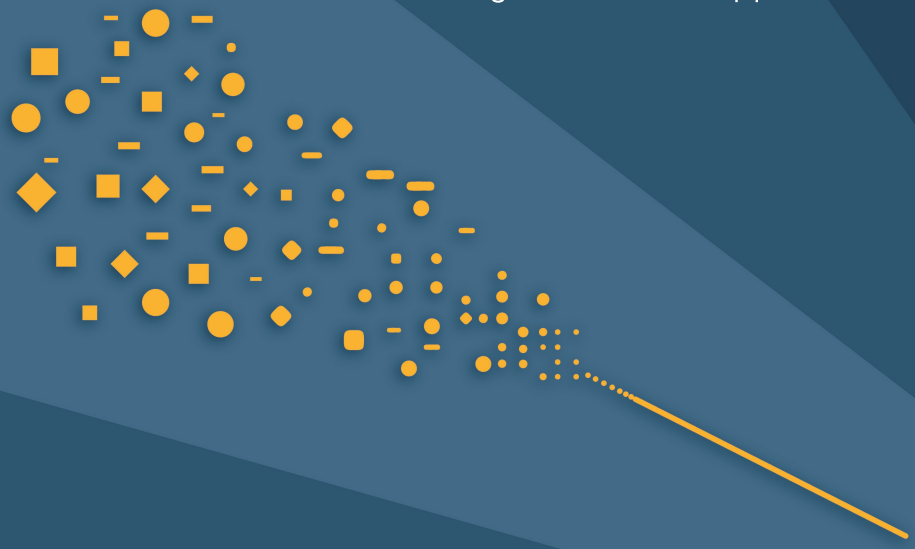
FROM DATA TO INSIGHT

Assessing the decontamination efficacy of mild food processing technologies using data-driven approaches



GEORGE PAMPOUKIS

2026

Propositions

1. Performing meta-analysis demands a scientific detective mindset.
   (this thesis)

2. In predictive modeling, carefully defined and consistently collected parameters outweigh additional model complexity.
   (this thesis)

3. Greater dietary variety increases household food waste.

4. Consumer preferences undervalue food safety.

5. In complex problems, identifying the right question is a bigger bottleneck than finding a technical solution.

6. Effective communication alternates diplomacy with strategic directness.

Propositions belonging to the thesis, entitled

From data to insight: Assessing the decontamination efficacy of mild food processing technologies using data-driven approaches

George Pampoukis
Wageningen, 16 October 2026

From data to insight

Assessing the decontamination efficacy of mild food processing technologies using data-driven approaches

George Pampoukis

Thesis committee

Promotors
Prof. Dr H.M.W. den Besten
Professor of Food Microbiology
Wageningen University & Research

Prof. Dr M.H. Zwietering
Professor of Food Microbiology
Wageningen University & Research

Other members
Prof. Dr K.A. Hettinga, Wageningen University & Research
Prof. Dr Kemal Aganovic, German Institute of Food Technologies (DIL), Germany
Dr Karin I. Beekmann-Metselaar, NIZO
Dr R.A.H. Timmermans, Wageningen University & Research

This research was conducted under the auspices of the VLAG Graduate School (Biobased, Biomolecular, Chemical, Food and Nutrition Sciences).

From data to insight

Assessing the decontamination efficacy of mild food processing technologies using data-driven approaches

George Pampoukis

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 16 October 2026
at 3.30 p.m. in the Omnia Auditorium.

George Pampoukis

From data to insight

Assessing the decontamination efficacy of mild food processing technologies using data-driven approaches

191 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2026)
With references, with summary in English

DOI https://doi.org/10.18174/681689

Table of contents

Chapter 1	General Introduction	1
Chapter 2	Ranking factors affecting the decontamination efficacy of non-thermal plasma: The approach of dissipated power per plasma volume through machine learning modeling	13
Chapter 3	A metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications	51
Chapter 4	The decontamination efficacy of pulsed electric field in food: Establishing global kinetic parameters	77
Chapter 5	High-pressure processing inactivation of foodborne bacteria and viruses in foods	111
Chapter 6	General Discussion	135
References		155
Appendix		181
	Summary	182
	Acknowledgements	185
	About the author	189
	Overview of completed training activities	190

Chapter 1

General Introduction

1.1 Global goals for safe food, lower food waste, and more sustainable production

The growing world population increases the need for safe, nutritious, and high-quality food that is accessible worldwide and produced with minimal environmental impact and minimal food loss and waste. This is a multidimensional challenge that becomes harder when socioeconomic and ideological factors shape consumer choices (Thirumdas et al., 2018). The United Nations 2030 Agenda of Sustainable Development Goals (SDG) is related in multiple points to food production. SDG 2 for zero hunger targets reliable access to safe, nutritious food through resilient, sustainable agriculture. SDG 12 for sustainable consumption and production aims the reduction of food loss and food waste and the adoption of sustainable practices from companies, and SDG 13 for climate action aims to combat climate change and its impacts by integrating measures into policies (United Nations, 2015). Food systems are responsible for a third of global anthropogenic GHG emissions (~18 Gt CO_2-eq/yr). Distribution, processing, consumption and end-of-life disposal account for 29% of this amount (~5.2 Gt CO_2-eq/yr), whereas processing itself accounts for ~0.59 Gt CO_2-eq/yr within the food system (Crippa et al., 2022, 2021). Food processing is essential in the food supply chain to ensure a safe and nutritional food supply, contributing at all stages from raw materials and ingredients to the final product. It enables preservation, palatability, transport stability, and convenience (Knorr and Watzke, 2019). However, many processed foods currently consumed are unhealthy, often containing high levels of salt, refined starch, sugar, and fat (Augustin et al., 2016). Mild or non-thermal processing technologies, such as high-pressure processing (HPP), have the potential to provide alternative preservation methods that enhance food safety and improve shelf life, reduce the need for added preservatives such as salt, and, for certain products, help to retain fresh-like sensory properties (O' Neill et al., 2018). In addition, emerging thermal technologies (radiofrequency, microwave, and ohmic heating) and non-thermal technologies (pulsed electric fields (PEF), HPP) may provide environmental benefits by improving overall energy efficiency or reducing the use of non-renewable resources (Pereira and Vicente, 2010).

1.2 Consumer preference and the case of mild processing

Greater processing and worldwide supply networks have increased the sense of distance and information gaps between consumers and food business operators about

food's "how" and "where". This has contributed to a consumer trend towards clean-label foods, characterized by front-of-pack cues such as natural or organic logos and "free-from" claims, and by ingredient lists that, upon back-of-pack inspection, are short, simple, familiar, and perceived as free of artificial additives or preservatives (Asioli et al., 2017). Mild processing technologies such as non-thermal plasma (NTP), PEF, and HPP are often presented as potential tools to meet these expectations because they can enhance safety and shelf life with a lower thermal load and, in some cases, reduced need for added preservatives (Zhang et al., 2019). This thesis therefore evaluates the decontamination efficacy of NTP, PEF, and HPP using available literature data, providing a quantitative basis to judge when, and to what extent, such mild technologies can realistically substitute or complement conventional preservation processes such as thermal treatments or the use of food additives.

1.3 The different mild processing technologies

NTP is an ionized gas that is commonly described as a fourth state of matter in which electrons possess much higher energy than ions, neutrals, and atoms. Plasmas are thermal when electron and heavy-particle temperatures are similar and non-thermal when these temperatures differ. In the non-thermal case, electron temperatures can approach about 10,000 °C while the bulk gas remains comparatively cool, typically below 70 °C, and loses heat quickly because it is a gas (Liao et al., 2017; Schnabel et al., 2021a). For food applications, microbial inactivation by non-thermal plasma involves multiple pathways that include UV-sourced DNA damage, lipid oxidation, protein modification, apoptosis, electrostatic disruption, and electroporation (Liao et al., 2017). Non-thermal plasma has a broad range of applications in food processing. It enables dry decontamination of food surfaces on meat, poultry, fish, and freshly harvested produce, and it can treat granular and particulate foods such as dried milk, herbs, spices, and sprouting seeds (Misra et al., 2011). In general, dry decontamination is preferable for dry food processing facilities, and NTP can be useful due to a broad antimicrobial and application spectrum (Dhaliwal et al., 2025). Another interesting application of NTP is the use of plasma-treated water (PTW) for washing fresh-cut produce (Schnabel et al., 2021b).

PEF applies short high-voltage pulses between electrodes while the food flows through the treatment zone. The treatment increases membrane permeability through electroporation, oxidative damage, and changes to membrane proteins, which leads to microbial inactivation (Kotnik et al., 2019). Because the product flows through the treatment zone, PEF can be applied as a continuous in-line process. However, PEF can only

work for liquid products for decontamination purposes. For industrial-scale applications, typically an electric field strength of 10-20 kV/cm (Toepfl, 2012) and an energy input up to 200 kJ/kg is used for fresh fruit juices and acidified vegetable juices to balance between the effectiveness of treatments and costs (Kempkes, 2017; Timmermans et al., 2022). Interesting applications of PEF for microbial inactivation include fruit juices (Timmermans et al., 2019), milk and dairy products (Cavalcanti et al., 2023), alcoholic products (González-Arenzana et al., 2015), egg products (Monfort et al., 2010), and plant-based beverages (Thamsuaidee et al., 2024).

HPP involves the use of pressures in the range of 100-800 MPa to inactivate microorganisms through altering their membrane lipid state, shifting intracellular pH, and disrupting essential physiological function (Balasubramaniam et al., 2015). Currently, food industry applications of HPP involve processing at hydrostatic pressures of 400 to 600 MPa for up to 6 minutes to inactivate pathogenic and spoilage microorganisms in various food products (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022). The process is typically in-container and batch. The product portfolio of HPP has a wide variety, including both solid and liquid products such as ready-to-eat products, fruit and vegetable juices and purees, beverages, seafood, etc. A lot of these product categories, such as deli meats, fruit, and vegetable purees, juices, and beverages, are already available in the global food market (Ghafoor et al., 2020). Additionally, EFSA has recognized high-pressure processing as effective for milk pasteurization and for control of *L. monocytogenes* in ready-to-eat products (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022).

The three mild processing technologies share the common feature of operating under non-thermal or mildly thermal conditions, but they also differ in important ways. All three can inactivate pathogenic and spoilage vegetative bacteria, yeasts, molds, viruses, and parasites. However, for spore-forming microorganisms, substantial inactivation by PEF and HPP generally requires the combination with thermal effects, whereas non-thermal plasma has been shown to inactivate spores without such combinations under prolonged exposure time to the reactive oxygen and nitrogen species (RONS) produced (Battaggia et al., 2024; EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Siemer et al., 2014). Among the three technologies, HPP is currently the most successfully commercialized non-thermal process for decontamination purposes (Farkas, 2016; Huang et al., 2017). PEF is also widely commercialized, but mainly for structural modification of plant tissues in the potato industry (fries and chips) to reduce oil uptake and breakage and to obtain smoother cutting surfaces and higher yield, rather than primarily for microbial inactivation (Hill et al., 2022). NTP applications are largely at laboratory and pilot scale and, although upscaling efforts are ongoing, industrial implementation for food decontamination remains limited (Weihe et al., 2024).

Despite these advantages and opportunities, several issues limit the broader adoption of mild-processing technologies. These include high investment costs, a limited range of products with demonstrated suitability, engineering challenges at industrial scale, limited knowledge about variability in microbial resistance, low consumer awareness, the absence of generic predictive models for efficacy, incomplete risk assessment for large-scale use, and, in some cases, unclear regulatory status.

1.4 Quantifying efficacy with data-driven methods

1.4.1 The approach of meta-regression

Food business operators need to anticipate the performance of their decontamination processes in order to design and validate effective control measures. EU food hygiene legislation, notably Regulation (EC) No 852/2004 and its associated guidance on food safety management systems, requires food business operators to validate HACCP-based procedures and recognises that validation can be supported by scientific evidence such as predictive microbiology or challenge testing, provided that this demonstrates that the chosen critical limits achieve the intended effect on the hazard (European Commission, 2016; European Parliament and Council of the European Union, 2004). Predictive modelling contributes to such validation by quantifying the survival or growth of relevant microorganisms under defined product and process conditions and by explicitly accounting for variability in product, organism, and processing factors. Data for predictive models can be obtained from dedicated laboratory experiments, from publicly available repositories and databases, or from the broader scientific literature. Scientific literature is growing exponentially, with an average doubling period of 15 years (Fortunato et al., 2018). In order to make sense of the rapidly expanding literature and provide stakeholders with digestible overviews, meta-analysis can be a useful tool.

Meta-analysis is a statistical method that combines a large collection of analysis results from individual studies for the purpose of integrating the findings (Glass, 1976). In medical sciences, formal meta-analyses can sometimes substitute for a pivotal clinical trial or extend and strengthen the generalizability of evidence on drug efficacy (Berlin and Colditz, 1999). In food sciences, meta-analyses have been applied across diverse sub-domains. They have been used to study willingness to pay for local, sustainable, or functional foods (Dolgopolova and Teuber, 2018; Printezis et al., 2019; Tran et al., 2024). They also synthesize evidence on sensory acceptance of fortified foods and on how sensory and physical characteristics affect satiation (Appleton et al., 2021; Grigor et al., 2016). Other applications include evaluating the effectiveness of consumer and

food-handler food-safety education (Insfran-Rivarola et al., 2020; Young et al., 2015) and assessing the efficacy of interventions and processing stages along the food chain on microbial safety (Dogan et al., 2022; Torrents-Masoliver et al., 2025; Zdolec et al., 2022).

Although a meta-analysis can assess the study effects that induce variability/heterogeneity in the data, meta-regression is a tool that merges meta-analysis with linear regression principles to better explore and quantify these effects (Baker et al., 2009). Meta-regression models have gained popularity in predictive food microbiology as a way to create more generic models for Microbial Risk Assessments that also include variability (Den Besten and Zwietering, 2012; Garre et al., 2023; Gonzales-Barron and Butler, 2011). In food processing, meta-regression can be used to obtain global inactivation kinetic parameters. For example, in thermal processing, a global estimate of the *D*-value, which is the time needed for a ten-fold reduction of the microbial population, can be calculated using data from multiple laboratories. This global *D*-value can then be refined by stratifying the data for different microbial species and food matrices. Additionally, because the *D*-value corresponds to a single processing condition (temperature level), it is useful to estimate the *z*-value, the temperature increase needed to reduce the *D*-value by tenfold, to predict *D*-values at other temperatures (Bigelow, 1921). The same principle extends to other technologies, such as HPP, where pressure replaces temperature to estimate the z_P-value. Reported *D*-values are typically $\log_{10}$-transformed and fitted with a linear regression on temperature (or pressure for HPP), so that the transformation stabilizes variances and improves the normality of residuals (Box and Cox, 1964), and the *z*- or z_P-value is then obtained as the negative inverse of the regression slope. The described meta-regression framework has been implemented in thermal processing (Diao et al., 2014; Rigaux et al., 2013; Van Asselt and Zwietering, 2006) and high-pressure processing (Santillana Farakos and Zwietering, 2011; Torrents-Masoliver et al., 2025).

1.4.2 The approach of data science

While meta-analysis and meta-regression describe how quantitative evidence can be synthesised and modelled, data science provides the broader framework for how such data are found, curated, explored, and ultimately analysed. Data science can be defined as the interdisciplinary science that combines statistics, data mining, databases, and distributed systems and focuses on extracting useful insights from data through programming and domain expertise (Van der Aalst, 2016). Extracting value from data requires a data processing pipeline, which typically includes data retrieval, data wrangling (cleaning) procedures, duplicate elimination, storage in a centralized repository,

pre-processing for a given analysis, and statistical analysis through predictive approaches (e.g., machine learning) or graph analytics. In practice, most effort is spent on data pre-processing rather than on fitting the final models. Therefore, data science needs to be paired with appropriate data management practices to ensure that datasets are ready for statistical analysis and require less ad hoc pre-processing. In this direction, data engineering, which primarily focuses on data management and on building systems that standardise and partly automate pre-processing and other error-prone steps in the pipeline, can be particularly useful (Romero and Wrembel, 2020).

Moving from data management to analysis, exploratory data analysis (EDA) is usually the first formal step in a data science project (Chatfield, 1986). EDA summarizes and visualizes data to reveal structure, detect anomalies, and generate hypotheses before modeling. It relies on simple plots and robust summaries. Interactive data-visualization tools can further support EDA by allowing users to examine data in real time and reduce information overload (Yigitbasioglu and Velcu, 2012). Once the dataset has been compiled, cleaned, and explored so that its structure and main patterns are understood, the analysis can proceed to formal statistical and regression modeling and, where appropriate, to machine-learning approaches.

The meta-regression framework was introduced in Section 1.4.1 as a linear regression concept. However, the same concept can be extended to a multivariable framework when several factors influence the outcome. For example, for thermal processing, temperature, microbial species, and food matrix can all affect decontamination efficacy as expressed by the D-value. This can be addressed either by stratifying the dataset (e.g., fitting separate linear models within combinations of species and matrices) or by using multivariable regression models, alternative parametric equations, or machine-learning algorithms that incorporate all relevant predictors simultaneously. Multiple linear regression models a continuous response (e.g., a D-value or $\log_{10}D$) as a linear combination of several explanatory variables plus random error (Montgomery et al., 2021). Each explanatory variable has its own regression coefficient (and z-value, i.e., the negative inverse of this coefficient/slope), and multiplicative (cross) terms combining two explanatory variables can be introduced to allow the model to capture non-additive, nonlinear patterns.

Machine learning (ML) refers to a family of algorithms that learn patterns from data to make predictions or discover structure, with minimal or no reliance on previous knowledge. Core categories include supervised learning (e.g., regression and classification), unsupervised learning (e.g., clustering and dimensionality reduction), and neural-network-based models, often referred to as deep learning (Bishop, 2006; Hastie et al., 2009). In contrast to classical linear regression, which assumes a fixed functional form between predictors and response specified a priori, many ML methods learn flexible mappings from high-dimensional inputs to outputs, which can improve predictive

performance but typically reduce model interpretability (Hastie et al., 2009).

In practice, meta-regression and data science approaches are complementary rather than competing. EDA and interactive visualization clarify the structure of complex datasets and help identify relevant predictors. Then, based on the outcome, meta-regression can be implemented using simple or multiple linear models, or extended with machine-learning algorithms, depending on the research question and data structure. In this context, greater model complexity does not necessarily yield better or more complete insight. From an information-theoretic perspective, model selection is a bias-variance trade-off, and the statistical principle of parsimony implies that models should include no more parameters than are needed to describe a problem, because overly complex models inflate variance and tend to overfit (Burnham and Anderson, 2002; Hastie et al., 2009). Following Box, "since all models are wrong", they cannot be made correct by "excessive elaboration"; instead, one should seek models that are useful for the purpose at hand (Box, 1976). In line with this view, Shmueli emphasizes that model complexity and choice should be guided by whether the primary goal is explanation or prediction, and that more complex models do not inherently provide better predictive performance (Shmueli, 2010).

Together, these concepts, meta-regression, and data science, provide the methodological backbone of this thesis. The next section translates this combined framework into the concrete aims, research questions, and contributions that structure the subsequent chapters.

1.4.3 Aims, research questions, and novelty of the thesis

Using the meta-regression and data-science framework described above, this thesis aimed to obtain quantitative, generic insights into the decontamination efficacy of NTP, PEF, and HPP under industrially relevant applications. The goal was not to optimize a single product or process, but to organize and exploit existing literature so that decontamination efficacy and its sources of variability can be described in terms of global kinetic parameters, matrix effects, and microbial species effects, and to translate these into benchmarking insights that can inform process designers and risk assessors. The sources of variability were explored, such as matrix category, pH, microbial genus or hazard group, shape resistance across these technologies, and how these relationships can be summarized in a way that is directly usable in practice.

Within this aim, the thesis addressed a series of interrelated research questions. A first question concerned non-thermal plasma (**Chapter 2**): which process, product, and microbial factors most strongly affect its decontamination efficacy, and whether an integrated descriptor of power input can capture a substantial part of the variability

in $\log_{10}D$-values across different plasma configurations. However, the available NTP literature was too heterogeneous and inconsistently reported to support robust estimation of global kinetic parameters that would be comparable to those derived for PEF and HPP, and it was therefore not taken forward into the subsequent global-parameter benchmarking and efficiency comparisons.

The thesis then examined how improved data management and metadata standardization, particularly for non-thermal plasma, can enhance comparability, reproducibility, and reusability of decontamination data and thereby support more robust modelling and the establishment of global kinetic parameters in future work (**Chapter 3**).

Building on this foundation, a next question was whether and to what extent global kinetic parameters for mild technologies can be derived from heterogeneous literature data, and how these parameters change when separating the microbial species and food matrix effects. For pulsed electric fields (**Chapter 4**) and high-pressure processing (**Chapter 5**), this involved estimating global and matrix-specific parameter estimates of decontamination efficacy and quantifying how matrix characteristics, such as acidity, shift the required processing intensity for a given $\log_{10}$-reduction target. Lastly, thermal processing (TP) was compared to HPP and PEF in terms of efficacy but also efficiency (**Chapter 6**). The outline of the thesis is presented in Figure 1.1.

The thesis is novel in several respects. It applied a unified quantitative strategy across three different mild technologies, showing that global kinetic parameters and generic patterns can be extracted from dispersed literature once data are carefully curated and harmonized. For non-thermal plasma, the most important factors affecting decontamination efficacy were ranked using machine learning to help in process optimization. The literature screening was supported by active learning (a subfield of machine learning) to improve the efficiency of the procedure. The approach of active learning for making literature screening was introduced and evaluated for the first time in food microbiology. It also introduced a dedicated metadata schema for non-thermal plasma decontamination in food-related applications, implemented as an extension of an existing plasma metadata schema, and illustrated with concrete, human- and machine-readable examples. For pulsed electric fields, the thesis compiled a large, harmonised database across multiple microbial groups and matrices and expressed decontamination efficacy consistently in terms of the volumetric energy input (estimated by multiple process and matrix components) per $\log_{10}$-reduction, linking these results to a decision tree that provides first-order estimates of the energy needed to achieve targets such as 5-$\log_{10}$-reductions and their prediction intervals in different product categories. It also developed an interactive dashboard that can be used to visualize different parameters of the underlying database in real time. For high-pressure processing, it extended previously developed databases, aggregated $\log_{10}D$-values obtained at non-thermal temperatures, compared them to an industrially relevant pres-

sure level, and ranked microbiological hazards into resistance categories while taking into account the matrix effects on inactivation. **Chapter 6** brings all pieces together by comparing PEF and HPP with conventional TP at matched microbiological targets and by linking the resulting efficacy benchmarks to energy use and greenhouse gas emissions. In doing so, the general discussion brings the meta-regression models developed in the previous chapters to a broader food safety and sustainability framework and outlines how the same data-driven approach can be extended to other emerging technologies (Figure 1.1).

These contributions are relevant as benchmarks for several groups of stakeholders. For risk assessors, the global kinetic parameters, hazard resistance rankings, and matrix effects provide transparent, literature-based inputs for quantitative microbial risk assessments and for evaluating whether specific non-thermal processes can meet required levels of risk reduction in different products. For food business operators, the parameter estimates can serve as practical starting points for process design and validation, allowing them to position their own data against broader evidence and to perform scenario analyses without having to build models from scratch. For data scientists working in food and microbiology, the curated datasets, metadata schema, and analysis workflows provide concrete examples of how active learning, meta-regression, and machine learning can be integrated into domain-specific pipelines, highlighting the gains that can be achieved when experimental data are made compliant with the FAIR (findable, accessible, interoperable, reusable) principles. For society at large, the thesis contributes to a more transparent evidence base on mild processing technologies, clarifying under which conditions and to what extent they can deliver microbial safety with lower processing severity, and how such processes might support cleaner labels, reduced food waste, and more sustainable production. In this way, the work helps to connect methodological advances in data-driven modelling with practical decision-making in food safety management and with broader discussions on the role of emerging technologies in future food systems.

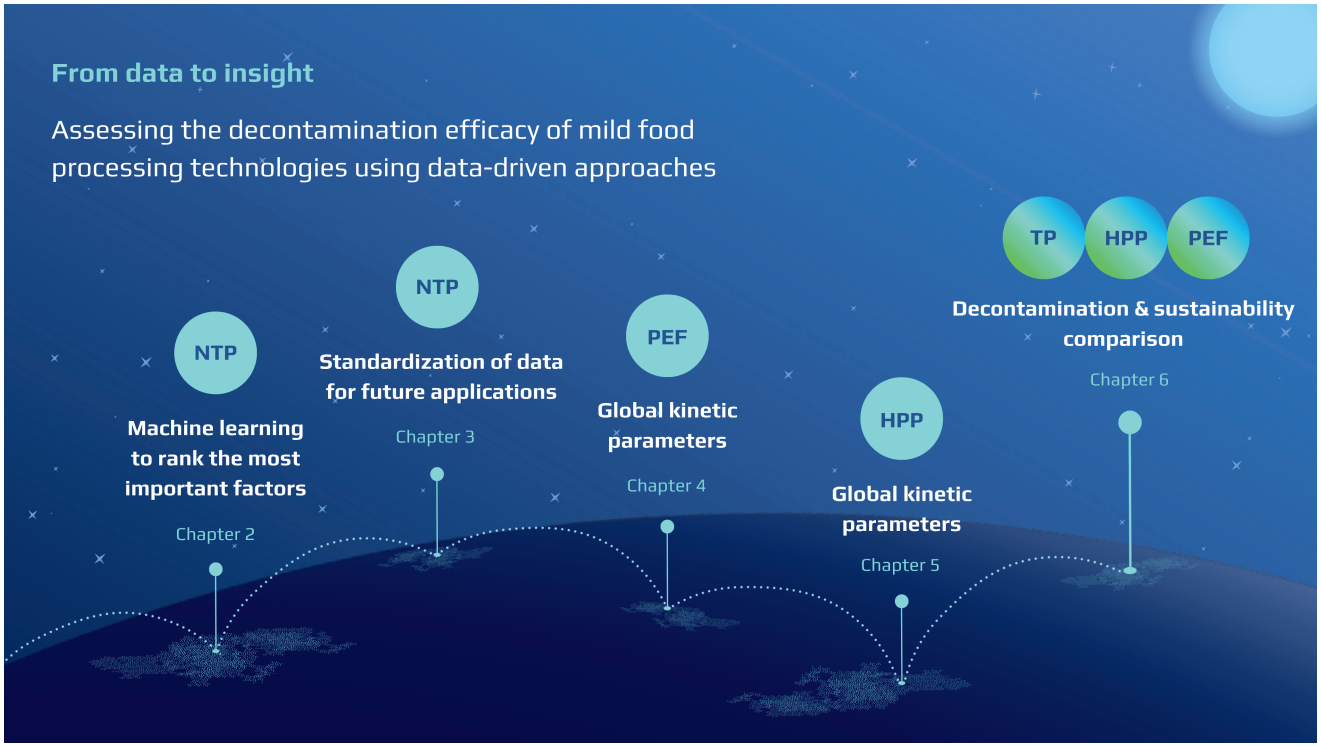


Figure 1.1: An overview of the thesis chapters after the General Introduction. Chapter 2: Ranking factors affecting the decontamination efficacy of non-thermal plasma: The approach of dissipated power per plasma volume through machine learning modeling; Chapter 3: A metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications; Chapter 4: The decontamination efficacy of Pulsed Electric Field in food: Establishing global kinetic parameters; Chapter 5: High-pressure processing inactivation of foodborne bacteria and viruses in foods, and Chapter 6: General Discussion.

Chapter 2

Ranking factors affecting the decontamination efficacy of non-thermal plasma: The approach of dissipated power per plasma volume through machine learning modeling

George Pampoukis, Marcel H. Zwietering, Heidy M.W. den Besten

Food Microbiology, Wageningen University & Research, Wageningen, the Netherlands

Published in *Innovative Food Science & Emerging Technologies*, Volume 96, August 2024, 103773. https://doi.org/10.1016/j.ifset.2024.103773

Abstract

Non-thermal plasma treatment can preserve food, but a meta-analysis assessing its efficacy has not been performed. This study retrieved the inactivation kinetics parameter $\log_{10}D$ (n = 519), which varied largely among different plasma setups. Atmospheric pressure plasma jet, corona discharge, surface barrier discharge, dielectric barrier discharge, and inductively-coupled plasma had the highest efficacy with median D-values under 2 minutes. Dielectric barrier discharge, the most frequent setup (n = 160), was analyzed using dissipated power per plasma volume (W/cm^3) as an integrated predictor of decontamination efficacy. Using conventional and machine learning approaches the most correlated parameters to the $\log_{10}D$ were: dissipated power per plasma volume and matrix category, followed by microbial genus and pH. This study uses active learning to improve literature screening for data collection and various data analysis techniques for data treatment and ranking of the factors affecting non-thermal plasma decontamination.

2.1 Introduction

Food safety accounts annually for 33 million disability-adjusted life years (DALY) and 420,000 deaths, globally (World Health Organization, 2015). Additionally, a lot of food spoils and is wasted. In Europe, 186 Mt CO_2-eq., 1.7 Mt SO_2-eq., and 0.7 Mt PO_4-eq. can be attributed to food waste which means that food waste is responsible for 15% of the total climate impacts of the entire food supply chain (Scherhaufer et al., 2018), posing major challenges to ending global hunger and malnutrition (FAO et al., 2023). To enhance food safety and reduce food spoilage, various methods have been proposed in the past to preserve food, and thermal processing and food additives prevailed in the food industry. However, some food additives face consumer acceptance issues (Bearth et al., 2014), while others can affect the organoleptic characteristics of food. The latter stands true also for thermal processing, where for certain applications the thermal effects are not desirable (e.g., fresh produce). Along with the trends toward the application of sustainable food processing technologies, the development of non-thermal processing methods such as non-thermal plasma (NTP) has emerged.

Plasma is a partially or fully ionized gas that can be considered the fourth state of matter with the highest energy level followed by gas, liquid, and solid. The categorization of thermal and non-thermal is based on the relationship of electrons' temperature with the temperature of ions, neutral molecules, and atoms. When these temperatures are almost the same, then the plasma is categorized as thermal, while when these are different it is categorized as non-thermal. In the latter case, the electrons can reach temperatures up to 10,000 °C (Liao et al., 2017; Scholtz et al., 2015). However, this is not the overall gas (plasma) temperature which remains below 70 °C (Schnabel et al., 2021a) and cools down rapidly after its production due to its gaseous state. Possible mechanisms of microbial inactivation through NTP include DNA damage by UV radiation, lipid oxidation, protein modulation, apoptosis, electrostatic disruption, and electroporation (Liao et al., 2017). The different devices, the ambient conditions, and the mode of treatment (i.e., direct or indirect) can affect the production and stability (i.e., concentration, life cycle) of the reactive oxygen and nitrogen species (RONS) present in the plasma gas. Examples of indirect treatment include applications where the plasma gas is produced distantly from the sample or applications related to plasma-treated water that is then used to decontaminate the sample. The NTP devices that are described in the literature are dielectric barrier discharge (DBD), atmospheric pressure plasma jet (APPJ), inductively-coupled plasma (ICP), gliding arc discharge (GAD), point-to-plate (PTP), corona discharge (CD), glow discharge (GD), surface barrier discharge (SBD), and surface micro-discharge (SMD). Additional categorization methods exist based on the type of source (i.e., microwave plasma source (MPS), radio-frequency (RF)), the pressure (i.e., low pressure, atmospheric pressure), and the discharge type

(i.e., filamentary, diffuse). Several studies have used machine learning for more efficient screening in systematic reviews (Ferdinands et al., 2023; Miwa et al., 2014), but also in the non-thermal plasma domain. Some of these studies aimed to predict the best non-thermal plasma parameters' combinations for seed priming (Shelar et al., 2022), or to predict the microbial inactivation of plasma-activated liquids for biomedical applications (Özdemir et al., 2023), and for real-time plasma diagnostics (Bonzanini et al., 2022).

Our study used active learning for the retrieval of relevant NTP studies and investigated the overall differences in the decontamination efficacy of ten different setups. Then, the concept of dissipated power per plasma volume was selected as a basis to prioritize and rank the parameters that affected the decontamination efficacy of NTP using a DBD setup through conventional and machine learning approaches.

2.2 Materials and methods

2.2.1 Data collection

The data were collected using five databases, namely, Scopus, Web of Science, SciFinder, Food Science and Technology (FSTA), and CAB Abstracts. Three sets of keywords were used, one for plasma, one for inactivation/decontamination, and one for the microbial species of interest (Supplementary Material A). The microorganisms of interest were *Escherichia coli* (including total coliform counts), *Listeria* spp., *Saccharomyces cerevisiae* (including total yeast counts), lactic acid bacteria, and *Bacillus* spp. These microorganisms were selected as a broad set of representants of typical pathogenic or spoilage Gram-negative and Gram-positive bacteria, spore formers, and yeasts. Each search was performed independently for each microbial genus, before merging, and the duplicates were removed. The data screening was conducted in two steps. For the first step, the studies were screened based on the information on the type of plasma device, achieved microbial inactivation > 0.5-$\log_{10}$-reduction, and reported power expression such as applied, forward, reverse, reflected, active, apparent, operating, generating, discharge, dissociated, transferred, input, or dissipated power. The threshold of 0.5-$\log_{10}$-reduction was selected because the inherent uncertainty of the measurement below this level is highly affected by the detection limit of the method and could mean that there was no actual inactivation. For the second step, the studies were screened in more detail for one specific plasma device setup, namely, the DBD setup, and expressions of input or dissipated power, electrodes' dimensions, and distance between the electrodes were compiled. The rationale behind this was a first investigation of the individual studies that demonstrated that

the different plasma setups induce a high variation of different parameters, that were not consistently reported. For NTP, a defining parameter equivalent to temperature for thermal processing and pressure for high-pressure processing could not be easily found. Multiple expressions of power exist in the literature such as transferred, active, applied, and apparent power, which cannot be compared with certainty, as well as different approaches for estimating the RONS either in the sample or in the plasma gas, different analytical methods, and different chemical components involved. Parameters such as voltage and frequency, although often reported in the literature, are not descriptive because they do not provide information regarding the waveform characteristics of the applied voltage. Furthermore, in most cases, it was not clear if the voltage reported was the peak-to-peak or the maximum voltage and if the frequency was the frequency of the powered device or the frequency related to the actual discharges. Thus, the second step of the screening procedure was performed by narrowing it down to only include DBD setups, because this setup was most frequently used in the studies, with clear expressions of power (i.e., input or dissipated power), electrodes' dimensions, and distance, to standardize the setup and compare the studies more consistently. In-package DBD setups where the powered electrode is placed inside a packaged food product were excluded due to the different approach of this setup and the difficulties in standardizing when compared to the remainder of the setups. Also, studies that reported synergistic effects when combining NTP with other technologies or substances were excluded, except for the studies that provided the individual effect of NTP. Lastly, studies referring to plasma-treated water or plasma-activated water were excluded due to the difficulties of comparing the microbial inactivation with the remainder of the setups. The workflow of our study is illustrated in Figure 2.1.

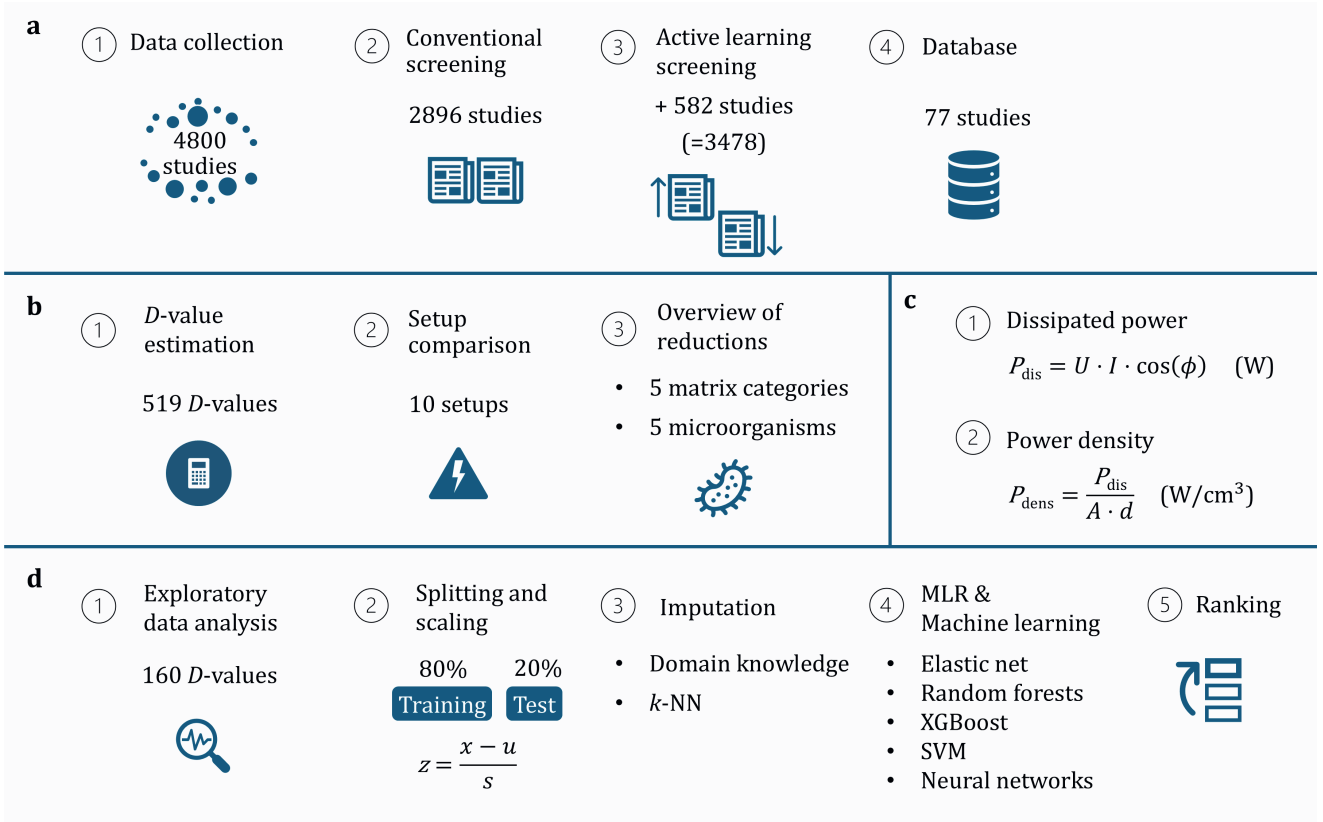


Figure 2.1: a: Data collection, and screening workflow. b: Overview of all data points across different setups. c: The concept of dissipated power per plasma volume. d: Exploratory data analysis and quantitative modeling workflow.

2.2.2 Screening with an active learning approach

In order to optimize the screening procedure, active learning was used for the second step of the screening process after narrowing down the scope of research to the DBD setup. Active learning is a subfield of machine learning in which the learning algorithm is allowed to choose the data from which it learns (Settles, 2009). This approach can work for various concepts such as speech recognition and information extraction, but in screening literature studies, the goal is to change the original order of the studies that are to be screened as relevant or irrelevant, choosing first the studies that have a higher probability of being relevant according to the previous researcher's choices (Settles, 2009; Van de Schoot et al., 2021). The researcher still has full control of the screening process, since active learning only changes the order of the studies to be reviewed and does not label them as relevant or irrelevant. This approach was used and evaluated in various studies (Harrison et al., 2020; Marshall and Wallace, 2019; O'Mara-Eves et al., 2015; Van de Schoot et al., 2021; Wallace et al., 2010), but, to our knowledge, not yet in the field of food microbiology. As the first step of the active learning procedure, an initial screening (prior knowledge) of some studies is conducted manually by the researcher to point out some relevant and irrelevant studies and then this knowledge is used to train machine learning algorithms. In our case, these algorithms were term frequency-inverse document frequency (TF-IDF) as the extractor, Naive Bayes as the classifier, and Dynamic resampling as the data rebalancing strategy that makes the process more efficient for datasets that potentially include many irrelevant and a few relevant studies. The query strategy was set to maximum so that the algorithms put first for screening the study with the highest probability of being relevant based on the title and the abstract (Van de Schoot et al., 2021). Based on the researcher's choice for this study (relevant or irrelevant) the algorithms are trained again including the new knowledge and put another study first etc. The procedure stops after reaching the stopping criterion that has to be defined by the researcher and is the number of consecutive irrelevant studies after which the process stops. This approach is assumed to save time as the number of studies to be screened is more realistic, and at the same time, it also does not make the search too precise and ensures that relevant studies are not missed (Van de Schoot et al., 2021). The latter is crucial for non-thermal plasma since the term is undoubtedly broad and making it more specific for certain applications or setups can induce selection bias, already from the first step. Our total study list included 4800 studies (after removing the duplicates) for all plasma setups and *Escherichia coli* (including total coliform counts), *Listeria* spp., *Saccharomyces cerevisiae* (including total yeast counts), lactic acid bacteria, and *Bacillus* spp. From the first step 2896 (of 4800) studies were randomly selected and screened manually (without active learning) for information on the type of plasma device, achieved microbial inactivation > 0.5-$\log_{10}$-reduction, and reported power expression. At this point 1904 (4800

minus 2896) studies were not screened yet. From the 2896 studies, that were initially screened manually, 57 were labeled as relevant and 2839 as irrelevant. To start the active learning procedure to screen the remaining 1904 studies and train the active learning algorithms for the second step, these 2896 labeled studies were used as prior knowledge to start with. The active learning procedure was performed using the ASReview LAB software (v1.2.1) (ASReview LAB developers, 2024). It should be noted that this prior knowledge can be established by using a way more limited number of studies (e.g., 10) just to "warm-up" the active learning algorithms. The "warm-up" in the context of an active learning procedure is the input of several starting data (like 10, or in our case 2896, and not like a "seed" which is one starting value). For the second step, the 1904 studies that were not screened previously were used as input for the active learning procedure with a focus on DBD setup. The studies that were screened with active learning were 582 (from 1904), before reaching the stopping criterion of 200 consecutive irrelevant studies. From the second step of screening, 32 out of the 582 additional relevant studies were retrieved. The reviewing process with active learning is illustrated in Figure 2.2. So in total, 3478 out of 4800 (2896 plus 582) studies were screened from which 89 (32 plus 57) were relevant, 3389 were irrelevant and the remaining 1322 (1904 minus 582) studies were not screened. The stopping criterion of 200 was an arbitrary choice, but was expected to be a conservative one in the manner that in this sequence of 200 studies not even one relevant study was found, suggesting that active learning successfully had put already on the top of the order the relevant studies (Figure 2.2). In addition to the 4800 studies, 294 *Salmonella* studies were searched, screened, and compiled retrospectively, when it was perceived that there were a lot of valuable studies available for *Salmonella* that could further assist in exploring NTP decontamination efficacy using the DBD setup. The search for *Salmonella* spp. was conducted using keywords only relevant to the DBD setup and not for all plasma setups as was done for the remainder of the microorganisms (Supplementary Material A). From this search, 7 more studies (in addition to the 89) were found to be relevant. In total, after the finalized screening for the inclusion of parameters in the database, 19 studies from both steps of the preliminary screening (i.e., manual or active learning assisted) were excluded because information on power or achieved reduction, or both, was incomplete. As a result, 77 studies (519 *D*-values or 517 microbial reduction entries (for two *D*-values the reductions were not reported)) were assessed qualitatively (for all setups), of which 9 studies on the DBD setup (160 *D*-values) were used for the quantitative modeling analyses (the 77 studies included in our database can be found in Supplementary Material B). The studies that reported the input power instead of the dissipated one, could not be utilized because the power loss can vary significantly for different DBD devices.

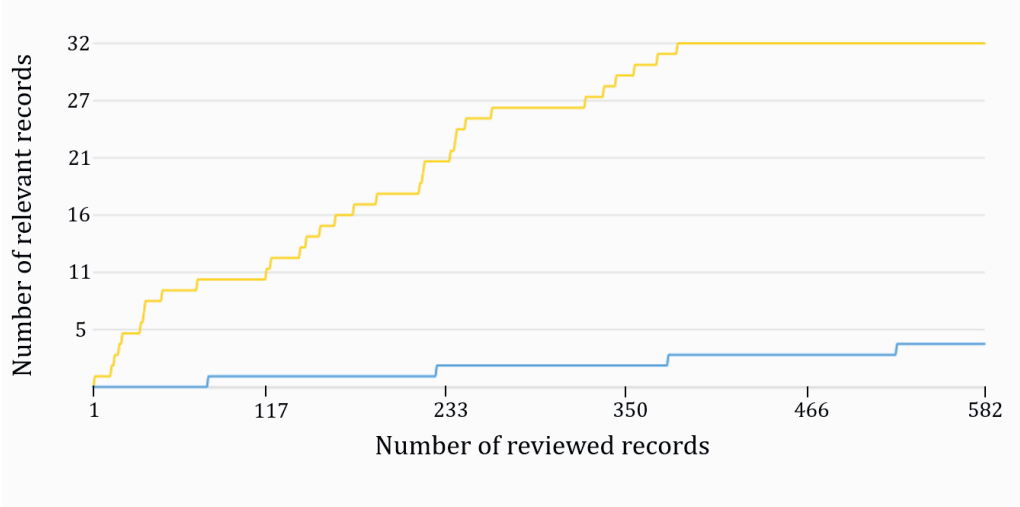


Figure 2.2: A visualization of the two different approaches to screening i.e., the manual screening procedure (blue line) and the screening procedure as assisted by active learning (yellow line). The random relevant studies (blue line) are the number of studies that are expected to be relevant in a random dataset during a manual screening procedure and are estimated based on a generic simulation of this procedure as provided by ASReview LAB software (v1.2.1). The relevant studies that are illustrated in the figure (yellow line) are based on our dataset.

2.2.3 Overview of data points and qualitative analysis

The decontamination efficacy of the different plasma setups was evaluated using the *D*-value, which is the time needed to reduce a microbial population with a factor of 10 or 1 $\log_{10}$. The *D*-values were estimated by fitting the Bigelow model (Bigelow, 1921) to the $\log_{10}$-counts data (mean values per time point) versus time, using the software "Bioinactivation" (Garre et al., 2017), or calculated when the maximum rate of inactivation (k_{max}) was reported (i.e., using the equation $\ln 10/k_{max}$). When inactivation kinetics were only reported as a graph, the tool "WebPlotDigitizer" (Rohatgi, 2025) was used to extract the $\log_{10}$-counts. It should be noted that the inactivation kinetics were also fitted using the Weibull model where possible (Mafart et al., 2002), but due to the limited number of observations in most of the NTP experiments (only 165/519 entries reported more than 4 observations, including the zero time point, and only 135/519 reported more than 5 observations), tail and shoulder effects could not be accurately estimated. Thus, the Weibull shape parameter estimates were not further used in this study but can be shared upon request. All $\log_{10}D$ values were standardized to $\log_{10}$ min. The $\log_{10}D$ values of the different setups found in the literature were compared, namely, ICP: Inductively coupled plasma, APPJ: Atmospheric pressure plasma jet, DBD: Dielectric barrier discharge, LPDBD: Low-pressure dielectric barrier

discharge, SBD: Surface barrier discharge, RFLPCP: Radiofrequency low-pressure cold plasma, CD: Corona discharge, LPMPS: Low-pressure microwave plasma source, PTP: Point-to-plate plasma, MPS: Microwave plasma source. The overview of the achieved reductions was assessed for five different matrix categories: i., solid food; ii., liquid food (treated on an abiotic surface); iii., solid medium; iv., liquid medium (treated on an abiotic surface) and, v., (dry) abiotic surface. The corresponding products for each category can be found in Supplementary Table 2.1.

2.2.4 The concept of dissipated power per plasma volume

In this research, the concept of dissipated power per plasma volume is proposed as a defining descriptor of NTP inactivation efficacy that can be easily estimated for DBD devices. This is because RONS concentrations, although directly associated with inactivation efficacy, are difficult to define (which species, where), measure (which device, which method, trained personnel) and report consistently (concentration, relative concentration, or spectrograph), making plasma diagnostics very complex. Additionally, the approach of RONS could be too specific or not include reactive species that could have an effect and are not measured accurately with the current methodology (e.g., short-lived reactive species). Thus, we used the power that is used for the creation of plasma (dissipated power, P_{dis}), divided by the volume of the plasma produced, i.e. power density (P_{dens}), as an integrated plasma descriptor. Dissipated power was chosen, because it is a more comprehensive parameter compared to other devices' properties as the peak-to-peak voltage and frequency, as explained above. The method to estimate the dissipated power can be found in Equation 2.1.

$$P_{\text{dis}} = U \cdot I \cdot \cos(\phi) \qquad \text{(W)} \tag{2.1}$$

where U is the voltage (in volts, V) and I is the current (in amperes, A), and ϕ (-) is the phase shift between current and voltage (Golda et al., 2016). Some modifications of this equation to make estimations easily applicable for some setups can be found in Golda et al. (2019).

For DBD devices this concept can be approximated using the dissipated power, divided by the powered electrode area and the distance between the powered and the ground electrode (Equation 2.2).

$$P_{\text{dens}} = \frac{P_{\text{dis}}}{A \cdot d} \qquad \text{(W/cm}^3\text{)} \tag{2.2}$$

where A is the powered electrode area (in cm^2) and d is the distance (in cm) between the powered and ground electrode.

It should be noted that the true plasma area can be smaller than the powered electrode's area, because the effective area in which plasma is created can be smaller than the area that is defined by the electrodes' dimensions. In the same manner, the width can be smaller than the distance between the electrodes, especially as the distance increases (Souza et al., 2016). However, these two system parameters together with the dissipated power are expected to affect most of the plasma creation and the corresponding inactivation efficacy.

2.2.5 Exploratory data analysis

An exploratory data analysis was conducted using Python programming language (v3.11) (Van Rossum and Drake, 2009) and the libraries "numpy" (Harris et al., 2020), "pandas" (McKinney, 2010), "seaborn" (Waskom, 2021), "matplotlib" (Hunter, 2007), "scikit" (Pedregosa et al., 2011), and "scipy" (Virtanen et al., 2020). Firstly, all the relevant power and energy expressions (dissipated power (W), dissipated power per plasma area (W/cm^2), dissipated power per plasma volume (W/cm^3), dissipated energy (W·min), dissipated energy per plasma area ($W{\cdot}min/cm^2$), dissipated energy per plasma volume ($W{\cdot}min/cm^3$)) were separately checked for their correlation with the $\log_{10}D$ ($\log_{10}$ min) using the Spearman's rank correlation coefficients (Supplementary Figure 2.1) to evaluate which could function as an integrated plasma descriptor. Then the other numerical and categorical variables that were included in the database for the DBD studies were evaluated, namely, the numerical variables peak-to-peak voltage (kV), frequency (kHz), relative humidity (%), product temperature (°C) after treatment, sample area (cm^2), pH (-), and the categorical variables ground electrode shape (circular plate, rectangle), ground electrode material (copper, stainless steel), upper electrode shape (circular plate, rectangle), upper electrode material (copper, stainless steel), dominant working gas (air, Ar, He, N_2, N_2/O_2), matrix category, food (yes, no), microbial genus (*Listeria*, *Salmonella*, *Escherichia*, and *Bacillus* (no data for the non-mentioned genera)), species (*monocytogenes*, *innocua*, *enterica*, *coli*, *cereus*), strain (LMG 23775, ATCC 33090, ATCC 25922, Amp-res-mutant, LMG 13305, LMG 14933, SL 1344, ATCC 13311), type of cells (vegetative, biofilm), mode of inoculation (surface, whole sample), selective recovery medium (yes, no). For the categorical variables, the Kruskal-Wallis H test (non-parametric alternative to one-way ANOVA) was used to test if there are statistically significant differences between the different groups' $\log_{10}D$ median value for each categorical variable. The Kruskal-Wallis H test provides two outputs, the H-statistic (the higher, the greater the differences) and the p-value, to see if the differences between the subgroups are statistically significant. Cramér's

V was also performed as a measure of association between the categorical variables to test if the categorical variables are associated with each other. Cramer's V provides the φc value which is the intercorrelation coefficient and can take values from 0 (no association) to 1 (perfect association). The summaries of statistics and violin plots were also used in this case to explore potential parameters affecting the $\log_{10}D$.

2.2.6 Quantitative modeling

2.2.6.1 Data splitting and scaling of numerical variables

Unless mentioned differently, the data were first split into a training and a test set. This step was implemented before scaling to avoid data leakage (from the training to the test set during the scaling step) (Lones, 2021). The test set was set to 20% of the full dataset (n_{training} = 128, n_{test} = 32). The training and test sets were selected randomly using the "random_state" function (pseudo-random numbers) of the "scikit" library (Pedregosa et al., 2011). The numerical variables of the training set were then scaled using the standard score (z) of a sample value (x) with mean (u) and standard deviation (s), and the estimated mean and standard deviation of each of the variables were used for scaling also the test data (Equation 2.3).

$$z = \frac{(x - u)}{s} \tag{2.3}$$

Subtracting the mean from a sample value typically improves the interpretation of main effects in the presence of interactions, while dividing by the standard deviation puts all variables on a common scale (Gelman, 2008). After scaling, the variable coefficients indicate the expected difference in the dependent variable for each standard deviation change in the variable's value. This was done because the scale i.e., the range of values of the numerical variables was varying significantly. For example, the dissipated power per plasma volume (W/cm^3) values ranged from 0.02 W/cm^3 to 2 W/cm^3, whereas the peak-to-peak voltage (V) ranged from 4.5 kV to 25 kV, and the relative humidity from 0% to 80%. With scaling the statistical significance of the variables does not change i.e., the p-values of the predictors' set will remain the same, but the interpretation and comparison of the relative importance (comparing the coefficients instead of their p-value) is improved (Schielzeth, 2010).

