria* spp. easily contaminate the milk. Poor hygiene may lead to a higher degree of contamination, e.g. with faecal bacteria like *Escherichia coli*. Composition of dairy products with other ingredients, e.g. fruit preparations, bares risk of contamination, too.

Contamination level changes during dairy processing. Centrifugation and filtration attenuate or accumulate small particles like spores and bacteria in resulting products, respectively. Process conditions like temperature, pH value, a_w value, salt and fat content may boost growth or lead to dying off of bacteria. Fat content has an influence on the tenacity of present bacteria. A recontamination is possible in most processes, e.g. hairline cracks in heat exchanger sections may form a bypass for raw milk and its flora into the already heat treated milk.

Until August 2007 the regulation *'Verordnung über Hygiene- und Qualitätsanforderungen an Milch und Erzeugnisse auf Milchbasis (Milchverordnung)'* was the legal framework of dairy production in Germany. In August 2007 the food hygiene regulation, *'Lebensmittelhygiene-Verordnung, LMHV'* replaced the milk regulation. The *'LMHV'* regulates specific topics of food hygiene on national level as required in the framework in the European Community and European Union. Some paragraphs regulate dairy production and processing. In dairy plants in Germany, applied process parameters are still those previously fixed in the *'Milchverordnung'* [**Anonymous, 2010a**].

In Germany there are several bacteriological quality criteria for different types of milk and dairy products. All forbid the presence of pathogenic bacteria.

In food processing safety of the final product is ensured by strict application of *´Hazard Analysis and Critical Control Points´* (HACCP) concepts. Their use is mandatory in the European Union according to Regulation (EU) Nr. 852/2004. The food industry uses HACCP concepts to systematically identify food safety hazards at all stages of food production and preparation processes. After critical control points are determined, monitoring and responses are established. Prediction of microbial growth is an important part of risk assessment techniques used in development of HACCP concepts. To model growth and decay of microorganisms during processing, both characteristics of processing and microorganisms must be known. The latter is described in chapter 3.

TABLE 4.1. Composition of raw milk, adapted from [**Bylund, 2003**]

Component	Minimum	Average	Maximum
Water	85,5	87,5	89,5
Total solids	10,5	13,0	14,5
Fat	2,5	3,9	6,0
Protein	2,9	3,4	5,0
Lactose	3,6	4,8	5,5
Minerals	0,6	0,8	0,9

4.2. Composition, physical and chemical properties of raw milk

Main components of raw milk are water, fat, lactose and proteins. Additionally, milk contains different enzymes, e.g. alkaline phosphatase, peroxidase and lactoperoxidase, and small amounts of minerals, pigments, vitamins, and phospholipids. The general composition of raw milk is listed in table 4.1. It depends on different factors, e.g. fat content is influenced by the breed of the cow and stage of lactation, feed, and season [**Sienkiewicz and Kirst, 2006**]. In the 1950s about 200 milk components were known. Today more than 2000 substances are identified [**Töpel, 2004**]. Natural fluctuations in milk composition are of technological and economical interest.

Milk is an oil-in-water emulsion. The fat droplets have an average diameter of 3-4μm. They are covered with a thin membrane. One ml milk contains 15 billion fat globules [**Bylund, 2003**]. Usually, in processing fat droplets are downsized by homogenisation to prevent the two phases from separating during storage. At a temperature of 37°C milk fat is a liquid [**Bylund, 2003**], [**Fox and McSweeney, 1998**]. Sugar, salts and vitamins are dissolved in the water phase; proteins are dispersed.

Milk contains lactose, a disaccharid of glucose and galactose. Infants are able to digest lactose by means of the enzyme lactase. Naturally expression of this enzyme ceases in young children, the resulting lactose intolerance shows up as indigestion after ingestion of lactose. In societies where consumption of dairy products is traditional, lactase is still produced by adults. In northern Europe about 85% of the population shows lactase persistence [**Swagerty et al., 2002**].

Coprecipitation of milk proteins due to acidification or addition of rennet is of importance in processing of some dairy products. Raw milk has a pH value of 6.5-6.7 [**Bylund, 2003**]. At a pH of 4.6 and a temperature of 30°C about 80% of the proteins fall out. These proteins are called casein, and consist of colloidal aggregates with a diameter between 50 and 600nm. The other 20% of milkprotein are very small molecules and are called non-casein nitrogen, '*Molkenproteine*' [**Fox and McSweeney, 1998**].

Usually milk has a density of 1.028 - 1.038 [**Bylund, 2003**]. Milk and blood are isotonic, with lactose accounting for 46% of osmotic pressure [**Bylund, 2003**]. The freezing point ranges between -0.54°C and -0.59°C [**Bylund, 2003**]. Precipitation of phosphates due to heating and watering both elevate the freezing point. In the dairy the freezing point of raw milk is determined to detect a fraudulent addition of water. Concentrations of protein and fat are measured, too. Payment per volume for the dairy farmer depends on quality, e.g. a high fat content, of milk [**Bylund, 2003**].

4.3. Milk flora

Milk flora consists of different microorganisms. Acidifying bacteria like *lactobacilli* lead to fermentation of milk. Certain acidifying bacteria are cultivated and used for specific processing of fermented milk products. Spoilage bacteria, yeasts and molds produce unwanted flavours and changes of texture. Some yeasts are used for fermentation, e.g. in kefir production. Bacteriophages may kill off bacteria necessary in cheese production. Pathogenic bacteria and viruses can cause foodborne illness. Spore forming bacteria are of special interesting in cheese production, because spores are very heat resistant and can cause late blowing in hard and semi-hard cheese. In the dairy, concentrations of bacteria and somatic cells in the raw milk is determined. It is one factor in definition of raw milk price.

Bacteria found in raw milk are of two different origins: the cow itself and the environment. A healthy cow produces nearly sterile milk. A mastitis will lead to a bigger bacterial load in the milk. Additionally, somatic cells can be found. Often the texture of milk changes, too. Milk from cows diagnosed with mastitis has to be discarded. Some pathogenic bacteria are secreted from the udder, e.g. *Brucella* spp. and *Mycobacterium avium* spp. *paratuberculosis*. They usually lead to symptoms in cows, but may not influence the texture of milk or lead to spoilage in the processed product.

Most bacteria present in raw milk originate from the environment. Some are of faecal origin, like *Escherichia coli*. Others are ubiquitous, e.g. *Bacillus cereus*, or may reside in milking equipment, e.g. *Listeria* spp.. Bacterial load of raw milk highly depends on hygiene standards on the dairy farm. One ml of raw milk may contain between a few thousand and several million bacteria [**Bylund, 2003**].

Contamination of milk with bacteriophages does not cause spoilage or foodborne illness, but may lead to serious trouble in production of cheese and fermented milk products. Natural milk flora plays an essential role in production of raw milk cheese (but not in other milk products). In production of cheese made from heated milk and in production of fermented milk products, milk is inoculated with starter cultures to compensate for the killed natural flora. Presence of bacteriophages or antibiotics leads to serious problems in fermentation steps, because some of the helpful bacteria do not grow or are killed off. Bacteriophages are ubiquitous and

much smaller than bacteria. They are not removed by microfiltration or bactofugation [**Bartel et al., 1996**], but can be inactivated by harsh heat treatment. Milk from cows treated with antibiotics is not marketable, nevertheless, sometimes there are antibiotic residues in dairy milk.

A natural antimicrobial system based on the enzyme lactoperoxidase retards spoilage by both Gram-negative and Gram-positive bacteria in raw milk. The lactoperoxidase system is more active in raw milk after storage for three to four days [**Lewis, 1999**]. A study by Marks, Grandison and Lewis ([**Marks et al., 2001**]) showed that the lactoperoxidase system retains most of its activity at normal pasteurisation temperature of 72°C, but is deactivated at heating temperatures of 80°C. Other bacteriostatic compounds present in raw milk are destroyed during pasteurisation [**Sienkiewicz and Kirst, 2006**].

Quality of raw milk, processing conditions like temperature and time, and storage temperature are main control points for refrigerated dairy products [**Lewis, 1999**]. Psychotrophic bacteria can even grow at refrigerator temperature. After an adaption that usually takes 48-72h, bacteria start to grow exponentially. Some thermophilic bacteria are able to form biofilms on stainless steel surfaces in areas related to heat treatment of milk, e.g. *Geobacillus* spp. [**Burgess et al., 2010**]. They can cause high bacterial counts and spoilage in the end product, even if the raw milk was of high quality and contained only a few bacteria. Microbiological quality of water used for washing or in brines is critical in manufacture of a safe and stable product, too. Pasteurised milk contains a spoilage flora totally different to that found in raw milk. It mainly consists of sporeformers and thermoduric bacteria, possibly accompanied by post-pasteurisation contaminants [**Marks et al., 2001**].

Several pathogenic bacteria need an increased partial pressure of CO_2 for growth. In milk and milk products CO_2 content only increases as a consequence of metabolic activity. Fermentation leads to a decrease of pH value and therefore to diminished growth conditions for most pathogens. The following zoonotic, human pathogenic or toxin producing bacteria are related to milk and milk products (this list makes no claim to be complete): *Bacillus cereus*, *Brucella* spp., *Campylobacter* spp., *Clostridium perfringens*, *Corynebacterium diphteriae*, *Escherichia coli*, *Listeria monocytogenes*, *Mycobacterium avium* spp. *paratuberculosis*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Salmonella* spp., *Staphylococcus aureus* and *Yersinia enterocolitica*.

4.4. Production processes

There is a great variety of processed milk products, see also section 4.5. Main process steps in manufacture of dairy products are heating and cooling, separation processes, homogenisation, storage and ripening, fermentation, enzymatic coagulation, composition, and packaging. In Germany

food processing does not contain the use of irradiation because it is not acceptable to the public. Additionally, it is expensive and some toxins are only destroyed by high doses of irradiation [**Havelaar et al., 2008**]. In the dairy plant, milk is always centrifuged and heated. For some products the milk is homogenised. All processes may have an influence on chemical and physical properties of the resulting milk product. A schematic presentation of relations between milk products in dairy processing is given in figure 4.1.

During production processes a variety of different factors affect milk components and milk flora, causing wanted and unwanted, sometimes reversible effects. Protein denaturation is one effect of heat. It is unwanted in milk proteins, but wanted in bacterial cellular proteins, because this kills off bacteria. When milk is pumped through narrow pipelines or pressed through small slits, shear forces occur. This leads amongst other things to smaller fat droplets, a wanted effect in homogenisation. The smaller fat globules together have a bigger surface than the original ones, and therefore are no longer totally included in a membrane. Milk fat now is exposed to lipolytic enzymes naturally present in milk; this changes the taste of milk. Under influence of light, milk acquires an unwanted taste (´*light flavour tainting*´) due to transformation of methionine to methional.

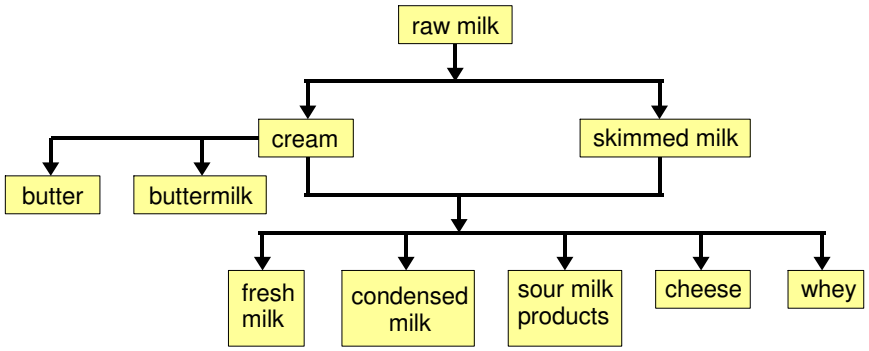


FIGURE 4.1. Schematic presentation of relations between milk products in dairy processing, from [**Belitz et al., 2009**]

Most processes in modern dairy plants are continuous-flow processes, but also in batch processes process parameters always follow probability distribution´s, and are not constants. There is only little specific information about probability distributions of process parameters [**Anonymous, 2010a**]. Additionally, exact control parameters are company secrets. Hence, in this thesis parameters of process steps belonging to process chains are described as uniform distributions of time and temperature.

4.4.1. Heat treatments. Processing of the majority of milk products contains a heating step. As milk is stored at lower temperatures, it often is preheated for economic reasons. The goal is to heat the product to final

temperature as fast as possible, and after a sufficient holding time, to cool it down rapidly. Most heat treatments are used to inactivate spoilage-, or pathogenic microorganisms, bacterial enzymes and toxins. Common assumption is that after proper pasteurisation no pathogenic bacteria are left alive [**Weber, 2006**]. Milk before heat treatment is called upstream milk, milk after heating is called downstream milk [**Spreer, 2005**]. A recontamination downstream in the process chain during further processing or during final packaging must be avoided. An overview over common heat treatments in Germany can be found in table 4.2.

TABLE 4.2. Common heat treatments in milk processing, [**Bylund, 2003**], [**Sienkiewicz and Kirst, 2006**], [**Töpel, 2004**] and [**Lewis, 1999**]

Heat treatment	**Temperature**	**Time**
Heat treatment before fermentation	90-95°C	2-10min
Heat treatment of cream for butter	105-118°C	-
Pasteurisation (Batch)	60-65°C	30-32min
Pasteurisation (ESL)	127°C	1-2s
Pasteurisation (HTH)	85-127°C	>1s
Pasteurisation (HTST)	72-75°C	15-30s
Sterilisation	>110°C	10-30min
Sterilisation (Batch)	109-120°C	10-40min
Thermisation	57-68°C	10-30s
Ultra-high temperature processing (UHT)	135-150°C	2-4s

Heat treatments lead to physical and chemical changes in the dairy product, which may hamper the acceptance by the consumer. At higher temperatures Maillard reactions between lactose and proteins may cause browning and a caramelish flavour. Heating causes fat to leak out of emulsified droplets. Milk for fermentation is heated to denaturate milk proteins. Milk for cheese production should not be heated too high, because a heating higher than pasteurisation temperatures causes denaturation of proteins leading to a softer coagulum in further cheese processing [**Bylund, 2003**].

Heat can be applied by direct or indirect heating. Usually, indirect heating by heat exchangers is used [**Bylund, 2003**], [**Anonymous, 2010a**]. A thin layer of milk (product) is led over a heated surface. Heat transfer in the product is done by convection. Transfer kinetics depend on viscosity of the fluid. Heat exchangers have to be cleaned regularly. Despite high temperatures, biofilms may form on the surfaces leading to a high bacterial load of the dairy product. Due to hairline cracks in the heat exchanger a recontamination may occur, see figure 4.2. For more-viscous products indirect heating can not be used.

Direct heating is done by injection of hot steam into milk or by infusion of milk into a steam chamber. The steam has a pressure of about 4 bar and a temperature of 145 °C [**Foissy, 2005**]. A rapid pressure reduction cools the milk quickly down and removes added water as cooler steam. During

cheese manufacture cheese curd is directly heated with warm water. During direct heating the temperature changes in the products are faster, but direct heating is more energy intensive than indirect heating.

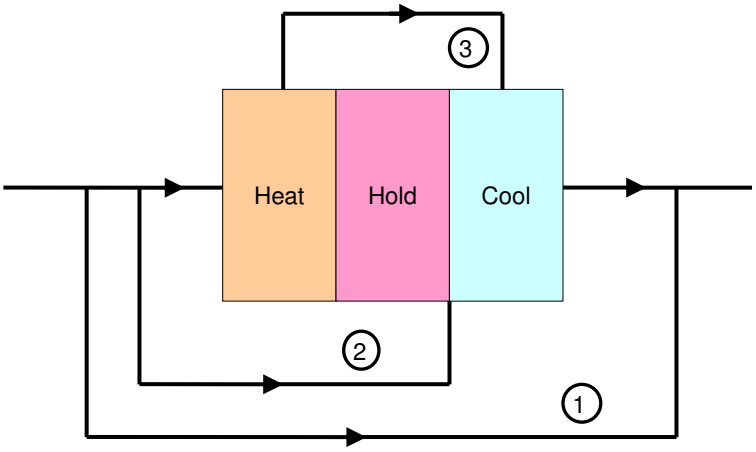


FIGURE 4.2. Bypass routes in milk pasteurisation: (1) via cleaning routes, (2) via reversal of flow direction, (3) via the heat exchanger (e.g. hairline cracks in heat exchanger plates), from [**Lewis, 1999**]

The most important heating process in any dairy plant is pasteurisation [**Marth and Steele, 2001**]. It is a relatively mild heat treatment. Chemical damage of the product is minimal; there are only small changes to sensory characteristics. Heat-labile microorganisms (vegetative bacterial cells, yeasts, molds) that may cause spoilage are killed off. Most pathogenic bacteria are inactivated, too [**Weber, 2006**], including *Listeria* spp. [**Smit, 2003**]. In Germany *´high temperature, short time´* pasteurisation (HTST) (traditional pasteurisation) and *´high temperature heating´* pasteurisation (HTH) are common [**Sienkiewicz and Kirst, 2006**]. Both are continuous processes.

HTST pasteurisation is done at 72-75°C for 15-30s. It reduces the total bacterial load whilst the ratio of acidifying bacteria and other bacteria remains almost the same [**Sienkiewicz and Kirst, 2006**]. Spores survive this treatment to a high degree. Dormant cells of *Mycobacterium avium* spp. *paratuberculosis* may survive pasteurisation, too [**Grant et al., 2005**].

HTH pasteurisation is done at 85-127°C for more than 1s. General spoilage flora is killed off, but spores may survive [**Sienkiewicz and Kirst, 2006**]. HTH-pasteurised milk is more photosensitive then traditional pasteurised milk [**Sienkiewicz and Kirst, 2006**]. HTH pasteurisation is often used in production of ESL (*´extended shelf life´*) milk. ESL is a general name

for heat treated products with a prolonged storage time due to different production procedures.

Effectiveness of HTST pasteurisation is determined by investigation of inactivation of alkaline phosphatase [**Belitz et al., 2009**]. In milk products with a fat content of more than 8%, alkaline phosphatase may be reactivated. HTH pasteurisation destroys both alkaline phosphatase and peroxidase [**Bylund, 2003**].

Another pasteurisation process formerly used is batch pasteurisation at 60-65°C for 30-32min. It is also referred to as *´low temperature, long time´* (LTLT) pasteurisation [**Bylund, 2003**]. Today in Germany it is only found in small dairies, e.g. in production of fresh organic goats milk. Drawbacks of batch heating are slow kinetics for heating and cooling, and pressure building up in the packaging during heating [**Smit, 2003**].

Aseptic milk can be produced with more rigorous temperature-time combinations. One goal in sterilising foodstuff is to kill off all bacteria including *Clostridium botulinum* which is the most heat resistant pathogenic bacterium [**Lewis and Heppell, 2000**]. A heat treatment of 121°C for 3min yields a log reduction of *C. botulinum* of about twelve [**Lewis and Heppell, 2000**], i.e., by 99.9999999999%. Due to strict heat treatment chemical alterations of milk components affect the taste of the end product. There are two different heat treatments that yield aseptic milk: ultra-high temperature processing and sterilisation.

Ultra-high temperature (UHT) processing is done at 135-150°C for 2-4s. It is a continuous process. Heating can be direct or indirect. Due to comprehensive security switching mechanisms, a mixing of upstream milk and downstream milk is not possible [**Bylund, 2003**], [**Spreer, 2005**].

Sterilisation is done at temperatures of at least 110°C for a duration of 10-30min. Thermised, pasteurised, or UHT milk may be used for production of sterilised milk [**Sienkiewicz and Kirst, 2006**]. Sterilisation traditionally is a batch process, but can also be applied as continuous process. Evaporated milk is tinned, and then sterilised in batches. Batch processing has the drawback of slow heating and cooling kinetics and a rising pressure during heating inside cans or bottles.

Both UHT treatment and sterilisation reduce the bacterial load about 10 log units [**Foissy, 2005**], i.e., by 99.99999999%. *Pseudomonas* spp. produces proteolytic and lipolytic enzymes that are not destroyed by UHT treatment. These enzymes may lead to spoilage of UHT milk despite bacteria being killed off [**Bylund, 2003**].

Pasteurisation of cream is done at 100-120°C for 14-120s [**Weber, 2006**]. Cream may also be UHT heated at 135-150°C for 2-8s. Sterilisation can be done at 109-115°C for 20-40min; all microorganisms, including spores, are

killed off, and all enzymes are inactivated [**Weber, 2006**].

Heat treatments are also used for technological reasons. A thermisation at 49-68°C for 5-40s weakens the raw milk flora and allows for longer storage at low temperatures [**Bylund, 2003**], [**Sienkiewicz and Kirst, 2006**], [**Weber, 2006**], [**Lewis and Heppell, 2000**]. It does not destroy the enzyme peroxidase present in raw milk. Therefore, thermised milk can easily be distinguished from pasteurised milk. Heat triggers germination of spores so milk must be rapidly cooled to a maximum of 4°C to prevent spoilage [**Bylund, 2003**]. Thermisation is also used as preparation for milk for production of cheese. It restores the protein equilibrium of refrigerated milk [**Sienkiewicz and Kirst, 2006**]. In production of farmer cheese ('*Frischkäse*') thermisation deactivates rennet and reduces further acidification [**Sienkiewicz and Kirst, 2006**]. In production of fermented milk products milk proteins often are partially denaturated before the actual fermentation process by heating to 90-95°C for 2-10min.

4.4.2. Separation processes. Separation is a major operation in food processing. In cheese production sometimes the curd is heated for faster separation from whey. Filtration and centrifugation are used to separate milk fat, denaturated proteins, dispersed particles, bacteria, spores and somatic cells. For both processes milk fat should be liquid so the milk usually is warmed up. Microfiltration and bactofugation are used to lessen the bacterial load and to remove other dispersed particles. Solids are separated from liquid milk, and accumulate in the sediment. Milk can contain up to 1kg of non-milk particles per 10,000 litres [**Bylund, 2003**]. In Germany it is forbidden to use sediment from separation processes for food or feed [**Sienkiewicz and Kirst, 2006**].

Bactofugation is a centrifugal separation. Most microorganisms and all spores are removed from the milk [**Spreer, 2005**]. It is used in cheese production to prevent cheese from late blowing during ripening due to sporeformers. It is also used in production of milk powder and other preserved milk products.

Skimmed milk can be pressed through microfilters with a pore size of about 1.4μm. This mechanical process reduces the bacterial load by more than 99.9% [**Sienkiewicz and Kirst, 2006**], [**Bylund, 2003**]. Spores are removed, too. Whole milk can not be microfiltrated because fat globules would plug the pores of the filter. Microfiltration is sometimes used for production of ESL milk traditionally pasteurised, skimmed milk is microfiltrated after separation. Cream is heat treated and mixed with the skimmed milk [**Bylund, 2003**], [**Spreer, 2005**].

When not homogenised milk is allowed to stand without stirring, a layer of cream settles on the surface due to gravity. Separation by centrifugation uses gravitational forces to separate particles and fat droplets from the watery milk phase, too. Centrifugal forces depend on mass of the particles and on

radius and speed of rotation [**Hui, 2006**]. Usually raw milk is separated into skimmed milk and cream by centrifugation. Coarse contaminating material is removed, too. Then skimmed milk and part of the cream are remixed, yielding milk with a clearly defined fat content. In Germany these are usually 3.5% (whole milk, '*Vollmilch*') and 1.5% (semi skimmed milk, '*Fettarme Milch*'). Separation has no influence on tenacity of bacteria, but they accumulate in the cream.

4.4.3. Homogenisation. In several countries creaming of pasteurised fresh milk is regarded as a quality characteristic, but not in Germany. Bigger fat droplets part faster from the emulsion than smaller ones. Homogenisation reduces the diameter of fat globules. When milk or cream is pressed through a small valve, fat globules are broken down from approximately 3,5μm to 1μm by a mixture of shear forces, cavitation and microturbulence. Homogenisation of cold milk or cream has almost no effect, hence, it is done at temperatures of 55-80°C [**Bylund, 2003**]. Pressure in the fluid is between 10 MPa and 25 MPa (100-250 bar), yielding velocities up to 250m/s [**Lewis and Heppell, 2000**].

The effect increases with a lower fat content and higher pressure and temperature [**Spreer, 2005**]. Usually, milk with a defined fat content (after separation and remixing) is homogenised; this is done in production of UHT milk and milk for fermentation [**Bylund, 2003**]. For pasteurised milk cream is diluted with skimmed milk to a fat content of 13-20% and then homogenised [**Spreer, 2005**]. The volume that passes the homogenisation facility is drastically reduced, saving up to 80% of energy [**Bylund, 2003**]. It is also possible to homogenise cream before remixing it with skimmed milk. To prevent re-aggregation, a second valve can be placed a small distance after the first one. In milk products with a high fat content, often such a two-step homogenisation process is used [**Spreer, 2005**]. Usually, the homogenisation facility is located upstream of the heating section [**Bylund, 2003**].

Advantages of homogenisation of milk include reduced creaming, a whiter colour of the product, a lower risk of oxidation, a better flavour and a higher stability of fermented milk products. Homogenisation is used in production of fresh milk, condensed milk, cream, yoghurt and milk mix beverages. In Germany most fresh milk is homogenised. Homogenised milk can not be processed to cheese, because the curd is too soft. Another disadvantage is the higher susceptibility for changes in taste due to light. The smaller fat globules are not totally covered with a membrane. Hence, lipolytic enzymes may degrade milk fat. Homogenised milk curdles faster than non-homogenised milk [**Sienkiewicz and Kirst, 2006**]. When homogenised cream is heated, it flocculates fast, an unwanted effect in coffee cream [**Sienkiewicz and Kirst, 2006**]. A second homogenisation may partly reverse these effects [**Sienkiewicz and Kirst, 2006**].

4.4.4. Storage and ripening. Characteristics and sensory quality of dairy products should be preserved during storage. In general, preservation is better at lower temperatures [**Smit, 2003**]. But even low temperatures do not stop psychrophilic bacteria from growing. Raw milk is always stored at a low temperature, optimal at 4°C. After milking raw milk is stored in tanks. Transport to the dairy plant is done between twice a day up to all four days. In small-scale manufacture milk may be produced and processed in the same facility. When raw milk has to be stored for several days it is thermised to reduce the number of psychrophilic bacteria. A deep-cooling at maximal 2°C allows for longer storage of raw milk, too [**Lewis and Heppell, 2000**].

For fresh milk products storage temperature has a big impact on keeping quality. Without cooling in a domestic refrigerator, keeping time of pasteurised milk is about 24h [**Lewis, 1999**]. There are milk products that allow for a storage at room temperature. These products are UHT processed, sterilised or got their content of water highly reduced. Examples are UHT milk, evaporated milk and dried whey.

Milk should be stored dark, because the energy of ultraviolet light changes its taste. First it gets a ´burnt feather´ taste, after longer light irradiation taste changes to ´board card´ [**Töpel, 2004**]. This is due to degradation of the amino acid methional [**Töpel, 2004**], [**Bylund, 2003**]. If the membrane of fat droplets is damaged, e.g. by homogenisation, naturally present lipases degrade the fat. A taste unwanted in most milk products is the result of such lipolysis [**Bylund, 2003**]. Other changes due to microbial metabolic activity may occur, e.g. souring.

Ripening is a storage of dairy products at special environmental conditions, e.g. time, temperature and humidity that intentionally leads to a desired alteration of flavour and texture. During ripening of whipping cream milk fat crystallises yielding cream that can be whipped. Spreadability of butter can be influenced by a controlled fat crystallisation. Temperature-time-combinations used in cream ripening depend on composition of the milk fat. In cheese ripening, microorganisms metabolise milk components, therewith producing special texture and flavours, e.g. lipolysis leads to changes in taste. Environmental conditions during cheese ripening depend on the type of cheese. Its low a_w value protects cheese from spoilage during ripening. Lactic acid fermentation is usually finished after 1-2 weeks [**Weber, 2006**]. In general it is assumed that undesirable bacteria are killed off in the long period of hard cheese ripening.

4.4.5. Fermentation. Fermentation of milk is done by bacteria and/or yeasts that digest the disaccharid lactose. One product is lactic acid that lowers the pH value of the milk product, leading to coagulation of casein. Decomposition of proteins by fermenting microorganism provides other bacteria with essential nitrogen and sulfur compounds [**Weber, 2006**]. Bacteria also produce characteristic flavours. The population of living

starter bacteria inhibits growth of spoilage bacteria. Fermentation is one process step in production of sour milk products like yoghurt, buttermilk or quark but also in cheese production.

In former times, the natural flora of raw milk was used for fermentation. Now commercial starter cultures provide standardised products. Starter cultures are certain strains of bacteria or compositions of different microorganisms, e.g. *Lactobacilli*, *Lactococci* and *Bifidobacteria*, yeasts and moulds [**Cogan et al., 2007**], [**Weber, 2006**]. Starter cultures are specially designed for production of one single dairy product in a plant. There are various incubation temperatures and methods of application. In yoghurt production temperatures between 32°C and 46°C are used, depending on starter culture and characteristics of the end product [**Weber, 2006**]. Starter cultures used in production of Swiss cheese must survive heating of the curd and have to grow at temperatures close to their maximum growth temperature [**Stern and Frazier, 1941**]. Presence of antibiotics or bacteriophages may cause problems in fermentation, see section 4.3.

4.4.6. Coagulation. For manufacture of some dairy products a major part of milk protein has to be separated from whey. Proteins can be precipitated by acidification, by addition of enzymes (rennet), or by a combination of both. Acidification usually is done by fermentation. Calves produce rennet to digest milk. Today rennet is obtained from slaughtered calves or from recombinant bacteria. In some cheeses extracts from molds, plants or bacteria are used as substitute for animal rennet. Curdling and separation after coagulation lead to a redistribution of fat and protein. Some cheeses are made from whey protein that coagulates due to heating, e.g. Ricotta cheese. Yoghurt is a gel formed by coagulated milk protein.

4.4.7. Concentration of anhydrous mass. In production of condensed milk or certain fermented milk products like Greek yoghurt, dry matter is increased. This can be done by centrifugation, evaporation, ultra-filtration or reverse osmosis. Addition of milk powder, milk- or whey protein that was obtained by such processes also increases the dry matter. In production of lactose, casein or other milk components the anhydrous mass is concentrated, too.

Milk and whey powder are produced by drum drying or spray drying. Milk destined for yoghurt production is evaporated to increase the dry matter content from 13% up to 16.25% [**Bylund, 2003**]. During evaporation pasteurised milk or whey is heated in a vacuum two times. Heating in a vacuum lowers the boiling temperatures to 70 °C and 40 °C, respectively [**Bylund, 2003**]. Therewith, chemical changes due to heat are reduced. The product is a concentrated fluid.

4.4.8. Composition. A variety of dairy products is composed of dairy and non-dairy ingredients like fruit preparation, sugar or salt. Adding of

ingredients always contains a risk of contamination. Microorganisms from the environment may be transferred to the product. If one of the ingredients is contaminated it pollutes the mixed end product. Ingredients are bought either as prefabricated mixtures or as components that are composed in the plant. In some dairy plants, fruit preparations for yoghurt or white cheese are cooked on a daily basis [**Dairy SachsenMilch, 2009**].

Addition of ingredients changes the conditions for microbial growth and survival. Fruit preparations may lower the pH value, but also add the nutrient sugar. Some ingredients are added for taste and food preservation, e.g. salt. [**Smittle, 2000**] suggests that addition of nisin or other bacteriocins would extend the shelf life of pasteurised milk. Addition of milk or whey powder or milk protein enhances the dry mass of milk for yoghurt or white cheese production. In cheese production milk protein powder can be added to increase the yield of curdling.

4.4.9. Packaging. Packaging serves different purposes. It disconnects the packaged food from an environment full of possible contaminations. Milk is photosensitive, therefore, packages should be opaque. Bottles used for fresh milk are often made of brown glass. To prevent a recontamination by polluted packaging material, bottles, tins, cans, cups, tubular packagings, and Tetra Paks® or other cartons must be clean or, for preserved milk products, aseptic. Sterilisation of packaging material can be done with hydrogen peroxide, superheated steam or irradiation [**Smit, 2003**].

Most dairy products are packaged after manufacture. For others packaging is not the last step in production; the product is portioned and then incubated or heated in the package. A recontamination of the product by package should be avoided. Pasteurised products are packed in clean packages. To guarantee the long shelf life of ESL and UHT products, aseptic packages are used. Aseptic packaging systems are complex, because a recontamination of the product by packaging must be prevented. Evaporated milk is tinned before heat treatment. This excludes a recontamination after heating. Sweetened condensed milk is tinned after heat treatment to prevent browning of the product by means of Maillard reactions. Set yoghurt is incubated in the packaging.

4.5. Milk products

Raw milk as such, without processing, is sold as raw milk on the farm, or as attest milk (´*Vorzugsmilch*´) in groceries. However, only a small amount of milk is marketed via this route. Some hard cheeses are made directly from raw milk. All this raw milk products are monitored. Nevertheless, there are foodborne outbreaks caused by raw milk products. In 1981 five of 10 and in 1982 six of 11 foodborne outbreaks of campylobacteriosis in the United States of America were associated with consumption of raw milk [**Anonymous, 1983**]. In general it is assumed that the low a_w value, the salt content and the cheese flora kill off undesirable bacteria in

TABLE 4.3. Approximate composition of some dairy products, in % and partially rounded, from [**Fox and McSweeney, 1998**]

Product	**Moisture**	**Protein**	**Fat**	**Sugars**[a]	**Ash**
Light whipping cream	63,5	2,2	30,9	3,0	0,5
Butter	15,9	0,9	81,1	0,1	2,1
Anhydrous butter oil	0,2	0,3	99,5	0,0	0,0
Ice-cream[b]	60,8	3,6	10,8	23,8	1,0
Evaporated whole milk	74,0	6,8	7,6	10,0	1,5
Sweetened condensed milk	27,1	7,9	8,7	54,4	1,8
Whole milk powder	2,5	26,3	26,7	38,4	6,1
Skim milk powder	3,2	36,2	0,8	52,0	7,9
Whey powder[c]	3,2	12,9	1,1	74,5	8,3
Casein powder	7,0	88,5	0,2	0,0	3,8
Cottage cheese, creamed	79,0	12,5	4,5	2,7	1,4
quark	72,0	18,0	8,0	3,0	-
Camembert cheese	51,8	19,8	24,3	0,5	3,7
Blue cheese	42,4	21,4	28,7	2,3	5,1
Cheddar cheese	36,7	24,9	33,1	1,3	3,9
Emmental cheese	36,0	28,9	30,0	-	-
Parmesan cheese	29,2	35,7	24,8	3,2	6,0
Mozzarella cheese	54,1	19,4	31,2	2,2	2,6
Processed cheese[d]	39,2	22,1	31,2	1,6	5,8
Acid whey	93,9	0,6	0,2	4,2	-

[a] Total carbohydrate
[b] Hardened vanilla, 19% fat
[c] Cheddar (sweet) whey
[d] American pasteurized processed cheese

the long period of cheese ripening. Nevertheless, in the Netherlands in 2006 hard aged raw-milk cheese caused a large-scale *Salmonella* outbreak [**van Duynhoven et al., 2009**].

There is a wide range of different milk products made from raw milk. They have different compositions, physical and chemical properties. Most of these differences have an influence on microbial growth and tenacity. Fresh milk, cream, coffee cream and condensed milk differ from each other only in fat and dry matter content. The approximate composition of some dairy products is listed in table 4.3.

Process chains for production of the various milk products distinctly differ from each other. For certain milk products the succession of process steps is outlined in this section. The following milk products are not taken into account in this thesis: reconstituted milk made from butterfat or plant fat emulsified in milk powder, casein, lactose, powder coffee whitener, milk powder, whey powder and chocolate products. In figure 4.3 an overview of the variety of dairy foodstuff common in Europe is given.

4.5.1. Fresh milk. Most fresh milk is produced from raw milk by separation and fat standardisation, heat treatment and homogenisation.

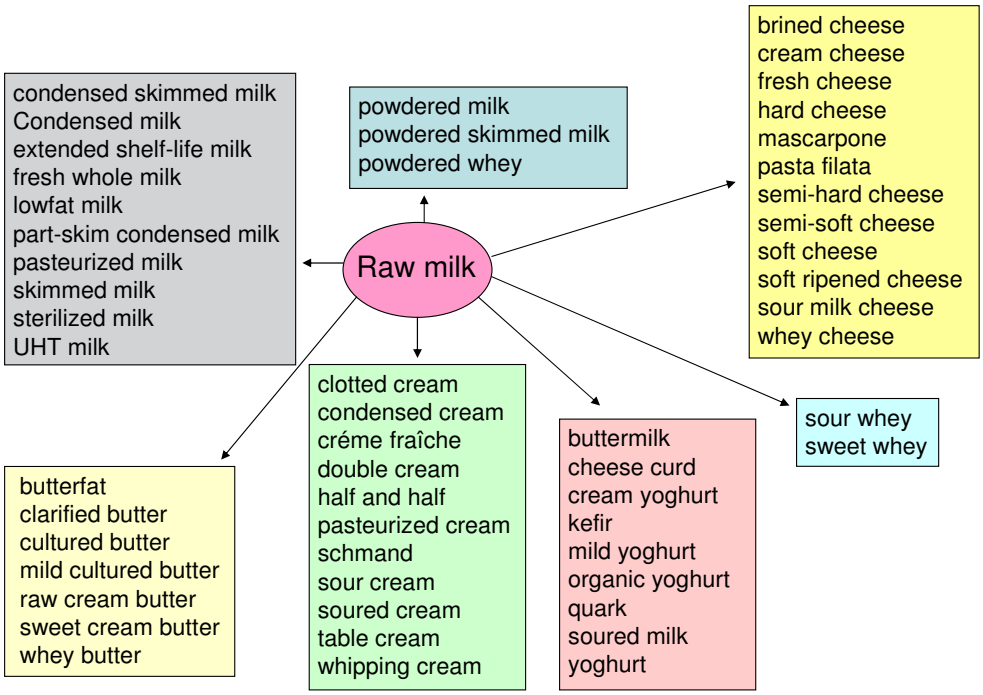


FIGURE 4.3. Diversity of milk products

Process lines differ even for the same type of fresh milk from dairy to dairy [**Bylund, 2003**]. Common fat content standards in Germany are 0.1% (skimmed milk), 1.5% (semi skimmed milk) and 3.5% (whole milk). Pasteurised milk may be sold with a natural fat content of 3.8%. Fresh milk is almost always heat treated to kill off spoilage- and pathogenic microorganisms. In Germany only a tiny fraction of all fresh milk is consumed as raw milk without heat treatment, it is sold as raw milk on the farm or as attest milk in groceries. Some (organic) HTST pasteurised fresh milk is manufactured without homogenisation.

Pasteurised milk is HTST-pasteurised, usually homogenised, and cooled. It is packed in clean bottles, tubular packagings, Tetra Paks® or other cartons. Cool storage is mandatory for a shelf life of up to 16 days [**Bylund, 2003**]; without cooling keeping time of pasteurised milk is about 24h [**Lewis, 1999**]. To discriminate HTST pasteurised milk from ESL milk in Germany it is named ´traditionally manufactured´.

ESL milk is HTH-pasteurised, homogenised and cooled. HTH-pasteurisation can be substituted by a microfiltration of skimmed milk, combined with a traditional HTST-pasteurisation after fat standardisation. Aseptic packaging is used to prevent contamination by package, and to assure the extended shelf life. Cool storage at 6-10°C is necessary to maintain a keeping time of up to 21 days [**Sienkiewicz and Kirst, 2006**]. In Germany the market share of ESL pasteurised milk has exceeded that of

traditional pasteurised milk.

UHT milk is UHT heated, homogenised and cooled. To guarantee a long storage time aseptic packaging is used. UHT milk can be stored at room temperature; it has a keeping time of at least six weeks [**Spreer, 2005**]. Raw milk quality is crucial for production of UHT milk [**Bylund, 2003**]. Sterilised milk is homogenised, packaged, heated and cooled. A continuous processing with aseptic packaging after sterilisation is possible. Sterilised milk has a keeping time of up to one year at room temperature [**Spreer, 2005**].

Milk mix beverages consist of whole milk, partially skimmed milk, or skimmed milk, and up to 30% of flavouring and colouring ingredients [**Spreer, 2005**], [**Sienkiewicz and Kirst, 2006**]. Other dairy mix beverages are made from whey, buttermilk, sour milk, kefir or yoghurt. Depending on heat treatment, these mix beverages must be stored cool, or can be stored at room temperature.

