

Article

A Comparative Study of Reliability Screening Predictions in Multifactorial DOE Industrial Trials

Helen C. Sereti ¹, Panagiotis Tsarouhas ^{1,2}  and George Besseris ^{1,3,*}

¹ Quality Management and Technology Postgraduate Program, Hellenic Open University, 26335 Patras, Greece; std511dipeap@gmail.com (H.C.S.); ptsarouhas@uth.gr (P.T.)

² Department of Computer Science and Biomedical Informatics, University of Thessaly, 35131 Lamia, Greece

³ Mechanical Engineering Department, University of West Attica, 12241 Aigaleo, Greece

* Correspondence: besseris@uniwa.gr

Abstract

This paper examines the reliability of a product in terms of its lifetime, based on the investigation of certain factors within the framework of maintaining the principles of the circular economy and the sustainable design of the experiment. The study aims to investigate the critical factors that could affect the product's lifespan and long-term efficiency. Initially, a statistical analysis of interactions was achieved through an interaction diagram using the Minitab software and for trials scheduled using the Plackett–Burman planner. Subsequently, the most significant interactions were selected and further analysis was conducted using the screening model, a combination of Pareto chart and the Lenth method to filter out weaker performing controlling factors and their associated two-way interactions. This was followed by a life-data regression analysis, and the results identified statistically strong factors and interactions based on various commonly used reliability distributions. Such findings are important in improving the life extent of a product, which in the specific case was a manufactured industrial-level thermostat, while reducing the risk for early product failure. In conclusion, the study demonstrates the variability and multiplicity in prediction accuracy of the screened effects when considering different reliability distributions in research which employs statistical reliability tools. Such tools are regularly encountered in product design and improvement efforts, aiming to successfully prevent early failures while ameliorating the environmental footprint in production.

Keywords: reliability distribution; censored data; screening designs; life-data regression; experimental design; Plackett–Burman planner; Pareto analysis; Lenth test

Academic Editors: Raul D. S. G. Campilho, Nhat Luong Nhieu, Phan Van-Thanh, Hector Tibo and Chian-Nan Wang

Received: 8 June 2026

Revised: 30 July 2026

Accepted: 6 August 2026

Published: 7 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Modern industry has acquired new evolutionary significance through the gradual advancement of its production systems, utilizing advanced manufacturing structures with specialized on-demand functionalities. Beyond the specific predefined characteristics of efficiency and flexibility in application, major focal points include the use of advanced information technologies and the design of materials with high performance, suitable for demanding applications requiring increased strength and durability [1].

The automation of numerous production functions and data exchange processes has led to the Fourth Industrial Revolution, or Industry 4.0. This automation is implemented through “smart factories.” According to Sahoo and Lo [2], smart factories can operate efficiently by achieving the following steps:

- i. The collection of a wide range of relevant data from equipment.
- ii. The extraction and analysis of information that is highly useful for supporting and establishing business initiatives.
- iii. The automation of processes, systemic optimization, and the updating of business strategy.

A smart factory may therefore include:

- i. “Smart” manufacturing through the intelligent interconnection of tools, machinery, and operators.
- ii. “Smart” services that ensure real-time transparency of production and maintenance data, as well as support for a wide range of predictive and preparatory analyses, and
- iii. “Intelligent” machines within highly efficient systems, which may involve Manufacturing Execution Systems (MES), shop floors, and Material Requirements Planning (MRP) systems with reduced waste generation.

Such efficient systems include real-time control systems, which are considered preferred solutions in modern industries, as they constitute advanced systems designed with emphasis on operational correctness and extremely low response-time variation. A fundamental requirement of most machines and devices is their effective performance within an extremely short time interval from the moment stimulus data are received. The examined time parameter represents the verified interval within which each control operation is executed. All control systems operate deterministically, meaning that they complete an operation within a predefined deadline [3].

Design of Experiments (DOE) constitutes a highly efficient problem-design technique aimed at minimizing parallel data, input–output, and result interactions [4,5].

The extended lifetime of a product is highly important for the circular economy and the achievement of sustainability. The literature suggests that the strategic methods and challenges capable of overcoming the factors that render product lifetime unmanageable and short remain largely undefined [6].

The product life cycle includes stages ranging from the sourcing of raw materials, design, production, marketing, distribution, purchase, use, and disposal, to, in certain model-specific cases, collection and reuse [7].

Many models are used to describe product lifetimes, which may be described by a wide range of statistical distributions [8]. Klion [9] argued that the prediction, study, and determination of reliability are based on lifetime distribution data and, more specifically, on data derived from a particular historical background.

The issue of evaluating the effective operating time of the thermostat until the first failure is part of the Life Cycle Inventory (LCI) stage, which, together with three other stages, defines the Life Cycle Assessment (LCA). The remaining three stages are the definition of Goal and Scope (G&S), Life Cycle Impact Assessment (LCIA), and the interpretation phase. All these stages contribute to the final Life Cycle Sustainability Assessment (LCSA), which calculates and predicts consequences based on a predefined value and/or concern referred to as an Area of Protection (AoP). These consequences are modeled and ultimately lead to cause–effect inventory flows with quantified impact indicators, which may be related not only to the environment and natural resource management, but also to human health. For this reason, the methodology followed in the present study opens pathways toward improved approaches for addressing such problems and aligns with the European Union’s vision for solutions whose environmental performance is already disclosed [10].

There are two principal methods commonly encountered in regression problems involving time-to-event data—in the present case, time until the occurrence of the first fail-

ure. The first method is the Accelerated Failure Time Model, and the second is the Proportional Hazards Model. Although numerous additional proportional models exist, ambiguity has generally been observed regarding the precise implementation of these proposed methods, as well as the diagnostic tools accompanying them for conducting the corresponding evaluations [11]. Since the investigated case study which is undertaken in this work ensures an adequate lifetime data volume, the Proportional Hazards Model is not assessed because the approach is less efficient and prone to loss of full survival time information since it replaces exact event time to ordering of events.

The Weibull distribution, in particular, has been used extensively in the past as a reliability model for studying time-to-failure data. In general, it is considered a generalization of the exponential model with variable failure-rate functions and provides results both for early burn-in failures and for end-of-life failures caused by wear-out mechanisms [12]. The DOE analysis or reliability datasets have a specialized toolbox for conducting life extension studies [13]. The DOE approach employed in the present study concerns the design of a product, specifically a manufactured thermostat, while the process details of production were originally established by Bullington et al. [14]. The case study sought to improve lifetime time performance under censoring conditions. The experimental design was carried out in the same manufacturing plant so as to optimize the operation of an industrial thermostat. The value of this study lies in the need to solve a real problem under interactive production conditions. Through Design of Experiments, the main factors examined in the experiment were clarified; these factors mainly concerned the manufacturing and processing procedure of the thermostat during its assembly stage. The factors selected for analysis resulted from an examination of all flow steps of the production process and inspection procedure and were derived from the judgments and analyses of specialized scientists and experts. The initial evaluation criterion for these factors was, on the one hand, their ability to influence the system response value and, on the other hand, the designers' ability to identify and maintain the levels of these factors within the experiment. The different design experiments required the consideration of two different factor levels, corresponding to high and low values, in order to establish a factorial variability on the lifetime response. Thus, the Plackett–Burman trial design method [15] was used, which included the construction of a two-level L_{12} array with ten observations per trial. Initially, the number of identified factors was quite high, but it was subsequently limited to eleven factors, at which point the array became saturated, with all columns fully assigned to individual factors. In practice, therefore, no empty columns could appear for possible interactions. This creates the possibility of confounding individual main effects with higher-order interactions. For this reason, the initial assumption was made that interaction effects were negligible and were not included. This also provided the possibility of estimating the main effects without confounding with two-factor interactions through a simple fold-over extension of the Plackett–Burman design.

