****

TRANSIT

**Deliverable D10.2**

Predictive models on decontamination treatment efficiency

Work package 10

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement TRANSIT No 955431

[](https://ec.europa.eu/programmes/horizon2020/en)

Project Acronym: TRANSIT

Project full title: Training Network Sustainable Technologies

Call: H2020-MSCA-ITN-2020

Type of Action: MSCA-ITN-ETN – European Training Network

Grant Number: 955431

Project website: www.transit-itn.com

|  |  |
| --- | --- |
| Deliverable number | D10.2 |
| Deliverable title: | Predictive models on decontamination treatment efficiency |
| Deliverable description: | Predictive models on decontamination treatment efficiency of HHP, PEF, US, NTP |
| Deliverable nature: | Report |
| Dissemination level: | Public |
| Work package | WP10 |
| Author: | George Pampoukis – WU |
| Lead Beneficiary: | WU |
| Contract Delivery date: | 31 August 2024 |
| Number of pages: | 20 |
| Peer review: | Prof. Heidy M.W. den Besten, Prof. Marcel H. Zwietering |

**1. Background**

Non-thermal food processing technologies such as high-pressure processing (HPP), pulsed electric field (PEF), non-thermal plasma (NTP), and ultrasounds (US) have gained attention for their ability to enhance food safety, control food quality, maintain the nutritional value, and potentially reduce the energy use compared to conventional pasteurization methods. As described in detail in deliverable D10.1, WP10 collected literature data and developed four structural databases (one for each technology) focusing on the microbial inactivation kinetics of the model organisms *Escherichia coli*, *Listeria monocytogenes*, *Lactiplantibacillus plantarum*, *Bacillus subtilis*, and *Saccharomyces cerevisiae*. For some technologies additional microorganisms were included in the databases, to provide a more complete view, as explained below. For HPP, the database developed was merged with a database that was developed within the framework of the EU project SAFFI (<https://www.saffi.eu/>) to increase the impact of both projects. The microbial inactivation was expressed in the *D*-value (the time needed for tenfold reduction of the microbial population) ([Equation 1](#eq-primary)) and the *z*-value (the change of the independent variable of interest needed for tenfold reduction of the *D*-value) ([Equation 2](#eq-secondary)) (Van Asselt & Zwietering, 2006). In addition, the processing-, food-, and microbial-related parameters were compiled for each *D*-value (min) for further predictive model development in this deliverable 10.2.

where (log10 cfu/g (or ml)) is the initial microbial population, and (log10 cfu/g (or ml)) is the microbial population after processing time *t* (min).

Below is an example equation for the effect of temperature (independent processing variable of interest) on the *D*-value with a log-linear secondary model:

where is the *D*-value at a reference temperature .

**2. Materials and methods**

**2.1 Exploratory data analysis**

The first step of the data analysis procedure was to establish a metric for comparing the decontamination efficacy and to perform an exploratory data analysis. The metric for HPP, NTP, and US was the log10 transform of the inactivation kinetics parameter i.e., log10*D*. For PEF, the microbial reductions were directly used as the dependent variable instead, since the treatment time is in the range of milli/micro-seconds which would make the comparisons more difficult. Therefore, the *D*-value was expressed in energy input units (kJ/L) instead of the processing time ([Equation 3](#eq-pef)).

where is the PEF energy input in kJ/L, and is the energy input (kJ/L) needed for tenfold reduction of the microbial population.

The exploratory data analysis was performed using Python programming language (v3.11) (Van Rossum & 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). The numerical variables were checked for their correlation with the log10*D* (log10 min) (or the microbial log10 reductions (-), for PEF) using the Spearman’s rank correlation coefficients to evaluate their potential as predictors in the microbial inactivation models. 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’ log10*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 log10*D*.

**2.2 Linear models**

Except for the statistical analysis that was explained above, all linear secondary models were fitted for all numerical parameters of interest and the corresponding z-values were estimated ([Equation 2](#eq-secondary)). For example, for HPP the main numerical variable of interest was the applied pressure, so the pressure needed to reduce the log10*D* by 1 (tenfold reduction of the *D*-value) was estimated, assuming log-linear inactivation kinetics. The equation to describe this was [Equation 2](#eq-secondary), where instead of and , and were used. For PEF, a *z* value was not established since the dependent variable is the microbial log10 reductions and not the log10*D*. Therefore, the equivalent of the *D*-value for the numerical parameter of interest was estimated instead. The exploratory data analysis and the linear model fitting were conducted on the full database for each technology to gain generic insights into the efficacy of each technology across all microbial genera and treatment matrices. Additionally, they were conducted at different levels of “zooming in” such as focusing on specific microbial genera, specific matrix categories, and, where sufficient data were available, specific microbial genera within specific matrix categories.