4.5.2. Cream. Today cream is obtained by centrifugal separation. In fresh milk production it accrues when the fat content is standardised. The various cream products differ in their fat content, and they may be soured or not. Additionally, cream is used for production of butter, buttermilk and farmers cheese. In production of ripened and unripened cheese it may be added to adjust the fat content.

It is assumed that the high fat content and increased viscosity of cream lead to a higher tenacity of bacteria during heating. This is the reason why cream usually is treated to higher temperatures than those in milk pasteurisation [**Weber, 2006**]. Most cream products are homogenised.

Whipping cream contains about 30% fat. A fat content of 35% is optimal for foam quality [**Spreer, 2005**]. Cream with a fat content as low as 20% can be whipped when certain additives are added [**Spreer, 2005**]. As acidity of cream has an influence on whipping, it sometimes is slightly fermented to reach a pH value of 6.2-6.4 [**Spreer, 2005**].

Whipping cream may be manufactured as follows. Cream is stored at 3-5°C. Then it is heated in several steps to denaturate whey proteins and to kill off all microorganisms and spores. Now homogenisation follows. Homogenised whipping cream yields a foam more stable [**Spreer, 2005**]. During ripening that follows, partial crystallisation of milk fat is obtained, allowing the cream to be whipped [**Bylund, 2003**]. Depending on heat treatment, whipping cream must be stored cool, or can be stored at room temperature.

The process chain of coffee cream equals the one for whipping cream. Additionally, cream is mixed with skimmed milk to reduce the fat content to 10-18%. It can be aseptically packed after heat treatment or sterilised

in bottles. Coffee cream is a preserved milk product that can be stored at room temperature.

4.5.3. Evaporated milk. Evaporated milk is made from whole milk, skimmed milk or milk recombined from milk powder. Milk is heat treated and then concentrated in an evaporator. Common concentrations are twofold (with 7,5% fat content and 17,5% fat-free dry mass) or threefold (with 33% fat-free dry mass und 4-10% fat content). Then it is homogenised and cooled. Sometimes stabilisers are added. The evaporated milk is now UHT-heated and aseptically packed or canned and sterilised. At a temperature of 0-15°C it can be stored for an unlimited time [**Bylund, 2003**].

4.5.4. Sweetened condensed milk. Sweetened condensed milk is made from whole milk, skimmed milk or milk recombined from milk powder. Like evaporated milk, sweetened condensed milk is produced in different concentrations and fat contents. The milk is heat treated. During concentration in an evaporator treacle is added. Therewith the a_w value of the product is reduced to a point that inhibits growth of most microorganisms [**Marth and Steele, 2001**], [**Bylund, 2003**]. Then the sweetened condensed milk is homogenised. Finally it is aseptically portioned and packed. The high sugar content of 62,5-64,5% enhances the possibility of Maillard reactions during heating. Therefore, sweetened condensed milk is not batch sterilised like evaporated milk.

4.5.5. Butter. Butter is one of the first dairy products that was traded internationally [**Marth and Steele, 2001**]. Butter contains 80-90% milk fat. Water, salt and carotene for colour may be added. Semi-fat butter contains only 39-41% fat. Sorbic acid, gelatine, carotene, emulsifiers and citric acid may be added [**Weber, 2006**].

Usually, butter is manufactured from sweet or soured cream. Butter made from sweet cream can be spiked with lactic acid or acidifying cultures, yielding butter with a mild sour flavour (*´Mildgesäuerte Butter´*). For production of 1kg butter, 2.5 litres of cream or 25 litres of raw milk are needed, respectively.

For butter production first cream is heated to at least 85°C to kill off microorganisms and lipolytic enzymes, but not homogenised [**Spreer, 2005**]. Cream for raw milk butter is not heated. Spreadability of butter is influenced in cream ripening, during which the milk fat partially crystallises. Composition of the milk fat determines temperature-time-combinations used. Now the cream is battered. This can be done in either a continuous or a batch process. Agitation damages the membranes around fat globules. Fat agglutinates, encasing water droplets; the two phases of the fat-in-water emulsion cream are reversed. Optionally salt is added, and finally kneading yields a soft and homogeneous butter.

Up to 18 billion droplets of water are dispersed in 1g of the water-in-oil emulsion. The diameter of water droplets in conventionally made butter has been reported at <1 to $>30\mu$m. The microbiology of butter reflects the microflora present in pasteurized cream from which it is made, and water added at time of salting butter, sanitary conditions of process equipment, manufacturing environment, and conditions under which the product is stored. All of these activities impact on microflora of the final product. The main source of microorganisms in butter made under excellent sanitary conditions is cream.

4.5.6. Fermented milk products. For manufacture of fermented milk products mostly pasteurised milk, usually with a standardised fat content, is used. Milk may be heated to 90-95°C to denaturate the milk proteins. This yields a more viscous end product [**Bylund, 2003**]. Often dry matter is increased by means of evaporation or addition of milk protein, whey powder or milk powder. This yields a superior texture of the fermented product, an improved taste and better growth of fermenting microorganisms. Then the milk is warmed and a starter culture is added. The culture-milk mix is incubated; duration and temperature both depend on microorganisms and end product. When the desired level of acidity is reached the fermented milk is cooled to stop further growth and metabolism of acidifying microorganisms. The culture-milk mix can be incubated in tanks or portioned in the final packaging.

Milk is fermented by lactic acid bacteria and yeasts. Growth of bacteria with capsules renders milk more viscous and slimy, a wanted effect only in some sour milk products. Lactic acid bacteria degrade lactose and produce lactic acid. Milk sours to a pH value of 4-4.5, causing precipitation of casein and curdling. Most other bacteria can not cope with the low pH of fermented milk products and die off after a short time. Falenski *et al.* report a survival of *Brucella* spp. in yoghurt with 3.5% or 10.0% fat for four and two days, respectively [**Falenski et al., 2010**].

Yoghurt is one of the eldest sour milk products [**Weber, 2006**]. Bacteria traditionally used for production of yoghurt are *Streptococcus thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*. Milk is pasteurised, may be it is homogenised and its dry mass is increased. Then starter cultures are added and the culture-milk mix is incubated at temperatures between 32°C and 46°C [**Weber, 2006**]. The mix can be filled into plastic cups or glass jars and then incubated, yielding set yoghurt. Creamy yoghurt is incubated in tanks; it is stirred and portioned after curdling. In section 8.8 a process chain for production of yoghurt is described in detail.

Buttermilk is a by-product of butter manufacture. It contains 0.3-0.5% fat. In production of sour cream butter most starter culture bacteria are retained in buttermilk [**Marth and Steele, 2001**]. Kefir is milk fermented by *lactobacilli* and yeasts. The result is a simultaneous acidification and alcoholic fermentation of milk. It has a fat content of 0.5-6% [**Bylund, 2003**].

Process steps are similar to yoghurt production, except incubation that is done in two steps. Other fermented milk products are sour cream, créme frâiche and sour milk products. They are made from cream, milk, sweet buttermilk or whey, respectively.

4.5.7. Unripened cheese. Unripened cheese is made from milk curdled by acidification and enzyme activity. The term includes cottage cheese (*´Hüttenkäse´*), farmers cheese (*´Frischkäse´*), and white cheese (quark *´Quark´*). Processing is partly similar to production of ripened cheese, but lacking ripening (as the name suggests). In Germany unripened cheese must not be made from raw milk.

For production of cottage cheese curd is thoroughly washed. Usually this is done in three steps at 30°C, 16°C and 4°C, respectively [**Bylund, 2003**]. Lactose and lactic acid are eluted, yielding a low acidity of the curd. Now cooled, salted pasteurised cream is added.

Usually skimmed milk is used for production of white cheese. Curdled milk is thermised at 56-60°C and then cooled down to 37-45°C. Then the curd is separated from sour whey. To enhance the fat content sweet or sour cream is added. The content of dry matter in white cheese varies between different countries between 14% and 24%.

4.5.8. Cheese. Most cheeses are made from pasteurised milk. Some are made from raw milk; for certain varieties milk must not be heated above 40°C to leave the natural milk flora intact. Often microorganisms and especially spores are removed by bactofugation or microfiltration to prevent late blowing in hard and semi-hard cheese. Milk may be standardised in fat and protein content by addition of cream and concentrated milk protein [**Bylund, 2003**]. Cheese cultures and rennet are added to warm milk. Then the mixture is left unstirred until proteins are coagulated to the desired degree by fermentation and enzyme activity. Temperature, amount of rennet, milk composition, starter culture and optimal consistency of the curd differ for every cheese variety. Now the curd is sliced in pieces. Form of slicing and size of the curd chunks depend on the cheese variety produced. For some cheeses the curd is washed or heated. Then the curd pieces are bailed into cheese moulds. Cheese loafs are pressed and salted. Ripening is done at defined temperature-humidity combinations, depending on cheese variety. During ripening cheese components are modified by cheese cultures and enzymatic activity. One process chain for hard cheese production is pictured in figure 4.4.

Sour milk cheese it made from sour milk quark, sometimes quark coagulated with rennet is added. Production process differs from that of rennet cheeses [**Bylund, 2003**]. Processed cheese is made from cheese mixed with emulsifying salts and other ingredients like acidifying substances, cream, whey powder, salt, colourings, flavourings, spices, mold inhibitors and

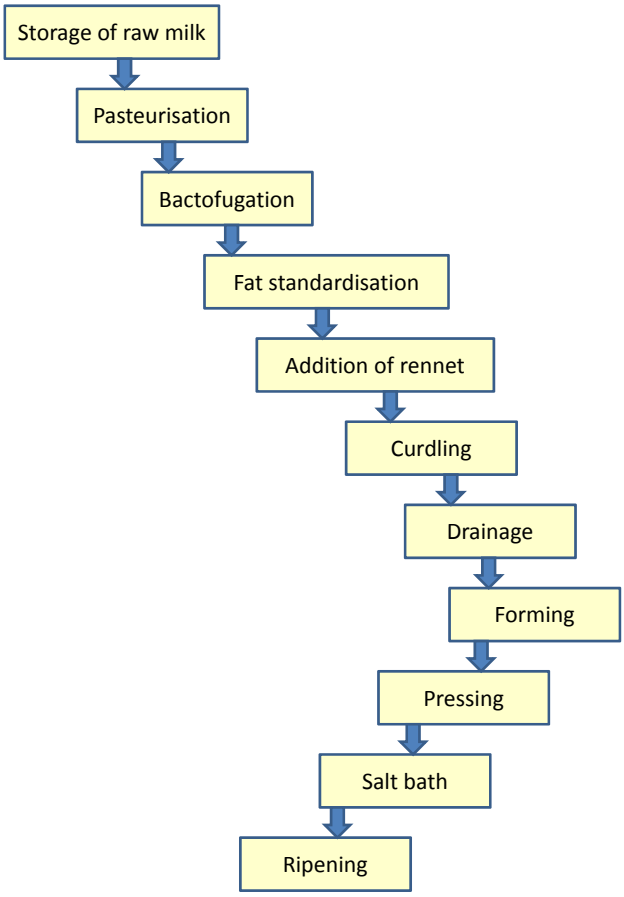


FIGURE 4.4. One process chain for hard cheese production

anti-sticking agents [**Kapoor and Metzger, 2008**]. Cheese is grounded and melted, then the other ingredients are added. The mixture is heated to 70-95°C, cooled and packed [**Bylund, 2003**]. Usually processed cheese contains 30-45% fat in dry matter.

CHAPTER 5

Predictive Microbiology

A model is a formalised description of a natural phenomenon. The complex reality is simplified to fit into an idealised hypothetical system. Predictive microbiology is the development of models to predict growth, survival or inactivation of microorganisms.

In predictive microbiology the term ´model´ is used for different structures: for a mathematical function describing bacterial population kinetics, and for the algorithm calculating the bacterial counts under various conditions. The former is a model of one growth or survival curve obtained in one experiment with certain (not necessarily static) conditions, or of dependencies of growth or survival parameters on conditions. The latter is a model of responses of bacterial populations to intrinsic and extrinsic conditions. It contains the mathematical functions described above, and is in the following referred to as predictive model.

´Predictive microbiology is based upon the premise that the responses of populations of microorganisms to environmental factors are reproducible, and that by considering environments in terms of identifiable dominating constraints it is possible, from past observations, to predict the responses of those microorganisms.´

[**Ross and McMeekin, 1994**]

Predictive models are used to optimise manufacture, or to determine monitoring points in the food or feed supply chain. Calculation of shelf life of foods, estimation of microbial stability of newly developed products, and prediction of risk of persistence of pathogens during storage are applications of predictive microbiology. It is also utilised in development of HACCP programs. In case of an intentional or unintentional contamination of food, the consequence, depending on matrix, microorganism, inoculum and process step in the supply chain, can be assessed. The dairy industry uses predictive modelling for forecasting keeping quality, to estimate growth and toxin production of bacteria, and to determine optimal pasteurisation conditions [**Griffiths, 1994**].

Another use of predictive microbiology is the quantitative description of conjugation of bacteria leading to resistances against antibiotics. Such predictive models can be used in optimisation of antimicrobial dosage regiments to minimise resistance development. The interested reader may refer to [**Gehring et al., 2010**].

In chapters 2-4 the biological knowledge required for modelling kinetics of bacterial populations is provided. Now the modelling process and some of the common models in predictive microbiology are highlighted. First, steps in development of a predictive model are described. Then different types of models are classified. After that, the terms variability and uncertainty are discussed. Finally, a selection of models used in predictive microbiology is presented.

5.1. Modelling process

Development of a predictive model is an iterative process. The question(s) the model should answer must be fixed first, *´Modelling itself is not a goal; it is a means of achieving a goal.´* [**Annino and Russell, 1979**]. Then microorganism(s) and environmental conditions have to be characterised in order to determinate relevant experimental settings.

Next step is the selection of a mathematical model appropriate for the specific problem, and the design and performance of experiments to obtain data for fitting and validation. The influence of different models and the effect of process variations on the output must be compared. If the latter has a more profound effect, the simplest model available should be chosen [**van Gerwen and Zwietering, 1998**]. Van Asselt and Zwietering [**van Asselt and Zwietering, 2006**] reported that variability in a set of more than 4,000 D-values that was collected from the literature was bigger than the influence of most factors.

The goal is to describe kinetics of growth or inactivation processes accurately and with as few parameters as possible (Ockham's razor) [**Ratkowsky, 1993**], [**Van Boekel, 2002**], [**Barbosa-Cánovas et al., 2005**]. The more parameters are included in a model, the more data are needed for fitting. Usually, parsimonious models make better predictions [**Baranyi et al., 1996b**], [**Delignette-Muller et al., 1995**]; and the amount of non-linearity is smaller in simpler models [**Barbosa-Cánovas et al., 2005**].

A simple model needs only few parameters as input and makes general statements, whilst a complex model needs a large amount of data for fitting and is applied to certain scenarios with defined conditions. If there is only little information about environmental conditions and the relevant microorganism, uncertainty is increased, and only a simple model can be applied. Models that are too simple are not able to fit the data, yielding consistent deviations between experimental data and predicted values [**Barbosa-Cánovas et al., 2005**].

A complex model may yield more precise predictions, but the output of one simulation suites only the conditions of this particular simulation. Models that are too complex may yield non-unique solutions, and fit is not improved

by changes of parameter values [**Barbosa-Cánovas et al., 2005**].

For obtaining a high fitting precision, it may be necessary to change the mathematical model after analysis of experimental data. In this case, additional data may be needed. Model performance may also be approved by reparameterisation [**Ratkowsky, 1993**].

High quality data from well-designed experiments is essential in predictive modelling [**Barbosa-Cánovas et al., 2005**]. Some measurements may carry more information than others. For reduction of uncertainty and improvement of accuracy without increasing costs, distribution of data points should be considered during experimental design. For example Ernstberger [**Ernstberger, 2008**] showed that data points from different phases of a sigmoid growth curve differ in their information content as to different growth parameters, e.g. measurements done during lag phase contain no information about maximum growth rate. Usually, data from the literature lack important information, e.g. strain or biovar of the species or history of the cells.

The selected model is fitted to the data, i.e., model parameter values are calculated for a best fit of model curve to data points. In general, this is done by regression techniques [**Barbosa-Cánovas et al., 2005**]. In ordinary linear regression a linear function is fitted by method of least squares. Method of least squares minimises the sum of squared residuals.

Reliability of the model output must be ensured. It is essential that a model reproduces the data it was built with. Model performance in reproduction of experimental data other than that used for previous fitting should be analysed. A model must be maintained to stay useful. New experimental data is included by further fitting and validation.

Figure 5.1 pictures the optimal process of development of a model in predictive microbiology. The interested reader may refer to Annino and Russel in [**Annino and Russell, 1979**] for an overview over lurking pitfalls in simulation analysis.

5.2. Variability and Uncertainty

Variations in outcome of an experiment may be due to a variability of the measured factor itself, to experimental impreciseness (uncertainty), or to both. The term variability refers to characteristics of the system observed. It can only be reduced by a change of the system itself. Uncertainty is due to the observer. Sir David Cox characterised variability and uncertainty as follows:

'Variability is a phenomenon in the physical world to be measured, analysed and where appropriate explained. By contrast, uncertainty is an aspect of knowledge.'

found in [**Vose, 2008**], page 47

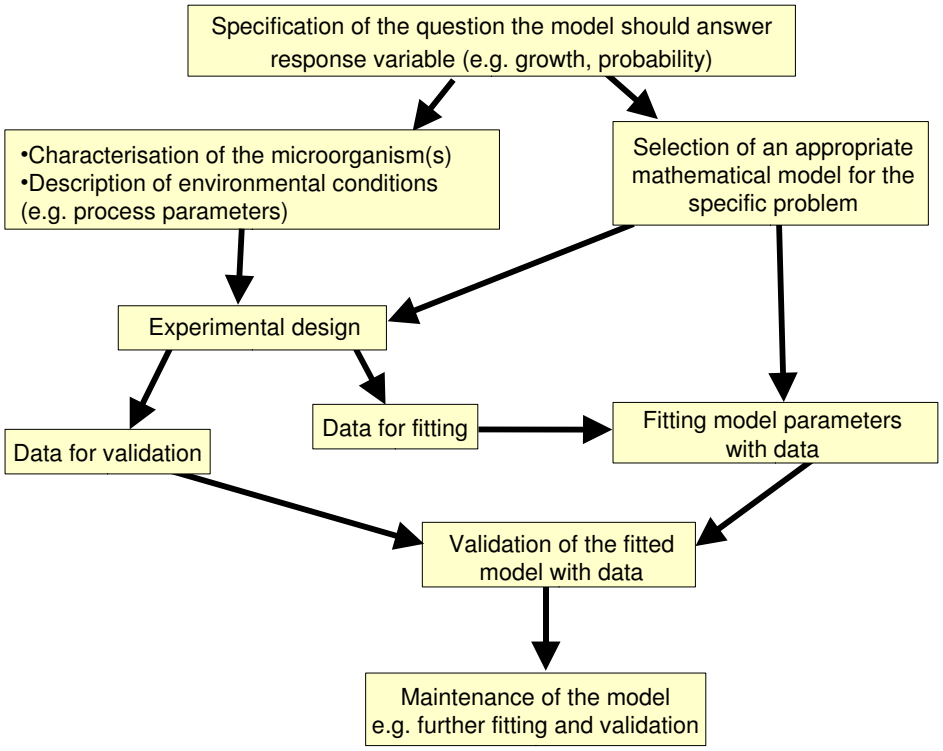


FIGURE 5.1. Development of a predictive model - a best case

Biological processes often are highly variable, e.g. generation times of individual cells in a growing population vary over a wide range [**Banks, 1994**], [**Kelly and Rahn, 1932**], [**Prescott, 1959**]. Pretreatment of the cells can drastically change the shape of the survival curve, e.g. for *Listeria monocytogenes* [**Págan et al., 1997**], [**Rowan and Anderson, 1994**]. Even slight differences in the matrix can lead to crucial differences in the survival curve, e.g. *Mycobacterium tuberculosis* dies off in 15-20min at 60°C in milk, but survives at the same conditions in the skin on the milk [**Hesse, 1960**]. In most manufacturing processes parameters like time and temperature are not set values, but vary over certain ranges.

Individual bacteria differ in their kinetics, but the large number of microorganisms in a population leads to averaged population kinetics. Kinetics of large populations can be modelled by continuous functions. Models for small populations or single cells have to be discrete.

'Variety, and variability, are the spice of life.'

[**Kot, 2001**], page viii

Variability is not reduced by further information about the system. When the proportion of variability is high and there is almost no uncertainty, the model performance can not be improved by collecting more data.

There is no such thing than a ´true´ growth or survival curve for fixed parameters, but a set of families of curves with variable parameters that follow probability distributions. Predictive models that do not include variability of the modelled system only yield point estimates, e.g. of the mean, and can not reproduce the full range of possible kinetics.

Uncertainty is due to the observer's lack of knowledge. Sometimes it can be reduced by further studies. Uncertainty may be increased by influencing factors of the system that are not included in the model. Data from the literature often lack information about stem and biovar, and of pretreatment of the cells, therewith increasing the uncertainty.

For inclusion into predictive models, variability and uncertainty can be represented by probability distributions that provide a mathematical structure for propagation during calculation [**Ghanem and Red-Horse, 1999**].

It is useful to keep variability and uncertainty separate. Otherwise, it can not be determined, if more data lead to a better prediction. Vose [**Vose, 2008**] suggests to combine uncertainty and variability in the same model, but in distinct probability distributions. The comparison of one simulation, in which all distributions are sampled, with another simulation, in which all uncertainty distributions are set to their mean value, gives an estimation of the proportion of uncertainty.

5.3. Model classification

Models are classified according to different criteria. An overview of common model classifications is given in figure 5.2.

According to how much *a priori* information on the system is available, there are two different approaches to build a mathematical model. Theory-based modelling is based on physical, chemical or biological theories. These models are also called mechanistic or white-box models. Mechanistic models *´(...) provide interpretation of the response observed in terms of the underlying mechanisms and are more amenable to refinement as knowledge of the system increases.´*, [**McMeekin and Ross, 2002**] and [**McMeekin et al., 2002**]. Construction of mechanistic models requires a detailed knowledge of the physiology of bacteria.

If the available data is the basis of model building, the model is called empirical, data-dependent, or black-box model. The underlying mechanisms are not considered, e.g. they are not known. In an empirical model a mathematical formula is derived from data analysis. Empirical models do not contain any information about the kinetics outside the region in which measurements are made. They should only be used for interpolation [**Baranyi et al., 1996b**].

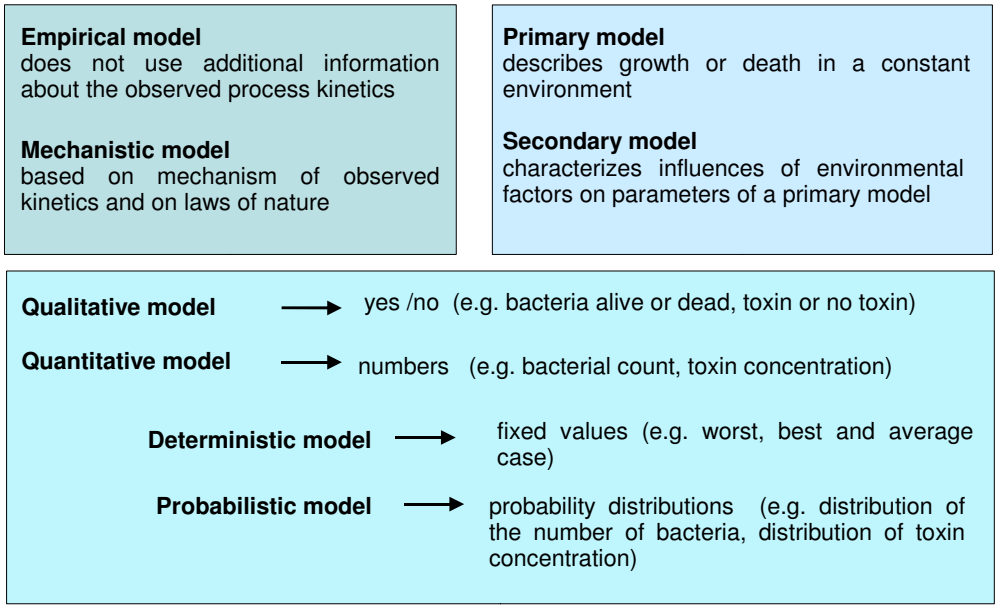


FIGURE 5.2. Basic model types

Generally, mechanistic models are regarded as superior to empirical models [**Barbosa-Cánovas et al., 2005**]. Most models in predictive microbiology are partly empirical and partly mechanistic [**Hao, 2007**].

Whiting and Buchanan [**Whiting and R., 1993**] classified models into three categories. This classification is widely used today. Models describing changes in bacterial counts over time for certain environmental conditions are called primary models. They describe the response of a population of microorganisms to one set of conditions. Primary models are used to estimate kinetic parameters, e.g. generation time or lag time duration. Response of this kinetic parameters to changes in environmental conditions is described by secondary models. Common secondary models are response surface (polynomial) models [**Box and Draper, 1987**], [**Li et al., 2008**] using statistics. There are approaches that use artificial intelligence [**Barbosa-Cánovas et al., 2005**]. The term model as defined above refers to primary and secondary models. Applications that provide predictions by use of primary and secondary models are called tertiary models. They are an interface between scientist and the end-user. The term predictive model defined above refers to a tertiary model.

A distinction can be made between qualitative and quantitative models. Qualitative growth/no growth models determine if a microorganism will grow under certain environmental conditions. This is of special interest for the food industry, because it helps to identify the processing and storage conditions that kill off, or eliminate growth of, spoilage, or pathogenic bacteria. Examples are the probabilistic growth/no growth interface of Ratkowsky and Ross [**Ratkowsky and Ross, 1995**], the Microbial

Responses Viewer database [**Koseki, 2009**], and the logistic growth/no growth model for *Listeria monocytogenes* developed by Vermeulen *et al.* [**Vermeulen et al., 2007**]. Quantitative models use empirical values for calculation and yield a quantitative output.

When the output of a model is definitely determined by the input parameters, it is called deterministic. Variability and uncertainty in the parameters are not taken into account. Deterministic models are sometimes also referred to as kinetic models [**Barbosa-Cánovas et al., 2005**].

The inclusion of variability and uncertainty into predictive models is done in two different ways. First, the parameters of deterministic models can be replaced by probability distributions. This approach of imposing probabilistic structures on model parameters, e.g. growth and death rates, leads to probabilistic models. If the process of growth or death itself is considered to be random, the model is called stochastic. In both approaches the output of one calculation is not definitely determined by the input parameters, i.e., repeated simulation with the same input can yield different outputs. Banks *et al.* [**Banks et al., 2009**] compared a probabilistic and a stochastic approach for modelling increase of individual sizes in populations. With a proper choice of variability, the outcome of both formulations was comparable.

5.4. Deterministic models

The first models developed for prediction of microbial growth were developed in the early 20th century [**McKellar and Lu, 2004**]. These were deterministic, and were used to describe the inactivation kinetics of pathogenic bacteria during thermal processing. In predictive microbiology the majority of models are deterministic [**McKellar and Lu, 2004**].

5.4.1. Exponential model. The simplest approach is to assume first-order kinetics for population size. Let $N(t)$ be the number of bacteria at time t, and let $N(0) = N_0$. Under the assumptions that reproduction is continuous with growth rate β, that all organisms are identical, and that the environment is constant in space and time, e.g. that resources are unlimited, change in population size can be described by the differential equation

$$\frac{dN}{dt} = \beta N. \tag{5.4.1}$$

The solution of equation 5.4.1 is exponential growth, i.e., equation 3.2.1. Negative β yields exponential inactivation. Equation 3.2.1 is also called the exponential model. It does not contain an upper limit for growth and never reaches zero.

Often kinetics are described in terms of logarithm of bacterial counts. Equation 3.2.1 yields

$$\log(N) = \log(N_0) + \beta t. \tag{5.4.2}$$

The logarithm can easily be changed according to

$$\log_c a = \frac{\log_b a}{\log_b c}. \tag{5.4.3}$$

5.4.2. Lag-exponential model. Neglecting the lag phase results in fail-save predictions [**van Gerwen and Zwietering, 1998**], i.e., bacterial counts are overestimated. Adding of a lag phase with length λ yields the lag-exponential model [**van Gerwen and Zwietering, 1998**]:

$$\log(N) = f(x) = \begin{cases} \log(N_0), & \text{for } t < \lambda, \\ \log(N_0) + \beta(t-\lambda), & \text{for } t \geq \lambda\ . \end{cases} \tag{5.4.4}$$

5.4.3. Exponential model for two sub-populations. Pruitt and Kamau [**Pruitt and Kamau, 1993**] modelled survival kinetics for a population that consists of two sub-populations of sizes $S_1(t)$ and $S_2(t) = N(t) - S_1(t)$ with specific death rates μ_1 and μ_2, respectively. Normalised population size is calculated as

$$\frac{N(t)}{N_0} = \frac{S_1(t)}{N_0} \exp(\mu_1 t) + \left(1 - \frac{S_1(t)}{N_0}\right) \exp(\mu_2 t)\,. \tag{5.4.5}$$

Logarithmic calculus of equation 5.4.5 does not yield a linear model like in simple first-order kinetics.

5.4.4. Bigelow model. One of the first applications of predictive microbiology is the description of microbial death due to heating in production of canned food. Bigelow stated the importance of canned food for supply of food for the civilian population, Army, and Navy in times of World War I [**Bigelow, 1918**]. Spoilage of tinned food due to germination of spores during storage was named as one problem of the canning industry. Two years later Bigelow and Esty [**Bigelow and Esty, 1920**] presented thermal death points for spores of several microorganisms as a result of a large amount of experimental work. In 1921 Bigelow [**Bigelow, 1921**] reported log-linearity of thermal death time curves. This was the basis for the development of the concept of D- and z-values for bacterial inactivation described below. This concept is still used and refined, e.g. in [**van Asselt and Zwietering, 2006**].

Under isothermal conditions the relationship between logarithm of bacterial counts and heating time is often considered to be linear, i.e., equation 5.4.2 holds true. Decimal reduction time D in min states how long it takes at a given temperature to reduce the bacterial population by 90%.

$$\log(N) = \log(N_0) - \frac{1}{D}t \tag{5.4.6}$$

Comparison with equation 5.4.2 yields $D = \frac{1}{\beta}$. The relation between D-value and temperature is (locally) log-linear. For non-isothermal conditions the dependencies between D-value and temperature T are described by the thermal inactivation coefficient z in min^{-1}. The z-value is the temperature increase necessary for decreasing the D-value by 90% , i.e., for reduction of the bacterial load by 90% ten times faster. The D-value at an arbitrary

temperature can be calculated from the reference D-value at one specific temperature, D_0, and the z-value according to the Bigelow model equation 5.4.7.

$$\log(D) = \log(D_0) - \frac{1}{z}T. \tag{5.4.7}$$

Usually, the D- and z-values of bacteria in food are determined by graphical methods [**Juneja et al., 2006**]. Time needed for killing off certain bacteria at a certain temperature is called thermal death time or F-value. The F-value is a multiple of the D-value.

The traditional concept for heating of canned food aims at a reduction of 12 log cycles (a reduction by a factor of 10^{-12}, or by 99.9999999999%). This approach is based on extrapolation, because in experiments only six to seven decimal reductions are determined [**Van Boekel, 2002**]. Therefore, calculated conditions for a reduction of 12 log cycles may not be precise [**Van Boekel, 2002**], [**Brannen, 1968**], [**Peleg, 1998**]. For example, the minimal temperature at which growth of *Bacillus cereus* is reported in the literature overestimates the theoretical minimal temperature systematically [**Zwietering et al., 1996**].

The concept of D- and z-values assumes instantaneous heating and a homogeneous distribution of microorganisms in the matrix. Solid food is heated from the outside inwards, and adjustments must be made to take into account cold spots in the matrix.

5.4.5. Deviations from log-linear kinetics. There are four common types of logarithmised growth and survival curves: linear curves following first order kinetics, curves with an initial lag period (growth) or a shoulder (survival), biphasic curves with tailing, and sigmoidal curves [**Barbosa-Cánovas et al., 2005**]. Deviations from first order kinetics can have many reasons. One mechanistic explanation for bacterial inactivation following first-order kinetics is the existence of a critical enzyme with first-order inactivation kinetics. But it is unlikely that a single event causes instantaneous death. The inactivation times of individual cells are not identical; they follow a distribution. There are various examples for non-linear survival curves [**Moats et al., 1971**], [**Baranyi et al., 1996a**], [**Peleg, 1998**], [**Peleg, 2003**]. A collection of experiments with varying shapes of survivor curves can be found in [**Withell and Pharm, 1942**]. Geeraerd *et al.* formulated requirements for modelling of non-linear survival curves, and they compared the performance of several deterministic primary models [**Geeraerd et al., 2000**].

During growth, a single cell doubles its mass and divides into two identical daughter cells. In a population, cells usually are at random points of growth. Cells can be synchronised in their generation periods, but this synchronisation fastly gots lost (see section 3.4). Plotting the logarithm of size of a synchronised population over time does not result in a smooth

curve, but in a series of discrete steps.

In a culture that contains more than one type of bacterium (mixed culture), the kinetics of the sub-populations are not independent. Inactivation curves often show a tailing effect, indicating that a small fraction of bacteria shows an increased heat resistance. In a pure culture sometimes a shoulder effect is observed. This may be due to torpidness of the heating process or to an energy threshold for the lethal effect [**Juneja et al., 2006**].

Common assumption is that for constant environmental conditions the lag time and the generation time are proportional [**Delignette-Muller, 1998**]. But variation of growth rate, lag phase and stationary population may be completely independent [**Dean and Sir Hinshelwood, 1966**]. Robinson *et al.* discussed in [**Robinson et al., 1998**] that the relationship between growth environment and lag time is more complex than the corresponding relationship between growth environment and maximum specific growth rate. They concluded that predictions of lag time will be less accurate than predictions of the maximum growth rate. Delignette-Muller *et al.* [**Delignette-Muller, 1998**] further suggest to assume a constant ratio anyway, despite the fact that it leads to a neglecting of possible environmental effects on the relation between lag time and generation time.

5.4.6. Logistic growth model. In a growing population increase rate β is composed of a birth rate μ and a death rate η. This yields

$$\frac{dN}{dt} = \mu N - \eta N. \tag{5.4.8}$$

There is an upper limit of population growth. The maximal number of bacteria K is called carrying capacity. It can be written as $K = \frac{\mu}{\eta}$ [**Metzler, 1987**]. Death rate may depend on population density, in the simplest case $\eta = \mu N$. This yields the initial value problem

$$\frac{dN}{dt} = \mu \left(1 - \frac{N}{K}\right) N, \tag{5.4.9}$$

$$N(0) = N_0. \tag{5.4.10}$$

Separation of variables, and partial fraction decomposition yields the solution of the initial value problem, the logistic growth model or Verhulst (Verhulst-Pearl) equation with maximum growth rate μ_{max}:

$$N(t) = \frac{N_0 K}{N_0 + (K - N_0)\exp\left(-\mu_{max} t\right)}, \tag{5.4.11}$$

$$= \frac{K}{1 + \exp\left(-\mu_{max} t + \ln\left(\frac{K}{N_0} - 1\right)\right)}. \tag{5.4.12}$$

The logistic growth model was introduced by Verhulst [**Verhulst, 1838**] and rediscovered by Pearl and Reed in 1920 [**Pearl and Reed, 1920**]. At low densities, the population growth rate is maximal and equals μ_{max}. It declines to zero for $N \to K$. The initial growth rate controls not only population growth rate, but also population decline rate (at N > K), i.e.,

organisms with a low reproduction rate die at the same slow rate. In reality, growth rate and decline rate are independent. If population numbers exceed the carrying capacity, the population growth rate becomes negative and population numbers decline.

The second derivative

$$\frac{d^2N}{d^2t} = (\mu - 2\eta N)(\mu - \eta N) N \tag{5.4.13}$$

is positive for $N < \frac{\mu}{2\eta} = \frac{K}{2}$ (the curve is concave up) and negative for $N > \frac{K}{2}$ (the curve is concave down). The inflection point is at $N = \frac{K}{2}$. This S-shaped curve is also called logistic curve. It is pictured in figure 5.3 for parameters $K = 9$, $N_0 = 2$ and $\mu = 0.2$ (growth) and parameters $K = 9$, $N_0 = 8$ and $\mu = -0.2$ (survival).
The logistic growth model is not appropriate if something preys on the population.

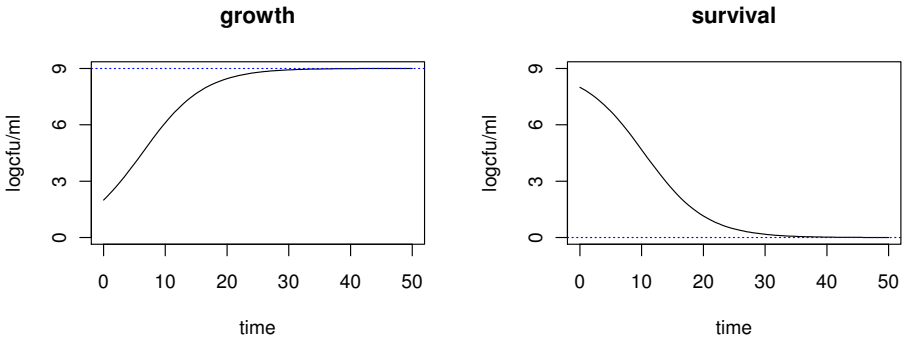


FIGURE 5.3. Growth curve (left) and a survival curve (right) calculated by the logistic growth model

The logistic growth model does not model lag time duration λ. Wolf and Venus [**Wolf and Venus, 1992**] suggest the multiplication of equation 5.4.9 with a dampening factor $1 - \exp\left(-\frac{t}{6\lambda}\right)$ for including a lag phase. With boundary condition $N(0) = N_0$, the solution yields the extended logistic growth function

$$N(t) = \frac{K}{1 + \left(\frac{K}{N_0} - 1\right) \exp\left(-\mu\left(t + 6\lambda \exp\left(-\frac{t}{6\lambda} - 1\right)\right)\right)}. \tag{5.4.14}$$

Zwietering *et al.* [**Zwietering et al., 1990**] modified the logistic model for biological relevant parameters. Instead of $N(t)$, relative population size $y(t) = \ln \frac{N(t)}{N_0}$ is calculated. With lag time duration λ, growth rate μ, and maximum relative population size $A = \ln \frac{K}{N_0}$, $y(t)$ is approximated by

$$y(t) = \frac{A}{1 + \exp\left(2 + (\lambda - t)\frac{4\mu_{max}}{A}\right)}. \tag{5.4.15}$$

5.4.7. Gompertz model. In 1925 Gompertz [**Gompertz, 1825**] presented a model for calculation of human mortality rates in actuarial science. He assumed an exponential dependency of growth rate μ and population size. With maximum relative population size $A = \ln \frac{K}{N_0}$, time at which the absolute growth rate reaches its maximum t_μ, and relative growth rate ν at time t_μ, population size is approximated by

$$\log N(t) = N_0 + A \exp\left(-\exp\left(-\nu\left(t - t_\mu\right)\right)\right). \tag{5.4.16}$$

Equation 5.4.16 is one possible parameterisation of the Gompertz model. Modification of the Gompertz model for biological relevant parameters is given in [**Gibson et al., 1988**] and [**Zwietering et al., 1990**].

There are several modifications of the Gompertz model. One modification allows for an additional increase in cell numbers after a first stationary phase [**Kahm et al., 2010**], i.e., it models diphasic growth (diauxie, see figure 3.2). Gil *et al.* [**Gil et al., 2006**] introduced a modified Gompertz model that models microbial inactivation at changing temperature.

5.4.8. Modelling of lag phase duration. None of the previously described models include a lag phase. Lag phase and growth rate are nearly independent of each other. Length of lag phase highly depends on a combination of history of the cells and environmental conditions (see section 3.2 and assumption 5 in section 6.3). Hence, effects of the previous environmental conditions on bacterial response to current conditions must be taken into account for prediction of lag phase duration [**Muñoz Cuevas et al., 2010**].