2.2.6.2 Imputation strategies

A selected set of values for some studies was imputed so as not to have missing values for the modeling part. For the relative humidity and pH, the imputation strategy was quite straightforward. For the cases with missing information, the relative humidity was assumed to be 50% (the average of 40-60% which is assumed to be a common relative humidity range at room temperature), whilst the pH of abiotic surfaces was assumed to be 7 (including the eggshells' surface). The pH of Trypticase Soy Broth with Yeast Extract was assumed to be 7.3 (U.S. Food and Drug Administration, 1998), with or without the addition of xanthan gum. For the sample area, a more complex approach was implemented. The approach was to use *k*-nearest neighbor (*k*-NN). The *k*-NN is a method that tries to classify an unknown sample based on the known classification of its neighbors (Mucherino et al., 2009). In our case, to predict the missing values of the sample area we used the matrix category, the sample size, and the research team (on the assumption that research teams often work with the same product dimensions based on their equipment) neighbor values to estimate the non-reported sample area based on the value of the sample area that the neighbor values of the aforementioned parameters were pointing to. The number of neighbors to be used was set at 8. This is an arbitrary choice that in our case was based on the density plot that compared the distribution of the original values with the imputed ones. This number of neighbors yielded the best overlap between the two distributions (Figure 2.3).

2.2.6.3 Multiple linear regression and regularization techniques

The approach of multiple linear regression (MLR) was implemented due to its simplicity and its high interpretability. This approach was performed to the full dataset (n = 160), as well as to the training dataset used for the machine learning approaches (80% i.e., n_{train} = 128), to compare the different methods. Ordinary least squares regression was used for this step using backward elimination with a threshold of p-value = 0.05. The dependent variable was the $\log_{10}D$ ($\log_{10}$ min) and the potential predictors were, as determined by the exploratory data analysis, the dissipated power per plasma volume (W/cm^3), the peak-to-peak voltage (kV), the relative humidity (%), the sample area (cm^2), the pH (-), the upper electrode shape (circular plate, rectangle), the upper electrode material (copper, stainless steel), the matrix category (solid medium, liquid medium, abiotic surface), and the microbial genus (*Listeria*, *Salmonella*, *Escherichia*). The categories that are not mentioned here, i.e., solid and liquid foods, *Saccharomyces cerevisiae* and yeasts, lactic acid bacteria, and coliforms were not included in the dataset used for the regression because they could not meet the criteria set i.e., DBD setup, achieved microbial inactivation > 0.5-$\log_{10}$-reduction, and reported dissipated power per plasma volume (W/cm^3) parameters. For *Bacillus cereus*, there

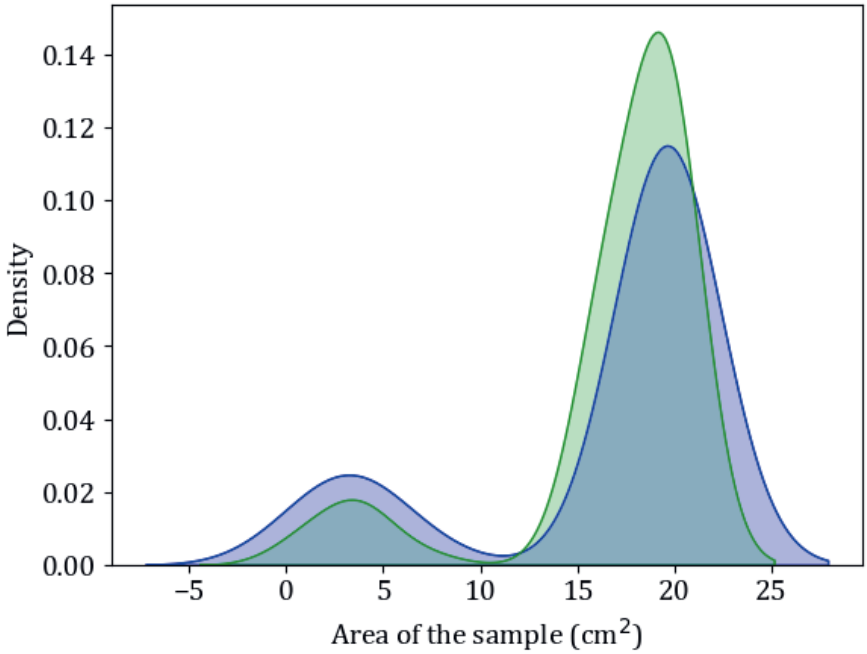


Figure 2.3: The density results of the imputation strategy applied to the sample area. The density of the reported sample area is colored with purple, whereas the imputed one is colored with green.

was only one entry (additional to the 160) that met the criteria and thus it was excluded. The categorical variables were converted to dummy variables (0, 1) and the first category of each variable, which was most often reported in the studies, was set as the base to avoid the dummy variable trap which would lead to multicollinearity between the dummy variables (Suits, 1957), although the effect of the base selection for each category was evaluated. In an attempt to further improve variable selection while retaining interpretability, regularization techniques were applied, as a different approach to backward elimination. Regularization is a technique used in statistics to help deal with multicollinearity and predictors' selection, through two main steps: parameter approximations' construction, and approximation selection based on bias-variance trade-off (Bickel et al., 2006), instead of relying on the p-value of each independent variable as it is done with backward elimination. Thus, lasso (Tibshirani, 1996), ridge (Hoerl and Kennard, 1970), and elastic net (Zou and Hastie, 2005) were used for their ability to perform regularization L1 (Supplementary Material C, equation S2.1), L2 (Supplementary Material C, equation S2.2), or both (Supplementary Material C, equation S2.3), respectively. In general, ridge performs better than lasso for datasets with a substantial number of (collinear) variables, but of the two only lasso can

shrink some variables' coefficients to 0, and thus perform variable selection (Neto et al., 2014), and elastic net combines both approaches. As with every machine learning concept, there are two different parameter categories involved i.e., the ones that can be initialized and updated through the data learning process (i.e., model parameters) and the ones that cannot be directly estimated from data learning and must be set before training because they define the model architecture (i.e., hyperparameters) (Yang and Shami, 2020). The model parameters in this case are the intercept and the coefficients of the multiple linear regression model, while the hyperparameters are constants that control the regularization penalties from which the model architecture is defined (the starting point from which the "machine" starts to "learn") and the iterative learning process starts to minimize the least squares penalty. The hyperparameter selection tuning is an arbitrary choice that involves trial-error simulations, using the training set. In order to make this approach more systematic the hyperparameter tuning was conducted using Optuna which is a software that uses sampling algorithms to find more promising hyperparameters based on past trial results and pruning algorithms to eliminate the less promising ones (Akiba et al., 2019). In our case, we split the training set into 5 and in each iteration, the four parts were used to train the model and the one was used to validate it. The procedure using Optuna continues for 5000 iterations to find the optimal combination that yields the best model fit with the lowest Root Mean Square Error (RMSE).

2.2.6.4 Machine learning approaches

To unveil more complex relationships between the dependent and the independent variables, four machine learning algorithms were tested, namely, random forests, extreme gradient boosting, support vector machines, and artificial neural networks. Random forests are a combination of tree predictors such that each decision tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001). Each decision tree consists of the root node, internal nodes, leaf nodes, and branches. The root node is the initial decision after which the data is split. The internal nodes are the points where the data is split further, and the leaf nodes are the nodes where the data cannot be split any further. The branches are the links between the nodes that represent the outcome of each test. The hyperparameters that were chosen for tuning for our case were the number of trees in the forest, the maximum depth of each tree, the minimum number of samples required to split an internal node, and the minimum number of samples required to be a leaf node. Gradient boosting, as random forests, is also a form of ensemble learning i.e., using a set of hypotheses instead of one to choose a set of weights to construct the "voted" classifier (or regressor) (Dietterich, 2002). This means that also for this case we compose our model based on multiple decision trees. The difference is

that the approach of gradient boosting is sequential and can formulate additive expansions for any fitting criterion (Friedman, 2001). Extreme gradient boosting (XGBoost) algorithm was used with tree booster, where each new tree attempts to correct the errors made by the previous ones. The hyperparameters that were chosen for tuning were the number of gradient-boosting trees, the maximum depth of each tree (from the root to the leaf node), the learning rate (shrinkage factor) that scales the contribution of each tree corrections to the next one to avoid overfitting, and the subsample ratio of the instances used for training before growing each tree that also helps to avoid overfitting (Chen and Guestrin, 2016). Support vector machines (SVM) use the SV algorithm that regulates the capacity by considering the function corresponding to the hyperplane that separates, according to the labels, the given training data and it is maximally distant from them (Boser et al., 1992; Cristianini and Ricci, 2016; Hearst et al., 1998). As such, SVM tries to find a decision boundary (i.e., hyperplane) that separates distinctly different data points into categories. It focuses on maximizing the distance (margin) between the data points of each category and the decision boundary, ensuring a clear separation of the data points into categories. The hyperparameters that were chosen for tuning were the regularization parameter (C), the epsilon tube, and the kernel coefficient. Regularization parameter C controls the balance between minimizing the misclassification and maximizing the margin between categories. The epsilon tube refers to a margin of tolerance within which some prediction errors are not penalized, to avoid overfitting i.e., a larger tube allows more errors, and a lower forces to a higher sensitivity which may lead to overfitting (Chang and Lin, 2011). Lastly, the kernel coefficient chooses the way to transform the data in higher dimensional space to make it possible to perform linear separation, when the original data are not linearly separable. Artificial neural networks simulate networks of multiple interconnected building blocks called model neurons. Each model neuron receives input data, assigns a weight factor to each input, and then sums the weighted inputs. This sum is then passed through an activation function which converts it to a modified output. For instance, an activation function can be used to introduce non-linearity to the model enabling the network to learn complex patterns in the data. The output from the activation function is then passed to the next layer and the process is repeated for multiple layers and data combinations. Back-propagation is implemented in every pass from the input to the output data to correct the weight factors in each layer. This is done based on the loss function change concerning each weight value in the network (Krogh, 2008). In this study, the number of model neurons in the input layer corresponded to the number of parameters of the database, and each row of the database constituted the input data. Each model neuron in one layer was connected to every model neuron in the next layer. This connection between the layers was made possible by assigning different weight factors for each parameter's value, summing up the weighted values for all parameters, and passing them to the next layer through an

activation function. The activation function used was the activation function rectifier linear activation which converts the non-positive sums of the inputs to 0 and the positive ones to their actual value, essentially creating a piecewise linear function with two linear pieces (Goodfellow et al., 2016). The loss function that was used in this study for the back-propagation step was the mean square error. The hyperparameters that were chosen for tuning were the number of hidden layers (the layers between the input and the output layers), the number of model neurons on each hidden layer, the dropout rate which selects a fraction of the model neurons not to be included in every pass (to avoid overfitting), and the learning rate which controls how much the weight factors change in each pass. All the above machine learning approaches have a lot more hyperparameters that can be tuned for better results, but given the limited number of data points (n = 160), this was unnecessary. The hyperparameters that were not tuned, were set to their default values, as determined by the "Scikit-learn" software that was used for random forests, XGBoost, and SVM (Pedregosa et al., 2011). For neural networks, the "TensorFlow" software was used (Abadi et al., 2015). For the hyperparameters that were tuned, Optuna (Akiba et al., 2019) was used for all machine learning approaches, as explained in Section 2.2.6.3. For the interpretation of the parameters that contribute the most to the machine learning models, the Shapley Additive exPlanations (SHAP) approach was followed and the Shapley values were estimated for each independent variable. The SHAP framework is based on the game theory and identifies the class of additive feature importance methods and shows that there is a unique solution in this class that adheres to desirable properties (Lundberg and Lee, 2017). In a typical SHAP summary plot, each dot represents the SHAP value for each parameter for each data point of the test set which means that the number of dots matches the number of the data points in the test set. The SHAP value scale (x-axis) is relative and centered around 0 and the dots that are on the right side push the dependent variable to positive correlations with the independent variable (higher $\log_{10}D$), whereas the dots on the left push it to the negative correlations with the independent variable (lower $\log_{10}D$). The variables' colors are based on the actual value of each parameter i.e., when the color is leaning to red then this corresponds to the data points that have higher values of this parameter, while blue corresponds to lower values of this parameter. If the red dots of a parameter are scattered on the left side and the blue dots are scattered on the right side this could imply a negative correlation of the independent variable with the dependent variable, whereas when the blue dots of a parameter are on the left side, and the red dots are on the right side, this could imply a positive correlation. If the red or blue data points or both are scattered on both sides, then this could imply that the correlation exists but the positive or negative effect is condition-dependent. If both dots are centered around 0 (low mean absolute SHAP value) this could imply that the parameter has low or no contribution to the model's output. For dummy variables, there are only two possible outcomes (colors),

where red corresponds to the presence of the effect, and blue to the absence of the effect (base), but their interpretation is analogous; for instance, red dots on the left side could imply a negative correlation, and so forth.

2.3 Results

2.3.1 Overview of *D*-value entries for all setups

The $\log_{10}D$ values obtained for the different NTP setups are shown in Figure 2.4 and are in line with Baier et al. (2013), who reported that there is significant variation between the different types of plasma sources and their mode of application. Our results showed that the APPJ was the setup with the lowest median *D*-value (0.5 minutes), and LPDBD had the highest median *D*-value (15 minutes). Differences in the extent of variability also exist, for instance, in the $\log_{10}D$ of the LPDBD setup where the variability was higher than LPMPS despite the more data points of the latter. It should be noted that the reliability of these estimates is not the same for all setups since the number of data points differs for the different plasma setups (Figure 2.4).

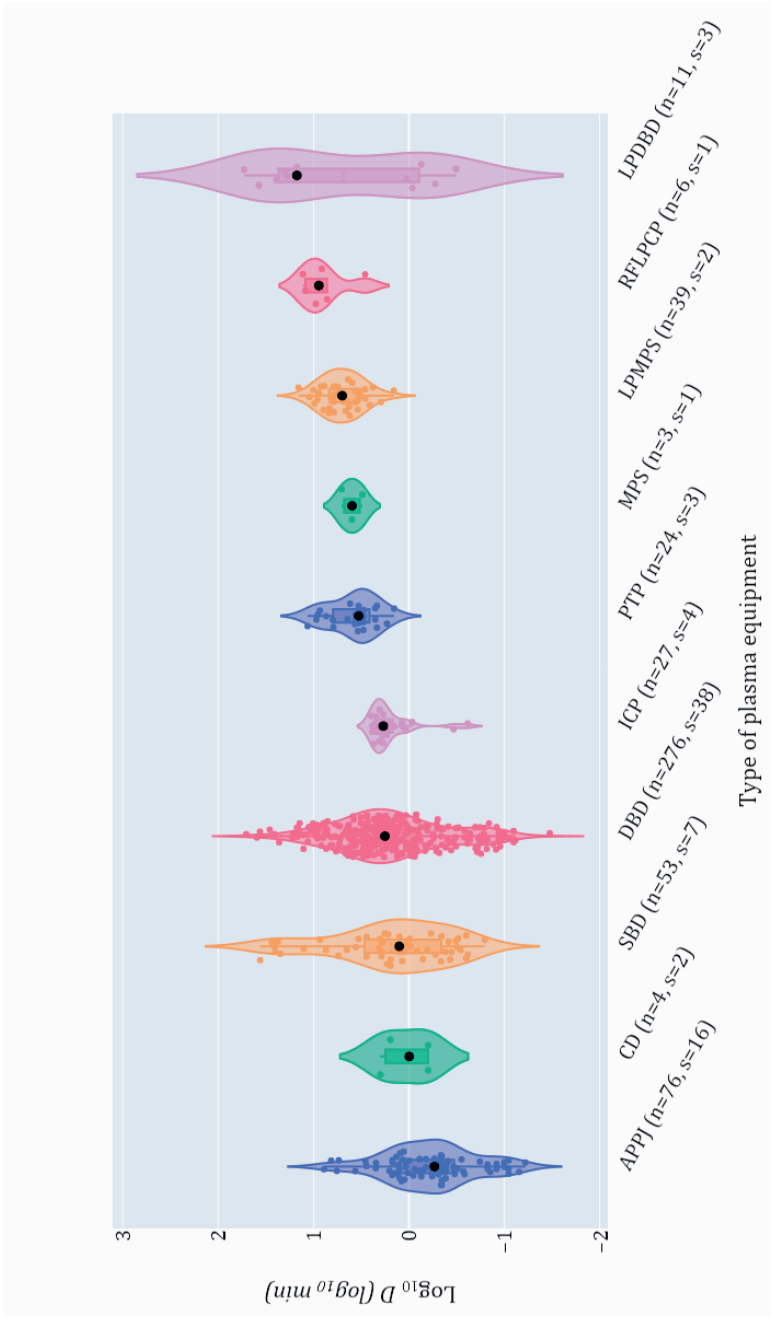


Figure 2.4: Overview of all *D*-values obtained for different plasma setups. APPJ: Atmospheric pressure plasma jet (median *D*-value = 0.54 min), CD: Corona discharge (median *D*-value = 1.0 min), SBD: Surface barrier discharge (median *D*-value = 1.3 min), DBD: Dielectric barrier discharge (median *D*-value = 1.8 min), ICP: Inductively coupled plasma (median *D*-value = 1.9 min), PTP: Point-to-plate plasma (median *D*-value = 3.4 min), MPS: Microwave plasma source (median *D*-value = 3.9 min), LPMPS: Low-pressure microwave plasma source (median *D*-value = 5.0 min), RFLPCP: Radiofrequency low-pressure cold plasma (median *D*-value = 8.8 min), LPDBD: Low-pressure Dielectric barrier discharge (median *D*-value = 14.9 min). The medians of each group are represented with a black dot. The n represents the number of *D*-values for each type of equipment and the s represents the corresponding number of studies. The violin plots are ordered based on the median $\log_{10} D$ for each type of equipment.

The overview of the microbial reductions achieved with NTP against various microorganisms and matrices is shown in Table 2.1. For lactic acid bacteria, there were no studies in the literature that met the eligibility criteria (information on the type of plasma device, achieved microbial inactivation > 0.5-$\log_{10}$-reduction, and reported power expression). This table can serve as a map of what decontamination is to be expected in general for most NTP setups. A huge variation between the different combinations of matrix-microorganism can be observed with the median reduction using NTP ranging from 1.1 to 7.9 $\log_{10}$. This variation entails the effect of the setup as well as other effects such as the specific matrix (e.g., the particular food product), the power, and the processing time used, but it is informative in terms of answering simple questions regarding NTP decontamination efficacy according to the current literature, without specific criteria. It should be noted that these values were most probably an underestimation of the efficacy since in experiments where NTP processing led to microbial counts below the detection limit, the previous time-point was used. However, considering also the inactivation entries that led to microbial counts below the detection limit using only the entries with an initial inoculum of ≥ 5 $\log_{10}$, 20.1% (71 of 352) of the total entries led to an inactivation ≥ 5 $\log_{10}$ or below the detection limit with entries for all microorganisms and matrix categories and with DBD, APPJ, ICP, PTP and SBD setups (data not shown).

Table 2.1: Overview of microbial $\log_{10}$-reductions across different matrix categories

Matrix category	*Listeria* spp.[a], [b]	*Salmonella* spp.[a], [b]	*Bacillus* spp.[a], [c]	Coliforms[a]	Yeasts[a]
solid food	1.7, 2.0±1.3 ($n = 39$)	7.9, 7.9 ($n = 1$)	1.7, 2.2±1.3 ($n = 12$)	1.5, 1.9±1.3 ($n = 67$)	2.4, 2.4±1.0 ($n = 6$)
liquid food	5.4, 4.4±1.7 ($n = 3$)	-	5.2, 4.7±0.9 ($n = 3$)	3.4, 3.2±1.1 ($n = 9$)	1.1, 1.1±0.7 ($n = 2$)
solid medium	1.7, 2.0±1.5 ($n = 25$)	1.5, 1.6±0.6 ($n = 14$)	-	3.5, 3.4±1.2 ($n = 22$)	-
liquid medium	1.4, 2.6±2.7 ($n = 18$)	2.2, 2.0±0.7 ($n = 14$)	3.6, 3.4±1.5 ($n = 14$)	4.5, 4.4±1.4 ($n = 40$)	3.9, 3.7±2.4 ($n = 24$)
abiotic surface	1.9, 2.1±1.0 ($n = 69$)	2.2, 2.4±1.0 ($n = 51$)	4.0, 3.9±1.2 ($n = 35$)	2.7, 2.5±1.3 ($n = 46$)	2.6, 3.1±1.0 ($n = 3$)
Total	154	80	64	184	35

[a]: The values correspond to the reductions above the detection limit and are reported as median, mean ± SD

[b]: For *Listeria* spp. and *Salmonella* spp. both vegetative and biofilm cells are reported

[c]: For *Bacillus* spp. both vegetative and spore cells are reported

From a modeling perspective, the most important issue in assessing the sources of variability within each setup was the lack of a parameter present in all studies, that is equivalent for example to temperature for thermal processing or pressure for high-pressure processing to standardize the results. We focused on the DBD setup because

according to the findings of this study (Figure 2.4) and previous research (Asl et al., 2022; Liao et al., 2017), it is the most common setup with the most data reported on. Atmospheric pressure plasma jet (APPJ) was the second most frequent setup in our database. It should be noted that DBD offers a short but wide discharge in which the sample is part of it, whereas the APPJ can treat a sample from a longer distance (the sample is not part of the discharge), but in a smaller area (Yan et al., 2016). DBD devices can produce stable and uniform NTP while working at relatively low temperatures under atmospheric pressure (Liao et al., 2017). Additionally, for the DBD setup, it was easier to estimate the plasma volume and thus the dissipated power per plasma volume (W/cm^3). This concept is not so easily transferable to APPJ or other setups.

2.3.2 Exploratory data analysis for DBD

2.3.2.1 Power expression selection and evaluation

Moving to the DBD setup, 160 *D*-values were estimated that reported all the required power density equation components (dissipated power, electrodes' dimensions, distance between the electrodes) (Equation 2.2) concerning *Listeria* spp., *Salmonella enterica*, *Escherichia coli*. For yeasts, lactic acid bacteria, and coliforms (different than *E. coli*) no entries were reporting the aforementioned parameters. The entries for *Listeria* spp., *Salmonella enterica*, and *Escherichia coli* were 70, 79, and 11, respectively. The corresponding products for each category can be found in Supplementary Table 2.2. The exploratory data analysis indicated that from the power- and energy-related expressions, the most negatively correlated (with the $\log_{10}D$ ($\log_{10}$ min)) were the dissipated power per plasma volume (W/cm^3) and the dissipated power per plasma area (W/cm^2) with the highest Spearman's rank correlation coefficients (Supplementary Figure 2.1). The Spearman's rank correlation coefficient was -0.16 for both expressions which was rather low but still the highest among the power expressions. From the two expressions, the dissipated power per plasma volume (W/cm^3) was selected, because it incorporates one more parameter that is crucial for DBD i.e., the electrodes' distance, in a rational and parsimonious way (Liu et al., 2004; Souza et al., 2016).

2.3.2.2 Variable selection

Based on the exploratory data analysis, from the numerical variables, the product temperature (°C) was excluded based on the scatterplot with the $\log_{10}D$ ($\log_{10}$ min) and the low Spearman's rank correlation coefficient (Supplementary Figure 2.2). The peak-to-peak voltage was considered for the next steps even though it showed a low Spearman's rank correlation coefficient, because of a slight trend in its scatterplot with the

$\log_{10}D$, whereas the frequency (kHz) was not considered because no trend could be observed in its scatterplot (Supplementary Figure 2.3). The dissipated power per plasma volume (W/cm^3), sample area (cm^2), relative humidity (%), peak-to-peak voltage (kV), and pH (-) were considered for the imputation and modeling steps. It should be noted that the RONS, although important, were not reported in any of these entries, neither as a whole nor as individual chemical compounds. Moving to the categorical variables, from the Kruskal-Wallis H-test and the violin plots, dominant working gas, type of cells, food (yes, no), and selective recovery medium (yes, no) did not seem to have a significant influence (both statistically and visually) on the $\log_{10}D$ ($\log_{10}$ min). From the results of Cramér's V, the shape and material of the ground and upper electrodes were intercorrelated with intercorrelation coefficient $\varphi c > 0.95$ respectively, hence, the upper electrode shape and the upper electrode material were selected as categorical variables, because the upper electrode is also the powered one. The microbial genus was intercorrelated with the species ($\varphi c = 0.99$) and the strain ($\varphi c = 0.98$) and thus only the microbial genus was used because it has fewer categories (i.e., fewer dummy variables) and thus sets a more parsimonious basis. The mode of inoculation was intercorrelated with the matrix category ($\varphi c = 0.99$) and thus only the matrix category was used. To sum up, from the numerical variables, dissipated power per plasma volume (W/cm^3), sample area (cm^2), relative humidity (%), peak-to-peak voltage (kV), and pH were considered, and from the categorical variables, upper electrode shape, upper electrode material, matrix category, and microbial genus were considered as potential predictors for the next steps, respectively. These potential predictors were selected to be imputed where needed so as not to have missing values for the modeling part. From the numerical variables, sample area, relative humidity, and pH had 77, 155, and 84 missing values, respectively. From the categorical variables, only the upper electrode material had 149 missing values but was not imputed, and therefore excluded from the next steps, because no possible strategy could be established. All other variables were imputed as explained in Section 2.2.6.2.

Multiple linear regression was performed for the parameters power density (W/cm^3), sample area (cm^2), upper electrode shape, matrix category, pH, microbial genus, peak-to-peak voltage (kV), and relative humidity (%). The upper electrode shape of the circular plate (one of the two categories), the matrix category of the abiotic surface (one of the three categories), and the microbial genus of *Salmonella* (one of the three categories) were the most frequent categories and were set as the base for all regression analyses. The categorical variables with three categories (microbial species, matrix category) were evaluated per category, and the effect of base selection on the significance evaluation of the variables was checked through separate trials. The order of the variable coefficients changed as a result of scaling all numerical variable values using Equation 2.3 (Table 2.2), which underlined the need for scaling for better insights into the relative importance based on the coefficients. The mean and the standard devi-

ation used for scaling the numerical variables can be found in Supplementary Table 2.3. Using all data points (n = 160) (Figure 2.5) and backward elimination using scaled variables, the peak-to-peak voltage, and the relative humidity were excluded as non-significant variables (Table 2.2). It should be noted that the *Bacillus* entry that was excluded because it was the only one for this microbial genus (although it matched the eligibility criteria, see Section 2.2.6.3) had the same power density level as the *E. coli* at 2 W/cm^3 (Figure 2.5). In addition, the model was tested by excluding the *E. coli* entry at 2 W/cm^3 and the effect of power density was still prominent. The results suggested that the upper electrode shape, the matrix category, the sample area, and the power density significantly affected the $\log_{10}D$, next to the pH and the microbial genus, with the latter two variables being just significant. The microbial genus of *Listeria* was significantly different from *Salmonella*, while *Escherichia* was not significant for the model's output. For the matrix category, the matrix category of the liquid medium was significantly different from both the abiotic surface and the solid medium, whereas the abiotic surface and the solid medium were not significantly different from each other.

Table 2.2: The effect of scaling in the order of the coefficients using all data points ($n = 160$) and multiple linear regression with backward elimination. The coefficients are ordered based on the absolute coefficient values with scaling. The R^2 of the multiple linear regression model with or without scaling was 0.342

Parameters	Coefficients with scaling	SE	Coefficients without scaling	SE	p-value
Upper electrode shape of rectangle	-1.31	0.36	-1.31	0.36	< 0.001
Matrix category of liquid medium	0.47	0.10	0.47	0.10	< 0.001
Sample area (cm^2)	-0.34	0.09	-0.07	0.02	< 0.001
Power density (W/cm^3)	-0.31	0.05	-1.52	0.23	< 0.001
Microbial genus of *Listeria*	-0.16	0.07	-0.16	0.07	0.022
pH	-0.07	0.03	-0.12	0.06	0.033
Peak-to-peak voltage (kV)	-	-	-	-	> 0.05
Matrix category of solid medium	-	-	-	-	> 0.05
Relative humidity (%)	-	-	-	-	> 0.05
Microbial genus of *Escherichia*	-	-	-	-	> 0.05
Intercept	0.33	0.06	2.89	0.58	< 0.001

When we split the data, to consistently compare the multiple linear regression results with the machine learning approaches, in training and test i.e., 80:20 (n_{train} = 128, n_{test} = 32), then the coefficients with scaling were slightly different but with the same order as with the full dataset (Figure 2.6). Then, elastic net was selected to perform the multiple linear regression with regularization due to its ability to penalize the variables' coefficients (reduce their value) in case of multicollinearity and even shrink them to zero (exclusion of the variable from the model), in case needed. This approach is more beneficial for datasets with a lot of variables but not so many observations and deals with the predictor selection in a different way than using the p-values of the coefficients

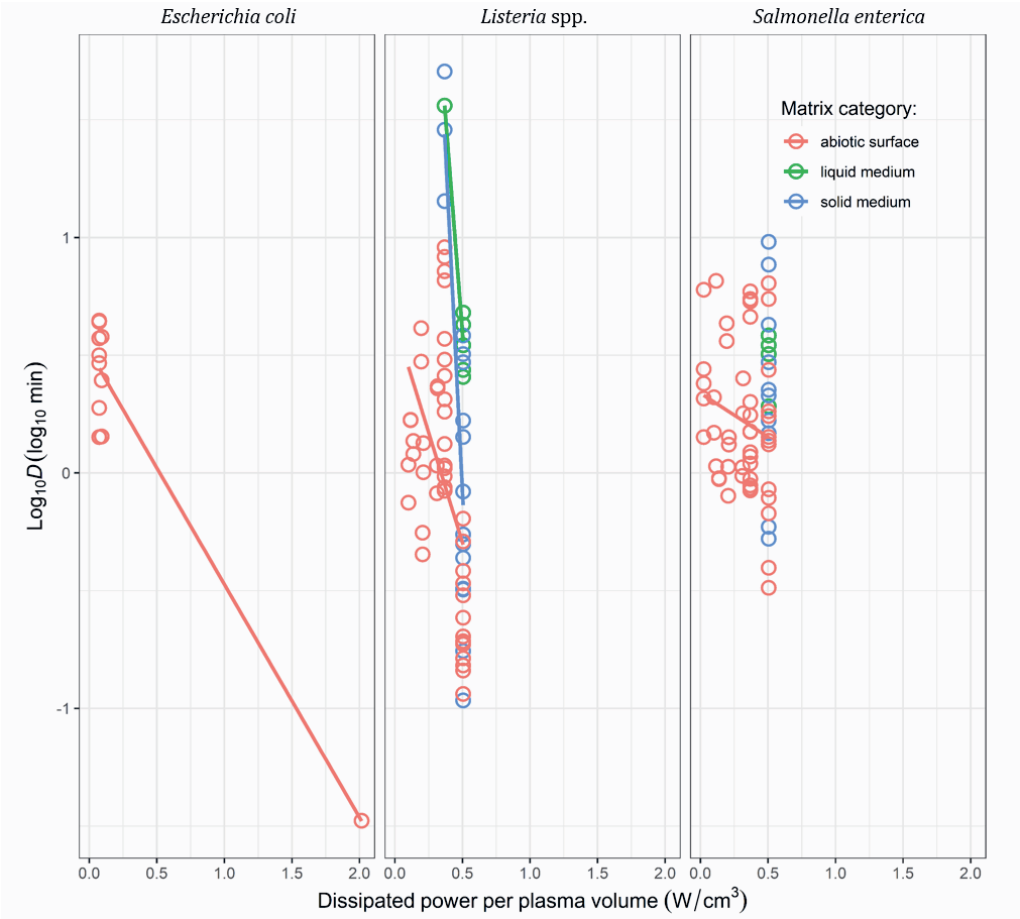


Figure 2.5: Overview of the $\log_{10}D$ values of the DBD setup that was used for modeling purposes. The regression lines of the $\log_{10}D$ ($\log_{10}$ min) are fitted as a linear function of dissipated power per plasma volume (W/cm^3) for each matrix category.

in the elimination procedure (backward elimination). For the optimal combination of the hyperparameter α and ρ, Optuna was used as explained in Section 2.2.6.3. The optimal hyperparameter α and ρ were found 0.0005 and 0.00002, respectively. The low α indicated that a minimal regularization was applied, while the low ρ indicated that from the two components of the elastic net regularization equation (i.e., lasso and ridge), the ridge component was mostly used (Supplementary Material C, equation S2.3). All variables were selected from the regression as predictors and none of them was shrunk to 0, and the R^2 on the training and test dataset were 0.40 and 0.07, respectively while the RMSE on the test dataset was 0.46 $\log_{10}$ min. Since the elastic net did not improve the model's performance much as the R^2 only slightly increased

(from 0.38 (for the training dataset) to 0.40), we used multiple linear regression with backward elimination (with scaling and data split in training and test) as the basis for comparing the multiple linear regression with the machine learning approaches.

Parameters	Coefficients	SE	p-value
Upper electrode shape of rectangle	-1.57	0.39	< 0.001
Matrix category of liquid medium	0.46	0.11	< 0.001
Sample area (cm^2)	-0.36	0.09	< 0.001
Power density (W/cm^3)	-0.34	0.05	< 0.001
Microbial genus of *Listeria*	-0.20	0.07	0.007
pH	-0.10	0.04	0.005
Peak-to-peak voltage (kV)	-	-	> 0.05
Matrix category of solid medium	-	-	> 0.05
Relative humidity (%)	-	-	> 0.05
Microbial genus of *Escherichia*	-	-	> 0.05
Intercept	0.37	0.06	< 0.001

Parameters	SHAP
Power density (W/cm^3)	0.19
Microbial genus of *Listeria*	0.12
Sample area (cm^2)	0.11
pH	0.11
Matrix category of liquid medium	0.08
Peak-to-peak voltage (kV)	0.07
Matrix category of solid medium	0.02
Relative humidity (%)	< 0.01
Microbial genus of *Escherichia*	< 0.01
Upper electrode shape of rectangle	< 0.01

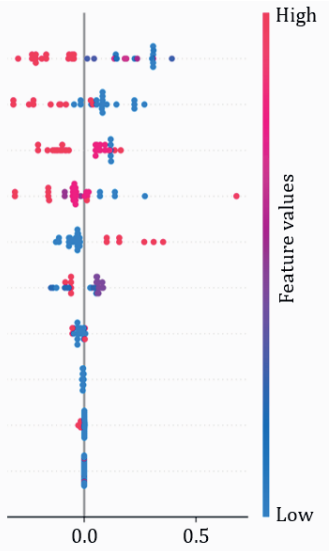


Figure 2.6: The output of multiple linear regression and the SHAP values of the XGBoost model using the scaled dataset split in training and test. In multiple linear regression, the parameters with p-value > 0.05 were excluded from the model (model R^2 was 0.384). In the XGBoost model that is visualized through the SHAP values, each dot represents the SHAP value for each parameter for each data point of the test set. The SHAP value scale is relative and centered around 0 and the dots that are on the right side push the $\log_{10}D$ ($\log_{10}$ min) to higher values, whereas the dots on the left push it to the lower values. The parameters are ordered based on the mean absolute SHAP value. The parameters' colors are based on the actual value of each parameter i.e., when the colour is leaning to red then this corresponds to the data points that have higher values of this parameter, while blue corresponds to lower values of this parameter. For dummy variables, red corresponds to the presence of the effect, and blue to the absence of the effect.

2.3.3 Machine learning approaches

To unveil more complex relationships between the dependent and the independent variables, random forests, eXtreme Gradient Boosting (XGBoost), support vector machines (SVM), and neural networks were used. Based on cross-validation in the training set, Optuna used 5000 iterations to determine the combination of hyperparameters with the lowest RMSE. The best-fitted model was with the use of the XGBoost algorithm (R^2 on test dataset = 0.71) followed closely by random forests (R^2 on test dataset = 0.70) and then SVM (R^2 on test dataset = 0.65), and neural networks (R^2 on test dataset = 0.34). Therefore, the XGBoost model was used for further comparison with multiple linear regression. The corresponding hyperparameters for XGBoost were: number of estimators (the number of gradient-boosted trees) = 69, maximum depth (the maximum depth of each tree) = 3, learning rate (the shrinkage factor of new trees added to the model) = 0.29, and subsample (the fraction of samples to be used for fitting each tree) = 0.63. The test dataset was used for interpretations in all machine learning approaches. Based on the mean absolute SHAP values in the test dataset (Figure 2.6), the power density, the sample area, and the matrix category of liquid medium were in the top five parameters with the highest effect on the $\log_{10}D$, which were also the highly significant parameters of the multiple linear regression. The pH and the microbial genus of *Listeria* which were just significant with the multiple linear regression, were the remaining two of the top five based on the mean absolute SHAP values of the XGBoost model. The microbial genus of *Listeria* had lower $\log_{10}D$ values than *Salmonella*, as indicated by the majority of the red data points scattered on the left side, and the blue data points on the right side of the SHAP figure. The microbial genus of *Escherichia* had a minimal contribution to the model with all data points scattered around zero. For the matrix category of liquid medium, the red dots on the right side and the blue data points on the left side indicated that the matrix category of liquid medium had higher $\log_{10}D$ values than the abiotic surface (in accordance with the positive coefficient of multiple linear regression), while the matrix category of solid medium had a minimal contribution to the model with all points scattered on the left side and close to zero. The insignificant variables of the multiple linear regression were also the variables with the lowest mean absolute SHAP values of the XGBoost model i.e., peak-to-peak voltage (kV), and relative humidity (Figure 2.6). The direction of the effect on the efficacy (positive or negative) was also consistent (i.e., the negative correlation coefficients of multiple linear regression matched with the red dots on the left side of the SHAP figure) except for the effect of sample area, where the negative coefficient of multiple linear regression was "translated" to red dots in both sides of the SHAP figure, pushing the model to either a positive or negative direction depending on the context which occurs, and not in a uniform way in the whole dataset. The main difference between the two methods was regarding the effect of the

upper electrode shape which was highly significant based on multiple linear regression analysis, but with almost zero contribution to the XGBoost model. This could be due to the huge differences in the frequency of the two shapes i.e., 145 data points for the circular plate shape and 15 data points for the rectangle shape, which can induce uncertainty in the parameter estimate (Supplementary Figure 2.4). Using the hyperparameters determined with Optuna, a reduced model with only the 5 parameters with the highest mean absolute SHAP values, i.e., power density, microbial genus of *Listeria*, sample area, pH, and the matrix category of liquid medium, was tested. In this case, the power density was the most important factor with a mean absolute SHAP value of 0.2, whereas the rest of the parameters were scattered around 0.1. The direction of the effect on the efficacy was the same as with the full model. Combining the results of MLR and XGBoost, it can be concluded that power density and matrix category are important parameters affecting the $\log_{10}D$, followed by microbial genus and pH, while peak-to-peak voltage and relative humidity have lower importance. The impact of sample area and electrode shape could not be well defined based on the current data set. The fact that the XGBoost SHAP values were estimated from the test data, while the linear coefficients were obtained based on the training data, and still align to mostly the same ranking of parameters makes the conclusions more robust.

2.4 Discussion

Active learning demonstrated to be a useful tool for making the literature screening more efficient, in terms of time but also for exploring a broad domain where keywords cannot make the search specific enough (e.g., plasma). The selection of the stopping criterion for this study was based on pragmatic criteria that fit the purpose of having the overview of the technology's literature before narrowing it down to the DBD setup (Callaghan and Müller-Hansen, 2020). However, a more complex approach for its optimal selection such as statistical stopping criteria should also be considered for fully systematic literature reviews (Callaghan and Müller-Hansen, 2020).

The different setups of NTP showed huge variations in the decontamination efficacy. As mentioned in previous research, different plasma systems generate different plasmas with different reactive plasma cocktails that in general cannot be compared (Niquet et al., 2018; Shimizu et al., 2017). Based on the data collected, the top 50% of setups (i.e., 5 out of 10) in terms of decontamination efficacy were APPJ, CD, SBD, DBD, and ICP with a median *D*-value below 2 minutes for all cases. The lowest 50% had median *D*-values higher than 3 minutes for all cases reaching up to 15 minutes for the case of LPDBD (Figure 2.4). As mentioned, some parameters and the lack of data for some setups could change this outcome but at the same time, our research during the first step

was performed with simple criteria and without restrictions to any particular setup.

The dissipated power was selected as an accurate enough basis since RONS are rarely reported in the literature and the results are not easily comparable. Not only was power rarely reported in the literature, but the terminology was also not consistent in the studies, because power was reported as applied, forward, reverse, reflected, active, apparent, operating, generating, discharge, dissociated, transferred, input, or dissipated power and also their mean, min, and max values or terms like power density without further explanation. Hence, dissipated power was used for consistency reasons, because it is correlated to the inactivation efficacy according to the literature (Gabriel et al., 2016; Govaert et al., 2019) and was more often reported than the other variants e.g., discharge. It should be noted that in most of the studies, the power that is dissipated to the plasma is the same for different working gas compositions, for instance, Georgescu et al. (2017) pointed out that power dissipation in the plasma gas was the same for different working gas compositions (Georgescu et al., 2017). However, Govaert et al. (2019) reported a disparity in the power that is dissipated to the plasma when using different working gas compositions (Govaert et al., 2019).

The concept of dissipated power per plasma volume that is established in the current study, although easily approximated for DBD devices using the dissipated power, electrodes' dimensions, and the distance between them, can be more challenging to apply for other setups such as plasma jets, where the estimation of the plasma volume is more complex to be approximated. According to most of the existing literature, the power that is applied to the gas can predict the production of RONS. However, due to inconsistencies in reporting and the difficulties of measuring the RONS, this correlation could not be quantified in the current study. From our results, the dissipated power per plasma volume seems to have an important effect on the decontamination efficacy of NTP with the DBD setup.

The majority of studies in the literature report an increase in the decontamination efficacy of NTP through the increase in power density. However, there are a few studies that determine a decrease in decontamination efficacy with increasing power, through the decrease of ozone species, in favor of other species that are less effective (Bauer et al., 2017; Pavlovich et al., 2013). These studies suggest that there should be a threshold for maximum inactivation efficacy, however, this threshold is not consistent, so if it exists it should be system and ambient conditions specific. Also, according to Wan et al. (2021), the other ROS and RNS formed with higher voltage (and power) are more efficient than ozone and it could be that strong electric fields could also be influencing bacterial inactivation except for the RONS (Wan et al., 2021). Thus, the power density (dissipated power per plasma volume (W/cm^3)) has been proposed in the current study to be reported in the experiments related to NTP as a basis for comparison of decontamination efficacy results.

The modeling framework started from the simple approach of multiple linear regression and was compared to more complex approaches using machine learning. The final modeling strategies are rather complex for the number of observations that are available at the moment, but so are the parameters and the variability involved in NTP inactivation. The linear regression with or without regularization showed a poor fit, which could be mainly attributed to the high variance of the dataset, a significant amount of which cannot be explained with a linear model using the predetermined parameters. The elastic net regularization did not improve the obtained results, which could be due to the limited parameters selected that were already determined from the exploratory data analysis. XGBoost and random forests, followed by support vector machines, showed higher R^2 values on the test dataset when compared to artificial neural networks. This aligns with existing literature which suggests that while deep learning has achieved remarkable success in various fields such as image, audio, and text processing, it often underperforms on small, tabular datasets (Grinsztajn et al., 2022; Shwartz-Ziv and Armon, 2022).

Based on the exploratory data analysis, the frequency, the product temperature, the dominant working gas, the type of cells, and the selective recovery medium (yes, no) were excluded from the set of potential predictors, before the modeling steps. Our results showed that dissipated power per plasma volume and matrix category, followed by the microbial genus and the pH were the most important parameters affecting decontamination efficacy. Regarding the effect of pH, higher pH values seemed to increase the decontamination efficacy, which seemed contrary to what would be expected based on other technologies such as thermal treatment but given the mean pH value of 6.8 of our dataset (Supplementary Table 2.3), both higher and lower pH values would be expected to increase the decontamination efficacy. Also, the pH could partially mask the effect of other correlated parameters for which there were not enough data or variation to accurately separate them. For example, in one study included in the database, the NaCl concentration changed along with the pH, and this adaptation to osmotic stress and suboptimal pH before CAP treatment led to lower inactivation efficacy (Smet et al., 2019). XGBoost showed some differences in this order, but these four were still the most important parameters affecting decontamination efficacy, and with the same direction, together with the sample area. However, the effect of the sample area was condition-dependent according to XGBoost, and opposite than expected through multiple linear regression (i.e., higher sample area, higher decontamination efficacy), so it was not considered in the list of the most important parameters according to both models i.e., dissipated power per plasma volume and matrix category, pH and the microbial genus. The pH and the microbial genus are ranked lower than dissipated power per plasma volume and matrix category, because they were just significant in the multiple linear regression model. The effect of the upper electrode shape was not consistent for both models. The least correlated parameters accord-

ing to both models were: peak-to-peak voltage, frequency, relative humidity, product temperature, dominant working gas, type of cells (vegetative, biofilm), and selective recovery medium (yes, no).

The unexplained variability of the quantitative models for DBD (Section 2.3) has to do with multiple reasons. Some of this part has to do with the experimental bias included in the original studies (Garre et al., 2023), and more importantly with unspecified practices of each research team and the specific goals of each study that are difficult to capture with meta-regression models. The specific goals that were not captured with the current study are the generation cycle of the biofilms before treatment (from 1st to 5th generation cycle), if the cells were grown before treatment in preculture, planktonic or surface mode, different proportions of working gases (even though the dominant gas was remaining the same) and the effect of dual-species presence in the matrix. Regarding the different salt concentrations that were used in some matrices, it should be noted that for these cases the authors listed also different pH levels for the corresponding cases i.e., pH 7.4 for 0% NaCl concentration, pH 6.5 for 2% NaCl concentration, and pH 5.5 for 6% NaCl concentration. All the aforementioned parameters are expected to be highly associated with the parameters that were already reported (microbial genus, strain, growth stage, type of cells, working gas, pH). Although other parameters that are assumed to affect the decontamination efficacy, i.e., flow rate, temperature, and water activity (Anuntagool et al., 2023), were reported in our structural database, there was not enough variation in our dataset to see their actual contribution. Useful parameters related to the solid surface properties (e.g., roughness, porosity) were not reported in any of the studies included in our database, while information regarding the protein and lipid profile and concentration was limited and difficult to compare consistently. Thus, we expect that the establishment of guidelines regarding the crucial parameters that should be reported in NTP microbial inactivation studies such as the dissipated power per plasma volume, and other technical characteristics of the setup used for the experiment as well as more detailed information about the product intrinsic and extrinsic characteristics would improve the research quality and would make comparisons more robust, while the implication of the produced results would make the technology more applicable.

2.5 Conclusion

This study used active learning, for the first time in the field of food microbiology for an efficient literature review and investigated the decontamination efficacy of NTP, qualitatively for various setups and quantitatively for DBD. It established the concept

of dissipated power per plasma volume as a valuable metric of microbial inactivation. Using conventional and machine learning approaches, it was shown that dissipated power per plasma volume and matrix category, followed by the microbial genus and the pH affect significantly the decontamination efficacy and explain a significant amount of the variability. The machine learning framework that was established in this study could be a valuable tool for future research.

Supplementary Material

Supplementary Material A and Supplementary Material B are not reproduced below. They remain available with the published article (DOI: 10.1016/j.ifset.2024.103773).

Supplementary Material C: The Lasso, Ridge and Elastic net regression approaches

The LASSO estimate minimizes the least-squares penalty with the $\alpha\|w\|_1$, where α is a constant and the hyperparameter that controls the regularization penalty and $\|w\|_1$ is the ℓ_1-norm of the coefficient vector w of the model parameters (Equation S2.1). The approach for the ridge is the same, with the same hyperparameter α but with $\|w\|_2^2$ instead, as the ℓ_2-norm of the coefficient vector w (Equation S2.2). LASSO can penalize a parameter's coefficient high enough and essentially reduce it to zero (exclusion). With Ridge, this is not possible. The elastic net combines both approaches with ℓ_1 and ℓ_2, i.e., $\alpha\rho\|w\|_1 + \frac{\alpha(1-\rho)}{2}\|w\|_2^2$, with two hyperparameters: α, as explained above, and ρ which is a constant and the hyperparameter that controls the ratio between the two penalties (Equation S2.3).

$$\min_w \left(\frac{1}{2n_{\text{samples}}} \|y - Xw\|_2^2 + \alpha\|w\|_1 \right) \qquad \text{(S2.1)}$$

$$\min_w \left(\frac{1}{2n_{\text{samples}}} \|y - Xw\|_2^2 + \alpha\|w\|_2^2 \right) \qquad \text{(S2.2)}$$

$$\min_w \left(\frac{1}{2n_{\text{samples}}} \|y - Xw\|_2^2 + \alpha\rho\|w\|_1 + \frac{\alpha(1-\rho)}{2}\|w\|_2^2 \right) \qquad \text{(S2.3)}$$

where $\min_w$ is the minimization objective i.e., to find the coefficient vector w (weight), that minimizes the expression that follows, $\frac{1}{2n_{\text{samples}}}$ is a scaling factor that uses the number of samples (data points), $\|y - Xw\|_2^2$ represents the sum of the squared residual errors i.e., the sum of the differences between the observed values (y) and the values fitted by the multiple linear model (X) multiplied by the coefficient vector w, α is a hyperparameter that controls the regularization strength, $\|w\|_1$ is the ℓ_1-norm of the coefficient vector w, $\|w\|_2^2$ is the ℓ_2-norm of the coefficient vector w, and ρ is the hyperparameter that controls the ratio between the two penalties.