Twelve factor combinations were carried out, and each of them corresponded to the testing of ten thermostats [14]. Overall, 120 thermostats were examined in operating cycles until failure. The maximum number of cycles at which the operating tests were terminated was 7,342,000. For the thermostats that continued to operate after the end of the cycle experiments, right-censored data were identified, corresponding to additional operating cycles. The best response was observed in trials 1, 6, and 11, with factors coded as E and H performing at the same level.

In order to focus on identifying failure during the early life of the product, the research team decided to take all these data into account but to use only the first two ordered statistics in each treatment combination, applying a signal-to-noise ratio analysis.

Due to the incomplete use of the data, it is generally accepted that a different analysis is necessary, one with broader data utilization and results anchored not only in the early stages of product operation but also in the long-term durability of the thermostat.

This study demonstrated that the appropriate experimental design contributed to the identification of factors causing non-compliance with process requirements and revised the necessary specifications of the thermostat production process in order to reduce the probability of failure. The expected lifetime increased by one order of magnitude, from the initial number of cycles to more than 7,000,000 cycles, with operational verification of 20 newly manufactured power elements. Because of censoring and non-fully randomized data, the challenge of developing a new robust procedure that would allow lifetime extension under more difficult conditions arose. A Bayesian variable selection procedure offered different follow-up predictions in the past making this case study interesting for further investigations [16]; it included the possibility of an interaction.

One way to study the cause-and-effect relationships non-parametrically between examined effects and Weibull distribution parameters was demonstrated by Besseris [17]. Using the super-ranking concept scale and shape parameters were individually ranked and concurrently optimized after fusion. Using Taguchi-type nonlinear orthogonal arrays for small samples, Weibull parameter multifactorial screening was confined to minimum replications such as to promote quick research results while retaining information about variability between replicates and extracting statistical conclusions. The approach placed the emphasis on robustly estimating the two Weibull-parameter responses and hence then to provide a simultaneous solution for optimizing robustly both parameters against four controlling factors by forming a single order-based response.

A frequently used method that contributed to the approach to the problem was Taguchi's SNR (signal-to-noise ratio) [5]. This is used to record simultaneous contributions of location and variation from trial replications. At the same time, the data were processed before applying ANOVA multiple-contrast capabilities. However, this method was clearly not sufficient, because it does not take into account all the required criteria for evaluating the response behavior of the product. The technique proposed in the Besseris work [17] can also be applied to the two-parameter Weibull model. The classical Taguchi technique can handle trial replications but was not applied to demonstrate the corresponding influences of shape and scale variations in the product's lifetime behavior.

The two-parameter Weibull distribution links the MTTF with the scale estimator in a linear manner through a gamma function. Its maximum value is obtained by simultaneously maximizing scale and shape, which is a critical prerequisite because it can lead to reliability optimization. Using the Besseris [17] technique, the shape and scale values are decoupled in the nonlinear Taguchi-type array and then recombined using the super-ranking concept. In fact, for the case of examining nonlinearity by evaluating three or more factorial settings, the non-parametric solver was selected to be the Kruskal–Wallis test. Essentially, Taguchi's SNR approach was replaced by a method based on the shape and scale parameters of the Weibull distribution, combined with super-ranking to condense individual reliability streams into a single response assessed through the non-parametric Kruskal–Wallis comparison test. Consequently, this research advanced a flexible technique for computing statistical tests of multiple settings for saturated and non-normal data. The case study in this field of industrial operations, specifically in the can-making sector, applied Taguchi's $L_9(3^4)$ OA in fused and saturated form. The examined factors were Mn and Mg content, the number of hot mill passes, and the cold mill reduction rate. The significance level of the effects of Mn and Mg indicated non-normal behavior. Meanwhile, the Taguchi mean estimation deviations led to the application of the non-parametric Kruskal–Wallis inference method. Only the Mn-content was found to be

statistically significant at a level of 0.10. Confirmatory tests showed estimated response limits with a deviation below 7% from the actual buckle strength values after the optimized settings were incorporated.

Wang et al. [18] examined how reliability can be improved through DOE and attempted to detect the factors that strongly affect percentile lifetime when the scale parameter is not constant. The data of Bullington et al. [14] were re-examined in an attempt to identify the factors that largely determine the location parameter and the levels of the significant factors using this time the smallest extreme value distribution. The technique employed maximum likelihood estimations which were screened using three different types of bootstrap methods—percentile bootstrapping (PB), accelerated percentile bootstrapping (BCa), and bias-corrected percentile bootstrapping (BCBP). The authors were mainly interested at obtaining estimations at lifetime probabilities of $p = 0.1$ and $p = 0.6321$. They noted that the location parameter is a special case of an estimation which identifies a particular percentile at $p = 0.6321$. The researchers focused on finding the factors that determine the 100p-th percentile. The surprising results were that all three methods found as strong effects fewer than the original study, which found that all 11 effects could be considered strong [14]. However, there was also disagreement among the bootstrap methods; BCPB and BCa found eight active factors, while PB reduced the active factors to five [18]. BCIs were not estimated because the implemented screening methods render them computationally expensive—a major problem facing engineers in reliability improvement projects.

Lv et al. [19] focus on reliability improvement through DOE while taking into account the constrained randomization of the experiments. The model they used assumed randomization effects expressed through MTTF under Weibull and SEV distributions. Through the thermostat example [14], it was shown that the traditional model assigns statistical significance to all factors, and this leads to unreliable predictions. In conclusion, it was found that the most important factors that should be reduced to lower levels are E, H, and their interaction. The authors identified the need for further examination of accelerated life tests, possibly with a non-log-location-scale distribution, such as the gamma distribution. However, an exhaustive screening of all possible interactions was not attempted.

Subsequently, Besseris and Tsarouhas [20] re-examined the problem of studying failure in the lifetime of the thermostat in order to discover the factors that affect the time to failure in low-percentile analysis. They focus on the 10th percentile of the thermostat paradigm [14], but disagreed with the arbitrary selection of a distribution, arguing that other types of distributions could also be included in the set of tests, especially for small, fractionated, and censored datasets. To remove the need for searching to a specific parametric reliability model, they employed a non-parametric treatment, relinquishing the dependence on the accuracy of identifying a proper parametric reliability model. They emphasized the computational convenience while searching for active effects when a multiplicity of lifetime distribution functions appears to offer approximately the same information about the nature of the datasets. They therefore proposed the sequential use of two non-parametric methods, Kaplan–Meier and Wilcoxon–Mann–Whitney, to calculate the 10% lifetime percentile. The Kaplan–Meier method can provide non-parametric estimates of survival probabilities for each factor combination in the Plackett–Burman array. With this method, an attempt was made to estimate the 90% survival rate, corresponding to the 10th percentile of lifetime. This method was applied as many times as there were controlling factors in the array. The ten measurements for each trial are compressed each time into a non-replicated vector and provide the measurement for the 10th percentile. The calculation of lifetimes at 90% survival was performed with the Non-

parametric Distribution Analysis (Right Censoring) module in software package Minitab [21]. This choice of data analysis compressed the dataset to a unreplicated-saturated form for the location and dispersion responses. For this reason, they resorted to the blocker-and-profiler method. This is parameterized through Kaplan–Meier and Wilcoxon–Mann–Whitney. The Kaplan–Meier calculation at 90% resolves the problem of equal lifetimes in the 9th trial in two ways. The first is expressed as the scenario of including a missing element in the 9th position of the 10th-percentile vector, or by extrapolation through curve fitting of lifetimes. The completion of the missing element in the 10th-percentile vector was replaced by an extrapolated value. It was calculated to be 97,000 cycles on the Kaplan–Meier survival probability curve at the cut-off value of 0.9. Factors E and H emerged as statistically active. It was demonstrated that even a single missing datum can lead to an incorrect detection of significant factors.