Baranyi *et al.* (see [**Baranyi and Roberts, 1994**], [**Baranyi et al., 1993a**], [**Baranyi et al., 1993b**]) and Hills *et al.* (see [**Hills and Mackey, 1995**], [**Hills and Wright, 1994**]) developed mechanistic models that include a lag phase. The model published by Baranyi *et al.* covers lag and log phase. Mirroring yields a model of inactivation [**Baranyi et al., 1996a**]. It includes Michaelis-Menten kinetics of a critical substance responsible for a bottleneck of growth, and influences of environmental conditions before inoculation. With maximum growth rate μ_{max}, concentration of critical substance in units per cell $P(t)$, Michaelis-Menten constant C_p and $q(t) := \frac{P(t)}{C_p}$, the Baranyi model is given as set of two differential equations 5.4.17 and 5.4.18:

$$\frac{dq}{dt} = \mu_{max} q(t), \tag{5.4.17}$$

$$\frac{dN}{dt} = \mu_{max} \left(\frac{q(t)}{q(t)+1}\right)\left(1 - \frac{N(t)}{K}\right) N(t). \tag{5.4.18}$$

The outcome of the Baranyi model depends explicitly on time. A model is called autonomous if its differential equations do not contain coefficients that are explicit functions of time that means they form an autonomous (time-invariant) system. The Baranyi model is non-autonomous. Vadasz and Vadasz [**Vadasz and Vadasz, 2007**] presented an autonomous version, with transient growth depending on initial cell concentration and

initial growth rate.

5.4.9. Other approaches. Neural networks have been applied to model bacteria l growth and death kinetics, too. Basheer and Hajmeer constructed an artificial neural network for modelling growth of *Shigella flexneri* (see [**Basheer and Hajmeer, 2000**]). The growth/no growth interface of *Staphylococcus aureus* was modelled with a neural network by Fernández-Navarro *et al.* [**Fernández-Navarro et al., 2010**]. Panagou [**Panagou, 2008**] used a neural network approach to model survival of *Listeria monocytogenes* in greek soft cheese.

5.4.10. Secondary models. Secondary models describe the relation between the environmental conditions and the parameters of the primary models. To describe the dependencies of inactivation rate k and heating temperature T in °K, conventionally the Arrhenius equation

$$\log(k) = \log(A) - \frac{E_a}{RT} \tag{5.4.19}$$

is used [**Juneja et al., 2006**], where A is the frequency factor, E_a is the activation energy in kJ, and $R = 8.314 \times 10^{-3} \frac{\text{kJ}}{\text{mol}}$ is the gas constant. According to [**Juneja et al., 2006**], a combination of first-order kinetics and equation 5.4.19 performs well in describing thermal inactivation of microorganisms in food.

Another secondary model is the Ratkowsky model or square root model [**van Gerwen and Zwietering, 1998**]. It assumes a linear dependency of the square root of specific growth rate and conditions, e.g. temperature. With regression coefficient b, experimental temperature T, minimum temperature for cell growth T_{min}, experimental pH value pH, minimum pH value for growth pH_{min}, experimental a_w value a_w, and minimum a_w value for growth a_{wmin}, an extended version of the square root model is given by

$$\sqrt{\mu} = b\,(T - T_{min})\,\sqrt{(a_w - a_{wmin})\,(pH - pH_{min})}. \tag{5.4.20}$$

The square root model usually performs better than the Arrhenius model [**Stannard et al., 1985**], [**Li et al., 2008**]. It does not take into account interactions between the conditions.

5.5. Probabilistic and stochastic models

A common use of deterministic models is calculation of best, average, or worst case scenarios. In figure 5.4 outputs of a best case, an average case and a worst case scenario of exponential bacterial growth are pictured. Figure 5.5 shows the output of a probabilistic calculation of exponential bacterial growth. Output from probabilistic and stochastic models yields information about the probability distribution of possible outcomes. Probabilistic models include variability and uncertainty by means of parameters that follow probability distributions. Stochastic models assume a random process

of procreation or dying, see section 5.3.

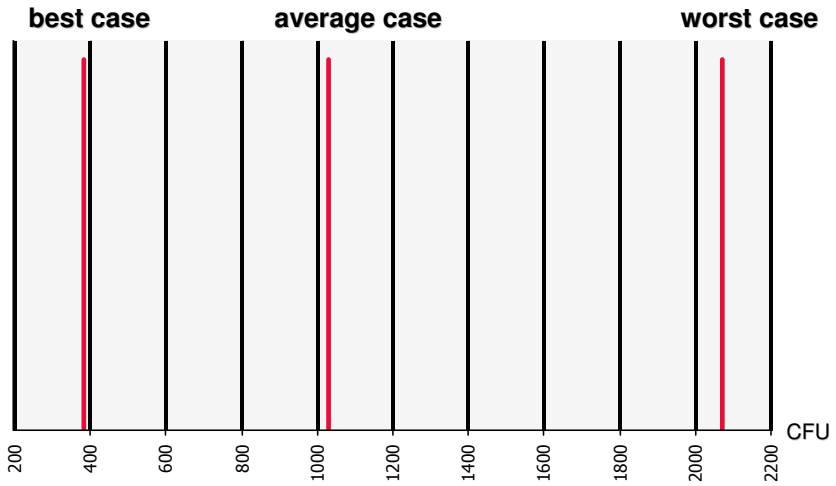


FIGURE 5.4. Comparison of best case, average case and worst case simulation

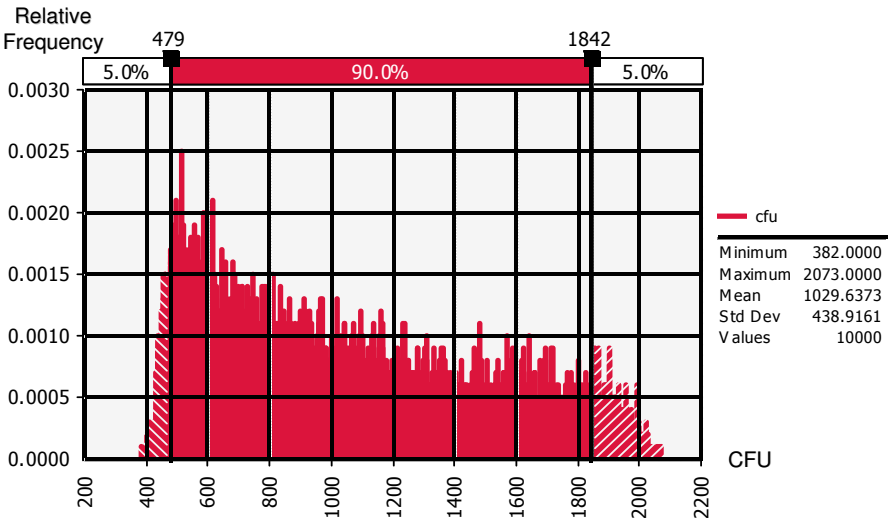


FIGURE 5.5. Example for a probability distribution as output from a probabilistic model

Deviations from log-linear kinetics may be due to variability of individuals in a population. Growth rate is not a constant, but was shown to follow a probability distribution (see assumption 2 in section 6.3). Probability of dying differs between individuals [**Van Boekel, 2002**].

Several authors confirm that both variability and uncertainty are very important in predictive microbiology and that it is essential to use probabilistic models [**Delignette-Muller and Rosso, 2000**], [**Standaert et al., 2007**], [**Standaert et al., 2007**], [**Koutsoumanis, 2008**], [**Van Boekel, 2002**], [**Membre and Lambert, 2008**]. The benefit of a probabilistic or stochastic model is that it combines probability distributions of the model's parameters for calculation of a probability distribution of possible model outcomes. Therefore, use of probabilistic or stochastic models in predictive microbiology supports a more profound risk assessment of food or feed supply chains than deterministic models. For a sufficient number of iterations of a probabilistic or stochastic model, the mean of the output equals the average case of its deterministic counterpart. Minimum and maximum correspond to deterministic best and worst case, respectively. Until now, there are only few probabilistic and stochastic models for growth and survival used in predictive microbiology.

There are several approaches of probabilistic and/or stochastic modelling of bacterial population kinetics.

5.5.1. Weibull model. The response to a harmful influence can be assumed to differ between individuals of a population, i.e., survival times follow a probability distribution. Then the cumulative distribution function of survival times describes the fraction of the population still alive. One approach of modelling survival curves is to consider them to be cumulative distribution functions of resistance or sensitivity [**Peleg, 1998**], [**Peleg, 2003**]. Those models include probability distributions, but their output is deterministic. One example is the widely used Weibull model. It assumes that resistance to stress, and therewith death times, follow a Weibull distribution [**Corradini et al., 2010b**], [**Coroller et al., 2006**], [**Van Boekel, 2002**], and it models survival curves as cumulative distribution curves of cell death times.

The Weibull model is an empirical model. It can be written in different ways, for example as power law model [**Peleg, 2003**], [**Peleg, 2006**]

$$\log \frac{N(t)}{N_0} = -bt^a. \tag{5.5.1}$$

with dimensionless shape parameter a, and scale parameter b in time units. These coefficients depend on the microorganism, the history of the cells, the matrix, and the inactivation agent, e.g. temperature, or the concentration of a substance. For $a = 1$ it yields first order kinetics. For $a < 1$ the curve is upper concave, and for $a > 1$ it is downward concave.

The Weibull model can be fitted to many survival curves [**Peleg et al., 2008**], [**Van Boekel, 2002**], [**Hassani et al., 2006**], and it can be adapted for modelling survival under non-isothermal conditions [**Chen et al., 2007**], [**Hassani et al., 2006**]. Coroller *et al.* [**Coroller et al., 2006**] modelled kinetics of sub-populations by mixing

two Weibull distributions.

5.5.2. Probabilistic growth/no growth model. Ratkowsky and Ross [**Ratkowsky and Ross, 1995**] developed a probabilistic growth/no growth model that takes into account temperature, pH value, a_w value, and sodium nitrite concentration. It is given as

(5.5.2)
$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \ln(T - T_{min}) + b_2 \ln(pH - pH_{min}) + b_3 \ln(a_w - a_{wmin}) + b_4 \ln(NO_{2max} - NO2),$$

with regression coefficients $b_0, \cdots, b_4$, experimental temperature T, minimum temperature for cell growth T_{min}, experimental pH value pH, minimum pH value for growth pH_{min}, experimental a_w value a_w, minimum a_w value for growth a_{wmin}, experimental nitrite concentration (in ppm) NO_2, and maximum nitrite concentration for growth (in ppm) NO_{2max}. This model is log-linear.

5.5.3. Stochastic approaches. Pin and Baranyi determined the distribution of division intervals from experiments done with single cells (see [**Pin and Baranyi, 2005**]). Population kinetics were modelled as result of multiplication of single cells. The stochastic model generated from measurements of the first few individual division times was able to predict bacterial growth at low concentrations. Kilsby *et al.* [**Kilsby et al., 2000**] modelled inactivation kinetics by assuming a distribution of inactivation time.

A process that evolves randomly in time can be mathematically described as a stochastic (or random) process. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A stochastic process is a set $\{X_t : t \in T\}$ of random variables X on Ω with state space S that is indexed by a set T ('time'). A stochastic process in which the next state depends solely on the current state is called a Markov process. One Markov process on non-negative integers is the Poisson process. It is a counting process, and the number of events until time t follows a Poisson distribution. The interested reader may refer to [**Klenke, 2008**] or [**Wengenroth, 2008**] for a brief introduction.

The Poisson process is the prototype of a pure birth process [**Taylor and Karlin, 1998**]. Takahashi [**Takahashi, 1968**] used a Poisson process to describe the transition of phases in the cell cycle of eucaryotic cells. Assuming that the chances of reproduction and mortality are independent of the previous history of the individual cell, including the time since ´birth´, kinetics follow a discontinuous Markov process [**Kendall, 1948b**].

The stochastic counterparts of the deterministic exponential model are the Yule model for growth, and the Pure Death model for survival. These

models were developed by Yule and Kendall at the beginning of the 20th century [**Yule and Kendall, 1950**]. Both models assume that all individuals in the bacterial population are identical concerning growth or death, respectively, and that procreation or death of different cells is independent and follows a Poisson process. Unlike the exponential model that can be applied to both growth and death curves, its statistic counterparts assume different processes for procreation and death.

5.5.4. Yule model. Let γ be the expected number of offspring per time step. Then the number of offspring of one single cell $M(t)$ can be described by the geometric distribution:

$$M(t) = geometric\left(\exp(-\gamma t)\right).$$

Now quantity of the population at some time t is calculated as sum of initial bacterial count and numbers of offspring of all cells in the initial population. The sum of a finite set of independent geometric distributions follows a negative binomial distribution [**Vose, 2008**], therefore, the sum of numbers of offspring is negative binomially distributed:

$$N(t) = N_0 + \sum_{i=1}^{N_0} geometric\left(\exp(-\gamma t)\right), \tag{5.5.3}$$

$$= N_0 + negbin\left(N_0, \exp(-\gamma t)\right). \tag{5.5.4}$$

Equation 5.5.4 is called the Yule model.

5.5.5. Pure Death model. With expected death ζ at time t the probability for surviving of an individual until time t is calculated as

$$p(t) = 1 - \exp\left(-\zeta t\right).$$

For an initial population of N_0 bacteria at time $t = 0$, the number of bacteria that are left at time t follows a binomial distribution:

$$N(t) = bin\left(N_0, \exp\left(-\zeta t\right)\right). \tag{5.5.5}$$

Equation 5.5.5 is called the Pure Death model.
Growth and death rate for application of Yule model and Pure Death model must be calculated from experimental data. In both models these rates are assumed to be constants. This assumption is false, because growth and death rate are probabilistic variables.

5.5.6. Other probabilistic and stochastic approaches. Feller modelled bacterial growth and survival with a Markov process [**Feller, 1939**], too. In 1948 Kendall presented extensions of Feller´s model with birth and death rates as functions in time [**Kendall, 1948a**], or with growth and death rates following probability distributions [**Kendall, 1948b**], respectively. In 1966 Takahashi [**Takahashi, 1966**] published a stochastic model for eucaryotic cells similar to the birth process of Kendall. The duration of the phases of the cycle was simulated by probability distributions. Corradini *et al.* recently developed a stochastic model for activation and inactivation of dormant spores of certain *Bacillus* spp. [**Corradini et al., 2010a**].

Modelling of inactivation of individual cells as a stochastic process was also done in [**Corradini et al., 2010b**] and [**Horowitz et al., 2010**].

In 1962 Koch and Schaechter [**Koch and Schaechter, 1962**] introduced a probabilistic model for cell division. It is deterministic for single cells, but includes variability in the distribution of cell sizes at division. Delignette-Muller and Rosso [**Delignette-Muller and Rosso, 2000**] included the variability of growth among strains of *Bacillus cereus* in pasteurised milk into another probabilistic model. Only temperature as environmental factor influencing growth was considered. Biological parameters were characterised by probability distributions. Biological variability had a great impact on the accuracy of the results.

Cao *et al.* [**Cao et al., 2010**] presented a new stochastic primary model. Using non-parametric and bootstrapping methods, prediction bands were developed that cover a prefixed confidence interval instead of growth curves. This inspired our use of sets of appropriate series of measured values to estimate the parameters of a log-normal distribution for the bacterial counts.

Stochastic modelling of the lag phase by a stochastic birth model was done by Baranyi [**Baranyi, 2002**].

The `NPMPM` follows a new approach to take into account for variability and uncertainty. None of the probabilistic or stochastic models presented in this section is used. Instead, deterministic models are fitted to the data.

5.6. Example of a risk assessment for toxin inactivation

Miscellaneous predictive models are used for optimisation of food processing, i.e., in designing production processes that yield food without pathogenic organisms, with a long shelf life, and with certain organoleptic characteristics. Predictive models are also used in risk assessment for estimating outcome and impact of an intentional or unintentional contamination of the food supply chain with bacteria or toxins.

One widely regarded risk assessment was published in 2005 by Wein and Liu: ´Analyzing a bioterror attack on the food supply: The case of botulinum toxin in milk´ [**Wein and Liu, 2005**]. They computed a probability distribution for the outcome of an intentional contamination of raw milk with botulinum toxin. In this section the modelling approach of Wein and Liu is described as an example of a risk assessment for contamination of the supply chain for fresh milk from cow to consumer, with estimated number of affected people as endpoint. This example highlights the procedure of modelling a process chain as succession of process steps with different effects on toxin concentration.

Wein and Liu investigated an imaginary bioterror attack in which the purified toxin is introduced into the milk supply chain. Presence of

toxin producing microorganisms is excluded. Therefore, only dilution and inactivation had to be taken into account. Inactivation is due to heating during pasteurisation. Wein and Liu roughly estimated the proportion of inactivation, because there are no sufficient data concerning the effect of heat pasteurisation on botulinum toxin in milk. A dilution may occur at different points in processing by mixing of milk from different tanks or trucks. Effects of manufacture on the concentration of botulinum toxin were linear, and therefore, use of models introduced previously in this chapter was not necessary.

Concentration of botulinum toxin during processing depends amongst others on the amount of toxin introduced into the process chain. The milk from dairy trucks is pumped into silos that are simultaneously filled and drained into the pasteurisation facility. Therefore, the possible toxin concentration in the silo depends on the point in time during the filling and replenishment intervals at which the contaminated milk is pumped into the tank. Because this point in time is random, the toxin concentration in the silo follows a probability distribution. After pasteurisation the milk is pumped into holding tanks. In this tanks the toxin concentration depends on the time interval during which the milk was drained from the silo. Therefore, it follows a probability distribution, too.

Average packing size, proportion of children and adults that drink milk, average portion size and the dose-response curve were taken into account for calculation of the number of poisoned people. Several assumptions made in model construction are due to lack of data. The interested reader may read the supplement to the paper of Wein and Liu [**Wein and Liu, 2005**] for further information.

CHAPTER 6

Assumptions, Material and Methods

In the previous chapters 2-5 the knowledge necessary for modelling bacterial growth and decay is presented. Now the assumptions made in development of the NPMPM, and the components and methods used in modelling and model validation are introduced. First, the method of Monte Carlo simulation is described. After that, the assumptions the NPMPM is based on are specified and discussed. This is followed by a characterisation of probability distributions contained in the NPMPM. The statistical tests used for data analysis and model validation are described, and the experimental data used for fitting are presented. Then the programming language is noted. Now the methods used in performance evaluation are established. Finally, the difficulties in communicating the output of probabilistic model are depicted.

6.1. Monte Carlo method

The NPMPM uses a Monte Carlo method. This method is an approach for computing results of models containing probability distributions. Random values are drawn from the probability distributions, and a result is computed from this values. Such a calculation is called an iteration. Iterations are repeated, and results are collected, until a stop criterion is fulfilled. Then the set of calculated values, i.e., a distribution of results, is returned. One calculation of a distribution of results in a succession of iterations is called one simulation.

The term Monte Carlo method was introduced by Metropolis and Ulam in 1949 in [**Metropolis and Ulam, 1949**]. The method as such was used before in time of World War II in development of the atomic bomb [**Law, 2007**], [**Gentle, 2003**]. The Monte Carlo method is not a single method, but a class of numerical approaches. A deterministic algorithm is stocked with probability distributions for input and parameters. In every iteration of the Monte Carlo simulation random values are drawn from the probability distributions. Then a deterministic computation is done with the obtained input and parameter values. Hence, for every iteration the calculation is done for one possible scenario. The results form a frequency distribution that approximates the probability distribution of possible output values. The Monte Carlo method is fast and easy, but it yields only approximations. There is no need to use it, if a closed form of the solution can be calculated. The interested reader may refer to [**Gentle, 2003**] for further information on this subject.

Monte Carlo simulation can be done with a lot of different software; e.g. @Risk, ModelRisk, Berkeley Madonna, and Crystal Ball are special designed for this task. Algorithms using the Monte Carlo method can be programmed in many programming languages, e.g. matlab or R.

'To calculate the probability of a successful outcome of a game of solitaire (we understand here only such games where skill plays no role) is a complete intractable task. On the other hand the laws of large numbers and the asymptotic theorems of the theory of probabilities will not throw much light even on qualitative questions concerning such probabilities. Obviously the practical procedure is to produce a large number of examples of any given game and then to examine the relative proportion of success.´

[Metropolis and Ulam, 1949]

In every iteration of a simulation random values are drawn. Random numbers from a certain probability distribution are calculated using uniformly distributed random numbers u_i that are easily calculated. One possibility of transformation is classical Monte Carlo sampling, e.g. to use (an approximation of) the inverse function of the cumulative distribution function of the probability distribution of interest, $F^{-1}(u)$. A computer cannot generate true random numbers without the connection to an external device. For generation of pseudo-random numbers a seed as starting point is needed. For most purposes the use of pseudo-random number is sufficient.

The method of classical Monte Carlo sampling leads to uniformly distributed random numbers u_i, and to clusters of the $F^{-1}(u_i)$. If the number of iterations n used in the Monte Carlo simulation is previously fixed, the probability distribution can be split in n intervals with equal probability, and then every interval is sampled once. This is called Latin hypercube sampling. It yields a plausible sample of the values $F^{-1}(u_i)$. The interested reader may refer to [**McKay et al., 1979**] for further information.

In stratified sampling proportions of the total sample are taken from specified regions of the sample space to ensure that all regions are covered. In importance sampling some sample regions are sampled more heavily. The sample density may change continuously. If some measurements carry more information than others, an adjustment of sample density is reasonable. When several factors are to be investigated simultaneously, a full factorial design that includes all possible combinations of the factors, leads to a large number of trials. Hence, a subset is chosen to expose information about the most important features of the problem studied. This approach is called fractional factorial design; it allows for studying of several factors, while keeping the number of experiments small. The interested reader can find an overview of common sampling methods used in Monte Carlo simulation in [**Gentle, 2003**].

Sampling methods that base on a fixed number of iterations can not be used in the `NPMPM`, because the number of iterations in one simulation is adaptive.

Additionally, there is no previous information about sample regions. An approach suiting the special conditions of the NPMPM applied to bacterial growth and decay was developed.

If the distribution of the output is already known before the simulation, there are different possible stop criteria, e.g. the length of a 95% confidence interval. The interested reader may consult [**Law, 2007**] for an overview. The probability distribution of the output of the NPMPM depends on the data and can not be specified, because the calculated values each derive from lognormal distributions with slightly different mean and standard deviation. So another approach is needed.

In the NPMPM an *a posteriori* error estimator determines if the number of iterations is sufficient for providing a given accuracy of the output. First, the minimal number of iterations is set to 100 to prevent too small sample sizes. From the 100th iteration on the arithmetic mean $\bar{x}$ of all heretofore obtained cfus is calculated. In the ith iteration, $i \geq 100$, the difference of the current and the previous mean, $\bar{x}_i - \bar{x}_{i-1}$, is calculated. If this difference is not greater than a previously fixed error bound, the simulation stops. The error bound is relative to the mean $\bar{x}_i$. The factor for calculation of the error bound is given with the input. The default value is 1% . The pseudocode is given in algorithm 1.

Algorithm 1 Stop criterion for one simulation with the NPMPM

`numberiter(i,error,mean)`

Input: number of iterations i, error bound *error*, vector *mean* with mean of first j *final-cfu/ml*s as jth component

if i $<$ 100 **then**
  stop = FALSE
else
  if $|mean(i) - mean(i-1)| > error \cdot mean(i)$ **then**
    stop = FALSE
  else
    stop = TRUE
  end if
end if

Output: stop

Values of cfu $= 0$ and cfu $= 10^9$ with $p(0) = p(10^9) = \frac{1}{2}$ yield a worst case fluctuation of the mean of cfus calculated. In this case the number of iterations with outcome 10^9 follows a binomial distribution, e.g. it is a random variable $X \sim Bin(n, p)$ with total number of iterations n and probability $p = \frac{1}{2}$. The expected value of X is $E(X) = \frac{n}{2}$ and the variance is $Var(X) = \frac{n}{4}$. The mean of the output-cfus of n iterations can be represented as the random variable

$$Y = \frac{10^9}{n} X.$$

With the linearity of the expected value it holds true that

$$E(Y) = \frac{10^9}{2}.$$

The expected value does not depend on n, and for the variance of Y it holds true that

$$Var(Y) = \frac{10^{18}}{4n}.$$

Hence, the variance of the mean of the calculated cfus is in $O(\frac{1}{n})$. Therefore, the *a posteriori* estimator described in algorithm 1 is reasonable as a breakpoint, because the variance of the mean decreases with increasing number of iterations.

The error estimator compares two results with different accuracy to get an approximation of the true error. A similar approach is used in adaptive numerical integration methods, where the difference between the results of the applied integration method and a method of lower order is used as an estimator for the true approximation error (see for example [**Deuflhard and Bornemann, 2002**]). The error estimator is also a form of bootstrap method, because it compares the mean of the whole sample of cfu to the mean of a sub-sample without replacement (´jackknife´).

6.2. Experimental data

The experimental data used for fitting the `NPMPM` were extracted from the ComBase [**Anonymous, a**]. The ComBase is a large database containing microbial response data in laboratory growth medium and food environments. The associated ComBase Predictor provides predictive tools on microbial responses to food environments.

The data used for running the model is stored in tables in a MySQL database [**Anomymous, 2006**]. The data needed for the simulation must be transferred to data frames in the global environment of the R session. Figure 6.1 gives an overview of the links between tables in the database, input of the model, and calculation.

The structure of the two tables in the MySQL database pictured in figure 6.2 is similar to the structure in the ComBase. Not all information provided by the ComBase is used in this thesis. The environmental conditions that have the main influence on growth and death kinetics are the temperature, the pH value and the a_w value (see section 3.6). Fortunately these are the information from the ComBase congruent with the accessible parameters of the process steps.

In one table the name of the microorganism and specification like stem or biovar are stored with the temperature and the content of CO_2. The pH and a_w value and a free-text field can provide information about matrix characteristics. If no exact values are available, the assumed values of temperature, pH and a_w value can be noted down. Every series of measured

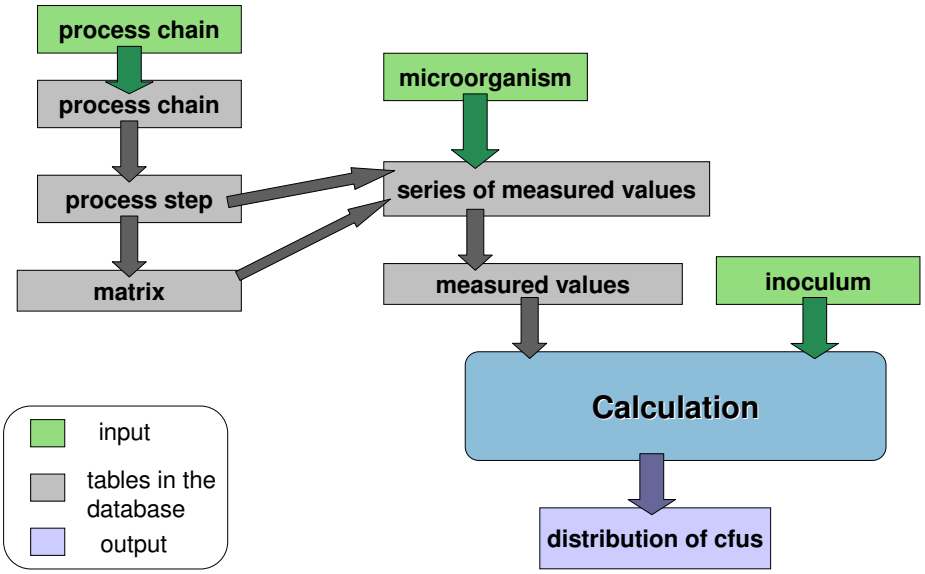


FIGURE 6.1. Overview of the links between the information stored in the database, input, algorithm and output

values has a unique id that links the first table with the second table that provides the raw data. The second table contains the time in hours, and the measured values at the specified points in time as decadic logarithm of the cfu. A graphical representation of this connection can be found in figure 6.2.

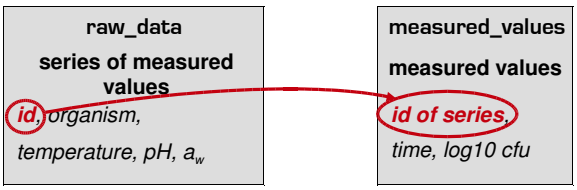


FIGURE 6.2. Overview of the links between the tables with the experimental conditions and the measured cfus, respectively

In this thesis a contamination of the milk supply chain with *Listeria* spp. is used as an example. Hence, only the series of measured values for *Listeria* spp. where downloaded from the ComBase. In table 6.1 and figure 6.3, a summary of the temperature range of this data is pictured. A summary statistics is provided in section E.1.

Some of the series of measured values were discarded, because there were multiple values at time zero. In the NPMPM the cfu are assumed to be log-normally distributed, see assumption 2. Hence, the logarithm of the cfu was tested for normality with the Shapiro-Wilk test. Sets of series

TABLE 6.1. Summary of the temperature distribution in the series of measured values for *Listeria* spp.

Minimum	1st Quantile	Median	Mean	3rd Quantile	Max
-20.00	5.00	12.00	20.34	28.00	90.00

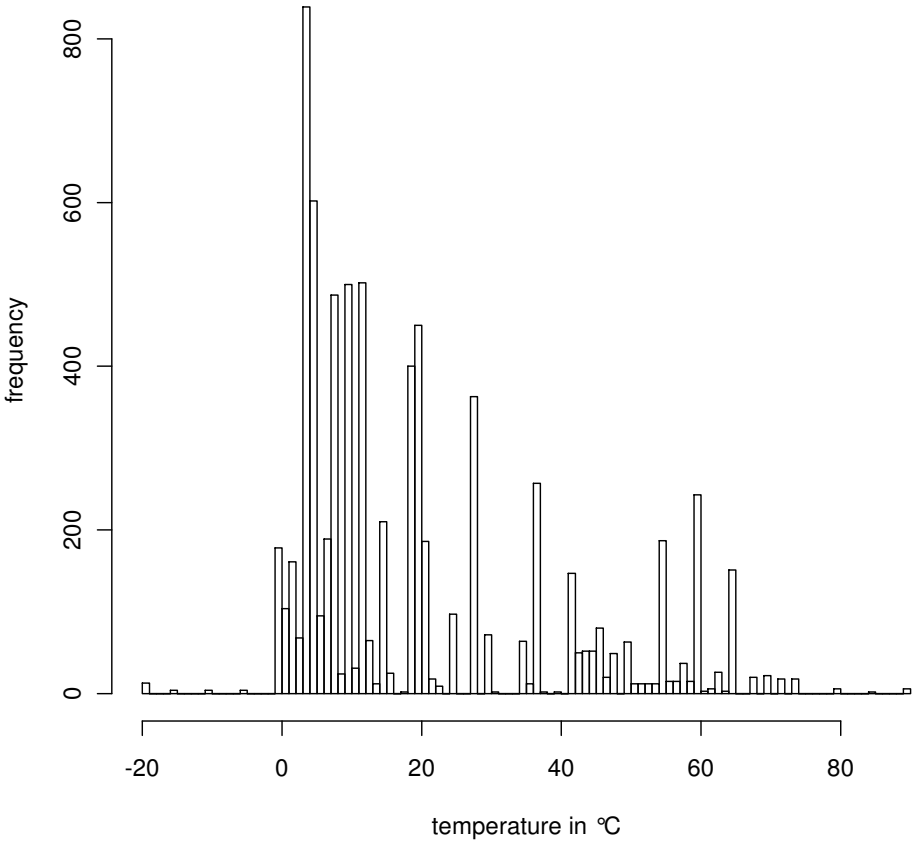


FIGURE 6.3. Histogram of the temperature of the series of measured values for *Listeria* spp.

of measured values obtained at the same temperature and with the same inoculum size were chosen. Only sets of at least 30 series of measured values were tested. The number of time points at which the tests were performed are listed in table 6.2. Unfortunately, in 77 of 80 sets the null hypothesis (normally distributed set) was rejected. This may be partly due to the broad range of pH and a_w values at most temperatures, pictured in figures 6.4 and 6.5.

Information about the dairy process chains was extracted from the literature, and some details are based on expert opinion. A process chain can be described as succession of process steps with defined combinations of time, temperature and pressure. The parameters of the process steps,

TABLE 6.2. Temperatures with more than 30 series of measured values with the same inoculum size, and the number of time points at which the cfu are tested for log-normality. (4 and 4a, and 20 and 20a, are both done at 4°C or 20°C, respectively, but with different inocula)

Temperature in °C	**number of time points**
0	10
2	7
4	12
4(a)	8
7	7
10	12
15	10
20	5
20(a)	7

the characteristics of the dairy staple, intermediate milk products and the end product are stored in different tables in a MySQL database. The tables, and the links between them, were developed to guarantee unique assignments, and to avoid redundant information. Linking is done by unique ids. All tables contain one column for references and one for free-text notes. Not all of the information stored in the tables is used by the NPMPM.

Growth and death of microorganisms are influenced by the environmental conditions. The biggest influence have temperature, pH value and aw value (see section 3.6). The influence of pressure on growth and death kinetics is neglected in the NPMPM, because the data in the ComBase were obtained under atmospheric pressure. Nevertheless, pressure data can be stored in the MySQL database. Time and temperature are technical parameters of one process step, whilst pH value and a_w value are characteristics of the matrix that may be present in other process steps, too. Therefore, the information is stored in different tables.

One table contains the process chains with a name and the id of the end product. In another table the process parameters for the process steps are stored as minimum and maximum of time, temperature, and pressure (if available). The id of the (intermediate) milk product associated with the process step links to the table containing matrix characteristics (see figure 6.6). The id of the corresponding process chain is used for linking, and the position of the process step in the process chain is stored, too. Additionally, a factor to take into account for a possible dilution or accumulation of bacteria in the process step is memorised. In the table containing the matrix characteristics, the minima and maxima of pH and a_w value, water, salt and fat content are stored. Free-text fields for information about preservatives and synonyms are provided.

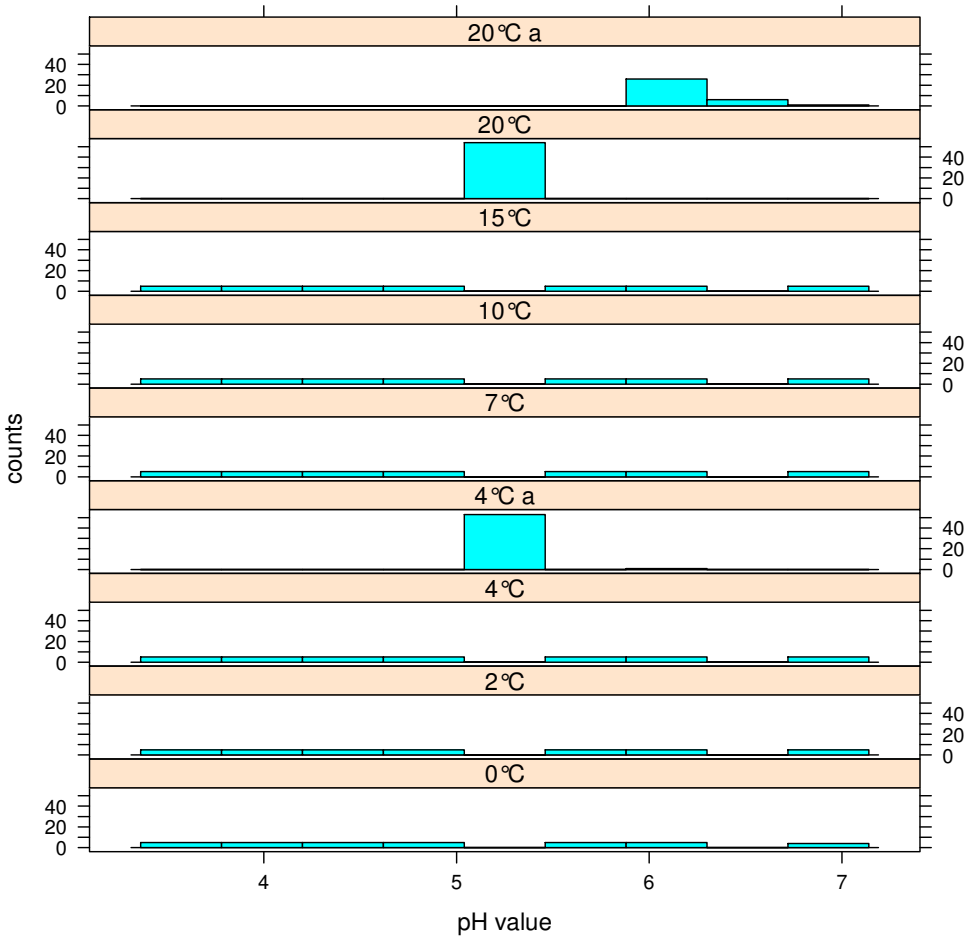


FIGURE 6.4. Histograms of pH values for the sets of series of measured values with similar temperature and inoculum size, sorted by temperature(4 and 4a, and 20 and 20a, are done at 4°C and 20°C, respectively, but with different inocula)

The references are stored in a table containing an id, entry type, author, title, year of publication, journal, publisher, editor, chapter, pages, URL, and how it was published. This structure is similar to the one used by LaTeX in the bibfile.

6.3. Assumptions

A model is always a simplification of the reality, because the entanglements between the causalities included are less complex than the causalities in reality. The preconditions made in development of a model must be satisfied when using it. Otherwise, the output of the model is useless. The default methods in the `NPMPM` are based on the following assumptions.

Assumption 1. *Growth and death kinetics are exponential.*

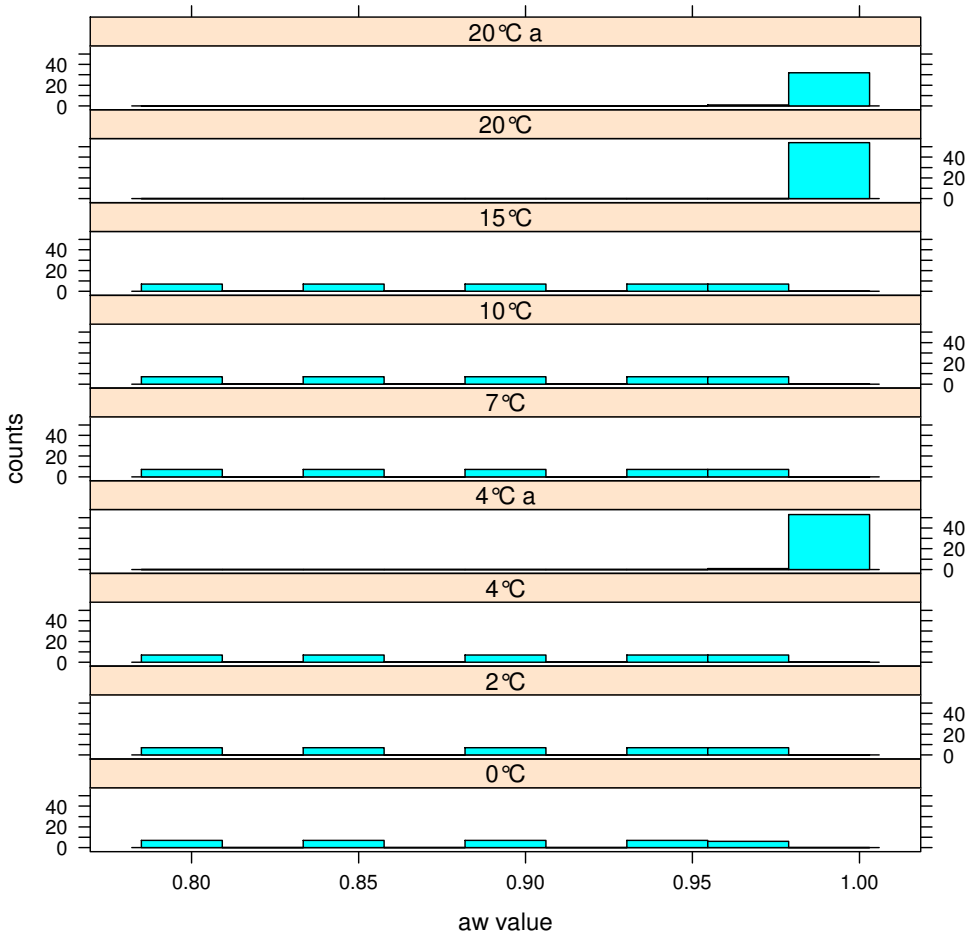


FIGURE 6.5. Histogram of the a_w values for the sets of series of measured values with similar temperature and inoculum size, sorted by temperature (4 and 4a, and 20 and 20a, are done at 4°C and 20°C, respectively, but with different inocula)

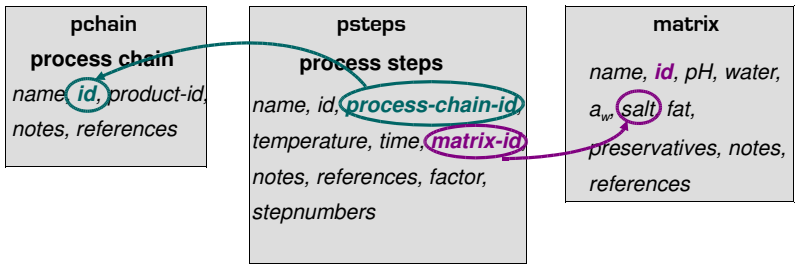


FIGURE 6.6. Overview of the links between the process chain, the process steps and the associated matrices

Exponential growth and death of all bacteria present in the population is the simplest assumption for kinetics, see section 3.2. For such kinetics the logarithm of population size depends linearly on time. The resulting log-linear curve can be approximated by ordinary linear regression, see section 5.1. Approximation by ordinary linear regression will be referred to as global linear interpolation. Global linear interpolation is the default interpolation method currently provided by the `NPMPM`.