Supplementary Tables

Supplementary Table 2.1: Overview of the products per matrix category for all the studies collected in the database (n = 519)

Matrix category	Products
Solid food	ham, bacon, beef, chicken, pork, frozen pork, red pepper, apple, cabbage, corn salad, cucumber, fig, lettuce, tomato, almond, hazelnuts, pine nuts, walnut, brown cooked rice, brown raw rice, mixed grains, wheat grain, white cooked rice, allspice berries, allspice dried berries, black pepper, dried peppermint, dried stinging nettle, red pepper powder, saffron, cheese, powder infant formula, flour
Liquid food	apple juice, tomato juice, orange juice, coconut juice, sour cherry juice, milk, tiger nut milk
Solid medium	gelatin brain heart infusion (BHI) with different salt concentrations, gelatin-solidified tryptic soy broth (TSB) with different salt concentrations, agar, tryptic soy broth with yeast extract solidified with xanthan gum
Liquid medium	water with different salt concentrations, phosphate-buffered saline (PBS) suspensions, sorbitol suspensions in various concentrations, BHI with different salt concentrations, TSB with different salt concentrations, TSB with yeast extract, $NaNO_3$ suspensions
Abiotic surface	dried BHI with different salt concentrations dried on polycarbonate, TSB with different salt concentrations dried on polycarbonate surface, egg shell, glass, low density polyethylene, paper, nylon, polyethylene terephthalate, polyamide, gluconate surgical suture, polystyrene with or without background microbiota, polypropylene, sporestrips, stainless steel, textile surface

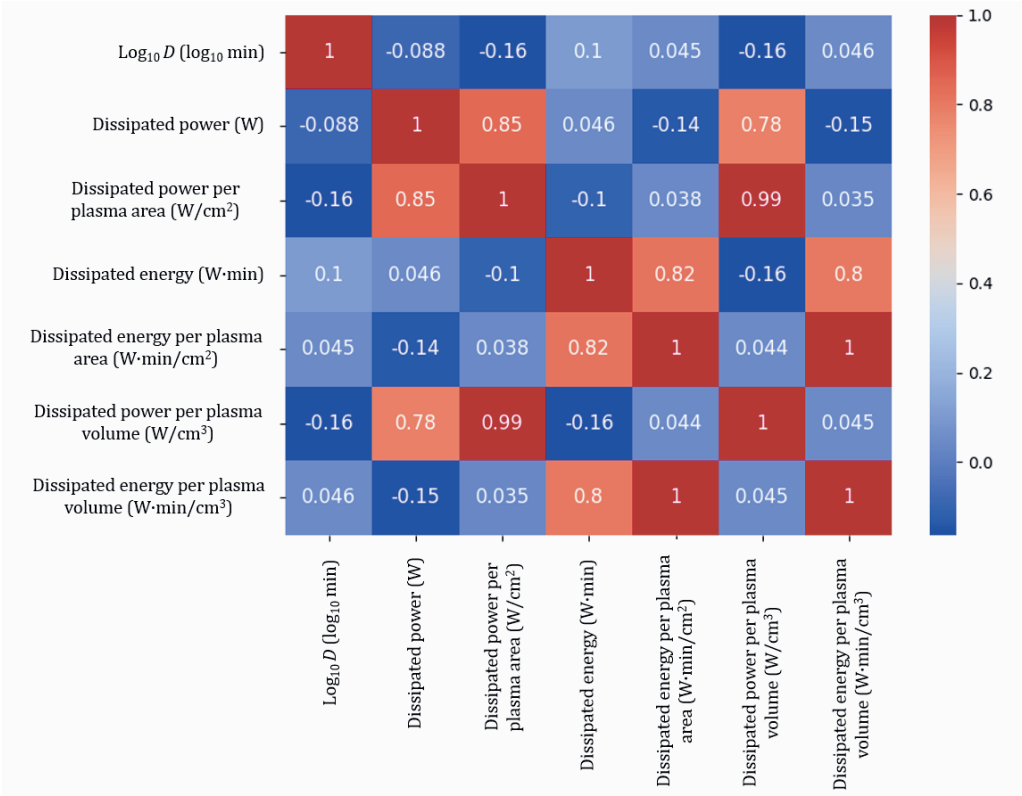
Supplementary Table 2.2: The overview of the products per matrix category for the studies used for the quantitative analysis (n = 160)

Matrix category	Products
Solid medium	gelatin brain heart infusion (BHI) with different salt concentrations, gelatin tryptic soy broth (TSB) with different salt concentrations, tryptic soy broth with yeast extract solidified with xanthan gum
Liquid medium	BHI with different salt concentrations, TSB with different salt concentrations, TSB with yeast extract
Abiotic surface	dried BHI with different salt concentrations dried on polycarbonate, TSB with different salt concentrations dried on polycarbonate surface, eggshell, glass, polystyrene with or without background microbiota

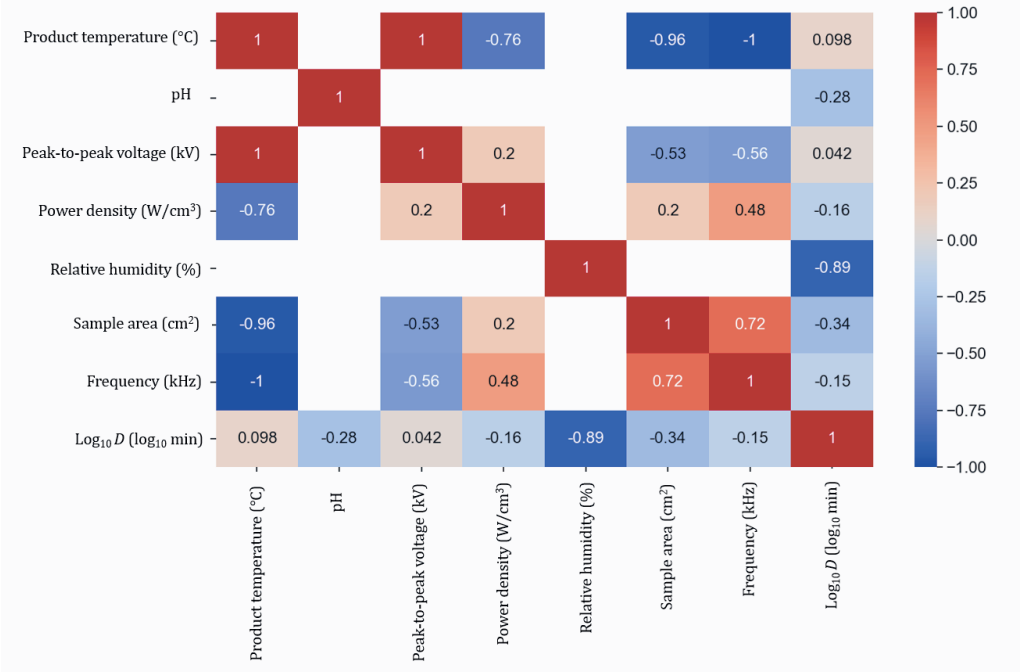
Supplementary Table 2.3: The mean and the standard deviation of each significant numerical variable, as used for scaling with Equation 2.3

Parameters	Mean	SD
Sample area (cm^2)	16.9	5.06
Power density (W/cm^3)	0.38	0.21
pH	6.8	0.57
Peak-to-peak voltage (kV)	7.5	3.30
Relative humidity (%)	49.2	7.27

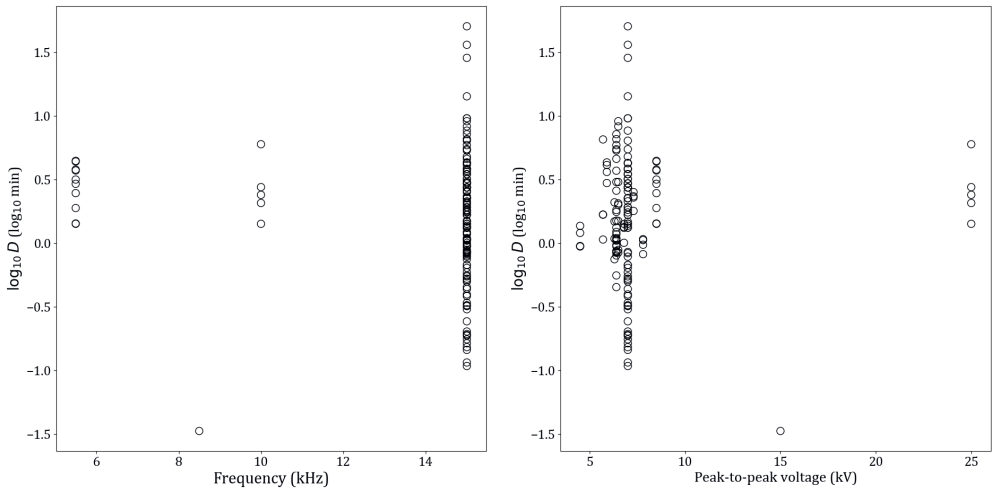
Supplementary Figures



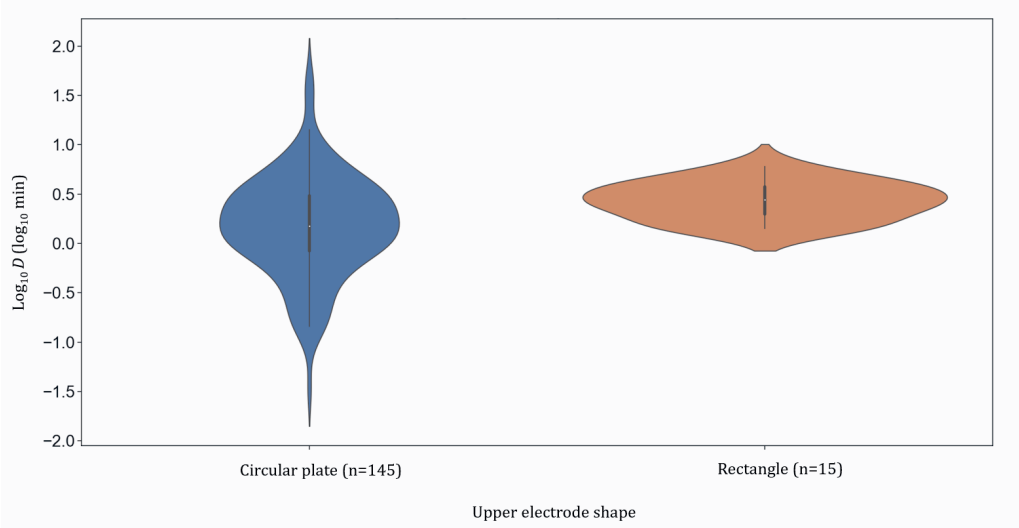
Supplementary Figure 2.1: The power- and energy-related expressions heatmap using Spearman's rank correlation coefficients.



Supplementary Figure 2.2: The numerical variables heatmap using Spearman's rank correlation coefficients. The empty cells indicate that there is not enough data available for the corresponding pair of numerical variables in question.



Supplementary Figure 2.3: The paired scatterplot of frequency (kHz) and peak-to-peak Voltage (kV) versus the $\log_{10}D$ ($\log_{10}$ min).



Supplementary Figure 2.4: The violin plot for the upper electrode shape using the DBD setup.

CRediT authorship contribution statement

George Pampoukis: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. **Marcel H. Zwietering**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration. **Heidy M.W. den Besten**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

We thank Vaiva Mikalkenaite (Food Microbiology, Wageningen University) for her support in the first literature search and data collection during her Bachelor's thesis.

Data availability

Supplementary Material C, Supplementary Table 2.1, Supplementary Table 2.2, Supplementary Table 2.3, Supplementary Figure 2.1, Supplementary Figure 2.2, Supplementary Figure 2.3, and Supplementary Figure 2.4 are reproduced in the thesis body. Supplementary Material A and Supplementary Material B are not reproduced below and remain available with the published article (DOI: 10.1016/j.ifset.2024.103773). The Chapter 2 supplementary dataset is also available via 4TU.ResearchData (DOI: 10.4121/aa957c13-9748-4af9-b8fc-b05986759034).

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N° 955431.

Chapter 3

A metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications

George Pampoukis[a], Thomas Weihe[b], Robert Wagner[b], Markus M. Becker[b], Yijiao Yao[b,c], Masja Nierop Groot[d], Uta Schnabel[b]

[a] Food Microbiology, Wageningen University & Research, Wageningen, the Netherlands
[b] Leibniz Institute for Plasma Science and Technology (INP), Greifswald, Germany
[c] Department of Food & Nutritional Sciences, University of Reading, Reading, RG6 6AD, UK
[d] Wageningen Food & Biobased Research (WFBR), Wageningen University & Research, Wageningen, the Netherlands

Published in *Scientific Data*, Volume 12, Article number 838, 21 May 2025. https://doi.org/10.1038/s41597-025-05203-5

Abstract

Non-thermal plasma research has gained significant attention as a promising technology for food-related decontamination. However, the lack of standardization in reporting key parameters restricts the applicability and reusability of results in this field. To address this, the present study introduces a metadata schema (MDS) that standardizes critical parameters and provides operational details for non-thermal plasma microbial inactivation in food-related applications. The metadata schema is implemented as an extension to Plasma-MDS, a metadata schema for plasma science, and aims to apply the FAIR (findable, accessible, interoperable, and reusable) data principles to food-related applications. The framework highlights relevant parameters related to the plasma source, the medium in which the plasma is ignited, the target (food- and microbial-related parameters) treated with the plasma, and the diagnostic methods used to investigate these properties. Concrete examples of accurate metadata descriptions of all relevant parameter categories are provided in a human and machine-readable format to simplify adoption and dissemination. The introduced metadata schema strives to improve the comparability, reproducibility, and reuse of non-thermal plasma research data for food-related applications, paving the way for more generalizable insights to legal authorities and the food industry.

3.1 Introduction

Non-thermal plasma is a partially ionized gas with a broad range of practical applications. It has received much attention in a wide variety of food-related applications including microbial inactivation on food and food processing surfaces (Katsigiannis et al., 2021) or their simulants, enzyme inactivation, seed germination, and physicochemical modifications of food products (Thirumdas et al., 2015). This is mainly due to its gentle and mild character, low temperature, antimicrobial efficacy, integrability, and upscaling possibilities in existing food processing chains (Laroque et al., 2022). For microbial inactivation purposes, there are three main fields of research interest: (i) the decontamination of food product surfaces (including spore inactivation); (ii) the antimicrobial treatment of fresh products using plasma-treated water, and (iii) the decontamination of abiotic surfaces such as packaging materials and food processing surfaces, including the inactivation of biofilms (Misra et al., 2011; Scholtz et al., 2015). This research interest led to the production of a vast amount of experimental data related to non-thermal plasma microbial inactivation. However, in multiple research articles published in the literature, various non-thermal plasma processing conditions crucial for understanding the experimental setup are not indicated precisely (Ozen and Singh, 2020). One explanation for this issue could be that many studies lack multidisciplinary teams with expertise in multiple scientific fields such as plasma physics, chemistry, and microbiology.

The importance of well-reported scientific data and the resulting data quality has received more attention in recent years (Keller et al., 2017; Pampoukis et al., 2024). Frameworks for assessing data quality have become established, particularly in the area of big data (Batini et al., 2015; Cichy and Rass, 2019). To improve reproducibility and replicability, standard research practices must be improved. In addition to improving study design and data analysis, it is also important to provide relevant experimental details that are needed to increase transparency in research and therefore reproducibility and reuse (National Academies of Sciences and Medicine, 2019). For novel technologies, there can be challenges to accurate descriptions of the processing conditions. For example, with Pulsed Electric Fields (PEF), inconsistencies in reporting crucial parameters led to the establishment of guidelines by the International Society for Electroporation-Based Technologies and Treatments (Raso et al., 2016).

Plasma classification remains challenging because by definition it is an inherently broad term and multiple categorization methods for plasmas and their applications do exist in the literature. Briefly, plasma can be categorized as thermal and non-thermal based on the thermodynamic equilibrium status. The plasma is categorized as thermal when the electron temperature matches the temperature of heavy particles in the gas, while in non-thermal plasmas the temperature of electrons is significantly higher

than the gas temperature (Liao et al., 2017; Scholtz et al., 2015). Non-thermal plasma categorization is often based on the mode of treatment (direct, indirect), the type of the device used for generating the plasma (e.g., dielectric barrier discharge (DBD)), the voltage frequency to generate the plasma (e.g., microwave plasma source (MPS), radio-frequency (RF)), the operating pressure level, and the discharge type (e.g., filamentary, diffuse) (Franke et al., 2020; Pampoukis et al., 2024). These options are just an excerpt as one type of device can be combined with a different one in the same setup, but also the different power supplies and operating pressures can form new categories.

In recent years, efforts have been started to develop a scheme for the categorization and structured description of plasma studies, including the adoption of the FAIR data principles (Wilkinson et al., 2016) in plasma-related research. This has led to the creation of the metadata schema for plasma (Plasma-MDS), which is an extension of generic metadata schemata to cover all plasma applications (Franke et al., 2020). Additionally, a data platform has been developed to serve as a repository for the stored data (Leibniz Institute for Plasma Science and Technology, 2019).

In the current study, Plasma-MDS was tailored specifically to non-thermal plasma decontamination data and metadata for food-related applications, covering (i) the decontamination of food product surfaces and their simulants (including spore inactivation); (ii) the antimicrobial treatment of food products and their simulants using plasma-treated liquids, and (iii) the decontamination of abiotic surfaces such as packaging materials and food processing surfaces, including the inactivation of biofilms. This extension includes recommendations for specific data and metadata descriptions and examples of which processing, microbial, and food-related parameters are critical, in a machine-readable format that adheres to the FAIR data principles. This aims at helping researchers in the field to better describe the key parameters, making results in the literature more comparable and reusable. Additionally, it forms the foundation for providing more generalized insights into the feasibility and decontamination efficacy of non-thermal plasmas for legal authorities and the food industry.

3.2 Current standards and missing gaps

The challenge to provide understandable, reproducible, and interoperable data of non-thermal plasma sources, diagnostics (physical and chemical), and application properties is among the largest hurdles to improving the communication of scientific outcomes and supporting the industry to implement non-thermal plasma into their value chains. Recently, the standards needed for plasma sources and their process parameters, diagnostics, and properties have been discussed by Alves et al. (2023). Based on the experience obtained at the Gaseous Electronics Conference (GEC) Reference

Cell (GECRC) and the Cooperation in Science and Technology (COST) Jet, the authors have recommended developing new standards which include best practices and available standards as well as new inventions to come to shared practices in reporting data sets to enhance the exchange between researchers and industry, to increase the reproducibility and reuse of experiments and models, to strengthen the transparency of methods and data, and to increase the reliability of models and digital data. With better alignment of scientists on reporting practices within the plasma physics community, the fundamental findings within research may be transferred more efficiently to practice (Alves et al., 2023).

3.2.1 Standardization of parameters in plasma medicine

Nowadays, the use of non-thermal plasma at atmospheric pressure in medical therapy is widely accepted and partly co-financed by health insurance companies. In 2022, the first guideline on the therapeutic use of cold physical plasma in the medical field in Germany was published (Deutsche Gesellschaft für Mund-, Kiefer- und Gesichtschirurgie (DGMKG), 2022). However, over 15 years ago, when research in the field of plasma medicine began, the requirements from physical, chemical, and biological points of view had not yet been defined (Oehmigen et al., 2011). Therefore, during this time frame, the necessary knowledge and understanding of mechanisms of action were investigated by researchers and physicians (Bernhardt et al., 2019). During this time, physical properties like plasma temperature, UV radiation, and electromagnetic fields were characterized, as well as the toxicological aspects of the plasma effluents for patients and therapists, given that direct contact was the preferred application in medical therapies. Additionally, the antimicrobial and in vitro / in vivo cell biological effects and their mode of action were analyzed. Some of these investigation requirements resulted in national standards like DIN SPEC 91315 (DIN SPEC 91315:2014-06 General requirements for plasma sources in medicine: https://dx.doi.org/10.31030/2141389, accessed 29.07.2024) (Mann et al., 2016; Tiede et al., 2016). However, as the therapeutical application of plasma in the field of plasma medicine is dependent on national guidelines, challenges in harmonization across different countries still exist.

3.2.2 The challenges of standardization in food-related applications

In the agricultural field, Waskow et al. (2022) attempted to standardize the reporting of parameters for plasma treatment of seeds to aid reproducibility and comparison between experiments. They developed a protocol organized into four main parts: (i)

plasma device characteristics; (ii) seed pre-treatment; (iii) seed plasma treatment, and (iv) seed post-treatment. For food-related applications, similar to the challenges in plasma medicine, the different plasma sources and application modes make it necessary to define specific requirements for describing the physical and chemical properties of the non-thermal plasma itself as well as the treated target. Differences in applications for food processing involve the consensus of the carrier gas used, which is in most cases atmospheric or compressed air, due to the acceptable consumable costs in the food industry. Another unique aspect is the broad range of applicable temperatures, from sub-zero temperatures (as low as -20 °C) to pasteurization temperatures (up to +70 °C) depending on the application. Therefore, more plasma sources can be used, including DBDs, jets, and microwave discharges (Laroque et al., 2022). Depending on the energy input, variations in reactive species within the gas composition can be expected. In combination with humidity (originating from the food itself, processing conditions, or intentionally added water) this results in a variety of reactive species forming a bouquet of antimicrobial active components. These components must be understood from a physicochemical and biochemical point of view to understand their mode of action. Further, the impact of plasma treatment on the sensorial aspects of foods, and the toxicological risks need to be assessed to ensure the safety of the technology to the consumers and applicants. In non-thermal plasma applications for food processing and for extending products' shelf life, packaging is also a crucial factor to consider, as the process can impact the material surface properties (Štěpánová et al., 2021). The selected plasma parameters must be compatible with the materials and surfaces used in the food industry, including the processing line and packaging materials. This is particularly important for ready-to-eat food products, where the packaging comes into direct contact with the food, which will not undergo further processing before consumption. As most review articles and current guidelines for non-thermal plasma applications in food processing do not provide consistent parameters and terminology for plasma sources, gases, and chemical / biological characterizations, the comparability of the different processes is often lacking, which inhibits the upscaling process and the successful transfer from lab to industry. Inconsistencies in descriptions exist even for the same plasma sources or even for the "same" parameters. A good example of this issue is the term power, where multiple expressions can be found in the literature without being clear about what they represent and how they were estimated, such as applied, forward, reverse, reflected, active, apparent, operating, generating, dissociated, transferred, input, or dissipated power (Pampoukis et al., 2024). These inconsistencies undermine the general understanding of stakeholders, potentially leading to less acceptance and realization / applicability within the industry where conventional alternatives are still being used despite their disadvantages. Following the example set by Raso et al. (2016) for PEF, where the key information to be reported in microbial inactivation and PEF-assisted processing studies has been

outlined, a similar approach was applied here for non-thermal plasma applications to improve the comparability of results.

Therefore, this study introduces recommendations and a practical scheme extending Plasma-MDS for standardized documentation of data and metadata that should accompany non-thermal plasma decontamination research in food-related applications. It is worth noting that this also contributes to the implementation of the FAIR data principles in this field of research and development, which is an urgent challenge in many areas. Moreover, explanations and examples are provided through a case study simulation to support the use and dissemination of the suggested metadata schema.

3.3 A metadata schema for food-related applications

3.3.1 Extension of Plasma-MDS

In order to extend Plasma-MDS for the description of plasma research studies related to decontamination and food-related applications, the scheme presented in Figure 3.1 and detailed in Table 3.1, Table 3.2, Table 3.3, Table 3.4, Table 3.5, and Table 3.6 was developed. Based on this metadata schema, benchmarking examples were also developed, as presented in Table 3.7, Table 3.8, Table 3.9, and Table 3.10. These should assist researchers in selecting the critical parameters along with their metadata. Plasma-MDS consists of five main schema elements that should be reported in plasma studies if applicable, namely, (i) "Source" (information about the plasma source, i.e., the device used to generate the plasma); (ii) "Medium" (information about the medium in which the plasma is generated); (iii) "Target" (information about the target treated with the plasma); (iv) "Diagnostics" (information on the diagnostic methods and instruments used to investigate plasma, medium or target properties); and (v) "Resource" (information about the file and data types of the research data belonging to the plasma study) (Franke et al., 2020). The suggested metadata schema (Figure 3.1) for food-related applications of non-thermal plasmas specifically extends the "Properties" and "Procedure" attributes of the schema elements "Source", "Medium", "Target", and "Diagnostics", while "Resource" remains unchanged since it applies to all metadata documents. Further attributes of the main schema elements, i.e., "source.name", "source.application", "source.specification", "medium.name", "target.name", and "diagnostics.name" also remain unchanged and should be used as intended for general naming and categorization of the data. While for the sake of generality, Plasma-MDS does not further specify the attributes "Properties" and "Procedure" for the different schema elements and considers them free text fields, the metadata schema proposed here specifies and structures the information for the dedicated application field. "Properties"

should include specific information and measurable parameters and "Procedure" defines the steps involved, such as those required to stabilize the plasma.

Figure 3.1 summarizes all relevant categories of parameters that affect the decontamination efficacy of non-thermal plasmas as an extension to the core elements of Plasma-MDS for food-related applications. Those parameters are organized into four levels, which cover the source (i) used in the process, the medium (ii) used for the process, the target (iii) of the process, and possible diagnostics (iv) to mirror the process. The extended metadata schema for all parameters is also publicly available in a machine-readable JSON-schema format (Pampoukis, 2025). The JSON-schema representation was developed using Adamant: a JSON schema-based metadata editor, allowing the user-friendly collection of standardized metadata in accordance with the given schema (Chaerony Siffa et al., 2022).

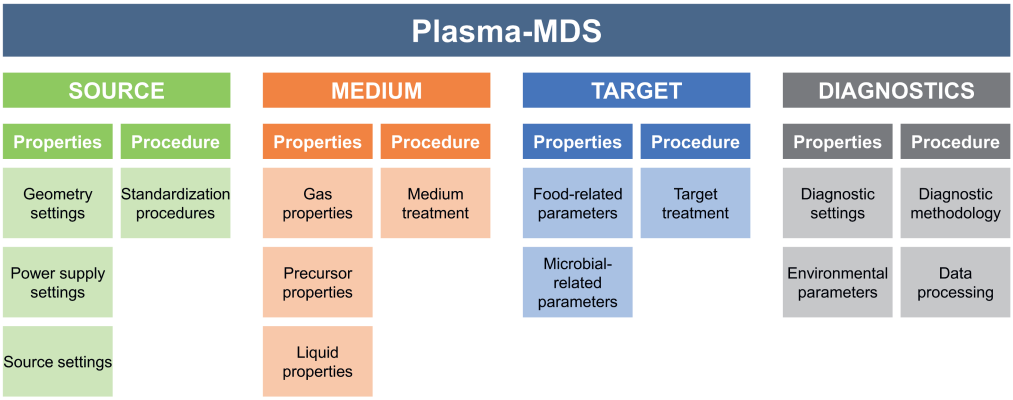


Figure 3.1: Extension of the Plasma-MDS core schema elements (solid color) with categories of parameters that are critical in food-related plasma decontamination applications.

3.3.2 Plasma source-related parameters

Examples of plasma sources, which are about to enter the field of food processing or agriculture are summarized in various reviews (Birania et al., 2022; Keener et al., 2012; Laroque et al., 2022; Misra et al., 2016; Rajan et al., 2023; Thirumdas et al., 2015). For instance, Misra et al. (2019) reviewed devices for plasma generation inside sealed packages. In general, DBD is the most frequent setup in food-related applications (Asl et al., 2022; Liao et al., 2017; Pampoukis et al., 2024). That being said, DBD is not always ideal for food applications particularly those involving matrices with complex 3D shapes. These shapes are difficult to position between two electrodes that require a relatively short distance for effective plasma generation. In these cases, an indirect

mode of treatment using plasma-processed gases or plasma-treated liquids, or alternative plasma sources such as plasma jets, or corona discharges may be more suitable.

Plasma source-related parameters are critical in all non-thermal plasma applications and can significantly affect the decontamination efficacy. Additionally, several plasma source properties can serve as substitutes for plasma diagnostics in some applications, as the latter are often complex and difficult to obtain. Specifically, the dissipated power and the estimated dimensions of the generated plasma (e.g., based on the source geometry settings) were used in previous research as an indirect method for summarizing the plasma process intensity for microbial inactivation experiments in the food domain (Pampoukis et al., 2024). Table 3.1 presents the relevant plasma source parameters that are critical in food-related applications. The main properties include the geometry settings, the power supply settings, and the source settings. Source geometry can define the volume and the composition of the generated plasma and thus its decontamination efficacy. Additionally, electrode materials can also play a role, as seen in DBD plasma sources (Ragni et al., 2016). The power supply specifications are important because as mentioned in Section 3.1, different plasma categories can be defined based on them such as RF and MPS, and the chosen settings can lead to different reactive oxygen and nitrogen species in the plasma medium. In addition, non-thermal plasma processes can be categorized based on the operating pressure: (i) low-pressure ($\leq 10^3$ Pa); (ii) medium-pressure (10^3 to 10^5 Pa); (iii) atmospheric-pressure ($\approx 10^5$ Pa), and (iv) high-pressure ($\geq 10^5$ Pa), as considered by the "source.specification" attribute of Plasma-MDS. Among these, atmospheric-pressure plasma processes are the most interesting for industrial applications, including food processing, because they do not need any complex vacuum techniques and they are frequently able to ignite plasma in ambient air.

Table 3.1: Minimal set of plasma source-related parameters that are critical in non-thermal plasma research for decontamination purposes

Id	Field name	Description	Field type
1.1.1	**Geometry settings**	**Object including all attributes defining the configuration of the plasma setup such as the type of plasma device used (e.g., DBD, plasma jet, corona discharge, etc.), its physical shape, structure, dimensions, and materials.**	**Object**
1.1.1.1	Geometry description	Free text description of the source geometry and materials.	String
1.1.2	**Power supply settings**	**Object including all attributes defining the electrical power supply used in plasma generation.**	**Object**
1.1.2.1	Power supply [AC, DC]	The electrical power used to generate the plasma.	Enumerated list
1.1.2.2	Waveform [sinusoidal, square, ramp, pulse, arbitrary]	The shape of the electrical signal.	Enumerated list
1.1.2.3	Input power [W]	Total power supplied to the plasma system.	Number
1.1.2.4	Peak-to-peak discharge voltage [kV]	The voltage difference between the highest and lowest points of the discharge.	Number

Id	Field name	Description	Field type
1.1.2.5	Frequency [kHz]	Frequency of the electrical discharge.	Number
1.1.3	**Source settings**	**Object including all attributes of the plasma source used to generate the desired plasma.**	**Object**
1.1.3.1	Discharge current [A]	Current passing through the plasma discharge.	Number
1.1.3.2	Dissipated power [W]	Power actually used by the plasma discharge, excluding losses.	Number
1.1.3.3	Specific energy input [J/cm^3]	Energy supplied per unit volume of the treated medium.	Number
1.2.1	**Standardization procedures**	**Object including all attributes defining how the plasma source is standardized to ensure consistency across experimental runs.**	**Object**
1.2.1.1	Waiting time [min]	The time needed to reach a stable plasma state before starting the treatment.	Number

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

Not only the specifications and the corresponding categorization of the plasma source are important, but also the chosen experimental settings (properties). The power supply settings that can affect the plasma decontamination efficacy, and are therefore critical, include input power (Song et al., 2015), peak-to-peak discharge voltage (Govaert et al., 2019), and frequency (Stulić et al., 2021). The source settings that can affect the plasma decontamination efficacy include the discharge current, the dissipated power (the power that goes to plasma generation) (Bauer et al., 2017), and the specific energy input (combination of the dissipated power, time, and volume of the plasma gas).

When it comes to plasma source procedures, the standardization procedures, and especially the waiting time required to achieve a stable plasma, are a critical factor for ensuring comparability between experiments conducted across different laboratories, especially in cases where no plasma diagnostic procedures are implemented.

Where applicable, the parameters reported in Table 3.1 should also be accompanied by a description of the method used for their estimation in terms of diagnostics (see Section 3.3.5 below) to make the parameters of different research studies transparent and traceable.

3.3.3 Plasma medium-related parameters

The medium of the process transports the active compounds generated by the source to the site of action (target). Table 3.2 presents the relevant plasma medium-related parameters that are critical in food-related applications. The measurable properties of the plasma medium include gas properties, precursor properties, and liquid properties (only relevant for plasma-treated liquids). Regarding gas properties, the pressure level can affect the decontamination efficacy, especially in low-pressure applications (Tanino et al., 2020). Plasma can be generated in a static gas environment (enclosed

reactor, no gas flow), where the plasma volume is defined by the reactor's size, or in a continuous mode, where the gas flows at a predetermined rate. The flow rate is critical since it affects the gas residence time, which in turn affects the plasma chemistry (Waskow et al., 2022). The temperature of the gas during plasma operation can also be important, especially if it affects the target's temperature. Additionally, temperature, carrier gas composition, and humidity can all affect the concentration of reactive oxygen and nitrogen species generated in the plasma gas. Moving to the precursor properties, the addition of precursors can create different plasma processes, with the extent of variation dependent on the carrier gas and the amount of added compounds or precursors. For example, Saremnezhad et al. (2021) observed variations in the efficacy of non-thermal plasma in reducing food allergens and mycotoxins depending on changes in the plasma medium. For decontamination purposes, Majumdar et al. (2009), Hertwig et al. (2017), and Patil et al. (2014) are only a few of the studies that evaluated the effect of the plasma medium (carrier gas) on the decontamination efficacy. Moreover, the influence of added compounds or precursors on a plasma process can be adjusted to achieve the desired target reaction. Numerous studies, for instance, have examined the role of inducer gases with varying oxygen levels, while many contemporary studies continue to employ helium or argon to achieve desired plasma properties (Cullen et al., 2018; Kim et al., 2017; Moritz et al., 2017).

Table 3.2: Minimal set of plasma medium-related parameters that are critical in non-thermal plasma research for decontamination purposes

Id	Field name	Description	Field type
2.1.1	**Gas properties**	**Object including all attributes defining the carrier gas.**	**Object**
2.1.1.1	Gas description	Free text description of the carrier gas.	String
2.1.1.2	Pressure [Pa]	Pressure of the carrier gas during plasma operation.	Number
2.1.1.3	Flow rate [L/min]	The rate at which the gas is supplied to the plasma source.	Number
2.1.1.4	Carrier gas composition	The type and composition of the gas mixture used in the plasma (e.g., 90% N_2, 10% O_2).	String
2.1.1.5	Purity [%]	Purity level of the gases used (e.g., 99.99% purity).	Number
2.1.1.6	Temperature [°C]	Temperature of the gas during plasma operation.	Number
2.1.1.7	Humidity [% RH]	Relative humidity level of the plasma medium.	Number
2.1.2	**Precursor properties (where applicable)**	**Object including all attributes defining the presence of any precursors or added compounds in the plasma medium.**	**Object**
2.1.2.1	Precursor description	Free text description of the precursor added in the carrier gas.	String
2.1.2.2	Amount of precursor added to the medium [ml or g]	Quantity of the precursor added to the plasma medium.	Number
2.1.3	**Liquid properties (where applicable)**	**Object including all attributes defining the plasma-treated liquid.**	**Object**
2.1.3.1	Liquid description	Free text description of the liquid being treated by the plasma.	String
2.1.3.2	Volume [L]	The volume of the liquid being treated.	Number
2.1.3.3	Distance [cm]	The distance between the plasma source and the liquid surface.	Number

Id	Field name	Description	Field type
2.1.3.4	Stirring [yes, no]	Indicates whether the liquid was stirred during plasma treatment.	Enumerated list
2.1.3.5	pH	The pH level of the liquid before and after treatment.	Number
2.1.3.6	Oxidation-reduction potential [mV]	The oxidation-reduction potential of the liquid.	Number
2.1.3.7	Minimum conductivity [S/m]	The minimum electrical conductivity needed for plasma ignition.	Number
2.1.3.8	Conductivity [S/m]	The electrical conductivity of the liquid.	Number
2.1.3.9	Treatment time [min]	The duration for which the plasma source was applied to treat the liquid, creating a plasma-treated liquid (PTL) intended for subsequent decontamination purposes.	Number
2.1.3.10	Reactive oxygen and nitrogen species	Free text description of the reactive oxygen and nitrogen species measured in the plasma liquid (e.g., hydroxyl radical, hydrogen peroxide, nitric oxide, nitrite, nitrate).	String
2.1.3.11	Temperature [°C]	The temperature of the liquid during plasma treatment.	Number
2.2.1	**Medium treatment**	**Object including all attributes defining the plasma medium.**	**Object**
2.2.1.1	Medium modification (where applicable)	The way the medium was modified before the plasma generation (e.g., humidification of the carrier gas).	String
2.2.1.2	Medium treatment time (where applicable) [min]	The treatment time in which the medium (gas, liquid, or both) is treated by a plasma source that will then be used to treat a target. Only applicable in the indirect mode of treatment.	Number

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

In cases where a plasma-treated liquid is used as the medium for decontamination, the properties of the plasma-treated liquid must also be described in addition to the previously mentioned parameters. These properties include the liquid's composition, volume, distance from the plasma source, stirring conditions, pH, oxidation-reduction potential (ORP), minimum conductivity to ignite the plasma, conductivity, treatment time, reactive oxygen and nitrogen species (RONS), and temperature. These parameters are crucial since they can determine the antimicrobial compounds present in the liquid and therefore its decontamination efficacy (Weihe et al., 2023; Zhao et al., 2020). In addition to these parameters, it is essential to report the actual volume of the plasma medium that was used to treat the target.

As for the plasma medium procedures, the medium treatment consists of two main attributes: (i) medium modification, and (ii) medium treatment time. As discussed in Section 3.2.2, but also in other studies (Georgescu et al., 2017; Patil et al., 2014), medium modification, such as humidification procedures, can result in a variety of different reactive species, and therefore it is crucial to be described. The medium treatment time applies to indirect plasma applications, such as plasma-processed gas or plasma-treated liquid. It refers to the duration during which the plasma source treats the medium (gas, liquid, or both). For these applications, this is distinct from the target treatment time, which refers to the duration the treated medium is applied to the target.

3.3.4 Plasma target-related parameters

The target treated with plasma is the focus of any application whether it is a food product, packaging material, food processing surface, or a simulant of these. As mentioned in Section 3.3.3, plasma-processed gases or plasma-treated liquids are considered part of the plasma medium (and not of the target), as they are used to treat a target in an indirect mode of treatment. Table 3.3 presents the relevant plasma target-related parameters that are critical in food-related applications. In addition to specific food- and microbial-related parameters which will be discussed in detail in the following section, other important properties applicable to all food-related matrices include the distance between the plasma source and the target surface and the surface temperature during treatment.

Table 3.3: Minimal set of plasma target-related parameters that are critical in non-thermal plasma research for decontamination purposes

Id	Field name	Description	Field type
3.1.1	**Food-related parameters**	**Object including all attributes defining the food matrix, the packaging material, or the food processing surface being treated.**	**Object**
3.1.1.1	Matrix description	Free text description of the matrix being treated.	String
3.1.1.2	Distance to surface [cm]	The distance from the plasma source to the target surface.	Number
3.1.1.3	Surface temperature [°C]	The temperature of the target surface during plasma treatment.	Number
3.1.2	**Microbial-related parameters**	**Object including all attributes defining the microbial properties of the target.**	**Object**
3.2.1	**Target treatment**	**Object including all attributes defining the methodology followed to treat the target.**	**Object**
3.2.1.1	Mode of treatment [direct, indirect]	Specifies whether the plasma treatment was applied to the target in direct (UV / vacuum ultraviolet (VUV) irradiation, plasma particles) or indirect (plasma-processed gas / plasma-treated liquid) way.	Enumerated list
3.2.1.2	Spatial location relative to the plasma source	Free text description of the spatial position of the target with respect to the electrodes or the plasma source geometry (how the target is exposed to the plasma medium (UV / VUV irradiation, plasma particles, plasma-processed gas / plasma-treated liquid)).	String
3.2.1.3	Target treatment time [min]	The duration for which the plasma medium was applied to the target matrix. For direct treatments the medium treatment time and the target treatment time match.	Number

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

As for the plasma target procedures, the target treatment consists of three main attributes: (i) the mode of treatment, (ii) the spatial location relative to the plasma source, and (iii) the target treatment time. Regarding the first attribute, the mode of treatment is considered a plasma target-related parameter because it defines how the plasma source is acting on the target (Franke et al., 2020). In this aspect, two main categories can be formed, i.e., (i) direct (including direct and semi-direct), and (ii) indirect treat-

ment. In direct treatments, plasma is in direct contact with the substrate, and the interaction is based on irradiation (UV, VUV), charged molecules, radicals, and reactive species. In semi-direct treatments, the distance between plasma and substrate is larger, and therefore, only irradiation, long-lived radicals, and metastable inhibitory substances reach the target. In this case, the antimicrobial effect is due to irradiation, long-lived radicals as well as metastable and inhibitory substances, but no interactions with charged particles take place (Schlüter et al., 2013). In the food industry and agriculture, products with a resistant matrix are predominantly treated with a direct treatment. Numerous studies have used direct treatment to decontaminate various products including black peppercorn (Mošovská et al., 2018), fresh vegetables, fruits, nuts, powdered food samples (Han et al., 2020), and sliced cheese (Huang et al., 2020). Additionally, direct treatment has been applied to decontaminate food processing surfaces (Alessandria et al., 2019; Tanino et al., 2005). However, direct treatments are predominantly surface treatments that can be applied for treatment goals without any vulnerable surface. Indirect treatments can be applied through irradiation with UV and VUV (where applicable) light (e.g., UV lamps) where plasma is enclosed in a UV / VUV-transparent reactor without any interaction with plasma particles or by using previously generated plasma to process a gas or treat a liquid (Schlüter et al., 2013). The decision between direct and indirect treatment is mainly based on the desired goal and the matrix properties (including geometry) of the target. Vulnerable products are better treated with an indirect plasma-based method, whereas more resistant products with regular surfaces can also be subjected to direct treatment. Schnabel et al. (2012) described the differences between direct and indirect treatment of food products for decontamination purposes. Indirect treatments are regularly applied for treatment goals with vulnerable surfaces that may change color, texture, or taste. Whichever process type, direct or indirect, is chosen, the effect of such a process is always realized by the chemical components that are generated by an ignited plasma. In contrast to direct treatment, in which the plasma-produced reactive compounds directly act on the treatment target, the reactive compounds used in indirect treatments do not act at their production site. Consequently, processes designed upon indirect plasma treatments need a carrier for the reactive compounds such as RONS. This carrier can be either a plasma-processed gas (PPG) or a plasma-treated liquid (PTL), as explained in Section 3.3.3. Among gases, plasma-processed air (PPA) is the most extensively studied category in recent literature, while among liquids, plasma-treated water (PTW) also referred to as plasma-activated water (PAW) or plasma-functionalized water (PFW) is the primary focus. For instance, PTW can be used to improve the microbiological safety of fresh-cut produce (mainly leafy greens) (Gülenç et al., 2024; Laurita et al., 2021; Schnabel et al., 2021a). Indirect treatments offer the possibility of mild plasma treatments for vulnerable treatment goals like food products or other sensitive goods. Additionally, indirect treatments are realized on the laboratory scale for water treat-

ments (Piskarev and Ivanova, 2021). It should be noted that in most cases indirect treatments involve additional parameters to be described, for example, the plasma-treated liquid parameters (Table 3.2). Also, the description of the design of the setup, the definition of the distance of the plasma source, and how plasma is transferred to the target are often more challenging. Clearly defining the spatial location of the target relative to the plasma source is essential for both direct and indirect treatments, as it impacts the treatment uniformity due to variations in plasma source geometry. Additionally, this spatial relationship complements the mode of treatment as it helps clarify whether the target is exposed to UV / VUV irradiation, plasma particles, reactive species within the plasma effluent, or a combination of these factors. The last attribute of the plasma target procedures refers to the target treatment time because as with all processing technologies, the treatment time of the target is essential for determining the decontamination efficacy.

3.3.4.1 Food-related parameters

Based on the target matrix, important parameters or prerequisites and restrictions may enter the process design. Product properties such as the structure (Muhammad et al., 2018), vulnerable ingredients (Dharini et al., 2023; Gavahian et al., 2018), or the water content (Shanker et al., 2023) may play a crucial role in the process design since they may alter the whole process chemistry. In Table 3.4, there is a summary of the food-related parameters that can affect the decontamination efficacy of non-thermal plasma.

Table 3.4: Minimal set of food-related parameters that are critical in non-thermal plasma research for decontamination purposes

Treatment matrix	Parameters
Abiotic surface	- Surface material - Surface metrology [roughness, waviness in cm] - Dimensions [e.g., L x W x H or radius, in cm]
Liquid medium	- Type - Volume [L] - Salt content [% w/v] or a_w - Viscosity [Pa·s] - Reflection index [°Brix] - Conductivity [S/m] - Oxidation-reduction potential [mV] - pH - (Commercial brand)

Treatment matrix	Parameters
Solid medium	- Type - Weight [g] - Medium shape - Dimensions [e.g., L x W x H or radius, in cm] - Treated surface area [cm^2] - Contact angle (where applicable) [°] - Salt content [% w/v] or a_w - pH - (Commercial brand)
Liquid food	- Volume [L] - Salt content [% w/v] or a_w - Viscosity [Pa·s] - Reflection index [°Brix] - Conductivity [S/m] - Oxidation-reduction potential [mV] - pH - Fat percentage [%] - Protein content [%] - Density [kg/m^3]
Solid food	- Weight [g] - Food shape - Dimensions [e.g., L x W x H or radius in cm] - Treated surface area [cm^2] - Contact angle (where applicable) [°] - Salt content [% w/v] or a_w - pH - Fat percentage [%] - Protein content [%]

It should be noted for clarification purposes that an abiotic surface is different from the other treatment matrices due to the condition of being dry. The surface material (Table 3.4) refers to the part of the abiotic surface that is contaminated with microorganisms and treated with non-thermal plasma. If more than one material constitutes the abiotic surface then all of them should be described. The matrix category of solid / liquid medium corresponds to those experimental setups where the microorganism is inoculated to a growth medium, a buffer, or water and then treated directly with non-thermal plasma. Solid medium is the category that covers all the experiments conducted in growth media contaminated with microorganisms, either inoculated and mixed before solidification or inoculated as a drop, a spread, a stamp, or a spray after the medium has solidified. In contrast, the liquid medium is the equivalent of an abiotic surface but when treated in a wet mode (e.g., as a droplet on an abiotic surface), and can correspond to a growth medium (e.g., brain heart infusion), or water (e.g., distilled or saline). Additionally, the shape of a solid matrix (food or medium), is critical, as the dimensions can be used to estimate the treated area during a plasma process. This aspect is particularly important for direct plasma experiments, where treatment predominantly occurs at the surface. Lastly, for some of the parameters mentioned, it needs to be clear where they were measured (location in the product or the processing environment) and if they were measured before, after treatment, or both (e.g., temperature, conductivity, pH).

3.3.4.2 Microbial-related parameters

In Table 3.5, there is a summary of the relevant microbial-related parameters that can affect the decontamination efficacy of non-thermal plasma. These parameters relate to how the microorganism of interest was cultured and inoculated in the matrix before plasma treatment. The importance of these parameters has been demonstrated across various non-thermal processing technologies such as pulsed electric fields (PEF) (Lytras et al., 2024b; Raso et al., 2016; Thamsuaidee et al., 2024), high-pressure processing (HPP) (Serra-Castelló et al., 2021; Tsagkaropoulou and Karatzas, 2024), and non-thermal plasma (Battaggia et al., 2024).

Table 3.5: Minimal set of microbial-related parameters that are critical in non-thermal plasma research for decontamination purposes

Microbial forms	Parameters
Vegetative bacteria	- Genus - Species - Serotype - Single- or multi-culture - Strain [ATCC / DSM / NCTC No, or equivalent] - Growth medium before and after treatment - Incubation time before and after treatment [h] - Incubation temperature [°C] - Growth stage [exponential, stationary] - Mode of inoculation [spraying, spotting, spreading, stamping, mixing] - Inoculation procedure - Microbial counts before and after treatment [$\log_{10}$ CFU/(mL, or cm^2, or g)] - Enumeration / quantification procedure
Bacterial cells / spores in biofilms	- Biofilm age [days] - Biofilm formation protocol - Sporulation protocol - All parameters from "Vegetative bacteria"
Yeasts and molds	- Ascospores or endospores [molds] - All parameters from "Vegetative bacteria"

3.3.5 Plasma diagnostics

In food-related applications of plasma decontamination, diagnostics usually refer to the processes that take place to monitor plasma generation and stability throughout processing. Table 3.6 presents the relevant plasma diagnostics parameters that should be described in food-related applications. The main properties include the diagnostic settings and the environmental parameters, whereas the important procedures involved are the diagnostic methodology and the data processing procedures.

Table 3.6: Minimal set of parameters that should be described in plasma diagnostics procedures taking place in non-thermal plasma research for decontamination purposes

Id	Field name	Description	Field type
4.1.1	**Diagnostic settings**	**Object including all attributes defining the diagnostic method used to assess plasma source-, medium-, or target-related properties.**	**Object**
4.1.1.1	Setting description	Free text description of the diagnostic settings (e.g., spectroscopy settings, sensor calibration).	String
4.1.2	**Environmental parameters**	**Object including all attributes defining the conditions in the measurement chamber during diagnostics (e.g., humidity, temperature, pressure).**	**Object**
4.1.2.1	Temperature [°C]	The temperature of the processing environment.	Number
4.1.2.2	Humidity [% RH]	Relative humidity of the processing environment.	Number
4.1.2.3	Pressure [Pa]	The pressure of the processing environment.	Number
4.2.1	**Diagnostic methodology**	**Object including all attributes defining the experimental setup of the diagnostic procedures implemented, such as location, targets, and quantification methods.**	**Object**
4.2.1.1	Methodology description	Free text description of the experimental setup of the diagnostic procedures.	String
4.2.1.2	Location of the diagnostic application	The specific location where the diagnostic method was applied (e.g., on the plasma source, in the plasma medium, on the target).	String
4.2.1.3	Targets of qualitative / quantitative analysis	The specific components being analyzed (e.g., reactive species, target surface properties).	String
4.2.1.4	Quantification method	The technique used for the quantification of the analytes (e.g., simulation with online databases, calibration curves using standards, machine built-in library, sensor direct output). For direct outputs, the limit of detection (LOD) should be provided.	String
4.2.2	**Data processing**	**Object including all attributes defining how raw data were processed and analyzed.**	**Object**
4.2.2.1	Data processing description	Free text description of the data processing and analysis procedures (e.g., spectral deconvolution, data smoothing, software used for data interpretation).	String

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

The most representative application of plasma diagnostics involves processes used to monitor the generation of reactive species. In this respect, the key information on reactive oxygen and nitrogen species (RONS) in the plasma gas should be provided not only in relative terms but also as absolute values with clear metric units. Spectrographs, though commonly reported in the literature, are often challenging to compare across studies due to their qualitative nature. Therefore, they should be accompanied by detailed metadata, including the conditions and devices used for sample analysis, as well as the raw data used to create the corresponding graphs. Additionally, the methodology used to identify the reactive species (including the analytical method, specific species measured, and location in the product or the processing environment) as well as environmental conditions (humidity, temperature, pressure) must also be documented.

3.4 Examples of data descriptions

To make the schema-based collection of the parameters more accessible to the reader, the aforementioned metadata schema was used to generate a complete, hypothetical non-thermal plasma case study. This case study includes tables for the elements of Plasma-MDS and the corresponding parameters. The scenario involves the decontamination of black peppercorns contaminated with *Salmonella* (plasma target), treated with a DBD device (plasma source) in a direct mode of treatment using ambient air (plasma medium). Plasma diagnostics were also simulated to assess the RONS in the plasma gas. This case study is summarized in the tables below and is also made publicly available in a machine-readable JSON format, generated by completing only the relevant fields of the metadata schema for this example (Pampoukis, 2025). Again, the "Adamant" tool was used to facilitate this task (Chaerony Siffa et al., 2022).

3.4.1 Plasma source- and plasma medium-related parameters

In Table 3.7, an example of plasma source- and plasma medium-related parameters is presented. For this example, a DBD device was simulated, but the same approach could be applied to generate equivalent tables for other plasma sources and devices. Some parameters from Figure 3.1 for the plasma medium were not reported in Table 3.7 (i.e., precursor properties, liquid properties, medium treatment) as they were irrelevant to this scenario (see also: Table 3.2). If the plasma medium was a plasma-treated liquid (e.g., water), then its properties would also need to be reported (Table 3.2).

Table 3.7: An example of accurately describing the plasma source- and plasma medium-related parameters of non-thermal plasma using a DBD device

Id	Field name	Value
1.1.1	**Geometry settings**	
1.1.1.1	Geometry description	DBD plasma reactor treats the sample between two circular electrodes; a circular dielectric barrier covers the upper electrode. The diameter of the electrodes, as well as the diameter and thickness of the dielectric barrier, were measured.
1.1.1.2	Upper electrode diameter [cm]	5.5
1.1.1.3	Ground electrode diameter [cm]	5.5
1.1.1.4	Dielectric barrier diameter [cm]	7.5
1.1.1.5	Dielectric barrier thickness [cm]	0.1
1.1.1.6	Electrode distance [cm]	0.8
1.1.1.7	Upper electrode material	Stainless steel
1.1.1.8	Ground electrode material	Stainless steel
1.1.1.9	Dielectric material	Glass
1.1.2	**Power supply settings**	
1.1.2.1	Power supply	AC

Id	Field name	Value
1.1.2.2	Waveform	sinusoidal
1.1.2.3	Input power [W]	30
1.1.2.4	Peak-to-peak discharge voltage [kV]	9
1.1.2.5	Frequency [kHz]	9
1.1.3	**Source settings**	
1.1.3.1	Discharge current [A]	0.01
1.1.3.2	Dissipated power [W]	25
1.1.3.3	Specific energy input [J/cm^3]	300
1.2.1	**Standardization procedures**	
1.2.1.1	Waiting time [min]	5
2.1.1	**Gas properties**	
2.1.1.1	Gas description	The plasma medium (gas) was ambient air without any further treatment or precursors.
2.1.1.2	Pressure [Pa]	1.01×10^5
2.1.1.3	Flow rate [L/min]	4
2.1.1.4	Carrier gas composition	NA
2.1.1.5	Purity [%]	NA
2.1.1.6	Temperature [°C]	35
2.1.1.7	Humidity [% RH]	40

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

3.4.2 Plasma target-related parameters

3.4.2.1 Food-related parameters

In Table 3.8, the matrix (solid food)-related parameters and the target treatment procedures are presented, as sub-schemes of the plasma target-related parameters. In this example, a solid food matrix (black peppercorn) that is treated in a direct mode of treatment was simulated, but the table could also be adapted to describe the information of other matrices, for example, an abiotic surface such as a packaging material or a food processing surface (Table 3.4). The "distance to surface" was reported as "NA" (not applicable) because of the direct treatment scenario where the product is placed between the two electrodes. Also, the treated surface was not quantified in this example due to the complexity of the food product (multiple spheres being treated simultaneously) which would require further diagnostics for precise assessment. This highlights the flexibility of the presented metadata schema, allowing adaptation based on the resources available in each laboratory.

Table 3.8: An example of accurately describing the food-related parameters and the target treatment procedures (plasma target-related parameters) in non-thermal plasma decontamination studies

Id	Field name	Value
3.1.1	**Food-related parameters**	
3.1.1.1	Matrix description	The target matrix was black peppercorns bought from a local supermarket and sterilized before treatment.
3.1.1.2	Distance to surface [cm]	NA
3.1.1.3	Surface temperature [°C]	30
3.1.1.4	Weight [g]	5
3.1.1.5	Food shape	spherical
3.1.1.6	Dimensions [cm]	~0.5 diameter
3.1.1.7	Treated surface area [cm^2]	NA
3.1.1.8	a_w	0.6
3.1.1.9	pH	5.5
3.1.1.10	Fat percentage [%]	3
3.1.1.11	Protein content [%]	10
3.2.1	**Target treatment**	
3.2.1.1	Mode of treatment [direct, indirect]	direct
3.2.1.2	Spatial location relative to the plasma source	The peppercorns were placed directly between the two electrodes of the DBD device. The product was exposed simultaneously to plasma-generated reactive species, charged particles, and UV irradiation emitted uniformly from the electrode surfaces.
3.2.1.3	Target treatment time [min]	5

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

3.4.2.2 Microbial-related parameters

In Table 3.9, an example of the microbial-related parameters is presented, as a subscheme of the plasma target-related parameters. In this example, a vegetative bacterium, i.e., *Salmonella* Typhimurium, is simulated but the table could also be adapted to report the information of other microorganisms, for example, *Listeria* cells in a biofilm or *Bacillus* spores (Table 3.5). Regardless of the microbial form, ensuring FAIR data in microbial-related parameters requires that preculture conditions are clearly defined and thoroughly discussed. This example was based on the assumption that the product of interest, i.e., black peppercorn, was sterilized before the inoculation with *Salmonella*, and then treated with non-thermal plasma.

Table 3.9: An example of accurately describing the microbial-related parameters (plasma target-related parameters) in non-thermal plasma decontamination studies

Id	Field name	Value
3.1.2	**Microbial-related parameters**	
3.1.2.1	Microbial form	Vegetative bacteria

Id	Field name	Value
3.1.2.2	Genus	*Salmonella*
3.1.2.3	Species	*enterica*
3.1.2.4	Serotype	Typhimurium
3.1.2.5	Single- or multi-culture	single-culture
3.1.2.6	Strain	LMG 14933
3.1.2.7	Growth medium before and after treatment	From -80 °C, the stock culture was grown in Tryptic Soy Agar (TSA) for 48 h and then a single colony was inoculated to Tryptic Soy Broth (TSB). TSA was used for the enumeration after treatment.
3.1.2.8	Incubation time before treatment [h]	24
3.1.2.9	Incubation temperature [°C]	37
3.1.2.10	Growth stage	stationary
3.1.2.11	Mode of inoculation	mixing
3.1.2.12	Inoculation procedure	1 ml of the initial inoculum of known concentration was mixed with the treatment matrix and let dry for 30 min at room temperature.
3.1.2.13	Microbial counts before treatment [$\log_{10}$ CFU/g]	8.0
3.1.2.14	Enumeration / quantification procedure	Samples were immediately analyzed after treatment through homogenization with physiological salt solution (PPS) followed by serial dilution. A 0.1 ml aliquot of each dilution was plated on TSA, and incubated.
3.1.2.15	Incubation time after treatment [h]	48
3.1.2.16	Microbial counts after treatment [$\log_{10}$ CFU/g]	1.5

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

While the example provided is comprehensive, more complex experimental setups would require additional parameters. For instance, if the microbial form were spores, then the sporulation protocol would also need to be reported (Table 3.5).

3.4.3 Plasma diagnostics-related parameters

In Table 3.10, an example of the plasma diagnostics-related parameters is presented. In this example, Fourier Transform Infrared Spectroscopy (FTIR) was used to analyze reactive oxygen and nitrogen species (RONS) generated directly in the gas phase within the discharge region, immediately above the peppercorn surface.