Wang et al. [22] attempted to maximize product reliability using the dual-response method. Through their article, they sought to use dual-response regression models for the lifetime response, focusing mainly on percentile lifetime and the scale parameter. They also aimed at an iterative computational method for assessing uncertainty in the identification of significant factors affecting percentile lifetime and standard deviation. They attempted to extend lifetime with the prospect of maintaining it in a stable manner. They took the warranty period into account in order to arrive at the selection of critical factors that provide the most optimized solution over time. They used the thermostat example of Bullington et al. [14] and implemented the above method with a simulated experiment on ready-made artificial data. They concluded that factor E is predominantly significant at the 10th percentile. However, with respect to the scale parameter, screening variability identified that the most critical factors were E and H.

Lv et al. [23] noted that the objective of most existing research material is reliability improvement through maximization of product lifetime. However, in real manufacturing, the objective is not only this but also the minimization of changes in product performance. An efficient way to achieve this is through robust parameter design. Nevertheless, the existence and observance of specific experimental protocols ignores the presence of certain experimental randomization factors. In this study, the authors attempt to achieve real optimization when the experiment is not fully randomized. More specifically, they used a Bayesian-based Weibull-fitted model for batch factors or experimental constraints. Bayesian posterior intervals declared whether a factor is significant or not. The optimal factor levels were determined with a multi-criteria optimization model that accounts, in reliability, for the reduction in total manufacturing cost, maximization of thermostat lifetime, and minimization of lifetime variance. The most important parameters appeared to be C, E, G, H, and I with respect to the scale parameter, while C, E, H, I, and K were important with respect to the shape parameter. Finally, they found that the best factor levels differ for each warranty period. These results may not have direct practical application since the product formulation is determined once, during product and process design, and it cannot be modified continuously afterwards while the product is manufactured.

Besseris [24] proposed a distribution-free method for full or fractional product-reliability and yield improvement. Reliability performance estimation is often correlated to product yield, thus knowing the performance of one characteristic may also imply the behavior of the other. With this proposed method, a multifactorial problem for reliability or yield optimization can be approached regardless of whether a full or fractional factorial experimental design is studied, in replicated or unreplicated form, while simultaneously handling censoring. The technique uses modified Wilcoxon–Mann–Whitney order statistics (the Halperin approach) and simultaneously examines factor effects and their

interactions. The technique is applied to four case studies, one of which is the thermostat case [14]. The flexibility of the method is demonstrated in moving from reliability-data to yield-data assessment without the need for computational iterative and resampling solvers, while bypassing the reliability distribution identification phase.

Wang et al. [25] introduced a posterior preference articulation approach for reliability improvement experiments, taking into account the loss costs of the manufacturer and the customer. As they report, the Design of Experiments in combination with the warranty period contributes to identifying the significant experimental factors for the product life cycle. A renewal process is presented within the warranty period to describe the manufacturer's loss cost. At the same time, a stress–strength model was used to describe the loss cost within the lifetime predicted by the design. The loss cost for the manufacturer and for the customer was calculated. They used the example of Bullington et al. [14] to explain the difference between the warranty time, which concerns the product reliability pursued by the manufacturer, and the design lifetime, which concerns the reliability pursued by the customer. They investigated the optimal factor levels that keep the manufacturer's cost reduced and the customer's cost below a certain value.

2. Methodology

2.1. Data and Purpose of the Research

This work introduces the compounding of interaction effects in the factorial screening of product/process lifespan behavior. Inserting interaction effects in multifactorial experiments permits the model to take into account interdependent, nonadditive effects among process controls, and hence to greatly expand the main-effects model, aiming to identify collaborative factor dependencies that govern the product lifespan. Guiding the product reliability requires the synergistic discovery of joint mechanisms which in turn promotes accurate profiling by considering factor interdependence, and thus prevents biased or incomplete process control adjustments. This work examines the influence of the 'core five' reliability distributions (Weibull, lognormal, normal, exponential and smallest extreme value) on the effect activeness. A "triple sieving" approach is enforced on the added interaction terms so as to arrive to a final effect profiling that contains a solution with active controlling factors and two-factor interactions. No new algorithm is needed to be created. The sequence of required statistical treatments has been devised such that the data analysis is performed on a mainstream statistical software to enhance the reproducibility of the predictions. The combination of applying successively diagnostic tools and techniques is unique to this study.

The response of the study was the lifetime of a product. More specifically, a review of the previous literature was carried out in order to determine the evolution of the way the experiment concerning product time-to-failure analysis, originally set by Bullington et al. [14], has been addressed. In particular, the product examined was an industrial thermostat, and the statistical analyses were conducted within the context of the preliminary design of the multistage manufacturing and assembly process for its production. The data used in the present study are based on the Bullington et al. [14] study. It required the execution of 12 experimental recipes, involving 11 different factors that could affect the lifetime of each thermostat. The 11 coded controlling factors and their two settings in parenthesis in brief were (A) diaphragm plating rinse (correct rinse/contaminated), (B) current density (60/15 amps), (C) sulfuric acid cleaning (3/30 s), (D) diaphragm electroclean (2/12 min), (E) beryllium copper gran size (0.008"/0.018"), (F) stress orientation (perpendicular/in-line), (G) humidity ("wet"/"dry"), (H) heat treatment (45 min/4 h), (I) brazing machine (no/excess cooling water), (J) power element electroclean (correct rinse/contaminated), and (K) power element plating rinse (correct

rinse/contaminated). The response data involved 10 repetitions for each run. For each trial run, the failure times were recorded with censoring. In the trial design the Plackett–Burman L_{12} planner was adopted at two-settings, for eleven controlling factors, requiring 12 factorial combinations. Otherwise, 10×2^{11} (=20,480) factorial experiments would have had to be conducted. Exploiting the DOE rapid data collection and analysis tactics promotes faster research completion. The factors evaluated and their responses under various combinations of levels are presented in Table 1 of Bullington et al. [14]. The researchers had set a censoring limit at 7,342,000 cycles in an accelerated form of trial execution so as to complete the experiments in a reasonable time frame.

The statistical analysis and the extraction of results were performed using Minitab software [21], while applying different models and distributions. The distributions applied were Weibull, lognormal, exponential, normal, and smallest extreme value.

The objective of the study was to select the factor effects or their interactions that yield the maximum lifetime until product failure, as well as to provide a comparative overview of the results among all the applied models.

2.2. Identification of Significant Candidate Interactions in Lifetime

The presentation of the dataset case study commences by the use of the Interaction Plot in the software package Minitab [21] to examine candidate interactions that may be of potential significance, from which certain possible interactions were screened and selected. Quantifying interactions in production trials is pivotal because it reveals concealed, nonlinear relationships between factors that one-variable tests ignore. This is because the potency of one independent factor is not the same for all settings of the other independent factor. Interaction is a complex multivariable influence which contributes more accurate information than a simplified main effect. Identifying the presence of interactions prevents costly production defects, and optimizes true product/process performance. Dismissing interaction effects may lead to false optimization, overlooking the true peak performance adjustment that only exists at a specific combination of both factors. Two-factor interactions are easily depicted in the interaction plot.

The interaction plot quickly summarizes the core visual evidence of potentially many two-way interaction combinations. In the illustrating thermostat example, there are 11 controlling factors and thus 55 two-way interactions to consider. Applying a regression model to a total of 66 potential effects with only 120 censored datapoints—of unknown distribution—might instigate serious overfitting concerns. The interaction plot furnishes a qualitative portrayal with respect to the degree of interaction present between two controlling factors. Visual depiction of an interaction relays the non-parallelism of parameter effects. When strong interactions occur between the parameters, lines cross in the individual interaction subplots. This type of empirical information quickly eliminates those factorial pairs that do not fit the strong interaction motif and hence greatly shortens the search list of effects for the multifactorial screener.