**2.3 Multiple linear regression models**

To account for multiple numerical and categorical parameters, multiple linear regression models were fitted for US and NTP. This approach was necessary for these technologies because the variance of the log10*D* could not be sufficiently explained by a single main processing parameter, even after filtering the data to separate the effects of microbial genus and matrix. These models utilized a limited set of parameters that were previously identified during the exploratory data analysis. The numerical variables were 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 4](#eq-scaling)).

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. 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).

**2.4 Machine learning models**

For NTP, in addition to the multiple linear regression model that was developed, to unveil more complex relationships between the dependent and the independent variables, an extreme gradient boosting model was also developed. Gradient boosting is 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). The algorithm works similarly to random forests, where tree predictors are combined in such a way 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). This means that we compose our model based on multiple decision trees. 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. In gradient boosting the ensemble learning is sequential and can formulate additive expansions for any fitting criterion (Friedman, 2001). For our purpose, the 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 & Guestrin, 2016). 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). The hyperparameters that were not tuned, were set to their default values, as determined by the “Scikit-learn” software that was used for all machine learning approaches (Pedregosa et al., 2011). 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 & 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 (i.e., higher log10*D*), whereas the dots on the left push it to the negative correlations with the independent variable (i.e., lower log10*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.

**3. Results and discussion**

**3.1 HPP models**

The database which was developed in collaboration with the SAFFI project consisted of 3890 *D*-values for 22 microbial hazards. The dataset was then filtered for vegetative cells and non-thermal processing temperatures (≤ 45 ºC) (n=3450) because it was found that the inactivation of spores was primarily heat-dependent, and thus only possible when combined with thermal processing, which was out of the current scope. Based on the exploratory data analysis, five groups of microorganisms were formed, namely, i., sensitive (*Toxoplasma*, *Cyclospora*, *Cryptosporidium*, *Trichinella*, *Rotavirus*, *Aeromonas*, *Cronobacter*, *Campylobacter*): < 1 min; ii., moderately resistant (Norovirus, Hepatitis A, *Listeria*, *Salmonella*, *Yersinia*): 1 min ≤ < 3 min; iii., resistant (Hepatitis E, vegetative *B. cereus*, *S. aureus*, *E. coli*): 3 min ≤ < 5 min; iv., highly resistant (*Mycobacterium*, *Shigella*): ≥ 5 min; and v., spores (*B. cereus*, *C. botulinum*, *C. perfringens*): no inactivation at 400 MPa. [Table 1](#tbl-hpp-micro) provides an overview of the expected *D*-values in a variety of microorganisms using all data available (n=3450). However, for some microorganisms e.g., *Toxoplasma gondii* the data were limited (n=5) which leads to no trends or trends opposite to those expected for the effect of the pressure ([Table 1](#tbl-hpp-micro)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: The decimal reduction time (*D*-value) for 3 different reference pressure levels for merged microbial data from TRANSIT and SAFFI projects. The dataset was filtered for vegetative cells and non-thermal processing temperatures (≤ 45 ºC) (n= 3450). The parameter is the pressure increase needed to reduce tenfold the *D*-value and was estimated as the negative inverse of the slope of the linear regression model. “NA” values indicate that fitting a regression line was not possible. The results are in ascending order based on the   | Microorganism | *D*400MPa (min) | *D*500MPa (min) | *D*600MPa (min) | *zp* ± SE*zP* (MPa) | n | | --- | --- | --- | --- | --- | --- | | *Toxoplasma gondii* | 0.24 | 0.92 | 3.5 | -173 ± 56a | 5 | | *Vibrio* spp. | 0.28 | 0.14\* | 0.072\* | 342 ± 32 | 259 | | Rotavirus | 0.50 | 0.32 | 0.19\* | 499 ± 49 | 104 | | *Aeromonas caviae* | 0.56\* | 0.18\* | 0.063\* | 205 ± 23 | 18 | | *Cryptosporidium* spp. | NA | NA | NA | NA | 1 | | *Cyclospora cayetanensis* | NA | NA | NA | NA | 2 | | *Trichinella* spp. | 0.62 | 0.29 | 0.13 | 303 ± 20 | 3 | | *Cronobacter* spp. | 0.72 | 0.35 | 0.17 | 317 ± 25 | 65 | | *Campylobacter* spp. | 0.81 | 0.23\* | 0.068\* | 185 ± 23 | 75 | | Norovirus | 1.1 | 0.89 | 0.69 | 924 ± 177 | 363 | | Hepatitis A | 1.5 | 1.02 | 0.71 | 623 ± 184 | 119 | | *Listeria monocytogenes* | 2.3 | 1.4 | 0.79 | 434 ± 22 | 752 | | *Salmonella enterica* non-Typhi | 2.3 | 1.8 | 1.4 | 951 ± 140 | 385 | | *Yersinia enterocolitica* | 2.8 | 1.99 | 1.5\* | 740 ± 273 | 89 | | Hepatitis E | 3.02 | 2.2 | 1.6 | 747 ± 385a | 36 | | *Bacillus cereus* | 3.6 | 6.03 | 10\* | -457 ± 189 | 25 | | *Staphylococcus aureus* | 3.9 | 3.2 | 2.6 | 1182 ± 260 | 339 | | *Escherichia coli* | 3.9 | 3.1 | 2.5 | 963 ± 146 | 758 | | *Mycobacterium tuberculosis* | 5.1 | 1.9 | 0.66 | 225 ± 27 | 44 | | *Shigella* spp. | 7.4 | 6.6 | 6.03 | 2079 ± 2254a | 8 | |