Table 3.10: An example of accurately describing the plasma diagnostics-related parameters in non-thermal plasma decontamination studies

Id	Field name	Value
4.1.1	**Diagnostic settings**	
4.1.1.1	Setting description	RONS in the plasma-processed medium were determined with FTIR spectroscopy with the following settings: absorption path length 0.25 m, resolution 0.6 cm^{-1}, wavenumber range of 4000 – 400 cm^{-1}, 64 scans in each measurement.
4.1.2	**Environmental parameters**	

Id	Field name	Value
4.1.2.1	Temperature [°C]	22
4.1.2.2	Humidity [% RH]	40
4.1.2.3	Pressure [Pa]	1.01×10^5
4.2.1	**Diagnostic methodology**	
4.2.1.1	Methodology description	After the plasma discharge had stabilized, FTIR measurements of reactive species were performed in the gas phase adjacent to the peppercorns placed between the electrodes, using a device (specify brand and model).
4.2.1.2	Location of the diagnostic application	The plasma gas phase was sampled 1 cm above the peppercorn surface, within the discharge region between the two electrodes.
4.2.1.3	Targets of qualitative / quantitative analysis	Hydrogen peroxide (H_2O_2), Ozone (O_3), Nitrate (NO), Nitrite (NO_2). FTIR spectroscopy, specific peak centers: 1400 cm^{-1} and 876 cm^{-1} for H_2O_2; 1050 cm^{-1} for O_3; 1875 cm^{-1} for NO; and 1600 cm^{-1} and 2900 cm^{-1} for NO_2.
4.2.1.4	Quantification method	Calibration curves based on known standards for H_2O_2, O_3, NO, NO_2.
4.2.1.5	H_2O_2 [mol/m^3]	1.6×10^{-4}
4.2.1.6	O_3 [mol/m^3]	2.04×10^{-3}
4.2.1.7	NO [mol/m^3]	1.2×10^{-4}
4.2.1.8	NO_2 [mol/m^3]	4.08×10^{-5}
4.2.2	**Data processing**	
4.2.2.1	Data processing description	Spectral deconvolution and baseline correction using a data analysis software (specify), and Savitzky-Golay smoothing. Peaks were identified via an atomic spectra database (specify).

Note: Bold rows denote parent categories; regular rows denote subordinate fields within those categories.

It should be noted that although FTIR can characterize the plasma-generated reactive species that were mentioned in Table 3.10, other plasma-induced effects, such as UV irradiation may require additional diagnostic techniques for a comprehensive assessment.

3.5 Additional considerations for better data descriptions

Besides the specific challenges related to describing parameters in non-thermal plasma research, as discussed in Section 3.2.2, it is important to note that the principle of "findable" data can be challenged when authors reference one or more previously published studies to describe certain experimental parameters. Although in general this approach is considered valid to avoid repetition, it can be problematic when one or more experimental parameters are already different when compared to the study of interest because it becomes then unsure if also other parameters have changed over time or not. Therefore, it is advisable for authors to clearly define the reported parameters and specify precisely which aspects remain the same and which differ from previous research.

The parameters discussed in this study can set the basis for researchers to determine which information is critical to document in non-thermal plasma decontamination experiments for food-related applications, and how this information can be effectively described. However, these parameters can be expanded to cover certain applications if needed. A practical way to facilitate process validation and regulatory approval is to distinguish between parameters that directly affect plasma chemistry and those that are more device-specific. In this view, environmental and operational conditions such as humidity, temperature, and pressure are different from factors such as electrodes' material or geometry that are essential to replicate the setup but are not always directly comparable across different systems. This approach can help identify and quantify the most critical processing conditions (e.g., a "dose") in a standardized manner. Nevertheless, the metadata schema that is presented in this study covers all use cases either when comparing the same or different systems since the relevance of each parameter (and the determination of the "dose") highly depends on the context. The more comprehensively the data are documented, the broader the range of potential reuse scenarios, ultimately bringing non-thermal plasma applications closer to regulatory approval. That being said, we acknowledge that not all parameters, such as plasma diagnostics-related parameters, are equally accessible and can be limited by expensive methods and the need for trained personnel. Therefore, in this study, we provide definitions and guidance on parameters ranging from easily accessible to more complex, to assist researchers in clearly and accurately describing their experimental data, regardless of their laboratory equipment. In this way, parameters that are generally more easily accessible to researchers could be used by other researchers as indirect methods of plasma diagnostics through the estimation of more complex metrics. For example, as mentioned in Section 3.3.2, the dimensions and geometry of the electrodes (e.g., in DBD devices), when clearly defined, can be used to calculate their respective areas, volumes, and other relevant physical properties depending on the shape. These properties can then be combined with the plasma source-related parameters (e.g., the dissipated power and the processing time) and be used as metrics for comparison of the decontamination efficacy (Pampoukis et al., 2024).

3.6 Conclusion

The purpose of this study is to provide a metadata schema to guide the consistent and traceable descriptions of the relevant parameters and to improve the interoperability, reproducibility, and reuse within the non-thermal plasma community for food-related applications. This would allow for more accurate overviews, and comparisons of the results in the literature, for example in the context of systematic reviews and meta-analyses. With this study, the Plasma-MDS platform that was previously developed

(Franke et al., 2020) is given a specific context for decontamination in food-related applications, to become easier to implement. That is because it provides specific guidance with examples regarding what is expected to be reported for each case while still maintaining some flexibility for additional or less information. In this way, researchers without specific expertise in non-thermal plasma can still conduct more reproducible research. Similarly, non-thermal plasma experts can use this metadata schema to accurately describe relevant food- and microbial-related parameters. From a data management perspective, like all schemes, Plasma-MDS will benefit from continuous updates as knowledge advances in the field of non-thermal plasma and new applications emerge. This will ensure its relevance and effectiveness in capturing essential information for future applications. The parameters provided in this study are expected to also be applicable to other non-thermal plasma applications.

CRediT authorship contribution statement

George Pampoukis: Conceptualization, Methodology, Software, Investigation, Data Curation, Writing - Original Draft, Visualization. **Thomas Weihe**: Conceptualization, Methodology, Writing - Original Draft. **Robert Wagner**: Conceptualization, Methodology, Software, Writing - Review & Editing, Visualization. **Markus M. Becker**: Conceptualization, Methodology, Resources, Writing - Review & Editing. **Yijiao Yao**: Writing - Review & Editing. **Masja Nierop Groot**: Writing - Review & Editing, Supervision, Project administration. **Uta Schnabel**: Conceptualization, Methodology, Resources, Writing - Original Draft, Supervision, Project administration.

Acknowledgements

We thank Dr. Jörg Ehlbeck (Leibniz Institute for Plasma Science and Technology (INP)) for valuable discussions and suggestions regarding the need for harmonization and standardization of data in the non-thermal plasma field. We also thank Prof. Heidy M. W. den Besten (Food Microbiology, Wageningen University) and Prof. Marcel H. Zwietering (Food Microbiology, Wageningen University) for their critical review of the manuscript which significantly improved its clarity and readability.

Code availability

No custom code or algorithms were used to generate or process the data described in this manuscript.

Data availability

The metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications is publicly available in a machine-readable JSON format on GitHub: https://github.com/gpampoukis/plasma-food-metadata-schema (Pampoukis, 2025). The repository includes both the complete metadata schema and the case study presented in this manuscript.

Funding

The projects of George Pampoukis and Yijiao Yao are part of the TRANSIT project that received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N° 955431. Additionally, funding has been received from the Federal Republic of Germany, Federal Ministry of Education and Research under the program "PlaVir", and funding reference: 03COV05A. Further funding for the work has been provided by the Federal Republic of Germany and the country's State Mecklenburg-Western Pomerania, Ministry of Science, Culture, Federal and European Affairs under the grant agreement N° VIII-0639-INP00-2023/004-002.

Chapter 4

The decontamination efficacy of pulsed electric field in food: Establishing global kinetic parameters

George Pampoukis, Jan Ham, Marcel H. Zwietering, Heidy M.W. den Besten

Food Microbiology, Wageningen University & Research, Wageningen, the Netherlands

Published in *International Journal of Food Microbiology*, Volume 447, 16 February 2026, 111547. https://doi.org/10.1016/j.ijfoodmicro.2025.111547

Abstract

Pulsed electric field (PEF) is a promising technology for microbial decontamination of liquid products such as fruit juices and beverages. This study developed a database with 569 PEF experiments for *Escherichia coli*, *Listeria* spp., *Pseudomonas* spp., *Staphylococcus aureus*, *Saccharomyces cerevisiae*, *Salmonella* spp., and lactic acid bacteria (LAB) in a range of food matrices. The efficacy of PEF for different microbial species was evaluated through the decimal-reduction volumetric energy input $D_{E_{\text{vol}}}$, which is the energy input required to achieve a one $\log_{10}$-reduction. The matrix composition had a significant impact on the $D_{E_{\text{vol}}}$, decreasing the $D_{E_{\text{vol}}}$ in high-acid fruit juices and increasing it in high-acid alcoholic products, revealing effects beyond acidity alone. Data for most species were available for high-acid fruit juices, and in this matrix, *S. cerevisiae* was the most sensitive species. A decision tree with eight matrix categories was developed to estimate the volumetric energy input needed for a 5-$\log_{10}$-reduction for different target microbial species in different matrices. The global kinetic parameters reported in this study provide a quantitative benchmark for food business operators and risk assessors, and can serve as a first indication of PEF treatment efficacy.

4.1 Introduction

Assurance of food safety and quality remains a fundamental goal of the food-processing sector. The increasing demand for foods with high nutritional value, a clean label, and fresh-like sensory characteristics led to the development of mild food treatment processes to enhance or substitute conventional techniques such as heat treatment and chemical preservatives (Toepfl et al., 2005). Pulsed electric field (PEF) has therefore gained more attention as a mild processing technology that can retain "fresh-like" sensory and nutritional properties. PEF is powered by electricity that can be generated from renewable sources, and avoids natural gas boilers and steam generators, which reduces wastewater in food plants (Pereira and Vicente, 2010). However, life cycle assessments indicate higher CO_2 emissions than thermal processing, because of greater electricity consumption and limited energy-recovery options, yet lower emissions than high-pressure processing (Aganovic et al., 2017; Aganovic and Smetana, 2022; Sampedro et al., 2014). Compared with other alternative processing technologies, such as high-pressure processing, PEF operates in continuous flow, making it attractive for industrial scale-up (Andreou et al., 2016; Lytras et al., 2024a; Niu et al., 2020). However, its use is restricted to liquid products such as fruit juices and beverages.

When PEF raises the transmembrane potential significantly, irreversible membrane permeabilization will impair vital physiological control systems and cause cell death (Heinz et al., 2001). In practice, irreversible microbial membrane permeabilization is usually obtained with electric field strengths of 15-40 kV/cm (Raso et al., 2016), total pulse durations of 20-2000 μs (Timmermans et al., 2019), and specific energy inputs of 5-250 kJ/kg (Toepfl, 2011). In the United States, based on the Code of Federal Regulations (CFR) for HACCP systems (21 CFR Part 120) (U.S. Food and Drug Administration, 2001), juice processors must demonstrate that at least a 5-$\log_{10}$-reduction of the pertinent pathogen is achieved, immediately before or after packaging (U.S. Food and Drug Administration, 2004), making this a starting reference for process validation and to determine the process parameters that are needed to achieve it.

A meta-analysis is a systematic analysis of a large collection of data from individual studies aiming to integrate the findings to produce a global estimate of the effect of a particular intervention or treatment (Den Besten and Zwietering, 2012). In food microbiology, meta-regression has been applied to derive global kinetic parameters for thermal processing (Halder et al., 2010; Muñoz-Cuevas et al., 2012; Van Asselt and Zwietering, 2006) and high-pressure processing (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Santillana Farakos and Zwietering, 2011; Torrents-Masoliver et al., 2025), and to rank factors affecting non-thermal plasma decontamination (Pampoukis et al., 2024). By combining data from multiple laboratories, meta-regression

captures variability arising from differences in strains, matrices, and equipment, and yields parameter estimates that single-laboratory experiments cannot provide (Garre et al., 2025). In this way, the "study" effects are averaged, and the insights are based on the available evidence. Nevertheless, a meta-regression quantifying PEF decontamination efficacy across microbial species and food matrices has not yet been performed.

Therefore, the aims of this work were to (i) compile available PEF inactivation data of different microbial species in diverse foods, (ii) establish global and matrix-specific kinetic parameters through meta-regression, and (iii) construct a decision tree that estimates the volumetric energy input required for a 5-$\log_{10}$-reduction, thereby supporting food-business operators, risk assessors and regulators.

4.2 Materials and methods

4.2.1 Data collection & active learning

To assure FAIR (findable, accessible, interoperable, and reusable) data, we defined explicit eligibility criteria and a three-block searching query. No restriction on publication year was applied. The searching query consisted of (i) pulsed-electric-field descriptors (e.g., "electric field", "PEF", etc.), (ii) processing terms (e.g., "processing", "treatment", etc.), and (iii) the taxonomic name of the target microorganism (e.g., *E. coli*, *Escherichia*, etc.). Eligible studies had to involve continuous flow processing (industrial applications require continuous flow (Raso et al., 2016)), report a microbial reduction > 0.5-$\log_{10}$, and provide the variables needed to calculate the volumetric energy input, which was used in this study as an integrated metric for decontamination efficacy comparisons. Literature searches were performed in Scopus, Web of Science, SciFinder, FSTA & CAB Abstracts for *E. coli*, *Listeria* spp., *Pseudomonas* spp., *S. aureus*, *S. cerevisiae*, *Salmonella* spp., *Bacillus* spp., and lactic acid bacteria (LAB). These microbial species were selected because they cover Gram-negative and Gram-positive vegetative cells, spore-forming bacteria, yeasts, and pathogenic, spoilage, fermentation, and hygiene indicator microorganisms. To build meta-regression models, sufficient, suitable data must be available in the literature, which is not always the case for all microbial species. A PRISMA flow diagram (Page et al., 2021) with detailed eligibility criteria is provided in Supplementary Material A (section "PRISMA flow diagram for data collection and eligibility criteria"), together with the complete search queries and the data-extraction protocol. Each microbial genus was searched independently; studies were then merged, and duplicates were removed. The studies were screened as relevant or irrelevant with support from active learning (Settles, 2009) using the ASReview software (Van de Schoot et al., 2021). After an initial set of studies had been

decided as relevant or irrelevant by the reviewer, the software (TF-IDF features + Naive Bayes classifier + dynamic resampling) continuously retrained on the accumulated decisions and re-ordered the remaining studies so that studies with the highest predicted relevance were presented first for review by the researcher. The screening procedure was terminated when 200 consecutive irrelevant studies were encountered, indicating near-exhaustive retrieval of relevant studies. The approach of active learning was explained in detail in previous research (Pampoukis et al., 2024). Out of the 2156 studies retrieved from the literature, 78 were relevant and contributed 1725 $\log_{10}$-reduction values to the database. To maintain industrially realistic conditions, studies with volumetric energy inputs above 300 kJ/L were excluded, given that process parameters suitable for PEF commercial scale-up typically fall within 5-250 kJ/kg (Toepfl, 2011). Mild thermal effects may already occur above ~200 kJ/kg: assuming an average liquid specific heat capacity C_p ~4 kJ/(kg·°C), this corresponds to an estimated temperature rise of about 50 °C over a short residence time (μs to ms), especially if no cooling is applied. Information on outlet temperatures and the use of cooling is provided in Supplementary Table 4.1. The highest temperature recorded in our database was 64 °C, observed in a single study with a short residence time of 0.43 s before cooling to room temperature. To ensure comparable pulse conditions and more accurate volumetric energy input estimations, only studies employing square-wave pulses were retained; studies using exponential decay pulses were discarded because the pulse width and, therefore, the volumetric energy input cannot be accurately estimated from the PEF settings, as will be explained in the following sections. Square wave pulses are also considered to be the most energy-efficient form of PEF delivery (Beveridge et al., 2005). Application of these filters yielded a curated set of 569 $\log_{10}$-reduction values for *E. coli*, *Listeria* spp., *Pseudomonas* spp., *S. aureus*, *S. cerevisiae*, *Salmonella* spp., and lactic acid bacteria (*Lactiplantibacillus*; *Lacticaseibacillus*; *Lactococcus*; *Leuconostoc*; *Oenococcus* and *Pediococcus*), derived from 28 studies; no eligible data were found for *Bacillus* spp. (vegetative cells or spores). In general, inactivation of bacterial spores, such as *Bacillus* spores, can only be achieved when PEF is combined with substantial thermal treatment, with outlet temperatures typically exceeding 100 °C (Siemer et al., 2014).

4.2.2 Matrix classification

Once microbial targets and PEF-processing boundaries had been standardized, each experimental condition was annotated with product type and intrinsic acidity. To evaluate food-matrix effects while retaining statistical parsimony, all 569 data points were grouped into eight categories that combine pH category (high-acid ≤ 4.5, low-acid > 4.5) with product type. The resulting groups were: (i) high-acid fruit juice (apple juice, model solution simulating strawberry juice, orange juice, pomegranate juice); (ii) low-

acid fruit juice (Kyoho grape juice); (iii) high-acid liquid medium (McIlvaine buffer, phosphate buffer, both with pH adjusted to ≤ 4.5); (iv) low-acid liquid medium (0.1% NaCl water solution, *Chlorella vulgaris* suspension, citrate-phosphate buffer, deionised water, McIlvaine buffer, phosphate buffer, tryptic soy broth with yeast extract); (v) low-acid animal milk (raw goat milk, skim milk); (vi) low-acid plant-based beverage (green tea extract); (vii) high-acid whey (50% acid whey) and (viii) high-acid alcoholic product (Chardonnay, Grenache, Tempranillo wines). This classification allowed for developing quantitative models that capture not only acidity but also physicochemical factors that differ among juices, broths, milk, plant-based beverages, whey, and alcoholic products, such as soluble solids in the fruit-juice categories, proteins and fats in milk and whey, polyphenols in the plant-based beverages, and ethanol in the alcoholic products. Grouping by category rather than by individual composition avoided the extensive missing data that would arise if explicit pH, protein, fat, ethanol, or sugar contents were required for every study, thereby retaining the complete set of 569 observations.

4.2.3 The volumetric energy input as a metric for comparisons

To standardize the data obtained from PEF studies and develop quantitative meta-regression models, the present study expresses the treatment dose as the volumetric energy input (kJ/L), a metric that integrates electric-field strength, pulse-repetition rate, pulse width, conductivity and residence time, and can be reconstructed from variables commonly reported in the literature. Therefore, we expressed the treatment dose as the volumetric energy input E_{vol} (Equation 4.1), with residence time (R_t) and volumetric flow rate (q) obtained from Equation 4.2 and Equation 4.3, respectively (Mendes-Oliveira et al., 2020).

$$E_{\text{vol}}\ [\text{J/cm}^3] \;=\; E_f^2 \cdot f \cdot \tau \cdot \sigma \cdot R_t \cdot n_c \qquad (1\,\text{J/cm}^3 = 1\,\text{kJ/L}) \tag{4.1}$$

$$R_t\ [\text{s}] \;= \frac{V}{q} \tag{4.2}$$

$$q\ [\text{cm}^3/\text{s}] \;= \frac{V \cdot f \cdot \tau \cdot n_c}{t_{\text{PEF}}} \tag{4.3}$$

where E_f is the electric-field strength (kV/cm); f is the frequency/pulse-repetition rate (pulses per second); τ is the pulse width (μs); σ is the electrical conductivity of the food at the inlet temperature (S/cm); R_t is the residence time in the PEF unit (s); n_c is the

number of PEF chambers in series (-); V is the chamber volume as estimated from its dimensions (cm³); q is the volumetric flow rate (cm³/s); t_{PEF} is the total effective treatment time (µs); and E_{vol} is the volumetric energy input (J/cm^3 or kJ/L). Equation 4.1 was used because it was the most frequent in the literature when compared to other expressions, such as those that involved the voltage and current signals. It is more accurate when compared to the calorimetric equation ($q_p = m \cdot C_p \cdot \Delta T$), which can also be used for PEF (Jaeger et al., 2010), but assumes that all supplied electricity is converted to heat within the product, and does not take into account device-specific energy losses or cooling procedures. Equation 4.1 is generally applicable because the estimations rely on variables that are widely reported and well-defined in the literature. It should be noted that in some studies, the reported energy input did not fully match this estimation due to the reasons mentioned above, and therefore, only the estimates from Equation 4.1 were used to ensure consistency and comparability across the different studies. Every data point (*n*) reported in the database corresponds to one experimental condition at a given energy input level as estimated from Equation 4.1 and the reduction achieved under this processing condition.

4.2.4 Meta-regression modeling to establish global kinetic parameters

4.2.4.1 Simple linear regression

The microbial reduction upon PEF treatment in different microbial species was assessed using Equation 4.4:

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) = \frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} \tag{4.4}$$

where N_0 is the initial microbial population (CFU/ml), N is the microbial population after treatment, E_{vol} is the volumetric energy input (kJ/L), and $D_{E_{\text{vol}}}$ is the decimal-reduction volumetric energy input (kJ/L) i.e., the energy input needed for a one $\log_{10}$-reduction. Equation 4.4 was fitted (i) to the full dataset, (ii) separately to the data of each microbial species, and (iii) to the data of every microbial species-matrix category combination. Regression models were fitted in Python (v3.11.2) (Van Rossum and Drake, 2009) using *scikit-learn* (LinearRegression(fit_intercept = False)) (Pedregosa et al., 2011), enforcing the theoretical constraint that 0 kJ/L yields 0 $\log_{10}$-reduction. For each subset, the slope $\beta = 1/D_{E_{\text{vol}}}$ was obtained by ordinary least squares, and the estimated reductions were $y_i = \beta x_i$.

The residual sum of squares (RSS) and the root-mean-square error (RMSE) were calculated as (Eisenhauer, 2003):

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(\log_{10}\left(\frac{N_0}{N(E_{\mathrm{vol}})}\right)_i - \beta \cdot E_{\mathrm{vol},i}\right)^2 \tag{4.5}$$

$$\mathrm{RMSE} = \sqrt{\frac{\mathrm{RSS}}{n-1}} \tag{4.6}$$

with n the number of observations and $n-1$ the degrees of freedom of a regression through the origin. The standard error of the slope was calculated as (Montgomery et al., 2021):

$$\mathrm{SE}_\beta = \sqrt{\frac{\mathrm{RSS}}{(n-1) \cdot \sum_{i=1}^{n} E_{\mathrm{vol},\,i}^2}} \tag{4.7}$$

The 95% confidence upper and lower limits of $D_{E_{\mathrm{vol}}}$ were obtained by inverting the t-based limits of the slope (Montgomery et al., 2021):

$$D_{E_{\mathrm{vol,low}}} = \frac{1}{\beta + t_{0.975,\,n-1} \cdot \mathrm{SE}_\beta}, \qquad D_{E_{\mathrm{vol,high}}} = \frac{1}{\beta - t_{0.975,\,n-1} \cdot \mathrm{SE}_\beta} \tag{4.8}$$

The resulting 95% confidence interval was therefore:

$$\left[D_{E_{\mathrm{vol,low}}},\ D_{E_{\mathrm{vol,high}}}\right] \tag{4.9}$$

The mean estimate $D_{E_{\mathrm{vol}}}$, its 95% confidence interval (Equation 4.9), and the RMSE of the regression (Equation 4.6) were reported for every model to compare goodness-of-fit and variability among microbial species and matrix categories.

To estimate the expected reduction of microbial species in different matrix categories at a reference energy input level, the fitted model (Equation 4.4) was used to estimate the reductions at $E_{\mathrm{vol}} = 100$ kJ/L:

$$\log_{10}\left(\frac{N_0}{N(100)}\right) = \frac{100}{D_{E_{\mathrm{vol}}}} \tag{4.10}$$

The 95% confidence interval for the estimated microbial reduction at $E_{\text{vol}} = 100$ kJ/L was (Montgomery et al., 2021):

$$\log_{10}\left(\frac{N_0}{N(100)}\right) \pm t_{0.975,\,n-1} \cdot RMSE \cdot \sqrt{\frac{100^2}{\text{SSX}}} \tag{4.11}$$

where $\log_{10}(N_0/N(100))$ is the expected reduction as estimated from Equation 4.10, $t_{0.975,\,n-1}$ is the 97.5% quantile of the Student-t distribution with n-1 degrees of freedom, RMSE is the root mean square error, and SSX $= \sum_{i=1}^{n} E_{vol,i}^2$. These limits were used to report the uncertainty of the predicted $\log_{10}$-reduction at 100 kJ/L for every species-matrix combination.

4.2.4.2 Multiple linear regression

To determine the effect of all food matrices simultaneously, multiple linear regression was performed for each microbial species using *statsmodels* (Seabold and Perktold, 2010) to fit Equation 4.12. With the intercept fixed at zero, matrix effects were encoded by multiplying each matrix dummy with E_{vol}. This strategy allows matrix-specific slopes while preserving the mechanistic requirement that $E_{\text{vol}} = 0$ must produce no microbial inactivation. Therefore, the multiple linear regression equation was defined as:

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) = E_{\text{vol}} \cdot \sum_{j=1}^{8} \frac{z_j}{D_{E_{\text{vol}}j}}, \qquad z_j \in \{0,1\} \tag{4.12}$$

where E_{vol} is the volumetric energy input (kJ/L), $z_j \in \{0,1\}$ is the dummy indicator for matrix category j (eight categories; Section 4.2.2), and $D_{E_{\text{vol}}j}$ is the matrix category-specific decimal-reduction energy.

The fitted D-values exactly match those from the simple linear models per matrix category (Section 4.2.4.1). The multiple-linear regression fit, however, estimates a single (pooled) residual variance across the matrix categories for a given microbial species and yields matrix-specific confidence intervals based on that pooled error term. This enables direct statistical contrasts across matrices from a single fit. Separate per-matrix regressions can also provide intervals, but they do not share a common error term unless one is explicitly pooled. All parameter estimates were reported as $D_{E_{\text{vol}}j}$ with 95% confidence intervals computed from the pooled-error t distribution as:

$$\left[\frac{1}{\beta_j + t_{0.975,\,n-p} \cdot \mathrm{SE}_{\beta_j}},\ \frac{1}{\beta_j - t_{0.975,\,n-p} \cdot \mathrm{SE}_{\beta_j}}\right] \tag{4.13}$$

where (β_j) is the slope for matrix (j) (so $D_{E_{\mathrm{vol}}j} = 1/\beta_j$), ($n$) is the total number of observations for that species, and (p) is the number of estimated matrix coefficients for that species (i.e., the number of matrix categories with data; ($p \leq 8$)), giving residual degrees of freedom ($n - p$). To test whether any two matrix categories differed in resistance, pairwise two-sided t tests were performed in the $\beta_j = 1/D_{E_{\mathrm{vol}}j}$ for every matrix category with available data within each species. The p-values provided from the two-sided t tests were adjusted for multiple hypothesis testing with both Bonferroni and Holm-Bonferroni corrections for multiple testing (Aickin and Gensler, 1996); a difference was deemed significant when both adjusted p-values were statistically significant at a significance level $\alpha = 0.05$ (p-value ≤ 0.05). Because the multiple- and simple-linear regression $D_{E_{\mathrm{vol}}}$ values coincide, subsequent analyses used the simpler form unless explicit matrix contrasts were required.

4

4.2.4.3 Energy input required for a 5-$\log_{10}$-reduction

The volumetric energy input for a 5-$\log_{10}$-reduction was obtained by inverting the simple-regression model for each microbial species-matrix category combination.

Therefore, the forward 95% prediction limits for a 5-$\log_{10}$-reduction at volumetric energy input E_{vol} were estimated as (Montgomery et al., 2021):

$$\frac{E_{\mathrm{vol}}}{D_{E_{\mathrm{vol}}}} \ \pm\ t_{0.975,\,n-1} \cdot RMSE \cdot \sqrt{1 + \frac{E_{\mathrm{vol}}^2}{\mathrm{SSX}}} = \log_{10}\left(\frac{N_0}{N(E_{\mathrm{vol}})}\right) = 5 \tag{4.14}$$

where E_{vol} is the energy input needed to achieve a 5 $\log_{10}(N_0/N(E_{\mathrm{vol}}))$-reduction given a $D_{E_{\mathrm{vol}}}$ estimated from fitting Equation 4.4 for this microbial species-matrix category combination, $t_{0.975,\,n-1}$ is the 97.5% quantile of the Student-t distribution with n-1 degrees of freedom, RMSE is the root mean square error, and SSX equals to $\sum_{i=1}^{n} E_{vol,i}^2$. Setting $\log_{10}(N_0/N(E_{\mathrm{vol}})) = 5$ and solving the quadratic for E_{vol} gives the best-case ($E_{\mathrm{vol,low}}$) and worst-case ($E_{\mathrm{vol,high}}$) levels of volumetric energy input to achieve a 5-$\log_{10}$-reduction. The interval $[E_{\mathrm{vol,low}}, E_{\mathrm{vol,high}}]$ was reported as the 95% inverse-prediction interval. A worked example of this inversion procedure is provided in Supplementary Material B.

4.3 Results

4.3.1 The overview of the datapoints

The microbial reductions collected in the database are summarized in Figure 4.1. For all microbial species, the decimal microbial reductions showed an increasing trend with the volumetric energy input, but this trend was different across the different microbial species and matrix categories.

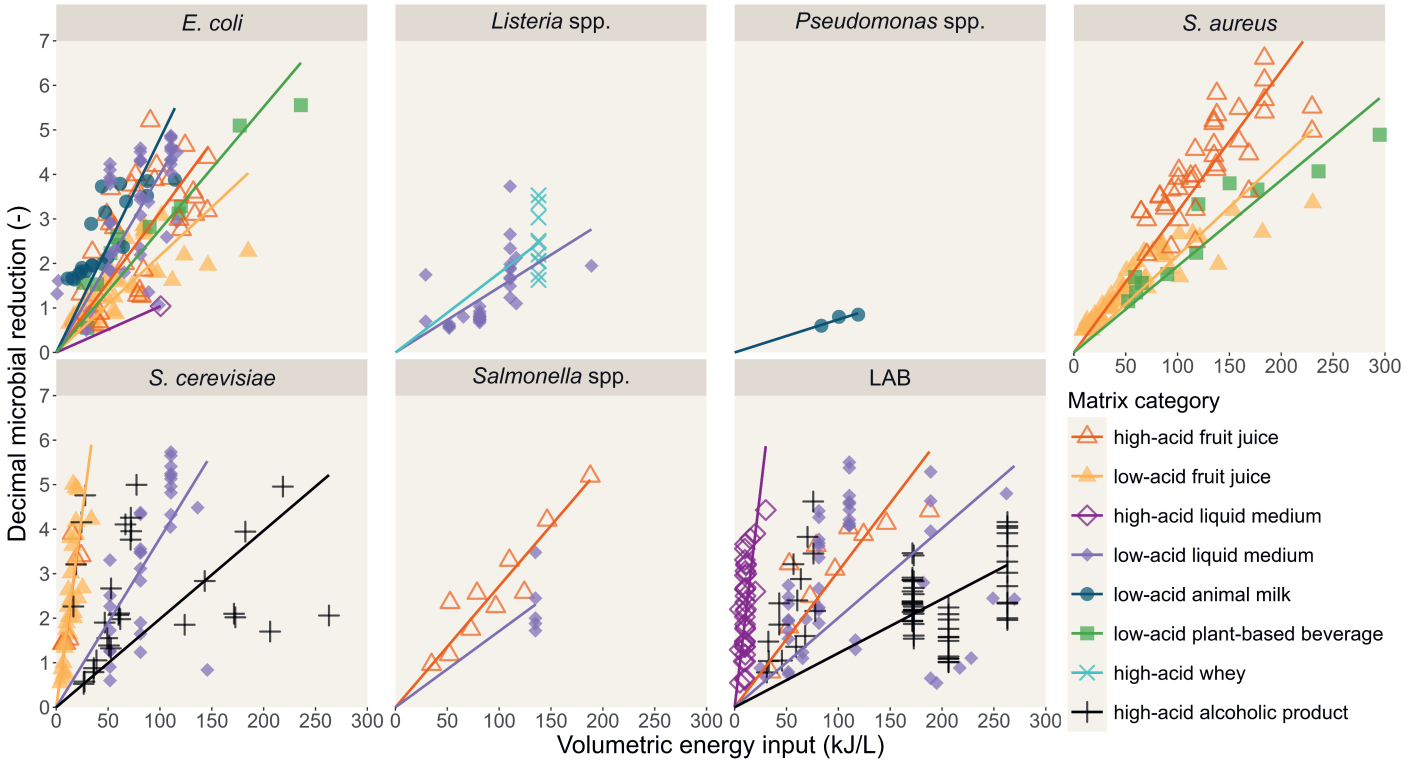


Figure 4.1: Overview of the datapoints (n = 569, studies = 28). Solid lines show no-intercept linear fits for each microbial species-matrix category combination.

Although high-acid matrix categories (fruit juice, liquid medium) differed from their low-acid counterparts in most cases, the patterns were not consistent across all microbial species, indicating that factors beyond acidity influence PEF decontamination. In high-acid matrix categories, excluding high-acid alcoholic products, *S. aureus* and LAB exhibited greater reductions when compared to low-acid matrix categories, whereas *S. cerevisiae* revealed no clear acidity effect (Figure 4.1). For *E. coli*, the acidity effect was only observed in fruit juice. For LAB, the different matrix categories had a wide range for inactivation, especially for the matrix categories of low-acid liquid medium and high-acid alcoholic product. These two matrix categories showed higher variability in the inactivation across all microbial species. For high-acid whey, all data points collected were corresponding to one microorganism (*Listeria* spp.) at a single energy input level, and therefore, the trend line has less explanatory value. The same limitation also constrained *Salmonella* spp. in low-acid liquid medium and *E. coli* in high-acid liquid medium fits, and the latter was only based on one data point. A user-friendly, interactive version of Figure 4.1 is provided in https://github.com/gpampoukis/pef-decontamination-efficacy-database (https://pef-database.onrender.com/), allowing readers to inspect each data point with its source study, matrix description, and -where reported- the exact pH and microbial strain tested. The quantitative effects of microbial species and matrix category are described in the following sections.

4.3.2 $D_{E_{\text{vol}}}$ for each microbial species

In Table 4.1, the volumetric energy input $D_{E_{\text{vol}}}$ values obtained from 569 experimental conditions of 28 studies are presented. Because each estimate pools all eligible data for the different microbial species, these $D_{E_{\text{vol}}}$ values can serve as global kinetic parameters across all matrix categories.

Table 4.1: Overall linear models for studies with volumetric energy input ≤ 300 kJ/L ($n = 569$, studies = 28). The number of studies (28) is lower than the sum of individual categories because some studies included more than one microbial species

Microbial species	$D_{E_{\text{vol}}}$ [95% CI] (kJ/L)	RMSE (log 10 -counts)	n	studies
E. coli	31 [29, 33]	1.1	153	16
Listeria spp.	63 [56, 71]	0.65	40	4
Pseudomonas spp.	135 [117, 160]	0.047	3	1
S. cerevisiae	34 [29, 41]	2.1	98	9
Salmonella spp.	44 [38, 53]	0.81	15	2
S. aureus	37 [35, 40]	0.90	104	4
LAB	66 [58, 76]	1.8	156	10

It should be noted that the LAB dataset integrated 6 genera, i.e., *Lactiplantibacillus* (n = 106); *Lacticaseibacillus* (n = 2); *Lactococcus* (n = 9); *Leuconostoc* (n = 4); *Oenococcus* (n = 23) and *Pediococcus* (n = 12). Additionally, the *Pseudomonas* spp. estimate was based on only 3 data points coming from a single study and therefore, although presented, it was not as robust as the aforementioned estimates and was not further processed in the next sections.

4.3.3 Multiple linear regression as a method to determine the effect of the food matrix

Although the $D_{E_{\text{vol}}}$ values shown in Table 4.1 can provide the overall trends of how the different microbial species are inactivated with PEF at a given volumetric energy input using all data collected from literature, they also include all different sources of variability, such as the effect of the food matrix. In order to get more precise estimations of the inactivation kinetics in specific food matrices, multiple linear regression was performed. The effect of the matrix category on the different microbial species' $D_{E_{\text{vol}}}$ values is presented in Figure 4.2. *Pseudomonas* spp. was excluded because, as shown in Figure 4.1, it had only 3 data points available, all in (low-acid) animal milk. In Figure 4.2, matrix-specific attributes for each microbial species can be evaluated, e.g., the contribution of acidity can be visualized by comparing high- and low-acid counterparts of the same matrix category alongside other matrix attributes.

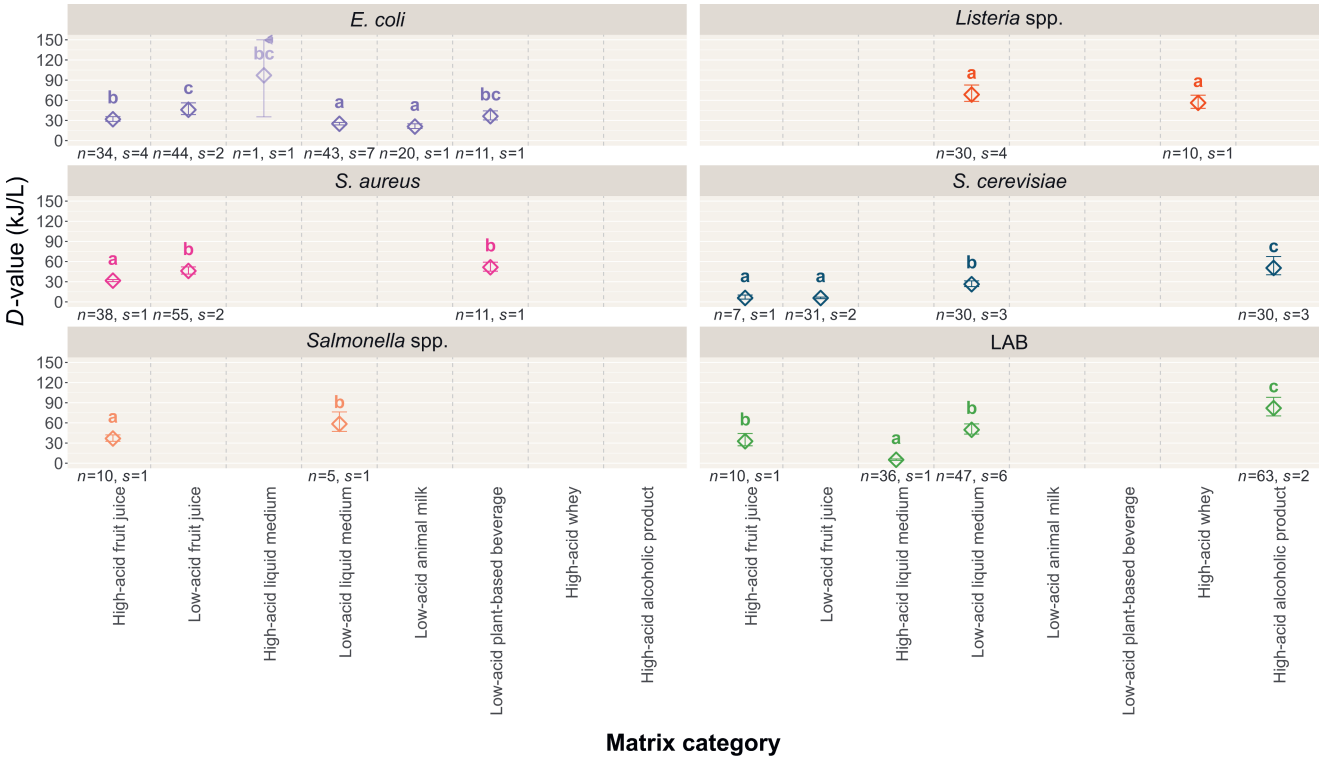


Figure 4.2: The effect of matrix category on the inactivation of the different microbial species. The "n" and "s" values correspond to the number of data points (experiments) and studies for each combination of microbial species-matrix category, respectively. The letters above each coefficient in each panel indicate statistically significant differences within that species at a significance level $\alpha = 0.05$ (p-value ≤ 0.05). The coefficient for *E. coli* in high-acid liquid medium is shaded because its 95% confidence interval of the slope included 0, leaving the upper bound of the corresponding D-value indeterminate.

For *E. coli*, the low-acid animal milk and the low-acid liquid medium were the matrix categories where the inactivation was higher with $D_{E_{vol}}$ < 30 kJ/L, followed by the high-acid fruit juice and the low-acid plant-based beverage categories. Inactivation of *E. coli* in low-acid plant-based beverages was in-between the high-acid and low-acid fruit juice categories but not significantly different from either of them. A single data point was available for the high-acid liquid medium category; although its point estimate suggested a higher $D_{E_{vol}}$, the 95% confidence interval for the slope included zero, and the upper bound of its $D_{E_{vol}}$ could not be defined. *Listeria* spp. showed comparable reductions in low-acid liquid medium and high-acid whey. *S. aureus* was most susceptible in high-acid fruit juice, with higher $D_{E_{vol}}$ values in low-acid fruit juice and low-acid plant-based beverages. For *S. cerevisiae*, the order from lower to higher $D_{E_{vol}}$ was high-acid fruit juice and low-acid fruit juice < low-acid liquid medium < high-acid alcoholic products. *Salmonella* spp. were more effectively inactivated in high-acid fruit juice than in low-acid liquid medium. LAB exhibited the highest inactivation in high-acid liquid medium, intermediate reductions in high-acid fruit juice and low-acid liquid medium, and the lowest reductions in high-acid alcoholic product (Figure 4.2). In terms of acidity effects, the high-acid matrix categories revealed a clear difference for *E. coli* and *S. aureus* (fruit juice), LAB (liquid medium), and *Salmonella* spp. (juice vs medium), whereas the $D_{E_{vol}}$ values for *Listeria* spp. and *S. cerevisiae* overlapped across acidity classes, indicating little or no acidity effect for these two species. Across the six species, the lowest $D_{E_{vol}}$ values were found in high-acid liquid medium for LAB; both in high- and low-acid fruit juice for *S. cerevisiae*; in low-acid animal milk and low-acid liquid medium for *E. coli*, and in high-acid fruit juice for *S. aureus* and *Salmonella* spp. High-acid alcoholic products consistently yielded the highest $D_{E_{vol}}$ values (for *S. cerevisiae* and LAB). Low-acid plant-based beverage, low-acid fruit juice, low-acid liquid medium, and high-acid whey categories yielded intermediate and species-dependent $D_{E_{vol}}$ values.

Multiple linear regression was also performed in the matrix categories datasets using the microbial species as a dummy variable instead, in order to rank the microbial species in each matrix category (Supplementary Figure 4.1). This representation makes the between-species differences within a given matrix visible (e.g., contrasts between Gram-negative and Gram-positive bacteria), although future studies could formalize such comparisons by explicitly grouping species on this or other biological criteria. The $D_{E_{vol}}$ values were identical to those in Figure 4.2 for all microbial species-matrix category combinations, but the uncertainty around the estimates was different based on the data points that each matrix category had available. *S. cerevisiae* was the least resistant for all matrix categories with available data. There were no matrix categories where *S. aureus*, LAB, and *Salmonella* spp. were different from each other, but they were always more resistant than *S. cerevisiae*, but not always more resistant than *E. coli* (i.e., low and high-acid fruit juices where they were just as resis-

tant). The magnitude of these differences was matrix-dependent. In fruit juices, the microbial species differences were higher than in alcoholic products. Across-matrix contrasts, within the same microorganism at the same pH class, were substantial: $D_{E_{vol}}$ for *E. coli* was 21 kJ/L in low-acid animal milk but 46 kJ/L in low-acid fruit juice (~2-fold), while for *S. cerevisiae*, it was 5.9 kJ/L in high-acid fruit juice and 50 kJ/L in high-acid alcoholic product (~eight-fold), showing that matrix composition, not pH alone, can significantly affect PEF efficacy. The data for *E. coli* in high-acid liquid medium, *Listeria* spp. in high-acid whey, and *Salmonella* spp. in low-acid liquid medium were not further processed in the next sections, because they originated from a single volumetric energy input level (Figure 4.1) and therefore the reduction estimates at different volumetric energy inputs would not be as robust.

4.3.4 The microbial reductions of PEF under common processing conditions

In Figure 4.3, the microbial reductions achieved for different combinations of matrix categories and microbial species at 100 kJ/L are presented. The level of 100 kJ/L was selected because it belongs to the range of typical conditions of PEF processing for the microbial inactivation of heat-sensitive liquid products (80-120 kJ/kg) (Aganovic and Smetana, 2022; Toepfl, 2011).

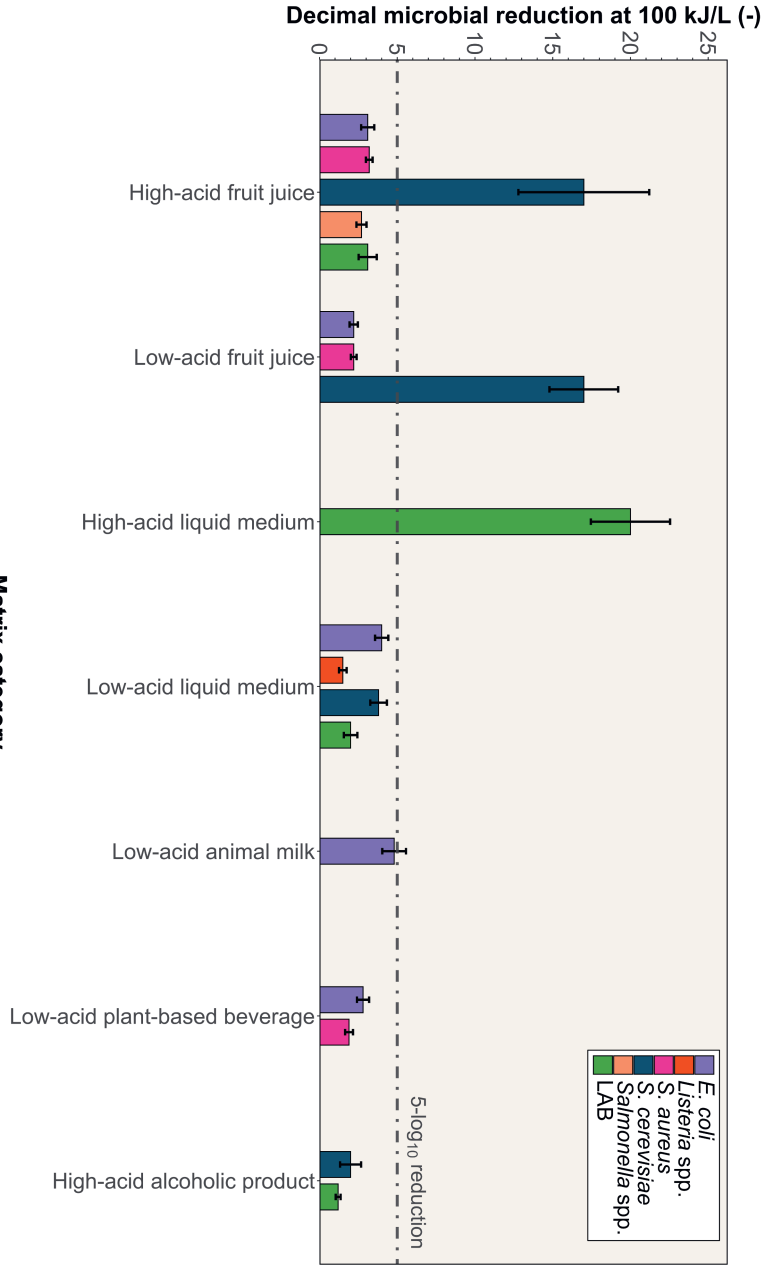


Figure 4.3: Estimated microbial reductions at a volumetric energy input of 100 kJ/L for each microbial species-matrix category combination. The error bars indicate the 95% CI for the reduction at 100 kJ/L. The number of data points (n) and studies (s) for each combination is reported in Figure 4.2 and Supplementary Table 4.1, which also provides detailed model-fit statistics.

Although in Figure 4.3 the achieved microbial reductions at 100 kJ/L were estimated, using the $D_{E_{\text{vol}}}$ values estimated from the regression lines of Figure 4.1 using Equation 4.10, reductions at other energy levels can also be estimated using the same equation. The complete set of regression equations for every microbial species-matrix category combination is provided in Supplementary Table 4.1. For example, Supplementary Table 4.1 reports a reduction of 2.7 ± 0.3 $\log_{10}$ for *Salmonella* spp. in high-acid fruit juices at E_{vol} = 100 kJ/L and therefore 200 kJ/L of volumetric energy input would result in a double reduction, i.e., 5.4 ± ~0.6 $\log_{10}$. As mentioned in Section 4.3.3, matrix acidity itself cannot sufficiently explain the variability related to the food matrix. For example, LAB is more susceptible in high-acid liquid media when compared to high-acid fruit juices. Overall, a volumetric energy input of 100 kJ/L failed, on average, to achieve a 5-$\log_{10}$-reduction for most microbial species-matrix category pairs, whereas 200 kJ/L would meet this threshold for the majority of the combinations.

4.3.5 Energy input needed to achieve 5-$\log_{10}$-reduction

Figure 4.4 illustrates, in three levels, the volumetric energy input necessary to achieve a 5-$\log_{10}$-reduction: (i) all microbial species and matrices combined; (ii) each microbial species pooled across matrices; and (iii) every microbial species-matrix category combination. For every segment, the mean volumetric energy input is reported together with the 95% inverse-prediction interval. Supplementary Material B provides a worked example of the inversion procedure. By using this example as a basis, the volumetric energy input needed to achieve a different level of reduction, such as 6- or 7-$\log_{10}$-reduction, can easily be estimated, while recognizing that these calculations may involve extrapolation beyond the fitted range.

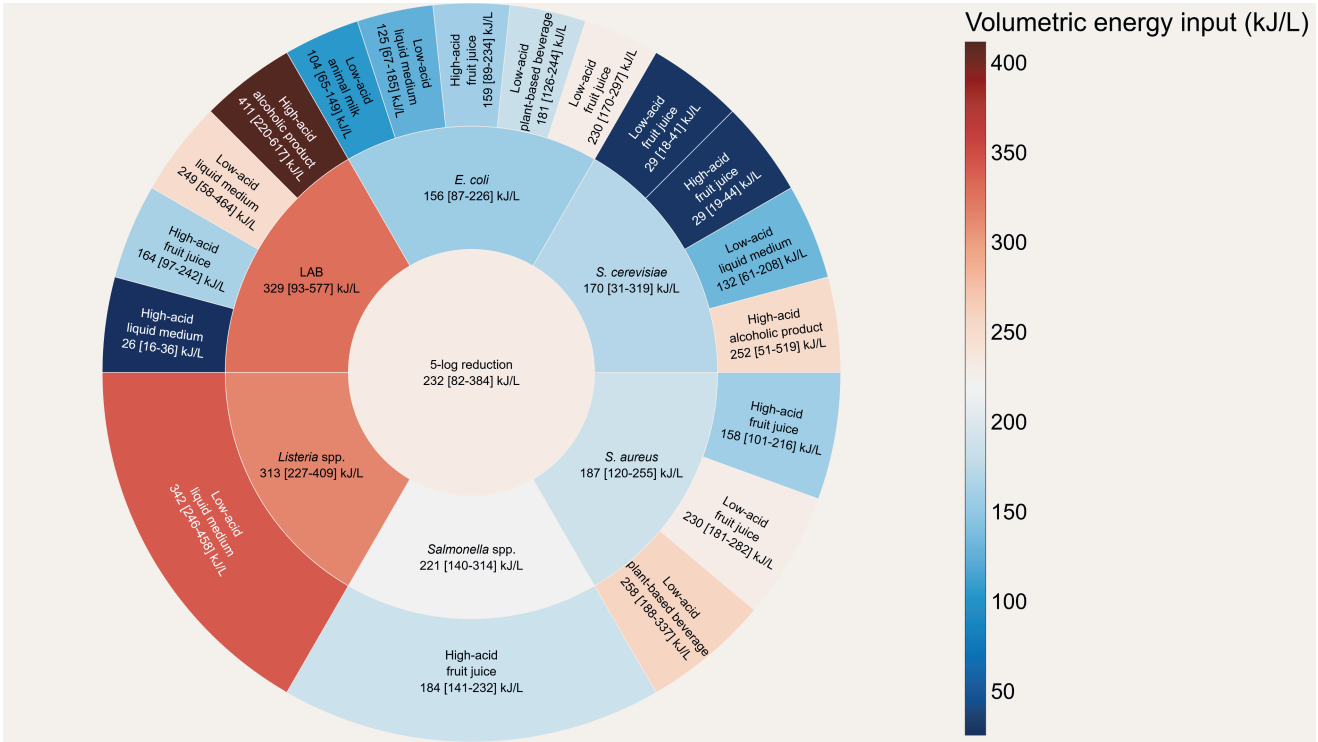


Figure 4.4: The volumetric energy input (kJ/L) needed to achieve a 5-$\log_{10}$-reduction. The center combines all data, the middle ring separates the microbial species, and the outer ring shows each microbial species-matrix category pair. Each segment lists the mean volumetric energy input and its 95% inverse prediction interval. The number of data points (n) and studies (s) for each combination is reported in Figure 4.2 and Supplementary Table 4.1, which also provides detailed model-fit statistics.

Across all 569 data points, the overall average energy input required to reach a 5-$\log_{10}$-reduction was 232 kJ/L (95% inverse prediction interval: 82-384 kJ/L). In line with Table 4.1 for the average $D_{E_{\text{vol}}}$, also when targeting the 5-$\log_{10}$-reductions, the species effects were evident: *E. coli* (156 [87-226] kJ/L), *S. cerevisiae* (170 [31-319] kJ/L), *S. aureus* (187 [120-255] kJ/L), *Salmonella* spp. (221 [140-314] kJ/L), *Listeria* spp. (313 [227-409] kJ/L), and LAB (329 [93-577] kJ/L). When coupled with the food matrix, the lowest requirement was recorded in LAB in high-acid liquid medium (26 [16-36] kJ/L) and *S. cerevisiae* in high-acid fruit juice (29 [19-44] kJ/L) and low-acid fruit juice (29 [18-41] kJ/L). These combinations yielded the lowest energy input requirements, and their prediction intervals did not overlap with any other microbial species-matrix category combination. At the opposite extreme were LAB in high-acid alcoholic products with 411 [220-617] kJ/L, and *Listeria* spp. in low-acid liquid medium with 342 [246-458] kJ/L. Within *E. coli*, low-acid animal milk with 104 [65-149] kJ/L and low-acid liquid medium with 125 [67-185] kJ/L were overlapping and required on average less than 125 kJ/L, whereas low-acid fruit juice with 230 [170-297] kJ/L and low-acid plant-based beverage with 181 [126-244] kJ/L were also overlapping but required on average above 180 kJ/L.