Based on the objective set for this study, all factors first had to be examined in terms of their two-way interactions, in order to investigate which of these candidate interactions might be nominated for further statistical examination. For this reason, statistical processing was carried out in Minitab using ANOVA, and the interaction plot was obtained. The interaction plot is a 12×12 matrix in which the mean values of the dependent (lifetime) variable are shown for all factors and at all their levels. The axis scales are appropriately configured to allow a quick identification of each two-way interaction on display. In general, the appearance of parallelism between the two lines indicates no interaction, whereas an indicative crossing between same-factor levels implies that a strong interaction may be present. Thus, through the interaction plot, it was possible to

filter out those candidate interactions that could receive subsequent data analysis.

2.3. Examination of Interactions and Factors Using the Screening Model

The candidate interactions selected in the previous step were entered into the screening model (Minitab). The combination of the Lenth method in a Pareto chart generated the statistically significant effects and interactions based on the Lenth statistical estimator [26]. This screening design removes the remaining model terms from further consideration on the individually selected reliability distribution. This method was chosen because it allowed the exploration of all factor effects with a moderate number of trials, while allowing to enter in the next step the influence in the reliability predictions due to interactions [27].

The calculation of the required response for lifetime response until failure (LF) can be treated, based on the screening model, as a linear equation containing interaction terms. The relationship describing it is based on least-squares regression and is:

$$LF = \beta_0 + \sum_i \beta_i X_i + \sum_{i < j} \beta_{ij} X_i X_j \quad (1)$$

where β_0 is the constant term, β_i is the i th factorial coefficient, β_{ij} is the ij^{th} interaction coefficient, X_i are the main factorial settings, $X_i X_j$ are the settings for the candidate interactions between two factors, and ϵ is the experimental error. In a least-squares regression model, the regression coefficients indicate the change per unit change in a factor variable, while the interactions indicate the changes between high and low levels [4,5,13]. The analysis was performed using ‘Stepwise Selection of Terms’, with levels of overall significance: (1) ‘ α to enter’ = 0.15 and (2) ‘ α to remove’ = 0.15. The stepwise regression is geared toward early-stage screening with its selection criteria to be judged against the default threshold value of 0.15—automatically set by the MINITAB software—because it counterbalances the exploratory model building while preventing active regressors from being prematurely excluded and simultaneously avoiding overfitting.

The ‘ α -to-enter’ and ‘ α -to-remove’ values determine, in combination with the p -value of a variable, whether that variable will be removed from or added to the process of adding or deleting predictors in the model, according to the F-test up to a specific number of degrees of freedom. The ‘ α -to-enter’ value denotes the maximum p -value a variable (factor or interaction) is allowed to possess so as to be retained in the model. The ‘ α -to-remove’ is a threshold in which if the p -value of a variable is higher than that limit, it does not possess the statistical significance to be included in the regression model. Thus, stepwise selection of factors and interactions is performed through successive forward selection and backward elimination procedures until the degrees of freedom are fully utilized [26]. The Anderson–Darling (adjusted) statistic was used as the main criterion for determining the goodness of fit owing to the screening model, and according to the “Analyze Screening Design” option in the “DOE-Screening” MINITAB module.

2.4. Examination of Candidate Interactions and Factors Using the Regression with Life-Data Model

At a subsequent stage of the study, the “Regression with Life Data” model was used to determine the probability that certain factors and candidate interactions identified in the screening phase affect the product’s time to failure in association to particular reliability distribution. This model contributes to the straightforward explanation of the rapid failures due to some factor settings and the long-term survival of others. This regression model has the advantage of being able to accommodate censored data and mul-

tiple distribution forms.

The relationship describing a life-data regression model is:

$$Y_p = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \sigma \Phi^{-1}(p) \quad (2)$$

where Y_p is the failure time or its logarithm, β_0 is the constant representing the intercept with the y-axis, $\beta_1 \dots \beta_k$ are regression coefficients, $x_1 \dots x_k$ are predictor values, σ is the inverse of the shape for the Weibull distribution or the scale for other distributions, and $\Phi^{-1}(p)$ is the p -th quantile of the standardized life distribution [8,13,27].

Five reliability distributions were assessed using regression modeling. Specifically, the Weibull, lognormal, exponential, normal, and smallest extreme value distributions were examined.

More specifically, for the probability density function for the Weibull distribution (α , λ) is defined by the parameters $\alpha > 0$ and $\lambda > 0$, being the shape and the inverse of the scale, respectively. It is expressed as follows:

$$f(t) = \begin{cases} \alpha \lambda t^{\alpha-1} e^{-(\lambda t)^\alpha}, & \text{if } t \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The failure-rate function of this distribution is given in Equation $\lambda(t) = \alpha \lambda t^{\alpha-1}$. When $0 < \alpha < 1$, $\lambda(t)$ decreases, and when $\alpha > 1$, it increases. If $\alpha = 1$, $\lambda(t)$ has a constant value corresponding to the exponential distribution [8].

The smallest extreme value distribution represents the behavior of a random variable t which follows a Weibull distribution and its data points have been converted to logarithmic values ($\ln(t)$). The probability density function of the smallest extreme value distribution can be described as follows if the location and scale parameters are μ and σ ($\sigma > 0$), respectively:

$$f(t) = \frac{1}{\sigma} \exp\left(\frac{t-\mu}{\sigma}\right) \exp\left[-\exp\left(\frac{t-\mu}{\sigma}\right)\right] \quad (4)$$

The exponential distribution is a continuous distribution of a variable t with the following probability density function and rate parameter λ :

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{if } x > 0 \text{ and } \lambda > 0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The normal distribution of a variable t is defined by the following probability density function with parameters the mean value μ and standard deviation σ :

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right) \quad (6)$$

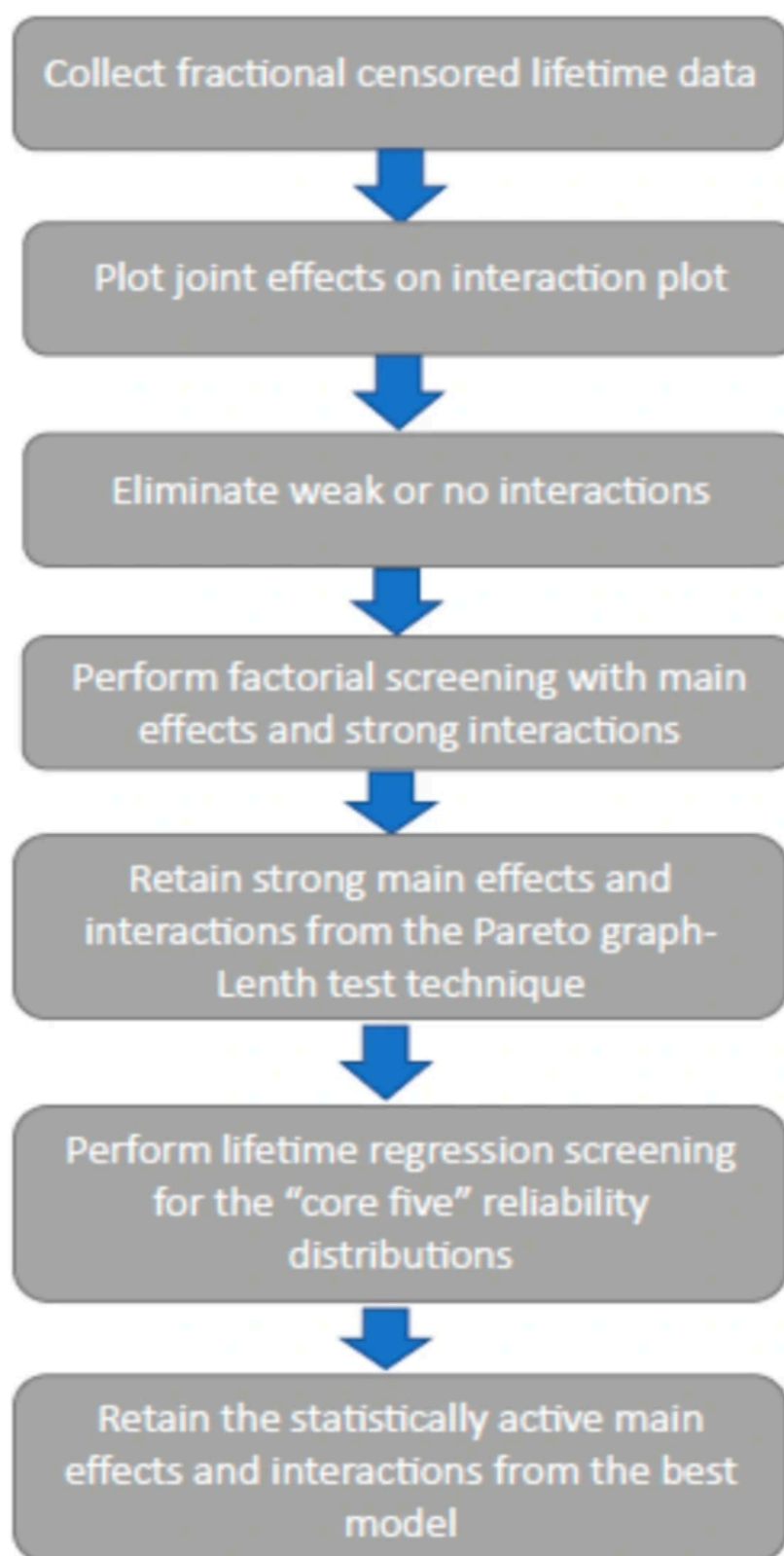
The probability density function of the lognormal distribution is:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right) \quad (7)$$