a: The corresponding slope was not statistically significant (p-value ≥ 0.05)

\*: Extrapolated value

To achieve a more precise estimation of the kinetic parameters while accounting for the food matrix effect, the same approach was applied, this time zooming on different food categories where sufficient *D*-values per organisms were available. An example of this approach for modeling *E. coli* and *Listeria monocytogenes* can be found In [Table 2](#tbl-hpp-meat-fruit). These two microorganisms were selected because *E. coli* is a Gram-negative bacterium that includes pathogenic serotypes, while *Listeria monocytogenes* is a Gram-positive environmental pathogen known for its ability to survive in ready-to-eat food products and cause foodborne disease with a high case fatality (EFSA, 2023). For this table only the *D*-value at 500 MPa was estimated, but a (log10) *D*-value in any other pressure level can be estimated using [Equation 2](#eq-secondary), and replacing and , and , with , , and instead.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: The decimal reduction time (*D*-value) at 500 MPa when zooming in at two microorganisms and two different matrix categories   | Microorganism | Matrix category | *D*500MPa (min) | *zP* ± SE*zP* (MPa) | n | | --- | --- | --- | --- | --- | | *E. coli* | meat & meat products | 1.5 | 389 ± 41 | 253 | | *E. coli* | fruit juice | 0.62 | 304 ± 54 | 44 | | *L. monocytogenes* | meat & meat products | 1.2 | 356 ± 22 | 243 | | *L. monocytogenes* | fruit juice | 1.1 | 444 ± 157 | 33 | |

Additionally, with sufficient data present, we could zoom in further to better standardize the effect of the food product on HPP inactivation. An example of this modeling approach can be found in [Table 3](#tbl-fruit). However, it should be noted that as we zoom in, the dataset per category also becomes more limited.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Decimal reduction time (*D*-value) at 500 MPa for *E. coli* in fruit juices of different acidity. This table categorizes fruit juices into high acidic pH ≤ 4.6 and low acidic pH > 4.6 groups.   | Microorganism | Matrix category | *D*500MPa (min) | *zP* ± SE*zP* (MPa) | n | | --- | --- | --- | --- | --- | | *E. coli* | high-acidic fruit juice | 0.48 | 270 ± 45 | 41 | | *E. coli* | low-acidic fruit juice | 2.01 | 464 ± 10 | 3 | |