Note that for most microbial species-matrix category combinations, the term $\sqrt{1+\frac{E_{\text{vol}}^2}{\text{SSX}}}$ from Equation 4.14 was close to 1 (see column "term" in "stratification results" of Supplementary Table 4.1). Consequently, E_{vol} levels to achieve a 5-$\log_{10}$-reduction can be approximated by the linear expression ($\frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} \pm t_{0.975,\,n-1} \cdot RMSE = 5$) without solving the full quadratic as shown in Supplementary Material B. The only exceptions were the cases where very few data points were available or when extrapolation was taking place, or both (Supplementary Table 4.1). These results showed that for most microbial species-matrix category combinations, the main driver behind the variability observed was the generic variation in the data. The estimates presented in Figure 4.4 define the best- and worst-case energy inputs needed to ensure a 5-$\log_{10}$-reduction across the range of microbial species-matrix category combinations, while taking into account the variability between different laboratories and PEF devices.

4.4 Discussion

This manuscript describes the available data from literature that met our quality criteria to establish, for the first time, the global kinetic parameters for PEF processing based on 569 inactivation trials. These global kinetic parameters can be used as a benchmark for future validation. As explained in Section 4.2.3, the volumetric energy

input was the variable selected to derive inactivation parameters from the developed database. Although E_{vol} incorporates the electric field strength and the pulse width, identical energy levels can be delivered with different combinations of these parameters. Earlier work showed that short pulses with high electric field strength induce higher inactivation when compared to pulses with higher width but lower electric field strength at the same energy input level (Álvarez et al., 2003; Fox et al., 2008). In Equation 4.1, the electric field strength term is already squared, and therefore, this effect is captured, as with the increase of the electric field strength, E_{vol} increases disproportionately. However, there could be an additional or different effect with the change of other parameters, such as temperature (Liu et al., 2017) or variability sourced from device-specific characteristics. Additionally, although conductivity is included in Equation 4.1, its specific effect on inactivation would be worth investigating in dedicated studies. Such studies would need to account for the fact that, at constant input voltage, higher conductivity increases medium temperature, making it difficult to separate any direct effect of conductivity on electroporation from its indirect effects via temperature and other process parameters (Delso et al., 2022).

Lytras and co-authors showed that variability in PEF treatments is strain and species-dependent, with *L. monocytogenes* being identified as the most resistant microbial species compared to *E. coli*, *L. plantarum,* and *S. cerevisiae* (Lytras et al., 2024b). In the same research, *L. monocytogenes* L6 strain was identified as the most resistant strain among the strains tested (Lytras et al., 2024b). In our research, without isolating the strain and matrix effects, *E. coli*, *S. cerevisiae*, and *S. aureus* showed the lowest resistance, followed by *Salmonella* spp. and *Listeria* spp. (see Table 4.1); however, this order was affected by the availability of data for each microbial species in the different matrices. LAB, which involved multiple microbial species, and *Pseudomonas* spp., which had limited data, showed the highest resistance. Depending on the matrix category, the order of microbial species changed, highlighting that the food environment can shift microbial resistance. Therefore, matrix-specific models refined the model estimates showing that differences in animal milk, fruit juices, plant-based beverages, liquid media, whey, and alcoholic products can yield strong matrix effects on the microbial resistance beyond acidity. These effects were up to eight-fold within a single microbial species, and they did not necessarily follow the same pattern for all microbial species. As shown in previous studies, the effect of food factors such as pH and protein concentration is prominent in PEF decontamination (García et al., 2007; Sampedro et al., 2011; Schottroff et al., 2019; Timmermans et al., 2014). However, in this study, acidity seemed to have a species-specific effect and, in some cases, to be less important than other matrix effects. For example, for *S. cerevisiae*, the difference in resistance was higher between high-acid fruit juice and high-acid alcoholic products when compared to the difference with the low-acid fruit juice. For the low-acid liquid medium category, a high variability in the resistance was observed for all microbial

species (see Figure 4.1), which could be partially attributed to the multiple matrices involved, such as saline water solutions, liquid buffers, and growth media. On the other hand, high deviation from the fitted trendlines (Figure 4.1) was also observed in the high-acid alcoholic product category, while the data were derived from only three different wine varieties (red and white), indicating that the phenolic content or other wine properties might also affect the inactivation. Independent validation experiments in different matrices and devices would help confirm these matrix-specific effects. These results indicate that while global $D_{E_{\text{vol}}}$ values provide a useful starting point (since they are based on the data available), matrix-specific models can provide more realistic estimates.

After isolating the matrix effects, *S. aureus*, LAB, and *Salmonella* spp. showed no significant differences from one another in high-acid fruit juices, which was the only matrix category with available data for all three microorganisms (see Supplementary Figure 4.1). *S. aureus*, LAB, and *Salmonella* spp. were also consistently more resistant than *S. cerevisiae*, whereas *E. coli* was more resistant than *S. cerevisiae* in low- and high-acid fruit juice but not in low-acid liquid medium. The magnitude of these differences was matrix-dependent. In order to obtain useful insights about LAB, *Lactiplantibacillus*, *Lacticaseibacillus*, *Lactococcus*, *Leuconostoc*, *Oenococcus*, and *Pediococcus* had to be grouped together. Although this grouping helped to provide useful insights about LAB in general by using multiple microbial species, it also entailed higher variability since the microbial responses to PEF differ not only between lactic acid bacteria but also between strains of the same species (Thamsuaidee et al., 2024). As shown in Figure 4.1, the LAB dataset was the most highly variable, partially as a result of this grouping, but also due to the matrix categories involved. For example, LAB in high-acid fruit juice (n = 10) and in high-acid liquid medium (n = 36) were all *Lactiplantibacillus plantarum* strains and were not so variable within the same matrix, but the two matrices still differed from each other (Figure 4.2). For the high-acid alcoholic products, the lower efficacy contradicts the expectation that alcohol-containing products would result in higher inactivation since ethanol has the ability to alter the fluidity of microbial membranes (Puligundla et al., 2018). This effect could not be confirmed in our database. Even after isolating the microorganism effects, there were other sources of variability that might be responsible for the lower decontamination efficacy of PEF in alcoholic products. For example, for LAB in high-acid alcoholic products, the microbial reductions above 150 kJ/L were all coming from the same study where the authors reported that the inactivation was not affected by the energy input level (González-Arenzana et al., 2015). The same study was also responsible for 4 of the data points of *S. cerevisiae* in high-acid alcoholic products at 171, 173, 206, and 263 kJ/L (Figure 4.1), where the inactivation was also not affected by the energy input level. When the inactivation was not affected by the energy input level for some data points, then the overall estimated $D_{E_{\text{vol}}}$ also increased. In contrast to other studies, this study also reported that

the energy input needed to inactivate bacteria was not always higher than the energy input needed for yeasts (González-Arenzana et al., 2015), which could explain why *S. cerevisiae* was just as resistant as *E. coli* in low-acid liquid media (see Supplementary Figure 4.1).

Although in previous research 80-120 kJ/kg was mentioned as the range of typical conditions of PEF processing for the microbial inactivation of heat-sensitive liquid products (Aganovic and Smetana, 2022; Toepfl, 2011), this range of processing conditions would not provide a 5-$\log_{10}$-reduction for most of the microbial species-matrix category combinations when putting all available evidence together (see Figure 4.3). Increasing the energy input to higher levels, such as 200 kJ/L or more, would provide the desired reductions, but these levels might also pose some thermal effects given that based on the calorimetric equation for an average specific heat capacity C_p of liquids at ~4 kJ/(kg·°C) this would translate to 50 °C of temperature increase in a matter of seconds, especially if no cooling procedure is applied (Thamsuaidee et al., 2024). That being said, several studies in the literature use PEF devices that operate at energy levels much higher than 300 kJ/L without mentioning thermal effects through the application of cooling procedures (Amiali et al., 2006; Mosqueda-Melgar et al., 2008). The aggregated mean dose for a 5-$\log_{10}$-reduction was 232 kJ/L (95% inverse prediction, 82-384 kJ/L). This E_{vol} estimate was lower for most of the microbial species-matrix category combinations and with increasing model specificity (see Figure 4.4). For most microbial species-matrix category combinations, the results were supported by multiple data points from several laboratories (see Figure 4.2), but a few estimates relied on limited evidence. For LAB, all observations in high-acid liquid medium originated from a single study that explored 5 to 30 kJ/L volumetric energy input levels, whereas the high-acid fruit juice dataset from another study explored 35-190 kJ/L. This difference in the process settings design and the corresponding differences in the $D_{E_{\text{vol}}}$ values might also imply differences in the efficacy of different PEF devices, in addition to the matrix effects. Nevertheless, meta-regression models of this type are generally conservative at the high-exposure end since the heterogeneity of large pooled datasets and the experimental bias in the source studies tend to make the fitted slopes less steep and the corresponding *D*-values higher. The opposite occurs at the low-exposure end, where resistance can be underestimated (Garre et al., 2023).

As with all meta-analyses and meta-regression models, by definition, when standardizing for effects, by filtering the dataset for certain conditions, the evidence will be weaker due to a lower number of data points to support a trend observed. Additionally, when standardizing for an intended effect, there are always unintended effects that are filtered out at the same time. For example, while filtering for a specific food matrix, we also filter for certain pH levels, but also maybe for specific microbial species, and also for specific research teams and their equipment and practices, which is a form

of selection bias. Therefore, the analyst must find the balance between zooming in the dataset enough to isolate the main sources of variability, while retaining sufficient observations to produce robust, generalizable trends from multiple studies; filter too little and key effects remain hidden, filter too much and the evidence base becomes too "thin" to be generic and considered as a meta-approach. The balance in this study was found in the use of the volumetric energy input as the main independent numerical variable, and microbial species and matrix category as the main independent categorical variables that affect treatment efficacy. The decision tree in Figure 4.4 enables volumetric energy input estimation from readily available product attributes (pH and broad matrix category). Because the energy doses are expressed in kJ/L, they provide a common quantitative scale that can, in combination with device- and process-specific information, be linked to overall energy use, life-cycle assessments, and other sustainability indicators such as greenhouse gas emissions. In this way, the models offer a quantitative basis for future multi-criteria evaluation and optimization of PEF processing. For process designers, they offer a practical starting point to select the worst-case bound of the 95% inverse prediction interval and then refine operating conditions as device- and product-specific validation data become available.

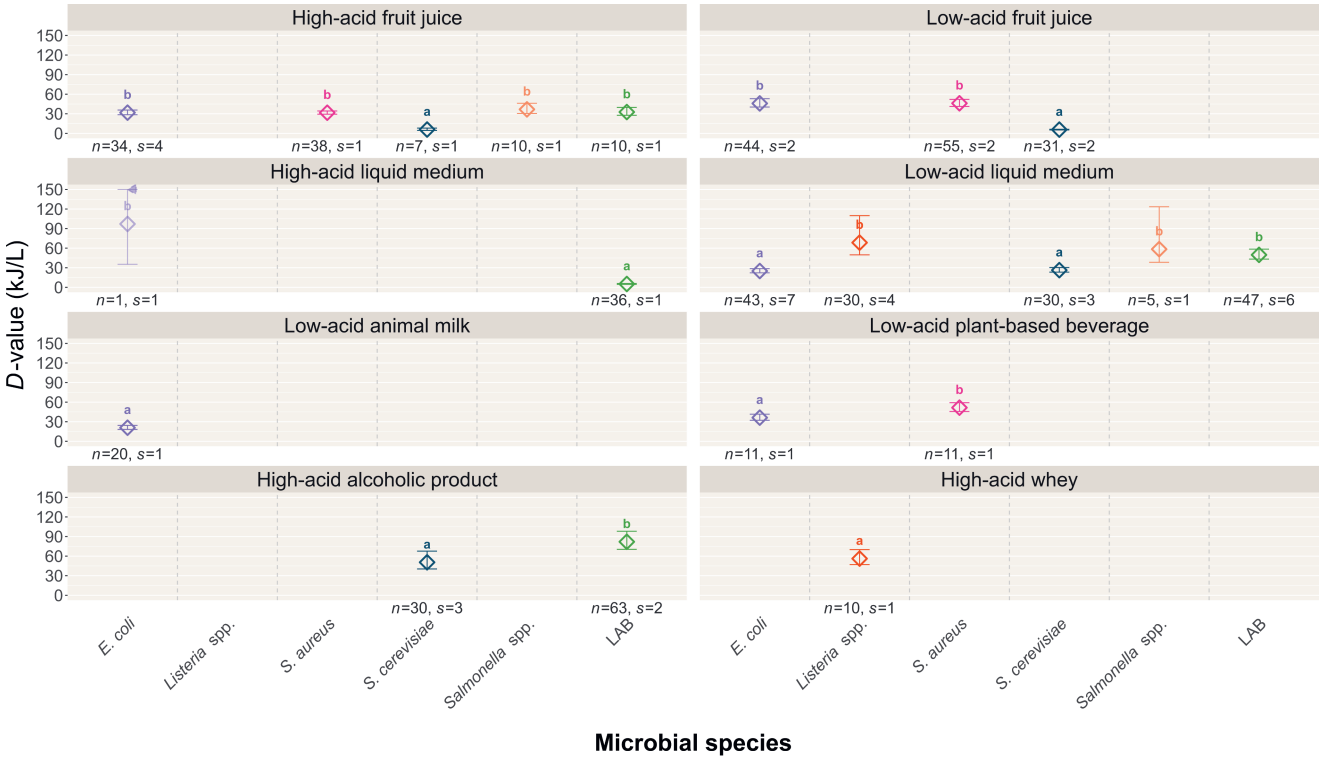
4.5 Conclusions

This study developed a database of 569 inactivation experiments from 28 studies and performed meta-regression to establish the global kinetic parameters for PEF based on seven microorganisms and various food matrices. Different food matrices induced variability in decontamination efficacy, which could not be fully described with the effect of acidity alone. While 100 kJ/L would provide a 5-$\log_{10}$-reduction of *S. cerevisiae* in low- and high-acid fruit juices, most microbial species-matrix category combinations require more than 150 kJ/L; the overall mean dose was 232 kJ/L (95% inverse prediction interval: 82-384 kJ/L). Future work should expand the database to additional microbial species and periodically update the models. As with all models, validation of the achieved reductions would also need to be performed before industrial applications, and it would be useful to link decontamination efficacy to sensory quality, nutrient retention, and sustainability metrics such as energy demand and greenhouse gas emissions. The kinetic parameters reported here provide a quantitative benchmark for food business operators, life cycle and risk assessors, and regulatory authorities.

Supplementary Material

The Supplementary Material included below reproduces Supplementary Figure 4.1 and Supplementary Material B. Supplementary Material A and Supplementary Table 4.1 are not included in the thesis body. They remain available with the published article (DOI: 10.1016/j.ijfoodmicro.2025.111547).

Supplementary Figures



Supplementary Figure 4.1: The results of multiple linear regression conducted for each matrix category dataset with the microbial species as a dummy variable. The "n" and "s" values correspond to the number of data points (experiments) and studies for each combination of microbial species-matrix category, respectively. The letters above each coefficient in each panel indicate statistically significant differences within that matrix category at a significance level $\alpha = 0.05$ (p-value ≤ 0.05). The coefficient for *E. coli* in high-acid liquid medium is shaded because its 95% slope confidence interval include 0, leaving the upper bound of the corresponding D-value indeterminate.

Supplementary Material B

Worked example: energy dose that can achieve a 5-$\log_{10}$-reduction

As mentioned in Section 2.4.1, log-linear fits in the manuscript used:

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) = \frac{E_{\text{vol}}}{D_{E_{\text{vol}}}}$$

Data used in this example

For the *Salmonella* spp. in high-acid fruit juice dataset (the complete set of regression equations for every microbial species-matrix category combination is provided in Supplementary Table 4.1):

Variables	Meaning	Value
n	observations	10
$D_{E_{\text{vol}}}$	decimal-reduction volumetric energy	36.7692 kJ/L
RSS	$\sum_{i=1}^{n}\left[\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)_i - \frac{E_{\text{vol},i}}{D_{E_{\text{vol}}}}\right]^2$	2.1
$RMSE$	$\sqrt{\text{RSS}/(n-1)}$	0.47818
SSX	$\sum_{i=1}^{n} E_{\text{vol},i}^2$	$1.12 \cdot 10^5$ (kJ/L)^2
$t_{0.975,\,9}$	two-sided 95 % t-value	2.2622

Forward prediction limits

The 95% prediction limits for $\log_{10}$-reduction at any volumetric energy input E_{vol} were:

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) \pm t_{0.975,\,n-1} \cdot RMSE \cdot \sqrt{1 + \frac{E_{\text{vol}}^2}{\text{SSX}}},$$

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) = \frac{E_{\text{vol}}}{D_{E_{\text{vol}}}}$$

This results in two nonlinear equations:

$$\frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} \; - \; t_{0.975,\,n-1} \cdot RMSE \cdot \sqrt{1 + \frac{E_{\text{vol}}^2}{\text{SSX}}} = \log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)$$

$$\frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} \; + \; t_{0.975,\,n-1} \cdot RMSE \cdot \sqrt{1 + \frac{E_{\text{vol}}^2}{\text{SSX}}} = \log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)$$

Solving the inverse-prediction equations by turning them into a quadratic

If we isolate the square-root term in the worst-case (minus sign) equation and then square both sides:

$$\begin{aligned}\left(\frac{1}{D_{E_{\text{vol}}}}\right) \cdot E_{\text{vol}} - \log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) &= t_{0.975,\,n-1} \cdot RMSE \\ &\quad \cdot \sqrt{1 + \frac{E_{\text{vol}}^2}{\text{SSX}}} \\ \Rightarrow \quad \left(\frac{1}{D_{E_{\text{vol}}}}\right)^2 \cdot E_{\text{vol}}^2 \; - \; &\frac{2 \cdot \log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)}{D_{E_{\text{vol}}}} \\ &\quad \cdot E_{\text{vol}} \; + \; \left[\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)\right]^2 \\ &= t_{0.975,\,n-1}^2 \cdot RMSE^2 \\ &\quad \cdot \left(1 + \frac{E_{\text{vol}}^2}{\text{SSX}}\right)\end{aligned}$$

Because taking either the "+" or "-" prediction-limit equation yields the same quadratic after squaring, the two real roots of that quadratic are the lower- and upper-bound energy inputs required for the specified $\log_{10}$-reduction.

Re-arranging gives a quadratic in E_{vol}:

$$\left[\left(\frac{1}{D_{E_{\text{vol}}}}\right)^2 - \frac{t^2_{0.975,\,n-1}\cdot RMSE^2}{\text{SSX}}\right]\cdot E^2_{\text{vol}} - \frac{2\cdot\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)}{D_{E_{\text{vol}}}}\cdot E_{\text{vol}} + \left[\left(\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)\right)^2 - t^2_{0.975,\,n-1}\cdot RMSE^2\right] = 0$$

Therefore, the quadratic can be written in the standard form $a\cdot E^2_{\text{vol}} + b\cdot E_{\text{vol}} + c = 0$ with coefficients:

$$a = \left(\frac{1}{D_{E_{\text{vol}}}}\right)^2 - \frac{t^2_{0.975,\,n-1}\cdot RMSE^2}{\text{SSX}},$$

$$b = -\frac{2\cdot\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)}{D_{E_{\text{vol}}}},$$

$$c = \left(\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)\right)^2 - t^2_{0.975,\,n-1}\cdot RMSE^2$$

For the present data set and 5-$\log_{10}$-reduction:

$$a = (1/36.7692)^2 - \frac{2.2622^2\cdot 0.47818^2}{1.12\cdot 10^5} = 7.29\cdot 10^{-4},$$
$$b = -\frac{10}{36.7692} = -2.72\cdot 10^{-1},$$
$$c = 25 - 2.2622^2\cdot 0.47818^2 = 23.83$$

The discriminant is:

$$\begin{aligned}\Delta &= b^2 - 4\cdot a\cdot c\\ &= (2.72\cdot 10^{-1})^2 - 4\cdot 7.29\cdot 10^{-4}\cdot 23.83\\ &= 4.45\cdot 10^{-3} > 0,\end{aligned}$$

so two real roots exist:

$$E_{\text{vol}} = \frac{-b \pm \sqrt{\Delta}}{2 \cdot a} = \frac{0.272 \pm 0.0668}{2 \cdot (7.29 \cdot 10^{-4})}$$

Hence, the two real roots of the quadratic give the volumetric-energy bounds that can assure a 5-$\log_{10}$ reduction:

$$E_{\text{vol,low}} = 141 \text{ kJ/L}, \qquad E_{\text{vol,high}} = 232 \text{ kJ/L}$$

The a, b, c terms and the discriminant Δ together with the estimated $E_{\text{vol,low}}$ and $E_{\text{vol,high}}$ for all microbial species-matrix category combinations can be found in Supplementary Table 4.1.

Because $E_{\text{vol}} = 5\, D_{E_{\text{vol}}}$ for a 5-$\log_{10}$-target, the corresponding $D_{E_{\text{vol}}}$ values are:

$$D_{E_{\text{vol,best}}} = \frac{E_{\text{vol,low}}}{5} = \frac{141}{5} = 28.2 \text{ kJ/L}, \qquad D_{E_{\text{vol,worst}}} = \frac{E_{\text{vol,high}}}{5} = \frac{232}{5} = 46.4 \text{ kJ/L}$$

The mean volumetric energy input to achieve 5-$\log_{10}$-reduction is:

$$E_{\text{vol}} = 5 \cdot D_{E_{\text{vol}}} = 5 \cdot 36.8 = 184 \text{ kJ/L}$$

The 95% inverse-prediction interval is:

$$[E_{\text{vol,low}},\ E_{\text{vol,high}}] = [\,141,\ 232\,] \text{ kJ/L}$$

Therefore, the 95% inverse prediction interval was estimated at 141-232 kJ/L for *Salmonella* spp. in high-acid fruit juice. The same calculation was applied to every microbial species-matrix category combination; the resulting mean volumetric energy inputs to achieve a 5-$\log_{10}$-reduction ($5/D_{E_{\text{vol}}}$) and the inverse prediction intervals were used to generate Figure 4.4.

CRediT authorship contribution statement

George Pampoukis: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. **Jan Ham**: Validation, Formal Analysis, Investigation, Data Curation. **Marcel H. Zwietering**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision. **Heidy M.W. den Besten**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project Administration, Funding Acquisition.

Acknowledgements

We thank Dr. Rian Timmermans (Wageningen Food & Biobased Research, Wageningen University & Research) for her helpful feedback regarding PEF principles. We thank Philipa Sindhunata (Food Microbiology, Wageningen University) and Aruna Murali (Food Microbiology, Wageningen University) for their support in the literature search and data collection during their Master's thesis projects. We thank Dr. Vasilis Valdramidis (Department of Chemistry, National and Kapodistrian University of Athens) for arranging an in-person visit to a PEF facility. We thank Dr. George Dimopoulos (Laboratory of Food Chemistry and Technology, National Technical University of Athens) and Prof. Petros Taoukis (Laboratory of Food Chemistry and Technology, National Technical University of Athens) for presenting their PEF device in operation and their energy input estimations. We thank Arisa Thamsuaidee (ELEA Technology GmbH, Germany & Department of Chemistry, National and Kapodistrian University of Athens) and Dr. Claudia Siemer (ELEA Technology GmbH, Germany) for providing information regarding the ELEA PEF devices and their energy input estimations.

Data availability

Supplementary Figure 4.1 and Supplementary Material B are reproduced in the thesis body. Supplementary Material A and Supplementary Table 4.1 are not included in the thesis body and remain available with the published article (DOI: 10.1016/j.ijfoodmicro.2025.111547). The Chapter 4 supplementary dataset is also available via 4TU.ResearchData (DOI: 10.4121/1ab51010-9b1c-4ad7-a86e-142bb16ae2e5).

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N° 955431.

Chapter 5

High-pressure processing inactivation of foodborne bacteria and viruses in foods

Kah Yen Claire Yeak[a,+], George Pampoukis[a,b,+], Cristina Serra-Castelló[a], Sara Bover-Cid[c], Heidy M.W. den Besten[a], Marcel H. Zwietering[a]

[a] Food Microbiology, Wageningen University & Research, Wageningen, the Netherlands
[b] Institute of Food, Nutrition and Health, ETH Zurich, Zurich, Switzerland
[c] Institute of Agrifood Research and Technology (IRTA), Food Safety and Functionality, Monells, Spain

[+] These authors contributed equally to this work.

Submitted for publication.

Abstract

High-pressure processing (HPP) is a non-thermal food processing technology that can inactivate microbiological hazards (MHs) while preserving product quality. This study compiled 2692 *D*-values across 15 MHs pressurized at different pressures at a temperature ≤ 45 °C. The Bigelow model was used to estimate D_{ref} values at P_{ref} of 500 MPa. Based on these values, MHs were grouped into three HPP resistance categories across all food matrices: 1) Sensitive (D_{500MPa} < 1 min): *Vibrio* spp., *Aeromonas caviae*, *Campylobacter* spp., Rotavirus, *Cronobacter* spp., Hepatitis A, Norovirus; 2) Moderate (1 ≤ D_{500MPa} < 3 min): *Yersinia enterocolitica*, *Listeria monocytogenes*, *Mycobacterium tuberculosis* var. *bovis*, *Salmonella enterica* non-Typhi, Hepatitis E, *Escherichia coli*; and 3) Resistant (D_{500MPa} ≥ 3 min): *Staphylococcus aureus*, *Bacillus cereus* (vegetative cells). However, food matrix composition strongly influenced microbial resistance to HPP: low water activity (a_w ≤ 0.95) conferred protection, while acidic (pH ≤ 4.5) or high a_w matrices (a_w > 0.95) enhanced inactivation. Multiple linear regression identified microbial species, a_w, and pH as significant predictors of the inactivation; other product characteristics, such as protein and carbohydrate content, showed limited effects. A decision-support framework was developed to evaluate MH resistance to HPP processing. The framework integrates three complementary approaches: 1) MH resistance ranking at 500 MPa, 2) matrix-stratified inactivation estimates by food category, and 3) multiple linear regression modeling using a_w and pH as independent variables. These approaches complement one another and support MH resistance evaluation under different data availability scenarios.

5.1 Introduction

Ensuring food safety is essential to protect consumers from foodborne microbiological hazards (MHs), including bacteria, viruses, parasites, physical hazards, and harmful chemical contaminants (World Health Organization, 2021, 2015). To achieve this, food safety regulations and food safety management systems are in place (Codex Alimentarius Commission, 2019; European Union, 2002). A key way to control microbial food safety is the use of processing treatments to inactivate MHs. Traditionally, thermal treatments, such as pasteurization, boiling, and sterilization, have been the most widely applied methods in the food industry (Espinosa et al., 2020; Juneja, 2001; Van Asselt and Zwietering, 2006). However, consumers increasingly demand not only safe food but also the preservation of nutritional quality and fresh-like sensory attributes. This shift has driven research into milder food processing techniques, such as ultraviolet light, non-thermal plasma, pulsed electric fields, and high-pressure processing (HPP) (Den Besten and Valdramidis, 2025; Silva et al., 2024). These technologies offer the advantage of inactivating MHs while better preserving sensory freshness traits and nutrients such as vitamins, antioxidants, and essential micronutrients, which could be reduced during traditional thermal processing (Niu et al., 2022; Rodríguez-Roque et al., 2016). Among these, HPP is one of the most commercially successful non-thermal food processing technologies, typically operating at temperatures below 40 °C (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Huang et al., 2017).

Currently, food industry applications of HPP involve processing at hydrostatic pressures of 400 to 600 MPa for a few minutes to inactivate MHs and spoilage microorganisms in various food products (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022). HPP is not strictly a novel technology, as the first commercial applications entered the Japanese market in 1980 and included yogurts, jams, and jellies (Aganovic et al., 2021). HPP is currently an emerging food preservation technology with an increasing number of food product applications. In 2019, more than 550 commercial HPP machines for food processing were in operation worldwide (Aganovic et al., 2021), increasing to 700 industrial units by the end of 2025 (Hiperbaric, 2025). Despite this growth, the installed base remains modest compared with established thermal processing technologies, reflecting the substantial capital investment required for HPP equipment. Nevertheless, uptake continues to expand, supported in part by contract "tolling" facilities that provide HPP services without requiring in-house ownership of a unit (Aganovic et al., 2021). Today, HPP is more frequently applied to meat and meat products, cheese and other dairy products, juices, smoothies, fish and fish products, and ready-to-eat products (Alemán et al., 1994; O'Reilly et al., 2000; Patterson et al., 1996; Pérez et al., 2023; Serra-Castelló et al., 2022; Simpson, 1998). Recently, EFSA evaluated HPP for raw milk and concluded that pathogen reductions achieved

under current industrial HPP conditions are lower than those required for thermal pasteurization. However, minimum pressure-time combinations were defined to meet specified performance criteria for hazard reduction (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022). Beyond efficacy considerations, wider adoption is mainly constrained by high capital investment, when operating costs also matter, with electricity demand being the major contributor (Cacace et al., 2020).

The efficacy of HPP depends on several factors, including processing parameters (pressure, initial temperature of the pressurization fluid, come-up time, and holding time) (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Santillana Farakos and Zwietering, 2011) and food matrix characteristics (pH, water activity (a_w), fat, protein, and carbohydrate content) (Sehrawat et al., 2020; Serra-Castelló et al., 2022). Microbial physiology also plays a crucial role, as resistance varies depending on factors such as growth phase (i.e., physiological state), cellular complexity, cell shape, and cell wall structure (Mañas and Mackey, 2004). For instance, despite the strain variability, Gram-positive bacteria (particularly cocci) generally exhibit higher pressure tolerance than Gram-negative bacteria, while parasites are thought to be more susceptible (Mañas and Mackey, 2004; Rolfe et al., 2023). However, spores of spore-forming bacteria such as *Bacillus cereus*, *Clostridium botulinum*, and *Clostridium perfringens* are highly resistant to non-thermal HPP treatments and often require additional thermal inputs to achieve sufficient inactivation (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022).

Despite the growing application of HPP in the food industry, knowledge gaps remain. Most existing literature focuses on specific MHs or food matrices, with the majority of available data referring to *Salmonella*, *Listeria* spp., pathogenic *Escherichia coli*, and *Staphylococcus aureus* (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Evelyn and Silva, 2019; Santillana Farakos and Zwietering, 2011; Torrents-Masoliver et al., 2025). Furthermore, some research studies report inactivation data under unrealistic processing conditions for industrial equipment, such as pressures exceeding 600 MPa or prolonged holding times (Lenz and Vogel, 2014; Margosch et al., 2004; Shao et al., 2022), that do not reflect standard industrial practices, which typically use 400-600 MPa for up to 6 minutes (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022).

Building on our earlier work on microbiological hazard identification and risk ranking in infant foods Yeak et al. (2024a, 2024b), this study compiles and analyzes HPP inactivation data for 15 foodborne MHs, including bacteria and viruses, across a range of food products. To support decision-making across pressure levels and food matrices, MH resistance was quantified and ranked using $\log_{10}D$-values standardized at a reference pressure of 500 MPa. Matrix effects were addressed using a tiered analysis, first stratifying by food category and, where needed, further evaluating intrinsic product characteristics such as pH and a_w.

5.2 Methods

5.2.1 List of microbiological hazards and food categories

Based on literature coverage and our earlier hazard identification work (Yeak et al., 2024b, 2022), an initial list of 34 MHs was targeted. Of these, 22 MHs were represented in the compiled dataset, and 15 had sufficient usable data for quantitative analysis, namely, *Aeromonas caviae*, *Bacillus cereus* (vegetative cells), *Campylobacter* spp., *Cronobacter* spp., *Escherichia coli*, Hepatitis A, Hepatitis E, *Listeria monocytogenes*, *Mycobacterium tuberculosis* var. *bovis*, Norovirus, Rotavirus, *Salmonella enterica* non-Typhi, *Staphylococcus aureus*, *Vibrio* spp., and *Yersinia enterocolitica.* The remaining hazards were excluded due to no or limited usable data, incomplete reporting, or inconsistent experimental conditions. Similar limitations have been documented for other non-thermal technologies, such as non-thermal plasma (Pampoukis et al., 2024) and pulsed electric fields (Pampoukis et al., 2026), where incomplete and inconsistent reporting of experimental metadata constrains cross-study comparisons and reuse of data (Pampoukis et al., 2025). The literature review and data curation procedures are described in detail in Section 5.2.2.

The food categories used in this study follow the classification approach discussed in Yeak et al. (2024a, 2024b), and were mapped to EFSA's FoodEx2 system for food classification (EFSA, 2015). These included: fish and fish products; meat and meat products; egg and egg products; milk and milk products, including two subgroups (milk and dairy products, and cheese and cheese products); food of non-animal origin (FoNAO), including subgroups of fruits and vegetable products, cereal products and rice and seeds, herbs and spices, and plant-based formulas.

HPP inactivation data for the 15 MHs across these defined food categories were collected through database searches using tailored search strategies, as described in Section 5.2.2.1. All relevant records were screened and curated according to specific inclusion and exclusion criteria.

5.2.2 Literature review and data curation

5.2.2.1 Databases and search strings

A literature search in five established databases, namely, Scopus, PubMed, Web of Science, FSTA, and CAB Abstracts, was conducted to retrieve relevant studies. Search strings were composed of keywords and their synonyms related to HPP, MHs, and food

matrices. Boolean operators such as "AND," "W/ (within)," and "PRE/" were adapted to suit each database's syntax and search logic. To maximize coverage and minimize duplication, all retrieved records were pooled, and duplicates were removed. Below is an example of a search string used in Scopus for *Salmonella* spp. in pork:

"high-pressure processing" OR "high pressure processing" OR "HPP" OR "HHP" OR "high hydrostatic pressure" OR "high-hydrostatic pressure") W/7 ("treatment" OR "inactivat*" OR "decontaminat*" OR "steriliz*" OR "pasteuriz*" OR "sterilis" OR "pasteuris*") AND ("*Salmonella*") AND ("pork").

The same search terms were used across databases, with adjustments made only for database-specific syntax and operators.

5.2.2.2 Literature review and exclusion criteria

A total of 13717 articles were retrieved from the five databases using the search query described in Section 5.2.2.1. After removing the duplicates, 3529 articles remained. Articles were then screened based on the exclusion criteria summarized in Figure 5.1, which were: **1)** Non-English articles; **2)** Title, abstract, and keywords not indicating high-pressure processing, HPP, or their synonyms; **3)** Articles about HPP but not for microbial inactivation; **4)** Temperature during/after HPP processing > 45 °C; **5)** HPP in combination with other lethal treatments, such as irradiation; **6)** *D*-values or $\log_{10}$-reduction values were not reported, or data could not be used to derive the *D*-value, e.g., articles reporting "not-detected" or "no-detection" for a hazard after HPP; **7)** The total microbial reduction was < 0.5 $\log_{10}$ units; **8)** Treatment was applied to spores, and **9)** Microbiological hazards with insufficient *D*-values for meta-regression ($n \leq 6$). After applying these criteria, 184 articles were selected containing extractable, relevant data, which were included in the final analysis of non-spore-forming MHs that can be inactivated by non-thermal HPP (≤ 45 °C).

5.2.2.3 Data extraction, imputation, and cleaning

Data from seven informational categories were extracted from the 184 articles: **1)** Scientific article: title, DOI, and publication year; **2)** Product characteristics: food matrix data (pH, a_w, fat, NaCl, protein and carbohydrate content) that may influence HPP efficacy; **3)** Hazard characteristics: strain/s, cocktail, surrogate/s, or serotype/s of MHs showing differences in HPP inactivation efficacy in certain food products; **4)** Hazard inoculum: inoculum levels used; **5)** HPP processing parameters: applied pressure (MPa), come-up time, processing (holding) time, initial and actual processing temperature; **6)** Adaptation before HPP treatment and growth phase of cultures; and **7)**

Inactivation data: $\log_{10}$-reduction values, *D*-values, and information about the number of observations used to estimate the *D*-values. For studies presenting data only in the form of graphs, the numerical values were extracted using the WebPlotDigitizer software (version 4.6) (Rohatgi, 2025). The full database is available in Supplementary Table 5.1.

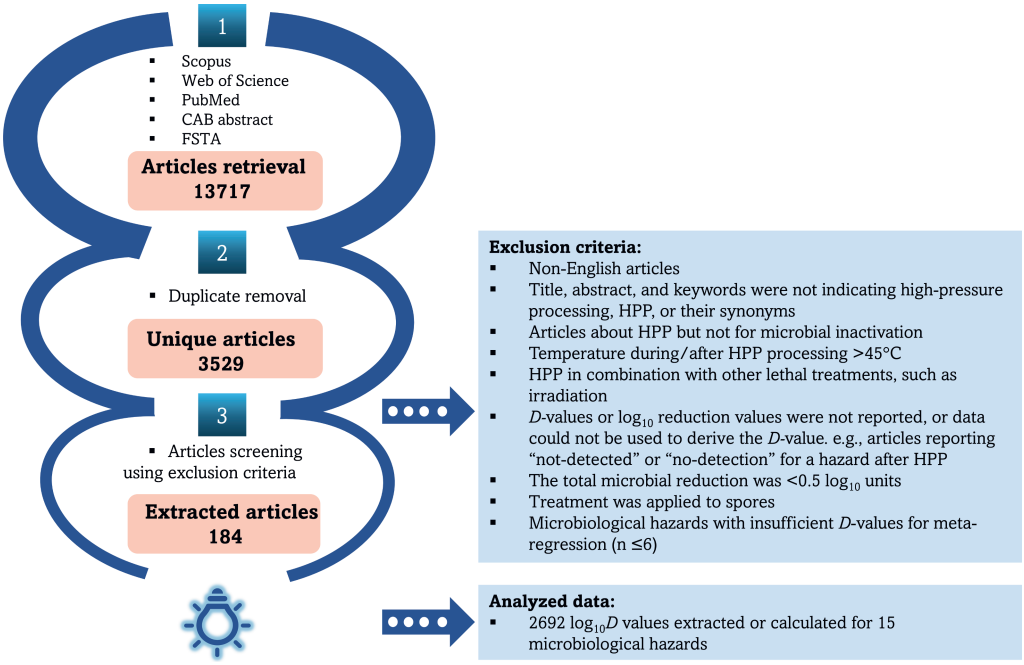


Figure 5.1: Literature review procedure and retrieved number of articles, including exclusion criteria and total data obtained (final quantitative dataset: 15 microbiological hazards).

When product characteristics were not reported, values for pH, a_w, fat, proteins, and carbohydrates were imputed for this food product using data from the FoodData Central USDA database, available at https://fdc.nal.usda.gov/. Additionally, pH, a_w, and fat content were categorized based on thresholds relevant to HPP resistance: pH (≤ 4.5 and > 4.5) (Guillou and Membré, 2019), a_w (a_w: ≤ 0.95 and > 0.95) (Guillou and Membré, 2019; Serra-Castelló, 2023; Valdivielso, 2023) and fat (≤ 10% and > 10%) (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Sehrawat et al., 2020; Shad et al., 2024). The actual processing temperature during HPP was also imputed when not reported, using an estimated increase of 3 °C per 100 MPa (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022).

To prepare a clean dataset suitable for statistical analysis and model fitting, the extracted raw data were reviewed, sorted, and refined to remove inconsistencies, formatting errors, and irrelevant entries. Subsequently, *D*-values, i.e., the pressurization

time required to achieve a 1-$\log_{10}$-reduction in an MH, were estimated per study and condition set by fitting a primary linear model to the available $\log_{10}$-reduction values and their corresponding treatment times at a given pressure level. Inactivation kinetic datasets consisting of at least three $\log_{10}$-reduction values for different treatment times at a given pressure level were not always available. In these cases, *D*-values were calculated from a single survival $\log_{10}$-count or $\log_{10}$-reduction value as the inverse of the slope between the reported initial level and the level after HPP. The *D*-values were subsequently $\log_{10}$-transformed into $\log_{10}D$-values.

5.2.3 Data analysis

To characterize MH resistance to HPP across pressure levels, the pressure-adapted Bigelow model (Bigelow, 1921) (Equation 5.1) was fitted to the $\log_{10}$-transformed *D*-values collected for MHs at different pressure levels:

$$log_{10}D = \log_{10} D_{ref} + \frac{(P_{ref} - P\ )}{z_P} \tag{5.1}$$

where $\log_{10}D$ is the time ($\log_{10}$ minutes) required to achieve a 1-$\log_{10}$-reduction of MHs at the applied pressure; $\log_{10}D_{\text{ref}}$ is the time ($\log_{10}$ minutes) required to achieve a 1-$\log_{10}$-reduction of MHs at reference pressure, i.e., 500 MPa; *P* is the applied pressure during holding time (MPa); P_{ref} is the reference pressure (500 MPa), and z_P is the increase in pressure necessary to reduce $\log_{10}D$ by one (MPa).

The Bigelow model was fitted separately to each MH dataset and to stratified MH-food category subsets. Model fit was summarized by the standard error of the regression (root mean square error (RMSE)) and the standard errors of the parameter estimates. In addition, 95% prediction intervals (PI) were calculated to quantify uncertainty and the expected variability around the predicted $\log_{10}D$-values. Full details of the statistical calculations, including the estimation of standard errors and prediction intervals, are provided in Supplementary Material A.

To better understand the food matrix effects on the inactivation of all MHs and avoid focusing only on certain MHs that have enough data for multiple food characteristics, multiple linear regression was implemented. The multiple linear regression approach used pressure (MPa) and the intrinsic food characteristics such as pH, protein content (%), carbohydrate content (%) as numerical independent variables, and MH, a_w (≤ 0.95, > 0.95), and fat (≤ 10%, > 10%) as categorical independent variables. The a_w and fat variables were treated categorically to minimize the impact of data loss from

missing values, enabling inclusion of a wider dataset and supporting more generalizable inference on their effects on inactivation. Most foods (liquids) in the dataset had a_w values above 0.95, with a smaller subset falling below this threshold. The 10% fat cut-off was determined based on the dataset, where most lean foods contained less than 10% fat, whereas higher-fat foods, such as cheese, clearly stood out above this value. In previous research, 10% of fat was also used as the starting point to evaluate the ability of fat to induce high-pressure resistance against *L. monocytogenes* in dry-cured ham (Bover-Cid et al., 2015). The MHs were converted to dummy variables (0, 1), and *L. monocytogenes* was selected as the baseline species because it was the most frequently represented MH in the dataset used for the multiple linear regression. To avoid multicollinearity, only variables with a Variance Inflation Factor (VIF) ≤ 5 were used. Additionally, backward elimination was performed to exclude variables that were not significantly different from 0 for numerical variables, their counterparts for the categorical variables, or the baseline species for the MH dummy variable, at a significance level $\alpha = 0.05$ (p-value > 0.05). Candidate multiple linear regression models were compared using Akaike's Information Criterion (AIC) (equation in Supplementary Material A), and the model with the lowest AIC was selected, while prioritizing inclusion of the largest possible number of observations. For completeness, models and model parameters that were excluded after these screening steps are reported in Supplementary Table 5.4. The data analysis was performed in Python 3.11.2 (Van Rossum and Drake, 2009). To give the multiple linear regression model better interpretability, pressure was fitted as a reparametrized variable as 500-P, and the pH as 7-pH. Therefore, the intercept of the resulting model corresponds to the $\log_{10}D$-value when all variables are set to zero, corresponding to the inactivation of *L. monocytogenes* (baseline species) at a matrix of pH 7, a_w category > 0.95 at a pressure level of 500 MPa.

5.3 Results and Discussion

5.3.1 Estimation of $\log_{10}D$ at a reference pressure of 500 MPa

A total of 2692 $\log_{10}D$-values were initially compiled from 184 published HPP inactivation articles. Although 34 MHs were initially targeted based on literature coverage and hazard identification work, only 15 were retained for quantitative analysis after data curation. Several hazards, including parasites, were not represented because the available outcomes did not meet the eligibility criteria, and therefore, they could not support a reliable $\log_{10}D$ estimation. The resulting evidence base in the literature remains concentrated on a small set of well-studied pathogens, mainly *Salmonella* non-Typhi,

L. monocytogenes, *E. coli*, and *S. aureus*, highlighting the need for more consistent and quantitative inactivation data for underrepresented hazard groups to strengthen future HPP risk evaluation frameworks.

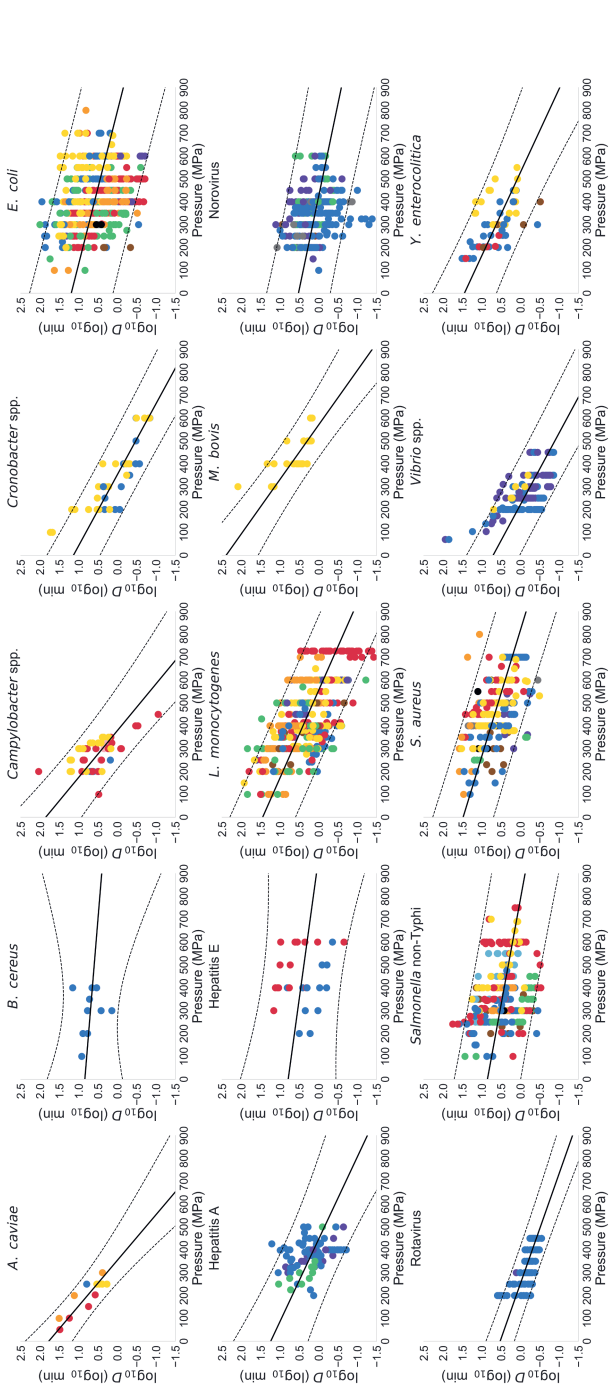


Figure 5.2: $\log_{10}D$-values collected for all microbial species across the different food matrices: Lab media (blue), Milk and milk products (yellow), Cheese and cheese products (orange), Fruits and vegetables (green), Fish and fish products (purple), Meat and meat products (red), Egg and egg products (brown), Plant-based formula (pink), Herbs and spices (light blue), Cereal products, rice and seeds (black), and Other foods (grey). The black solid lines represent the fitted Bigelow model (Table 5.1), while the dashed lines represent the 95% prediction interval for each model.

To enable standardized comparisons of the MHs in terms of pressure level, the Bigelow model was fitted to the $\log_{10}D$-values collected for each MH to describe their inactivation resistance under HPP (Figure 5.2). A higher $\log_{10}D_{\text{ref}}$ (at P_{ref} = 500 MPa) together with a higher z_P value indicated greater resistance to HPP. The estimated $\log_{10}D_{500\text{MPa}}$-values representing the average time (in $\log_{10}$ minutes) required to achieve a 1-$\log_{10}$-reduction at 500 MPa were used as a first step to rank MHs in terms of HPP resistance (Table 5.1). Based on these values, the resistance ranking from least to most resistant was: *Vibrio* spp. < *A. caviae* < *Campylobacter* spp. < *Rotavirus* < *Cronobacter* spp. < Hepatitis A < *Norovirus* < *Y. enterocolitica* < *L. monocytogenes* < *M. bovis* < *Salmonella* non-Typhi < Hepatitis E < *E. coli* < *S. aureus* < *B. cereus* (vegetative cells). MHs were further categorized based on their $\log_{10}D_{500\text{MPa}}$: 1) Sensitive ($D_{500\text{MPa}} \leq 1$ min), 2) Moderate ($1 < D_{500\text{MPa}} \leq 3$ min), and 3) Resistant ($D_{500\text{MPa}} > 3$ min) (Table 5.1).

Table 5.1: Microbiological hazard ranking based on the $\log_{10}D$ value at a reference pressure level (500 MPa). The dataset is filtered for non-spores and processing temperature ≤ 45 °C ($n = 2692$)

Microbiological hazards pressure resistance group	Microbiological Hazard	$\log_{10} D_{500\text{MPa}}$ ($\log_{10}$ min) ± SE	z_P ± SE (MPa)	Root Mean Square Error (RMSE) ($\log_{10}$ min)	*n*
Sensitive ($D_{500\text{MPa}} \leq 1$ min)	*Vibrio* spp.	-0.89 ± 0.07	314 ± 28	0.34	217
	A. caviae	-0.73 ± 0.26	200 ± 33	0.21	12
	Campylobacter spp.	-0.59 ± 0.18	205 ± 33	0.40	56
	Rotavirus	-0.51 ± 0.04	481 ± 49	0.17	88
	Cronobacter spp.	-0.47 ± 0.06	311 ± 28	0.32	56
	Hepatitis A	-0.15 ± 0.09	360 ± 87	0.41	87
	Norovirus	-0.09 ± 0.05	811 ± 167	0.41	274
Moderate ($1 \text{ min} < D_{500\text{MPa}} \leq 3$ min)	*Y. enterocolitica*	0.08 ± 0.11	365 ± 66	0.38	64
	L. monocytogenes	0.15 ± 0.02	383 ± 17	0.43	608
	M. bovis	0.30 ± 0.07	238 ± 37	0.29	35
	Salmonella non-Typhi	0.31 ± 0.03	924 ± 152	0.43	289
	Hepatitis E	0.38 ± 0.10	1230 ± 1152	0.49	28
	E. coli	0.44 ± 0.03	668 ± 83	0.54	609
Resistant ($D_{500\text{MPa}} > 3$ min)	*S. aureus*	0.58 ± 0.03	553 ± 54	0.39	259
	B. cereus (vegetative cells)	0.60 ± 0.22	2066 ± 4151	0.30	10

The **sensitive resistance group** included *Vibrio* spp., *A. caviae*, *Campylobacter* spp., *Rotavirus*, *Cronobacter* spp., Hepatitis A, and Norovirus with $D_{500\text{MPa}}$ < 1 min (or $\log_{10}D_{500\text{MPa}} < 0$ $\log_{10}$ min), suggesting a short processing time is required for effective inactivation (Table 5.1).

The **moderate resistance group** included *Y. enterocolitica*, *L. monocytogenes*, *M. bovis*,

Salmonella non-Typhi, Hepatitis E, and *E. coli* with $\log_{10}D_{500\text{MPa}}$ values ranging from 0.08 to 0.44 (Table 5.1). Among these, *E. coli* (n = 609), *L. monocytogenes* (n = 608), and *Salmonella* non-Typhi (n = 289) were the most extensively studied (Table 5.1), providing high confidence in their ranking. In contrast, *M. bovis* (n = 35) had fewer data points, reducing confidence in its exact ranking.

Interestingly, although both Hepatitis A and E are non-enveloped viruses, their estimated $\log_{10}D_{500\text{MPa}}$-values differed (Table 5.1). This discrepancy may partially reflect the matrices represented for each virus, including variability within the broad lab media category (Figure 5.2). Hepatitis A is primarily linked to fresh produce such as leafy greens and berries (Govaris and Pexara, 2021; Hirneisen and Kniel, 2013; Kingsley et al., 2005), which typically have lower pH, lower fat, and protein content, although it was also studied in fish and fish products in this dataset (Figure 5.2). In contrast, Hepatitis E is more often associated with raw or undercooked pork products (Emmoth et al., 2017; Nasheri et al., 2020), where higher fat and protein levels may offer protective effects to viral particles under pressure (Serra-Castelló et al., 2022). In this dataset, the available Hepatitis E data were concentrated in meat and meat products and lab media, with higher $\log_{10}D$-values in the former and an apparent separation across matrix categories (Figure 5.2). These observations underscore the vital role that matrix composition plays in shaping MH resistance, even among viruses with similar structural characteristics. Importantly, most MHs in the moderate resistance group were predominantly studied in specific food types. As a result, their observed inactivation levels may reflect the trends relevant to those matrices rather than generalizable across all food products, underscoring the need for broader cross-matrix investigations to draw generalizable conclusions.

The **resistant group** comprised *S. aureus*, *B. cereus* (vegetative cells), each with $D_{500\text{MPa}}$ > 3 min (or $\log_{10}D_{500\text{MPa}}$ > 0.48), indicating comparatively high resistance to HPP (Table 5.1). Among them, *S. aureus* (n = 259) had a larger dataset, providing robust estimations of prediction intervals (Figure 5.2). In contrast, *B. cereus* (vegetative cells) exhibited a high z_P value, suggesting that resistance was not affected by the pressure level, contrary to the expected pressure-inactivation trend. This result is likely explained by the small dataset (n = 10) obtained from experiments with single treatment times, rather than reflecting the true kinetic behavior, highlighting the impact of rough calculation of the D-values (i.e., based on one single $\log_{10}$-reduction) and the limited number of D-values in total (n = 10).