where μ is the mean of the natural logarithm of the variable, and σ is the standard deviation of the natural logarithm of the variable for $t > 0$ [28].

The maximum likelihood principle was used to choose among the different reliabil-

ity distribution models, and according to the “Regression with Life Data” option in the “Reliability/Survival” MINITAB module. Supportive information was also provided by using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The outline for the statistical analysis steps is shown in Scheme 1.



Scheme 1. The sequence of statistical processing of the factorial censored lifespan dataset.

3. Results

The presentation of results commences with the application of the interaction plot. Figure 1 illustrates all two-way candidate interactions in coded controlling factor mode, covering factors A to K (MINITAB 18). The visualization of the interaction strength level is empirically detected by selecting those relationships that manifest an evident crossing between factorial levels.

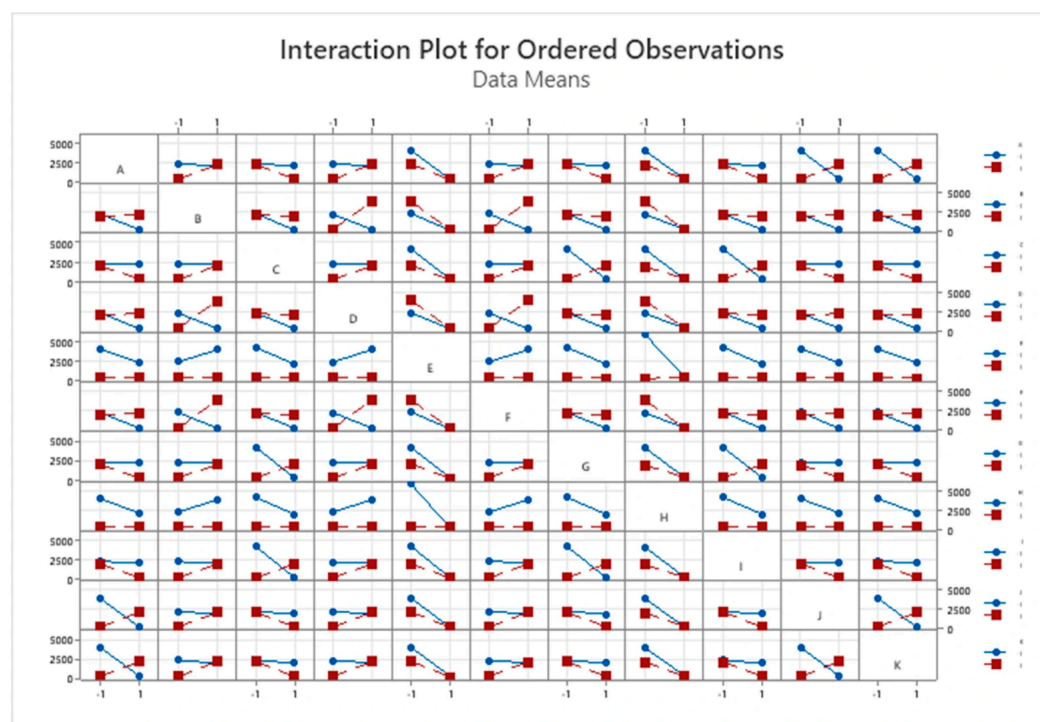


Figure 1. Interaction plot (MINITAB 18) for the eleven controlling factors of ref. [14] (vertical axis is the lifetime scale—all factorial levels are in coded form).

Based on the interaction plot, all two-way interaction combinations appear to be either weak or strong; there is no interaction relationship that may be deemed as “no interaction”. The candidate interaction combinations that could be retained for a subsequent analysis are a total of nine pairs (out of a total of 55 two-way interactions): AJ, AK, BD, BF, CG, CI, DF, GI, and JK.

The next step is to consider the potency of the original 11 effects along with the nine candidate interactions for screening. All 20 nominated terms were entered for processing in the “Screening Design Model” module (MINITAB 18). The screening analysis of the lifetime dataset generated a Pareto chart for the standardized effects (Figure 2). The Pareto chart provided two types of information: (1) the activeness of the effects/interactions, and (2) the hierarchy of the effects/interactions. The selection threshold has been estimated by the Lenth method on the absolute values of the standardized effects. The Pareto bar values correspond to Lenth modified t-statistics for each effect/interaction in order to test the null hypothesis of effect inactiveness. The nominated effect/interaction group has been dichotomized based on the Lenth cutoff estimate value of 1.450, represented by a reference line on the drawn Pareto chart. The estimated cutoff point corresponds to a selected significance level of $\alpha = 0.15$ (automatically assigned by the software package MINITAB 18), which is applied to all effects and candidate interactions. Factors or interactions located to the right of the reference line are considered statistically significant and are retained for further analysis, whereas those located to the left are considered weak, therefore, they are excluded from further investi-

gation. In summary, according to the Pareto chart analysis, the active group of controlling factors and candidate interactions are listed in terms of descending order of strength: E, G, H, BD, K, AJ, F, and CI. It is observed that 5 out of the 11 controlling factors and 3 out of the 9 interactions are retained for the next phase of the analysis.