Lag phase, stationary phase, or other deviations from log-linearity can not be approximated by global linear interpolation. A local (piecewise) linear interpolation usually yields a better approximation. Local linear interpolation corresponds to an interpolation of the *logcfu/ml* with splines of order one. It is the second interpolation method provided by the `NPMPM`. *Calculated logcfu/ml* at time s is computed from two adjacent measurement points by a simple linear equation. These two points are the *cfu/ml* at maximal measurement time smaller than s, and *cfu/ml* at mimimal measurement time bigger than s. Previous population kinetics is taken into account only indiretly as difference between *start-cfu/ml* and *cfu/ml* at maximal measurement time smaller than s. Local linear interpolation and other splines yield exact values at grid points.

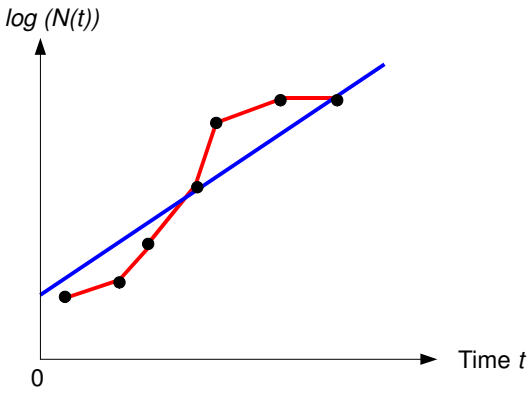


FIGURE 6.7. Comparison of approximation with global linear interpolation (blue line) and local linear interpolation (red line) for experimental data (black dots) from a population with a sigmoid growth curve

Assumption 2. *For one fixed parameter combination the bacterial counts follow a log-normal distribution.*

Several authors assume log-normal distributions for bacterial counts, e.g. [**Robertson, 1932**], [**Delignette-Muller and Rosso, 2000**]. In a growing cell population generation times of individual cells vary over a wide range [**Banks, 1994**], [**Schaechter et al., 1962**], there may also be a day-to-day variation in the mean generation time [**Kelly and Rahn, 1932**]. Under the assumption of exponential growth, the cfu of populations grown under the

very same conditions for a fixed time t follows nearly a log-normal distribution. Dean and Sir Hinshelwood [**Dean and Sir Hinshelwood, 1966**] found that the generation times are approximally normally distributed (see figure 6.8). Powell and Errington [**Powell and Errington, 1963**] reported that the growth medium influences the width of dispersion of generation times. Generation time is the time needed for growing by factor 2, and hence, a multiple of the time needed for growing by any other factor. Application of the decadic logarithm to equation 3.2.1 yields

$$\log N(t) = \log N_0 + \beta \cdot t \tag{6.3.1}$$

$$= c_1 + c_2 \cdot \beta. \tag{6.3.2}$$

With frequency β of growing by a factor of e following a normal distribution, $\log N(t)$ is normally distributed, and therefore, the cfus follow a log-normal distribution.

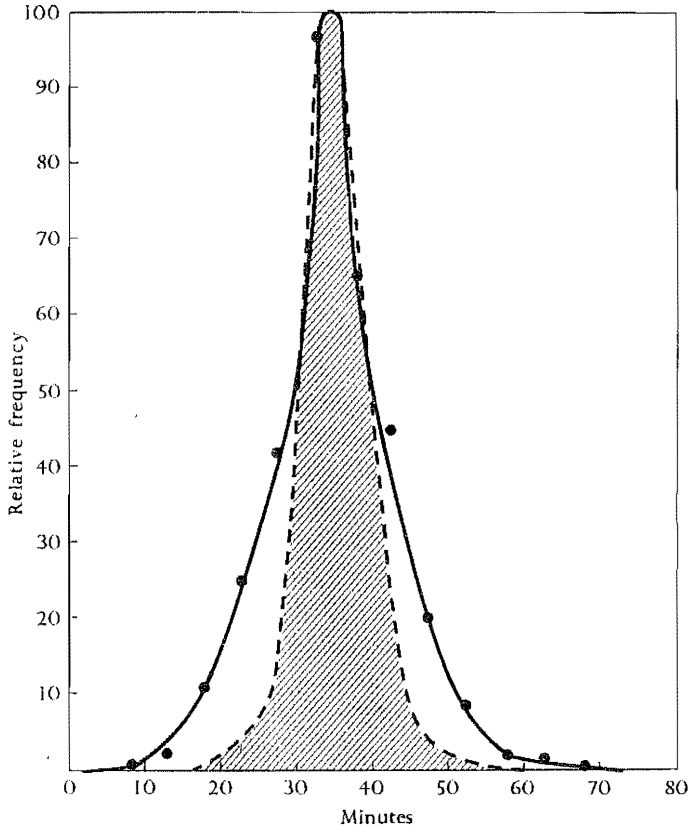


FIGURE 6.8. Distribution of generation times - the crosshatch area marks a normal distribution with the same mean, from [**Dean and Sir Hinshelwood, 1966**]

Assumption 3. *There are lower and upper bounds for bacterial counts per volume. The maximal bacterial count is* 10^9 *cfu per ml. More than one bacterium per ml must be present for growth.*

Under optimal growth conditions there are up to 10^9 cfu/ml. In the example of the milk supply chain concentrations of more than 10^9 cfu/ml will not occur, even in the case of an intentional contamination (see section 3.4). Counts are non-negative. The data from the ComBase provide the decadic logarithm of the cell count. A number of zero corresponds to a contamination below detection [**Baranyi, 2011**], but $1 = \log 0$. In practice this may make barely a difference, but during calculation it must be taken into account. Therefore, we assume that more than one bacterium is necessary for growth.

Assumption 4. *Dynamics of growth and decay does not depend on initial bacterial concentration (inoculum size).*

But the *start-cfu/ml* may have an effect on growth and survival kinetics, see section 3.3. Unfortunately, there are not enough experimental data in the ComBase to take into account for inoculum size. The default selection method does not choose appropriate series of measured values according to initial bacterial counts; the trivial case of an initial value of zero bacteria is excluded. Therefore, the default interpolation method makes adjustments for initial bacterial counts.

Assumption 5. *The history of the cells, including stresses and potential injury, is of no influence.*

Like described in chapter 3, the history of the cells does have an influence on growth and death kinetics. But the data from the ComBase do not contain information about the previous cultivation conditions. Therefore, the history of the cells can not be taken into account when choosing the appropriate series of measured values during a simulation of the NPMPM. When data containing the pre-experimental cultivation conditions are available, it is recommended to implement a method for selection of the appropriate series of measured values in the NPMPM that incorporates this information. A general framework for inclusion of memory models into survival models was introduced by Vaidya and Corvalan [**Vaidya and Corvalan, 2009**].

Assumption 6. *The parameters of the process steps are uniformly distributed.*

The distribution of process step parameters in dairy manufacture is not known [**Anonymous, 2010a**]. The uniform distribution is used when the only parameters known are the minimum and the maximum. When the mode is known, too, the use of a PERT distribution is recommended [**Vose, 2008**].

6.4. Probability distributions

In the `NPMPM` certain parameters follow certain probability distributions, namely the continuous uniform distribution, the normal distribution, the log-normal distribution and the Pert distribution. The interested reader may consult [**Vose, 2008**] or [**Bosch, 1998**] for detailed information about

probability distributions.

The process parameters are assumed to be uniformly distributed (assumption 6). Under fixed conditions the number of bacteria after a certain time duration is assumed to be log-normally distributed. Hence, the decadic logarithm of the cfu is assumed to follow a normal distribution. For modelling of expert opinions often the Pert distribution (a version of the beta distribution) is used. In the validation, the assumption of a normal distribution is compared with the Pert distribution.

A continuous random variable X is called uniformly distributed in $[a, b]$, if it has the probability density function

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

Its mean and variance are given by

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

The uniform distribution is determined by the interval limits of its domain. It is used when all possible outcomes have the same probability or when nothing but the minimum and maximum value is known about the underlying probability distribution. Figure 6.9 shows a plot of the probability density function of a continuous uniform distribution.

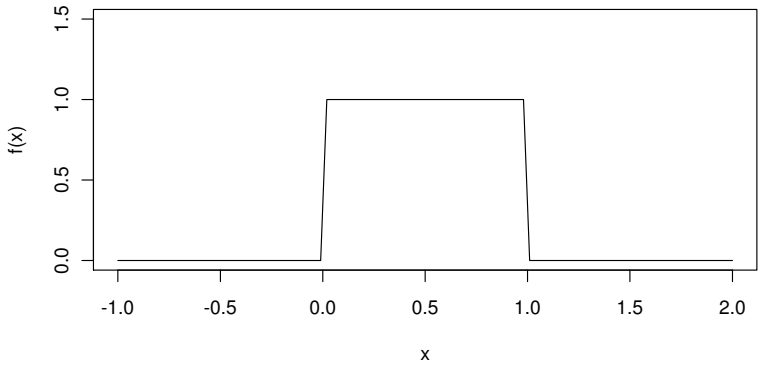


FIGURE 6.9. Probability density function of a continuous uniform distribution with $[a, b] = [0, 1]$

A continuous random variable X with mean μ and variance σ^2 is called normally distributed, $X \sim N(\mu, \sigma^2)$, if it has the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Its mean and variance are given by

$$E(X) = \mu,$$
$$Var(X) = \sigma^2.$$

The normal distribution is determined by its mean μ and variance σ^2. It is also called Gaussian distribution. Examples for normally distributed phenomena are milk production in cows, and deviations from target values in industrial processes [**Limpert et al., 2001**]. Figure 6.10 shows a plot of the probability density function of the standard normal distribution $N(0, 1)$.

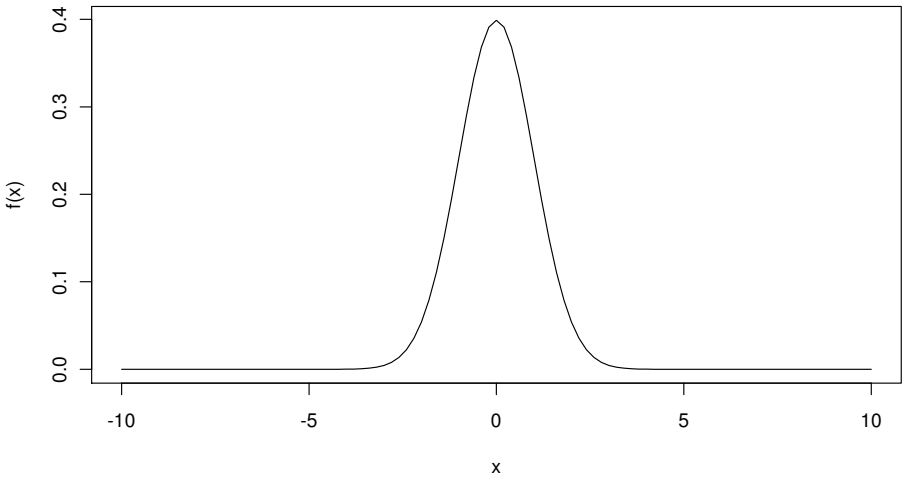


FIGURE 6.10. Probability density function of the standard normal distribution, $N(0, 1)$

When independent random variables interact additionally, the resulting random variable is normally distributed. The domain of the normal distribution is $\mathbb{R}$. But many random variables do not yield negative values, e.g. lengths and weights. When the distribution is not symmetrical, often the logarithm of the random variables is normally distributed. This is the case when independent, non-negative random variables interact multiplicatively.

A continuous random variable X with non-negative realisations is called log-normally distributed, if its logarithm is normally distributed, $\ln X \sim N(\mu, \sigma^2)$. Then X has the probability density function

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), & \text{for } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Its mean and variance are given by

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right),$$
$$Var(X) = \exp\left(2\mu + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right).$$

Figure 6.11 shows a plot of the probability density function of the log-normal distribution. Often latent periods of infectious diseases are log-normally

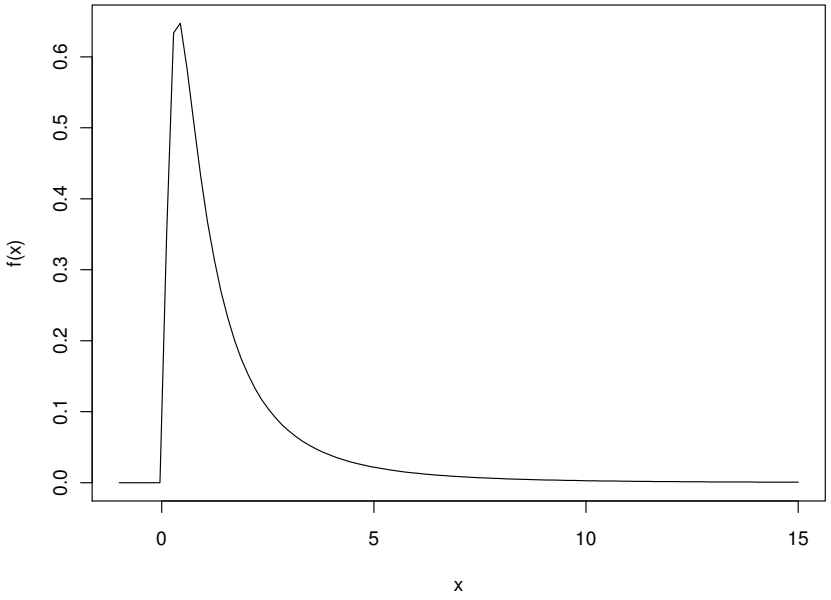


FIGURE 6.11. Probability density function of the standard log-normal distribution

distributed [**Limpert et al., 2001**]. An overview of the log-normal distribution in biological sciences is given in [**Koch, 1966**] and [**Koch, 1969**]. When the variation is small, both normal and log-normal distributions fit well [**Limpert et al., 2001**].

6.5. Statistical Tests

In this thesis, statistical tests are used to analyse the data and in model validation. The significance level was set to $\alpha = 0.05$. In the tests, a test statistic is calculated and compared with a critical value. The tabulated critical value depends on the chosen level of significance. Sets of series of measured values from the ComBase with matching parameters were tested with the Shapiro-Wilk test for normality of the log cfu. The two-sample Kolmogorov-Smirnov test is used to test if the `NPMPM` can reproduce the data it was built with.

The Shapiro-Wilk test tests for normal distribution. Let X be a continuous random variable with distribution function $F(x) = P(X \leq x)$. The null hypothesis is that the sample was taken from a normally distributed population.

$$H_0 : F(x) \sim N(\mu, \sigma^2).$$

The alternative hypothesis is that the corresponding population is not normally distributed,

$$H_1 : F(x) \nsim N(\mu, \sigma^2).$$

If the calculated test statistic is smaller than the critical value the null hypothesis is discarded. The Shapiro-Wilk test implemented in R calculates a p-value, too. The null hypothesis is discarded when the p-value equals, or is smaller than the level of significance.

The two-sample Kolmogorov-Smirnov test tests if two samples follow the same distribution. Let X and Y be continuous random variables with distribution functions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, respectively. Let the data sample and the output of the corresponding simulation consist of realisations of X and Y, respectively. The null hypothesis is that the random variables have the same distribution function,

$$H_0 : F(x) = G(x), \text{ for all x.}$$

The alternative hypothesis is inequality of the distribution functions,

$$H_1 : F(x) \neq G(x), \text{ for at least one x.}$$

The distance between the empirical distribution functions of the samples is quantified. If it is bigger than the critical value, the null hypothesis is discarded. The two-sample Kolmogorov-Smirnov test implemented in R (`ks.test(x,y,···)`) calculates a p-value, too. The null hypothesis is discarded when the p-value equals, or is smaller than the level of significance.

In theory, a sample from a population following a continuous probability distribution should not contain tied values. But experimental data may contain ties, because the accuracy of measurement is limited. The Kolmogorov-Smirnov test returns a warning message when there are ties. This can be prevented by addition of a small amount of noise, e.g. by the use of the R function `jitter`.

In validation of the `NPMPM` a significance level of 5% is chosen for both tests. Hence, a p-value bigger than 0.05 means that the null hypothesis is not rejected. In this thesis, the null hypotheses are the working hypotheses , i.e., a p-value bigger than 0.05 is requested. For the Shapiro-Wilk test that means the tested sample originated from a normally distributed population (with an error probability of 5%). For the Kolmogorov-Smirnov test that means the two tested samples originated from the same population (with an error probability of 5%).

6.6. Programming language

The `NPMPM` is implemented in the R programming language; an R package containing the model is downloadable from CRAN (http://cran.r-project.org/) [**R Development Core Team, 2010**]. The full documentation of the R package is included in this thesis in appendix D. R is often used in statistical software development and data analysis. It is part of the GNU project; its source code is freely available under the GNU General

Public License [**Anonymous, b**].

The decision to use R was based on the fact that spreadsheet-based program, like @Risk, Crystal Ball or Model Risk can not easily cope with large amounts of data and big numbers. For example, a test calculation of bacterial growth done with Excel and @Risk for an initial contamination of 100 bacteria and reasonable growth rate produced an error message in simulating 17h of growth.

6.7. The `grofit` R package

Methods from the `grofit` package [**Kahm et al., 2010**] are used for evaluation of experimental data. `grofit` was developed by Kahm *et al.* to derive dose-response curves. It contains methods for fitting four parametric and one non-parametric growth model to estimate characteristic growth parameters. Fitting is done by non-linear least squares. Available parametric models are the logistic growth model, the Gompertz model, the modified Gompertz model, and the Richards model. Smoothed cubic splines are used as model-free method. The length of the lag phase, the growth rate, and the maximum cell growth are estimated from the fitted curves.

6.8. Performance evaluation

Without an evaluation of its performance a model is useless, because there is no evidence about the relation of its output and the reality. There are several approaches for evaluation, depending on the type of model. It must be ensured that the model does not contain errors or bugs and that the algorithm has been properly implemented. Validation ensures that the model represents and correctly reproduces the behaviours of the modelled processes in the real world, and it examines if the model addresses the right problem. It must be assured that the data the model was built with can be recomputed (internal validation). The ability to reconstruct data not used to build the model should be investigated (external validation). Examination of the influence of fluctuations in model parameters on the output (sensitivity analysis) is relevant, too.

For evaluation the residuals of predicted and observed growth parameters are consulted. This can be done either as ratio, as a bias factor or accuracy factor introduced by Ross [**Ross, 1996**], or on means square differences [**Baranyi et al., 1999**]. Often the predicted growth parameters are plotted against the corresponding observed values, and the deviation from the identity line is taken as measure of inaccuracy. The interested reader may refer to chapter 12 of [**McCullagh and Nelder, 1999**] for a collection of other model checking approaches.

In sensitivity analysis the dependencies of the output on the choice of model parameters, and on fluctuations of the input are investigated. A traditional method is to use artificial data without noise and restrict the investigation to

certain regions of interest. Then for each set of output values the standard errors are computed. Delignette-Muller and Rosso calculated the Spearman rank correlation coefficients of the model response, with each variable input of the model [**Delignette-Muller and Rosso, 2000**]. The Spearman rank coefficient ρ is a measure for monotone dependence of two ordinal variables. The variation in one variable with respect to changes in another parameter or in the initial conditions can be quantified by traditional sensitivity functions (TSF) or by generalised sensitivity functions (GSF). The TSF depend on the parameters and the time. They are the first derivatives with respect to each parameter and the components of the initial condition, respectively. The TSF have local character, e.g. the sensitivity of the output corresponding to a certain parameter may vary in different time intervals [**Ernstberger, 2008**]. The GSF depend on the time. They are discrete cumulative functions [**Ernstberger, 2008**].

The output of the `NPMPM` is a distribution of *cfu/ml*s. Therefore, a direct comparison of corresponding values is impossible. We will use the statistical tests described in section 6.5 to compare experimental and simulated distributions.

6.9. Communication of probability information

The output of the simulations done with a predictive model often has to be communicated to diverse stakeholder groups, e.g. in context of risk assessment or risk analysis. These stakeholders can be industry professionals and their representatives, scientists, policy officers, programme administrators, economists, produce and quarantine inspectors, veterinary officers or trading partners [**Caponechia, 2009**]. In some cases an effective communication to the public or customers is needed, too. The output of a deterministic best-, average- or worst case simulation is one single number that can easily be discussed. But the communication and interpretation of the probability information obtained from a probabilistic model to people without mathematical background and not involved in the development of the model is a difficult task.

Caponecchia in [**Caponechia, 2009**] suggests graphical display of the key probabilities, a proper documentation and stakeholder workshops. He points out the difference in risk interpretation for low probabilities, presented in traditional probability estimates or as relative risk, respectively. The use of reference classes to ensure correct interpretation of probabilities, and a combined numerical and verbal presentation that links the numerical probabilities with terms like ‘average‘ or ´likely´ is recommended, too.

The R package `NPMPM` contains a function that produces a .pdf-file containing information about the current simulation. This function is named `makepdfoutput`. It takes a text as input, e.g. the names of the microorganisms and the process chain. The .pdf-file constructed contains

- a timestamp with its construction time,

- the text from the input,
- a histogram of the calculated cfus,
- a summary statistics,
- a list of the inoculum size(s) in the process steps,
- the number of iterations,
- the relative error bound for the fluctuation of the mean for stopping iterating,
- a density plot of the calculated cfus,
- the number of calculations for one process step with less than 30 appropriate series of measured values,
- and the information if there were process steps where the ranges of temperature, pH and a_w value had to be shifted to get appropriate series of measured values.

This .pdf-file will provide a helpful tool for systematic filing, and for communication of the results to stakeholders. For a brief description of the function and the R package `NPMPM` see appendix D. An example for a simulation of the `NPMPM` is given in section 8.8.

CHAPTER 7

A New Probabilistic Model in Predictive Microbiology

The new probabilistic model in predictive microbiology (NPMPM) computes the distribution of possible contamination sizes in a finished product at the end of a process chain. It uses Monte Carlo simulation. It starts at the first process step with a contamination, and calculates the contaminations in linear succession of the following process steps.

The NPMPM computes concentrations, i.e., numbers of bacteria in a reference volume of one ml. The terms *cfu/ml* and *logcfu/ml* refer to the number of bacteria in one ml that was measured in an experiment, and to the decadic logarithm of cfu/ml, respectively. Calculated bacterial concentration, and decadic logarithm of the calculated bacterial concentrations, are termed *calculated cfu/ml* and *calculated logcfu/ml*, respectively. The concentration of bacteria at the beginning of a process step is denoted by *start-cfu/ml*, the concentration of bacteria at the end of a process step is denoted by *end-cfu/ml*. The calculated concentration of bacteria in the end product, i.e., the *end-cfu/ml* of the last process step, is referred to as *final-cfu/ml*.

In chapters 2-6 knowledge necessary for modelling bacterial growth and decay is gathered. In chapter 2 causes and impact of foodborne diseases are described, showing the need for knowing the underlying mechanisms of bacterial growth and decay to prevent foodborne illness. These mechanisms are investigated in chapter 3. The kinetics of a population of bacteria depend on a multitude of interacting factors. Environmental conditions, e.g. the matrix in which bacteria are located, heavily influence population dynamics. Hence, as an example for food processing, the food supply chain of milk and dairy products is discussed in chapter 4. Options to mathematically describe the kinetics of bacterial growth and survival are discussed in chapter 5. In chapter 6 methods and assumptions used in development and validation of the NPMPM are characterised.

This chapter presents the NPMPM. First, the main idea is described in section 7.1. Then a top-down approach is used to introduce the algorithm in sections 7.2 - 7.5. Obligatory and optional input is characterised in section 7.6. In section 7.7 the output of the NPMPM is described. In section 7.8 the conformity of the algorithm to assumption 2 from section 6.3 is shown. A comparison to existing approaches in predictive microbiology is given in section 7.9.

7.1. Main idea

In food processing, predictive microbiology is used for calculation of bacterial counts in end products or intermediate products during processing. This is done for certain environmental conditions and certain points in time, e.g. after a heat treatment with a defined temperature-time combination.

The NPMPM is a probabilistic model, it does not include stochastic processes. Usually, probabilistic models use probability distributions of parameters for incorporating both variability and uncertainty, yielding a closed-form expression that describes the whole growth or survival curve, see section 5.5. Secondary models are included to make predictions under non-static conditions. These secondary models assume a certain dependence of growth or survival kinetics parameters on environmental conditions, see section 5.5.

The NPMPM is based on a new, different approach for including variability and uncertainty, and instead of describing the whole growth or survival curve, it aims at predicting the possible *cfu/ml* at one certain point in time s, e.g. at the end of a process step for calculation of *end-cfu/ml*. A set of *calculated cfu/ml* at time s is computed. This set is assumed to be a sample from a population of *cfu/ml* following a certain probability distribution, e.g. a log-normal distribution. The parameters of this probability distribution are estimated from the calculated sample. Finally, the *end-cfu/ml* is randomly drawn from this probability distribution. This procedure is pictured in figure 7.1.

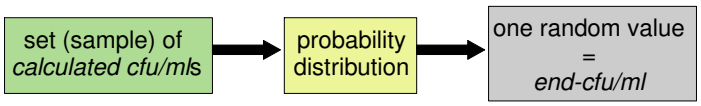


FIGURE 7.1. The main idea of the NPMPM

For computing a set of *calculated cfu/ml*, experimental data that match current process step conditions are gathered. One experiment is represented in this data set as series of measured values, i.e., one *cfu/ml* or *logcfu/ml* per measured time. From every single series of measured values one *calculated logcfu/ml* at time s is computed by fitting one deterministic primary model to its data points. An adaptive choice of primary models is possible, because not every series of measured values has to be fitted with the same model. One set of experimental data yields a set of *calculated cfu/ml*s at time s. This set is used to calculate the *end-cfu/ml* like described above. The whole procedure is pictured in figure 7.2.

Usually, process step conditions are not fixed values, but intervals (see assumption 6). In the NPMPM all experimental data available that match into these intervals is selected. Hence, the change of population kinetics due to environmental conditions is included without the use of a secondary model.

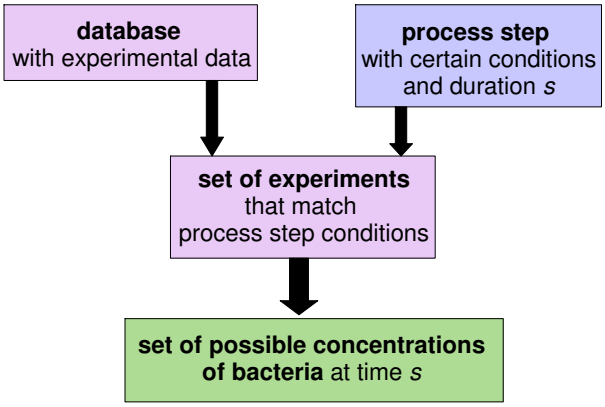


FIGURE 7.2

To model continuously changing process conditions, these changes have to be discretised, i.e., the process step is broken down in distinct process steps with constant conditions. An example for segmentation of a fermentation process during manufacture of yoghurt is pictured in figure 7.3.

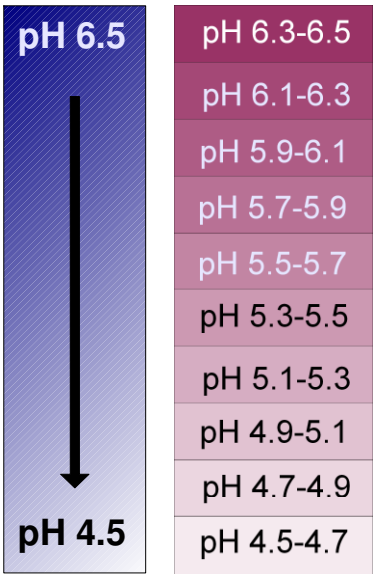


FIGURE 7.3. Segmentation of fermentation into sub-steps for modelling manufacture of yoghurt

7.2. General structure

The `NPMPM` uses Monte Carlo simulation. It contains three nested loops (see figure 7.4). The outer loop cycles through the iterations of the Monte Carlo simulation (see section 7.3). In every iteration of the

outer loop one *final-cfu/ml* is calculated. Therewith, during one simulation a set of *final-cfu/ml*s is computed. This set is the output of the NPMPM.

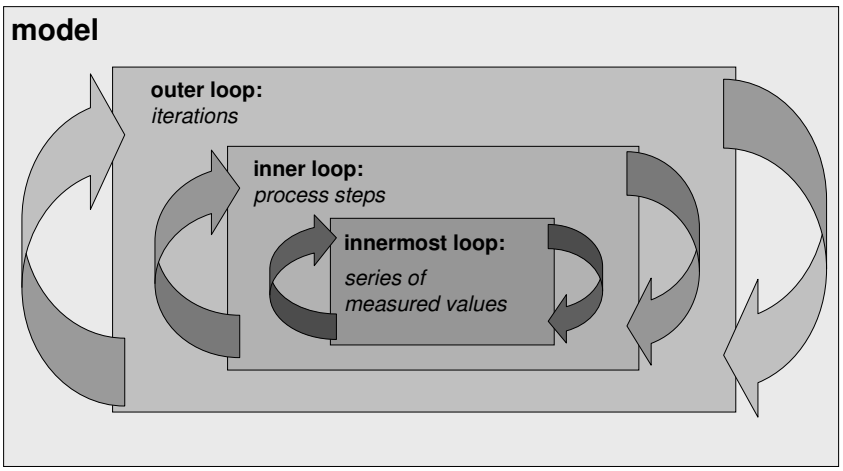


FIGURE 7.4. The three nested loops in the NPMPM

In every iteration of the outer loop, the inner loop cycles through all process steps of the process chain (see section 7.4). In every iteration of the inner loop one *end-cfu/ml* is calculated. For this purpose, a set of experimental data that match current process step conditions is selected. From this data a set of *calculated cfu/ml*s is computed. This set of bacterial concentrations is used for calculation of the *end-cfu/ml*. The process steps are handled in linear order. The *end-cfu/ml* of one process step is used for calculation of the *start-cfu/ml* of the next process step. The *end-cfu/ml* of the last process step in the process chain is the *final-cfu/ml* of one iteration of the outer loop.

In every iteration of the inner loop, the innermost loop cycles through all series of measured values in the set of experimental data selected previously in the inner loop (see section 7.5). In every iteration of the innermost loop, one *calculated cfu/ml* is computed from one series of measured values. This yields a set of *calculated cfu/ml*s that is used for calculation of the *end-cfu/ml* in the inner loop. The series of measured values do not follow a certain order, they are handled in the order provided by the inner loop.

The chosen structure of the model allows for a varying number of process steps, and for repeated contamination during processing. The input is checked for consistency, e.g. if all inoculum sizes are zero, the algorithm terminates with a warning. Then the first process step with a contamination is determined, and the calculation starts with this process step.

7.3. Outer loop

One iteration of the outer loop is one iteration of the Monte Carlo simulation. In one iteration the propagation of contamination during the whole process

chain is computed (in the inner loop, see algorithm 3). This calculation yields a *final-cfu/ml* that is stored in a vector as output. Additionally, in every iteration the mean of the hitherto calculated *final-cfu/ml*s is calculated and stored in another vector. The outer loop stops when the fluctuation of this mean between two consecutive iterations has become small enough. Termination is ensured by the stop criterion described in algorithm 1 in section 6.1: Worse case fluctuation is a $O\left(\frac{1}{n}\right)$. Hence, the algorithm always terminates.

The outer loop returns the vector with the set of *final-cfu/ml*s and the vector containing the mean of the first j *final-cfu/ml*s as jth component. Hence, in one simulation a distribution of *final-cfu/ml*s is computed. This distribution of *final-cfu/ml*s is also the output of the `NPMPM`. The pseudocode of the outer loop is given in algorithm 2.

Algorithm 2 Iterations of the Monte Carlo simulation (outer loop)

Input: error bound, inoculum size, methods used during the simulation {stop criterion, calculation for one process step, selection of experimental data, selection of process step duration, interpolation method}

1: iterations = 0 {initiate counter for number of iterations}
2: **while** stop criterion is not fulfilled **do** {algorithm 1 from section 6.1 returns FALSE}
3: increase the counter: iterations = iterations + 1
4: calculate the *final-cfu/ml* {in inner loop,see algorithm 3}
5: save *final-cfu/ml*
6: compute and save mean of hitherto calculated *final-cfu/ml*s
7: **end while**

Output: vector with *final-cfu/ml*s and vector with mean of first j *final-cfu/ml*s as jth component

7.4. Inner loop

In every iteration of the outer loop, the inner loop cycles through the process steps of the process chain. It starts with the first process step with a contamination, and iterates over the successional process steps in linear order. The *start-cfu/ml* of the current process step is calculated as sum of the *end-cfu/ml* of the previous process step (or zero in case of the first step), and the current inoculum size. The parameter intervals of the current process step, e.g. temperature, are determined as minimum and maximum values. One random point in time is drawn from the duration interval. Then the experiments matching the conditions of the process step (temperature, pH value and a_w value) are selected according to algorithm 5. From the set of series of measured values and the *start-cfu/ml*, one set of *calculated cfu/ml*s is computed in the innermost loop (see algorithm 6).

This set is assumed to be a sample from a normal distribution, according to assumption 2 in chapter 6. Parameters of a normal distribution are

estimated from this sample, and one value is drawn. This value is multiplied by the process step factor to take into account for accumulation and attenuation. If the calculated value is bigger than 10^9, it is rounded down to 10^9, and if it is equal or smaller than one, it is rounded down to zero, according to assumption 3. This calculations yield the *end-cfu/ml* of the current process step. The loop terminates when the *end-cfu/ml* of the last process step, i.e., the *final-cfu/ml*, is calculated. The *final-cfu/ml* is returned in a vector by the outer loop. In figure 7.5 one iteration of the inner loop (under the assumption that the cfus are log-normally distributed) is pictured. The pseudocode of this default inner loop is noted in algorithm 3.

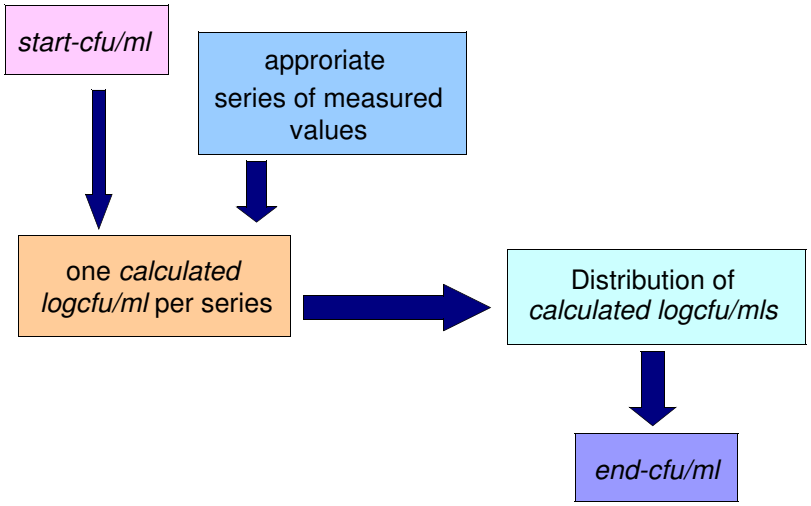


FIGURE 7.5. Calculation of one *end-cfu/ml* under the assumption that the cfus are log-normally distributed (inner loop)

If experimental data show no normal distribution of *logcfu/ml*s, and information about the actual probability distribution is lacking, using the frequency distribution of *calculated logcfu/ml* instead of fitting another probability distribution will usually yield a better approximation. In this case, it suffices to draw one random series of measured values, and to compute one *calculated logcfu/ml* from this series of measured values, instead of calculating a set of values from which one random value is drawn. This procedure reduces running time. The one *calculated logcfu/ml* is then used to calculate the *end-cfu/ml*, see algorithm 4.

One crucial step in the `NPMPM` is the selection of appropriate series of measured values. The default method is described in algorithm 5. It returns the selected series of measured values as list of ids.

The default method first selects all experiments with conditions matching the parameter intervals of the current process step. If the chosen set of

Algorithm 3 Inner loop cycling in linear order over the process steps in the process chain(default setting assuming a log-normal distribution of bacterial counts at one point in time)

Input: inoculum size, *end-cfu/ml* of previous process step, current process step, methods {selection of experimental data, selection of process step duration, interpolation method}

min = number of the first process step in the process chain with a contamination
max = number of the last process step in the process chain
i = min
while i < max+1 **do**
  if i = 1 **then**
    set *start-cfu/ml* = inoculum size
  else
    set *start-cfu/ml* = inoculum size + *end-cfu/ml* of previous process step
  end if
  if *start-cfu/ml* = 0 **and** there is no further contamination in following process steps **then**
    set *end-cfu/ml* = 0
  else
    get parameter intervals for current process step
    draw one random point in time from process step duration interval
    select appropriate series of measured values{see algorithm 5}
    compute a set of *calculated logcfu/ml*s from this set of series {in innermost loop, see algorithm 6}
    estimate parameters of the probability distribution of the population of this sample
    draw one random value x from this probability distribution
    set *end-cfu/ml* = factor·10^x
    if *end-cfu/ml* > 10^9 **then**
      set end-cfu/ml = 10^9
    end if
    if *end-cfu/ml* ≤ 1 **then**
      set *end-cfu/ml* = 0
    end if
  end if
end while

Output: *end-cfu/ml* (in last process step = *final-cfu/ml*)

appropriate series of measured values is too small, i.e., if there are less than 30 series of measured values, the parameter range is successively widened until a certain threshold. First, the interval of permitted a_w values is extended by ±0.05. Then, the range of the pH values is extended by ±1. After that, experiments with missing values in a_w values, pH values and finally in both are included. If the number of appropriate series of measured values still do not suffice, all series of measured values that fit in

Algorithm 4 Inner loop cycling in linear order over the process steps in the process chain(using the frequency distribution of *calculated logcfu/ml*), the part that differs from algorithm 3 is boldly printed in italics

Input: inoculum size, *end-cfu/ml* of previous process step, current process step, methods {selection of experimental data, selection of process step duration, interpolation method}

min = number of the first process step in the process chain with a contamination
max = number of the last process step in the process chain
i = min
while i < max+1 **do**
  if i = 1 **then**
    set *start-cfu/ml* = inoculum size
  else
    set *start-cfu/ml* = inoculum size + *end-cfu/ml* of previous process step
  end if
  if *start-cfu/ml* = 0 **and** there is no further contamination in following process steps **then**
    set *end-cfu/ml* = 0
  else
    get parameter intervals for current process step
    draw one random point in time from process step duration interval
    select appropriate series of measured values{see algorithm 5}
    draw one random series of measured values
    compute one **calculated logcfu/ml** = x ***from this series of measured values*** {in innermost loop, see algorithm 6}
    set *end-cfu/ml* = factor·10^x
    if *end-cfu/ml* > 10^9 **then**
      set end-cfu/ml = 10^9
    end if
    if *end-cfu/ml* ≤ 1 **then**
      set *end-cfu/ml* = 0
    end if
  end if
end while

Output: *end-cfu/ml* (in last process step = *final-cfu/ml*)

the temperature range of the process step are included, regardless of the other parameters. If yet not one single series of measured values is selected, the temperature range will be extended until at least one series of measured values is found; the other parameters are not considered. In this case for every extension of ±1°C a warning is printed on the screen; and the output of such a simulation must be interpreted very carefully.