Several MHs, such as Norovirus, *Salmonella* non-Typhi, Hepatitis E, and *B. cereus*, exhibited high z_P values when pooling the available data across all food matrices. Such patterns are common in meta-regression models. Meta-regression models are generally conservative at the high-exposure end since the heterogeneity of large, pooled datasets and the experimental bias in the source studies tend to make the fitted slopes

less steep and the corresponding z_P values higher. The opposite occurs at the low-exposure end, where resistance can be underestimated (Garre et al., 2023). At the same time, the pooled estimate represents an average response across diverse conditions, so individual observations may deviate substantially from the fitted line. For instance, for *E. coli* at 400 MPa, reported $\log_{10}D$-values vary roughly from -0.7 to 2.0 $\log_{10}$ min, while the pooled model predicts around 0.6 $\log_{10}$ min, illustrating that pooled fits can overestimate or underestimate inactivation under specific-matrix conditions.

In meta-regression models, there is always some unexplained variability involved, which partially originates from the variability in the characteristics of the food matrices, the microbial strains assessed, and the experimental bias in the original studies, but also from the unspecified practices of each research team and the specific goals of each study (Pampoukis et al., 2024). Since z_P reflects the rate of change in $\log_{10}D$ with pressure, increasing or decreasing the pressure may influence the relative classification of MHs. To assess the robustness of the 500 MPa-based resistance rankings, classifications were recalculated at 400 MPa and 600 MPa. The rankings remained largely consistent across these levels, with the MHs shifting up to 1-2 positions higher or lower when going from 500 MPa to 400 MPa or 600 MPa. The only exception was *M. bovis*, which shifted from moderate resistance positions at 500 MPa to higher resistance at 400 MPa, consistent with its relatively low z_P (stronger pressure dependence of $\log_{10}D$). The estimated $\log_{10}D$-values at 400, 500, and 600 MPa with their 95% PI, and other statistical metrics can be found in Supplementary Table 5.2. These findings confirm that the pressure-based resistance ranking of MHs is generally stable across the industrially relevant HPP range (400-600 MPa) (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022). Where EFSA reported log-linear parameter estimates for the same hazards (*Campylobacter* spp., *M. bovis*, *L. monocytogenes*, *E. coli*, *S. aureus*) in raw milk, the overall ranking of pressure sensitivity was broadly consistent with the pooled trends observed in the present dataset (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022). Since EFSA also applied a pooled analysis, we additionally estimated $\log_{10}D_{\text{ref}}$ at 600 MPa from our fitted models to enable direct comparison of D_{ref} and z_P values between the two assessments (Supplementary Table 5.3). Notably, EFSA estimates were derived from pooled inactivation data in raw milk, whereas the present fits reflect pooled observations across multiple food matrices (Table 5.1), which is expected to influence absolute parameter values. In this comparison, z_P values were broadly comparable for *M. bovis*, while EFSA reported higher D_{ref} at 600 MPa for *L. monocytogenes*, *E. coli*, and *S. aureus* than the pooled estimates derived here, whereas *Campylobacter* spp., although among the most sensitive in our dataset, showed substantially lower D_{ref} at 600 MPa in EFSA (Supplementary Table 5.3). These differences likely reflect the contrasting matrix composition and evidence base of the two pooled datasets. In addition, the present dataset included some *D*-values derived from single-$\log_{10}$-reductions, which support wider comparisons but can increase between-study

variability and parameter uncertainty in the pooled estimates.

Overall, these findings highlight that resistance classifications based solely on $\log_{10}D_{\text{ref}}$ or z_P values may result in big variations without context. Factors such as the food matrix variability, the number of observations used to derive the kinetic parameters, and the corresponding model fit must be considered when evaluating HPP resistance. This is especially critical when using pooled literature data for predictive modeling, as the extent of variability can significantly influence confidence in MH resistance classifications.

5.3.2 Impact of food matrix on $\log_{10}D$-values

The resistance ranking presented in Section 5.3.1 was standardized solely based on pressure but did not consider the influence of food matrices, which is an important limitation. Food characteristics are well-documented to significantly influence MH resistance to HPP (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022; Santillana Farakos and Zwietering, 2011; Sehrawat et al., 2020; Serra-Castelló et al., 2022; Torrents-Masoliver et al., 2025). Depending on food characteristics such as pH, a_w, and fat content, the food matrix can either enhance or reduce the efficacy of pressure-based MH inactivation. To explore this further, MH resistance was re-evaluated using stratified datasets based on food category, and the Bigelow model was fitted for each MH-food category combination (Table 5.2). MH-matrix category combinations that had fewer than 6 data points available from diverse sources and yielded negative z_P values or z_P values > 1000 MPa were excluded from Table 5.2 to avoid fits that might be too forced or misleading because they indicate either a weak pressure-inactivation relationship or even reduced inactivation at higher pressures, which is likely attributable to additional sources of variability that could not be explained through the effects of pressure-time, microbial hazard, and food category. This was primarily due to either a limited number of observations used to estimate a $\log_{10}D$-value (i.e., estimation from one single $\log_{10}$-reduction), a limited number of corresponding $\log_{10}D$-values in certain MH-food category combinations, or high variability present within these combinations driven by more specific MH characteristics (e.g., strain) or product characteristics. Food categories can vary in a wide range of pH and a_w values. For example, meat and meat products can vary from dry-cured meats with low a_w to cooked meats with high a_w, and fruits and vegetables can vary from low-acid to high-acid, which are key contributors to variability in HPP inactivation patterns.

Table 5.2: Microbiological hazard $\log_{10}D$ values at 500 MPa reference pressure in different food categories ($n \geq 6$, $0 < z_P \leq 1000$ MPa)

Microbiological Hazard	Food categories	$\log_{10} D_{500MPa}$ ($\log_{10}$ min) ± SE	z_P ± SE (MPa)	Root Mean Square Error (RMSE) ($\log_{10}$ min)	*n*
Campylobacter spp.	Meat and meat products	-0.97 ± 0.26	160 ± 30	0.46	23
Campylobacter spp.	Lab media	-0.40 ± 0.34	296 ± 113	0.18	10
Campylobacter spp.	Milk and dairy products	-0.15 ± 0.28	246 ± 77	0.30	19
Cronobacter spp.	Lab media	-0.51 ± 0.05	430 ± 51	0.20	26
Cronobacter spp.	Milk and dairy products	-0.42 ± 0.08	268 ± 28	0.33	30
E. coli	Fish and fish products	-0.12 ± 0.08	323 ± 73	0.39	23
E. coli	Meat and meat products	0.13 ± 0.04	328 ± 34	0.43	207
E. coli	Cheese and cheese products	0.31 ± 0.08	476 ± 112	0.50	78
E. coli	Lab media	0.66 ± 0.06	445 ± 78	0.44	72
E. coli	Milk and dairy products	0.80 ± 0.05	866 ± 248	0.42	90
Hepatitis A	Fish and fish products	-0.79 ± 0.19	137 ± 27	0.29	19
Hepatitis A	Fruits and vegetables	-0.36 ± 0.15	266 ± 54	0.23	13
Hepatitis A	Lab media	0.01 ± 0.13	592 ± 360	0.45	55
Hepatitis E	Lab media	0.10 ± 0.12	946 ± 705	0.35	14
Hepatitis E	Meat and meat products	0.63 ± 0.12	365 ± 148	0.43	14
L. monocytogenes	Egg and egg products	-0.13 ± 0.15	230 ± 46	0.32	15
L. monocytogenes	Lab media	-0.09 ± 0.10	668 ± 279	0.37	57
L. monocytogenes	Meat and meat products	0.07 ± 0.03	369 ± 27	0.43	182
L. monocytogenes	Fruits and vegetables	0.11 ± 0.13	410 ± 111	0.59	53
L. monocytogenes	Milk and dairy products	0.13 ± 0.06	372 ± 58	0.40	83
L. monocytogenes	Fish and fish products	0.22 ± 0.06	280 ± 47	0.37	33
L. monocytogenes	Cheese and cheese products	0.39 ± 0.03	472 ± 42	0.29	184
M. bovis	Milk and dairy products	0.30 ± 0.07	238 ± 37	0.29	35
Norovirus	Other foods	-0.65 ± 0.20	197 ± 45	0.39	20
Norovirus	Lab media	-0.26 ± 0.08	715 ± 210	0.42	129
Norovirus	Egg and egg products	0.13 ± 0.29	660 ± 600	0.37	9
Rotavirus	Lab media	-0.50 ± 0.04	507 ± 53	0.16	86
Salmonella non-Typhi	Fruits and vegetables	-0.81 ± 0.14	203 ± 29	0.35	43
Salmonella non-Typhi	Cheese and cheese products	0.08 ± 0.11	450 ± 139	0.13	15
Salmonella non-Typhi	Lab media	0.21 ± 0.14	538 ± 180	0.39	55
Salmonella non-Typhi	Meat and meat products	0.36 ± 0.04	757 ± 162	0.41	99
Salmonella non-Typhi	Milk and dairy products	0.42 ± 0.07	689 ± 264	0.35	28
Salmonella non-Typhi	Herbs and spices	0.47 ± 0.07	637 ± 217	0.31	28
S. aureus	Fruits and vegetables	-0.96 ± 0.61	121 ± 48	0.48	10
S. aureus	Lab media	0.50 ± 0.04	517 ± 53	0.30	73
S. aureus	Milk and dairy products	0.53 ± 0.05	481 ± 111	0.36	72
S. aureus	Cereal products, rice, and seeds	0.74 ± 0.13	837 ± 902	0.39	11
S. aureus	Meat and meat products	0.82 ± 0.05	270 ± 43	0.28	35
S. aureus	Cheese and cheese products	0.96 ± 0.06	873 ± 271	0.28	34
Vibrio spp.	Milk and dairy products	-1.11 ± 0.23	177 ± 32	0.12	6
Vibrio spp.	Lab media	-0.97 ± 0.11	305 ± 40	0.34	107
Vibrio spp.	Fish and fish products	-0.90 ± 0.08	291 ± 34	0.33	104
Y. enterocolitica	Meat and meat products	-0.77 ± 0.29	175 ± 34	0.28	7
Y. enterocolitica	Egg and egg products	-0.55 ± 0.39	215 ± 98	0.39	6
Y. enterocolitica	Lab media	-0.19 ± 0.19	264 ± 53	0.34	28
Y. enterocolitica	Milk and dairy products	0.40 ± 0.16	622 ± 459	0.39	19

The $\log_{10}D_{500MPa}$ values varied substantially for the same MH depending on the food category in which it was tested. For example, *S. aureus* exhibited the lowest resistance in fruits and vegetables ($\log_{10}D_{500MPa}$ = -0.96, n = 10) and the highest in cheese and cheese products ($\log_{10}D_{500MPa}$ = 0.96, n = 34) and meat and meat products ($\log_{10}D_{500MPa}$ = 0.82, n = 35). These findings suggest that fat- and protein-rich matrices may protect *S. aureus* against HPP, whereas low-fat or high-acid matrices (e.g., fruits and vegetables) support more effective microbial reduction.

Similar trends were observed for *Salmonella* non-Typhi, which showed higher resistance in milk and dairy products, and meat and meat products, and lower resistance in fruits and vegetables, highlighting the influence of matrix composition on HPP resistance (Table 5.2). For *L. monocytogenes*, resistance was highest in cheese and cheese products ($\log_{10}D_{500MPa}$ = 0.39, n = 184), while fruits and vegetables showed intermediate values that were comparable to meat and meat products and milk and dairy products (Table 5.2), indicating that matrix effects can be product-specific and cannot always be explained with broad food categories. On the other hand, *Campylobacter* spp. demonstrated a high sensitivity across all tested matrices, with negative $\log_{10}D_{500MPa}$ values in lab media ($\log_{10}D_{500MPa}$ = -0.40, n = 10), milk and dairy products ($\log_{10}D_{500MPa}$ = -0.15, n = 19), and meat products ($\log_{10}D_{500MPa}$ = -0.97, n = 23), indicating this MH is generally susceptible to HPP regardless of food matrix type.

It should be noted that in some cases, stratifying the dataset by food category did not provide additional insight beyond the generic approach presented in Section 5.3.1 for assessing MH resistance to HPP treatments, as several MH-matrix category combinations had limited or no data, preventing meaningful resistance ranking.

5.3.3 Multiple linear regression modeling of product characteristics

To quantify the combined influence of intrinsic food characteristics on microbial resistance to HPP and still retaining a generic approach independent of the food category, multiple linear regression modeling was performed using $\log_{10}D$-values as the response variable. The independent variables included pH (-), a_w category (≤ 0.95, > 0.95), fat category (≤ 10%, > 10%), and protein (%), and carbohydrate (%) content. In most studies, these food characteristics were not reported and therefore had to be assumed. However, for certain complex foods and laboratory media, imputation was not feasible. The data points with reported or imputed pH values, and a_w category were 2485 out of 2692, excluding *B. cereus* due to limited data with these variables reported (n < 6). The data points with also reported or imputed protein and carbohydrate content, and fat category were 2087. The backward elimination (see Section 5.2.3 for more

details) indicated that, despite including fat category, proteins, and carbohydrates as predictors, only pH and a_w were statistically significant in explaining the bacterial inactivation by HPP (Supplementary Table 5.4). While fat, proteins and carbohydrates are widely reported to influence bacterial resistance to HPP, the lack of significance in this study may be due to (i) imputed values not accurately reflecting the actual composition of products reported in the collected studies and/or (ii) the possibility that bacterial inactivation by HPP is more influenced by the type of proteins, and carbohydrates rather than their total amount and/or (iii) the effects are limited and overshadowed by other more influential factors. Therefore, the wider dataset with only reported or imputed pH values and a_w category was used for further analysis (n = 2485) (Table 5.3). In this dataset, *L. monocytogenes* was used as the baseline species since it was the most frequent microbial species (n = 608), followed by *E. coli* (n = 606) and *Salmonella* non-Typhi (n = 289).

Table 5.3: Estimates of the multiple linear regression for $\log_{10}D$ values according to: $\log_{10}D$ = 0.210 + 0.00216·(500 - P) - 0.0604·(7 - pH) + 0.143·(7 - pH)·($a_w \leq 0.95$) - 0.000750·(500 - P)·($a_w \leq 0.95$) + species effects (ΔP = 500 - P (MPa); pH centered as 7 - pH; ($a_w \leq 0.95$) is coded as 1 for low a_w and 0 otherwise). RMSE = 0.447, R^2adj = 0.387; AIC = 3075; $n = 2485$

Parameter	Coefficient ± SE ($\log_{10}$ min)	p-value	Predicted $\log_{10}$ D_{500MPa}, at pH 7; a_w > 0.95 ($\log_{10}$ min)	Classification (based on predicted $\log_{10}$ D_{500MPa}, at pH 7; a_w > 0.95)*
Intercept	0.210 ± 0.018	< 0.001	–	–
Vibrio spp.	-0.869 ± 0.038	< 0.001	-0.66	Sensitive
Cronobacter spp.	-0.544 ± 0.062	< 0.001	-0.33	Sensitive
Norovirus	-0.371 ± 0.033	< 0.001	-0.16	Sensitive
Hepatitis A	-0.266 ± 0.054	< 0.001	-0.056	Sensitive
Campylobacter spp.	-0.144 ± 0.068	0.035	0.066	Moderate
*L. monocytogenes***	0	–	0.21	Moderate
A. caviae	–	> 0.05	0.21	Moderate
Rotavirus	–	> 0.05	0.21	Moderate
Y. enterocolitica	–	> 0.05	0.21	Moderate
Salmonella non-Typhi	–	> 0.05	0.21	Moderate
Hepatitis E	–	> 0.05	0.21	Moderate
E. coli	0.229 ± 0.023	< 0.001	0.44	Moderate
M. bovis	0.251 ± 0.077	0.0012	0.46	Moderate
S. aureus	0.400 ± 0.033	< 0.001	0.61	Resistant
ΔP (500-P)	0.00216 ± 0.000075	< 0.001	–	–
pH (7-pH)	-0.0604 ± 0.011	< 0.001	–	–
$a_w \leq 0.95$	–	> 0.05	–	–
$pH \times a_w \leq 0.95$	0.143 ± 0.021	< 0.001	–	–
$\Delta P \times pH$	–	> 0.05	–	–
$\Delta P \times a_w \leq 0.95$	-0.000750 ± 0.00021	< 0.001	–	–

*: Classification thresholds: "Sensitive" if $\log_{10} D_{500\text{MPa}} \leq 0$ (D-value ≤ 1 min), "Moderate" if $0 < \log_{10} D_{500\text{MPa}} \leq 0.48$ ($1 < D$-value ≤ 3 min), and "Resistant" if $\log_{10} D_{500\text{MPa}} > 0.48$ (D-value > 3 min).

**: Baseline species.

The multiple linear regression modeling showed that pH, the interaction between pH and a_w, and the interaction between pressure and a_w were statistically significant variables, whereas a_w and the interaction between pressure and pH were not significant and were therefore not retained in the final model (Table 5.3). The variable a_w was always included in the models as a categorical variable with two levels (≤ 0.95, > 0.95) due to the low number of numerical values reported in the literature and the lack of possibility to impute them.

The negative coefficient for pH (centered as 7-*pH*) indicated higher $\log_{10}D$-values with increasing pH, while the influence of low a_w was captured through its interaction with pH and its interaction with pressure rather than through a standalone a_w term (Table 5.3). In practice, this means that the pH effect and the pressure effect differ between high a_w and low a_w products. These results were expected since acidic pH enhances the lethality of HPP (Torrents-Masoliver et al., 2025), and low a_w is known to exert a piezo-protective effect (Bover-Cid et al., 2015; Serra-Castelló, 2023). The separate effect of low a_w was not significant because it was already captured through its interaction with pH and its interaction with pressure. More specifically, if the interaction term of pH and a_w was not used as an independent variable, then low a_w becomes significant, but the final model has a higher AIC. The same happens when the interaction term between pressure and a_w is also excluded. Therefore, none of these options were preferred (Supplementary Table 5.4). The positive coefficient of the *pH* × low a_w ≤ 0.95 term partly offsets the standalone pH effect under low a_w conditions, indicating that the fitted pH dependence differs between low and high a_w foods and that the piezo-protective effect of reduced a_w can decrease the enhanced lethality typically associated with acidic conditions. However, under low a_w conditions, the fitted net pH slope becomes slightly positive, which likely reflects limited and uneven coverage of pH-a_w combinations in the underlying data rather than a true reversal of the pH effect (Table 5.3).

These findings indicate that multiple linear regression modeling can provide additional, descriptive insight into MH inactivation by HPP by accounting for intrinsic product characteristics that are not considered in the pooled ranking of Section 5.3.1 and are not limited to the food categories of Section 5.3.2. However, these associations should be interpreted as descriptive rather than mechanistic. In this framework, MHs were classified as sensitive ($\log_{10}D_{500MPa} \leq 0$; $D_{500MPa} \leq 1$ min), moderate ($0 < \log_{10}D_{500MPa} \leq 0.48$; $1 < D_{500MPa} \leq 3$ min), or resistant ($\log_{10}D_{500MPa} > 0.48$; $D_{500MPa} > 3$ min) under the reference condition of 500 MPa, pH 7, and high a_w (Table 5.3). Compared with the pooled ranking (Table 5.1), some MHs that were classified as sensitive based on pooled estimates were here classified as moderate once pH and a_w effects were accounted for. For example, *Campylobacter* spp. shifted from $\log_{10}D_{500MPa} = -0.59$ in the pooled analysis (Table 5.1) to a predicted $\log_{10}D_{500MPa} = 0.066$ under the reference

condition in the multiple linear regression (Table 5.3), which is only slightly above the sensitive threshold. In addition, several MHs, such as *A. caviae* and Rotavirus, were not statistically distinguishable from the baseline species (*L. monocytogenes*) in the multiple linear regression and were therefore assigned the baseline prediction under the reference condition (Table 5.3), shifting them from the sensitive group in Table 5.1 to the moderate group in Table 5.3. For completeness, the full model including all species terms as well as other multiple linear regression models with higher AIC or lower number of data points due to the addition of parameters that were not always reported in the corresponding studies is provided in Supplementary Table 5.4. Overall, these results indicate that incorporating intrinsic product characteristics reduces unexplained variability and can lead to more conservative and standardized resistance classification when compared to pooled estimates that do not account for matrix composition.

5.3.3.1 Application of results in a decision-support framework

The combined findings from the general resistance ranking, food category stratification, and multiple linear regression modeling presented in this study can be integrated into a decision-support framework for evaluating the inactivation of MHs by HPP. This framework, illustrated in Figure 5.3, consists of three complementary approaches that can be applied depending on data availability and the required level of specificity, with Approach 1 intended for a broad benchmark when the product is not well represented by the available food categories, Approach 2 for products that align with a represented category, and Approach 3 for pH- and a_w-adjusted predictions when pH and a_w information is available for bacterial hazards covered by the regression model (excluding *B. cereus*). In more detail, the three approaches are outlined below, moving from a broad benchmark to product-specific estimates.

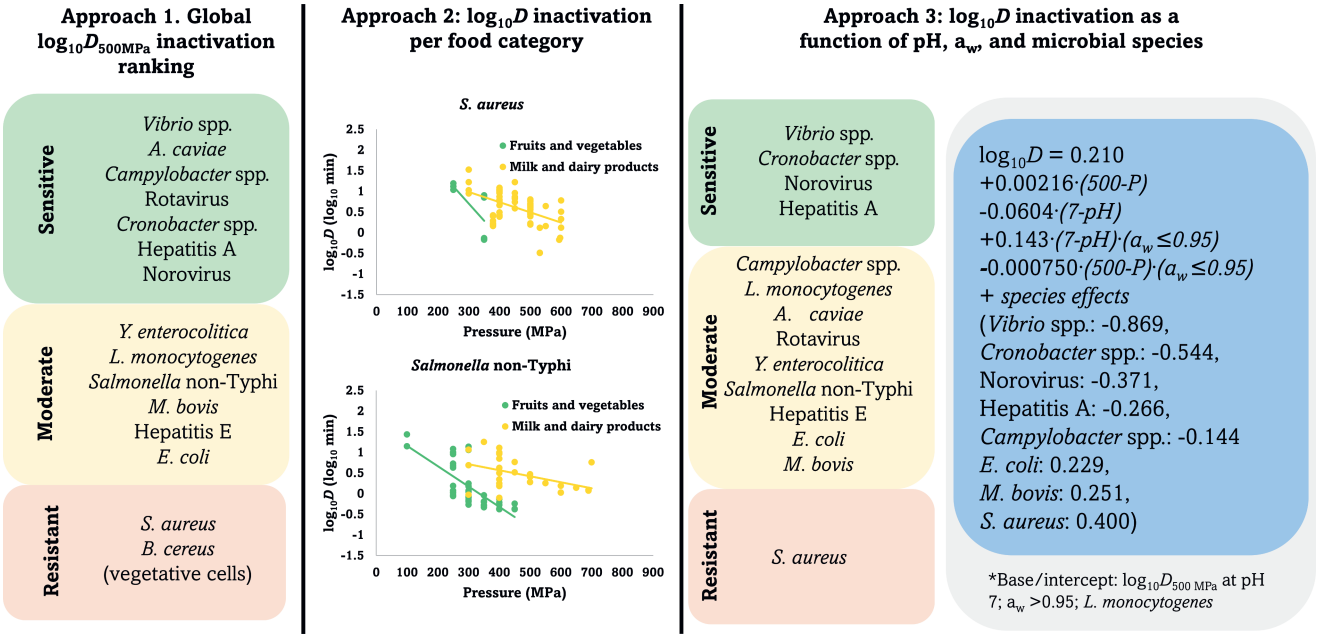


Figure 5.3: Microbiological hazard resistance to high-pressure processing based on three complementary approaches: **Approach 1:** Estimation of $\log_{10}D$-values at a reference pressure of 500 MPa for global MH resistance ranking (Section 5.3.1, Table 5.1); **Approach 2:** Matrix-specific $\log_{10}D$ inactivation values stratified by food category (Section 5.3.2, Table 5.2); **Approach 3:** $\log_{10}D$ inactivation values modeled as a function of pH, water activity (a_w) and microbial species using multiple linear regression (Section 5.3.3, Table 5.3).

Approach 1 provides a general estimation of MH inactivation trends and enables resistance ranking across bacteria and viruses using $\log_{10}D_{\text{ref}}$ values standardized at 500 MPa ($\log_{10}D_{500\text{MPa}}$) through simple linear regression (Bigelow model fit) to each microbial hazard's data. This approach offers a rapid, high-level overview and is useful for broad prioritization without making any assumptions about food-related categories. However, since the underlying data are compiled from diverse food matrices, the estimates may be influenced by overrepresented matrix types, potentially limiting their applicability in specific products.

Approach 2 allows for context-specific estimation of MH inactivation by fitting Bigelow models to stratified datasets based on food categories, thereby providing insight into matrix effects. This approach is most informative when sufficient data exist for the food category of interest, and the intended application aligns with that category. However, model robustness can be limited by small sample sizes, uncertainty when D-values are derived from only two observations (a single $\log_{10}$-reduction), and residual variability in product composition within each category. Importantly, these models are category-specific and cannot be used as generic models beyond the food categories represented in the dataset.

Approach 3 focuses on estimating bacterial inactivation as a function of defined product characteristics and microbial species using multiple linear regression with the available data. This method enables estimating the $\log_{10}D_{500\text{MPa}}$ values in one regression while accounting for pH and a_w effects, to reflect actual food matrix characteristics and the microbial species sensitivity, thereby supporting refined resistance ranking in defined product contexts.

Together, these three approaches can be used in a complementary manner to assess the efficacy and robustness of HPP treatments to inactivate MHs under different conditions. These layered frameworks support food safety decision-making and can be applied individually or in combination to evaluate HPP efficacy under varying levels of complexity, from generalized assessments to product-specific risk evaluation. Moreover, the framework highlights areas where data gaps persist, especially for viruses and parasites and underrepresented matrices, guiding future research priorities.

5.4 Conclusion

This study assessed the resistance of microbiological hazards (MHs), including foodborne bacteria and viruses, to high-pressure processing (HPP), considering both pressure intensity and the influence of the food matrix. The results confirm that vegetative

bacteria and viruses exhibit varying levels of resistance, which are significantly affected by the composition and characteristics of the food matrix. External validation through a product-oriented experimental approach would be useful for a better understanding of the food matrix effects. The decision-support framework established in this study, based on three complementary approaches (global resistance ranking, food category-stratified estimates, and multiple linear regression modeling accounting for matrix characteristics and microbial species), enables a structured assessment of MH resistance to HPP treatments under different data availability scenarios. These insights support food safety risk analysis by enabling generic hazard risk ranking and providing a quantitative benchmark to evaluate the efficacy of HPP as an intervention strategy for achieving required levels of microbial reduction, while emphasizing the need for consistent reporting of key metadata, including strain and physiological state, processing conditions, and food matrix characteristics. Future research should aim to expand available datasets with more kinetic data under various processing conditions and refine predictive models to improve the accuracy of microbial inactivation estimates across a broader range of food products.

CRediT authorship contribution statement

Kah Yen Claire Yeak: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing - Original Draft, Writing - Review & Editing, Validation, Visualization. **George Pampoukis**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Cristina Serra-Castelló**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing. **Sara Bover-Cid**: Conceptualization, Data Curation, Funding Acquisition, Supervision, Validation, Writing - Review & Editing. **Heidy M.W. den Besten**: Conceptualization, Data Curation, Funding Acquisition, Supervision, Validation, Writing - Review & Editing. **Marcel H. Zwietering**: Conceptualization, Data Curation, Funding Acquisition, Supervision, Validation, Writing - Review & Editing.

Acknowledgements

We thank Zhu Yujie for her assistance in reviewing some of the literature and extracting part of the inactivation data in this study.

Data availability

No supplementary material is reproduced in the thesis body. Supplementary Material A and Supplementary Table 5.1, Supplementary Table 5.2, Supplementary Table 5.3, and Supplementary Table 5.4 are available via 4TU.ResearchData (DOI: 10.4121/3cab8862-d524-4bcd-bee1-3318525a7311).

Funding

Kah Yen Claire Yeak conducted the work under the SAFFI project, with funding received from the European Union's Horizon 2020 research and innovation program under grant agreement No. 861917. The project of George Pampoukis is part of the TRANSIT project that received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N° 955431. Sara Bover Cid thanks the support of Consolidated Research Group (SEQUSAL 2021 SGR 00468) by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) and CERCA Program of the Generalitat de Catalunya.

Chapter 6

General Discussion

6.1 Consumer preferences and the processing dilemma

Consumers often have conflicting interests regarding food. They want to save cooking time and buy convenience foods, and at the same time, they prefer foods that are perceived as natural and minimally processed (Román et al., 2017). "Naturalness" and "clean label" are widely used concepts in consumer research and in product development, but they are not designed to function as scientific descriptors of nutritional quality. In practice, they are often associated with familiar ingredients and the limited use of chemical additives and processing aids, but they do not necessarily align with microbiological safety, nutrient retention, or overall nutritional value.

Processing remains essential for controlling hazards, extending shelf life, and reducing food loss. Its effects on the nutritional value depend on the product and conditions. Heat treatment can increase the bioavailability of some compounds, such as lycopene in tomatoes (Porrini et al., 1998), while it may reduce heat-sensitive vitamins and other constituents in other foods. In legumes (soy, beans, peas, etc.), thermal processing can inactivate trypsin inhibitors, which are among the most relevant naturally occurring antinutritional factors, and thereby improve protein digestion (Avilés-Gaxiola et al., 2018).

This creates a practical dilemma for product developers. Not accounting for consumer expectations for naturalness and clean label can reduce acceptance and market success (Román et al., 2017), yet these expectations do not necessarily align with evidence on microbiological risk control or on the nutritional consequences of processing. Mild or non-thermal processing technologies are often positioned as a more sustainable way to assure safety, shelf life, and fresh-like sensory attributes by reducing thermal load and, in some cases, reducing the use of chemical preservatives. Whether they deliver on this proposition depends on (i) demonstrated decontamination efficacy under relevant use conditions, (ii) the resource demands of the process, including energy use and greenhouse gas emissions, and (iii) the robustness and comparability of the supporting evidence.

This chapter first summarizes the main contributions of the thesis, then defines how evidence was made comparable across technologies, and then applies the fitted models in a case study that links estimated decontamination efficacy with published energy use and greenhouse gas emissions. Finally, it discusses implications for fair technology comparison and outlines future research directions.

6.2 Scope and main contributions of this thesis

This thesis aimed to obtain quantitative, generic insights into the decontamination efficacy of non-thermal plasma (NTP), pulsed electric fields (PEF), and high-pressure processing (HPP) under relevant food processing conditions, and to place these insights within a broader context of food safety and sustainability. Across the chapters, the main contributions can be summarized as follows.

First, for NTP, a curated database of decontamination studies was compiled and used to rank influential process, product, and microbial factors using machine learning (Chapter 2). Because the available literature was too heterogeneous to support robust global kinetic parameter estimation, this chapter focused on factor ranking rather than on deriving generic *D*-values. The analysis showed that dissipated power per plasma volume and matrix category, followed by microbial genus and pH, explain a substantial part of the observed variability, and it highlighted major reporting gaps that currently limit quantitative synthesis.

Second, these insights motivated the development of a dedicated metadata schema for NTP decontamination in food-related applications (Chapter 3). The schema extended an existing plasma metadata framework with food-specific descriptors and minimum information requirements, and it provided human- and machine-readable templates for reporting. Its purpose was to make future NTP studies more FAIR (findable, accessible, interoperable, and reusable) and to enable meta-regression once sufficiently consistent information has accumulated.

Third, for PEF, a harmonized database was assembled across multiple microbial groups and matrices (Chapter 4). Decontamination efficacy was expressed consistently in terms of the decimal-reduction volumetric energy input $D_{E_{\mathrm{vol}}}$, and meta-regression was used to estimate global and matrix-specific kinetic parameters for spoilage microorganisms and pathogens. The chapter also introduced an interactive dashboard to support exploratory analysis and developed a decision tree for estimating the volumetric energy input required to reach a 5-$\log_{10}$-reduction across microorganisms and matrix categories, including prediction intervals.

Fourth, for HPP, previously developed databases (Santillana Farakos and Zwietering, 2011; Torrents-Masoliver et al., 2025) were updated and extended with additional studies to cover 15 microbial hazards, including viruses (Chapter 5). Global $\log_{10}D$-values were estimated at non-thermal temperatures and linked to an industrially relevant reference pressure (500 MPa), allowing hazards to be grouped into resistance categories while accounting for matrix effects. The resulting benchmarks provided literature-based starting points for process design, validation, and quantitative microbiological risk assessment.

Finally, this general discussion connects the chapter outputs with thermal processing benchmarks and with published information on energy use and greenhouse gas emissions, and it applies the fitted models in a case study to interpret sustainability indicators together with expected decontamination performance.

6.3 Meta-regression case study: efficacy, energy use, and greenhouse gas emissions for thermal processing, HPP, and PEF

6.3.1 Rationale, scope, and comparison principle

Mild processing technologies differ in underlying principles and in the types of products they can treat. NTP is mainly a surface treatment for solid foods such as spices and powders, or it is applied indirectly via plasma-treated water for washing fresh produce. PEF targets liquid products such as fruit juices and plant-based beverages, and it is typically implemented as a continuous flow process. HPP is used for both liquid and solid foods, usually in batch mode and often in-pack. These differences in application domain and operational mode limit direct comparability, and they influence which technologies can be evaluated within the same product category. Similar technology-specific constraints apply to other mild processing technologies, such as ultrasound (Beitia et al., 2023). For this reason, the quantitative comparison in the present case study focuses on technologies that are commonly applied to fruit juices, namely thermal processing, HPP, and PEF, and on a product-microorganism combination for which sufficient data were available to fit the meta-regression models used below.

Building on the rationale above, a high-acid fruit juice application was selected (pH ≤ 4.5) with *E. coli* as the target microorganism. High-acid fruit juices are a relevant application for mild technologies because freshness and minimal processing are commonly emphasized in this product category (Martins et al., 2020; Tiwari et al., 2009). *E. coli* is a relevant hazard for juices and has been implicated in outbreaks linked to high-acid fruit juices and ciders (Martínez-Gonzáles and Castillo, 2016). In addition, regulatory guidance for juice HACCP uses *E. coli* among the pertinent microorganisms for demonstrating a 5-$\log_{10}$-reduction immediately before or after packaging (U.S. Food and Drug Administration, 2004).

Across emerging technologies, the supporting evidence also differs in completeness and consistency of reporting, which directly affects what can be synthesized and

benchmarked. In this thesis, reporting guidelines and available descriptors supported global parameter estimation for PEF and broader benchmarking for HPP, whereas heterogeneous reporting in NTP currently limits global *D*-value estimation and motivated the development of a dedicated metadata schema (Pampoukis et al., 2025; Raso et al., 2016; Santillana Farakos and Zwietering, 2011; Torrents-Masoliver et al., 2025). Together, these chapters show that technology-specific minimum information requirements are a prerequisite for synthesis, benchmarking, and process optimization.

To make the evidence base as comparable as possible, this thesis used a transparent, reversible data wrangling pipeline to standardize units and fields, apply domain filters, exclude ambiguous entries (for example, values below detection limits), validate reported *D*-values where possible, and de-duplicate records (Rattenbury et al., 2017). Exploratory analyses were used to detect outliers and guide model development (Chatfield, 1986). The modeling strategy then used meta-regression to estimate global kinetic parameters when supported by the data and focused on factor ranking to identify key drivers when data limitations and heterogeneity prevented robust quantitative parameter estimation.

Sustainability comparisons in the literature often report energy demands or greenhouse gas (GHG) emissions per liter or kilogram of product and then infer which technology is "more sustainable". When processing is used to assure food safety, this is only interpretable if the processes deliver comparable decontamination performance for the same microorganism and product category. In practice, published case studies often use different target reductions and different process constraints, so a technology can appear energy efficient per liter while delivering substantially lower lethality, or it can deliver a much higher lethality than intended.

In this thesis, energy demands and GHG emissions are interpreted together with the estimated decontamination efficacy for the same microorganism and product category. This is necessary because energy per liter can be compared meaningfully only when the achieved lethality is comparable.

Using the reported energy demand and greenhouse gas emissions from each study, for the specific processing conditions evaluated in that study, avoids extrapolating sustainability indicators to other operating points that were not assessed. An alternative would be to fix a common energy demand or greenhouse gas emissions level and compare the estimated reductions, but in applications, the starting point is typically a target reduction, often a 5- or 6-$\log_{10}$-reduction, and this would require extrapolating to operating points that were not evaluated in the original studies. This was not used here because energy demand and greenhouse gas emissions are not estimated in a comparable way across technologies and studies, and the underlying assumptions

differ. Additionally, for PEF, the dose metric chosen in this thesis is the volumetric energy input, whereas the reported electricity demand depends on system efficiency and losses that vary across setups. For thermal processing and HPP, reported energy demand depends on the full time-temperature or time-pressure profile and on equipment and heat recovery assumptions that are often not reported consistently. "Efficacy matched" therefore refers to comparisons in which the evaluated operating points are expected, on average, to deliver similar microbial $\log_{10}$-reductions for the same microorganism in the same product category.

For TP, the starting point was an existing database of *D*-values (Van Asselt and Zwietering, 2006), which was updated for *E. coli* in high-acid fruit juices (pH ≤ 4.5) (Van der Sluis, 2025). On this filtered subset (n = 86, 13 studies), primary and secondary linear inactivation models were fitted (Bigelow, 1921):

$$\log_{10}\left(\frac{N_0}{N}\right) = \frac{t}{D} \tag{6.1}$$

$$\log_{10} D = \log_{10} D_{\text{ref}} + \frac{T_{\text{ref}} - T}{z} \tag{6.2}$$

where $\log_{10}\left(\frac{N_0}{N}\right)$ are the $\log_{10}$-reductions of the microbial population after processing, t is the processing time, D is the time needed for a ten-fold reduction of the microbial population at the temperature T, D_{ref} is the D-value at a reference temperature T_{ref}, and z is the temperature increase required to reduce D by one $\log_{10}$. Combining both gives a single expression for the reduction as a function of temperature and time:

$$\log_{10}\left(\frac{N_0}{N}\right) = \frac{t}{D_{\text{ref}} \cdot 10^{\frac{T_{\text{ref}}-T}{z}}} = \frac{t}{D_{\text{ref}}} \cdot 10^{-\frac{T_{\text{ref}}-T}{z}}. \tag{6.3}$$

The fitted secondary model for TP in high-acid fruit juices was:

$$\log_{10} D = -1.4 + \frac{71 - T}{7.7} \tag{6.4}$$

which expresses $\log_{10} D$ (with D in minutes) as a linear function of temperature, with 71 °C as the reference temperature and a z-value of 7.7 °C. The temperature of 71 °C (160 °F) is a relevant temperature for pasteurization treatment of fruit juices since it is expected to achieve a 5-$\log_{10}$-reduction of the pertinent microorganism in a matter of seconds (U.S. Food and Drug Administration, 2004). These parameter estimates were

introduced in Equation 6.3 to obtain the microbial reductions under different time-temperature combinations (Figure 6.1).

6.3.2 High-pressure processing

For HPP, the corresponding subset of the HPP database compiled in Chapter 5, filtered for *E. coli* in high-acid fruit juices (pH ≤ 4.5) (n = 37, 6 studies), was used to fit an analogous secondary model in which pressure P replaces temperature. A reference pressure of 500 MPa was chosen because it lies in the middle of the pressure range typically applied in industrial practice (EFSA Panel on Biological Hazards (BIOHAZ) et al., 2022):

$$\log_{10} D = -0.39 + \frac{500 - P}{245} \tag{6.5}$$

with D in minutes. Here, 500 MPa is the reference pressure, and 245 MPa corresponds to the pressure z-value (z_P). An increase of 245 MPa reduces the D-value by one $\log_{10}$. As for TP, this secondary model and its parameter estimates can be combined with the primary inactivation equation to obtain $\log_{10}$-reductions for given pressure-time combinations (Equation 6.3) as visualized in Figure 6.1.

6.3.3 Pulsed electric field

For PEF, the case study used the PEF database compiled in Chapter 4, filtered for *E. coli* in high-acid fruit juices (pH ≤ 4.5) (n = 34, 4 studies). In this case, the inactivation was modeled directly as a function of the volumetric energy input E_{vol}, which aggregates the main process parameters into a single dose metric:

$$E_{\text{vol}}\,[\text{J/cm}^3] \;=\; E_f^2 \cdot f \cdot \tau \cdot \sigma \cdot R_t \cdot n_c, \qquad (1\,\text{J/cm}^3 = 1\,\text{kJ/L}) \tag{6.6}$$

where E_f is the electric-field strength (kV/cm); f is the frequency/pulse-repetition rate (pulses per second); τ is the pulse width (µs); σ is the electrical conductivity of the food at the inlet temperature (S/cm); R_t is the residence time in the PEF unit (s); n_c is the number of PEF chambers in series (-). On this basis, the primary inactivation model fitted was:

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) = \frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} \tag{6.7}$$

where $D_{E_{\text{vol}}}$ is the decimal-reduction volumetric energy input (kJ/L), i.e., the energy input needed for a one $\log_{10}$-reduction. Here, E_{vol} is a dose metric delivered to the product, and it should not be interpreted as the total electricity demand of a specific PEF system. For *E. coli* and the high-acid fruit juice subset, the fitted relationship was:

$$\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right) = \frac{E_{\text{vol}}}{32} \tag{6.8}$$

so that $D_{E_{\text{vol}}} = 32$ kJ/L for *E. coli* in high-acid fruit juices based on this subset of the PEF database. Because E_{vol} already combines multiple process parameters in one, such as electric field strength, pulse width, frequency, conductivity, residence time, and the number of chambers, no additional secondary model was required.

6.3.4 Model performance, uncertainty, and units

For all three technologies, model performance was quantified using the root mean square error (RMSE) of the fitted meta-regression models and the 95% prediction intervals (PI) for the microbial $\log_{10}$-reduction at given processing conditions.

For TP and HPP, the fitted models were the secondary meta-regression equations as shown in Equation 6.4 and Equation 6.5, where $\log_{10} D$ was expressed as a function of temperature T (TP) or pressure P (HPP) for *E. coli* in high-acid fruit juice datasets with an intercept and slope. Denoting the predictor generically by X_i (temperature or pressure), the RSS and RMSE were:

$$\text{RSS} = \sum_{i=1}^{n} \left(\log_{10} D_i - (\alpha + \beta X_i)\right)^2 \tag{6.9}$$

$$\text{RMSE} = \sqrt{\frac{\text{RSS}}{n-2}} \tag{6.10}$$

where α is the intercept, β the slope, and $n-2$ the degrees of freedom for a simple linear regression with two fitted parameters (Montgomery et al., 2021).

For a given processing condition X_0 (temperature or pressure) used in the case study, the 95% PI for a new $\log_{10} D$ observation at X_0 was:

$$\log_{10} D(X_0) \pm t_{0.975,n-2} \cdot \text{RMSE} \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\text{SSX}}} \tag{6.11}$$

where $\log_{10} D(X_0) = \alpha + \beta X_0$ is the fitted mean value of $\log_{10} D$ at X_0 (the point on the regression line), $\bar{X}$ is the mean of the predictor, and $\text{SSX} = \sum_{i=1}^{n}(X_i - \bar{X})^2$.

These prediction limits for $\log_{10} D$ were then back-transformed to obtain lower and upper bounds for the D-value at the condition in question (X_0):

$$D_{\text{lower}} = 10^{\log_{10} D(X_0) - \Delta}, \qquad D_{\text{upper}} = 10^{\log_{10} D(X_0) + \Delta} \tag{6.12}$$

with

$$\Delta = t_{0.975,n-2} \cdot \text{RMSE} \cdot \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\text{SSX}}} \tag{6.13}$$

For each specific time-temperature (TP) or time-pressure (HPP) combination used in the comparison table, these D_{lower} and D_{upper} values were substituted into the primary inactivation model (Equation 6.1) to obtain the corresponding lower and upper 95% prediction limits for the $\log_{10}$-reduction at that setting. The mean $\log_{10}$-reductions were obtained from Equation 6.1, with $D = 10^{\log_{10} D(X_0)}$.

For PEF, the primary model explained in Equation 6.7 was fitted as a regression through the origin, with $\log_{10}$-reductions as response and E_{vol} as predictor. The residual sum of squares (RSS) was:

$$\text{RSS} \;=\; \sum_{i=1}^{n} \left(\log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)_i - \beta \cdot E_{\text{vol},i}\right)^2 \tag{6.14}$$

where the slope $\beta = 1/D_{E_{\text{vol}}}$ was estimated by ordinary least squares, and the fitted values were $y_i = \beta x_i$. The RMSE was then:

$$\text{RMSE} \;=\; \sqrt{\frac{\text{RSS}}{n-1}} \tag{6.15}$$

with n the number of observations and $n - 1$ the degrees of freedom for a regression through the origin (Eisenhauer, 2003). The 95% PI for the $\log_{10}$-reductions at a given volumetric energy input E_{vol} were obtained as (Montgomery et al., 2021):

$$\frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} \pm t_{0.975,n-1} \cdot \text{RMSE} \cdot \sqrt{1 + \frac{E_{\text{vol}}^2}{\text{SSX}}} \tag{6.16}$$

where $\frac{E_{\text{vol}}}{D_{E_{\text{vol}}}} = \log_{10}\left(\frac{N_0}{N(E_{\text{vol}})}\right)$ is the predicted $\log_{10}$-reduction at E_{vol}, and $\text{SSX} = \sum_{i=1}^{n} E_{\text{vol},i}^2$. For the numerical examples in this chapter, this expression was evaluated at the E_{vol} values corresponding to each PEF condition, and the mean and 95% prediction limits were reported.

RMSE was 0.36 for TP and 0.37 for HPP on the $\log_{10}D$ scale, and 1.06 for PEF on the $\log_{10}$-reduction scale. The prediction intervals above show how this uncertainty propagates to expected reductions at specific processing conditions. The fitted models are visualized in Figure 6.1 as mean estimated $\log_{10}$-reductions across time-temperature combinations for TP, time-pressure combinations for HPP, and volumetric energy input levels for PEF, with contours for 3, 5, 7, and 9 $\log_{10}$-reductions. The corresponding 95% PI around the case study operating points, together with the RMSE and other regression diagnostics, are provided in the Supplementary Material.

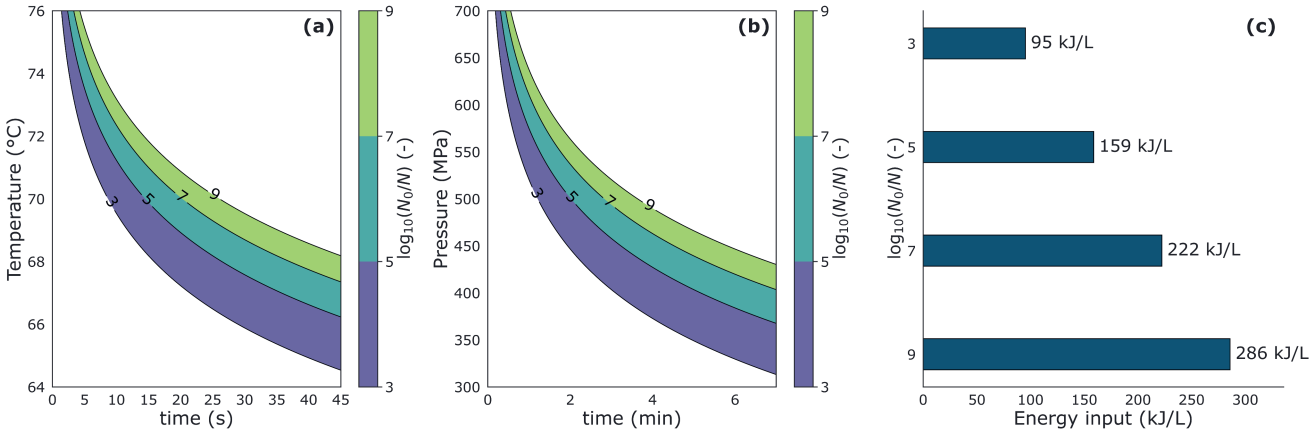


Figure 6.1: Decontamination efficacy for *E. coli* in high-acid fruit juices: (a) TP (n = 86, 13 studies), (b) HPP (n = 37, 6 studies), and (c) PEF (n = 34, 4 studies) to achieve 3, 5, 7, and 9 $\log_{10}$-reductions.

6.3.5 Linking sustainability indicators to predicted efficacy

Reported energy demand values expressed in kJ/kg in the sustainability comparison studies were converted to kJ/L assuming a density of approximately 1.03 kg/L for high-acid juices (Charrondiere et al., 2012). These converted values were then linked to the estimated reductions from the meta-regression models. For PEF, the reported energy demand, including losses, is study-specific and does not equal the volumetric energy input used as the dose metric in the meta-regression.

Aganovic et al. (2017) reported the electricity use for a tomato juice pasteurization scenario as the kWh consumed during one hour of operation, at a continuous production throughput of 120 L/h. To express this on a per-liter basis, the value was converted to kJ/L by multiplying by 3600 kJ/kWh and dividing by 120 L processed in that hour (Aganovic et al., 2017).

The reported operating points were not matched on energy demand. They reflect the specific pilot settings chosen in that study for each technology, so the energy per liter differs by design. The fair comparison in this thesis is therefore not based on equal energy input, but on interpreting the reported energy demand together with the estimated $\log_{10}$-reduction for the same microorganism and product category, and then examining the normalized indicators (energy per mean $\log_{10}$-reduction) within each study.

Under the Aganovic et al. (2017) scenarios, the electricity demand was 136.8 kJ/L for TP at 74 °C for 30 s with 70% heat recovery, 652 kJ/L for HPP at 600 MPa for 5 min with a 36.5% filling ratio, and 433 kJ/L for the PEF system. For PEF, the study also reported the treatment dose delivered to the product as 188 kJ/kg, which corresponds to 194 kJ/L assuming a juice density of 1.03 kg/L. The difference between 433 kJ/L and 194 kJ/L is not a contradiction. It reflects that electricity demand includes system inefficiencies and losses, while the volumetric energy input is the dose metric in the product used for the inactivation model (Aganovic et al., 2017). Because these values reflect the specific pilot setup, including heat recovery and filling ratio specifications, they are most informative for within-study comparisons.

To interpret these energy values in terms of food safety performance, each reported scenario was evaluated with the meta-regression models for high-acid fruit juices and *E. coli*. According to Aganovic et al. (2017), the aforementioned process conditions are expected to achieve at least 5-$\log_{10}$-reductions at the treated product. Predicted lethality and 95% PI using the meta-regression models were 34 [5.9, 202] $\log_{10}$-reductions for TP, 32 [4.6, 215] for HPP, and 6.1 [3.8, 8.4] for PEF. The wide asymmetric intervals for TP and HPP arose because uncertainty was estimated on $\log_{10}D$ and then back-transformed through the primary model, which yields multiplicative uncertainty in

reductions. For PEF, uncertainty was estimated directly on reductions, which yields symmetric prediction limits by construction.

Many sustainability studies compare technologies on an "energy per liter" basis under scenarios that each aim to meet a common microbial target (often 5-$\log_{10}$-reduction). However, scenarios that satisfy the same nominal target can still differ in the achieved lethality. In the Aganovic et al. (2017) operating points, TP, HPP, and PEF are all expected to meet at least 5-$\log_{10}$-reduction of *E. coli* in high-acid fruit juice, but the meta-regression models estimate markedly different mean reductions across the three technologies. Comparing energy or greenhouse gas emissions per liter alone would therefore mix two effects: process efficiency and the safety margin delivered under the chosen settings.

This thesis evaluates two quantities for each reported operating point: (i) the reported energy use and greenhouse gas emissions per liter (as given by the original study), and (ii) the estimated mean $\log_{10}$-reduction for the same microorganism and product category (from the fitted models). The resulting normalized indicators, "Energy per mean $\log_{10}$-reduction" and "GHG per mean $\log_{10}$-reduction", are the primary comparison metrics used here within each study. They answer the question: for the operating point reported in that study, how much energy or climate impact is associated with one estimated $\log_{10}$-reduction?

Table 6.1 summarizes these normalized indicators alongside the underlying energy/GHG values and estimated reductions. Under the Aganovic et al. (2017) scenarios, TP shows the lowest energy per mean $\log_{10}$-reduction, followed by HPP and then PEF, because TP combines relatively low energy per liter (with heat recovery) with high lethality, whereas PEF is closer to the 5-$\log_{10}$ target on average (Aganovic et al., 2017). Energy per mean $\log_{10}$-reduction and GHG per mean $\log_{10}$-reduction are not universal sustainability metrics, and they should not be interpreted as standalone measures of sustainability but as two of the metrics of efficacy that could be of relevance. Their role here is diagnostic, based on sustainability data available in the literature and the global kinetic parameters estimated in this thesis.