**3.2 PEF models**

The approach of fitting linear models and zooming into different matrix categories with sufficient data was also performed for PEF. In addition to the five model microorganisms of TRANSIT (i., *Escherichia coli*; ii., *Listeria monocytogenes*; iii., *Lactiplantibacillus plantarum;* iv. *Bacillus cereus*; v., *Saccharomyces cerevisiae*), the search was expanded to additional microorganisms due to the abundance of research articles exploring PEF microbial inactivation for these microorganisms. Therefore, data were also collected for all lactic acid bacteria, all *Bacillus* and *Listeria* species, as well as data for *Pseudomonas* spp., *Staphylococcus aureus*, *Salmonella* spp., and *Pseudomonas* spp. For PEF the microbial reduction was the dependent variable, and the energy input was the independent variable [Equation 3](#eq-pef). Only continuous systems and devices featuring square wave pulses were used, because they are more relevant for industrial applications (De Haan & Willcock, 2002). The energy input used to achieve microbial reductions in the various studies was estimated based on the reported information using the [Equation 5](#eq-energy-input).

where is the specific energy input in J/ml, is the electric field strength in kV/cm, is the pulse frequency in pulses per second (pps), is the pulse width in microseconds (μs), is the flow rate in cubic centimeters per second (cm³/s), is the volume of the chamber in cubic centimeters (cm³), is the number of treatment chambers, and is the medium conductivity in Siemens per centimeter (S/cm). The final dataset consisted only of articles that reported the energy input equation components described above. Regarding conductivity, because it is highly affected by the treatment temperature, only articles that reported conductivity at the temperature before treatment were used. The dataset was filtered regarding the maximum energy input (), and only studies with values below or equal to 300 kJ/L were used for the development of the models. This was done to provide results that are relevant for industrial purposes and to avoid the thermal inactivation that a higher energy level would pose. According to Aganovic & Smetana (2022), the energy input levels applied for the treatment of heat-sensitive liquids are usually between 80-120 kJ/kg. In [Table 4](#tbl-pef-micro), an example of a way to quantitatively look at the literature is presented (n=463). For this table, the microbial log10 reductions at 80, 120, and 180 kJ/L were estimated, but the microbial log10 reductions in any other energy input level can be estimated using the [Equation 3](#eq-pef). For bacterial spores, the inactivation was due to thermal effects and with energy input above 300 kJ/L. For *Pseudomonas* spp. the limited data available could not provide a reliable estimation of the .

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: The overview of the microbial log10 reductions achieved using PEF. The dataset is filtered for energy input entries below 300 kJ/L (n= 463). The parameter is the energy input needed to reduce the microbial population tenfold and was estimated as the inverse of the slope of the linear regression model. The results are in descending order based on the microbial log10 reductions at   | Microorganism | *Ein* (80 kJ/L) | *Ein* (120 kJ/L) | *Ein* (180 kJ/L) | *DEin* ± SE*Ein* (kJ/L) | n | | --- | --- | --- | --- | --- | --- | | *Saccharomyces cerevisiae* | 3.6 | 3.8 | 4.2 | 157 ± 104a | 80 | | *Salmonella* spp. | 2.6 | 3.5 | 5.0 | 41 ± 7.4 | 26 | | *S. aureus* | 2.5 | 3.2 | 4.3 | 54 ± 5.5 | 104 | | *E. coli* | 2.3 | 2.7 | 3.02 | 58 ± 6.5 | 125 | | Lactic acid bacteria | 2.3 | 2.4 | 2.6 | 390 ± 165 | 107 | | *Listeria* spp. | 1.6 | 2.06 | 2.7 | 95 ± 39 | 16 | | *Pseudomonas* spp. | 0.9 | 0.8\* | 0.6\* | -373 ± 643a | 5 | |

a: The corresponding slope was not statistically significant (p-value ≥ 0.05)

\*: Extrapolated value

As we did before for HPP, we could zoom in further for PEF to obtain more precise results for a specific food category, for example, we could zoom in on fruit juices and *E. coli* ([Table 5](#tbl-pef-fruit)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5: The microbial log10 reductions at 120 kJ/L when zooming in *E. coli* and the matrix category of fruit juice. The dataset is filtered for energy input entries below 300 kJ/L.   | Microorganism | Matrix category | *Ein* (120 kJ/L) | *DEin* ± SE*Ein* (kJ/L) | n | | --- | --- | --- | --- | --- | | *E. coli* | high-acidic fruit juice | 3.7 | 40 ± 6.6 | 36 | | *E. coli* | low-acidic fruit juice | 1.9 | 106 ± 22 | 44 | |