Algorithm 5 Selection of appropriate series of measured values

`apsomv(process step)`

Input: experimental data and process step {data frame `raw_data` and one row of data frame `psteps`, see table 7.2}

1: get all series of measured values with conditions that fit into the parameter intervals of the current process step
2: **while** less than 30 series of measured values are selected **do**
3: gradually widen the range of the permitted a_w values and pH values until at least 30 series of measured values are selected, or until all series of measured values in the specified temperature range are selected
4: **end while**
5: **while** there is yet not a single series of measured values selected **do**,
6: print a warning
7: widen the temperature range by ± 1°C (ignoring a_w and pH values)
8: **end while**

Output: ids of the selected series of measured values

7.5. Innermost loop

In every iteration of the inner loop, the innermost loop cycles through the selected experiments. It computes the decadic logarithm of bacterial concentration. From every appropriate series of measured values one *calculated logcfu/ml* is computed. Hence, it yields a distribution of *calculated logcfu/ml*s. If the *start-cfu/ml* is zero, the *end-cfu/ml* must be zero, too. Therefore, the *calculated logcfu/ml* is set to -10^{35}, yielding a *calculated cfu/ml* of approximately zero. The pseudocode of the innermost loop is noted in algorithm 6.

First, extrapolation is excluded. If the timespan of the experiment is smaller than the process step duration, a warning ´No Extrapolation´ is printed in the shell, and the *calculated logcfu/ml* is set to NA. For a positive *start-cfu/ml*, one *calculated logcfu/ml*is computed by interpolation. If the number of data points in the series is not sufficient for the selected interpolation method, no *calculated logcfu/ml* can be computed from this series of measured values. It may occur that there are no appropriate series of measured values with a sufficient timespan and a sufficient number of data points. In this case the simulation terminates with an error message due to missing data.

The default interpolation method is an ordinary linear regression of the *logcfu/ml* over time. It returns the *calculated logcfu/ml* at one point in time. This point in time was drawn from the uniform distribution of the process step duration, see algorithm 3. Another interpolation method provided by the R package is a local linear interpolation. An overview of the innermost loop with interpolation by linear regression is pictured in figure 7.6.

Algorithm 6 Calculation of *calculated logcfu/ml* for one series of measured values (innermost loop)

Input: *start-cfu/ml*, process step duration, id of one series of measured values, measured values {data frame `measured_values`, see table 7.2}

1: **for** all appropriate series of measured values **do**
2: **if** timespan of measurements is smaller than duration of process step **then**
3: set *calculated logcfu/ml* to NA
4: print warning ´No Extrapolation´
5: **else**
6: **if** *start-cfu/ml* = 0 **then**
7: set *calculated logcfu/ml* = -10^{35}
8: **else**
9: compute one *calculated logcfu/ml* from the series and the *start-cfu/ml*
10: **end if**
11: **end if**
12: **end for**

Output: *calculated logcfu/ml*s

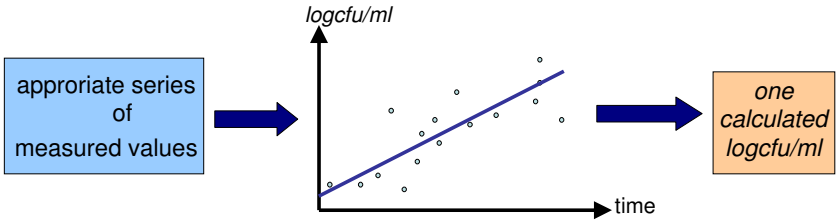


FIGURE 7.6. Calculation of one *calculated logcfu/ml* from one experiment with linear regression as interpolation method (innermost loop), with process step duration t

7.6. Input

Part of the information essential for a simulation must be passed as input to the main method `npmpm`, other information must be provided as data frames in the global environment of the R session. A data frame (`data.frame`) is a matrix able to contain different data types. Especially data frames containing experimental data might be big, and passing this information as entry to `npmpm` could significantly slow down simulation. An overview of the input essential for a simulation is given in table 7.1. Figure 6.1 pictures the links between the input and the calculation.

Process steps, corresponding (intermediate) matrices, and experimental data must be provided as data frames. Composition of the essential data frames and links in between them do match the structure given by tables in the MySQL database described in section 6.2. The intersection of processing

TABLE 7.1. Obligate and optional input of the `NPMPM`

information	provided as	obligate or optional
methods called during a simulation	entry	optional
error bound	entry	optional
inoculum sizes	entry	obligate
matrices	data frame	obligate
process steps	data frame	obligate
series of measured values	data frames	obligate

parameters and experimental parameters are the environmental conditions temperature, pH and a_w value. Therefore, these parameters are selected as essential parameters.

Experimental data is provided in two data frames. The data frame `raw_data` contains experiments done with the microorganism of interest, with experimental conditions temperature, pH and a_w value. The measured values data frame `measured_values` corresponds to the experiments. Storing the experimental data in one data frame only, would lead to redundant storage of experimental conditions. Additionally, the process of selecting series of measured values matching certain experimental conditions would be slowed down.

Restricting `raw_data` to the microorganism of interest is mandatory, because the default methods do not query `raw_data` during simulation to select for experiments done with a certain bacterium. It is recommended to restrict `measured_values`, too, but it is not compulsory. Selection of experiments matching the microorganism of interest before the simulation has some advantages. It speeds up simulation, because multiple searches during calculation are faster in a smaller data frame. In experiments pathogens are often substituted by non-pathogenic bacteria. For using the surrogate data for calculation, they only have to be provided in the data frames. An alteration of the microorganism of interest in `raw_data` can restrict the simulation to any number of certain strains or biovars, or to experiments done with a special matrix. If another method of selection of experiments than the default method is favoured, the name of the corresponding R function (written by the user) can be passed to `npmpm`. The link between `raw_data` and `measured_values` is pictured in figure 6.2.

The process chain of interest is not passed to `npmpm`, but is represented in data frames with the process step conditions and the characteristics of the corresponding (intermediate) matrices. The data frame `psteps` lists the process steps with the parameters temperature and duration, the position of the process step in the process chain, and a factor that describes attenuation or accumulation of bacteria in the process step, e.g. in dairy processing an accumulation in cream during separation. The data frame `matrix` describes the (intermediate) milk products by means of their pH and a_w value. The link between `psteps` and `matrix` is pictured in figure 6.6. It is mandatory to

TABLE 7.2. Mandatory data frames and their obligatory structure

data frame	column names	content
`psteps`	id	unique id
	temp_min	minimal temperature in °C
	temp_max	maximal temperature in °C
	time_min	minimal duration in s
	time_max	maximal duration in s
	matrix_id	id of the corresponding (intermediate) milk product for the process step
	factor	takes account for attenuation or accumulation of bacteria
	stepnumber	position in the sequence of process steps in the process chain
`matrix`	id	unique id
	pH_min	minimal pH value
	pH_max	maximal pH value
	aw_min	minimal a_w value
	aw_max	maximal a_w value
`raw_data`	id	unique id
	temperature_C	temperature in °C
	pH	pH value
	aw	a_w value
`measured_values`	raw_data_key	id of the corresponding series of measured values
	time	time in s
	logc	*logcfu/ml* at the specified point in time

restrict `psteps` to the process chain of interest, because the default methods do not query `psteps` during the simulation to select for the process chain. It is recommended to restrict `matrix`, too, but it is not compulsory.

The data used in this thesis is stored in a MySQL database. For one particular combination of process chain and microorganism the database must be queried for the matching records. These records have to be exported to data frames with the obligatory structures described in table 7.2. The name of the microorganism is not needed for the actual simulation; it must be ensured by the user that the data frames only contain series of measured values for the microorganism of interest.

The only mandatory input that must be passed to `npmpm` is the size of the inocula in list of numeric values. For every process step exactly one inoculum size must be provided; otherwise the simulation stops with a warning. Therewith, multiple contaminations in different process steps of the same process chain can be taken into account. A process step without a new contamination corresponds to an inoculum size of zero.

The specification of the methods called by the `NPMPM` is optional. For using other than the default methods, their names must be passed to `npmpm`. For

help in implementation of methods not included in the R package `NPMPM` see the package documentation in appendix D.

A simulation of the `NPMPM` stops when the calculated distribution of *final-cfu/ml*s is stable, like described in algorithm 1. Stability is measured as fluctuation of the mean of the hitherto calculated *final-cfu/ml*s. Calculation of the worst case in section 6.1 yields that the fluctuation is a $O\left(\frac{1}{n}\right)$, therefore, the algorithm always terminates. The upper limit of fluctuation tolerated is determined by a relative error bound. Its default value is 1%, but other values can be passed to `npmpm` as optional input.

7.7. Output

The `NPMPM` computes the distribution of possible contamination sizes in the end product. A vector with the *final-cfu/ml*s is returned.
The number of iterations is printed in the shell to show the progression of the simulation. Warnings are printed in the shell, too, e.g. when the number of matching experimental data is too small.

The R package provides the method `makepdfoutput` that generates a .pdf-file describing the simulation results. `makepdfoutput` needs a text as obligatory input, e.g. containing the names of the microorganism and the process chain. Additionally, it uses information that is assigned to the global environment during the simulation. The .pdf document constructed by `makepdfoutput` contains

- a timestamp with the construction time,
- the text from the input,
- a histogram of the *final-cfu/ml*s,
- summary statistics,
- a list of the inoculum sizes,
- the number of iterations,
- the relative error bound for the fluctuation of the mean for stopping the simulation,
- a density plot of the *final-cfu/ml*s,
- the number of process steps with less than 30 appropriate series of measured values (sum over all iterations),
- and the information if there where process steps where the ranges of temperature, pH and a_w value had to be shifted to get appropriate series of measured values (see algorithm 5).

An example of the .pdf is shown in figures 8.13 and 8.14.

7.8. Log-normally distributed cfus

The `NPMPM` was developed to make predictions of bacterial counts during milk processing. Calculation is done in linear order of process steps. The *end-cfu/ml* of one process step serves as *start-cfu/ml* of the next process step, from which the next *end-cfu/ml* is calculated. It is assumed that

calculated cfu/ml are log-normally distributed.

It still has to be shown that under certain basic conditions a log-normally distributed *start-cfu/ml* yields a log-normally distributed *end-cfu/ml*. Theorem 7.8.1 states that this is true, if all series of measured values have the same *start-cfu/ml* and the same measuring points in time $t_i = i = 1, \cdots, n$, and if the set of *cfu/ml*s from all series of measured values for one fixed point in time follows a log-normal distribution.

Theorem 7.8.1. *Let the decadic logarithm of the* start-cfu/ml *be a realisation of a normally distributed random variable,* $Y \sim N\left(\mu_*, \sigma_*^2\right)$. *Let all* m *series of measured values have the same* start-cfu/ml *and the same* n *measuring points in time* $t_i = i = 1, \cdots, n$. *Let the* logcfu/ml*s for one point in time be normally distributed random variables* $X_i, i = 1, \cdots, n$, $X_i \sim N\left(\mu_i, \sigma_i^2\right)$.

If interpolation is done by means of linear regression, the decadic logarithm of the end-cfu/ml *is a normally distributed random variable* Z. *Furthermore with process time* s *it holds true that*

$$Z \sim N\left(A, B\right). \tag{7.8.1}$$

with

$$A = \mu_* + \frac{12s}{n(n^2-1)} \sum_{i=1}^{n} \left(\left(i - \frac{n+1}{2} \right) \mu_i \right) \qquad \textit{and}$$

$$B = \sigma_*^2 + \frac{144s^2}{n^2(n^2-1)^2} \sum_{i=1}^{n} \left(\left(i - \frac{n+1}{2} \right)^2 \sigma_i^2 \right).$$

The following theorem takes part in the proof of theorem 7.8.1.

Theorem 7.8.2. *Let the random variables* X_i, $i = 1, \cdots, n$ *be independent and normally distributed,* $X_i \sim N\left(\mu_i, \sigma_i^2\right)$. *Then the linear combination*

$$X = \sum_{i=1}^{n} c_i X_i \qquad , c_i \in \mathbb{R}$$

is normally distributed, too. It has the expected value $\mu = \sum_{i=1}^{n} c_i \mu_i$ *and the variance* $\sigma^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2$.

The proof of theorem 7.8.2 can be found in basic stochastics books, e.g. [**Bosch, 1998**].

Proof. To proof theorem 7.8.1 it has to be shown that the linear regression curve that is fitted to the random variables Y and $X_i, i = 1, \cdots, n$, for fixed time s returns a normally distributed random variable Z. The slope b of the linear regression curve is defined as quotient of covariance of time and *logcfu/ml*, and variance of time. After computing covariance and variance, b is shown to be normally distributed. As Z is a linear combination of normally distributed intercept Y and normally distributed

slope b, it is also normally distributed.

Let $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$ denote the mean of the points in time t_i. Then the variance of time, var_t, can be calculated as

$$\begin{aligned}
var_t &:= \frac{1}{n-1}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2, \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(i - \bar{t}\right)^2, \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(i - \frac{1}{n}\sum_{i=1}^{n} t_i\right)^2, \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(i - \frac{1}{n}\sum_{i=1}^{n} i\right)^2, \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2, \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(i^2 - (n+1)i + \frac{(n+1)^2}{4}\right), \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n} i^2 - (n+1)\sum_{i=1}^{n} i + \frac{n\,(n+1)^2}{4}\right), \\
&= \frac{1}{n-1}\left(\frac{n\,(n+1)\,(2n+1)}{6} - \frac{n\,(n+1)^2}{2} + \frac{n\,(n+1)^2}{4}\right), \\
&= \frac{n\,(n+1)}{12}.
\end{aligned}$$

The variance depends on the number of measured values in one series, n. For a given n the variance var_t is a constant. With theorem 7.8.2 the mean of the *logcfu/ml*, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, is a normally distributed random variable,

$$\bar{X} \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i, \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2\right).$$

Hence, it holds true that $\left(X_i - \bar{X}\right)$ is normally distributed,

$$\left(X_i - \bar{X}\right) \sim N\left(\mu_i - \frac{1}{n}\sum_{j=1}^{n}\mu_j, \sigma_i^2 - \frac{1}{n^2}\sum_{j=1}^{n}\sigma_j^2\right).$$

Therefore, the covariance of time and *logcfu/ml* can be calculated as

$$
\begin{aligned}
cov_{tX} &:= \frac{1}{n-1}\sum_{i=1}^{n}\left((t_i-\bar{t})\left(X_i-\bar{X}\right)\right),\\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(\left(i-\frac{n+1}{2}\right)\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)\right),\\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(iX_i-\frac{n+1}{2}X_i-\frac{i}{n}\sum_{j=1}^{n}X_j+\frac{n+1}{2n}\sum_{j=1}^{n}X_j\right),\\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}iX_i-\frac{n+1}{2}\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\left(\frac{i}{n}\sum_{j=1}^{n}X_j\right)+\sum_{i=1}^{n}\frac{n+1}{2n}\sum_{j=1}^{n}X_j\right),\\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}iX_i-\frac{n+1}{2}\sum_{i=1}^{n}X_i-\frac{n+1}{2}\sum_{i=1}^{n}X_i+\frac{n+1}{2}\sum_{i=1}^{n}X_i\right),\\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}iX_i-\frac{n+1}{2}\sum_{i=1}^{n}X_i\right),\\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(\left(i-\frac{n+1}{2}\right)X_i\right)\right).
\end{aligned}
$$

The covariance depends on the number of measured values in one series, n, and on the random variables representing the *log cfu/ml* at the measured time points. For a given n the covariance $covar_{tX}$ is normally distributed.

The slope b of the linear regression curve now can be calculated as

$$
\begin{aligned}
b &:= \frac{cov_{tX}}{var_t},\\
&= \frac{12}{(n-1)\,n\,(n+1)}\sum_{i=1}^{n}\left(\left(i-\frac{n+1}{2}\right)X_i\right),\\
&= \frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(\left(i-\frac{n+1}{2}\right)X_i\right).
\end{aligned}
$$

Like the covariance $covar_{tX}$, the slope of the regression curve depends on n, and on the X_i. Therefore, b is normally distributed for a given n.

Theorem 7.8.2 yields

(7.8.2)

$$
b \sim N\left(\frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(\left(i-\frac{n+1}{2}\right)\mu_i\right), \frac{144}{n^2(n^2-1)^2}\sum_{i=1}^{n}\left(\left(i-\frac{n+1}{2}\right)^2\sigma_i^2\right)\right).
$$

The decadic logarithm of the *end-cfu/ml* Z depends on the duration s of the process step. For a given s it holds true that

$$
Z := Y + b\cdot s\ .
$$

Therefore, with theorem 7.8.2 Z is normally distributed, $Z \sim N(A, B)$, with

$$A = \mu_* + \frac{12s}{n(n^2-1)} \sum_{i=1}^{n} \left(\left(i - \frac{n+1}{2} \right) \mu_i \right) \quad \text{and}$$

$$B = \sigma_*^2 + \frac{144s^2}{n^2(n^2-1)^2} \sum_{i=1}^{n} \left(\left(i - \frac{n+1}{2} \right)^2 \sigma_i^2 \right).$$

□

Theorem 7.8.1 proofs that under certain basic conditions a log-normally distributed *start-cfu/ml* yields a log-normally distributed *end-cfu/ml*. When conditions are different, e.g. when experimental data do not yield log-normally distributed *cfu/ml*s, then the *end-cfu/ml* might not be log-normally distributed, despite of log-normally distributed *start-cfu/ml*s.

7.9. Comparison to existing approaches

The `NPMPM` uses a new approach for including variability and uncertainty into a predictive model. Existing probabilistic models gather experiments done under similar conditions, and then fit one family of curves to the set of series of measured values. Variability and uncertainty are usually included into such primary (and secondary) models by means of parameters that follow probability distributions. Such models consist of a deterministic and a stochastic or probabilistic part [**McMeekin et al., 1993**], i.e., a certain shape of kinetics is assumed. These approaches are used to model population kinetics over one time interval, i.e., the global approximation error is minimised.

In risk assessment, the local approximation error is more important than the global approximation error, e.g. for modelling survival of *Listeria monocytogenes* in the milk chain during a certain heat treatment, the concentration of bacteria must be exactly predicted at the end of the process step. The goal of probabilistic models used in risk assessment is to capture the whole range of possible outcomes, connected with their probabilities. Models that restrict this distribution without biological reasons, e.g. because it is necessary for reducing the global error in fitting, reduce the reliability of prognostications. McKellar and Lu analysed model comparisons from the literature; they state that there is no single growth model as most appropriate representation of bacterial growth [**McKellar and Lu, 2004**]. In section 8.3 experimental data from the ComBase [**Anonymous, a**] are presented. The experiments were done in one laboratory, under identical environmental conditions, and with the same microorganism. Nevertheless, kinetics of these series of measured values show a variety of shapes (see figures 8.6 and 8.7), including death and growth in one single curve. Therefore, models that assume a certain shape of kinetics, may fail in describing variations in population kinetics that is shown even in experiments with matching experimental conditions.
The `NPMPM` includes variability and uncertainty on another level than existing approaches. It does not reconstruct one general growth or survival curve, but aims at predicting the number of bacteria at one certain point in time.

Every curve is used to fit one deterministic primary model, allowing for an adaptive choice of primary models. Therewith, differences in population kinetics due to biological variability and uncertainty about experimental conditions can be taken into account. This adaptivity is not yet implemented, but planned for future work (see section 9.2). Currently, the R package `NPMPM` provides two interpolation methods. In every simulation the method selected by the user is used for calculation.

CHAPTER 8

Results and Discussion

In the last chapter the NPMPM was presented. Now the model is validated with artificial and experimental data. Simulation results and an example simulation for a particular dairy process chain are given.

Accuracy of the output from a model that was not yet validated can not be assessed. A model must reproduce the data it was built with. This is ensured by internal validation. Both artificial data and experimental data from the ComBase (see section 6.2) are used for internal validation of the NPMPM. Model performance must be evaluated under different settings. For example for fitting of the logistic growth model (equation 5.4.12), it is important that data for fitting do sample all parts of a sigmoid growth curve, because information content of data points from different parts of the growth curve varies [**Ernstberger, 2008**].

A sensitivity analysis is essential for detection of parameters with wide influence on the output. Fitting of these parameters has to be as accurate as possible. Identification of requirements that must be fulfilled for applying the model is crucial. Partly, these preconditions are given by the assumptions made during development of the model, e.g. the NPMPM can only be applied with its default settings when bacterial counts are log-normally distributed for fixed parameter combinations, according to assumption 2. Other requirements, like level of accuracy of input parameters for a certain accuracy of the output, are identified in sensitivity analysis, and in internal and external validation.

Evaluation of the NPMPM uses statistical tests. The Kolmogorov-Smirnov test is used for comparison of data and output. Normality of decadic logarithms of bacterial counts is tested with the Shapiro-Wilk test. A significance level of 5% is chosen for both tests. These statistical methods are described in chapter 6.

Like in chapter 7, the terms *cfu/ml* and *logcfu/ml* refer to the number of bacteria in one ml that was measured in an experiment, and to the decadic logarithm of cfu/ml, respectively. Calculated bacterial concentrations, and decadic logarithm of the calculated bacterial concentrations, are termed *calculated cfu/ml* and *calculated logcfu/ml*, respectively. The concentration of bacteria at the beginning of a process step is denoted by *start-cfu/ml*, the concentration of bacteria at the end of a process step is denoted by

end-cfu/ml. The calculated concentration of bacteria in the end product, i.e., the *end-cfu/ml* of the last process step, is referred to as *final-cfu/ml*.

In chapters 2-6 the background for development of the `NPMPM` is provided. The `NPMPM` is presented in chapter 7. Now model performance is investigated and discussed. The methods from the R package used during validation are presented. First evaluation uses artificial data. Then certain sets of series of measured values were selected from the ComBase-data for investigation of model performance with experimental data. Demands on experimental data are highlighted, and sensitivity of the `NPMPM` is analysed. Running time of the `NPMPM` is discussed, and constraints are reviewed. Finally, an example simulation for fresh milk processing is presented.

8.1. Methods

Validation uses default methods, and other methods provided with the R package (see appendix D). An overview of key procedures considered in validation is pictured in figure 8.1.

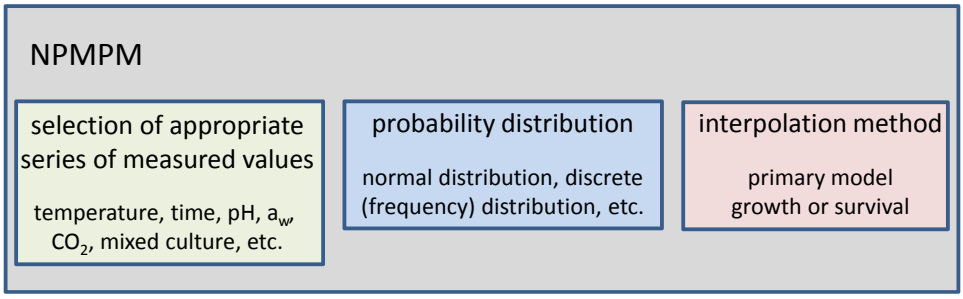


FIGURE 8.1. Key procedures in the `NPMPM`

Default method for selection of series of measured values in the inner loop is `apsomv` (see algorithm 5). It was not used in validation, because data sets used in validation have to be selected before simulation, in order to allow for a comparison of input data and output. Hence, a method that selects all series of measured values in the data frames was used. It is also provided in the R package as method `allsomv`, see appendix D.

Default assumption is a normal distribution of the set of *calculated logcfu/ml*s computed in the innermost loop (see assumption 2). Model performance is compared for use of a normal distribution (method `onestep`, see algorithm 3), and use of the (discrete) frequency distribution of the *calculated logcfu/ml*s (method `disconestep`). In the latter one series of measured values is randomly drawn and used to compute one *calculated logcfu/ml*. From this value the *end-cfu/ml* is calculated. If the experimental data used in the simulation are not log-normally distributed, the frequency distribution should yield a better approximation. Both methods

do not allow for extrapolation. Calculation of the *end-cfu/ml* by the method `onestep` includes rounding to a value between zero and 10^9, see algorithm 3. This is essential, because the normal distribution is not bounded, and therefore can yield values that are biological impossible, e.g. negative bacterial counts. This adjustment is not made by `disconestep`.

Default method for interpolation in the innermost loop is global linear interpolation (method `cfuinterpolation`, see section 7.5). Intercept and slope of a linear function are fitted to the data by ordinary least squares, see section 5.1. This linear function is used to compute the *calculated logcfu/ml*. The method `cfuinterpolation` is compared with local linear interpolation (method `simpleinterpolation`). If one measured value corresponds to the point in time at which the concentration of bacteria should be calculated, this value is used as result, i.e., `simpleinterpolation` yields an exact approximation at measuring times. Otherwise, intercept and slope of a linear function are calculated from the two measured values adjacent to the desired point in time. The *calculated logcfu/ml* is computed from this linear function.

8.2. Artificial data

First, internal validation was done with artificial data. Both log linear (exponential) data and series of measured values following a sigmoid curve were generated. The latter was calculated with the Verhulst equation (equation 5.4.12). Every data set was used as input in data frames `raw_data` and `measured_values` (see section 7.6. All data sets contained 30 artificial series of measured values.

The following sets of log linear artificial data were used:

- sets containing the same log linear series repeatedly,
- sets containing two log linear series with different *start-cfu/ml*s repeatedly (15 times each),
- sets with normally distributed *start-cfu/ml* and fixed slope, yielding normally distributed *logcfu/ml*s at every point in time, and
- sets of series with a fixed *start-cfu/ml* and normally distributed slopes, also yielding normally distributed *logcfu/ml*s at every point in time.

Tests with these artificial data satisfy all assumptions listed in section 6.3: log-normality of *cfu/ml* at one certain point in time under fixed conditions (assumption 1), log-linear series of measured values (assumption 2), bounded bacterial counts (assumption 3) due to constant parameters, no influence of inoculum size on kinetics (because of fixed inoculum size, assumption 4), no differences between series of measured values caused by the history of the cells (assumption 5), and uniformly distributed process step parameters (assumption 6), due to constant parameters). As expected, for all log linear series the `NPMPM` reproduced the data with both (linear) interpolation methods.

Then a set of 30 identical logistic curves was generated to investigate the influence of the shape of the curve on goodness of approximation. These curves covered lag phase, log phase and stationary phase. Such sigmoid curves are non-linear. Therefore, assumption 1 is violated. Every series of measured values consisted of 21 equidistantly distributed measured values. Because the method for local linear interpolation `simpleinterpolation` yields an exact approximation at the measuring time, interpolation was done between these measured values, e.g. for adjacent measuring points in time t_i and t_{i+1}, interpolation was done at $\frac{t_i+t_{i+1}}{2}$.

In figure 8.2 the output of the `NPMPM` and the logistic curve of the artificial input are plotted for different interpolation methods and probability distributions. Global linear interpolation (subfigures a, b) yielded a good fit in the area around the intersection of the regression curve and the logistic curve; outside this area approximation was poor. This is due to non-linearity of the sigmoid curve. Local linear interpolation (subfigures c, d) yielded a better approximation; goodness of fit of local linear interpolation depends on the distribution of grid points. The method that uses the frequency distribution of *calculated logcfu/ml* (method `disconestep`) makes no adjustments for allocating the *calculated cfu/ml* to values between 0 and 10^9 (subfigures b, d). Therefore, assumption of a normal distribution of *logcfu/ml*s (method `onestep`) yielded a better approximation.

Simulations with the two linear interpolation methods are able to reproduce the data if all assumptions listed in section 6.3 are satisfied. If the measured values do not depend log-linearly on time, i.e., assumption 1 is violated, global linear interpolation usually fits poorly and the data can not be reproduced. There are various reasons for deviations from log-linear kinetics (see subsection 6.3). Therefore, global linear interpolation will often lead to poor approximations. Approximation quality of local linear interpolation is usually better, but depends on grid points. If experiments and process step duration are on different time scales, e.g. hours versus seconds, the distance between grid points may be too big, and local linear interpolation fits poorly. In summary the two linear interpolation methods will not suffice for experimental data that are not log-linear, and for data and process steps on different time scales.

8.3. Experimental data

Further internal validation was done with different samples of experimental data extracted from the ComBase. These samples were used in input data frames (see section 7.6). The `NPMPM` was run with one single process step with conditions matching the experimental conditions of the data. This approach simulates the experiments in which the data were generated. To allow for calculation of the approximation error, calculation was done for measuring times of the experimental data. The method for local linear interpolation `simpleinterpolation` yields an exact approximation at measuring times (see section 8.1). For evaluation of the performance of local

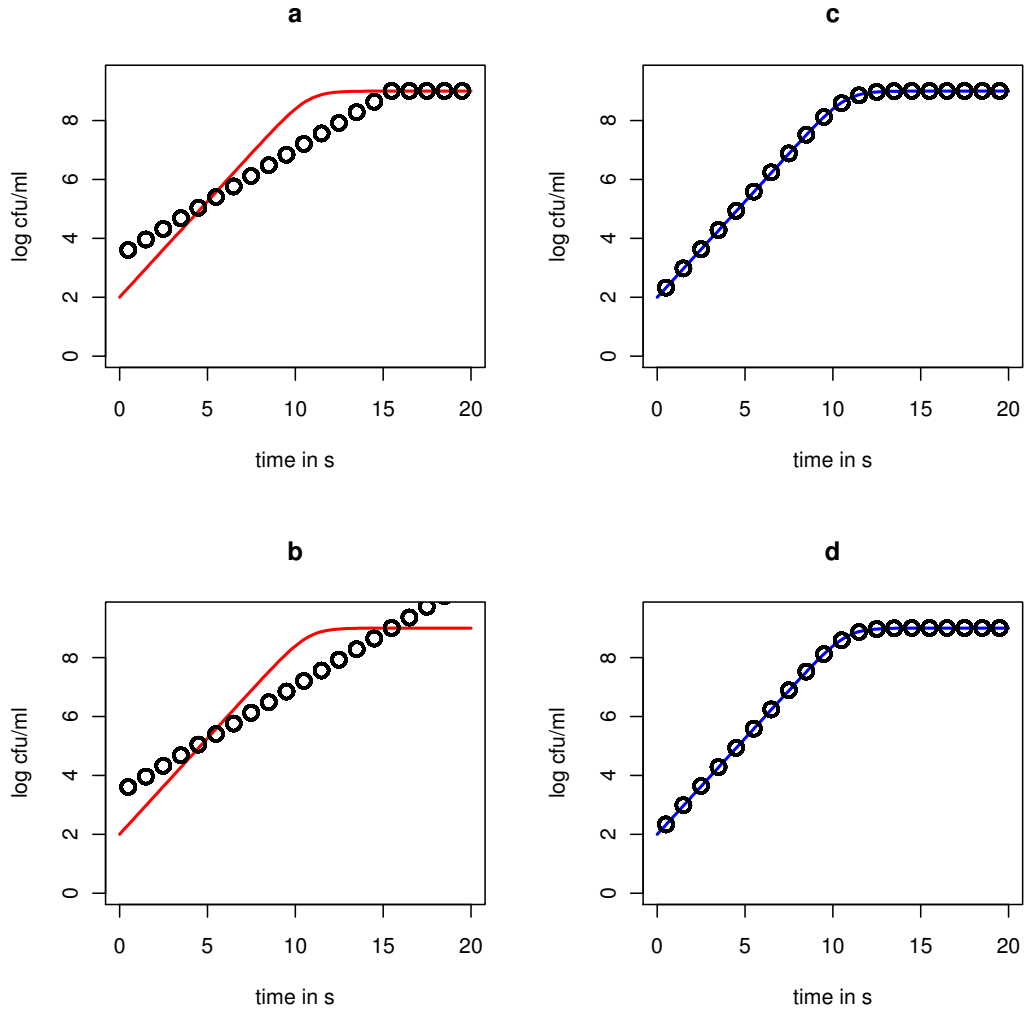


FIGURE 8.2. Output (circles) of interpolation for series of measured values generated with the logistic growth model (red line), **a** with global linear interpolation and assumption of log-normally distributed *cfu/ml*s (bacterial concentrations bigger than 10^9 are rounded to 10^9), **b** with global linear interpolation and discrete distribution of *cfu/ml*s (frequency distribution), **c** with local linear interpolation and assumption of log-normally distributed *cfu/ml*s, **d** with local linear interpolation and discrete distribution of *cfu/ml*s (frequency distribution)

linear interpolation at one measuring time t_i, the corresponding measured value was removed from all series of measured values in the input data frame. This manipulation yields a bigger interpolation interval and increases the approximation error, but it allows for assessing the interpolation error of the method `simpleinterpolation`. For ensuring comparability, removal of

TABLE 8.1. ComBase data: temperatures with a total number of more than 100 series of measured values, and associated *start-cfu/ml*s with at least 30 series of measured values

temperature in °C	**total number of series**	$\log_{10}$ ***start-cfu/ml***	**number of series**
0	146	6.49	34
2	151	6.55	35
4	821	6.29	35
		6.5	54
7	126	6.55	35
10	492	6.29	35
15	210	6.29	35
20	447	6.5	54
		6.55	33

measured values described above was done before corresponding simulations with global linear interpolation, too.

Sets of series of measured values with identical experimental conditions and similar *start-cfu/ml* were sampled. Figure 8.3 summarises the selection process. First, the database was queried for series with the same temperature. More than 100 series of measured values each were found for 0°C, 2°C, 4°C, 7°C, 10°C, 15°C, and 20°C. From this sets subsets with the same *start-cfu/ml* were sampled, see table 8.1. More than 30 series of measured values each were found for 4°C and a *start-cfu/ml* of $10^{6.29}$, 4°C and a *start-cfu/ml* of $10^{6.5}$, 20°C and a *start-cfu/ml* of $10^{6.5}$, and 20°C and a *start-cfu/ml* of $10^{6.55}$. From this sets subsets with the same pH value and a_w value were sampled, see figures 6.5 and 6.4. More than 30 series of measured values were found for 4°C and a *start-cfu/ml* of $10^{6.5}$ and a pH value of 6.5 and an a_w value of 0.98, and for 20°C and a *start-cfu/ml* of $10^{6.5}$ and a pH value of 6.5 and an an a_w value of 0.98, see table 8.2. These two samples will be referred to as 4° *C-sample* and 20° *C-sample*, respectively.

All experiments of both 4° *C-sample* and 20° *C-sample* were done in the same laboratory (Luchansky_95) with the same matrix (cheese). Nevertheless, these data show no log-normal distribution of *cfu/ml*s, according to the Shapiro-Wilk test. Hence, simulations with this data violate assumption 2. In figures 8.4 and 8.5 densities of *logcfu/ml*s at different measuring times are plotted.

Shape of the curves of both 4° *C-sample* and 20° *C-sample* is strongly divers (see figures 8.6 and 8.7). Simulations with this samples violate assumption

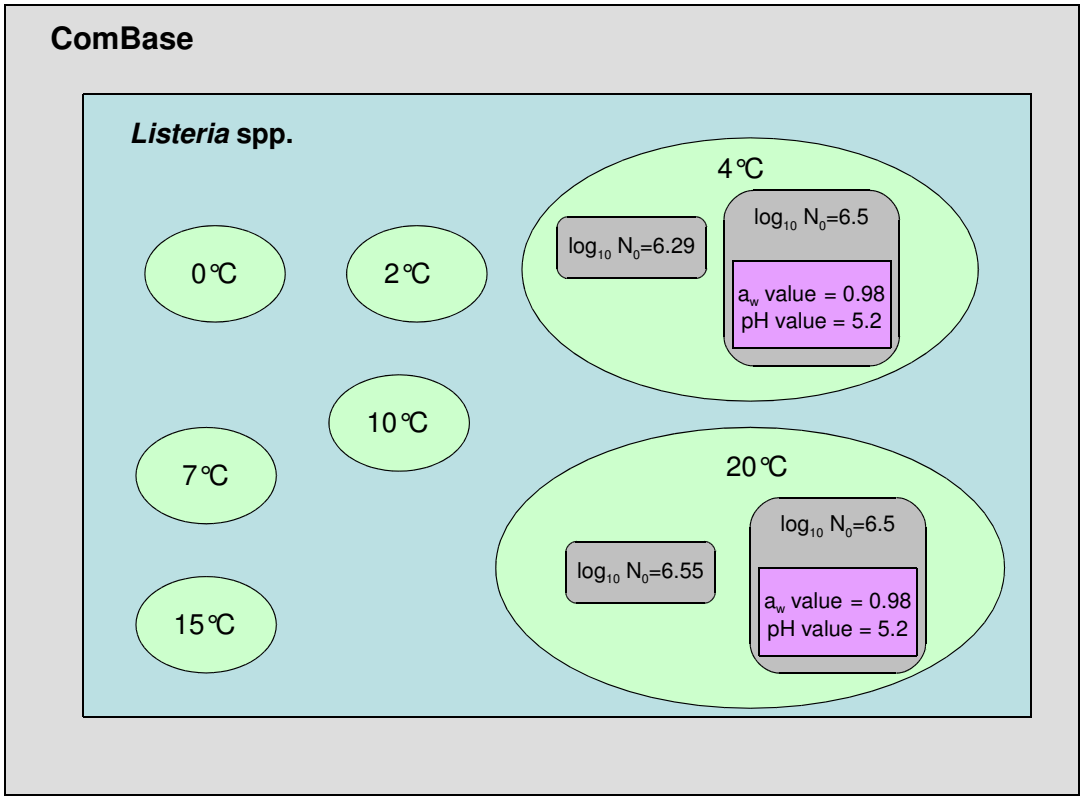


FIGURE 8.3. Selection of data subsets according to temperature, *start-cfu/ml*, a_w value, and pH value

TABLE 8.2. ComBase data: sets with the same temperature, *start-cfu/ml*, pH value and a_w value

	4°C	20°C
$\log_{10}$ ***start-cfu/ml***	6.5	6.5
a_w value	0.98	0.98
pH value	5.2	5.2
number of series	53	54
time points per series	8	5

1. Most of the series of measured values could not be approximated with the parametric models provided by the `grofit` R package [**Kahm et al., 2010**] (see 6.7), i.e., the logistic growth model, the Gompertz model, the modified Gompertz model, and the Richards model. This diversity in population kinetics in comparable experiments, demonstrates the big variability in biological processes, e.g. growth and survival. It highlights the need for probabilistic models, i.e., for including variability and uncertainty.

Simulations were done with both $4°C$-*sample* and $20°C$-*sample*, and methods `cfuinterpolation`, `simpleinterpolation`, `onestep`, and `disconestep`

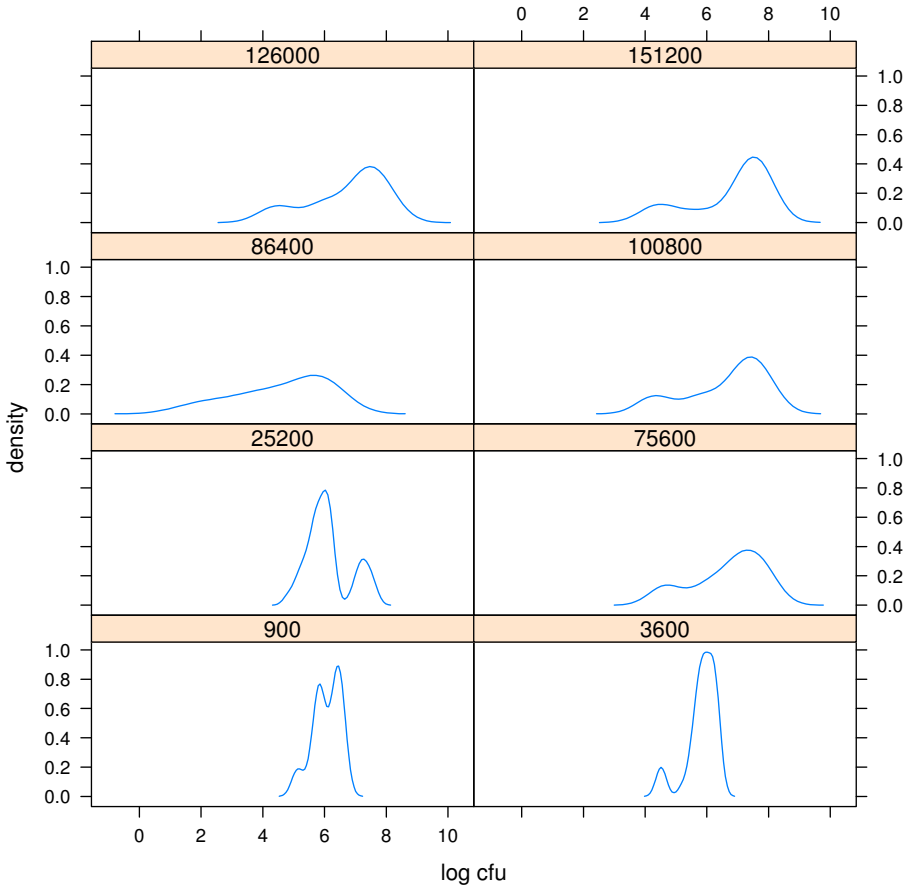


FIGURE 8.4. Experimental data from the ComBase for *Listeria* spp. at 4°C with *start-cfu/ml* $10^{6.5}$, pH 5.2, a_w 0.98: densities of *logcfu/ml*s at different measuring times (in s)

described in section 8.1. Output and input were tested for matching distributions of *calculated cfu/ml*s and *cfu/ml*s with the Kolmogorov-Smirnov test. Results of this statistical testing are listed in appendix E.