Table 6.1: Sustainability insights for TP, HPP, and PEF interpreted together with estimated decontamination efficacy. "Energy (kJ/L)" and "GHG (g CO_2-eq/L)" were taken from each study, and when studies reported these metrics in different units, the values were converted to kJ/L and g CO_2-eq/L for consistent presentation in this table. "Microbial reduction (mean [95% PI])" was estimated from the meta-regression models developed in this thesis. The columns "Energy per mean $\log_{10}$-reduction" and "GHG per mean $\log_{10}$-reduction" are the within-study comparison indicators used to interpret sustainability together with efficacy

Reference	Processing method	Settings / scenario	Microbial reduction (mean [95% PI])	Energy (kJ/L)	GHG (g CO_2-eq/L)	Energy per mean $\log_{10}$-reduction (kJ/L)	GHG per mean $\log_{10}$-reduction (g CO_2-eq/L)
Aganovic et al. (2017) (tomato juice)	TP	74 °C, 30 s; 70% heat recovery	34 [5.9, 202]	136.8	-	4.0	-
	HPP	600 MPa, 5 min; 36.5% fill	32 [4.6, 215]	652	-	20.4	-
	PEF	188 kJ/kg	6.1 [3.8, 8.4]	433	-	71.0	-
Ge et al. (2024) (orange juice)	TP	62 °C, 30 min	56 [11, 297]	373	-	6.7	-
	HPP	600 MPa, 3 min; 36% fill	19 [2.8, 129]	645	-	34	-
Atuonwu et al. (2018) (orange juice)	TP	76.8 °C, 15 s	40 [6.5, 245]	470	24	12	0.6
	HPP	600 MPa, 3 min; 95% fill	19 [2.8, 129]	256	15	13	0.79
Cacace et al. (2020) (orange juice)	TP	95 °C, 15 s (orange juice)	9540 [1058, 86021]	-	160	-	0.017
	HPP	600 MPa, 6 min	38 [5.6, 258]	-	47	-	1.2
Sampedro et al. (2014) (orange juice)[a]	TP	85 °C, 5 s; 96% heat recovery	157 [22, 1119]	-	5.5	-	0.42
	HPP	550 MPa, 1.5 min	5.9 [0.9, 38]	-	47	-	8.0
Landi et al. (2025) (orange juice)	TP	90 °C, 15 s; 35% heat recovery	1000 [133, 7517]	-	31	-	0.031
	PEF[b]	55 °C preheat + 38 kJ/kg	$\geq$1.2 [0, 2.6]	-	22	-	$\leq$18

[a]: To obtain the GHG (g CO_2-eq/L), the values provided in the article were divided by the yearly production of 16,500,000 L and expressed in g CO_2-eq/L.
[b]: The thermal effects in this study could not be neglected, and this is why inequality symbols were added to the estimates.

The same interpretation applies across studies: differences in reported energy per liter can reflect different safety margins rather than intrinsic process efficiency. The normalized indicators in Table 6.1 highlight where apparent advantages on a per-liter basis change once predicted lethality is taken into account (e.g., Atuonwu et al. (2018); Ge et al. (2024)).

GHG emissions data were also compiled from published life cycle assessment studies for the different technologies (Van der Sluis, 2025). Life cycle assessment quantifies environmental burdens across a defined system boundary, and climate impact is commonly reported in g CO_2-eq per functional unit, here per liter of product. GHG emissions are relevant because they are frequently used to summarize climate impact and because electricity-driven processes may change the climate impact ranking depending on the way in which electricity is produced. Reported values vary across studies due to differences in system boundaries, estimation methods, production scale, and equipment assumptions (Van der Sluis, 2025). Two studies reported lower greenhouse gas emissions for HPP than for TP under their analysis frameworks (Atuonwu et al., 2018; Cacace et al., 2020), whereas one study reported lower emissions for TP than for HPP (Sampedro et al., 2014). Our meta-regression models indicate that the compared scenarios were not efficacy matched, which means that GHG emissions per liter cannot necessarily be interpreted as evidence of superior sustainability unless the achieved safety performance is comparable and therefore the experimental conditions are equivalent (Table 6.1).

For PEF, the only life cycle assessment in the reference set that assessed the GHG emissions combined preheating at 55 °C with a PEF dose of 38 kJ/kg, and reported lower GHG emissions than TP at 90 °C for 15 s (Landi et al., 2025). This comparison requires additional care because thermal effects cannot be separated. The preheat contributes directly to lethality, Joule heating occurs during the PEF step, and the combined effects of mild heat and PEF occur. In Table 6.1, the estimated reduction attributed to the PEF dose alone is therefore expressed with inequality symbols, because the study reports a heat-assisted PEF process designed to meet a 5-$\log_{10}$-reduction target, and the available information does not allow strict attribution of the observed reductions to PEF alone.

Across the available studies, the normalized indicators in Table 6.1 tend to favor TP over HPP and HPP over PEF within each study, while highlighting that operational factors such as heat recovery for TP, filling ratio for HPP, and inlet temperature for PEF strongly influence this ranking.

A common strategy to increase efficacy while limiting energy demand is the use of hurdles (Giannoulis and Karatzas, 2024). Hurdle technology is common in food science applications, e.g., pressure-thermal combinations (Rastogi et al., 2007) or PEF with preheating as shown above, but also in other research (Siemer et al., 2014) because such combinations can induce higher inactivation (Singh and Shalini, 2016). The contribution of heat must be quantified when interpreting such combinations as mild processing. This is particularly important because microbial z-values, including those for spores, are typically lower than those of many quality factors such as vitamins (Awuah et al., 2007). High-temperature short-time concepts can therefore achieve a given lethality with less quality loss than lower-temperature, longer-time

treatments (Van Boekel, 2008). Comparisons between mild combinations and conventional thermal processing should be performed at matched decontamination targets while explicitly evaluating the food quality parameters that motivate the use of mild technologies.

6.4 Implications and future directions

Sustainability comparisons in the literature often use qualitative benchmarks such as 'can achieve 5-$\log_{10}$-reduction'. As shown in the case study above, this is not sufficient for fair comparison because the achieved reduction and its uncertainty can differ strongly across operating points. Although a 5-$\log_{10}$-reduction is a common benchmark, it is mandated as a criterion primarily for U.S. juices under HACCP (21 CFR Part 120), where the pertinent pathogen must be reduced by ≥5 logs immediately before or after packaging (U.S. Food and Drug Administration, 2004, 2001). Higher safety margins may target 6-, 7-, or 12-$\log_{10}$-reductions, or "commercial sterility" for shelf-stable, low-acid foods in hermetically sealed containers (U.S. Food and Drug Administration, 1979). Even after a 12-$\log_{10}$-reduction, a small residual risk of illness or spoilage remains (Zwietering et al., 2021). This highlights that fair comparison requires reporting the expected reduction and its uncertainty, and stating how targets are judged under uncertainty as a risk-management choice, for example, by requiring that the lower 95% prediction limit meets the target reduction as a conservative worst-case criterion, while recognizing that this may imply larger safety margins with consequences for cost, product quality, and environmental performance.

Fair comparison also extends beyond safety and climate indicators. Mild processing technologies are often adopted to preserve sensory quality, fresh-like attributes, or heat-sensitive nutrients, and these outcomes are product dependent. Conversely, thermal processing can remain preferable when heat recovery is efficient and when the product tolerates thermal load without unacceptable quality loss. Claims that a technology is "milder", "more natural", or "more sustainable" therefore need to be tied to a defined product category, target microorganism, and process, and should be evaluated across safety, quality, and sustainability factors.

Technology comparison should also be dynamic. Decarbonization will shift the relative sustainability of electricity-driven processes (Atuonwu et al., 2018). In the EU, renewables generated 47.3% of electricity in 2024, up 7.7% from 2023 (Eurostat, 2025). Mild processing technologies such as HPP, PEF, and NTP may therefore become more favorable over time from a greenhouse gas perspective, provided that their performance remains competitive at the required decontamination level. If fresh-like properties are not required and thermal effects are acceptable, electricity-driven thermal

processing through ohmic heating and continuous-flow microwave volumetric heating can also be competitive on emissions (Atuonwu et al., 2018). In any case, temperature profiles and the associated quality impacts must be evaluated explicitly, given the advantages of high-temperature short-time processing noted above. Non-thermal processing is therefore not inherently more eco-friendly. Its value lies in achieving safety with limited heat exposure and thereby preserving attributes that motivate its use.

The case study in this chapter illustrates how quantitative benchmarks can reduce the risk of misinterpretation. Energy use or greenhouse gas emissions per liter can only be interpreted as evidence of superior sustainability when the compared scenarios are efficacy matched, meaning that they deliver comparable $\log_{10}$-reductions for the same microorganism and product category, within uncertainty. When efficacy is not matched, rankings may reflect different safety margins rather than intrinsic process efficiency. A practical implication is that sustainability studies should quantify the achieved microbial reduction, or provide enough information to predict it, alongside energy and climate indicators.

Several research directions follow directly from the limitations identified across the chapters.

First, systematic external validation would give additional value to the global kinetic parameters established. For PEF and HPP, validation should use independent datasets generated under conditions representative of industrial operation, and should quantify both bias and prediction interval coverage across microorganisms, matrix categories, and operating modes. For NTP, the main bottleneck is still comparability. Wider uptake of the metadata schema and consistent reporting of core descriptors are needed before robust kinetic meta-regression can be pursued.

Second, a matched efficacy design should be integrated into life-cycle assessment. Future LCAs should evaluate multiple process settings per technology and relate energy demands and GHG emissions to the corresponding achieved or predicted microbial reductions under those settings. This would allow scenario analyses in which decontamination targets, throughput, and quality constraints are varied transparently, and it would reduce the risk of comparing technologies operated at different microbial inactivation levels.

Third, modeling should be extended to combined hurdles with explicit accounting for thermal contributions. Many practical applications rely on combinations such as mild heat plus PEF or pressure-assisted thermal processing. For these cases, time-temperature profiles, separation of thermal and non-thermal effects where possible,

and potential interactions between hurdles need to be quantified. Without this, combined processes risk being described as non-thermal while their performance is partly driven by heat.

Finally, the modeling and data infrastructure developed in this thesis can support a shift from technology-specific benchmarks to decision support. Quantitative meta-regression models can be combined with life-cycle assessment across multiple processing settings and with consumer-relevant quality criteria. This can support choices by process developers, manufacturers, quality managers, and regulators. Such a framework should integrate decontamination efficacy, processing time and throughput, uncertainty, energy use and GHG emissions, scale and mode of operation, and product constraints, including packaging and physicochemical properties. It should also account for consumer priorities such as sensory quality, nutritional value, cost, and label expectations. Extending this approach to additional technologies, including irradiation, continuous-flow microwave volumetric heating, and ohmic heating, and maintaining curated, interoperable datasets will be important for robust comparisons as evidence accumulates. In addition to database development and cross-technology comparison, a practical contribution of this thesis is the integration of established data science practices for food safety evidence synthesis, through reusable code, transparent wrangling pipelines, and semi-automated workflows that can be adapted to future studies and related research tasks.

Supplementary Material

No supplementary material is reproduced in the thesis body. The supplementary workbook supporting this chapter is available via 4TU.ResearchData (DOI: 10.4121/08c650c5-82be-468f-a878-b6ccbf93b7ca).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems [software]. Available at: https://www.tensorflow.org/.

Aganovic, K., Hertel, C., Vogel, R.F., Johne, R., Schlüter, O., Schwarzenbolz, U., Jäger, H., Holzhauser, T., Bergmair, J., Roth, A., Sevenich, R., Bandick, N., Kulling, S.E., Knorr, D., Engel, K.-H., Heinz, V., 2021. Aspects of high hydrostatic pressure food processing: Perspectives on technology and food safety. Compr. Rev. Food Sci. Food Saf. 20, 3225–3266. https://doi.org/10.1111/1541-4337.12763

Aganovic, K., Smetana, S., 2022. Environmental impact assessment of pulsed electric fields technology for food processing, in: Raso, J., Heinz, V., Alvarez, I., Toepfl, S. (Eds.), Pulsed Electric Fields Technology for the Food Industry: Fundamentals and Applications. Springer International Publishing, Cham, pp. 521–539. https://doi.org/10.1007/978-3-030-70586-2_19

Aganovic, K., Smetana, S., Grauwet, T., Toepfl, S., Mathys, A., Van Loey, A., Heinz, V., 2017. Pilot scale thermal and alternative pasteurization of tomato and watermelon juice: An energy comparison and life cycle assessment. J. Clean. Prod. 141, 514–525. https://doi.org/10.1016/j.jclepro.2016.09.015

Aickin, M., Gensler, H., 1996. Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. Am. J. Public Health 86, 726–728. https://doi.org/10.2105/ajph.86.5.726

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1–10.

Alemán, G., Farkas, D.F., Torres, J.A., Wilhelmsen, E., McIntyre, S., 1994. Ultra-high pressure pasteurization of fresh cut pineapple. J. Food Prot. 57, 931–934. https://doi.org/10.4315/0362-028X-57.10.931

Alessandria, V., Rantsiou, K., Cavallero, M.C., Cocolin, L.S., 2019. Effect of atmospheric pressure plasma on *Listeria monocytogenes* attached to abiotic surfaces. J. Food Prot. 82, 233–237. https://doi.org/10.4315/0362-028X.JFP-18-228

Álvarez, I., Pagán, R., Condón, S., Raso, J., 2003. The influence of process parameters for the inactivation of *Listeria monocytogenes* by pulsed electric fields. Int. J. Food Microbiol. 87, 87–95. https://doi.org/10.1016/S0168-1605(03)00056-4

Alves, L.L., Becker, M.M., Van Dijk, J., Gans, T., Go, D.B., Stapelmann, K., Tennyson, J., Turner, M.M., Kushner, M.J., 2023. Foundations of plasma standards. Plasma Sources Sci. Technol. 32, 023001. https://doi.org/10.1088/1361-6595/acb810

Amiali, M., Ngadi, M., Smith, J.P., Raghavan, V., 2006. Inactivation of *Escherichia coli* O157:H7 and *Salmonella* Enteritidis in liquid egg using continuous pulsed electric field system. Int. J. Food Eng. 1, 8. https://doi.org/10.2202/1556-3758.1058

Andreou, V., Dimopoulos, G., Katsaros, G., Taoukis, P., 2016. Comparison of the application of high pressure and pulsed electric fields technologies on the selective inactivation of endogenous enzymes in tomato products. Innov. Food Sci. Emerg. Technol. 38, 349–355. https://doi.org/10.1016/j.ifset.2016.07.026

Anuntagool, J., Srangsomjit, N., Thaweewong, P., Alvarez, G., 2023. A review on dielectric barrier discharge nonthermal plasma generation, factors affecting reactive species, and microbial inactivation. Food Control 153, 5–6. https://doi.org/10.1016/j.foodcont.2023.109913

Appleton, K.M., Newbury, A., Almiron-Roig, E., Yeomans, M.R., Brunstrom, J.M., de Graaf, K., Geurts, L., Kildegaard, H., Vinoy, S., 2021. Sensory and physical characteristics of foods that impact food intake without affecting acceptability: Systematic review and meta-analyses. Obes. Rev. 22, e13234. https://doi.org/10.1111/obr.13234

Asioli, D., Aschemann-Witzel, J., Caputo, V., Vecchio, R., Annunziata, A., Næs, T., Varela, P., 2017. Making sense of the ”clean label” trends: A review of consumer food choice behavior and discussion of industry implications. Food Res. Int. 99, 58–71. https://doi.org/10.1016/j.foodres.2017.07.022

Asl, P.J., Rajulapati, V., Gavahian, M., Kapusta, I., Putnik, P., Mousavi Khaneghah, A., Marszałek, K., 2022. Non-thermal plasma technique for preservation of fresh foods: A review. Food Control 134, 2–3. https://doi.org/10.1016/j.foodcont.2021.108560

ASReview LAB developers, 2024. ASReview LAB - A tool for AI-assisted systematic reviews. https://doi.org/10.5281/zenodo.11185216

Atuonwu, J.C., Leadley, C., Bosman, A., Tassou, S.A., Lopez-Quiroga, E., Fryer, P.J., 2018. Comparative assessment of innovative and conventional food preservation technologies: Process energy performance and greenhouse gas emissions. Innov. Food Sci. Emerg. Technol. 50, 174–187. https://doi.org/10.1016/j.ifset.2018.09.008

Augustin, M.A., Riley, M., Stockmann, R., Bennett, L., Kahl, A., Lockett, T., Osmond, M., Sanguansri, P., Stonehouse, W., Zajac, I., Cobiac, L., 2016. Role of food processing in food and nutrition security. Trends Food Sci. Technol. 56, 115–125. https://doi.org/10.1016/j.tifs.2016.08.005

Avilés-Gaxiola, S., Chuck-Hernández, C., Serna Saldívar, S.O., 2018. Inactivation methods of trypsin inhibitor in legumes: A review. J. Food Sci. 83, 17–29. https://doi.org/10.1111/1750-3841.13985

Awuah, G.B., Ramaswamy, H.S., Economides, A., 2007. Thermal processing and qual-

ity: Principles and overview. Chem. Eng. Process. 46, 584–602. https://doi.org/10.1016/j.cep.2006.08.004

Baier, M., Foerster, J., Schnabel, U., Knorr, D., Ehlbeck, J., Herppich, W.B., Schlüter, O., 2013. Direct non-thermal plasma treatment for the sanitation of fresh corn salad leaves: Evaluation of physical and physiological effects and antimicrobial efficacy. Postharvest Biol. Technol. 84, 81–87. https://doi.org/10.1016/j.postharvbio.2013.03.022

Baker, W.L., Michael White, C., Cappelleri, J.C., Kluger, J., Coleman, C.I., From the Health Outcomes, Policy, and Economics (HOPE) Collaborative Group, 2009. Understanding heterogeneity in meta-analysis: The role of meta-regression. Int. J. Clin. Pract. 63, 1426–1434. https://doi.org/10.1111/j.1742-1241.2009.02168.x

Balasubramaniam, V.M.(Bala)., Martínez-Monteagudo, S.I., Gupta, R., 2015. Principles and application of high pressure–based technologies in the food industry. Annu. Rev. Food Sci. Technol. 6, 435–462. https://doi.org/10.1146/annurev-food-022814-015539

Batini, C., Rula, A., Scannapieco, M., Viscusi, G., 2015. From data quality to big data quality. J. Database Manag. 26, 60–82. https://doi.org/10.4018/JDM.2015010103

Battaggia, D., Yao, Y., Nierop Groot, M.N., Abee, T., Den Besten, H.M.W., 2024. Impact of species and strain variability on non-thermal plasma decontamination efficacy. Innov. Food Sci. Emerg. Technol. 95, 103674. https://doi.org/10.1016/j.ifset.2024.103674

Bauer, A., Ni, Y., Bauer, S., Paulsen, P., Modic, M., Walsh, J.L., Smulders, F.J.M., 2017. The effects of atmospheric pressure cold plasma treatment on microbiological, physical-chemical and sensory characteristics of vacuum packaged beef loin. Meat Sci. 128, 77–87. https://doi.org/10.1016/j.meatsci.2017.02.003

Bearth, A., Cousin, M.-E., Siegrist, M., 2014. The consumer's perception of artificial food additives: Influences on acceptance, risk and benefit perceptions. Food Qual. Prefer. 38, 14–23. https://doi.org/10.1016/j.foodqual.2014.05.008

Beitia, E., Gkogka, E., Chanos, P., Hertel, C., Heinz, V., Valdramidis, V., Aganovic, K., 2023. Microbial decontamination assisted by ultrasound-based processing technologies in food and model systems: A review. Compr. Rev. Food Sci. Food Saf. 22, 2802–2849. https://doi.org/10.1111/1541-4337.13163

Berlin, J.A., Colditz, G.A., 1999. The role of meta-analysis in the regulatory process for foods, drugs, and devices. JAMA 281, 830–834. https://doi.org/10.1001/jama.281.9.830

Bernhardt, T., Semmler, M.L., Schäfer, M., Bekeschus, S., Emmert, S., Boeckmann, L., 2019. Plasma medicine: Applications of cold atmospheric pressure plasma in dermatology. Oxid. Med. Cell. Longev. 2019, 3873928. https://doi.org/10.1155/2019/3873928

Beveridge, J.R., MacGregor, S.J., Anderson, J.G., Fouracre, R.A., 2005. The influence of

pulse duration on the inactivation of bacteria using monopolar and bipolar profile pulsed electric fields. IEEE Trans. Plasma Sci. 33, 1287–1293. https://doi.org/10.1109/TPS.2005.852345

Bickel, P.J., Li, B., Tsybakov, A.B., Van de Geer, S.A., Yu, B., Valdés, T., Rivero, C., Fan, J., Van der Vaart, A., 2006. Regularization in statistics. TEST 15, 271–344. https://doi.org/10.1007/BF02607055

Bigelow, W.D., 1921. The logarithmic nature of thermal death time curves. J. Infect. Dis. 29, 528–536. https://doi.org/10.1093/infdis/29.5.528

Birania, S., Attkan, A.K., Kumar, S., Kumar, N., Singh, V.K., 2022. Cold plasma in food processing and preservation: A review. J. Food Process Eng. 45, e14110. https://doi.org/10.1111/jfpe.14110

Bishop, C.M., 2006. Pattern recognition and machine learning, Information Science and Statistics. Springer, New York.

Bonzanini, A.D., Shao, K., Stancampiano, A., Graves, D.B., Mesbah, A., 2022. Perspectives on machine learning-assisted plasma medicine: Toward automated plasma treatment. IEEE Trans. Radiat. Plasma Med. Sci. 6, 16–32. https://doi.org/10.1109/TRPMS.2021.3055727

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92. Association for Computing Machinery, New York, NY, USA, pp. 144–152. https://doi.org/10.1145/130385.130401

Bover-Cid, S., Belletti, N., Aymerich, T., Garriga, M., 2015. Modeling the protective effect of a_w and fat content on the high pressure resistance of *Listeria monocytogenes* in dry-cured ham. Food Res. Int. 75, 194–199. https://doi.org/10.1016/j.foodres.2015.05.052

Box, G.E.P., 1976. Science and statistics. J. Am. Stat. Assoc. 71, 791–799. https://doi.org/10.1080/01621459.1976.10480949

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26, 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd ed. Springer, New York.

Cacace, F., Bottani, E., Rizzi, A., Vignali, G., 2020. Evaluation of the economic and environmental sustainability of high pressure processing of foods. Innov. Food Sci. Emerg. Technol. 60, 102281. https://doi.org/10.1016/j.ifset.2019.102281

Callaghan, M.W., Müller-Hansen, F., 2020. Statistical stopping criteria for automated screening in systematic reviews. Syst. Rev. 9, 273. https://doi.org/10.1186/s13643-020-01521-4

Cavalcanti, R.N., Balthazar, C.F., Margalho, L.P., Freitas, M.Q., Sant'Ana, A.S., Cruz,

A.G., 2023. Pulsed electric field-based technology for microbial inactivation in milk and dairy products. Curr. Opin. Food Sci. 54, 101087. https://doi.org/10.1016/j.cofs.2023.101087

Chaerony Siffa, I., Schäfer, J., Becker, M.M., 2022. Adamant: A JSON schema-based metadata editor for research data management workflows. F1000Res. 11, 475. https://doi.org/10.12688/f1000research.110875.2

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 6–7. https://doi.org/10.1145/1961189.1961199

Charrondiere, U.R., Haytowitz, D.B., Stadlmayr, B., 2012. FAO/INFOODS density database. Version 2.0. Food and Agriculture Organization of the United Nations, Rome.

Chatfield, C., 1986. Exploratory data analysis. Eur. J. Oper. Res. 23, 5–13. https://doi.org/10.1016/0377-2217(86)90209-2

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cichy, C., Rass, S., 2019. An overview of data quality frameworks. IEEE Access 7, 24634–24648. https://doi.org/10.1109/ACCESS.2019.2899751

Codex Alimentarius Commission, 2019. Procedural manual: Twenty-seventh edition. Food; Agriculture Organization of the United Nations (FAO); World Health Organization (WHO), Rome.

Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F.N., Leip, A., 2021. Food systems are responsible for a third of global anthropogenic GHG emissions. Nat. Food 2, 198–209. https://doi.org/10.1038/s43016-021-00225-9

Crippa, M., Solazzo, E., Guizzardi, D., Van Dingenen, R., Leip, A., 2022. Air pollutant emissions from global food systems are responsible for environmental impacts, crop losses and mortality. Nat. Food 3, 942–956. https://doi.org/10.1038/s43016-022-00615-7

Cristianini, N., Ricci, E., 2016. Support Vector Machines, in: Kao, M.-Y. (Ed.), Encyclopedia of Algorithms. Springer, New York, NY, pp. 2170–2174. https://doi.org/10.1007/978-1-4939-2864-4_415

Cullen, P.J., Lalor, J., Scally, L., Boehm, D., Milosavljević, V., Bourke, P., Keener, K., 2018. Translation of plasma technology from the lab to the food industry. Plasma Process. Polym. 15, 1700085. https://doi.org/10.1002/ppap.201700085

Delso, C., Martínez, J.M., Cebrián, G., Condón, S., Raso, J., Álvarez, I., 2022. Microbial inactivation by pulsed electric fields, in: Raso, J., Heinz, V., Alvarez, I., Toepfl, S. (Eds.), Pulsed Electric Fields Technology for the Food Industry: Fundamentals and Applications. Springer International Publishing, Cham, pp. 169–207. https://doi.org/10.1007/978-3-030-70586-2_5

Den Besten, H.M.W., Valdramidis, V.P., 2025. Editorial: Sustainable food technologies. Innov. Food Sci. Emerg. Technol. 101, 103974. https://doi.org/10.1016/j.ifset.2025.103974

Den Besten, H.M.W., Zwietering, M.H., 2012. Meta-analysis for quantitative microbiological risk assessments and benchmarking data. Trends Food Sci. Technol. 25, 34–39. https://doi.org/10.1016/j.tifs.2011.12.004

Deutsche Gesellschaft für Mund-, Kiefer- und Gesichtschirurgie (DGMKG), 2022. Rationaler therapeutischer Einsatz von kaltem physikalischem Plasma.

Dhaliwal, H.K., Sonkar, S., V, P., Puente, L., Roopesh, M.S., 2025. Process technologies for disinfection of food-contact surfaces in the dry food industry: A review. Microorganisms 13, 648. https://doi.org/10.3390/microorganisms13030648

Dharini, M., Jaspin, S., Mahendran, R., 2023. Cold plasma reactive species: Generation, properties, and interaction with food biomolecules. Food Chem. 405, 134746. https://doi.org/10.1016/j.foodchem.2022.134746

Diao, M.M., André, S., Membré, J.-M., 2014. Meta-analysis of D-values of proteolytic *Clostridium botulinum* and its surrogate strain *Clostridium sporogenes* PA 3679. Int. J. Food Microbiol. 174, 23–30. https://doi.org/10.1016/j.ijfoodmicro.2013.12.029

Dietterich, T.G., 2002. Ensemble learning, in: Arbib, M.A. (Ed.), The Handbook of Brain Theory and Neural Networks. MIT Press, pp. 405–408.

Dogan, O.B., Aditya, A., Ortuzar, J., Clarke, J., Wang, B., 2022. A systematic review and meta-analysis of the efficacy of processing stages and interventions for controlling *Campylobacter* contamination during broiler chicken processing. Compr. Rev. Food Sci. Food Saf. 21, 227–271. https://doi.org/10.1111/1541-4337.12860

Dolgopolova, I., Teuber, R., 2018. Consumers' willingness to pay for health benefits in food products: A meta-analysis. Appl. Econ. Perspect. Policy 40, 333–352. https://doi.org/10.1093/aepp/ppx036

EFSA, 2015. The food classification and description system FoodEx 2 (revision 2). EFSA Support. Publ. 12, 804E. https://doi.org/10.2903/sp.efsa.2015.EN-804

EFSA Panel on Biological Hazards (BIOHAZ), Koutsoumanis, K., Alvarez-Ordóñez, A., Bolton, D., Bover-Cid, S., Chemaly, M., Davies, R., De Cesare, A., Herman, L., Hilbert, F., Lindqvist, R., Nauta, M., Peixe, L., Ru, G., Simmons, M., Skandamis, P., Suffredini, E., Castle, L., Crotta, M., Grob, K., Milana, M.R., Petersen, A., Roig Sagués, A.X., Vinagre Silva, F., Barthélémy, E., Christodoulidou, A., Messens, W., Allende, A., 2022. The efficacy and safety of high-pressure processing of food. EFSA J. 20, e07128. https://doi.org/10.2903/j.efsa.2022.7128

Eisenhauer, J.G., 2003. Regression through the origin. Teach. Stat. 25, 76–80. https://doi.org/10.1111/1467-9639.00136

Emmoth, E., Rovira, J., Rajkovic, A., Corcuera, E., Wilches Pérez, D., Dergel, I., Ottoson, J.R., Widén, F., 2017. Inactivation of viruses and bacteriophages as models for swine Hepatitis E virus in food matrices. Food Environ. Virol. 9, 20–34. https:

//doi.org/10.1007/s12560-016-9268-y
Espinosa, M.F., Sancho, A.N., Mendoza, L.M., Mota, C.R., Verbyla, M.E., 2020. Systematic review and meta-analysis of time-temperature pathogen inactivation. Int. J. Hyg. Environ. Health 230, 113595. https://doi.org/10.1016/j.ijheh.2020.113595
European Commission, 2016. Commission Notice on the implementation of food safety management systems covering prerequisite programmes (PRPs) and procedures based on the HACCP principles, including facilitation/flexibility of implementation in certain food businesses (2016/C 278/01).
European Parliament and Council of the European Union, 2004. Regulation (EC) No 852/2004 of the European Parliament and of the Council of 29 April 2004 on the hygiene of foodstuffs.
European Union, 2002. Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety.
Eurostat, 2025. Renewable energy supply grew by 3.4% in 2024.
Evelyn, Silva, F.V.M., 2019. Heat assisted HPP for the inactivation of bacteria, moulds and yeasts spores in foods: Log reductions and mathematical models. Trends Food Sci. Technol. 88, 143–156. https://doi.org/10.1016/j.tifs.2019.03.016
FAO, IFAD, UNICEF, WFP, WHO, 2023. The State of Food Security and Nutrition in the World 2023: Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum. FAO, Rome. https://doi.org/10.4060/cc3017en
Farkas, D.F., 2016. A short history of research and development efforts leading to the commercialization of high-pressure processing of food, in: Balasubramaniam, V.M., Barbosa-Cánovas, G.V., Lelieveld, H.L.M. (Eds.), High Pressure Processing of Food: Principles, Technology and Applications. Springer, New York, NY, pp. 19–36. https://doi.org/10.1007/978-1-4939-3234-4_2
Ferdinands, G., Schram, R., Bruin, J. de, Bagheri, A., Oberski, D.L., Tummers, L., Teijema, J.J., Van de Schoot, R., 2023. Performance of active learning models for screening prioritization in systematic reviews: A simulation study into the average time to discover relevant records. Syst. Rev. 12, 1. https://doi.org/10.1186/s13643-023-02257-7
Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., Barabási, A.-L., 2018. Science of science. Science 359, eaao0185. https://doi.org/10.1126/science.aao0185
Fox, M.B., Esveld, D.C., Mastwijk, H., Boom, R.M., 2008. Inactivation of *L. plantarum* in a PEF microreactor. Innov. Food Sci. Emerg. Technol. 9, 101–108. https://doi.org/10.1016/j.ifset.2007.06.007
Franke, S., Paulet, L., Schäfer, J., O'Connell, D., Becker, M.M., 2020. Plasma-MDS, a

metadata schema for plasma science with examples from plasma technology. Sci. Data 7, 439. https://doi.org/10.1038/s41597-020-00771-0

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29, 1189–1232.

Gabriel, A.A., Aba, R.P.M., Tayamora, D.J.L., Colambo, J.C.R., Siringan, M.A.T., Rosario, L.M.D., Tumlos, R.B., Ramos, H.J., 2016. Reference organism selection for microwave atmospheric pressure plasma jet treatment of young coconut liquid endosperm. Food Control 69, 74–82. https://doi.org/10.1016/j.foodcont.2016.04.034

García, D., Gómez, N., Mañas, P., Raso, J., Pagán, R., 2007. Pulsed electric fields cause bacterial envelopes permeabilization depending on the treatment intensity, the treatment medium pH and the microorganism investigated. Int. J. Food Microbiol. 113, 219–227. https://doi.org/10.1016/j.ijfoodmicro.2006.07.007

Garre, A., Fernández, P., Grau-Noguer, E., Guillén, S., Portaña, S., Possas, A., Vila, M., 2025. Chapter One - Predictive microbiology through the last century. From paper to Excel and towards AI, in: Sant'ana, A.S. (Ed.), Advances in Food and Nutrition Research. Academic Press, pp. 1–63. https://doi.org/10.1016/bs.afnr.2024.09.012

Garre, A., Fernández, P.S., Lindqvist, R., Egea, J.A., 2017. Bioinactivation: Software for modelling dynamic microbial inactivation. Food Res. Int. 93, 66–74. https://doi.org/10.1016/j.foodres.2017.01.012

Garre, A., Zwietering, M.H., Den Besten, H.M.W., 2023. The importance of what we cannot observe: Experimental limitations as a source of bias for meta-regression models in predictive microbiology. Int. J. Food Microbiol. 387, 110045. https://doi.org/10.1016/j.ijfoodmicro.2022.110045

Gavahian, M., Chu, Y.-H., Mousavi Khaneghah, A., Barba, F.J., Misra, N.N., 2018. A critical analysis of the cold plasma induced lipid oxidation in foods. Trends Food Sci. Technol. 77, 32–41. https://doi.org/10.1016/j.tifs.2018.04.009

Ge, L., Jia, F., Yu, J., Liu, H., Bai, R., Li, S., Wang, H., Li, Z., 2024. Hybrid pasteurization strategy for energy conservation and quality preservation in cloudy apple juice processing. J. Food Eng. 382, 112217. https://doi.org/10.1016/j.jfoodeng.2024.112217

Gelman, A., 2008. Scaling regression inputs by dividing by two standard deviations. Stat. Med. 27, 2865–2873. https://doi.org/10.1002/sim.3107

Georgescu, N., Apostol, L., Gherendi, F., 2017. Inactivation of *Salmonella enterica* serovar Typhimurium on egg surface, by direct and indirect treatments with cold atmospheric plasma. Food Control 76, 52–61. https://doi.org/10.1016/j.foodcont.2017.01.005

Ghafoor, K., Gavahian, M., Marszałek, K., Barba, F.J., Xia, Q., Denoya, G.I., 2020. Chapter 1 - an overview of the potential applications based on HPP mechanism, in: Barba, F.J., Tonello-Samson, C., Puértolas, E., Lavilla, M. (Eds.), Present and Future of High Pressure Processing. Elsevier, pp. 3–11. https:

//doi.org/10.1016/B978-0-12-816405-1.00001-7
Giannoulis, N., Karatzas, K.A.G., 2024. The combined effect of chitosan and high hydrostatic pressure on *Listeria monocytogenes* and *Escherichia coli*. Innov. Food Sci. Emerg. Technol. 94, 103693. https://doi.org/10.1016/j.ifset.2024.103693
Glass, G.V., 1976. Primary, secondary, and meta-analysis of research. Educ. Res. 5, 3–8. https://doi.org/10.3102/0013189X005010003
Golda, J., Held, J., Redeker, B., Konkowski, M., Beijer, P., Sobota, A., Kroesen, G., Braithwaite, N.S.J., Reuter, S., Turner, M.M., Gans, T., O'Connell, D., Gathen, V.S.D., 2016. Concepts and characteristics of the "COST Reference Microplasma Jet." J. Phys. D Appl. Phys. 49, 6–7. https://doi.org/10.1088/0022-3727/49/8/084003
Golda, J., Kogelheide, F., Awakowicz, P., Gathen, V.S.D., 2019. Dissipated electrical power and electron density in an RF atmospheric pressure helium plasma jet. Plasma Sources Sci. Technol. 28, 3–4. https://doi.org/10.1088/1361-6595/ab393d
Gonzales-Barron, U., Butler, F., 2011. The use of meta-analytical tools in risk assessment for food safety. Food Microbiol. 28, 823–827. https://doi.org/10.1016/j.fm.2010.04.007
González-Arenzana, L., Portu, J., López, R., López, N., Santamaría, P., Garde-Cerdán, T., López-Alfaro, I., 2015. Inactivation of wine-associated microbiota by continuous pulsed electric field treatments. Innov. Food Sci. Emerg. Technol. 29, 187–192. https://doi.org/10.1016/j.ifset.2015.03.009
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. The MIT Press, Cambridge, Massachusetts.
Govaert, M., Smet, C., Vergauwen, L., Ećimović, B., Walsh, J.L., Baka, M., Van Impe, J., 2019. Influence of plasma characteristics on the efficacy of cold atmospheric plasma (CAP) for inactivation of *Listeria monocytogenes* and *Salmonella* Typhimurium biofilms. Innov. Food Sci. Emerg. Technol. 52, 376–386. https://doi.org/10.1016/j.ifset.2019.01.013
Govaris, A., Pexara, A., 2021. Inactivation of foodborne viruses by high-pressure processing (HPP). Foods 10, 215. https://doi.org/10.3390/foods10020215
Grigor, J.M., Brennan, C.S., Hutchings, S.C., Rowlands, D.S., 2016. The sensory acceptance of fibre-enriched cereal foods: A meta-analysis. Int. J. Food Sci. Technol. 51, 3–13. https://doi.org/10.1111/ijfs.13005
Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on tabular data?
Guillou, S., Membré, J.-M., 2019. Inactivation of *Listeria monocytogenes*, *Staphylococcus aureus*, and *Salmonella enterica* under high hydrostatic pressure: A quantitative analysis of existing literature data. J. Food Prot. 82, 1802–1814. https://doi.org/10.4315/0362-028X.JFP-19-132
Gülenç, B., Özdemir, E., Başaran, P., Kartal, S., Akan, T., 2024. Impact on microbiology and product quality of plasma-activated water treatment on fresh green

leafy vegetables. Postharvest Biol. Technol. 217, 113087. https://doi.org/10.1016/j.postharvbio.2024.113087

Halder, A., Black, D.G., Davidson, P.M., Datta, A., 2010. Development of associations and kinetic models for microbiological data to be used in comprehensive food safety prediction software. J. Food Sci. 75, R107–R120. https://doi.org/10.1111/j.1750-3841.2010.01687.x

Han, J.-Y., Song, W.-J., Eom, S., Kim, S.B., Kang, D.-H., 2020. Antimicrobial efficacy of cold plasma treatment against food-borne pathogens on various foods. J. Phys. D Appl. Phys. 53, 204003. https://doi.org/10.1088/1361-6463/ab761f

Harris, C.R., Millman, K.J., Van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with NumPy. Nature 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Harrison, H., Griffin, S.J., Kuhn, I., Usher-Smith, J.A., 2020. Software tools to support title and abstract screening for systematic reviews in healthcare: An evaluation. BMC Med. Res. Methodol. 20, 7. https://doi.org/10.1186/s12874-020-0897-3

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: Data mining, inference, and prediction, 2nd ed, Springer Series in Statistics. Springer, New York.

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., 1998. Support vector machines. IEEE Intell. Syst. 13, 18–28. https://doi.org/10.1109/5254.708428

Heinz, V., Alvarez, I., Angersbach, A., Knorr, D., 2001. Preservation of liquid foods by high intensity pulsed electric fields—basic concepts for process design. Trends Food Sci. Technol. 12, 103–111. https://doi.org/10.1016/S0924-2244(01)00064-4

Hertwig, C., Reineke, K., Rauh, C., Schlüter, O., 2017. Factors involved in *Bacillus* spore's resistance to cold atmospheric pressure plasma. Innov. Food Sci. Emerg. Technol. 43, 173–181. https://doi.org/10.1016/j.ifset.2017.07.031

Hill, K., Ostermeier, R., Töpfl, S., Heinz, V., 2022. Pulsed electric fields in the potato industry, in: Raso, J., Heinz, V., Alvarez, I., Toepfl, S. (Eds.), Pulsed Electric Fields Technology for the Food Industry: Fundamentals and Applications. Springer International Publishing, Cham, pp. 325–335. https://doi.org/10.1007/978-3-030-70586-2_9

Hiperbaric, 2025. Hiperbaric expects to grow 12% in 2025 and achieve record sales.

Hirneisen, K.A., Kniel, K.E., 2013. Inactivation of internalized and surface contaminated enteric viruses in green onions. Int. J. Food Microbiol. 166, 201–206. https://doi.org/10.1016/j.ijfoodmicro.2013.07.013

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67. https://doi.org/10.2307/1267351

Huang, H.-W., Wu, S.-J., Lu, J.-K., Shyu, Y.-T., Wang, C.-Y., 2017. Current status and future trends of high-pressure processing in food industry. Food Control 72, 1–8. https://doi.org/10.1016/j.foodcont.2016.07.019

Huang, Y.-M., Chen, C.-K., Hsu, C.-L., 2020. Non-thermal atmospheric gas plasma for decontamination of sliced cheese and changes in quality. Food Sci. Technol. Int. 26, 715–726. https://doi.org/10.1177/1082013220925931

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.1109/MCSE.2007.55

Insfran-Rivarola, A., Tlapa, D., Limon-Romero, J., Baez-Lopez, Y., Miranda-Ackerman, M., Arredondo-Soto, K., Ontiveros, S., 2020. A systematic review and meta-analysis of the effects of food safety and hygiene training on food handlers. Foods 9, 1169. https://doi.org/10.3390/foods9091169

Jaeger, H., Meneses, N., Moritz, J., Knorr, D., 2010. Model for the differentiation of temperature and electric field effects during thermal assisted PEF processing. J. Food Eng. 100, 109–118. https://doi.org/10.1016/j.jfoodeng.2010.03.034

Juneja, V.K., 2001. Thermal inactivation of microorganisms, in: Juneja, V.K., Sofos, J.N. (Eds.), Control of Foodborne Microorganisms. CRC Press, pp. 13–54.

Katsigiannis, A.S., Bayliss, D.L., Walsh, J.L., 2021. Cold plasma decontamination of stainless steel food processing surfaces assessed using an industrial disinfection protocol. Food Control 121, 107543. https://doi.org/10.1016/j.foodcont.2020.107543

Keener, K.M., Jensen, J.L., Valdramidis, V.P., Byrne, E., Connolly, J., Mosnier, J.P., Cullen, P.J., 2012. Decontamination of *Bacillus subtilis* spores in a sealed package using a non-thermal plasma system, in: Machala, Z., Hensel, K., Akishev, Y. (Eds.), Plasma for Bio-Decontamination, Medicine and Food Security. Springer Netherlands, Dordrecht, pp. 445–455. https://doi.org/10.1007/978-94-007-2852-3_34

Keller, S., Korkmaz, G., Orr, M., Schroeder, A., Shipp, S., 2017. The evolution of data quality: Understanding the transdisciplinary origins of data quality concepts and approaches. Annu. Rev. Stat. Appl. 4, 85–108. https://doi.org/10.1146/annurev-statistics-060116-054114

Kempkes, M.A., 2017. Industrial pulsed electric field systems, in: Handbook of Electroporation. Springer, Cham, pp. 1–21. https://doi.org/10.1007/978-3-319-26779-1_211-1

Kim, J.E., Oh, Y.J., Won, M.Y., Lee, K.-S., Min, S.C., 2017. Microbial decontamination of onion powder using microwave-powered cold plasma treatments. Food Microbiol. 62, 112–123. https://doi.org/10.1016/j.fm.2016.10.006

Kingsley, D.H., Guan, D., Hoover, D.G., 2005. Pressure inactivation of Hepatitis A virus in strawberry puree and sliced green onions. J. Food Prot. 68, 1748–1751. https://doi.org/10.4315/0362-028X-68.8.1748

Knorr, D., Watzke, H., 2019. Food processing at a crossroad. Front. Nutr. 6, 85. https://doi.org/10.3389/fnut.2019.00085

Kotnik, T., Rems, L., Tarek, M., Miklavčič, D., 2019. Membrane electroporation and electropermeabilization: Mechanisms and models. Annu. Rev. Biophys. 48, 63–91. https://doi.org/10.1146/annurev-biophys-052118-115451

Krogh, A., 2008. What are artificial neural networks? Nat. Biotechnol. 26, 195–197. https://doi.org/10.1038/nbt1386

Landi, G., Benedetti, M., Sforzini, M., Eslami, E., Pataro, G., 2025. Comparative analysis of cost, energy efficiency, and environmental impact of pulsed electric fields and conventional thermal treatment with integrated heat recovery for fruit juice pasteurization. Foods 14, 2239. https://doi.org/10.3390/foods14132239

Laroque, D.A., Seó, S.T., Valencia, G.A., Laurindo, J.B., Carciofi, B.A.M., 2022. Cold plasma in food processing: Design, mechanisms, and application. J. Food Eng. 312, 110748. https://doi.org/10.1016/j.jfoodeng.2021.110748

Laurita, R., Gozzi, G., Tappi, S., Capelli, F., Bisag, A., Laghi, G., Gherardi, M., Cellini, B., Abouelenein, D., Vittori, S., Colombo, V., Rocculi, P., Dalla Rosa, M., Vannini, L., 2021. Effect of plasma activated water (PAW) on rocket leaves decontamination and nutritional value. Innov. Food Sci. Emerg. Technol. 73, 102805. https://doi.org/10.1016/j.ifset.2021.102805

Leibniz Institute for Plasma Science and Technology, 2019. INPTDAT - the data platform for plasma technology. Available at: https://doi.org/10.5281/zenodo.3500283.

Lenz, C.A., Vogel, R.F., 2014. Effect of sporulation medium and its divalent cation content on the heat and high pressure resistance of *Clostridium botulinum* type E spores. Food Microbiol. 44, 156–167. https://doi.org/10.1016/j.fm.2014.05.010

Liao, X., Liu, D., Xiang, Q., Ahn, J., Chen, S., Ye, X., Ding, T., 2017. Inactivation mechanisms of non-thermal plasma on microbes: A review. Food Control 75, 83–91. https://doi.org/10.1016/j.foodcont.2016.12.021

Liu, C., Cui, N., Brown, N.M.D., Meenan, B.J., 2004. Effects of DBD plasma operating parameters on the polymer surface modification. Surf. Coat. Technol. 185, 311–320. https://doi.org/10.1016/j.surfcoat.2004.01.024

Liu, Z.-W., Zeng, X.-A., Ngadi, M., Han, Z., 2017. Effect of cell membrane fatty acid composition of *Escherichia coli* on the resistance to pulsed electric field (PEF) treatment. LWT 76, 18–25. https://doi.org/10.1016/j.lwt.2016.10.019

Lones, M.A., 2021. How to avoid machine learning pitfalls: A guide for academic researchers. arXiv 2108.02497, 6–7. https://doi.org/10.48550/arXiv.2108.02497

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., p. 9.

Lytras, F., Psakis, G., Gatt, R., Cebrián, G., Raso, J., Valdramidis, V., 2024a. Exploring the efficacy of pulsed electric fields (PEF) in microbial inactivation during food processing: A deep dive into the microbial cellular and molecular mechanisms.

Innov. Food Sci. Emerg. Technol. 95, 103732. https://doi.org/10.1016/j.ifset.2024.103732

Lytras, F., Psakis, G., Gatt, R., Hummerjohann, J., Raso, J., Valdramidis, V., 2024b. Evaluation of strain variability of food microorganisms in response to decontamination by pulsed electric fields and thermal treatments. Innov. Food Sci. Emerg. Technol. 95, 103731. https://doi.org/10.1016/j.ifset.2024.103731

Mafart, P., Couvert, O., Gaillard, S., Leguerinel, I., 2002. On calculating sterility in thermal preservation methods: Application of the Weibull frequency distribution model. Int. J. Food Microbiol. 72, 107–113. https://doi.org/10.1016/S0168-1605(01)00624-9

Majumdar, A., Singh, R.K., Palm, G.J., Hippler, R., 2009. Dielectric barrier discharge plasma treatment on *E. coli*: Influence of CH_4/N_2, O_2, N_2/O_2, N_2, and Ar gases. J. Appl. Phys. 106, 084701. https://doi.org/10.1063/1.3236570

Mañas, P., Mackey, B.M., 2004. Morphological and physiological changes induced by high hydrostatic pressure in exponential- and stationary-phase cells of *Escherichia coli*: Relationship with cell death. Appl. Environ. Microbiol. 70, 1545–1554. https://doi.org/10.1128/AEM.70.3.1545-1554.2004

Mann, M.S., Tiede, R., Gavenis, K., Daeschlein, G., Bussiahn, R., Weltmann, K.-D., Emmert, S., Woedtke, T. von, Ahmed, R., 2016. Introduction to DIN-specification 91315 based on the characterization of the plasma jet kINPen® MED. Clin. Plasma Med. 4, 35–45. https://doi.org/10.1016/j.cpme.2016.06.001

Margosch, D., Ehrmann, M.A., Gänzle, M.G., Vogel, R.F., 2004. Comparison of pressure and heat resistance of *Clostridium botulinum* and other endospores in mashed carrots. J. Food Prot. 67, 2530–2538. https://doi.org/10.4315/0362-028X-67.11.2530

Marshall, I.J., Wallace, B.C., 2019. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. Syst. Rev. 8, 163. https://doi.org/10.1186/s13643-019-1074-9

Martínez-Gonzáles, N.E., Castillo, A., 2016. Safety of fresh-squeezed juices, in: Kotzekidou, P. (Ed.), Food Hygiene and Toxicology in Ready-to-Eat Foods. Academic Press, pp. 183–208. https://doi.org/10.1016/B978-0-12-801916-0.00011-X

Martins, I.B.A., Rosenthal, A., Ares, G., Deliza, R., 2020. How do processing technology and formulation influence consumers' choice of fruit juice? Int. J. Food Sci. Technol. 55, 2660–2668. https://doi.org/10.1111/ijfs.14519

McKinney, W., 2010. Data structures for statistical computing in Python, in: Van der Walt, S., Millman, J. (Eds.), Proceedings of the 9th Python in Science Conference. pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Mendes-Oliveira, G., Jin, T.Z., Campanella, O.H., 2020. Modeling the inactivation of *Escherichia coli* O157:H7 and *Salmonella* Typhimurium in juices by pulsed electric fields: The role of the energy density. J. Food Eng. 282, 110001. https://doi.org/10.1016/j.jfoodeng.2020.110001

Misra, N.N., Schlüter, O., Cullen, P.J., 2016. Cold plasma in food and agriculture: Fundamentals and applications. Academic Press.

Misra, N.N., Tiwari, B.K., Raghavarao, K.S.M.S., Cullen, P.J., 2011. Nonthermal plasma inactivation of food-borne pathogens. Food Eng. Rev. 3, 159–170. https://doi.org/10.1007/s12393-011-9041-9

Misra, N.N., Yepez, X., Xu, L., Keener, K., 2019. In-package cold plasma technologies. J. Food Eng. 244, 21–31. https://doi.org/10.1016/j.jfoodeng.2018.09.019

Miwa, M., Thomas, J., O'Mara-Eves, A., Ananiadou, S., 2014. Reducing systematic review workload through certainty-based screening. J. Biomed. Inform. 51, 246–252. https://doi.org/10.1016/j.jbi.2014.06.005

Monfort, S., Gayán, E., Raso, J., Condón, S., Álvarez, I., 2010. Evaluation of pulsed electric fields technology for liquid whole egg pasteurization. Food Microbiol. 27, 845–852. https://doi.org/10.1016/j.fm.2010.05.011

Montgomery, D.C., Peck, E.A., Vining, G.G., 2021. Introduction to linear regression analysis, 6th ed. John Wiley & Sons, Inc., Hoboken, NJ.

Moritz, M., Wiacek, C., Koethe, M., Braun, P.G., 2017. Atmospheric pressure plasma jet treatment of *Salmonella* Enteritidis inoculated eggshells. Int. J. Food Microbiol. 245, 22–28. https://doi.org/10.1016/j.ijfoodmicro.2017.01.004

Mošovská, S., Medvecká, V., Halászová, N., Ďurina, P., Valík, Ľ., Mikulajová, A., Zahoranová, A., 2018. Cold atmospheric pressure ambient air plasma inhibition of pathogenic bacteria on the surface of black pepper. Food Res. Int. 106, 862–869. https://doi.org/10.1016/j.foodres.2018.01.066

Mosqueda-Melgar, J., Raybaudi-Massilia, R.M., Martín-Belloso, O., 2008. Non-thermal pasteurization of fruit juices by combining high-intensity pulsed electric fields with natural antimicrobials. Innov. Food Sci. Emerg. Technol. 9, 328–340. https://doi.org/10.1016/j.ifset.2007.09.003

Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. k-nearest neighbor classification, in: Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (Eds.), Data Mining in Agriculture, Springer Optimization and Its Applications. Springer, New York, NY, pp. 83–106. https://doi.org/10.1007/978-0-387-88615-2_4

Muhammad, A.I., Xiang, Q., Liao, X., Liu, D., Ding, T., 2018. Understanding the impact of nonthermal plasma on food constituents and microstructure—A review. Food Bioprocess Technol. 11, 463–486. https://doi.org/10.1007/s11947-017-2042-9

Muñoz-Cuevas, M., Metris, A., Baranyi, J., 2012. Predictive modelling of *Salmonella*: From cell cycle measurements to e-models. Food Res. Int. 45, 852–862. https://doi.org/10.1016/j.foodres.2011.04.033

Nasheri, N., Doctor, T., Chen, A., Harlow, J., Gill, A., 2020. Evaluation of high-pressure processing in inactivation of the Hepatitis E virus. Front. Microbiol. 11, 461. https://doi.org/10.3389/fmicb.2020.00461

National Academies of Sciences, Engineering, Medicine, 2019. Reproducibility and

replicability in science. The National Academies Press, Washington, DC. https://doi.org/10.17226/25303

Neto, E.C., Bare, J.C., Margolin, A.A., 2014. Simulation studies as designed experiments: The comparison of penalized regression models in the "large p, small n" setting. PLoS One 9, 5–6. https://doi.org/10.1371/journal.pone.0107957

Niquet, R., Boehm, D., Schnabel, U., Cullen, P., Bourke, P., Ehlbeck, J., 2018. Characterising the impact of post-treatment storage on chemistry and antimicrobial properties of plasma treated water derived from microwave and DBD sources. Plasma Process. Polym. 15, 4–11. https://doi.org/10.1002/ppap.201700127

Niu, D., Zeng, X.-A., Ren, E.-F., Xu, F.-Y., Li, J., Wang, M.-S., Wang, R., 2020. Review of the application of pulsed electric fields (PEF) technology for food processing in China. Food Res. Int. 137, 109715. https://doi.org/10.1016/j.foodres.2020.109715

Niu, H., Yuan, L., Zhou, H., Yun, Y., Li, J., Tian, J., Zhong, K., Zhou, L., 2022. Comparison of the effects of high pressure processing, pasteurization and high temperature short time on the physicochemical attributes, nutritional quality, aroma profile and sensory characteristics of passion fruit purée. Foods 11, 632. https://doi.org/10.3390/foods11050632

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., Ananiadou, S., 2015. Using text mining for study identification in systematic reviews: A systematic review of current approaches. Syst. Rev. 4, 5. https://doi.org/10.1186/2046-4053-4-5

O' Neill, C.M., Cruz-Romero, M.C., Duffy, G., Kerry, J.P., 2018. Shelf life extension of vacuum-packed salt reduced frankfurters and cooked ham through the combined application of high pressure processing and organic acids. Food Packag. Shelf Life 17, 120–128. https://doi.org/10.1016/j.fpsl.2018.06.008

O'Reilly, C.E., O'Connor, P.M., Kelly, A.L., Beresford, T.P., Murphy, P.M., 2000. Use of hydrostatic pressure for inactivation of microbial contaminants in cheese. Appl. Environ. Microbiol. 66, 4890–4896. https://doi.org/10.1128/AEM.66.11.4890-4896.2000

Oehmigen, K., Winter, J., Hähnel, M., Wilke, C., Brandenburg, R., Weltmann, K.-D., Woedtke, T. von, 2011. Estimation of possible mechanisms of *Escherichia coli* inactivation by plasma treated sodium chloride solution. Plasma Process. Polym. 8, 904–913. https://doi.org/10.1002/ppap.201000099

Özdemir, M.A., Özdemir, G.D., Gül, M., Güren, O., Ercan, U.K., 2023. Machine learning to predict the antimicrobial activity of cold atmospheric plasma-activated liquids. Mach. Learn. Sci. Technol. 4, 1. https://doi.org/10.1088/2632-2153/acc1c0

Ozen, E., Singh, R.K., 2020. Atmospheric cold plasma treatment of fruit juices: A review. Trends Food Sci. Technol. 103, 144–151. https://doi.org/10.1016/j.tifs.2020.07.020

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville,

J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372, 1–6. https://doi.org/10.1136/bmj.n71

Pampoukis, G., 2025. Plasma-food-metadata-schema [GitHub repository]. Available at: https://github.com/gpampoukis/plasma-food-metadata-schema.