**3.3 Ultrasound models**

For ultrasounds, only data for the probe systems were analyzed. The ultrasonic intensity was used as the main parameter determining the decontamination efficacy (see [Equation 6](#eq-us-power) & [Equation 7](#eq-us-ui)).

where is the absolute ultrasonic power (W), m is the mass (kg), is the specific heat capacity (J/(kg·°C)) and *dT*/*dt* is the initial rate of change of temperature during sonication (°C/s) (Tiwari & Mason, 2012).

where is the ultrasonic intensity (W/cm2), and is the diameter of the emitter (probe) (cm) (Tiwari & Mason, 2012). In the database developed, the ultrasonic intensity could not be estimated based on the components of [Equation 7](#eq-us-ui) (to ensure consistency between the different studies, as we did for PEF), and thus the database was standardized differently by including only articles with reported ultrasonic intensity values. The filtered dataset consisted of 40 log10*D*-values for *E. coli*, *Listeria* spp., *Bacillus*, and *Salmonella* from a total of 12 articles. For *Lactiplantibacillus plantarum* (or any other lactic acid bacteria) and *Saccharomyces cerevisiae,* there were no data available to report the parameter. From the statistical summaries of the exploratory data analysis, it was found that the maximum was 986 W/cm2, while the 75th percentile was 401.5. Based on this insight the results were analyzed for all microorganisms in three ways i.e., i., the full dataset; ii., the data with ≤ 700 W/cm2; iii., the data with ≤ 500 W/cm2 ([Table 6](#tbl-us-overall)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 6: The decimal reduction time (*D*-value) for 3 different reference ultrasonic intensity levels. The dataset is filtered in three different ways (data frames). The parameter is the ultrasonic intensity increase needed to reduce tenfold the *D*-value and was estimated as the negative inverse of the slope of the linear regression model   | Data frame | *D*100W/cm2 (min) | *D*200W/cm2 (min) | *D*400W/cm2 (min) | *z*UI ± SE*zUI* (W/cm2) | n | | --- | --- | --- | --- | --- | --- | | full dataset | 9.3 | 7.8 | 5.5 | 1304 ± 601 | 40 | | ≤ 700 W/cm2 | 9.9 | 7.2 | 3.8 | 712 ± 242 | 38 | | ≤ 500 W/cm2 | 9.9 | 6.3 | 2.5 | 504 ± 137 | 36 | |

In [Table 6](#tbl-us-overall) it is shown that the linear model was highly affected by 4 values above 500 W/cm2 leading to high values, which are not clear if they were due to true effects of the high values or due to a miscalculation of these values in the corresponding articles. Therefore, these 4 values were excluded and the analysis proceeded with the ≤ 500 W/cm2 data frame (n=36). The numerical parameters with the highest Spearman’s rank correlation coefficients (|ρ| ≥ 0.6) with log10*D* were the processing temperature (inlet and outlet temperature), and the ultrasonic intensity. For ultrasounds, the effect of temperature was important not only for spores but also could not be omitted for the vegetative cells ([Table 7](#tbl-us-genus)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7: The decimal reduction time (*D*-value) for a reference ultrasonic intensity level and a reference temperature level. The *D*-value at 40 °C should not be confused with thermal processing inactivation since it refers to the effect of in the context of ultrasound inactivation experiments. The dataset is filtered for ≤ 500 W/cm2. “NA” values indicate that fitting a regression line was not possible   | Microorganism | *D*100W/cm2 (min) | *D*40°C (min) | *z*UI ± SE*zUI* (W/cm2) | *z*Tin ± SE*zTin* (°C) | n | | --- | --- | --- | --- | --- | --- | | *Bacillus* spp.b | 56 | 11 | 253 ± 40 | 29 ± 3.7 | 9 | | *E. coli* | 11 | 1.5\* | 626 ± 576a | 22 ± 6.3 | 14 | | *Listeria* spp. | 2.4 | 3.4 | 177 ± 15 | -83 ± 50a | 7 | | *Salmonella* spp. | 6.1 | NA | 944 ± 443a | NA | 6 | |