First, simulations were done without removal of the measured values corresponding to the interpolation point in time. Results of the Kolmogorov-Smirnov test are listed in section E.2. A summary of the test results is given in table 8.3. Like for sigmoidal artificial data, global linear interpolation (method `onestep`) was not able to reproduce the kinetics. Hence, performance of global linear interpolation did not depend on the probability distribution. For the 4°*C-sample* approximation quality was better when the frequency distribution (`disconestep`) was used, and for the 20°*C-sample* approximation quality was better when the normal distribution (`onestep`) was used.

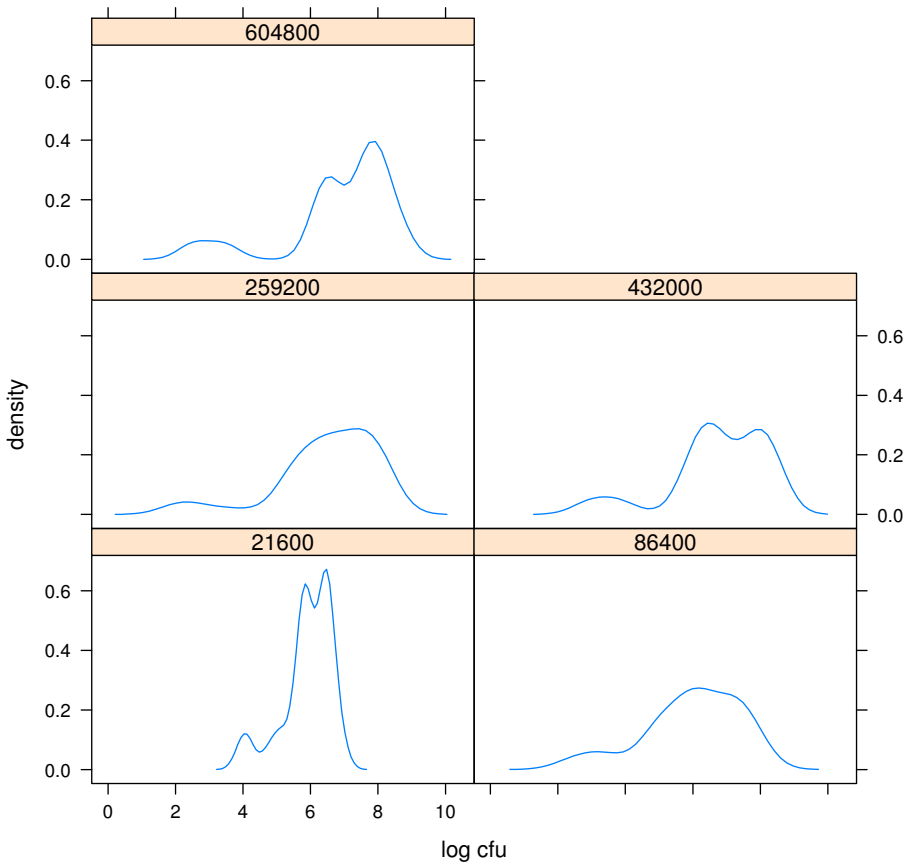


FIGURE 8.5. Experimental data from the ComBase for *Listeria* spp. at 20°C with *start-cfu/ml* $10^{6.5}$, pH 5.2, a_w 0.98: densities of *logcfu/ml*s at different times

Like expected, local linear interpolation (method `simpleinterpolation`) reproduced the kinetics. Although this method is exact at grid point, in two out of eight simulations for the $4°C$*-sample* with method `onestep` the Kolmogorov-Smirnov test yielded different distributions of input and output. This difference is due to the assumption of normally distributed *logcfu/ml*. If a simulation contains only one process step with identical *start-cfu/ml* in every iteration, the set of *final-cfu/ml*s calculated by method `onestep` is log-normally distributed. If the experimental data do not yield a log-normal distribution of *cfu/ml*s, i.e., if assumption 2 is violated, distributions of input and output are different. The Kolmogorov-Smirnov test is a conservative test that favours the null hypothesis, i.e., the hypothesis of identical distributions. Hence, for small samples like $4°C$*-sample* and $20°C$*-sample*, the null hypothesis is rarely discarded.

Then, measured values corresponding to the interpolation point in time were removed before simulations. This manipulation yields a bigger interval

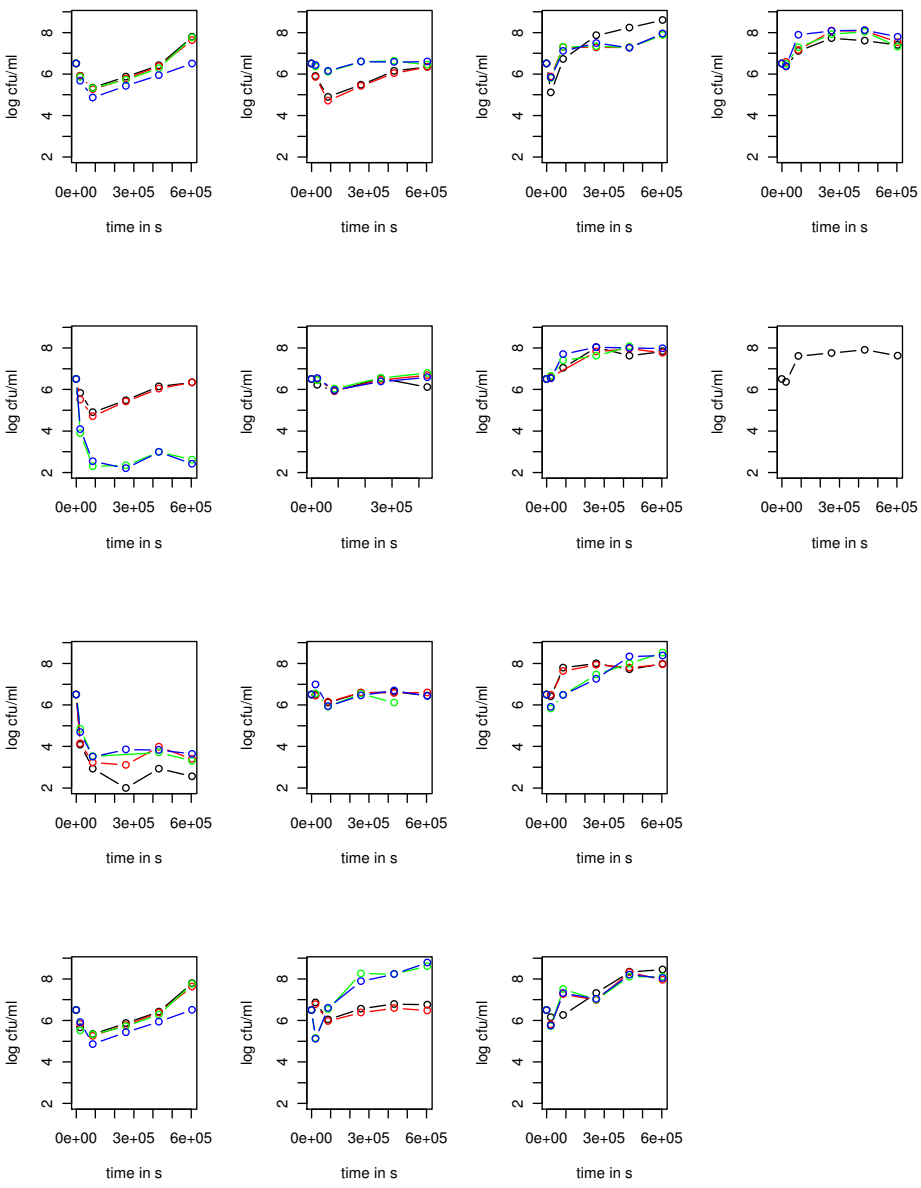


FIGURE 8.6. $4°C$-*sample*: variety in kinetics for similar conditions

between the adjacent measurement times. Hence, approximation quality of global linear interpolation worsens. Additionally, `simpleinterpolation` does not yield exact values like in first simulations described above. Therefore, approximation is inferior. Results of the Kolmogorov-Smirnov test are listed in section E.3. A summary of the test results is given in table 8.4.

Figures 8.6 and 8.7 picture the variability of kinetics in the data sets. Kinetics in the $4°C$-*sample* are more homogeneous than kinetics in the $20°C$-*sample*. Hence, performance of both interpolation methods was better

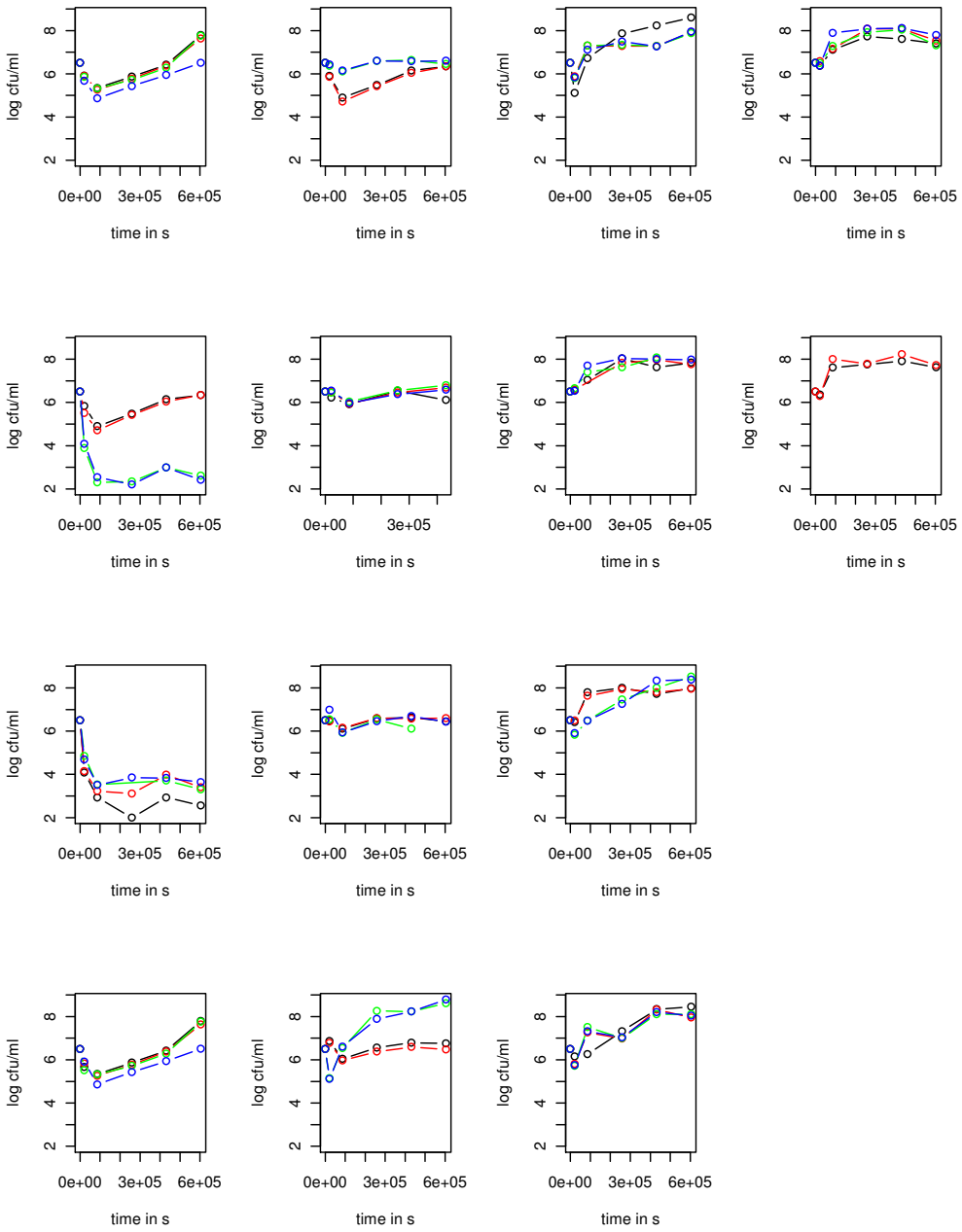


FIGURE 8.7. 20°*C-sample*: variety in kinetics for similar conditions

for the 4°*C-sample*. Additionally, kinetics of some experiments consist of an initial die back, followed by growth. Global linear interpolation is not able to approximate this non-linearity. Approximation of the minimal grid point in the bend of the curves by local linear interpolation (i.e., linear interpolation between the adjacent grid points) can not be exact. Therefore,

TABLE 8.3. Number of simulations that reproduced the data distribution (p-value >5%), simulations were done with 4°*C-sample* and 20°*C-sample*, with different interpolation methods and different assumed probability distributions, and without removal of measured values

		`cfuinterpolation`	`simpleinterpolation`
`onestep`	4°C	2 out of 8 (25%)	6 out of 8 (75%)
	20°C	4 out of 5 (80%)	5 out of 5 (100%)
`disconestep`	4°C	8 out of 8 (100%)	8 out of 8 (100%)
	20°C	2 out of 5 (40%)	5 out of 5 (100%)

TABLE 8.4. Number of simulations that reproduced the data distribution (p-value >5%), simulations were done with 4°*C-sample* and 20°*C-sample*, with different interpolation methods and different assumed probability distributions, and with removal of measured values corresponding to the current interpolation point

		`cfuinterpolation`	`simpleinterpolation`
`onestep`	4°C	2 out of 7 (29%)	6 out of 7 (86%)
	20°C	0 out of 4 (0%)	1 out of 4 (25%)
`disconestep`	4°C	4 out of 7 (57%)	4 out of 7 (57%)
	20°C	0 out of 4 (0%)	1 out of 4 (25%)

approximation of single experiments was poor.

Figures 8.4 and 8.5 show that the experimental data used for validation are not log-normally distributed. Therefore, assumption 2 is violated. Nevertheless, like in first simulations without removal of measured values, approximation quality seems to be independent from the assumed probability distribution, because the Kolmogorov-Smirnov test is a conservative test that rarely discards the null hypothesis for small samples like 4°*C-sample* and 20°*C-sample*.

The `NPMPM` was not able to reproduce the experimental data it was built with. Results from section 8.2 indicate that the default methods could not be fitted to the experimental data with reasonable accuracy. This result demonstrates the need for an adaptive choice from a selection of interpolation methods. This adaptivity will be provided in a future version of the `NPMPM` (see section 9.2).

During validation with experimental data the `NPMPM` was not run for a succession of process steps, because the model was not able to reproduce the data for single process steps. An example for a simulation with the `NPMPM` for a process chain with more than one process step is presented in

section 8.8.

8.4. Requirements on experimental data

Generally, model performance always depends on data quality. Performance of the NPMPM strongly depends on interaction of its parameter settings and experimental data provided as input. Experimental data must be consistent. Series of measured values with more than one measured value at time $t = 0$ have to be discarded.

The NPMPM excludes extrapolation. Therefore, series of measured values must completely cover modelled time intervals. Every series should contain a reasonable number of measured values for allowing a fitting of primary models. The number of measured values essential for proper fitting depends on the interpolation method. If the selected interpolation method does not match the shape of the growth or survival curve, approximation is poor. Therefore, the interpolation method should be chosen according to the data. Hence, it is important to analyse the experimental data before simulation.

In the ComBase [**Anonymous, a**] time is measured in hours with two decimal places. In dairy manufacture heating processes with short time spans in scale of seconds are of interest, see section 4.4. This difference in scales largely increases uncertainty. An approximation of kinetics on a small scale, e.g. seconds, with data on a bigger scale can not model possible oscillations in the smaller scale. Additionally, interpolation may largely underestimate or overestimate bacterial counts. This false estimation is pictured in figure 8.8 for linear interpolation. If data and process steps use the same time scale, the approximation error is smaller than in case of data on a big scale and calculation on a smaller scale.

If the appropriate series of measured values for one process step are not evenly spread over the process-parameter interval, but form clusters at one border, variance of the output may be small. This must not be interpreted as a reliable estimation of a small variability. A large number of data from experiments with conditions matching the process conditions yield a more reliable estimation, but may also yield a wider distribution of *final-cfu/ml*s. The NPMPM counts the process steps with less than 30 series of measured values that exactly match its process parameters, see section 7.6. Distributions of experimental conditions in the selected sets of series of measured values are not memorised. Mostly, if experimental conditions of input data do not reflect the exact process conditions, the output distribution will be spaced out. Hence, it is important to analyse the experimental data for interpretation of simulation results.

Quality and quantity of experimental data not only influence reliability of the output of the NPMPM, but also running time.

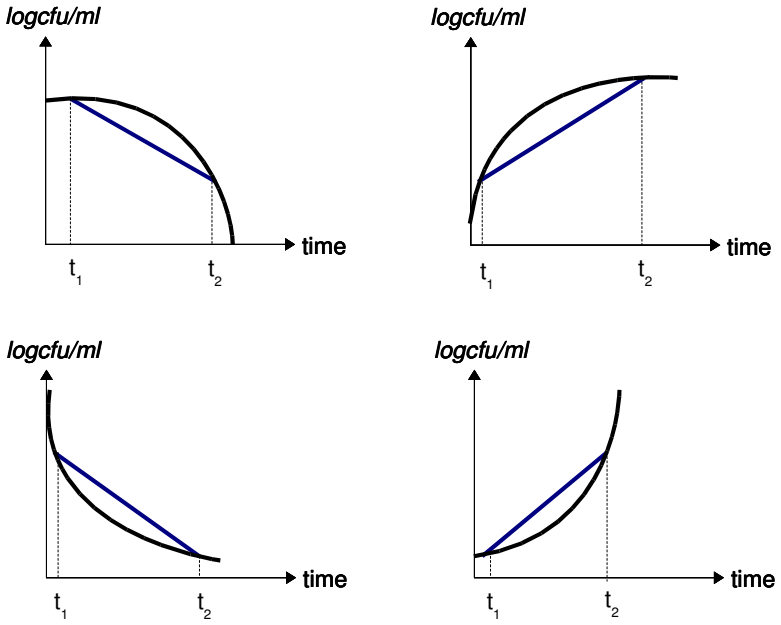


FIGURE 8.8. Scaling problems in linear interpolation, clockwise from upper left: underestimation due to downwards concavity, underestimation due to upwards concavity, overestimation due to upwards convexity, overestimation due to downwards concavity

Huge amounts of data are necessary for modelling bacterial growth and inactivation kinetics. Today, predictive microbiology basically relays on data obtained from non-standardised experiments. Laboratory equipment used may vary between laboratories, leading to slight differences in environmental conditions. There are several methods for counting bacteria that often yield different bacterial counts (see section 3.1). Complex media are not exactly reproducible, because they contain undefined ingredients like yeast extract. The use of a synthetic medium reduces the uncertainty added by a complex medium [**Schaechter, 2006**]. Different bacteria show different kinetics, and the same bacterial culture may show variation in kinetics depending on the history of the cells (see section 3.3). Hence, experimental data generated in different laboratories, with different growth media, and with various strains and biovars of the same species or genus, is afflicted with a lot of uncertainty. Even repeated experiments done in the same laboratory, by the same person, with the same batch of growth medium, and the same bacterial culture, yield varying bacterial counts. Figures 8.6 and 8.7 picture variability of kinetics in data sets obtained under similar conditions.

We recommend the definition of a standard in regard to data production for use in predictive microbiology. This would be a step towards homogeneous and comparable data for the development of reliable predictive models. We agree with Barbosa-Cánovas *et al.* [**Barbosa-Cánovas et al., 2005**] and Schaechter [**Schaechter, 2006**] that a publication of growth and

inactivation kinetics of bacteria should always contain a supplemental section with detailed growth conditions. Fortunately, it is not necessary to develop a new database for collecting standardised experimental data. The Combase [**Anonymous, a**] is a good approach of a centralised and open source platform for data collection that could be extended to fit this purpose.

8.5. Sensitivity analysis

Because output of the `NPMPM` is a distribution of *cfu/ml*s, a direct comparison of corresponding values is impossible. Interaction of experimental data with selected interpolation method, probability distribution selected for calculation of the *end-cfu/ml*, and process parameters is complex.

Quality and quantity of experimental data influence approximation behaviour (see section 8.4. If series of measured values can not be fitted with the selected interpolation method, performance is poor. Kinetics may be highly variable, as in experimental data used in section 8.3. Hence, selection of the best-fitting interpolation method for every series of measured values, i.e., an adaptive selection of interpolation methods, would yield a better approximation of the distribution of *final-cfu/ml*s.

Quantity and distance of measuring times, i.e., grid points, has an impact on interpolation quality. Different primary models need different quantities of measurement points for fitting. Distance of the measurement points influences the interpolation error. Large distances may be an effect of different scales in experimental data and process conditions, like shown in figure 8.8. In a typical sigmoid growth curve data points from different parts of the curve contain different information, e.g. *start-cfu/ml* N_0 does not yield information about carrying capacity K. If a series of measured values does not cover all parts of the growth or death curve, approximation by primary models like the logistic growth model is bad or even impossible. Hence, the method for selection of appropriate series of measured values must take into account requirements of the interpolation method, if selection of this method is not adaptive.

If there are only few series of measured values that match process conditions, or if distribution of series of measured values in the range of conditions is not homogeneous, uncertainty is big. Quantity of appropriate series of measured values distribution of experimental conditions in the database must be analysed before simulation, for allowing assessment of simulation output. Approximation error of the `NPMPM` highly depends on the experimental data.

The probability distribution selected for calculation of the *end-cfu/ml* in one iteration of the inner loop (algorithm 3) largely affects the distribution of the output. Default setting is a normal distribution of *calculated logcfu/ml*s, see assumption 2. Bacterial counts measured during different experiments with identical conditions at the same point in time may not follow a log-normal distribution, as was shown in section 8.3. If the data do not yield

a normal distribution of *logcfu/ml*s, the NPMPM may better perform with the frequency distribution of the *calculated logcfu/ml*s instead of a normal distribution. When the probability distribution of bacterial counts is known, it should be used, because this adds prior information and reduces uncertainty.

Errors may sum up along a succession of process steps. Therefore, number of process steps in the process chain influences approximation behaviour, too.

Uncertainties in process parameters, e.g. expert opinion, must be taken into account in interpretation of simulation results. Especially in quantitative microbial risk assessment it is essential to make uncertainties transparent [**Boone et al., 2005**]. Therefore, at times it may be useful to run the NPMPM twice with different parameters, e.g. with different temperature ranges in one process step, and to compare both outputs. When variability and uncertainty are high, usually the distribution of the output of the NPMPM expands.

8.6. Running time

A classical running time analysis of the NPMPM is not possible, because running time strongly depends on the relation of process conditions and experimental data. Running time of the NPMPM depends on the error bound given as input, number of process steps in the process chain, interpolation method, criteria for and method of search in the data frame containing series of measured values, and on quantity and quality of experimental data. In every process step the data frame with experimental data is queried. If there are enough applicable experiments for every process step of the process chain the data frame with the data must only be queried once. It is queried until a certain minimal number of series of measured values is found, see algorithm 5 for a description of the default settings. Hence, number of queries depends on search parameters (process step conditions), number of entries in the data frame, and number of series of measured values matching the process step parameters. If there are at least 30 series of measured values matching parameters of the process step, the data are queried once for every process step.

Running time of one simulation is a multiple of running time of one iteration (outer loop). The error bound influences the number of iterations. Running time of one iteration is a product of the number of process steps and the running times for calculation of *end-cfu/ml*s of process steps.

For every selected series of measured values an interpolation is done. Its running time depends on the method of interpolation. The NPMPM provides both global and local linear interpolation; with number of measured values in the series n, running time is in $O(n^3)$ and $O(n)$, respectively. If parameters of the interpolation curves are stored in the data frame `raw_data`, running time is linear and does not depend on the interpolation method (see also chapter 9).

8.7. Limitations

There are restrictions to applicability of the NPMPM. The model does not allow for parallel processes, e.g. different treatments of skimmed milk and cream after separation of raw milk. Those process steps have to be combined to one single process step with the process step factor (see section 7.4) taking into account for the combined effect of the parallel process steps. This combination of process steps with different conditions increases uncertainty.

The NPMPM can not map continuous changes in parameters. Process steps with changing conditions, e.g. heating or cooling steps, can be modelled in two different ways. One approach is segmentation of process steps into artificial sub-steps. For every sub-step a set of appropriate series of measured values is selected, therewith ensuring the coverage of the whole parameter range by experimental data. For example in yoghurt production the pH value of the milk inoculated with a starter culture drops from 6.5 to 4.5 during hours of incubation. This may be modelled as a succession of process steps with smaller pH ranges, like pictured in figure 7.3.

Change in conditions can also be taken into account as parameter range, e.g. heating from 50°C to 60 °C may be modelled as process step with a temperature range of 50°C to 60°C. This increases uncertainty, especially when experimental data fit only a small area of the parameter range as pictured in figure 8.9. A data analysis is necessary for detection of such an unfavourable distribution of series of measured values. The small number of series of measured values between 55°C and 60°C is detected during simulation if the process step is segmented into two successional process steps with temperature ranges from 50°C to 55°C and 55°C to 60°C, respectively.

Assumption 4 states that the inoculum size has no influence on population kinetics. This assumption is false, but it is necessary if there are not enough experimental data available to take into account for varying *start-cfu/ml*s. When series of measured values with different inoculum sizes are used to calculate the *end-cfu/ml* of one process step, the interpolation method must make adjustments for inoculum sizes.

A global linear interpolation yields a straight line that can easily be shifted by adjusting the intercept. For other interpolation methods this shift may yield serious underestimation or overestimation of bacterial counts. Simple shift of a logistic growth curve is pictured in figure 8.10. Maximal bacterial counts of down-shifted curves are underestimated. A local linear interpolation underestimates bacterial counts both in log phase and in stationary phase. Fitting of a sigmoid curve and adjusting the carrying capacity would lead to a better approximation. This adjustment does not include possible shape changes of the growth and survival curves due to different inoculum sizes. For survival experiments a simple up-shift of the

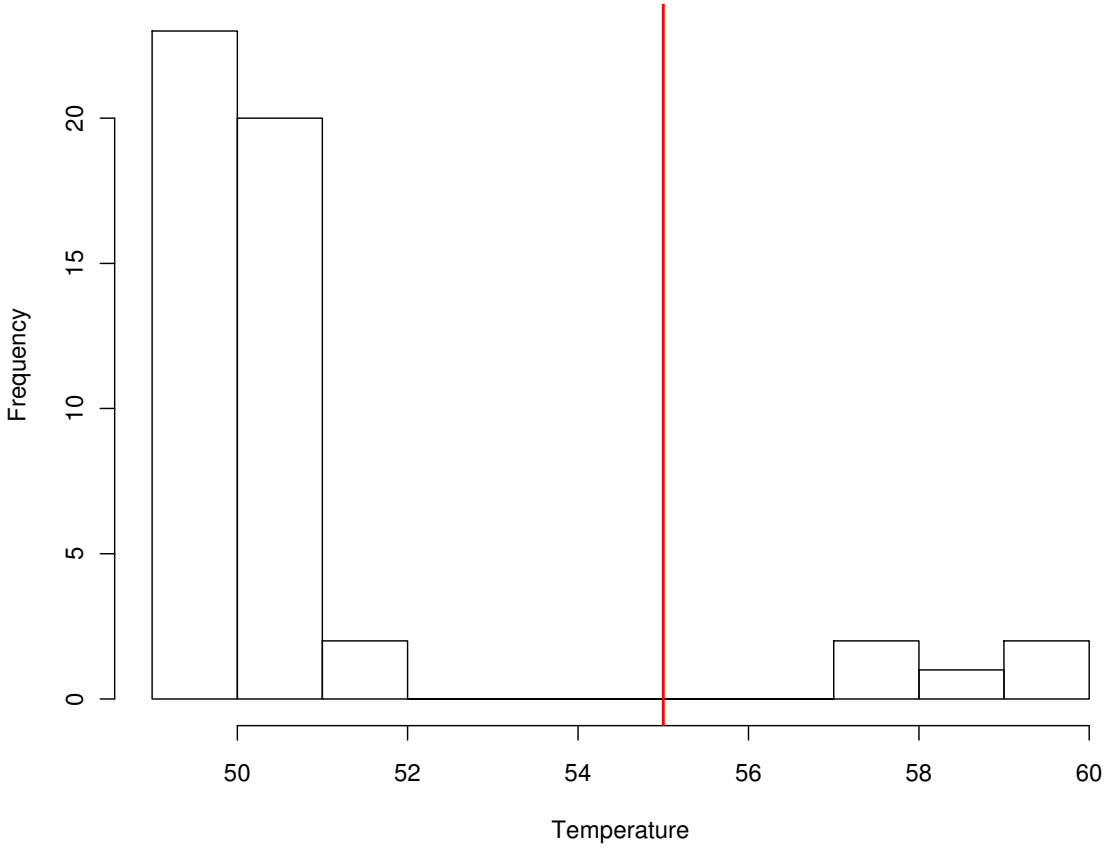


FIGURE 8.9. Unfavourable distribution of series of measured values between 50°C and 60°C, the red line marks the temperature mean

survival curve yields an ineradicable population, see figure 8.11.

The NPMPM models growth and death of viable cells from single species. Mixed cultures are only taken into account indirectly by incorporating parameter changes, e.g. due to fermentation of milk by *lactobacilli*. For modelling competition, experimental data for mixed cultures are needed. Spore and dormant cells are not taken into account, because kinetics of combination of decay of viable cells and spores, sporulation and germination of spores is usually very complex.

8.8. Example

In this section, an example for simulation with the NPMPM for a single process chain is presented. Output of the NPMPM for manufacture of proper pasteurised fresh milk is compared with the output for a process chain with malfunctioning pasteurisation. In the dairy plant, improper pasteurisation should be detected. Preparation of input data frames is demonstrated.

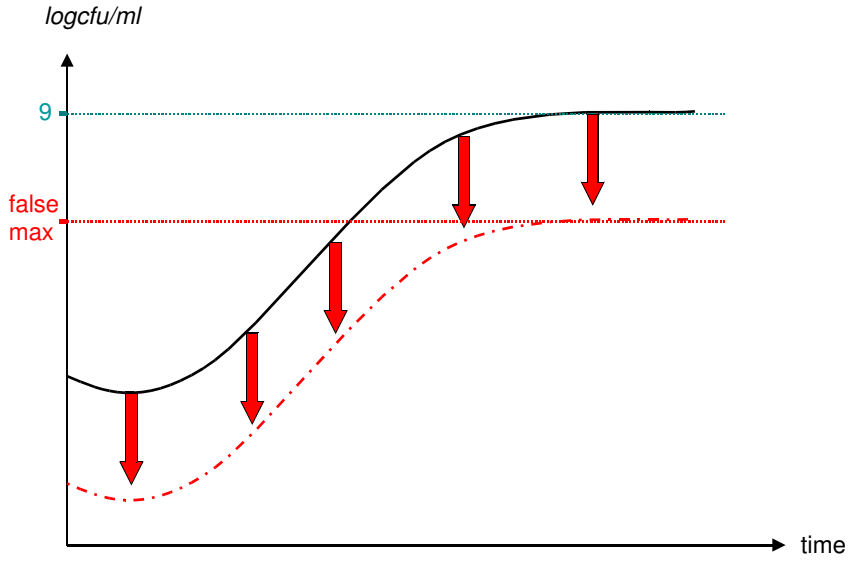


FIGURE 8.10. Simple down-shifting of the *start-cfu/ml* in a sigmoid growth curve

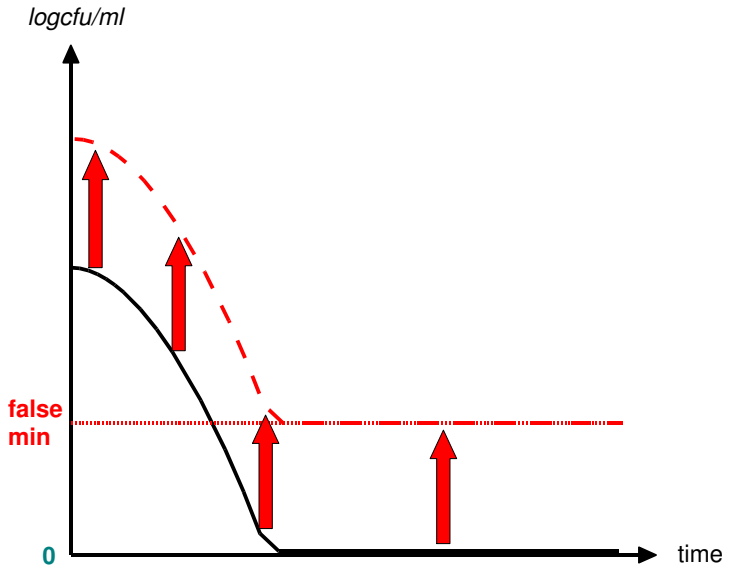


FIGURE 8.11. Simple up-shifting of the *start-cfu/ml* in a sigmoid survival curve

The general processing of fresh milk is described in section 4.5. In figure 8.12 a process chain for the manufacture of fresh milk content is pictured; its layout was taken from [**Bylund, 2003**]. We use it to generate data frames with the process steps and the associated (intermediate) milk products, respectively.

Literature	Model
Raw milk storage 4-6°C, 4-24h	Raw milk storage 4-6°C, 4-24h, pH 6.5-6.7
Separation, fat standardisation 55-65°C	Separation, fat standardisation 55-65°C, 5-10s, pH 6.5-6.7, factor 0.9 (dilution)
Homogenisation 55-80°C, 10-15µs	Homogenisation 55-80°C, 1-2s,pH6.5-6.7
Pasteurisation 72-75°C, 15-30s	Pasteurisation 72-75°C, 15-30s, pH 6.5-6.7

FIGURE 8.12. Schematic presentation of one possible process chain for pasteurised fresh milk from [**Bylund, 2003**], and the corresponding (intermediate artificial) process steps

Production parameters differ between dairy plants, and exact process parameters are company secrets. Additionally, there are fluctuations in processing conditions due to technical reasons. Hence, for some process steps parameters are not given explicitly in the literature. a_w values of dairy products are rarely found in the literature. Conversion from water content to the a_w value is not possible for such a complex matrix like milk. Therefore, the a_w values were set to NA. Changes in pH and a_w value during processing often are not quantified. For running the example simulations all missing parameters were estimated from literature sources or expert opinions. Process steps with highly, and continuously changing parameters have to be fragmented into distinct process steps, yielding moderately changing parameters in every intermediate artificial process step. All parameters in the process steps are assumed to follow uniform distributions, because there is no information about their probability distribution [**Anonymous, 2010a**].

A single contamination with 10^5 *Listeria* spp. per ml is assumed in the first process step, i.e., contamination of raw milk. Default obligatory input parameters were used, e.g. global linear interpolation and a log-normal distribution of *cfu/ml*s. Experimental data for calculation of the bacterial load at the end of each process step was extracted from the database (see section 7.6 for details), and the unit of time was converted from hours to seconds.

In figure 8.12 process steps and parameters are listed. Raw milk is stored at 4°C and may be stored between 4h and one day before processing. Manufacture continues with separation at 55-65°C and recombination of skimmed milk and cream to the desired fat content. For fat standardisation, a duration of 5-10s was taken from the literature. A dilution factor of 0.9 is estimated. In all other process steps there is no dilution or enrichment

(factor one). During following homogenisation the temperature is increased to 55-80°C, with an assumed duration of 1-2s. After that the milk is pasteurised at 72-75°C for 15-30s. In case of improper pasteurisation a temperature of 55°C (lowest possible temperature during homogenisation) is assumed.

For improper pasteurisation there were iterations with viable bacteria in the end product. For proper pasteurisation there were some iterations with viable bacteria in the end product, too. With information from chapter 4, after proper pasteurisation all *Listeria* are expected to be killed off. The reason for this discrepancy is that the default methods are not able to reproduce the experimental data (see section 8.3).

The .pdf-file built from the output of the simulation for improper pasteurisation is pictured in figures 8.13 and 8.14. It contains a short description of simulation settings. This description was passed as input text. The single contamination with 1000 bacteria per ml in the first process step is given as inoculum sizes.

A summary statistics of the distribution of *final-cfu/ml*s with minimum, maximum, quartiles and mean is provided. The minimum is zero, showing that in some iterations all bacteria were assumed to be killed off. In 25% of the iterations a contamination with less than 635 bacteria per ml was calculated. Another 25% of the iterations computed values between 635 and $3.2 \cdot 10^3$ bacteria per ml. In 25% of the iterations values between $3.2 \cdot 10^3$ and $4.4 \cdot 10^4$ bacteria per ml were computed. In the remaining 25% of iterations, values between $4.4 \cdot 10^4$ and $5.1 \cdot 10^6$ bacteria per ml were computed. Due to the high maximum value and third quartile, the mean of *final-cfu/ml*s is $1.2 \cdot 10^5$. The output distribution is pictured in a histogram and a density plot. The bandwidth given in the density plot is a smoothing parameter that is used in estimation of the density function.
The relative error bound of 1% that is used as stop criterion (see algorithm 1) is noted down. In this particular simulation, the stop criterion was fulfilled after the minimum amount of 101 iterations. There was no process step with conditions that could not be matched by experimental data (*NAwarning*). For every process step at least 30 appropriate series of measured values were selected, hence, there were no *countwarnings*. Finally, default settings and input settings of the simulation are listed.

It is essential to supplement simulation results with a data analysis. Without an analysis of the experimental data used for fitting, uncertainty due to data availability, and therewith model performance, can not be assessed (see section 8.4). The distribution of the simulation output can be smaller than the real probability distribution of bacterial concentrations, if there are only few series of measured values matching the process step conditions, or if the data are clustered and not homogeneously distributed in the parameter interval.

The information provided by the .pdf-file allows for a discussion of the results and a rerun of the simulation. If simulation results are coupled with consumption data, the impact of a contamination of a food supply chain with a certain microorganism can be assessed, e.g. as number of people affected, or as expected costs.