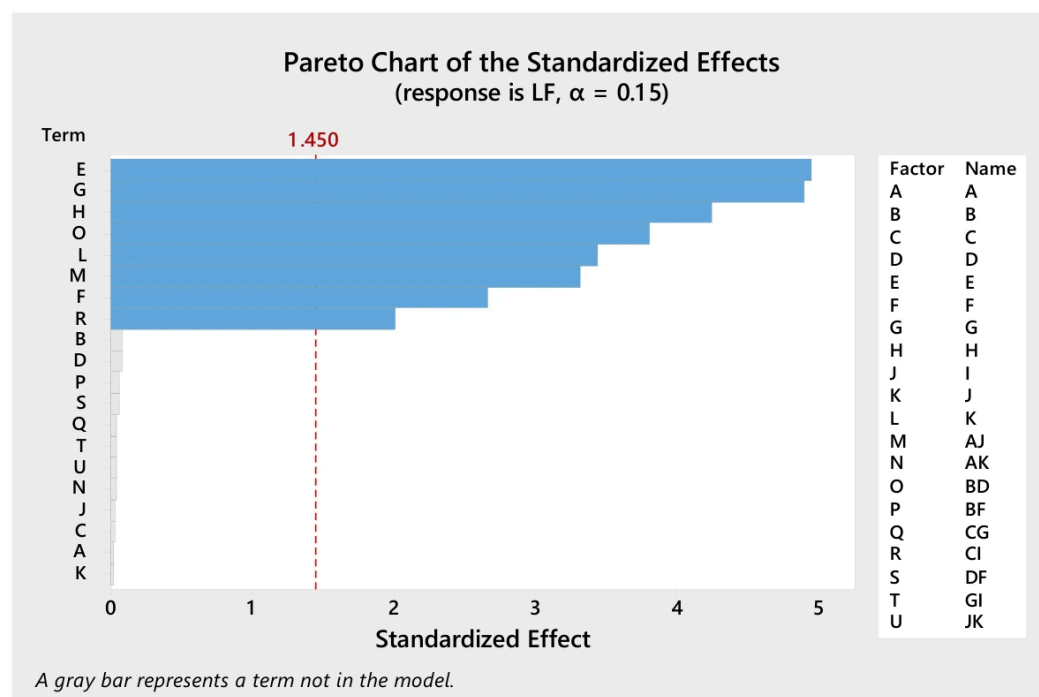


Figure 2. Visualization of screening statistically significant factors and interactions using a combination of a Pareto chart and the Lenth test.

The presence of increased dispersion, outliers, and certain deviations from the normal behavior of residuals is confirmed in Figure 3. The instability in the constructed normal probability plot is evident, as it is explained by the right skewness of the residuals in the histogram, as well as accentuated localities in the residuals versus fitted values plot, and residuals versus observation order plot. The Anderson–Darling test further confirmed the existence of severe deviations in the fitted model, with a p -value < 0.005 .

Furthermore, Table 1 demonstrates the statistical significance of each contribution against an individual significance level of $\alpha = 0.05$ in the multifactorial ANOVA treatment, which also indicated that there is no lack of fit in the proposed model (Table 1A). On the same note, the linear model appears to be statistically significant (p -value < 0.001). The dominant effects appear to be the beryllium copper grain size, the humidity and the heat treatment. Similarly, the predominant interaction is identified between the factors regarding the current density and the diaphragm electrocleaning. From Table 1B, it is shown that the existing model explains about 75% of the variability in the lifetime dataset. From Table 1C, it is also observed that out of the five effects, four of them have negative effect on controlling lifetime behavior, while all three candidate interactions have positive contribution. Finally, it is noted that the variance inflation factor estimations were maintained low ($VIF < 4$) for all considered effects and interactions, indicating that multicollinearity does not distort substantially the regression results.

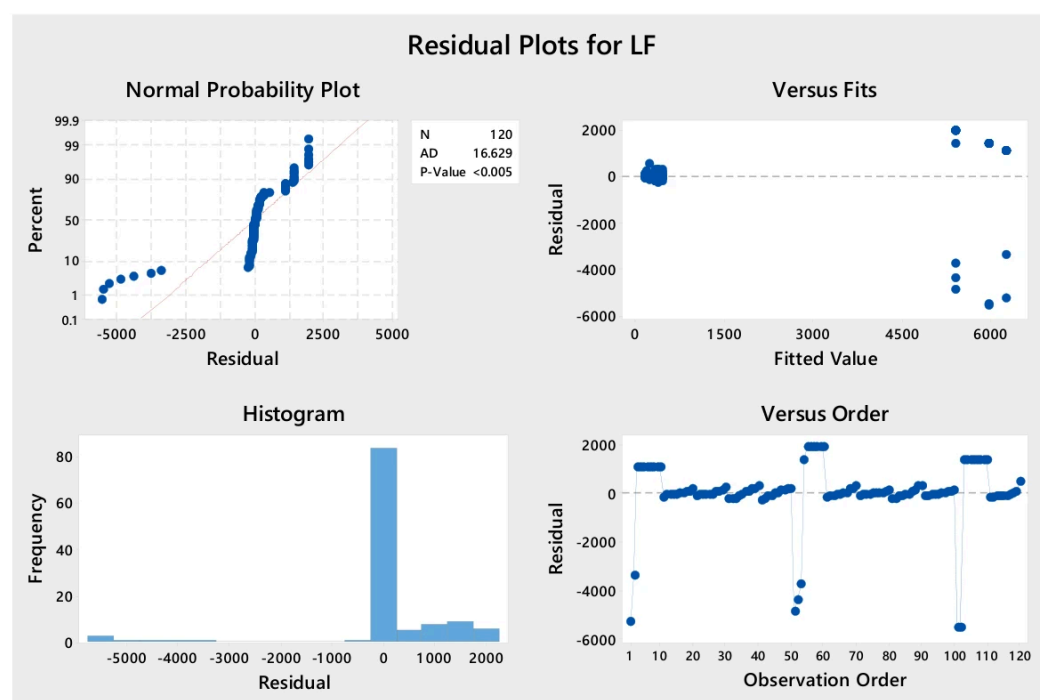


Figure 3. Four-way residual plots for the lifetime dataset screening process.

Table 1. Screening Design Model lifetime results using stepwise selection terms for $\alpha = 0.15$ (MINITAB 18).

(A) Analysis of Variance Results					
Source	DF	Adj SS	Adj MS	F-Value	p -Value
Model	8	701,013,513	87,626,689	46.19	0.000
Linear	8	701,013,513	87,626,689	46.19	0.000
E	1	46,480,564	46,480,564	24.50	0.000
F	1	13,435,776	13,435,776	7.08	0.009
G	1	45,391,040	45,391,040	23.93	0.000
H	1	34,109,548	34,109,548	17.98	0.000
K	1	22,395,830	22,395,830	11.81	0.001
AJ	1	20,815,481	20,815,481	10.97	0.001
BD	1	27,461,524	2,7461,524	14.48	0.000
CI	1	7,645,486	7,645,486	4.03	0.047
Error	111	210,562,898	1,896,963		
Lack-of-Fit	3	16,363	5454	0.00	1.000
Pure Error	108	210,546,535	1,949,505		
Total	119	911,576,411			
(B) Model Summary					
S		R ²	R ² (Adj)	R ² (Pred)	
1377.30		76.90%	75.24%	71.50%	
(C) Coefficients of Regression					
Term	Coef	SE Coef	T-Value	p -Value	VIF
Constant	1693	126	13.47	<0.001	
E	−730	147	−4.95	<0.001	1.38
F	516	194	2.66	0.009	2.37

G	−948	194	−4.89	<0.001	2.37
H	−625	147	−4.24	<0.001	1.37
K	−666	194	−3.44	0.001	2.37
AJ	765	231	3.31	0.001	3.38
BD	879	231	3.80	<0.001	3.37
CI	464	231	2.01	0.047	3.37

The final phase of the analysis was to perform lifetime regression by screening simultaneously the effects and the candidate interactions that were retained from the Pareto/Lenth analysis phase. To complete this phase the original lifetime dataset [14] was fitted to the five available reliability distributions using the MINITAB 18 software. The screened statistical distributions were modeled according to the following functions: (1) Weibull, (2) smallest extreme value, (3) exponential, (4) normal, and (5) lognormal.

The lifetime dataset has been fitted using regression for the Weibull distribution, which also accounted for the 22 censored observations (out of 120 datapoints). The maximum likelihood estimations of the predictors are listed in Table 2. Individually, all factors and candidate interactions, including the constant term, are significant on a level of $\alpha = 0.05$. It is noted that four of the factors have negative effect on controlling lifetime behavior, while all three candidate interactions have positive contribution. Factors E, G, H, K and F appear to be statistically significant, while interactions AJ, BD and CI are all comparably active. Similar observations may be drawn after fitting the lifetime dataset using the smallest extreme value distribution (Table 3). Increasing contributions are identified to factor F and all three candidate interaction terms for both reliability models. The results are differentiated when using the exponential distribution (Table 4). In this case, factors F and K appear inactive as well as the CI interaction. AJ remains the stronger interaction effect, whereas the factorial effects G and E are the most dominant. Assuming a normal distribution (Table 5) only eliminates the CI interaction, and still detects E, G, H and K as being on par in strength. Finally, assuming a lognormal distribution, factor H appears statistically insignificant (Table 6). Factors G, E and K are similarly strong and factor F should also be retained. All three candidate interactions appear influential and they should be included in the structural modeling.