\*: Extrapolated value

a: The corresponding slope was not statistically significant (p-value ≥ 0.05)

b: For *Bacillus* spp. both vegetative and spore cells are reported

Due to the significant effect of temperature and the limited dataset available for zooming in on different matrix categories, fitting separate models was challenging. Furthermore, because multiple microorganisms and matrix categories were involved, a multiple linear regression model incorporating all relevant parameters was fitted instead. The model was developed using all 33 *D*-values (for 3 entries was not reported). The parameters used were: i., (W/cm2); ii., (°C); iii., microbial genus (*Bacillus*, *E. coli*, *Listeria*, *Salmonella*) iv., matrix category (a., buffer and growth media; b., meat, egg and dairy, c., fruit and vegetables). In order to be included in the multiple linear regression, the microbial genus and matrix categories were converted to dummy variables, and the most frequent categories were set as bases (0) i.e., *Escherichia* and buffer and growth media, respectively. The numerical parameters were scaled using [Equation 4](#eq-scaling). Backward elimination was implemented to avoid over-fitting, and parameters’ coefficients with p-value > 0.05 were excluded, namely, the microbial genera of *Listeria* and *Salmonella*, and the matrix categories of meat, egg, and dairy, and fruit and vegetables. The multiple linear regression model is presented in [Equation 8](#eq-us-mlr) (scaled model):

**3.4 NTP models**

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 non-thermal plasma. 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). The overview of *D*-value entries for all setups can be found in [Figure 1](#fig-violin).

|  |
| --- |
| Figure 1: 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 log10*D* for each type of equipment. |

The overview of the microbial reductions achieved with NTP against various microorganisms and matrices is shown in [Table 8](#tbl-qualitative-overview). 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 log10 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 log10. 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 log10, 20.1% (71 of 352) of the total entries led to an inactivation ≥ 5 log10 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 8: Overview of microbial log10 reductions (-) across different matrix categories   | Matrix category | *Listeria* spp.a, b | *Salmonella* spp.a, b | *Bacillus* spp.a, c | Coliformsa | Yeastsa | | --- | --- | --- | --- | --- | --- | | 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 ± stdev

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 comprehensive parameter present in all studies, that is equivalent for example to temperature for thermal processing or pressure for HPP to standardize the results. Also, a parameter such as the energy input used for PEF or the ultrasonic intensity for US did not exist. We focused on the DBD setup because according to the findings of this study ([Figure 1](#fig-violin)) 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/cm3), which was assumed to provide an integrated descriptor of efficacy using the equations [Equation 9](#eq-power) and [Equation 10](#eq-powerdens). This concept is not so easily transferable to APPJ or other setups.

where *U* is the voltage (in volts, V), *I* is the current (in amperes, A), and (-) 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 10](#eq-powerdens)).

where *A* is the powered electrode area (in cm2) and *d* is the distance (in cm) between the powered and ground electrode.

From all database entries, 160 data points corresponded to the DBD setup, and the concept of power density, as explained above, could be implemented. However, the data were not sufficient to zoom in on certain matrix categories and microbial genera and fit separate, specific models.

Therefore, multiple linear regression was performed for the parameters power density (W/cm3), sample area (cm2), 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 (0) for all regression analyses. Using all data points (n=160) and backward elimination with scaled variables, the matrix category of solid medium, the genus of *Escherichia*, the peak-to-peak voltage and the relative humidity were excluded as non-significant variables. The results also suggested that the upper electrode shape, the matrix category, the sample area, and the power density significantly affected the log10*D*, next to the pH and the microbial genus. The multiple linear regression model is presented in [Equation 11](#eq-ntp-mlr) (scaled model).

To unveil more complex relationships between the dependent and the independent variables, eXtreme Gradient Boosting (XGBoost) was used. 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 (ntrain=128, ntest=32), then the coefficients with scaling were slightly different but with the same order as with the full dataset ([Figure 2](#fig-shap)). The on the test dataset of the XGBoost model was 0.71. 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 interpretation of the results. Based on the mean absolute SHAP values in the test dataset ([Figure 2](#fig-shap)), 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 log10*D*, which were also significant parameters of the multiple linear regression. The pH and the microbial genus of *Listeria* which were also used in the multiple linear regression model ([Equation 11](#eq-ntp-mlr)), 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 log10*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*, being more similar to *Salmonella*, 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 log10*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](#fig-shap)). 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. Using the same hyperparameters, 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 log10*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. The power density and matrix category, followed by microbial genus and pH were the most important factors affecting the decontamination efficacy of non-thermal plasma (Pampoukis et al., 2024).