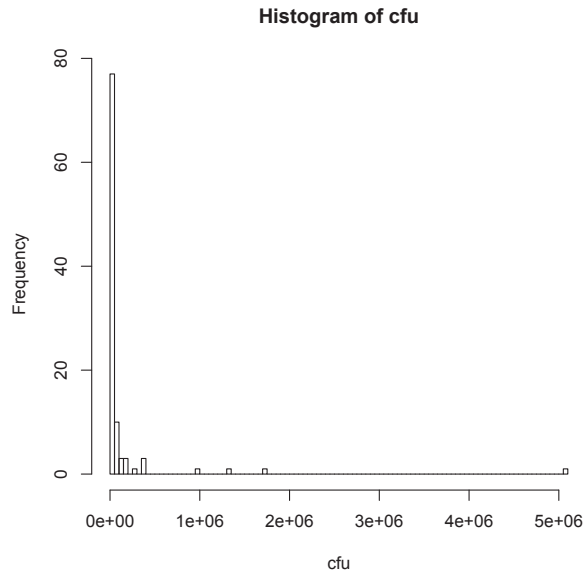
pdfoutput NPMPM
2011-03-07 12:18

```
> textt

[1] "Listeria monocytogenes in the process chain of fresh milk - improper pasteurisation"
```

Summary statistics of the calculated cfus

```
> summary(cfu)

     Min.   1st Qu.
      0.0     634.3
   Median      Mean
   3185.0  124800.0
  3rd Qu.      Max.
  44480.0 5092000.0
```

Histogram of the cfus calculated



Inoculumsizes

```
> inoculum

[1] 10000     0
[3]     0     0
```

Number of iterations in the simulation:

1

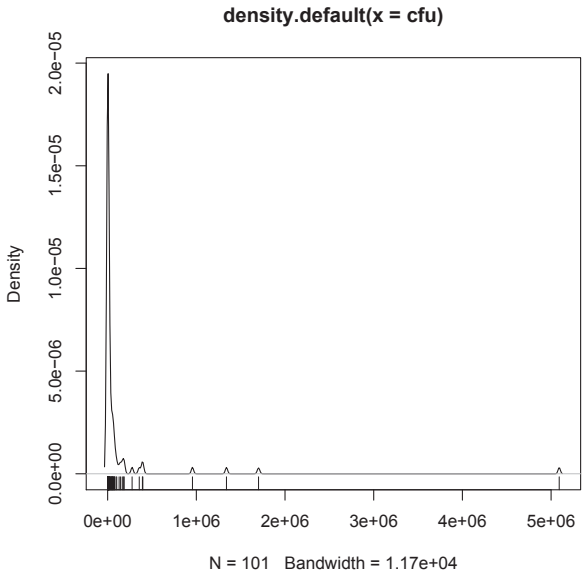
FIGURE 8.13. The first page of the .pdf-file generated from the output of one simulation with the NPMPM for *Listeria* spp. in the process chain for fresh milk with improper pasteurisation

```
> length(cfu)

[1] 101
```

Chosen relative upper error bound for the fluctuation of the mean of calculated cfus

```
> errorb

[1] 0.01
```

Density plot of the cfus calculated



Were there process steps with no appropriate series of measured values:

```
> NAwarning

[1] "No"
```

Number of other process steps with less than 30 appropriate series of measured values:

```
> countwarnings

[1] 0
```

Input settings default settings: `errorb=0.01, intexmethod=cfuinterpolation, apsomvmethod=apsomv, pdistpsteps=psteppar, lastiteration=numberiterations, calconestep=onestep`

```
> inputsettings

npmpm(inoculum = inoculum, intexmethod = cfuinterpolation)
```

2

FIGURE 8.14. The second page of the .pdf-file generated from the output of one simulation with the `NPMPM` for *Listeria* spp. in the process chain for fresh milk with improper pasteurisation

CHAPTER 9

Conclusion

In this thesis, a new probabilistic model in predictive microbiology, the NPMPM, was presented. It makes two major contributions to the field of predictive microbiology. It was designed for use in microbial risk assessment of food processing, i.e., for quantification of the probability of a contaminated end product in case of contamination during processing. An introduction into the biological and technical background was given, and existing approaches in predictive microbiology were introduced. The NPMPM was presented in chapter 7. Its performance was discussed in chapter 8. In this section, we summarise its contribution to predictive microbiology and indicate some future work.

9.1. Contribution

This thesis comprises two major contributions to the field of predictive microbiology.

- The NPMPM is a new approach for including variability and uncertainty into predictive models. Unlike in existing approaches, variability and uncertainty are not included in random processes or as model parameters that follow probability distributions. The NPMPM fits deterministic primary models to the experimental data, instead of fitting one probabilistic primary model to pooled series of measured values. This approach takes into account for differences in shape of growth and survival curves, and allows for an adaptive choice of the best-fitting primary model for every series of measured values. Therewith the amount of uncertainty introduced during model fitting is reduced in comparison to existing approaches.

- The NPMPM focuses on local approximation quality, i.e., it minimises the local approximation error. It is designed for prediction of the distribution of bacterial counts at one certain point in time. Usually, models in predictive microbiology aim at approximating and explaining kinetics of the whole growth or survival curve, i.e., the global approximation error is minimised. For modelling the propagation of a contamination during food processing, point estimation of the distribution of bacterial counts at the end of the process step must be as accurately as possible. It is not essential to exactly predict population kinetics during the entire process step and beyond process step duration. The new approach realised in the NPMPM will

be a particular benefit for risk assessment of contaminations in the food supply chain.

These two contributions both improve upon the current methods of the research community.

9.2. Future work

Next step in modelling will be to include an adaptive selection of interpolation methods. This adaptive choice will allow for using the best-fitting interpolation method for every single series of measured values. For this purpose, different primary models and a fitting routine, e.g. method of least squares, will be implemented. For a certain series of measured values, selection and parameter-calculation of the best-fitting interpolation method do not have to be repeated in every simulation, if results are saved.

Secondary models could be included into the `NPMPM`. If a secondary model is validated for certain conditions, e.g. dependency of maximal growth rate on temperature for a certain microorganism, or influence of inoculum size on kinetics, it may be used for calculation. In combination with corresponding primary models, a secondary model yields artificial series of measured values. In case of process step parameter ranges, artificial series of measured values can be generated for supplementing experimental data that do not homogeneously cover the whole parameter range. Use of the artificial series of measured values must be compared with use of original experimental series of measured values from which the secondary model was derived. For using secondary models, these methods have to be implemented and incorporated into the R package `NPMPM`.

Original experimental data are preferable to deterministic primary models, because a set of series of measured values contains uncertainty and variability, whilst deterministic models do not. Probabilistic and stochastic primary and secondary models yield distributions of series of measured values. It has to be investigated if use of probabilistic models is preferable to original experimental series of measured values. Effects on simulation output of combination of artificial series of measured values and experimental data have to be assessed.

Weights may be used to take into account for combination of artificial and experimental data, or for experimental data of different quality. Some experimental data may contain less uncertainty, e.g. because information about the history of the cells is available. This differences may be compensated by assigning higher weights to reliable data. When series of measured values are not homogeneously distributed over the range of process conditions, weights may be used to modulate the influence of certain series on *end-cfu/ml*. Investigation is necessary for decision how those weights may be calculated.

For some microorganism only few data are available. Particularly tenacity data for microorganisms of biosafety level 3 and 4 are rare. For modelling

of growth and survival of such microorganisms, sometimes surrogate germs are used, e.g. *Bacillus cereus* is used as substitute for *Bacillus anthracis*. If experimental data from surrogate bacteria are provided in the data frames with the series of measured values, these data are used by the `NPMPM` for simulation. For adding weights to assess global differences in behaviour, quantification of similarities between bacteria is essential.

Structure of the database will be extended. For shortening of running time, fitting information can be stored alongside environmental conditions, e.g. name of the best-fitting method and calculated parameters. If secondary models are included into the `NPMPM`, information about the microorganisms and conditions these models are validated for has to be saved in the database. Weights assessing similarities between microorganisms have to be stored in the database, too. Data from experiments with mixed cultures must be stored in a way that allows for identification of all microorganisms.

Use of probability distributions of inoculum sizes instead of deterministic values will be included into the `NPMPM`.

Methods in predictive microbiology rely on a large amount of high-quality data. For taking into account factors that influence population kinetics, data used for fitting a predictive model must provide information about these factors. These information are, in particular:

- **Species, stem and biovar** of the microorganism: Population kinetics largely differ between different species, but also between stems and biovars of the same species. Knowledge of the particular biovar used in experiments reduces uncertainty. Additionally, it allows for assessment of biological variability.
- **Previous cultivation conditions**: The history of the cells used in experiments largely influences kinetics. In survival experiments, tenacity can be enhanced or reduced by previous cultivation conditions. In growth experiments, previous cultivation under similar conditions shortens the lag phase, whilst a change in conditions yields a longer lag phase. Without information about the history of the cells used in experiments, these effects can not be modelled.
- **Environmental conditions**: Kinetics of growth and death depend on the environmental conditions. Temperature, water content and pH value of the matrix are usually provided with experimental data. The a_w value has a bigger influence on growth and death kinetics than the water content, but it is seldom provided with experimental data; for complex matrices like food it can not be calculated from water content. Information about these environmental conditions is essential for prediction of population kinetics. Additionally, salt and fat content, and presence of preservatives or concurrence flora should be provided to reduce uncertainty in predictive models.
- **Cultivation conditions**: Population kinetics may be different in cultivation on solid media or in liquid media. On solid media, gas, e.g. oxygen, is freely available for the bacteria, whilst nutrients

and metabolite are distributed very slowly by diffusion. In liquid media, nutrients and metabolites are distributed more quickly, but oxygen and other gas are distributed slower by diffusion. Bacterial cultures in liquid media can be incubated on a shaker that provides good mixture and ventilation, and prevents sedimentation of the cells. These different cultivation conditions can lead to differences in growth and death kinetics, even under similar environmental conditions.

- **Volume of bacterial cultures**: The volume of bacterial cultures is of special interest in experiments that investigate temperature dependencies. Small volumes respond quickly to temperature changes, e.g. heating of small tubes in a thermocycler. Bigger volumes, e.g. raw milk in tanks of a dairy, only allow for slower temperature changes. For avoiding scaling problems in predictive modelling, it is useful to take into account the volume of bacterial cultures.

Today, data collection is not standardised. When collecting data in the future, providing all these information will lead to an improvement of prediction quality: Existing probabilistic models may fail in describing variations in population kinetics, that may be even found in experiments with matching experimental conditions (see section 7.9). If more experimental parameters are taken into account, experiments could be better described. Experimental data could be partitioned into groups according to experimental parameters. This would lead to more homogeneous kinetics in this groups, and to a better fit of primary models.

APPENDIX A

Abstract (German)

In dieser Arbeit wird ein neues probabilistisches Modell in der prädiktiven Mikrobiologie, das `NPMPM`, präsentiert. Die Einbeziehung von Variabilität und Unsicherheit in die Modellierung folgt einem neu eintwickelten Ansatz. Das `NPMPM` wurde zur Verwendung in der Risikobewertung von bakteriellen Kontaminationen der Lebensmittelkette entwickelt. Es wird am Beispiel einer Kontamination der Milchkette mit *Listeria monocytogenes* vorgestellt.

Lebensmittel, die pathogene Keime oder bakterielle Toxine enthalten, können zu lebensmittelbedingten Erkrankungen führen. Eine Abschätzung von Wachstum und Tenazität von Bakterien während der Lebensmittelproduktion und während der Lagerung des fertigen Lebensmittels in Abhängigkeit von verschiedenen Faktoren ist deshalb von besonderem Interesse für eine Garantie der Lebensmittelsicherheit. Der Einsatz prädiktiver Modelle ist hierfür von großem Nutzen, da eine experimentelle Untersuchung der Populationskinetiken im Labor einen sehr hohen zeitlichen und personellen Aufwand erfordert und damit konstenintensiv ist.

Biologische Prozesse wie Wachstum und Absterben sind oft hochgradig variabel. Existierende Modellansätze berücksichtigen Variabilität und Unsicherheit, indem sie die biologischen Prozesse selbst als stochastische Prozesse modellieren, oder indem Parameter nicht als konstant, sondern als wahrscheinlichkeitsverteilt definiert werden. Das `NPMPM` nutzt einen anderen, neu entwickelten Ansatz. Es berechnet Stichproben möglicher Bakterienkonzentrationen mit Hilfe deterministischer Modelle. Aus einer solchen Stichprobe werden die Parameter der zugrundeliegenden Wahrscheinlichkeitsverteilung geschätzt.

Diese Arbeit ist interdisziplinär. Ihr erstes Kapitel erläutert die Motivation des Forschungsansatzes und gibt einen Überblick über ihren Aufbau. Das zweite Kapitel beschreibt Ursachen und Auswirkungen lebensmittelbedingter Erkrankungen. Im dritten Kapitel werden Einflussfaktoren auf die Wachstums- und Absterbekinetik von Bakterienpopulationen diskutiert. Das vierte Kapitel liefert eine Zusammenfassung der Grundlagen der Milchverarbeitung in der Bundesrepublik Deutschland. Im fünften Kapitel werden existierende Modelle in der prädiktiven Mikrobiologie dargestellt. Das sechste Kapitel gibt einen Überblick über die verwendeten Daten und Methoden und diskutiert die Modellannahmen. Im siebten Kapitel wird das `NPMPM` vorgestellt. Validierung des Modells, Vorstellung und Diskussion der

Ergebnisse erfolgen im achten Kapitel. Abschließend wird im neunten Kapitel der Beitrag zum Forschungsfeld zusammengefasst und ein Ausblick auf weitere Forschung gegeben.

APPENDIX B

Abstract (English)

In this thesis a new probabilistic model in predictive microbiology, the NPMPM, is presented. It is based on a new approach for including variability and uncertainty. The NPMPM was developed for risk assessment of bacterial contaminations in the food supply chain. It is introduced by the example of a contamination of the milk supply chain with *Listeria monocytogenes*.

Human illness that results from the consumption of contaminated food may be caused by bacteria and their toxins. Hence, assessment of growth and tenacity of bacteria during food production processes and storage is vital for food security. Predictive modelling is used to forecast the development of microorganisms in food, depending on different influence factors. Getting the desired information solely by means of laboratory experiments is time and cost intensive.

Biological processes like growth and death are highly variable. Existing approaches that take into account variability and uncertainty either assume parameters to follow probability distributions, or model a biological process as stochastic process. The NPMPM follows a newly developed approach. It calculates samples of possible bacterial concentrations by means of deterministic models. Such a sample is used for estimation of the probability distribution of the corresponding population.

This thesis is an interdisciplinary one. In the first chapter the research approach is motivated, and an overview of the organisation of this thesis is given. The second chapter describes causes and impact of foodborne diseases. Chapter three discusses influencing factors on growth and survival kinetics of bacterial populations. In chapter four dairy manufacturing in Germany is summarised. The fifth chapter presents existing models in predictive microbiology. Chapter six gives an overview of data and methods used, and a discussion of model assumptions. In chapter seven the NPMPM is introduced. The model is validated in chapter eight, and results are discussed. Finally, in chapter nine the contributions to the field of predictive microbiology are summarised, and some future work is suggested.

APPENDIX C

Terminology

β	growth rate
η	decline rate
λ	lag time duration
γ	expected number of offspring per time step
μ	rate of population change, includes both growth and decline
μ_m	maximum specific growth rate
σ	standard deviation
σ^2	variance
A	asymptote of a growth curve, $A = \ln \frac{K}{N_0}$
autonomous equation	a differential equation that does not depend on the independent variable, e.g. time
calculated cfu/ml	calculated bacterial concentration
calculated logcfu/ml	decadic logarithm of the calculated bacterial concentration
CDC	Centers of Disease Control (USA)
cfu	colony forming unit
cfu/ml	cfu in one ml (measured in an experiment)
closed-form expression	an expression that is a combination of a finite number of ´well-known´ functions

data frame	a matrix that may contain different data types
end-cfu/ml	concentration of bacteria at the end of a process step
ESL	Extended Shelf Life
FDA	U.S. Food and Drug Administration
final cfu/ml	calculated concentration of bacteria in the end product
generation time	time needed for doubling the number of cells
growth rate	number of cleavages per time
HACCP	*Hazard Analysis and Critical Control Points*, a system used to identify and prevent food-safety problems in production, processing, and distribution of foods
hazard	an agent or action that can cause adverse effects
HSP	heat shock protein
HTH	*high temperature heating* pasteurisation
HTST	*high temperature/short time* pasteurisation
inoculation	placement of something that will grow or reproduce
inoculum	microorganism used in inoculation
K	carrying capacity
kinetics	a term describing the dynamics of bacterial growth and decay, e.g. the velocity of changes in population size
logcfu/ml	decadic logarithm of cfu/ml
LTLT	*low temperature/long time* pasteurisation
matrix	medium in which bacteria are grown
$M(t)$	expected number of offspring of one single cell
mixed culture	a culture that contains more than one species (stem, biovar) of bacteria

N_0	start population size (at time $t_0 = 0$), $N_0 = N(0)$
non-autonomous equation	a differential equation that depends on the independent variable, e.g. time
`NPMPM`	new probabilistic model in predictive microbiology
$N(t)$	population size at time t
predictive model	model of the responses of bacterial populations to intrinsic and extrinsic conditions
pure culture	a culture that contains only one species (stem, biovar) of bacteria
start-cfu/ml	concentration of bacteria at the beginning of a process step
t	time
T	temperature in °C
t_0	start time, $t_0 = t = 0$
tenacity	persistence of bacteria
UHT	*ultra-high temperature* processing
y_0	relative population size at start, $y_0 = y(0)$
$y(t)$	relative population size, $y(t) = \ln \frac{N(t)}{N_0}$
YOPI	*young, old, pregnant, immunocompromised* people

APPENDIX D

The R package

The model NPMPM presented in this thesis is implemented in the R programming language [**R Development Core Team, 2010**]. It is downloadable from CRAN (http://cran.r-project.org/), and it is licensed under the GNU General Public License [**Anonymous, b**]. In this appendix, the documentation of the R package and its methods is given. The structure of dependencies (calls) of methods provided by the package is pictured in figure D.1. This methods are described in table D.1.

Part of the information essential for a simulation must be passed as input to the methods, other information must be provided as data frames in the global environment of the R session. The structure of this data frames is given in table 7.2 in section 7.6.

Documentation. `Package: NPMPM`

```
Type: Package
Title: tertiary probabilistic model in predictive microbiology for
use in food manufacture
Version: 1.0
Date: 2011-03-18
Author: Nadine Schoene
Maintainer: Nadine Schoene <nadine.schoene@fu-berlin.de>
Description: The main method npmpm calculates bacterial
concentrations during food manufacture after contamination.
Variability and uncertainty are included by use of probability
distributions and Monte Carlo Simulation. The model aims at
predicting possible bacterial concentrations at one certain
point in time s, e.g. at the end of a process chain. The process
steps of this process chain are run through in linear order.
Experimental data that match current process step conditions are
gathered, and one deterministic primary model is fitted to every
series of measured values. From every fitted curve one concentration
of bacteria at time s is computed, yielding a set of concentrations.
This sample of possible contamination sizes is assumed to follow a
certain probability distribution. After calculation of distribution
parameters, one value is randomly drawn from this probability
distribution. This value may be modified, and then serves as
contamination for the next process step.
License: GPL version 2 or newer
Depends: R(>=2.11), SweaveListingUtils, graphics
```

```
LazyLoad: yes
```

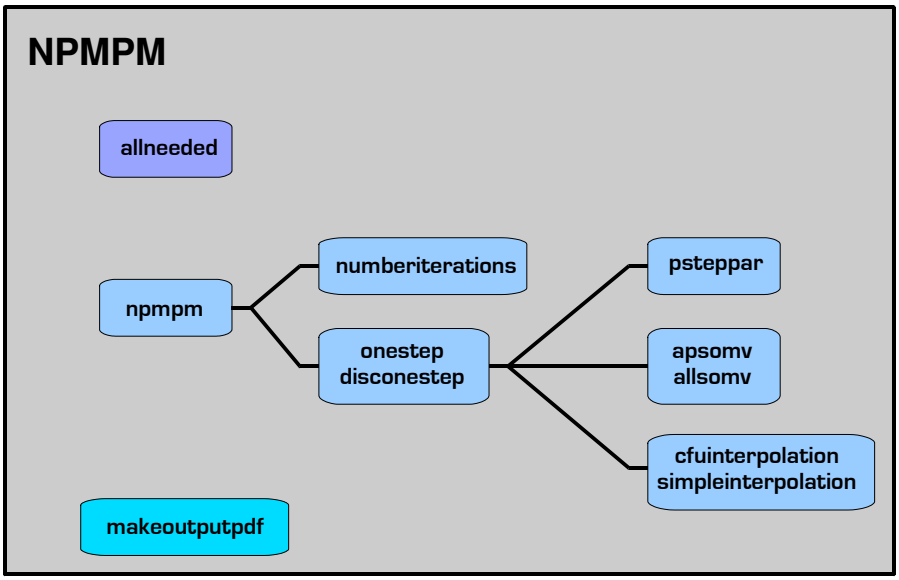


FIGURE D.1. R package NPMPM: methods and calls

Call of allneeded loads the libraries the package depends on, namely graphics and SweaveListingUtils. Package SweaveListingUtils [**Leisch, 2002**] allows for including and executing R code in LaTeX documents, and for generating a .pdf-file from the R session. Package graphics is used for picturing output. allneeded takes no input. This method should be called once in an R session before running the main method. The documentation is given in documentation D.

Documentation.
```
\name{allneded}
\alias{allneded}
\title{
loads all packages this package depends on
}
\description{loads all packages needed for running the package
NPMPM, i.e., packages SweaveListingUtils and graphics}
\usage{
allneded()
}
\value{
no return value
}
\author{
Nadine Schoene
}
\examples{
function(){
```

```
# LOAD PACKAGE FOR GENERATING OF .PDF OUTPUT
library(SweaveListingUtils)
# LOAD PACKAGE FOR PLOTTING
library(graphics)
  } # END FUNCTION
}
```

`npmpm` is the main method of the package NPMPM. It computes the concentration of bacteria in the end product of a process chain. Obligatory input is a vector with exactly one non-negative inoculum size for every process step of the process chain. `npmpm` takes the following facultative input: the relative error bound used for the stop criterion (default value 0.01), interpolation method (default method `cfuinterpolation`, see documentation D), method for selection of experiments (default method `apsomv`, see documentation D), method that returns one process step duration (default method `psteppar`, see documentation D), stop criterion for the simulation (default method `numberiterations`, see documentation D), method that calculates the bacterial concentration at the end of one process step (default method `onestep`, see documentation D). `npmpm` returns a vector with calculated bacterial concentrations (in cfu/ml). Additionally, `npmpm` initialises and sets some vectors in the global environment. These are used by method `makepdfoutput` (see documentation D.

Documentation.
```
\name{npmpm}
\alias{npmpm}
\title{
main method of the package NPMPM
}
\description{
computes the concentration of bacteria in the end product of a
process chain
}
\usage{
npmpm(inoculum, errorb = 0.01, intexmethod = cfuinterpolation,
apsomvmethod = apsomv, pdistpsteps = psteppar,
lastiteration = numberiterations,
calconestep = onestep)
}
\arguments{
  \item{inoculum}{
mandatory input, exactly one non-negative inoculum size per
process step
}
  \item{errorb}{
error bound for stop criterion (default value 0.01)}
  \item{intexmethod}{
interpolation method (default method cfuinterpolation)
}
  \item{apsomvmethod}{
```

```
method for selection of experiments (default method apsomv)
}
  \item{pdistpsteps}{
 method that returns one process step duration (default
 method psteppar)
}
  \item{lastiteration}{
stop criterion for the simulation (default method
numberiterations),
}
  \item{calconestep}{
method that calculates the bacterial concentration at
the end of one
process step (default method onestep)
}
}
\value{
vector with calculated bacterial concentrations (in cfu/ml)
}
\author{
Nadine Schoene
}
```

R package provides two methods for calculation of the bacterial count at the end of a process step. These methods only differ in the assumed probability distribution of cfus. Default method `onestep` assumes the set of cfus to be log-normally distributed. Method `disconestep` uses its frequency distribution. Both need the following input: initial bacterial concentration, the current process step (one row of data frame `psteps`), method for determining the process step duration, method for selection of series of measured values, and the interpolation method. The output is one bacterial concentration at the end of the process step.

Documentation. `\name{onestep}`

```
\alias{onestep}
\alias{disconestep}
\title{
calculates the bacterial count at the end of a process step
}
\description{calculates the bacterial count at the end of
a process step,
default method {\tt onestep} assumes the set of cfus to be
normally distributed,
method {\tt disconestep} uses its frequency distribution
}
\usage{
onestep(cfu_in, pstep, pdistpsteps, apsomvmethod, intexmethod)
disconestep(cfu_in, pstep, pdistpsteps, apsomvmethod, intexmethod)
```

```
}
\arguments{
  \item{cfu_in}{
initial bacterial concentration
}
  \item{pstep}{pstep is one row of
psteps <- data.frame(name, id, process_id, temp_min, temp_max,
time_min, time_max, matrix_id, factor, stepnumber)
}
  \item{pdistpsteps}{
method for determining the process step duration
}
  \item{apsomvmethod}{
method for selection of series of measured values
}
  \item{intexmethod}{
interpolation method
}
}
\value{
one bacterial concentration (in cfu/ml)
}
}
\author{Nadine Schoene
}
```

The package provides two methods for selection of experimental data. Both need the current process step as input, i.e., one row of data frame `psteps` (see table 7.2 in section 7.6). The output is a list of ids of selected series of measured values. Method `allsomv` selects the whole data.frame `raw_data`. Default selection-method apsomv selects experiments matching the parameters (temperature, pH value, and a_w value) of the current process step. Documentation of both methods is given in documentation D.

Documentation.
```
\name{apsomv}
\alias{apsomv}
\alias{allsomv}
\title{function that selects the series of measured values
}
\description{Selects series of measured values from the
data.frame raw_data.
Default selection-method apsomv selects experiments
matching the parameters (temperature, pH value, aw value)
of the current process step.
Method allsomv selects the whole data.frame.
}
\usage{
```

```
apsomv(pstep)
allsomv(pstep)
}
\arguments{
  \item{pstep}{pstep is one row of
psteps <- data.frame(name, id, process_id, temp_min, temp_max,
time_min, time_max, matrix_id, factor, stepnumber)
}
}
\value{
returns a list with the ids of selected series of
measured values
}
\author{Nadine Schoene
}
```

The package contains one method for extraction of process step duration. This method `psteppar` needs the current process step as input, i.e., one row of data frame `psteps` (see table 7.2 in section 7.6). From the uniform distribution of process step duration given as minimum and maximum, one random value is drawn. This process step duration is returned as output.

Documentation. `\name{psteppar}`

```
\alias{psteppar}
\title{
function that returns one process step duration
}
\description{
draws one random value from the uniform distribution
of process
step duration (given by minimum and maximum duration)
}
\usage{
psteppar(pstep)
}
\arguments{
  \item{pstep}{pstep is one row of
psteps <- data.frame(name, id, process_id, temp_min, temp_max,
time_min, time_max, matrix_id, factor, stepnumber)
}
}
\value{
process step duration
}
\author{Nadine Schoene
}
```

The package provides two interpolation methods. Method `simpleinterpolation` is piecewise linear interpolation between measurement points. Default-method `cfuinterpolation` approximates population kinetics of one single experiment by ordinary least squares. Both methods need the id of the series of measured values, process step duration, and the initial bacterial concentration (cfu/ml). The methods return one decadic logarithm of the concentration of bacteria (log10 cfu/ml).

Documentation.
```
\name{cfuinterpolation}
\alias{cfuinterpolation}
\alias{simpleinterpolation}
\title{
interpolation method
}
\description{
interpolation method - fits a function to the measured values
of one experiment and calculates one value (decadic logarithm
of cfu/ml) from this functiondefault method cfuinterpolation
uses ordinary linear regression
method simpleinterpolation interpolates linearly between
measurement points (interpolation by splines of order 2)
}
\usage{
cfuinterpolation(psteptime, series, cfu_in)
simpleinterpolation(psteptime, series, cfu_in)
}
\arguments{
  \item{psteptime}{
duration of the process step
}
  \item{series}{
id of the series of measured values
}
  \item{cfu_in}{
initial bacterial concentration
}
}
\value{
decadic logarithm of bacterial concentration (log10 cfu/ml)
}
\author{
Nadine Schoene
}
```

Method `makepdfoutput` generates a .pdf file with information about simulation settings and results. It needs a text as input, e.g. with names of process chain and microorganism. The method uses the file `pdfoutput.SNW` that is also provided with this package. An example for the result is shown

in figures 8.13 and 8.14 in section 8.8.

Documentation.
```
\name{makepdfoutput}
\alias{makepdfoutput}
\title{
generates a .pdf file with information about simulation
settings and results
}
\description{
generates a .pdf file that is hopefully useful for filing
and explaining simulation results
this file contains the input text, summary statistics and
graphics with the bacterial concentrations calculated in
the simulation, methods (from the package) used during
simulation,and information if quantity of data was sufficient
}
\usage{
makepdfoutput(text)
}
\arguments{
  \item{text}{
text that gives further description of the simulation,
e.g. the name of the microorganism and process chain
}
}
\value{
returns no value to the global environment, but generates
a .pdf file with information about simulation settings
and results
}
\author{
Nadine Schoene
}
\note{
this method depends on the packages SweaveListingUtils
and graphics it generates the .pdf file from the file
pdfoutput.SNW provided with this package
}
```

This method calculates if the stop criterion is fulfilled. It needs the number of iterations, the relative error bound, and a vector with mean of cfus calculated in the first j iterations, as jth component It calculates the difference of the mean of the results of all iterations hitherto done (current mean), and the mean of the results of the iterations done before the last one. If the absolute value of this difference is smaller as the ratio of the current mean determined by the error bound, the stop criterion is fulfilled. The method returns FALSE and the simulation stops. As long as the stop criterion is not

fulfilled yet, the method returns TRUE. stop criterion `numberiterations <- function(iterations,errorb,cfu_average)`, return boolean

Documentation.
```
\name{numberiterations}
\alias{numberiterations}
\title{
stop criterion
}
\description{
computes if the fluctuation of the mean of cfus
calculated in the previous iterations is bigger
than a ratio of the mean determined by the relative
error bound
}
\usage{
numberiterations(iterations, errorb, cfu_average)
}
%- maybe also 'usage' for other objects documented here.
\arguments{
  \item{iterations}{
number of iterations
}
  \item{errorb}{
error bound for stop criterion
}
  \item{cfu_average}{
vector with mean of cfus calculated in the first j
iterations, as jth component
}
}
\value{
TRUE (continue)
FALSE (stop)
}
\author{
Nadine Schoene
}
```

TABLE D.1. Methods provided by the R package `NPMPM`

method	description	input
`allneeded`	loads libraries the package depends on	-
`allsomv`	selects appropriate series of measured values for the current process step	process step
`apsomv`	selects appropriate series of measured values for the current process step (default method)	process step
`cfuinterpolation`	computes one *calculated logcfu/ml* by global linear interpolation (default method)	process step duration, series-id, *start-cfu/ml*
`disconestep`	calculates one *endcfu/ml*	*start-cfu/ml*, process step, method that extracts the process step parameters, method that selects the appropriate series of measured values for the current process step, intexmethod
`makepdfoutput`	generates an output .pdf file	text
`npmpm`	implementation of the new probabilistic model in predictive microbiology (`NPMPM`)	inoculum size
`numberiterations`	terminates the simulation	number of iterations, error bound, vector of means of *final-cfu/ml*s
`onestep`	calculates one *endcfu/ml* (default method)	*start-cfu/ml*, process step , method that extracts the process step parameters, method that selects the appropriate series of measured values for the current process step, intexmethod
`psteppar`	extracts the process step parameters (default method)	process step
`simpleinterpolation`	computes one *calculated logcfu/ml* by local linear interpolation	process step duration, series-id, *start-cfu/ml*

APPENDIX E

Statistics

In this appendix summary statistics of experimental data from the ComBase [**Anonymous, a**], and results of statistical test in validation are listed.

E.1. Summary statistics of *Listeria* data from the ComBase

We get the following summary statistics for the available data.

```
data_source  ComBase:9420
STAT-UP:   3
organism     Listeria monocytogenes/innocua:9423

environment
Culture medium (In: broth)                    :1946
Culture medium (In: TPB)                      :1311
Culture medium (In: BHIB)                     :1113
Pork (In: ham)                                : 449
Culture medium (In: TSYB (0.6% Yeast Extract)): 296
Culture medium (In: TSB)                      : 256
(Other)                                       :4052


b_f                         temperature_C
Culture medium:5474    Min.   :-20.00
Pork          : 615    1st Qu.:  6.00
Beef          : 511    Median : 13.00
Milk          : 483    Mean   : 19.92
Seafood       : 419    3rd Qu.: 28.00
(Other)       :1918    Max.   : 90.00
NA's          :   3


pH                aw
Min.   :0.000         Min.   :0.0000
1st Qu.:5.000           1st Qu.:0.0000
Median :6.000           Median :0.9710
Mean   :5.421           Mean   :0.6742
3rd Qu.:6.700           3rd Qu.:0.9950
Max.   :9.200           Max.   :0.9990
```

Only milk as medium:

```
listeria_milk <- subset(listeria_raw, b_f=="Milk" | b_f=="milk")
```

Summary statistics for *Listeria* spp.

```
data_source     ComBase:483
```

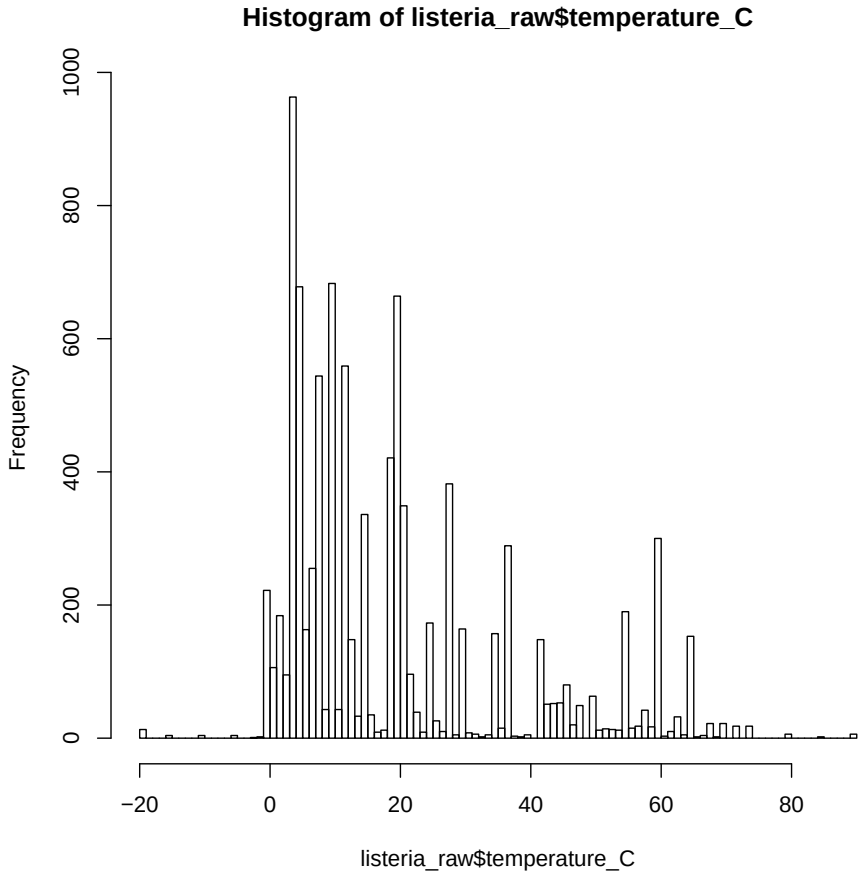


FIGURE E.1. Histogram for the absolute counts of series of measurements for different temperatures for *Listeria* spp.

```
STAT-UP:  0
organism   Listeria monocytogenes/innocua:483


id              source_id
B304_LM:  1    Chhabra_99   :122
B305_LM:  1    Chhabra_02   :109
BJ00_01:  1    Rajkowski_94a: 36
BJ00_02:  1    AFSCE        : 24
BJ00_03:  1    Donnelly_86  : 19
BJ00_04:  1    FSA-CCFRA    : 18
(Other):477    (Other)      :155


environment
Milk (In: formulated milk)               :231
Milk (In: milk)                          : 49
Milk (In: uht milk)                      : 47
```

```
Milk (In: whole milk)                      : 41
Milk (In: skimmed milk)                    : 32
Milk (In: UHT milk with added amino acids): 20
(Other)                                    : 63


b_f             b_f_details
Milk    :483 formulated milk               :231
milk                           : 49
uht milk                       : 47
whole milk                     : 41
skimmed milk                   : 32
UHT milk with added amino acids: 20
(Other)                        : 63


temperature_C   temperature_assumed pH         pH_assumed
Min.   : 0.00   0:483              Min.   :0.000   0:483
1st Qu.:13.00                      1st Qu.:6.000
Median :45.10                      Median :6.600
Mean   :38.53                      Mean   :6.166
3rd Qu.:60.00                      Max.   :7.200
3rd Qu.:6.700
Max.   :65.00


aw            aw_assumed   co_2    in_on
Min.   :0.0000    0:448       Min.   :0    in:483
1st Qu.:0.0000    1: 35       1st Qu.:0    on:  0
Median :0.0000               Median :0
Mean   :0.1042               Mean   :0
3rd Qu.:0.0000               3rd Qu.:0
Max.   :0.9970               Max.   :0


conditions
                                                                  :130
Sterilised before inoculation, Shaken (agitated, stirred)    : 76
Sterilised before inoculation, Shaken (agitated, stirred),
Fat in the environment(%):5                    : 65
Sterilised before inoculation, Shaken (agitated, stirred),
Fat in the environment(%):2.5                  : 64
Inoculation in/on previously heated (cooked, baked,
pasteurized, etc) but not sterilised food/medium : 41
Sterilised before inoculation                                  : 22
(Other)                                                        : 85
```

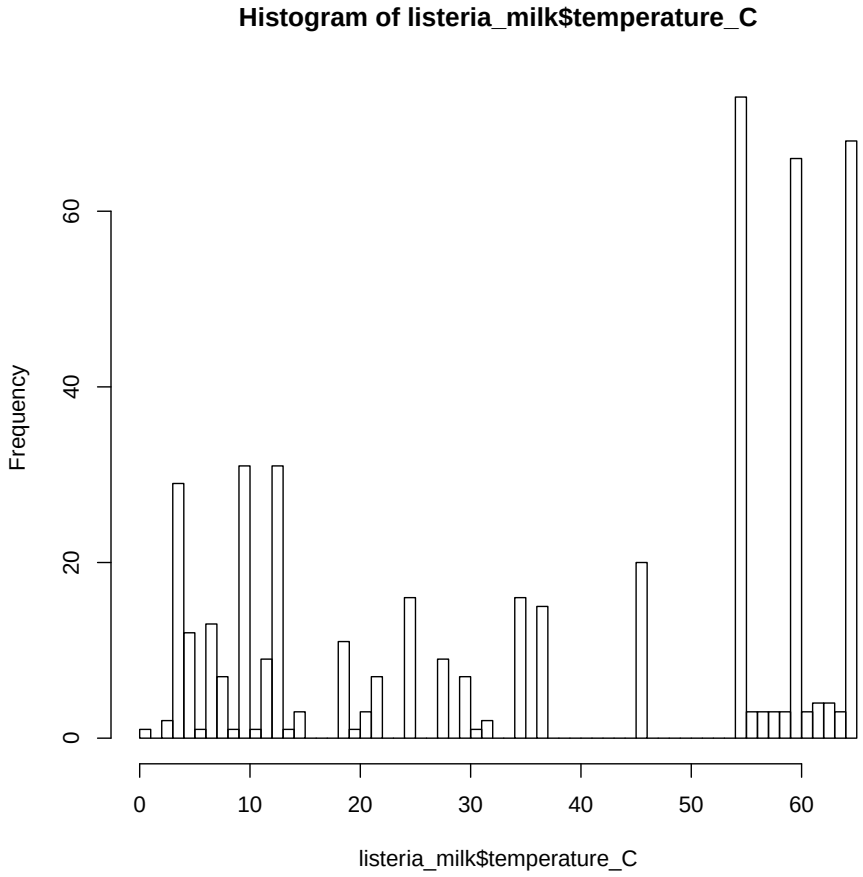


FIGURE E.2. Histogram for the absolute counts of series of measurements in milk for different temperatures for *Listeria* spp.

E.2. Results of the Kolmogorov-Smirnov test I

In this section the results of the Kolmogorov-Smirnov tests done in section 8.3 during validation in with experimental data without removal of measuring times are listed.