Table 2. Lifetime regression results using MLE for estimating effects: censored Weibull distribution (MINITAB 18).

Predictor	Coef	Standard	Z	p	95.0% Normal CI	
		Error			Lower	Upper
Intercept	6.68838	0.0658155	101.62	<0.001	6.55939	6.81738
E	−0.545427	0.0632514	−8.62	<0.001	−0.669397	−0.421456
F	0.446519	0.150139	2.97	0.003	0.152252	0.740787
G	−0.677562	0.141217	−4.80	<0.001	−0.954343	−0.400781
H	−0.215172	0.0619007	−3.48	0.001	−0.336495	−0.0938491
K	−0.452696	0.150166	−3.01	0.003	−0.747016	−0.158376
AJ	0.680422	0.164399	4.14	<0.001	0.358206	1.00264
BD	0.493650	0.165305	2.99	0.003	0.169658	0.817642
CI	0.354186	0.133858	2.65	0.008	0.0918292	0.616542
Shape	1.85668	0.150871			1.58332	2.17723

Table 3. Lifetime regression results using MLE for estimating effects: censored smallest extreme value distribution (MINITAB 18).

Predictor	Coef	Standard	Z	p	95.0% Normal CI	
		Error			Lower	Upper
Intercept	2210.87	61.1422	36.16	<0.001	2091.03	2330.71
E	−970.763	56.7938	−17.09	<0.001	−1082.08	−859.449
F	883.747	137.053	6.45	<0.001	615.129	1152.37
G	−1112.46	129.158	−8.61	<0.001	−1365.60	−859.311
H	−870.401	56.7183	−15.35	<0.001	−981.567	−759.235
K	−883.750	137.053	−6.45	<0.001	−1152.37	−615.132
AJ	1057.33	150.436	7.03	<0.001	762.477	1352.17
BD	1006.16	150.586	6.68	<0.001	711.016	1301.30
CI	795.672	122.346	6.50	<0.001	555.879	1035.46
Scale	491.701	47.4968			406.891	594.189

Table 4. Lifetime regression results using MLE for estimating effects: censored exponential distribution (MINITAB 18).

Predictor	Coef	Standard	Z	p	95.0% Normal CI	
		Error			Lower	Upper
Intercept	6.75291	0.122191	55.27	<0.001	6.51343	6.99240
E	−0.662282	0.114552	−5.78	<0.001	−0.886801	−0.437764
F	0.459865	0.278731	1.65	0.099	−0.0864374	1.00617
G	−0.880806	0.259373	−3.40	0.001	−1.38917	−0.372444
H	−0.268514	0.114552	−2.34	0.019	−0.493032	−0.0439959
K	−0.484461	0.278731	−1.74	0.082	−1.03076	0.0618412
AJ	0.771910	0.304911	2.53	0.011	0.174296	1.36952
BD	0.659514	0.304932	2.16	0.031	0.0618583	1.25717
CI	0.329861	0.248487	1.33	0.184	−0.157165	0.816887
Shape	1					

Table 5. Lifetime regression results using MLE for estimating effects: censored normal distribution (MINITAB 18).

Predictor	Coef	Standard	Z	p	95.0% Normal CI	
		Error			Lower	Upper
Intercept	1889.54	150.373	12.57	<0.001	1594.81	2184.26
E	−827.889	170.061	−4.87	<0.001	−1161.20	−494.577
F	519.782	249.356	2.08	0.037	31.0540	1008.51
G	−1181.80	245.543	−4.81	<0.001	−1663.06	−700.549
H	−723.272	170.061	−4.25	<0.001	−1056.58	−389.960
K	−721.981	247.895	−2.91	0.004	−1207.85	−236.116
AJ	905.130	288.143	3.14	0.002	340.381	1469.88
BD	1070.93	292.244	3.66	<0.001	498.142	1643.72
CI	425.933	278.452	1.53	0.126	−119.824	971.689
Scale	1575.40	118.626			1359.24	1825.93

Table 6. Lifetime regression results using MLE for estimating effects: censored lognormal distribution (MINITAB 18).

Predictor	Coef	Standard	Z	p	95.0% Normal CI	
		Error			Lower	Upper
Intercept	6.35446	0.0740355	85.83	<0.001	6.20936	6.49957
E	−0.586866	0.0829145	−7.08	<0.001	−0.749375	−0.424356
F	0.299717	0.125964	2.38	0.017	0.0528313	0.546602
G	−0.605316	0.123023	−4.92	<0.001	−0.846436	−0.364196
H	−0.120664	0.0829145	−1.46	0.146	−0.283174	0.0418450
K	−0.505040	0.123983	−4.07	<0.001	−0.748042	−0.262038
AJ	0.426607	0.142183	3.00	0.003	0.147934	0.705280
BD	0.552054	0.147672	3.74	<0.001	0.262622	0.841485
CI	0.332364	0.137946	2.41	0.016	0.0619949	0.602733
Scale	0.766255	0.0569572			0.662372	0.886431

The corresponding *p*-values associated with each factor and candidate interaction under different distribution fitting tests are listed in Table 7, also tabulating the log-likelihood score.

Table 7. Comparative results of the lifetime regression with various reliability distributions.

Effect <i>p</i> -Values per Distribution Type					
Regressor	Weibull	Smallest Extreme Value	Exponential	Normal	Lognormal
E	<0.001	<0.001	<0.001	<0.001	<0.001
F	0.003	<0.001	0.099	0.037	0.017
G	<0.001	<0.001	0.001	<0.001	<0.001
H	0.001	<0.001	0.019	<0.001	0.146
K	0.003	<0.001	0.082	0.004	<0.001
AJ	<0.001	<0.001	0.011	0.002	0.003
BD	0.003	<0.001	0.031	<0.001	<0.001
CI	0.008	<0.001	0.184	0.126	0.016
Log-Likelihood	−664.229	−806.779	−686.086	−879.211	−681.924
AIC	1344	1630	1388	1774	1380
BIC	1366	1652	1410	1796	1402

The statistical significance of factor E (beryllium copper grain size) is confirmed across all tested reliability distributions. The influence of factor G (humidity) is also expected to deliver the same level of influence in controlling the lifetime performance according to all modeling efforts. Factor K (power element electrocleaning) and H (heat treatment) appear active with respect to four out of five reliability models; several variations in significance values are observed depending on the choice of the distribution model—excluding the exponential model. On the other hand, interaction AJ (between diaphragm plating rinse and brazing machine) is consistently present in all reliability model fittings providing estimations with high significance values. Additionally, interaction BD (between current density and diaphragm electrocleaning) is considered a strong contribution in all five models. Factor F appears to be significant across all distributions, with varying estimations, except in the case of fitting the exponential distribution. Factor H appears significant in the fittings of Weibull, Smallest Extreme Value, and Normal dis-

tributions, while it is not significant with regard to the Lognormal distribution. Interaction CI is significant in three reliability models, but not when modeled to Exponential and Normal distributions. It is also noted that even between the Exponential and Lognormal models, which exhibit similar predictive performance, the values of the terms CI, F, and K are highly inconsistent in strength.

Based on the log-likelihood values, the Weibull distribution model shows the best predictive performance (−664.2). However, the predictions by the lognormal and exponential models are not statistically apart. If the decision scenario was involving a cumulative inference from those three reliability models then, factors A and G, as well as interactions AJ and BD should be included in further modeling considerations. The comparison of the models agrees to the same prioritization of the models using the decision rule for lower values of the AIC and BIC estimations in Table 7.