|  |
| --- |
| Figure 2: 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. 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 log10*D* (log10 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 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. For dummy variables, red corresponds to the presence of the effect, and blue to the absence of the effect. |

**4. Conclusion**

The models developed under WP10 utilized all literature data that met with the set quality criteria to describe the expected decontamination efficacy of each technology and identify the most important sources of variability. However, not all of the technologies had the same amount of information available in the literature, nor were they equally standardized. For HPP and PEF the amount of well-standardized data allowed for the development of various linear models for different food matrices, microorganisms, and their combinations. For example, these models can predict the expected decontamination efficacy for fruit juice contaminated with *E. coli*. For US and NTP, the limited availability of well-reported information yielded the need for developing a single model to describe all available data points simultaneously. This was achieved through multiple linear regression (for US and NTP), and complementary through machine learning (for NTP). The ranking of the parameters and the corresponding models developed for the different technologies could serve as valuable tools for guiding regulatory authorities and food business operators in quantitative microbial risk assessments (Den Besten & Zwietering, 2012). Additionally, they can assist industrial applications in prioritizing certain parameters over others and optimizing the procedures for maximum efficacy (Pampoukis et al., 2024).

**References**

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

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

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>

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

De Haan, S. W. H., & Willcock, P. R. (2002). Comparison of the energy performance of pulse generation circuits for PEF. *Innovative Food Science & Emerging Technologies*, *3*(4), 349–356. <https://doi.org/10.1016/S1466-8564(02)00069-3>

Den Besten, H. M. W., & Zwietering, M. H. (2012). Meta-analysis for quantitative microbiological risk assessments and benchmarking data. *Trends in Food Science & Technology*, *25*(1), 34–39. <https://doi.org/10.1016/j.tifs.2011.12.004>

Dietterich, T. G. (2002). Ensemble learning. *The Handbook of Brain Theory and Neural Networks*, *2*(1), 110–125.

EFSA. (2023). Prolonged multi-country cluster of *Listeria monocytogenes* ST155 infections linked to ready-to-eat fish products. *EFSA Supporting Publications*, *20*(12), 8538E. https://doi.org/<https://doi.org/10.2903/sp.efsa.2023.EN-8538>

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. <https://www.jstor.org/stable/2699986>

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*(15), 2865–2873. <https://doi.org/10.1002/sim.3107>

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.” *Journal of Physics D: Applied Physics*, *49*(8), 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 Science and Technology*, *28*(9), 3–4. <https://doi.org/10.1088/1361-6595/ab393d>

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>

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>

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, p. 9). Curran Associates, Inc. <https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>

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

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. *Innovative Food Science & Emerging Technologies*, *96*, 103773. <https://doi.org/10.1016/j.ifset.2024.103773>

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. *Journal of Machine Learning Research*, *12*, 2825–2830.

Suits, D. B. (1957). Use of dummy variables in regression equations. *Journal of the American Statistical Association*, *52*(280), 548–551. <https://doi.org/10.2307/2281705>

Tiwari, B. K., & Mason, T. J. (2012). Chapter 6 - Ultrasound Processing of Fluid Foods. In P. J. Cullen, B. K. Tiwari, & V. P. Valdramidis (Eds.), *Novel Thermal and Non-Thermal Technologies for Fluid Foods* (pp. 135–165). Academic Press. <https://doi.org/10.1016/B978-0-12-381470-8.00006-2>

Van Asselt, E. D., & Zwietering, M. H. (2006). A systematic approach to determine global thermal inactivation parameters for various food pathogens. *International Journal of Food Microbiology*, *107*(1), 73–82. <https://doi.org/10.1016/j.ijfoodmicro.2005.08.014>

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. van der, Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 1–4. <https://doi.org/10.21105/joss.03021>

Yan, D., Sherman, J. H., & Keidar, M. (2016). Cold atmospheric plasma, a novel promising anti-cancer treatment modality. *Oncotarget*, *8*(9), 15977–15995. <https://doi.org/10.18632/oncotarget.13304>