Pampoukis, G., Ham, J., Zwietering, M.H., Den Besten, H.M.W., 2026. The decontamination efficacy of pulsed electric field in food: Establishing global kinetic parameters. Int. J. Food Microbiol. 447, 111547. https://doi.org/10.1016/j.ijfoodmicro.2025.111547

Pampoukis, G., Weihe, T., Wagner, R., Becker, M.M., Yao, Y., Nierop Groot, M.N., Schnabel, U., 2025. A metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications. Sci. Data 12, 838. https://doi.org/10.1038/s41597-025-05203-5

Pampoukis, G., Zwietering, M.H., Den Besten, H.M.W., 2024. Ranking factors affecting the decontamination efficacy of non-thermal plasma: The approach of dissipated power per plasma volume through machine learning modeling. Innov. Food Sci. Emerg. Technol. 96, 103773. https://doi.org/10.1016/j.ifset.2024.103773

Patil, S., Moiseev, T., Misra, N.N., Cullen, P.J., Mosnier, J.P., Keener, K.M., Bourke, P., 2014. Influence of high voltage atmospheric cold plasma process parameters and role of relative humidity on inactivation of *Bacillus atrophaeus* spores inside a sealed package. J. Hosp. Infect. 88, 162–169. https://doi.org/10.1016/j.jhin.2014.08.009

Patterson, M.F., Quinn, M., Simpson, R., Gilmour, A., 1996. High pressure inactivation in foods of animal origin, in: Hayashi, R., Balny, C. (Eds.), Progress in Biotechnology, High Pressure Bioscience and Biotechnology. Elsevier, pp. 267–272. https://doi.org/10.1016/S0921-0423(06)80046-8

Pavlovich, M.J., Chang, H.-W., Sakiyama, Y., Clark, D.S., Graves, D.B., 2013. Ozone correlates with antibacterial effects from indirect air dielectric barrier discharge treatment of water. J. Phys. D Appl. Phys. 46, 4–8. https://doi.org/10.1088/0022-3727/46/14/145202

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pereira, R.N., Vicente, A.A., 2010. Environmental impact of novel thermal and non-thermal technologies in food processing. Food Res. Int. 43, 1936–1943. https://doi.org/10.1016/j.foodres.2009.09.013

Pérez, P., Hashemi, S., Cano-Lamadrid, M., Martínez-Zamora, L., Gómez, P.A., Artés-

Hernández, F., 2023. Effect of ultrasound and high hydrostatic pressure processing on quality and bioactive compounds during the shelf life of a broccoli and carrot by-products beverage. Foods 12, 3808. https://doi.org/10.3390/foods12203808

Piskarev, I.M., Ivanova, I.P., 2021. Comparison of chemistry induced by direct and indirect plasma treatment of water to the effect of UV radiation. Plasma Chem. Plasma Process. 41, 447–475. https://doi.org/10.1007/s11090-020-10127-6

Porrini, M., Riso, P., Testolin, G., 1998. Absorption of lycopene from single or daily portions of raw and processed tomato. Br. J. Nutr. 80, 353–361. https://doi.org/10.1017/S000711459800141X

Printezis, I., Grebitus, C., Hirsch, S., 2019. The price is right!? A meta-regression analysis on willingness to pay for local food. PLoS One 14, e0215847. https://doi.org/10.1371/journal.pone.0215847

Puligundla, P., Pyun, Y.-R., Mok, C., 2018. Pulsed electric field (PEF) technology for microbial inactivation in low-alcohol red wine. Food Sci. Biotechnol. 27, 1691–1696. https://doi.org/10.1007/s10068-018-0422-1

Ragni, L., Berardinelli, A., Iaccheri, E., Gozzi, G., Cevoli, C., Vannini, L., 2016. Influence of the electrode material on the decontamination efficacy of dielectric barrier discharge gas plasma treatments towards *Listeria monocytogenes* and *Escherichia coli*. Innov. Food Sci. Emerg. Technol. 37, 170–176. https://doi.org/10.1016/j.ifset.2016.07.029

Rajan, A., Boopathy, B., Radhakrishnan, M., Rao, L., Schlüter, O.K., Tiwari, B.K., 2023. Plasma processing: A sustainable technology in agri-food processing. Sustain. Food Technol. 1, 9–49. https://doi.org/10.1039/D2FB00014H

Raso, J., Frey, W., Ferrari, G., Pataro, G., Knorr, D., Teissie, J., Miklavčič, D., 2016. Recommendations guidelines on the key information to be reported in studies of application of PEF technology in food and biotechnological processes. Innov. Food Sci. Emerg. Technol. 37, 312–321. https://doi.org/10.1016/j.ifset.2016.08.003

Rastogi, N.K., Raghavarao, K.S.M.S., Balasubramaniam, V.M., Niranjan, K., Knorr, D., 2007. Opportunities and challenges in high pressure processing of foods. Crit. Rev. Food Sci. Nutr. 47, 69–112. https://doi.org/10.1080/10408390600626420

Rattenbury, T., Hellerstein, J.M., Heer, J., Kandel, S., Carreras, C., 2017. Principles of data wrangling: Practical techniques for data preparation, 1st ed. O'Reilly Media, Inc., Sebastopol, CA, USA.

Rigaux, C., Denis, J.-B., Albert, I., Carlin, F., 2013. A meta-analysis accounting for sources of variability to estimate heat resistance reference parameters of bacteria using hierarchical Bayesian modeling: Estimation of D at 121.1 °C and pH 7, z_T and z_{pH} of *Geobacillus stearothermophilus*. Int. J. Food Microbiol. 161, 112–120. https://doi.org/10.1016/j.ijfoodmicro.2012.12.001

Rodríguez-Roque, M.J., Ancos, B. de, Sánchez-Vega, R., Sánchez-Moreno, C., Cano, M.P., Elez-Martínez, P., Martín-Belloso, O., 2016. Food matrix and processing influ-

ence on carotenoid bioaccessibility and lipophilic antioxidant activity of fruit juice-based beverages. Food Funct. 7, 380–389. https://doi.org/10.1039/C5FO01060H

Rohatgi, A., 2025. WebPlotDigitizer (version 4.6). Available at: https://automeris.io.

Rolfe, C.A., Anderson, N.M., Black, D.G., Lee, A., 2023. Barotolerance of acid-adapted and cold-adapted bacterial isolates of *E. coli* O157:H7, *Salmonella* spp., and *L. monocytogenes* in an acidic buffer model. J. Food Prot. 86, 100116. https://doi.org/10.1016/j.jfp.2023.100116

Román, S., Sánchez-Siles, L.M., Siegrist, M., 2017. The importance of food naturalness for consumers: Results of a systematic review. Trends Food Sci. Technol. 67, 44–57. https://doi.org/10.1016/j.tifs.2017.06.010

Romero, O., Wrembel, R., 2020. Data engineering for data science: Two sides of the same coin, in: Song, M., Song, I.-Y., Kotsis, G., Tjoa, A.M., Khalil, I. (Eds.), Big Data Analytics and Knowledge Discovery. Springer International Publishing, Cham, pp. 157–166. https://doi.org/10.1007/978-3-030-59065-9_13

Sampedro, F., McAloon, A., Yee, W., Fan, X., Geveke, D.J., 2014. Cost analysis and environmental impact of pulsed electric fields and high pressure processing in comparison with thermal pasteurization. Food Bioprocess Technol. 7, 1928–1937. https://doi.org/10.1007/s11947-014-1298-6

Sampedro, F., Rodrigo, D., Martínez, A., 2011. Modelling the effect of pH and pectin concentration on the PEF inactivation of *Salmonella enterica* serovar Typhimurium by using the Monte Carlo simulation. Food Control 22, 420–425. https://doi.org/10.1016/j.foodcont.2010.09.013

Santillana Farakos, S.M., Zwietering, M.H., 2011. Data analysis of the inactivation of foodborne microorganisms under high hydrostatic pressure to establish global kinetic parameters and influencing factors. J. Food Prot. 74, 2097–2106. https://doi.org/10.4315/0362-028X.JFP-11-162

Saremnezhad, S., Soltani, M., Faraji, A., Hayaloglu, A.A., 2021. Chemical changes of food constituents during cold plasma processing: A review. Food Res. Int. 147, 110552. https://doi.org/10.1016/j.foodres.2021.110552

Scherhaufer, S., Moates, G., Hartikainen, H., Waldron, K., Obersteiner, G., 2018. Environmental impacts of food waste in Europe. Waste Manag. 77, 98–113. https://doi.org/10.1016/j.wasman.2018.04.038

Schielzeth, H., 2010. Simple means to improve the interpretability of regression coefficients. Methods Ecol. Evol. 1, 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schlüter, O., Ehlbeck, J., Hertel, C., Habermeyer, M., Roth, A., Engel, K.-H., Holzhauser, T., Knorr, D., Eisenbrand, G., 2013. Opinion on the use of plasma processes for treatment of foods. Mol. Nutr. Food Res. 57, 920–927. https://doi.org/10.1002/mnfr.201300039

Schnabel, U., Balazinski, M., Wagner, R., Stachowiak, J., Boehm, D., Andrasch, M.,

Bourke, P., Ehlbeck, J., 2021a. Optimizing the application of plasma functionalised water (PFW) for microbial safety in fresh-cut endive processing. Innov. Food Sci. Emerg. Technol. 72, 102745. https://doi.org/10.1016/j.ifset.2021.102745

Schnabel, U., Handorf, O., Stachowiak, J., Boehm, D., Weit, C., Weihe, T., Schäfer, J., Below, H., Bourke, P., Ehlbeck, J., 2021b. Plasma-functionalized water: From bench to prototype for fresh-cut lettuce. Food Eng. Rev. 13, 115–135. https://doi.org/10.1007/s12393-020-09238-9

Schnabel, U., Niquet, R., Krohmann, U., Winter, J., Schlüter, O., Weltmann, K.-D., Ehlbeck, J., 2012. Decontamination of microbiologically contaminated specimen by direct and indirect plasma treatment. Plasma Process. Polym. 9, 569–575. https://doi.org/10.1002/ppap.201100088

Scholtz, V., Pazlarova, J., Souskova, H., Khun, J., Julak, J., 2015. Nonthermal plasma — a tool for decontamination and disinfection. Biotechnol. Adv. 33, 1108–1119. https://doi.org/10.1016/j.biotechadv.2015.01.002

Schottroff, F., Gratz, M., Krottenthaler, A., Johnson, N.B., Bédard, M.F., Jaeger, H., 2019. Pulsed electric field preservation of liquid whey protein formulations – influence of process parameters, pH, and protein content on the inactivation of *Listeria innocua* and the retention of bioactive ingredients. J. Food Eng. 243, 142–152. https://doi.org/10.1016/j.jfoodeng.2018.09.003

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with Python, in: 9th Python in Science Conference. pp. 92–96.

Sehrawat, R., Kaur, B.P., Nema, P.K., Tewari, S., Kumar, L., 2020. Microbial inactivation by high pressure processing: Principle, mechanism and factors responsible. Food Sci. Biotechnol. 30, 19–35. https://doi.org/10.1007/s10068-020-00831-6

Serra-Castelló, C., 2023. Combined application of experimental and predictive modelling approaches towards the microbial safety of ready-to-eat meat products (PhD thesis). Universitat de Girona.

Serra-Castelló, C., Desriac, N., Jofré, A., Belletti, N., Coroller, L., Bover-Cid, S., 2022. Key factors determining the behavior of pathogens in dry-cured ham after high pressure processing. Appl. Sci. 12, 12732. https://doi.org/10.3390/app122412732

Serra-Castelló, C., Jofré, A., Belletti, N., Garriga, M., Bover-Cid, S., 2021. Modelling the piezo-protection effect exerted by lactate on the high pressure resistance of *Listeria monocytogenes* in cooked ham. Food Res. Int. 140, 110003. https://doi.org/10.1016/j.foodres.2020.110003

Settles, B., 2009. Active learning literature survey (Computer Sciences Technical Report No. 1648). University of Wisconsin-Madison Department of Computer Sciences.

Shad, E., Raninen, K., Podergina, S., Chan, L.I., Tong, K.P., Hälikkä, H., Huovinen, M., Korhonen, J., 2024. Impact of high-pressure processing on quality and safety of high-oil-content pesto sauce: A comparative study with thermal processing. Appl.

Sci. 14, 9425. https://doi.org/10.3390/app14209425

Shanker, M.A., Khanashyam, A.C., Pandiselvam, R., Joshi, T.J., Thomas, P.E., Zhang, Y., Rustagi, S., Bharti, S., Thirumdas, R., Kumar, M., Kothakota, A., 2023. Implications of cold plasma and plasma activated water on food texture- a review. Food Control 151, 109793. https://doi.org/10.1016/j.foodcont.2023.109793

Shao, Y., Ramaswamy, H.S., Bussey, J., Harris, R., Austin, J.W., 2022. High pressure destruction kinetics of *Clostridium botulinum* (Group I, strain PA9508B) spores in milk at elevated temperatures. LWT 154, 112671. https://doi.org/10.1016/j.lwt.2021.112671

Shelar, A., Vikram Singh, A., Dietrich, P., Singh Maharjan, R., Thissen, A., N. Didwal, P., Shinde, M., Laux, P., Luch, A., Mathe, V., Jahnke, T., Chaskar, M., Patil, R., 2022. Emerging cold plasma treatment and machine learning prospects for seed priming: A step towards sustainable food production. RSC Adv. 12, 10467–10488. https://doi.org/10.1039/D2RA00809B

Shimizu, T., Lachner, V., Zimmermann, J.L., 2017. Surface microdischarge plasma for disinfection. Plasma Med. 7, 175–187. https://doi.org/10.1615/PlasmaMed.2017019455

Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310. https://doi.org/10.1214/10-STS330

Shwartz-Ziv, R., Armon, A., 2022. Tabular data: Deep learning is not all you need. Inf. Fusion 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Siemer, C., Toepfl, S., Heinz, V., 2014. Inactivation of *Bacillus subtilis* spores by pulsed electric fields (PEF) in combination with thermal energy – I. Influence of process- and product parameters. Food Control 39, 163–171. https://doi.org/10.1016/j.foodcont.2013.10.025

Silva, A., Rocha, C., Ribeiro, J.C., Aganovic, K., Lima, R.C., Methven, L., Cunha, L.M., 2024. Consumer perception of risk towards new sustainable non-thermal food processing technologies: A cross-cultural study between Portugal, Germany, and the UK. Innov. Food Sci. Emerg. Technol. 96, 103772. https://doi.org/10.1016/j.ifset.2024.103772

Simpson, B.K., 1998. High pressure processing of fresh seafoods, in: Shahidi, F., Ho, C.-T., Van Chuyen, N. (Eds.), Process-Induced Chemical Changes in Food. Springer US, Boston, MA, pp. 67–80. https://doi.org/10.1007/978-1-4899-1925-0_7

Singh, S., Shalini, R., 2016. Effect of hurdle technology in food preservation: A review. Crit. Rev. Food Sci. Nutr. 56, 641–649. https://doi.org/10.1080/10408398.2012.761594

Smet, C., Baka, M., Steen, L., Fraeye, I., Walsh, J.L., Valdramidis, V.P., Van Impe, J.F., 2019. Combined effect of cold atmospheric plasma, intrinsic and extrinsic factors on the microbial behavior in/on (food) model systems during storage. Innov. Food Sci. Emerg. Technol. 53, 3–17. https://doi.org/10.1016/j.ifset.2018.05.016

Song, A.Y., Oh, Y.J., Kim, J.E., Song, K.B., Oh, D.H., Min, S.C., 2015. Cold plasma treatment for microbial safety and preservation of fresh lettuce. Food Sci. Biotechnol. 24, 1717–1724. https://doi.org/10.1007/s10068-015-0223-8

Souza, I.A. de, Nascimento Neto, A.B. do, Queiroz, J.C.A. de, Matamoros, E.P., Costa, T.H. de C., Feitor, M.C., Souza, J.M.L. de, Camara, N.T., Da Silva Severiano Sobrinho, V., 2016. Study of the influence of variation in distances between electrodes in spectral DBD plasma excitation. Mater. Res. 19, 202–206. https://doi.org/10.1590/1980-5373-MR-2015-0205

Štěpánová, V., Šrámková, P., Sihelník, S., Stupavská, M., Jurmanová, J., Kováčik, D., 2021. The effect of ambient air plasma generated by coplanar and volume dielectric barrier discharge on the surface characteristics of polyamide foils. Vacuum 183, 109887. https://doi.org/10.1016/j.vacuum.2020.109887

Stulić, V., Butorac, A., Pavičić, T.V., Ivanov, M., Herceg, Z., 2021. Stress response of *Escherichia coli* and *Listeria innocua* to plasma treatment. LWT 147, 111695. https://doi.org/10.1016/j.lwt.2021.111695

Suits, D.B., 1957. Use of dummy variables in regression equations. J. Am. Stat. Assoc. 52, 548–551. https://doi.org/10.2307/2281705

Tanino, M., Xilu, W., Takashima, K., Katsura, S., Mizuno, A., 2005. Sterilization using dielectric barrier discharge at atmospheric pressure, in: Fourtieth IAS Annual Meeting. Conference Record of the 2005 Industry Applications Conference, 2005. pp. 784–788 Vol. 2. https://doi.org/10.1109/IAS.2005.1518421

Tanino, T., Matsui, M., Uehara, K., Ohshima, T., 2020. Inactivation of *Bacillus subtilis* spores on the surface of small spheres using low-pressure dielectric barrier discharge. Food Control 109, 106890. https://doi.org/10.1016/j.foodcont.2019.106890

Thamsuaidee, A., Schaefer, E., Schneider, D., Siemer, C., Valdramidis, V.P., 2024. Disentangling the effects of electroporation and heat during pulsed electric field (PEF) processing of oat-based milk alternative: A case study on *Lactiplantibacillus plantarum* inactivation. Innov. Food Sci. Emerg. Technol. 94, 103691. https://doi.org/10.1016/j.ifset.2024.103691

Thirumdas, R., Kothakota, A., Annapure, U., Siliveru, K., Blundell, R., Gatt, R., Valdramidis, V.P., 2018. Plasma activated water (PAW): Chemistry, physico-chemical properties, applications in food and agriculture. Trends Food Sci. Technol. 77, 21–31. https://doi.org/10.1016/j.tifs.2018.05.007

Thirumdas, R., Sarangapani, C., Annapure, U.S., 2015. Cold plasma: A novel non-thermal technology for food processing. Food Biophys. 10, 1–11. https://doi.org/10.1007/s11483-014-9382-z

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288.

Tiede, R., Hirschberg, J., Viöl, W., Emmert, S., 2016. A μs-pulsed dielectric barrier discharge source: Physical characterization and biological effects on human skin

fibroblasts. Plasma Process. Polym. 13, 775–787. https://doi.org/10.1002/ppap.201500190

Timmermans, R.A.H., Mastwijk, H.C., Berendsen, L.B.J.M., Nederhoff, A.L., Matser, A.M., Van Boekel, M.A.J.S., Nierop Groot, M.N., 2019. Moderate intensity pulsed electric fields (PEF) as alternative mild preservation technology for fruit juice. Int. J. Food Microbiol. 298, 63–73. https://doi.org/10.1016/j.ijfoodmicro.2019.02.015

Timmermans, R.A.H., Nierop Groot, M.N., Nederhoff, A.L., Van Boekel, M.A.J.S., Matser, A.M., Mastwijk, H.C., 2014. Pulsed electric field processing of different fruit juices: Impact of pH and temperature on inactivation of spoilage and pathogenic micro-organisms. Int. J. Food Microbiol. 173, 105–111. https://doi.org/10.1016/j.ijfoodmicro.2013.12.022

Timmermans, R., Nierop Groot, M., Matser, A., 2022. Liquid food pasteurization by pulsed electric fields, in: Raso, J., Heinz, V., Alvarez, I., Toepfl, S. (Eds.), Pulsed Electric Fields Technology for the Food Industry: Fundamentals and Applications. Springer International Publishing, Cham, pp. 299–323. https://doi.org/10.1007/978-3-030-70586-2_8

Tiwari, B.K., O'Donnell, C.P., Cullen, P.J., 2009. Effect of non thermal processing technologies on the anthocyanin content of fruit juices. Trends Food Sci. Technol. 20, 137–145. https://doi.org/10.1016/j.tifs.2009.01.058

Toepfl, S., 2012. Pulsed electric field food processing: Industrial equipment design and commercial applications. Stewart Postharvest Rev. 8, 1–7. https://doi.org/10.2212/spr.2012.2.4

Toepfl, S., 2011. Pulsed electric field food treatment - scale up from lab to industrial scale. Procedia Food Sci. 1, 776–779. https://doi.org/10.1016/j.profoo.2011.09.117

Toepfl, S., Heinz, V., Knorr, D., 2005. 4 - overview of pulsed electric field processing for food, in: Sun, D.-W. (Ed.), Emerging Technologies for Food Processing. Academic Press, London, pp. 69–97. https://doi.org/10.1016/B978-012676757-5/50006-2

Torrents-Masoliver, B., Serra-Castelló, C., Jofré, A., Ribas-Agustí, A., Bover-Cid, S., 2025. Towards safer fruit and vegetable products: Global meta-regression modelling of vegetative pathogen inactivation by non-thermal high-pressure processing. Food Res. Int. 218, 116912. https://doi.org/10.1016/j.foodres.2025.116912

Tran, D., Schouteten, J.J., Gellynck, X., De Steur, H., 2024. How do consumers value food traceability? – A meta-analysis. Food Control 162, 110453. https://doi.org/10.1016/j.foodcont.2024.110453

Tsagkaropoulou, T., Karatzas, K.A.G., 2024. Microbial species and strain heterogeneity affect resistance to high pressure processing. Innov. Food Sci. Emerg. Technol. 94, 103645. https://doi.org/10.1016/j.ifset.2024.103645

United Nations, 2015. Transforming our world: The 2030 agenda for sustainable development.

U.S. Food and Drug Administration, 2004. Guidance for industry: Juice HACCP haz-

ards and controls guidance (Guidance for industry No. 2002D-0333). U.S. Department of Health and Human Services, Rockville, MD.

U.S. Food and Drug Administration, 2001. 21 CFR part 120 - hazard analysis and critical control point (HACCP) systems (Code of federal regulations No. 21 CFR Part 120). U.S. Department of Health and Human Services, Washington, DC.

U.S. Food and Drug Administration, 1998. BAM media M157: Trypticase soy broth with 0.6% yeast extract (TSBYE). U.S. Food and Drug Administration.

U.S. Food and Drug Administration, 1979. 21 CFR part 113 - thermally processed low-acid foods packaged in hermetically sealed containers (Code of federal regulations No. 21 CFR Part 113). U.S. Department of Health and Human Services, Washington, DC.

Valdivielso, V., 2023. High pressure processing (HPP): Suitable and not suitable products.

Van Asselt, E.D., Zwietering, M.H., 2006. A systematic approach to determine global thermal inactivation parameters for various food pathogens. Int. J. Food Microbiol. 107, 73–82. https://doi.org/10.1016/j.ijfoodmicro.2005.08.014

Van Boekel, M.A.J.S., 2008. Kinetic modeling of food quality: A critical review. Compr. Rev. Food Sci. Food Saf. 7, 144–158. https://doi.org/10.1111/j.1541-4337.2007.00036.x

Van de Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., Oberski, D.L., 2021. An open source machine learning framework for efficient and transparent systematic reviews. Nat. Mach. Intell. 3, 125–133. https://doi.org/10.1038/s42256-020-00287-7

Van der Aalst, W., 2016. Data science in action, in: Process Mining: Data Science in Action. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–23. https://doi.org/10.1007/978-3-662-49851-4_1

Van der Sluis, A., 2025. Exploring efficacy and sustainability aspects of high-pressure processing and pulsed electric field on the inactivation of *E. coli* in fruit juices and ciders (Master's thesis). Wageningen University & Research, Wageningen, The Netherlands.

Van Rossum, G., Drake, F.L., 2009. Python 3 reference manual. CreateSpace, Scotts Valley, CA.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nat. Methods 17,

261–272. https://doi.org/10.1038/s41592-019-0686-2
Wallace, B.C., Trikalinos, T.A., Lau, J., Brodley, C., Schmid, C.H., 2010. Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 11, 55. https://doi.org/10.1186/1471-2105-11-55
Wan, Z., Misra, N.N., Li, G., Keener, K.M., 2021. High voltage atmospheric cold plasma treatment of *Listeria innocua* and *Escherichia coli* K-12 on Queso Fresco (fresh cheese). LWT 146, 4–8. https://doi.org/10.1016/j.lwt.2021.111406
Waskom, M.L., 2021. Seaborn: Statistical data visualization. J. Open Source Softw. 6, 1–4. https://doi.org/10.21105/joss.03021
Waskow, A., Avino, F., Howling, A., Furno, I., 2022. Entering the plasma agriculture field: An attempt to standardize protocols for plasma treatment of seeds. Plasma Process. Polym. 19, 2100152. https://doi.org/10.1002/ppap.202100152
Weihe, T., Yao, Y., Opitz, N., Wagner, R., Krall, J., Schnabel, U., Below, H., Ehlbeck, J., 2023. Plasma-treated water: A comparison with analog mixtures of traceable ingredients. Microorganisms 11, 932. https://doi.org/10.3390/microorganisms11040932
Weihe, T., Yao, Y., Stachowiak, J., Ehlbeck, J., Schnabel, U., 2024. Microwave plasmas in food safety. A review. Innov. Food Sci. Emerg. Technol. 96, 103774. https://doi.org/10.1016/j.ifset.2024.103774
Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L.B. da, Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., Van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., Van der Lei, J., Van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18
World Health Organization, 2021. Food systems for health: Information brief. World Health Organization.
World Health Organization, 2015. WHO estimates of the global burden of foodborne diseases.
Yan, D., Sherman, J.H., Keidar, M., 2016. Cold atmospheric plasma, a novel promising anti-cancer treatment modality. Oncotarget 8, 15977–15995. https://doi.org/10.18632/oncotarget.13304
Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Yeak, K.Y.C., Dank, A., Den Besten, H.M.W., Zwietering, M.H., 2024a. A web-based microbiological hazard identification tool for infant foods. Food Res. Int. 178, 113940. https://doi.org/10.1016/j.foodres.2024.113940

Yeak, K.Y.C., Garre, A., Membré, J.-M., Zwietering, M.H., Den Besten, H.M.W., 2024b. Systematic risk ranking of microbiological hazards in infant foods. Food Res. Int. 192, 114788. https://doi.org/10.1016/j.foodres.2024.114788

Yeak, K.Y.C., Palmont, P., Rivière, G., Bemrah, N., Den Besten, H.M.W., Zwietering, M.H., 2022. Microbial and chemical hazard identification in infant food chains. Glob. Pediatr. 2, 100010. https://doi.org/10.1016/j.gpeds.2022.100010

Yigitbasioglu, O.M., Velcu, O., 2012. A review of dashboards in performance management: Implications for design and research. Int. J. Account. Inf. Syst. 13, 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Young, I., Waddell, L., Harding, S., Greig, J., Mascarenhas, M., Sivaramalingam, B., Pham, M.T., Papadopoulos, A., 2015. A systematic review and meta-analysis of the effectiveness of food safety education interventions for consumers in developed countries. BMC Public Health 15, 822. https://doi.org/10.1186/s12889-015-2171-x

Zdolec, N., Kotsiri, A., Houf, K., Alvarez-Ordóñez, A., Blagojevic, B., Karabasil, N., Salines, M., Antic, D., 2022. Systematic review and meta-analysis of the efficacy of interventions applied during primary processing to reduce microbial contamination on pig carcasses. Foods 11, 2110. https://doi.org/10.3390/foods11142110

Zhang, Z.-H., Wang, L.-H., Zeng, X.-A., Han, Z., Brennan, C.S., 2019. Non-thermal technologies and its current and future application in the food industry: A review. Int. J. Food Sci. Technol. 54, 1–13. https://doi.org/10.1111/ijfs.13903

Zhao, Y.-M., Patange, A., Sun, D.-W., Tiwari, B., 2020. Plasma-activated water: Physicochemical properties, microbial inactivation mechanisms, factors influencing antimicrobial effectiveness, and applications in the food industry. Compr. Rev. Food Sci. Food Saf. 19, 3951–3979. https://doi.org/10.1111/1541-4337.12644

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Zwietering, M.H., Garre, A., Wiedmann, M., Buchanan, R.L., 2021. All food processes have a residual risk, some are small, some very small and some are extremely small: Zero risk does not exist. Curr. Opin. Food Sci. 39, 83–92. https://doi.org/10.1016/j.cofs.2020.12.017

Appendix

Summary

Acknowledgements

About the author

Overview of completed training activities

Summary

The growing demand for safe, nutritious, and high-quality food that is produced with limited environmental impact and minimal food loss and waste requires preservation technologies that reliably control microbiological hazards without unnecessarily compromising product quality. Conventional thermal processing has a long history of food applications but can negatively affect sensory and nutritional attributes. Mild or non-thermal technologies such as non-thermal plasma (NTP), pulsed electric field (PEF), and high-pressure processing (HPP) are increasingly proposed as alternatives that may combine high levels of microbial safety with improved product quality and, in some cases, lower environmental burdens. At the same time, food business operators and risk assessors need quantitative evidence on decontamination performance across products, microorganisms, and process settings in order to design and validate control measures, perform risk assessments, and compare technologies on a fair basis.

This thesis aimed to obtain quantitative, generic insights into the decontamination efficacy of NTP, PEF, and HPP under food-relevant conditions by systematically collecting, structuring, and analyzing existing literature data. Rather than optimizing a single process or product, the work focused on extracting global kinetic parameters and identifying the key processing, matrix, and microbial factors that drive variability in inactivation outcomes. To do so, it combined meta-regression and other statistical approaches with recent data-science methods, including active learning for literature screening, transparent data-wrangling pipelines, and machine-learning algorithms to obtain generic insights. The resulting benchmarks are intended to support process design and validation, quantitative microbiological risk assessment, and technology comparison in a broader sustainability context.

Chapter 2 addressed non-thermal plasma decontamination in food-related applications. A literature-derived database of $\log_{10}D$-values was compiled across a wide range of plasma sources, matrices, and target microorganisms. Because the data were highly heterogeneous and reporting practices differed considerably between studies, classical kinetic meta-regression was not feasible at a generic level. Instead, the chapter introduced dissipated power per plasma volume as an integrated descriptor of process intensity for dielectric barrier discharge configurations and used machine-learning methods to rank the relative importance of process, product, and microbial factors. Power density and matrix category emerged as the most influential predictors of decontamination efficacy, followed by microbial genus and pH. The chapter also demonstrated, for the first time in food microbiology, the use of active learning to make literature screening more efficient, reducing the manual workload while maintaining coverage of relevant studies.

Chapter 3 responded to the lack of standardization that limited the analyses in **Chapter 2** by developing a dedicated metadata schema for NTP decontamination in food-related applications. Building on the existing Plasma-MDS framework from plasma science, the new schema defined a structured set of parameters covering the plasma source, working medium, food or target matrix, microbial characteristics, and diagnostic methods. The schema was implemented in human- and machine-readable format and illustrated with concrete examples how experimental parameters should be reported. By making key experimental details explicit and interoperable, this work supports FAIR (findable, accessible, interoperable, reusable) data practices for NTP and lays the foundation for future meta-analyses and predictive modelling once more consistently reported metadata have accumulated.

Chapters 4 and **5** focused on PEF and HPP, respectively, for microbial decontamination in food products. In **Chapter 4**, a harmonized database of 569 PEF experiments was compiled for seven microbial groups, including *E. coli*, *Listeria* spp., *Pseudomonas* spp., *S. aureus*, *S. cerevisiae*, *Salmonella* spp., and lactic acid bacteria (LAB), across multiple liquid food matrices. Decontamination efficacy was expressed using the decimal-reduction volumetric energy input ($D_{E_{\text{vol}}}$), defined as the energy per litre required for a one-$\log_{10}$-reduction. Meta-regression was used to estimate global and matrix-stratified $D_{E_{\text{vol}}}$ values, revealing that matrix composition can change the decimal-reduction energy by up to eight-fold. High-acid fruit juices showed comparatively low $D_{E_{\text{vol}}}$ values, whereas high-acid alcoholic products were more protective, demonstrating that acidity alone is insufficient to describe matrix effects. A decision tree was developed to provide estimates of the energy input needed to achieve a 5-$\log_{10}$-reduction for different microorganisms and matrices, and an interactive dashboard was created to allow users to visualize and explore the underlying data in real time.

In **Chapter 5**, high-pressure processing was examined using an extended database of 2692 $\log_{10}D$-values for 15 microbiological hazards treated at temperatures ≤ 45 °C. Using a Bigelow-type model fitted to the available data, global $\log_{10}D$-values were estimated at a reference pressure of 500 MPa, allowing microbiological hazards to be grouped into three resistance categories (sensitive, moderate, and resistant) at this pressure level. The analysis confirmed that matrix properties strongly influence HPP resistance: low water activity ($a_w \leq 0.95$) tended to protect microorganisms, while acidic or aqueous matrices (pH ≤ 4.5 and $a_w > 0.95$) facilitated inactivation. Multiple linear regression indicated that a_w and pH categories were the dominant product-level predictors of inactivation, whereas other variables such as protein and carbohydrate content had comparatively limited effects. Together, these results led to a decision-support framework combining (i) hazard resistance ranking at 500 MPa, (ii) matrix-stratified inactivation estimates by food category, and (iii) multiple linear regression models based on a_w and pH, providing flexibility depending on the level of information

available to users.

Chapter 6 integrated the technology-specific findings into a broader food safety and sustainability perspective by comparing thermal processing, HPP, and PEF for the inactivation of *E. coli* in high-acid fruit juices. Using meta-regression models developed in this thesis, the chapter derived efficacy benchmarks ($\log_{10}$-reductions) and their variability for each technology under various industrially relevant processing conditions reported in the literature. These efficacy estimates were linked to energy use and, where available, greenhouse gas emission data from the literature. The analysis illustrated that conclusions about "more sustainable" technologies depend critically on matching decontamination performance across processes and on considering both energy and GHG emissions per $\log_{10}$-reduction rather than per litre alone. It also highlighted that electricity-driven mild technologies may become relatively more favorable as electricity production relies more and more on renewable sources. However, although thermal processes with heat recovery compromise fresh-like properties, they can remain highly efficient from both an energy and a decontamination performance perspective. Finally, the chapter outlined how the meta-regression framework used here could be coupled with life-cycle assessment and extended to other technologies and product-pathogen combinations.

Taken together, this thesis shows that carefully curated databases, transparent data-processing pipelines, and quantitative synthesis methods can extract generic, practically useful information from heterogeneous literature on mild food processing technologies. For NTP, it clarifies which factors most strongly influence decontamination and introduces a metadata infrastructure to enable more robust comparisons in the future. For PEF and HPP, it delivers global kinetic parameters, hazard and matrix resistance rankings, and simple tools such as decision trees and dashboards that can support process design, validation, and risk assessment. By bringing these pieces together in a comparative framework that explicitly considers both microbiological performance and sustainability factors, the work links data-driven modelling to practical decision-making on process choice. In doing so, it contributes to the evidence base needed to deploy mild preservation technologies in a way that is scientifically justified, transparent, and aligned with broader ambitions for safe, lower-waste, and more sustainable food systems.

Acknowledgements

This thesis is the result of many years of work, but it also reflects the support, patience, and generosity of many people. I am deeply grateful to everyone who contributed, directly or indirectly, to making this work possible and to keeping me grounded throughout the process. I could not have completed it alone.

First, I would like to thank my promotors and supervisors at Wageningen University: **Prof. Heidy den Besten** and **Prof. Marcel Zwietering**. **Heidy**, thank you for trusting me and for bringing me to the forefront at conferences, workshops, and symposia. You supported our everyday process and kept my progress on track and realistic. I was always proud to be part of your team, and I felt secure knowing that you would keep our work carefully coordinated and on schedule. **Marcel**, thank you for being such an inspiring mentor and for consistently bringing a practical perspective to the scientific process. I will always carry with me your advice to "be critical, not cynical." You also taught me to come to meetings well prepared, with a clear idea of what I wanted to get out of them rather than simply accepting what was given. From you, I learned that making a long story short and digestible is often more difficult, but also more useful. Finally, you taught me to be ready to answer questions about my work, even when the answer was not yet certain. These lessons will stay with me.

I also sincerely appreciate that both of you gave me the opportunity and freedom to explore and develop Chapter 3, *A metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications*, even though this work was not originally part of the project and lay outside your direct areas of involvement. Your openness and support meant a great deal to me and allowed this work to become an important part of the thesis.

To my colleagues at FHM, both former and current, thank you for creating a daily environment that made work feel lighter. I am grateful for the conversations and small moments that kept things moving. **Gerda**, thank you for your support during your time as the group's secretary, for always being ready to help with administrative matters, and especially for your patience with my sometimes messy receipt submissions and post-trip expense claims. **Frank**, thank you for staying in touch with genuine interest and for your long, thoughtful messages. That support meant a lot. **Johanna**, I admire your amazing spirit and passion for life. I love the way you can criticize yourself and laugh at yourself without formality or taboo. It is an attribute that is becoming less and less common in our society. **Lukas** and **Domiziana (Domi)**, thank you for being such great gym coaches. Special thanks to **Luuk**, **Wiebe**, and **Koen**, who supported me greatly during the final stage of my PhD and through my transition to Switzerland.

To everyone involved in the TRANSIT project, including the early-stage researchers (ESRs), supervisors, and broader network that came together during our meetings and informal discussions afterward, thank you as well. Special thanks to **Assoc. Prof. Vasilis Valdramidis** for being a great mentor and for hosting me in his lab during my secondment in Athens. Your approach to life and science will continue to inspire me. **Nick** and **Fotis**, the two distinguished members of our "asylum seekers" group (yes, the one you could usually find in bars after midnight), thank you for being part of this journey.

I am also deeply grateful to all of my friends for the amazing times we spent together over the past years.

To my beloved paranymphs, **Alberto (Albi)** and **Domi**, thank you both. **Albi**, it was wonderful to spend fun evenings with you, and I valued your scientific insight as a point of reference. **Domi**, you combine clarity of thought with broad knowledge across many fields. I could always turn to you for thoughtful discussions, and I came away having learned something or gained a new perspective every time.

Linda, my whiskey connoisseur, I loved the gatherings you and **Alessandro** hosted. Preparing drinks and food together was always a laugh, and I cherish the meaningful conversations we shared during those evenings. **Alessandro**, thank you for being there and for your honest, unfiltered opinions. **Tamara**, I enjoy chatting and spending time with you. Thank you for being there whenever I needed help with hectic tasks. You were the one who got me back on calorie tracking, a change that helped me lose more than 20 kilos and return to a healthier weight. You and **Kais** were also the best and kindest hosts for our house gatherings. **Kais**, thank you for the programming insights and tools you shared with me, and for our discussions about life and software development. I also thank **Xuchuan** for joining those conversations and introducing me to Quarto and other useful tools. Quarto is what I used to write this section and the entire thesis. Most importantly, thank you for your inspiring way of thinking about science.

Claire, thank you for the fantastic gatherings, parties, festivals, and drinks, and for making everything so much fun. Together with **Albi**, you helped create some of my most wonderful memories. **Damiano**, your humor, honesty, and consistently good mood brought a lot of joy. **Cristina**, you are an inspiring scientist and an amazing friend. Thank you for being around and for all the fun times we shared both at work and outside of it.

Angela and **Panos**, my two beloved friends, thank you for all the long days we spent together without a single moment of boredom. No meeting of ours could ever last less than a full shift. **Angela**, thank you for supporting me and caring for your friends.

Panos, thank you for all the fun moments and for our insightful discussions about how companies work and are managed.

Jori, thank you for always finding time for long discussions that would often begin over a beer and carry on through the evening.

Soundarya, thank you for always being around. **Rebecca** and **Dimitri**, I always enjoy the time we spend together.

Elina, thank you for our Saturday coffee gatherings. **Paolo**, my first roommate in Wageningen and an ever-(un)reliable drinking buddy, thank you for everything.

To the members of our running group, **Erik**, **James**, **Karin**, **Moniek**, **Oscar**, **Selena**, and of course **Marcel**, thank you. Running has now become a weekly habit for me, whereas in the past I never ran at all. Thank you for helping me make that change and improve my 5 km time from 34 minutes to 26 minutes, even though I still have not managed to outrun **Marcel**.

Dimitris K., thank you for staying in touch, for sharing your insights into electronic music and festivals, and for your movie recommendations. **Maria**, thank you for your close friendship and for giving me extra support to keep going.

Danaos, thank you for being there for a phone call when I needed it.
Evangelos A., thank you for somehow keeping up with every night out and lifting everyone's spirits, even from a thousand kilometers away, and, more importantly, for truly being there when it mattered most and for your encouragement at the right moments.

Dimitra S. and **Frosso**, my best friends in Athens, thank you for the endless hours of discussion we shared, along with liters of tsipouro.

Giannis S., the president of our hearts, and **Dinos**, **Socrates**, and **Natassa**, thank you for our yearly holidays in Crete and for the Friday tavern gatherings whenever I come to Thessaloniki. It is always a joy to spend time with you and discuss life, everyday issues, and dilemmas for hours.

Stamatis and the Pargianas brothers, **Kostas** and **George**, thank you for the tsipouro and seafood dinners that I always make a priority whenever I return to Thessaloniki. **Grigoris**, thank you for helping me see things from a different, more realistic perspective and for setting a great example through your way of thinking and the career you have built.

Giannis X. and **Elpida**, thank you for being such amazing hosts and for making me feel like a king whenever I return. **Nikos L.**, thank you for being the best BBQ host. One of the main reasons I look forward to returning for holidays is to enjoy those

wonderful gatherings at your place. Cheers as well to your wife, **Marina**, and your daughter **Faidra**, for never kicking us out, even after a 12-hour BBQ.

To my cousin **Nikos P.**, thank you for checking on me abroad, for exchanging song lyrics, and for revisiting funny memories from the past. Cheers to **Sia** as well. To my cousin **Vangelis**, his wife **Polina**, my nephew **Theo**, and my niece **Dimitra**, thank you for hosting some wonderful weekends in Düsseldorf. I treasure those weekends and the time we spend together.

To everyone whose name is not written here but whose support helped me reach this point, thank you. I truly appreciate each and every one of you.

Iosif, my best friend since childhood, thank you for always being there for me during difficult times and for listening whenever I needed to talk things through. **Tasos**, one of my best friends, thank you for sharing both the funniest moments and the hard times with me. You are the reason I visited Ireland countless times.

Another special thank you goes to **Peggy**. Without you, I do not think I would have been able to earn this PhD. You truly deserve half of it. You supported me throughout the entire process, even before I came to the Netherlands. Moving to Athens with you to chase my dreams outside my hometown, Thessaloniki, and pursue opportunities abroad was a pivotal step. It may be bittersweet to admit it, but you have been the most influential person in my adult life so far. Thank you for everything.

Finally, a big thank you to my family, who supported me throughout this journey. Living abroad meant that most of our time together was during holidays and visits home, but I could still rely on you, especially during stressful moments. To my brother **Vangelis**, his wife **Melanie**, and their beloved family, thank you for the joyful moments we shared during those visits. I always looked forward to spending time with you and playing with my nieces **Adriana** and, more recently, her younger sister **Ioanna**. To my mother, **Marina**, thank you for always being there and for supporting me in every way you could. I also thank my father, who passed away far too young but shaped who I am and continues to inspire me to strive for excellence. His guidance remains with me whenever I make important decisions.

About the author



George Pampoukis was born on 16 August 1996 in Thessaloniki, Greece. He studied Agriculture and Food Science at Aristotle University of Thessaloniki (integrated BSc/MSc, 2014-2019), where his thesis work was carried out in the Laboratory of Food Engineering and Processing on the topic: "Effect of olive leaf extract on the physicochemical properties of fresh pork", supervised by Eugenios Katsanidis. During his studies he completed an internship at the Heineken Company (Malting and Brewing Departments, Thessaloniki plant, July-August 2018), supporting quality control and process monitoring. He then completed a second MSc in Food Quality Control and Food Safety at the Agricultural University of Athens (2019-2021), with a thesis titled: "Rapid methods for foodborne pathogens detection and quantification: a review", supervised by George-John Nychas. He also completed an internship at the Institute of Technology of Agricultural Products (ELGO-DEMETER, September-October 2020), working on microbiological analyses and food safety evaluation, and worked as a Research Fellow at the Agricultural University of Athens (April-September 2021, Laboratory of Food Microbiology and Biotechnology), contributing to EU Horizon 2020 projects on rapid microbiological assessment to map spoilage and pathogen growth. From 2021 to 2025 he was a PhD candidate at Wageningen University & Research, Laboratory of Food Microbiology, and a Marie Skłodowska-Curie Actions (MSCA) Early Stage Researcher under the EU Horizon 2020 programme, grant agreement No. 955431, where he quantified the efficacy of mild food processing technologies using data-driven approaches and developed databases, meta-regression models, and decision tools. Since December 2025, he has been a Postdoctoral Researcher in Machine Learning Operations for Food Science at ETH Zurich, Laboratory of Food Structure Engineering, working on data and models for solid-state fermentation to improve the nutritional value of plant-based products.

Mail: giorgospampoukis@gmail.com

LinkedIn: linkedin.com/in/george-pampoukis or scan the QR code

Overview of completed training activities

Discipline specific activities

Name of the course/meeting	Organizing institute(s)	City	Year
Course: Ecophysiology of food-associated micro-organisms: Roles in health and disease	VLAG	Wageningen (NL)	2021
Secondment – Cooperation on modelling (2 months)	University of Athens	Athens (GR)	2023
Workshop 1 & conference: State-of-the-art of novel food processing technologies	University of Reading	Dublin (IE)	2022
Workshop 2 & conference: Quantitative Microbial Risk Assessments	Wageningen University, Arla	Aberdeen (SC)	2022
Workshop 3: Consumer and risk perception	University of Malta, Sense Test	Ghent (BE)	2024
Workshop 4: Commercialisation of results	Wageningen University, DIL	Bruges (BE)	2024
International symposium + workshop: Novel food processing technologies	Wageningen University	Bruges (BE)	2024
Course: Quantitative Microbial Risk Assessment	University of Florida	Online	2023
IAFP European Symposium 2023	IAFP	Aberdeen (SC)	2023
EFFoST International Conference	EFFoST	Dublin (IE)	2022
Mikrobiokosmos conference	MBK	Athens (GR)	2021
FS4M Seminar 1 "Identification of Mycotoxins"	Inst. of Tech. of Agric. Products	Athens (GR)	2023
Data-analysis workshop	University of Athens	Athens (GR)	2023
Sustainable preservation of marine biomasses …	NTUA	Athens (GR)	2023
IAFP World Food Safety Day + GoToWebinar seminar	IAFP	Online	2023
EXCEL4MED Living-Lab event	Agricultural Univ. of Athens	Athens (GR)	2023
Training school – Modelling low-pH & other stresses	University of Athens	Athens (GR)	2023
9th PEF school	BOKU	Vienna (AT)	2023
KNVM Symposium Food Safety Risks of our Future Food	Wageningen University	Wageningen (NL)	2023
Eurosis conference FOODSIM'2024	KU Leuven	Ghent (BE)	2024
AFSG 4th Food Sciences symposium	AFSG	Wageningen (NL)	2025
13th ICPMF International Conference on Predictive Modelling in Food	ICPMF	Athens (GR)	2025

General courses

Name of the course/meeting	Organizing institute(s)	City	Year
VLAG PhD week	VLAG	Baarlo (NL)	2022
Summer School 1: Managing scientific research	Wageningen University, Unilever	Wageningen (NL)	2021
Summer School 2: Development of business skills	University of Malta, ELEA	Valletta (MT)	2022
Summer School 3: Communication skills	University of Reading, Sense Test	Reading (UK)	2023
Applied statistics	VLAG	Wageningen (NL)	2024
Career orientation	EPS	Wageningen (NL)	2025

Other activities

Name of the course/meeting	Organizing institute(s)	City	Year
Preparation of research proposal	FHM	Wageningen (NL)	2021
PhD study tour to Germany and Switzerland (incl. organisation)	FHM	DE & CH	2022
PhD study tour to Germany and Denmark (incl. organisation)	FHM	DE & DK	2025
Network meetings	TRANSIT consortium	Multiple	2021 – 2024
Department seminars	FHM	Wageningen (NL)	2021 – 2025
Organisation of a scientific event: "Application of PEF to sustainably produced foods: A scientific, technical, and personal perspective"	Wageningen University	Wageningen (NL)	2024
Review of scientific articles	IJFM, JFP	Multiple	2021 – 2025

Authorship statement

PhD candidate's name : George Pampoukis

First promotor : Prof. Heidy M.W. den Besten

Promotor: Prof. Marcel H. Zwietering

Title of PhD thesis : From data to insight: Assessing the decontamination efficacy of mild food processing technologies using data-driven approaches

Date of public defence : 16-10-2026 15:30

Chapter 1: General Introduction
The general research question and the aims of this thesis were defined together with my promotors. I defined the scope of the work, positioned non-thermal plasma (NTP), pulsed electric fields (PEF), and high-pressure processing (HPP) in the broader food-safety and sustainability context, and wrote and revised this chapter following comments from my promotors.

Chapter 2: Ranking factors affecting the decontamination efficacy of non-thermal plasma: The approach of dissipated power per plasma volume through machine learning modeling
The research question and study design were developed jointly with my promotors. I defined the eligibility criteria, implemented the active-learning workflow for literature screening, curated the NTP decontamination database, developed the exploratory data analysis and machine-learning workflow, interpreted the results, and drafted and revised the manuscript text. A BSc thesis contributed to the initial literature search and data extraction under my guidance. This chapter has been published in a peer-reviewed journal.

Chapter 3: A metadata schema to standardize non-thermal plasma decontamination parameters in food-related applications
The aims of this work were discussed with my promotors and international collaborators. I led the schema design, prepared the examples and documentation, and coordinated the integration with data-management tools in close collaboration with Thomas Weihe, Robert Wagner, Markus M. Becker, and Uta Schnabel. Robert Wagner contributed to software and visualization, and Masja N. Nierop Groot and Uta Schnabel to supervision and project coordination. I prepared the first manuscript draft and revised subsequent versions after feedback from my co-authors and promotors. This chapter has been published in a peer-reviewed journal.

Chapter 4: The decontamination efficacy of pulsed electric field in food: Establishing global kinetic parameters
The research question and modeling strategy were defined in consultation with my promotors. Together with Jan Ham, in the context of his MSc thesis project and a second MSc thesis project, I assembled and harmonized the PEF decontamination database; Jan Ham also contributed to the investigation, validation, formal analysis, and the interactive dashboard. I designed the exploratory data analysis, specified and fitted the meta-regression models, quantified the kinetic parameters, and drafted and revised the manuscript based on input of all co-authors. This chapter has been published in a peer-reviewed journal.

Chapter 5: High-pressure processing inactivation of foodborne bacteria and viruses in foods
The research questions were formulated together with my promotors and co-authors. Together with Kah Yen Claire Yeak and Cristina Serra-Castelló, we updated and extended the HPP decontamination database. Part of the data extraction was contributed through an MSc thesis project. The co-authors contributed to defining the modeling approach, and Cristina Serra-Castelló initiated the modeling procedures. This chapter was developed in close collaboration with Kah Yen Claire Yeak, with whom I share equal contribution to the underlying publication. I continued and implemented the data-wrangling pipelines, performed the meta-regression analyses, derived hazard-resistance rankings, and drafted the manuscript text together with Kah Yen Claire Yeak and Cristina Serra-Castelló. The analytical approach, interpretation, and manuscript were further refined through discussions and feedback from all co-authors.

Chapter 6: General Discussion
I proposed the outline of the general discussion, integrating the findings from Chapters 2, 4, and 5 with thermal-processing benchmarks and sustainability considerations. Sustainability indicators used in this chapter were collected from the literature with support from an MSc thesis project. I performed the additional comparative analyses (including the matched-efficacy case study for high-acid fruit juice and *E. coli*), interpreted the implications for food safety management, sustainability assessments, and future research on mild processing technologies, and wrote the first draft of this chapter. I revised the chapter and finalized the text after feedback from my promotors.

Date: 27-03-2026

Signature PhD candidate



Signature promotor for agreement



The research described in this thesis was performed at the Laboratory of Food Microbiology, Wageningen University & Research. George Pampoukis was part of the TRANSIT project that received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement N° 955431.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Matteo Gallo, based on a concept by George Pampoukis.

Thesis layout by George Pampoukis.

Printed by ProefschriftMaken.