TABLE E.1. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `onestep`, 4°*C-sample*

time in s	D	p-value
900	0.2157143	0.15245525
3600	0.2985714	0.01516392
25200	0.2728571	0.03371737
75600	0.1914286	0.26016017
86400	0.3500000	0.03540442
100800	0.2671429	0.03985237
126000	0.1731259	0.36643331
151200	0.2900000	0.02652370

TABLE E.2. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `onestep`, 20°*C-sample*

time in s	D	p-value
21600	0.2059259	0.08594705
86400	0.2196226	0.05799889
259200	0.1941509	0.12392381
432000	0.2000000	0.10231926
604800	0.2549020	0.02236404

TABLE E.3. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `onestep`, 4°*C-sample*

time in s	D	p-value
900	0.1787836	0.32893101
3600	0.1335219	0.68198725
25200	0.3014286	0.01382143
75600	0.2014286	0.21063589
86400	0.1733333	0.68131959
100800	0.2611033	0.04642245
126000	0.2127298	0.16078025
151200	0.2503094	0.07772675

TABLE E.4. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `onestep`, 20° *C-sample*

time in s	D	p-value
21600	0.2022222	0.09586457
86400	0.1203774	0.64142120
259200	0.1742948	0.20915437
432000	0.1500000	0.36396620
604800	0.2189866	0.06746808

TABLE E.5. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `disconestep`, 4° *C-sample*

time in s	D	p-value
900	0.1800000	0.326131050
3600	0.2900000	0.019949741
25200	0.3628571	0.001480385
75600	0.2657143	0.041540228
86400	0.3333333	0.051335712
100800	0.2242857	0.124482661
126000	0.1614286	0.456459097
151200	0.3587500	0.002700828

TABLE E.6. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `disconestep`, 20° *C-sample*

time in s	D	p-value
21600	0.2614379	0.012729096
86400	0.2718868	0.009082834
259200	0.2830189	0.005814734
432000	0.1886689	0.139190140
604800	0.1900602	0.152625337

TABLE E.7. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `disconestep`, $4°C$-*sample*

time in s	D	p-value
900	0.07949081	0.9906594
3600	0.08428571	0.9834955
25200	0.06874116	0.9985373
75600	0.15285714	0.5246406
86400	0.20297030	0.4863083
100800	0.10571429	0.8994932
126000	0.05714286	0.9999493
151200	0.09125000	0.9732949

TABLE E.8. Kolmogorov-Smirnov test for input and output of `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `disconestep`, $20°C$-*sample*

time in s	D	p-value
21600	0.14148148	0.4346953
86400	0.12254810	0.6169338
259200	0.07264151	0.9841216
432000	0.08814815	0.9187702
604800	0.09235294	0.9026703

E.3. Results of Kolmogorov-Smirnov tests done during validation II

In this section the results of the Kolmogorov-Smirnov tests done in section 8.3 during validation with experimental data from which certain measuring times were removed before simulation are listed.

TABLE E.9. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `onestep`, 4°*C-sample*

time in s	D	p-value
900	0.2900000	0.0199497414
3600	0.3500000	0.0024500227
25200	0.4142857	0.0001611965
75600	0.3142857	0.0089916308
86400	0.4405941	0.0032616266
100800	0.2242857	0.1244826607
126000	0.2178788	0.1554206671

TABLE E.10. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `onestep`, 20°*C-sample*

time in s	D	p-value
21600	0.3044444	0.002175458
86400	0.3207547	0.001097450
259200	0.3107547	0.001758317
432000	0.2840225	0.005906383

TABLE E.11. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `onestep`, $4°C$-*sample*

time in s	D	p-value
900	0.4857143	4.236233e-06
3600	0.3914286	4.489807e-04
25200	0.2093352	1.742710e-01
75600	0.2314286	1.044940e-01
86400	0.3960396	1.117395e-02
100800	0.1114286	8.648935e-01
126000	0.1296970	7.322901e-01

TABLE E.12. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `onestep`, $20°C$-*sample*

time in s	D	p-value
21600	0.4270370	2.708892e-06
86400	0.3207547	1.110724e-03
259200	0.2015094	1.014341e-01
432000	0.2545137	1.834236e-02

TABLE E.13. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `disconestep`, $4°C$-*sample*

time in s	D	p-value
900	0.2500000	0.0643957496
3600	0.2857143	0.0228492324
25200	0.3942857	0.0003967854
75600	0.3142857	0.0089916308
86400	0.4207921	0.0057798371
100800	0.1742857	0.3636312117
126000	0.2078788	0.1942329424

TABLE E.14. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `cfuinterpolation` and method for calculation of one step `disconestep`, 20° *C-sample*

time in s	D	p-value
21600	0.3044444	0.002175458
86400	0.3207547	0.001110724
259200	0.3207547	0.001110724
432000	0.2341176	0.040038581

TABLE E.15. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `disconestep`, with 4° *C-sample*

time in s	D	p-value
900	0.4857143	0.0000042
3600	0.4114286	0.0001835
25200	0.2000000	0.2169786
75600	0.2327323	0.0978446
86400	0.3900000	0.0133309
100800	0.1521924	0.5236976
126000	0.1791179	0.3492437

TABLE E.16. Kolmogorov-Smirnov test for input and output of interpolation (measured value that was removed from the data) with `npmpm` with interpolation method `simpleinterpolation` and method for calculation of one step `disconestep`, 20° *C-sample*

time in s	D	p-value
21600	0.4170370	0.0000051
86400	0.3207547	0.0011107
259200	0.2392453	0.0307318
432000	0.1945098	0.1306565

APPENDIX F

Curriculum Vitae

Aus Datenschutzgründen liegt der elektronischen Version dieser Dissertation kein Lebenslauf bei.

Because of data privacy the electronic version of this thesis does not contain a curriculum vitae.

List of Tables

List of Figures

List of Algorithms

Index

Bibliography

[Annino and Russell, 1979] Annino, J. and Russell, E. (1979). The ten most frequent causes of simulation analysis failure - and how to avoid them! *SIMULATION*, 32:137–140.

[Anomymous, 2006] Anomymous (2006). *MySQL 5.0 im Detail*. MySQL AB.

[Anonymous, a] Anonymous.

[Anonymous, b] Anonymous. Gnu general public license. http://www.gnu.org/licenses/.

[Anonymous, 1983] Anonymous (1983). Campylobacteriosis associated with raw milk consumption - pennsylvania.

[Anonymous, 2004] Anonymous (2004). Risk assessment of *Listeria monocytogenes* in ready-to-eat foods. Technical report, WHO / FAO.

[Anonymous, 2008] Anonymous (2008). The community summary report on trends and sources of zoonoses and zoonotic agents and food-borne outbreaks in the european union in 2008. Technical report, European Food Safety Authority (EFSA).

[Anonymous, 2010a] Anonymous (2010a). Milchwirtschaftliche Lehr- und Untersuchungsanstalt Oranienburg (personal communication).

[Anonymous, 2010b] Anonymous, Laboratory staff, F. I. f. R. A. (2010b). personal communication.

[Augustin et al., 2000] Augustin, J., Brouillaud-Delattre, A., Rosso, L., and Carlier, V. (2000). Significance of inoculum size in the lag time of *Listeria monocytogenes*. *Applied and Environmental Microbiology*, 66:1706–1710.

[Banks et al., 2009] Banks, H., Davis, J., Ernstberger, S., Hu, S., Artimovich, E., Dhar, A., and Browdy, C. (2009). A comparison of probabilistic and stochastic formulations in modelling growth uncertainty and variability. *Journal of Biological Dynamics*, 3:130–148.

[Banks, 1994] Banks, R. (1994). *Growth and Diffusion Phenomena*. Springer.

[Baranyi, 2002] Baranyi, J. (2002). Stochastic modelling of bacterial lag phase. *International Journal of Food Microbiology*, 73(2-3):203–206.

[Baranyi, 2011] Baranyi, J. (2011). personal communication.

[Baranyi et al., 1996a] Baranyi, J., Jones, A. Walker, C., Kaloti, A., and Robinson, T.and Mackey, B. (1996a). A combined model for growth and subsequent thermal inactivation of *Brochothrix thermosphacta*. *Applied and Environmental Microbiology*, 62:1029–1035.

[Baranyi et al., 1993a] Baranyi, J., McClure, P., Sutherland, J., and Roberts, T. (1993a). Modeling bacterial growth responses. *Journal of Industrial Microbiology*, 12:190–194.

[Baranyi et al., 1999] Baranyi, J., Pin, C., and Ross, T. (1999). Validating and comparing predictive models. *International Journal of Food Microbiology*, 48(3):159–166.

[Baranyi and Roberts, 1994] Baranyi, J. and Roberts, T. (1994). A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, 23(3-4):277–294.

[Baranyi et al., 1993b] Baranyi, J., Roberts, T., and P. McClure, P. (1993b). A non-autonomous differential equation to model bacterial growth. *Food Microbiology*, 10:43–59.

[Baranyi et al., 1996b] Baranyi, J., Ross, T., McMeekin, T., and Roberts, T. (1996b). Effects of parameterization on the performance of empirical models used in 'predictive microbiology'. *Food Microbiology*, 13:83–91.

[Barbosa-Cánovas et al., 2005] Barbosa-Cánovas, G., Tapia, M., and Pilar Cano, M., editors (2005). *Novel Food Processing Technologies*. CRC Press.

[Bartel et al., 1996] Bartel, B., Malczan, M., and Riemelt, I. (1996). *Milchwirtschaftliche Mikrobiologie : Nachdruck*. Behr's.

[Basheer and Hajmeer, 2000] Basheer, I. and Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1):3–31.

[Battista et al., 1999] Battista, J., Earl, A., and Park, M. (1999). Why is *Deinococcus radiodurans* so resistant to ionizing radiation? *Trends in Microbiology*, 7:362–365.

[Baumgartner, 1938] Baumgartner, J. (1938). Heat sterilized reducing sugars and their effect on the thermal resistance of bacteria. *Journal of Baceriology*, 36:369–382.

[Belitz et al., 2009] Belitz, H., Grosch, W., and Schieberle, P., editors (2009). *Food Chemistry*, chapter Milk and Dairy Products, pages 498–545. Springer.

[Beuchat and Ryu, 1997] Beuchat, L. and Ryu, J. (1997). Produce handling and processing practices. *Emerging Infectious Diseases*, 3:459–465.

[Bigelow, 1918] Bigelow, W. (1918). Scientific research in the canning industry. *Journal of the Franklin Institute*, 186:1–14.

[Bigelow, 1921] Bigelow, W. (1921). The logarithmic nature of thermal death time curves. *The Journal of Infectious Diseases*, 29:528–236.

[Bigelow and Esty, 1920] Bigelow, W. and Esty, J. (1920). The thermal death point in relation to time of typical thermophilic organisms. *Journal of Infectious Diseases*, 27:602–617.

[Boone et al., 2005] Boone, I., van der Stede, Y., Bollaerts, K., Vose, D., Daube, G., Aerts, M., and Mintiens, K. (2005). Evaluating the quality of data in quantitative microbial risk assessments: pedigree methods applied to a farm-to-fork risk assessment for salmonella in pork.

[Bosch, 1998] Bosch, K. (1998). *Statistik Taschenbuch*. Oldenbourg.

[Box and Draper, 1987] Box, G. and Draper, N. (1987). *Empirical Model-Building and Response Surfaces*. Wiley.

[Brannen, 1968] Brannen, J. (1968). On logarithmic extrapolation of microbial survivor curves for planetary quarantine requirementsthermal death point in relation to time of typical thermophilic organismsthermal death point in relation to time of typical thermophilic organisms. *Space Life Sciences*, 1:150–152.

[Buchanan, 1907] Buchanan, R. M. (1907). The carriage of infection by flies. *The Lancet*, 170:216–218.

[Burgess et al., 2010] Burgess, S., Lindsay, D., and Flint, S. (2010). Thermophilic bacilli and their importance in dairy processing. *International Journal of Food Microbiology*, 144:215–225.

[Bylund, 2003] Bylund, G. (2003). *Tetra Pak Handbuch der Milch- und Molkereitechnik*. Verlag Th. Mann GmbH & Co. KG.

[Cao et al., 2010] Cao, R., Francisco-Fernández, M., and Quinto, E. (2010). A random effect multiplicative heteroscedastic model for bacterial growth. *BMC Bioinformatics*, 11:–.

[Caponechia, 2009] Caponechia, C. (2009). Strategies to improve the communication of probability information in risk analyses. *International Journal of Risk Assessment and Management*, 12:380–395.

[Cavicchioli et al., 2003] Cavicchioli, R., Curmi, P., Saunders, N., and Thomas, T. (2003). Pathogneic archaea: do they exist? *BioEssays*, 25:1119–1128.

[Chen et al., 2007] Chen, G., Campanella, O., and Corvalan, C. (2007). A numerical algorithm for calculating microbial survival curves during thermal processing. *Food Research International*, 40:203–208.

[Chirife, 1978] Chirife, G. (1978). Prediction of water activity in intermediate moisture foods. *Journal of Food Technology*, 13:417–424.

[Cogan et al., 2007] Cogan, T., Beresford, T., Steele, J., Broadbent, J., Shah, N., and Ustunol, Z. (2007). Invited review: Advances in starter cultures and cultured foods. *Journal of Dairy Science*, 90(9):4005–4021.

[Conway de Macario and Macarioa, 2009] Conway de Macario, E. and Macarioa, A. (2009). Methanogenic archaea in health and disease: A novel paradigm of microbial pathogenesis. *International Journal of Medical Microbiology*, 299:99–108.

[Coroller et al., 2006] Coroller, L., Leguerinel, I., Mettler, E., Savy, N., and Mafart, P. (2006). General model, based on two mixed Weibull distributions of bacterial resistance,

for describing various shapes of inactivation curves. *Applied and Environmental Microbiology*, 72:6493–6502.

[Corradini et al., 2010a] Corradini, M., Normand, M., Eisenberg, M., and Peleg, M. (2010a). Evaluation of a stochastic inactivation model for heat-activated spores of *Bacillus* spp. *Applied and Environmental Microbiology*, 76:4402–4412.

[Corradini et al., 2010b] Corradini, M., Normand, M., and Peleg, M. (2010b). Stochastic and deterministic model of microbial heat inactivation. *Journal of Food Science*, 75:R59–R70.

[Dairy SachsenMilch, 2009] Dairy SachsenMilch, L. G. (2009). personal communication.

[Davis et al., 2005] Davis, K., Joseph, S., and Janssen, P. (2005). Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Applied and Environmental Microbiology*, 71:826–834.

[Dean and Sir Hinshelwood, 1966] Dean, A. and Sir Hinshelwood, C. (1966). *Growth, function and regulation in bacterial cells*. Oxford University Press.

[Delignette-Muller, 1998] Delignette-Muller (1998). Relation between the generation time and the lag time of bacterial growth kinetics. *International Journal of Food Microbiology*, 43:97–104.

[Delignette-Muller and Rosso, 2000] Delignette-Muller, M. and Rosso, L. (2000). Biological variability and exposure assessment. *International Journal of Food Microbiology*, 58:203–212.

[Delignette-Muller et al., 1995] Delignette-Muller, M., Rosso, L., and Flandrois, J. (1995). Accuracy of microbial growth predictions with square root and polynomial models. *International Journal of Food Microbiology*, 27:139–146.

[Deuflhard and Bornemann, 2002] Deuflhard, P. and Bornemann, F. (2002). *Numerische Mathematik II*. de Gruyter.

[Ernstberger, 2008] Ernstberger, S. (2008). *Sensitivity methods for dynamical systems*. PhD thesis, North Carolina State University.

[et al., 2008] et al., H. (2008). Ifcn dairy report 2008. Technical report, International Farm Comparison Network, Dairy Research Center.

[Falenski et al., 2010] Falenski, A., Mayer-Scholl, A., Filter M, Göllner, C., Appel, B., and Nöckler, K. (2010). Survival of *Brucella* spp. in mineral water, milk and yogurt. *International Journal of Food Microbiology*, pages 326–330.

[Farkas-Himsley, 1980] Farkas-Himsley, H. (1980). Bacteriocins-are they broad-spectrum antibiotics? *Journal of Antimicrobial Chemotherapy*, 6:424–426.

[Feller, 1939] Feller, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheoretica*, 5:11–40,.

[Fernández-Navarro et al., 2010] Fernández-Navarro, F., Valero, A., Hervás-Martínez, C., Gutiérrez, P., García-Gimeno, R., and Zurera-Cosano, G. (2010). Development of a multi-classification neural network model to determine the microbial growth/no growth interface. *International Journal of Food Microbiology*, 141:203–212.

[Foissy, 2005] Foissy, H. (2005). *Milchtechnologie - Eine vorlesungsorientierte Darstellung*. IMB-Verlag.

[Fox and McSweeney, 1998] Fox, P. and McSweeney, P. (1998). *Dairy Chemistry and Biochemistry*. blackie Academic & Professional, first edition.

[Fretz et al., 2010a] Fretz, R., Pichler, J., Sagel, U., Much, P., Ruppitsch, W., Pietzka, A., Stöger, A., Huhulescu, S., Heuberger, S., Appl, G., Werber, D., , Stark, K., Prager, R., Flieger, A., Karpíšková, R., Pfaff, G., and Allerberger, F. (2010a). Update: Multinational listeriosis outbreak due to 'quargel', a sour milk curd cheese, caused by two different l. monocytogenes serotype 1/2a strains, 2009-2010. *Eurosurveillance*, 15:1–2.

[Fretz et al., 2010b] Fretz, R., Sagel, U., Ruppitsch, W., Pietzka, A., Stöger, A., Huhulescu, S., Heuberger, S., Pichler, J., Much, P., Pfaff, G., Stark, K., Prager, R., Flieger, A., Feenstra, O., and Allerberger, F. (2010b). *Listeriosis* outbreak caused by acid curd cheese 'quargel', austria and germany 2009. *Eurosurveillance*, 15:–.

[Fuchs, 2007] Fuchs, G. (2007). *Allgemeine Mikrobiologie (Schlegel)*. Thieme.

[Geeraerd et al., 2000] Geeraerd, A., Herremans, C., and Van Impe, J. (2000). Structural model requirements to describe microbial inactivation during a mild heat treatment. *International Journal of Food Microbiology*, 59:185–209.

[Gehring et al., 2010] Gehring, R., Schumm, P., Youssef, M., and Scoglio, C. (2010). A network-based approach for resistance transmission in bacterial populations. *Journal of Theoretical Biology*, 262:97–106.

[Gentle, 2003] Gentle, J. (2003). *Random Number Generation and Monte Carlo Methods*. Springer.

[Gerner-Smidt and Whichard, 2007] Gerner-Smidt, P. and Whichard, J. (2007). Foodborne disease trends and reports. *Foodborne Pathogens and Disease*, 4:1–4.

[Ghanem and Red-Horse, 1999] Ghanem, R. and Red-Horse, J. (1999). Propagation of probabilistic uncertaintiy in complex physical systems using a stochastic finite element approach. *Physica D*, 133:137–144.

[Ghosh and Setlow, 2009] Ghosh, S. and Setlow, P. (2009). Isolation and characterization of superdormant spores of bacillus species. *Journal of Bacteriology*, 191:1787–1797.

[Gibson et al., 1988] Gibson, A., Bratchell, N., and Roberts, T. (1988). Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by ph, sodium chloride and storage temperature. *International Journal of Food Microbiology*, 6:155–178.

[Gil et al., 2006] Gil, M., Brandão, T., and Silva, C. (2006). A modified Gompertz model to predict microbial inactivation under time-varying temperature conditions. *Journal of Food Engineering*, 76:89–94.

[Gompertz, 1825] Gompertz, B. (1825). 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. *Philosophical Transactions of Royal Society of London*, pages 513–585.

[Grant et al., 2005] Grant, I., Williams, A., Rowe, M., and Muir, D. (2005). Efficacy of various pasteurization time-temperature conditions in combination with homogenization on inactivation of *Mycobacterium avium* subsp. *paratuberculosis* in milk. *Applied and environmental biology*, 71:2853–2861.

[Griffith, 1928] Griffith, F. (1928). The significance of pneumococcal types. *Journal of Hygiene*, 27:113–159.

[Griffiths, 1994] Griffiths, M. (1994). Predictive modelling: applications in the dairy industry. *International Journal of Food Microbiology*, 23(3-4):305–315.

[Guillier et al., 2005] Guillier, L., Pardon, P., and Augustin, J. (2005). Influence of stress on individual lag time distributions of *Listeria monocytogenes*. *Applied and Environmental Microbiology*, 71:2940–2948.

[Hao, 2007] Hao, L. (2007). Evolution and limitations of primary mathematical models in predictive microbiology. *British Food Journal*, 109:608–626f.

[Hassani et al., 2006] Hassani, M., Mañas, P., Condón, S., and Pagán, R. (2006). Predicting heat inactivation of *Staphylococcus aureus* under nonisothermal treatments at different ph. *Molecular Nutrition and Food Research*, 50:572–580.

[Havelaar et al., 2008] Havelaar, A., Brul, S., de Jong, A., de Jonge, R., Zwietering, M., and ter Kuile, B. (2008). Future challenges to microbial food safety. *International Journal of Food Microbiology*, 139:S79–S94.

[Hennessy et al., 1996] Hennessy, T., Hedberg, C., Slutsker, L., White, K., Besser-Wiek, J., Moen, M., Feldman, J., Coleman, W., Edmonson, L., MacDonald, K., and Osterholm, M. (1996). A national outbreak of *Salmonella enteritidis* infections from ice cream. the investigation team. *New England Journal of Medicine*, 334(20):1281–1286.

[Hesse, 1960] Hesse, W. (1960). Über das Verhalten pathogener Mikroorganismen in pasteurisierter Milch. *Medical Microbiology and Immunology*, 34:346–348.

[Hills and Mackey, 1995] Hills, B. and Mackey, B. (1995). Multi-compartment kinetic models for injury, resuscitation, induced lag and growth in bacterial cell populations. *Food Microbiology*, 12:333–346.

[Hills and Wright, 1994] Hills, B. and Wright, K. (1994). A new model for bacterial growth in heterogeneous systems. *Journal of Theoretical Biology*, 168(1):31–41.

[Horowitz et al., 2010] Horowitz, J., Normand, M., Corradini, M., and Peleg, M. (2010). Probabilistic model of microbial cell growth, division, and mortality. *Applied and Environmental Microbiology*, 76:230–242.

[Hui, 2006] Hui, Y. (2006). *Handbook of Food Science, Technology, and Engineering, Volume 3: Food Engineering and Food Processing*. Taylor & Francis.

[Juneja et al., 2006] Juneja, V., Huang, L., and Marks, H. (2006). Approaches for modeling thermal inactivation of foodborne pathogens. *American Chemical Society symposium series*, 931:235–251.

[Kahm et al., 2010] Kahm, M., Hasenbrink, G., Lichtenberg-Fraté, H., Ludwig, J., and Kschischo, M. (2010). grofit: Fitting biological growth curves with R. *Journal of Statistical Software*, 33:1–21.

[Kapoor and Metzger, 2008] Kapoor, R. and Metzger, L. E. (2008). Process cheese: Scientific and technological aspects - a review. *Comprehensive Reviews in Food Science and Food Safety*, 7:194–214.

[Kelly and Rahn, 1932] Kelly, C. and Rahn, O. (1932). The growth rate of individual bacterial cells. *Journal of Bacteriology*, 23:147–153.

[Kemmeren et al., 2006] Kemmeren, J., Mangen, M., van Duynhoven, Y., and Havelaar, A. (2006). Priority settings of foodborne pathogens. Technical Report 330080001/2006, RIVM.

[Kendall, 1948a] Kendall, D. (1948a). On the generalized "birth-and-death" process. *Annals of Mathematical Statistics*, 1:1–15.

[Kendall, 1948b] Kendall, D. (1948b). On the role of variable generation time in the development of a stochastic birth process. *Biometrika*, 35:316–330.

[Kennedy, 2008] Kennedy, S. (2008). Epidemiology. why can't we test our way to absolute food safety? *Science*, 322(5908):1641–1643.

[Kilsby et al., 2000] Kilsby, D., Davies, K., McClure, P., Adair, C., and Anderson, W. (2000). Bacterial thermal death kinetics based on probability distributions: the heat destruction of *Clostridium botulinum* and *Salmonella* bedford. *Journal of Food Protection*, 63:1197–1203.

[Klenke, 2008] Klenke, A. (2008). *Wahrscheinlichkeitstheorie*. Springer.

[Knolle, 1988] Knolle, H. (1988). Cell kinetic modelling and the chemotherapy of cancer. In *Lecture Notes in Biomathematics*. Springer.

[Koch, 1966] Koch, A. (1966). The logarithm in biology. i. mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology*, 12(2):276–290.

[Koch, 1969] Koch, A. (1969). The logarithm in biology. ii. distributions simulating the log-normal. *Journal of Theoretical Biology*, 23(2):251–268.

[Koch and Schaechter, 1962] Koch, A. and Schaechter, M. (1962). A model for statistics of the cell division process. *Journal of General Microbiology*, 29:435–454.

[Koseki, 2009] Koseki, S. (2009). Microbial responses viewer (mrv): a new combase-derived database of microbial responses to food environments. *International Journal of Food Microbiology*, 134(1-2):75–82.

[Kot, 2001] Kot, M. (2001). *Elements of mathematical ecology*. University Press, Cambridge.

[Kothary and Babu, 2001] Kothary, M. and Babu, U. (2001). Infective dose of foodborne pathogens in volunteers: a review. *Journal of Food Safety*, 21:49–73.

[Koutsoumanis, 2008] Koutsoumanis, K. (2008). A study on the variability in the growth limits of individual cells and its effect on the behavior of microbial populations. *International Journal of Food Microbiology*, 128:116–121.

[Law, 2007] Law, A. (2007). *Simulation Modeling & Analysis*. McGraw-Hill Internationa.

[Leisch, 2002] Leisch, F. (2002). Sweave, part I: Mixing R and LaTeX. *R News*, 2(3):28–31.

[Leistner and Gorris, 1995] Leistner, L. and Gorris, L. (1995). Food preservation by hurdle technology. *Trends in Food Science & Technology*, 6:41–46.

[Lewis, 1999] Lewis, M. (1999). Microbiological issues associated with heat treated milks. *International Journal of Dairy Technology*, 52:121–125.

[Lewis and Heppell, 2000] Lewis, M. and Heppell, N. (2000). *Continuous Thermal Processing of Foods - Pasteurization and UHT Sterilization*. AN Aspen Publication.

[Li et al., 2008] Li, H., Xie, G., and Edmondson, A. (2008). Review of secondary mathematical models of predictive microbiology. *Journal of Foods Products Marketing*, 14:57–74.

[Limpert et al., 2001] Limpert, E., Stahel, W., and Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. *BioScience*, 51:341–351.

[Margosch et al., 2004] Margosch, D., Gänzle, M., Ehrmann, M., and Vogel, R. (2004). Pressure inactivation of *Bacillus* endospores. *Applied and Environmental Microbiology*, 70, No.12:7321–7328.

[Mariani et al., 2007] Mariani, C., Briandet, R., Chamba, J., Notz, E., Carnet-Pantiez, A., Eyoug, R., and Oulahal, N. (2007). Biofilm ecology of wooden shelves used in ripening the French raw milk smear cheese Reblochon de Savoie. *Journal of Dairy Science*, 90(4):1653–1661.

[Marks et al., 2001] Marks, N., Grandison, A., and Lewis, M. (2001). Challenge testing of the lactoperoxidase system in pasteurized milk. *Journal of Applied Microbiology*, 91(4):735–741.

[Marth and Steele, 2001] Marth, E. and Steele, J., editors (2001). *Applied Dairy Microbiology*. Marcel Dekker, Inc.

[Mastronicolis et al., 1998] Mastronicolis, S., German, J., Megoulas, N.and Petrou, E., Foka, P., and Smith, G. (1998). Influence of cold shock on the fatty-acid composition of different lipid classes of the food-borne pathogen listeria monocytogenes. *Food Microbiology*, 15:299–306.

[McCullagh and Nelder, 1999] McCullagh, P. and Nelder, J. (1999). *Generalized Linear Models*. Chapman and Hall. Weibull - survival distribution times.

[McKay et al., 1979] McKay, M., Beckman, R., and Conover, W. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245.

[McKellar and Lu, 2004] McKellar, R. and Lu, X. (2004). *Modeling microbial responses in food*. CRC Press.

[McMeekin et al., 2002] McMeekin, T., Olley, J., Ratkowsky, D., and Ross, T. (2002). Predictive microbiology: towards the interface and beyond. *International Journal of Food Microbiology*, 73:395–407.

[McMeekin et al., 1993] McMeekin, T., Olley, J., Ross, T., and Ratkowsky, D. (1993). *Predictive Microbiology: Theory and Application*. John Wiley & Sons, New York.

[McMeekin and Ross, 2002] McMeekin, T. and Ross, T. (2002). Predictive microbiology: providing a knowledge-based framework for change management. *International Journal of Food Microbiology*, 78(1-2):133–153.

[Membre and Lambert, 2008] Membre, J. and Lambert, R. (2008). Application of predictive modelling techniques in industry: from food design up to risk assessment. *International Journal of Food Microbiology*, 128(1):10–15.

[Mermin and Griffin, 1999] Mermin, J. and Griffin, P. (1999). Public health in crisis: outbreaks of *Escherichia coli* O157:H7 infections in Japan. *American Journal of Epidemiology*, 150(8):797–803; discussion 804–5.

[Metropolis and Ulam, 1949] Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Society*, 44:335–341.

[Metzler, 1987] Metzler, W. (1987). *Dynamische Systeme in der Ökologie*. Teubner.

[Moats et al., 1971] Moats, W., Dabbah, R., and Edwards, V. (1971). Interpretation of nonlogarithmic survivor curves of heated bacteria. *Journal of Food Science*, 36:523–526.

[Muñoz Cuevas et al., 2010] Muñoz Cuevas, M., Fernández, P., Geroge, S., and Pin, C. (2010). Meling the lag period and exponential growth of *Listeria monocytogenes* under conditions of fluctuating temperature and water activity values. *Applied and Environmental Microbiology*, 76:2908–2915.

[Organization, 2008] Organization, W. H. (2008). Terrorist threats to food.

[Panagou, 2008] Panagou, E. (2008). A radial basis function neural network approach to determine the survival of *Listeria monocytogenes* in katiki, a traditional greek soft cheese. *Journal of Food Protection*, 71:750–759.

[Pearl and Reed, 1920] Pearl, R. and Reed, L. (1920). On the rate of growth of the population of the United States since 1770 and its mathematical representation. *Proceedings of the National Academy of Sciences*, 6:275–288.

[Peleg, 2003] Peleg, M. (2003). Microbial survival curves: Interpretation, mathematical modeling, and utilization. *Comments on Theoretical Biology*, 8:357–387.

[Peleg, 2006] Peleg, M., editor (2006). *Advanced quantitative microbiology for foods and biosystems : models for predicting growth and inactivation*. Taylor & Francis.

[Peleg et al., 2008] Peleg, M., Normand, M., and Corradini, M. (2008). Interactive software for estimating the efficacy of non-isothermal heat preservatin processes. *International Journal of Food Microbiology*, 126:250–257.

[Peleg, 1998] Peleg, M.and Cole, M. (1998). Reinterpretation of microbial survival curves. *Critical Reviews in Food Science and Nutrition*, 38(5):353–380.

[Págan et al., 1997] Págan, R., Condon, S., and Sala, F. (1997). Effects of several factors on the heat-shock-induced thermotolerance of listeria monocytogenes. *Applied and Environmental Microbiology*, 63:3225–3232.

[Pin and Baranyi, 2005] Pin, C. and Baranyi, J. (2005). Kinetics of single cells: observation and modeling of a stochastic process. *Applied and Environmental Microbiology*, 72:2163–2169.

[Powell and Errington, 1963] Powell, E. and Errington, F. (1963). Generation times of individual bacteria: some corroborative measurements. *Journal of General Microbiology*, 31:315–327. cell growth and division, distribution of generation times.

[Pratchett, 2010] Pratchett, T. (2010). *I shall wear midnight*. HarperCollins.

[Prescott, 1959] Prescott, D. (1959). Variations in the individual generation times of *Tetrahymena geleii* hs. *Experimental Cell Research*, 16:279–284.

[Price and Sowers, 2004] Price, P. and Sowers, T. (2004). Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proceedings of the National Academy of Sciences*, 101:4631–4636.

[Pruitt and Kamau, 1993] Pruitt, K. and Kamau, D. (1993). Mathematical models of bacterial growth, inhibition and death under combined stress conditions. *Journal of Industrial Microbiology & Biotechnology*, 12:221–231.

[R Development Core Team, 2010] R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Ratkowsky, 1993] Ratkowsky, D. (1993). Principles of nonlinear regression modeling. *Journal of Industrial Microbiology*, 12:195–1999.

[Ratkowsky and Ross, 1995] Ratkowsky, D. and Ross, T. (1995). Modelling the bacterial growth/no growth interface. *Letters in Applied Microbiology*, 20:29–33.

[Ray and Speck, 1973] Ray and Speck (1973). *Critical Reviews in Clinical laboratory Science*, page 161. CRC Press.

[Robertson, 1932] Robertson, A. (1932). Averaging bacterial counts. *Journal of Bacteriology*, 23:123–134.

[Robinson, 1982] Robinson (1982). Yogurt manufacture - some considerations of quality. *Dairy Industries International*, 46:31.

[Robinson et al., 1998] Robinson, T., Ocio, M., Kaloti, A., and Mackey, B. (1998). The effect of the growth environment on the lag phase of *Listeria monocytogenes*. *International Journal of Food Microbiology*, 44:83–92.

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

[Ross and McMeekin, 1994] Ross, T. and McMeekin, T. (1994). Predictive microbiology. *International Journal of Food Microbiology*, 23:241–264.

[Rowan and Anderson, 1994] Rowan, N. and Anderson, J. (1994). Effects of above-optimum growth temperature and cell morphology on thermotolerance of *Listeria monocytogenes* cells suspended in bovine milk. *Applied and Environmental Microbiology*, 64:2065 – 2071.

[Rowe and Grant, 2006] Rowe, M. and Grant, I. (2006). *Mycobacterium avium* ssp. *paratuberculosis* and its potential survival tactics. *Letters in Applied Microbiology*, 42:305–311.

[Schaechter, 2006] Schaechter, M. (2006). From growth physiology to systems biology. *International Microbiology*, 9:157–161.

[Schaechter et al., 1962] Schaechter, M., Williamson, J., Hood, J., and Koch, A. (1962). Growth, cell and nuclear divisions in some bacteria. *Journal of General Microbiology*, 29:421–434.

[Sergelidis and Abrahim, 2009] Sergelidis, D. and Abrahim, A. (2009). Adaptive response of *Listeria monocytogenes* to heat and its impact on food safety. *Food Control*, 20:1–10.

[Shleeca et al., 2010] Shleeca, M., Salina, E., and Kaprelyants, A. (2010). Dormant forms of mycobacteria. *Micobiology*, 79:1–12.

[Sienkiewicz and Kirst, 2006] Sienkiewicz, T. and Kirst, E. (2006). *Analytik von Milch und Milcherzeugnissen*, volume 1. Behr´s Verlag GmbH & Co. KG.

[Slater and Schaechter, 1974] Slater, M. and Schaechter, M. (1974). Control of cell division in bacteria. *Bacteriological Reviews*, 38:199–221.

[Smit, 2003] Smit, G., editor (2003). *Dairy processing*. CRC Press.

[Smittle, 2000] Smittle, R. (2000). Microbial safety of mayonnaise, salad dressings, and sauces produced in the united states: a review. *Journal of Food Protection*, 63:1144–1153.

[Spreer, 2005] Spreer, E. (2005). *Technologie der Milchverarbeitung*. B. Behr´s Verlag´GmbH & Co.

[Standaert et al., 2007] Standaert, A., Francois, K., Devlieghere, F., Debevere, J., Van Impe, J., and Geeraerd, A. (2007). Modeling individual cell lag time distributions for *Listeria monocytogenes*. *Risk Analysis*, 27(1):241–254.

[Stannard et al., 1985] Stannard, C., Williams, A., and Gibbs, P. (1985). Temperature/growth relationship for psychotrophic food-spoilage bacteria. *Food Microbiology*, 2:115–122.

[Stern and Frazier, 1941] Stern, R. and Frazier, W. (1941). Physiological characteristics of lactic acid bacteria near the maximum growth temperature: I. growth and acid production. *Journal of Bacteriology*, 42:479–499.

[Swagerty et al., 2002] Swagerty, D., Walling, A., and Klein, R. (2002). Lactose intolerance. *American Family Physicians*, 65:1845–1850.

[Takahashi, 1966] Takahashi, M. (1966). Theoretical basis for cell cycle analysis. *Journal of Theoretical Biology*, 13:202–211.

[Takahashi, 1968] Takahashi, M. (1968). Theroetical basis for cell cycle analysis. *Journal of Theoretical Biology*, 18:195–209.

[Taylor and Karlin, 1998] Taylor, H. and Karlin, S. (1998). *An Introduction to Stochastic Modeling*. Academic Press, 3rd edition.

[Todd et al., 2009] Todd, E., Greig, J., Bartleson, C., and Michaels, B. (2009). Outbreaks where food workers have been implicated in the spread of foodborne disease. Part 6. Transmissio and survival of pathogens in the food processing and preparation environment. *Journal of Food Protection*, 72:202–219.

[Tomé et al., 2005] Tomé, E., Todorov, S., Gibbs, P., and Teixeira, P. (2005). Partial charakterization of nine bacteriocins produced by lactic acid bacteria isolated from cold-smoked salmon with activity against *Listeria monocytogenes*. *Food Biotechnology*, 19(2):50–73.

[Töpel, 2004] Töpel, A. (2004). *Chemie und Physik der Milch*. Behr´s.

[Vadasz and Vadasz, 2007] Vadasz, P. and Vadasz, E. (2007). Biological implications from an autonomous version of baranyi and roberts growth model. *International Journal of Food Microbiology*, 114(3):357–365.

[Vaidya and Corvalan, 2009] Vaidya, N. and Corvalan, C. (2009). An integral model of microbial inactivation taking into account memory effects: power law memory kernel. *Journal of Food Protection*, 72:837–842.

[van Asselt and Zwietering, 2006] van Asselt, E. and Zwietering, M. (2006). A systematic approach to determine global thermal inactivation parameters for various food pathogens. *International Journal of Food Microbiology*, 107(1):73–82.

[Van Boekel, 2002] Van Boekel, M. (2002). On the use of the Weibull model to describe thermal inactivation of microbial vegetative cells. *International Journal of Food Microbiology*, 74(1-2):139–159.

[van Duynhoven et al., 2009] van Duynhoven, Y., Isken, L., Borgen, K., Besselse, M., Soethoudt, K., Haitsma, O., Mulder, B., Notermans, D., de Jonge, R., Kock, P., van Pelt, W., Stenvers, O., and van Steenbergen, J. (2009). A prolonged outbreak of *Salmonella* typhimurium infection related to an ucommon vehicle: hard cheese made from raw milk. *Epidemiology and Infection*, 137:1548–1557.

[van Gerwen and Zwietering, 1998] van Gerwen, S. and Zwietering, M. (1998). Growth and inactivation models to be used in quantitative risk assessment. *Journal of Food Protection*, 61:1541–1549.

[Verhulst, 1838] Verhulst, P. (1838). Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique*, 10:113–121.

[Vermeulen et al., 2007] Vermeulen, A., Gysemans, K., Bernaerts, K., Geeraerd, A., JVan Impe, J., Debevere, J., and Devlieghere, F. (2007). Influence of ph, water activity and acetic acid concentration on listeria monocytogenes at 7 °c: Data collection for the development of a growth/no growth model. *International Journal of Food Microbiology*, 114:332–341.

[Vose, 2008] Vose, D. (2008). *Risk Analysis: A Quantitative Guide*. Wiley & Sons, 3. edition.

[Weber, 2006] Weber, H., editor (2006). *Milch und Milchprodukte*. Mikrobiologie der Lebensmittel. Behr´s Verlag.

[Wein and Liu, 2005] Wein, L. M. and Liu, Y. (2005). Analyzing a bioterror attack on the food supply: The case of botulinum toxin in milk. *PNAS*, 102(28):9984–9989.

[Wengenroth, 2008] Wengenroth, J. (2008). *Wahrscheinlichkeitstheorie*. de Gruyter.

[Whiting and R., 1993] Whiting, R. and R., B. (1993). A classification of models for predictive microbiology. *Food Microbiology*, 10:175–177.

[Withell and Pharm, 1942] Withell, E. and Pharm, B. (1942). The significance of the variation in shape of time-survivor curves. *Journal of Hygiene*, 42:124.

[Wolf and Venus, 1992] Wolf, K. and Venus, J. (1992). Description of the delayed microbial growth by an extended logistic equation. *Acta Biotechnologica*, 12:405–410.

[Yule and Kendall, 1950] Yule, G. and Kendall, M. (1950). *An introduction to the theory of statistics*. Griffin.

[Zdolec et al., 2009] Zdolec, N., Kozačinski, L., Njari, B., Filipović, I., Hadžiosmanović, M., Mioković, B., Kuzmanović, Z., Mitak, M., and Samac, D. (2009). The antimicrobial effect of lactobacilli on some foodborne bacteria. *Archiv für Lebensmittelhygiene*, 60:115–119.

[Zwietering et al., 1996] Zwietering, M., deWit, J., and Notermans, S. (1996). Application of predictive microbiology to estimate the number of *Bacillus cereus* in pasteurised milk at the point of consumption. *International Journal of Food Microbiology*, 30:55–70.

[Zwietering et al., 1990] Zwietering, M., Jongenburger, I., Rombouts, F., and van ´t Riet, K. (1990). Modelling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56:1875–1881.