The normal, lognormal, Weibull, exponential, and smallest extreme value distributions were chosen as default distributions by the MINITAB software to initiate a distribution screening procedure. The five selected distributions are fundamental in reliability engineering because their mathematical properties map directly to physical failure physics, wear-out mechanics, and weakest-link behavior. Other reliability distributions (such as gamma, log-logistic, or generalized gamma) are often omitted as primary defaults because they add unnecessary mathematical complexity, require larger sample sizes to estimate extra parameters (censoring and parameter overhead), and offer redundant physical interpretations compared to this core five. Adding extra parameters (such as a threshold or third shape parameter) makes maximum likelihood estimation unstable or impossible with small sample sizes.

Different reliability models are limited by different theoretical assumptions. Controlling factors and their associated interaction terms, which have been identified as active predictors in all tested models, may be viewed as “surviving” effects because they influence the lifetime response: (1) under constant failure rate (memoryless trait) in random not-aged component failures (exponential distribution), (2) under symmetrically clustered failures with a low variability time-to-failure mean value (normal distribution), (3) under cumulative multiplicative degradation processes (lognormal distribution), (4) under a flexible shape-based lifetime behavior which adapts to a “weakest-link” physical model (Weibull distribution), and (5) under weakest-link modeling to extreme minimum stress/breakdown phenomena (smallest extreme value distribution). With respect to contradictory results, not-all-model-consensus effects may require further exploration by incorporating in the research methods sensitivity analysis exercises while integrating additional field data.

The results which were presented in this work concern operational situations in which quick information on the reliability condition of a product or process is needed under manufacturing constraints in time and resources. The use of fractional factorial designs expedites the lifetime data collection process. At times, this might be a single historical event, during which important decisions may be taken, to either immediately consider the predicted behaviors to future process/product modifications, or proceed to additional rounds of experimentation, if time, resources and engaged processing units are permitted to become available. Pragmatic industrial trials are conducted on the production line; the example that was elucidated in this case study involved a range of processes on the production line. This means that operational availability is a major issue that decisively restricts the extent and the duration of manufacturing research trials because it removes income-generating flows from the daily production schedule. Under a limited experimentation outlook then, factorial screening may be conducted and its generalizability should be assessed. The proposed approach demonstrated that it took ad-

vantage of lifespan multifactorial data and attempted to extract from them potential interaction contributions. Under ordinary conditions, a much larger amount of data would be required to question the potential influence of interactions. On the other hand, direct estimations of all two-factor interactions were not possible in an ordinary life-data regression model due to realizable software constraints. Furthermore, directly evaluating all possible interactions at once—in the absence of previous knowledge—would raise serious overfitting concerns due to their large number (55) of interactions against a restricted lifetime sample (120 data points). Therefore, the first step to nominate some interactions was to assess the visualized information which was generated from the classical interaction plot. Strong candidate pairs are easily identified within the exhaustive interaction subplot combinations, but they are not statistically confirmed. After this first filtering, factorial screening ensues using the complementary treatment of Pareto graph and Lenth test to identify which active terms to retain. The results may be generalizable if the active group of effects substantially separate from the weaker ones even in the case where model assumptions do not hold. At this stage, it is not expected to hold because the reliability reference law that governs the investigated process is still at large. Uncertainty in describing lifetime datasets begins with choosing a suitable distribution, because the normal behavior of the lifetime dataset is never guaranteed. Therefore, the final stage will identify those effects that manage to stand out in all five reliability distribution screenings, if such effects exist. Otherwise, further exploration may be needed with additional data to test deeper the corresponding model assumptions.

In a nutshell, the physical mechanisms behind the final main-effect regressor list may be summarized as follows [14]: (1) the active power element plating rinse is important in the tin-coating of the power element, in connection to electrolytic residue removal, (2) the grain size of the beryllium copper affects the resistance on stresses due to forming, convoluting and shearing, (3) the heat treatment induces tensile strength loss leading to premature diaphragm failures, (4) humidity interferes with contaminating the power element, and (5) stress orientation in the diaphragm influences failure modes in the power element. Correspondingly, the interaction-based regressors imply: (1) Diaphragm electrocleaning quality may interact with the tin plating current density in regulating nodularity, (2) preceding sulfuric acid cleaning quality may interact with the heat quenching process of the power element, and (3) the preceding successive primary and secondary rinse quality may interfere with power element electrocleaning effectiveness, since a contaminated bath may induce diaphragm corrosion.

4. Discussion

Realizing that the thermostat case study was not a trivial one, the researchers [14] attempted several methods to predict the significance of the examined controlling factors. In one version, using a lognormal distribution in regression analysis, it was shown that all 11 factors were active. However, by treating the lifetime dataset with pooled-ANOVA, only factors E and H were found significant and only those two were taken in account in follow-up process correction. Unlike the literature, we have demonstrated the potential effect of candidate interactions on screening.

The influence of candidate interactions was also not assessed in the bootstrap analysis of the same thermostat lifetime dataset [18]. Percentile bootstrapping inferred that factors C, E, G, H and J were important; there is disagreement with regard to our work on inclusion of factors C and J. Moreover, factors B, D and F are added in the previous list if the bias-corrected percentile bootstrapping and bias-corrected accelerated percentile bootstrapping are employed. Only F is to be retained in comparison to the screening results of our method. In order to clarify and better understand how the solution to

the same problem differs when using the aforementioned methods, an evaluation compared to previously published approaches is also required.

The present study employed Design of Experiments (DOE) in order to perform an appropriate statistical analysis of the factors and their candidate interactions that could cause variation in the product's operational reliability. Since failure data analysis is heavily influenced by suitably adopting a reliability distribution, the task of choosing from a broad range of available models may become challenging. Using the Bayesian variable selection procedure for the lognormal distribution, only factors E and H, as well as their two-way candidate interaction were found to be statistically significant [16]. This Bayesian solution agrees to the two factors when compared to the screening active effect list from our work, but there is disagreement as far as the respective effect of their interaction.

Censoring is an anticipated condition in planning to collect reliability data. This complication requires more specialized statistical approaches to be considered in order to accommodate any of the selected reliability distributions to complete the structural modeling task. Lv et al. [19] re-examined the thermostat dataset by incorporating in the statistical formulation process a random effect model, while assuming a smallest extreme value distribution, which also could handle censored datasets in maximum likelihood estimations. Surprisingly, this model found that all effects were significant, excluding only the influence of factor F. This finding was in conflict to the original prediction by Bullington et al. [14], Wu and Hamada [16], Wang et al. [18] as well as the results which were obtained from the present work. By intervening on the original thermostat dataset and replacing the “weak factor F” to the interaction EH, then, the new model estimations showed that factors E and H, along with the interaction EH are the influential model terms [19] which is also substantially differentiated from the original study [14] in addition to previous resampling efforts [18]. Posterior preference articulation may also rely on factors E and H being the strong regressors in reliability improvement against costs of manufacturer and customer [25]. However, customer's loss function cannot be assumed to be linear according to Taguchi's philosophy [5]. Meanwhile, the Weibull model assumption may not be relevant to all real scenarios according to Wang et al. [25], since in reliability problems it is the lifetime dataset that would dictate the behavior. This in turn may multiply the number of modeling schemes that an engineer needs to take in account in order to perhaps approximate a viable solution.

The residuals obtained from the above plots showed that the data distribution is not fully normal, mainly due to the presence of many scattered residual outliers, high variability in the values, and heteroscedasticity. For this reason, it may not be possible to accurately describe the data using a single simple linear model. However, according to the results obtained, our approach in this study may be viewed as beneficial since it considers the behavior of the thermostat dataset from the perspective of five different reliability distributions which remain popular in reliability engineering applications of improving product/process lifetime behavior.

In this work, there was no special treatment for censored values. We kept the methodology simple. In the interaction plot (Figure 1), each cell is a subplot. Each subplot comprises four graphed points. Each point on the subplot is a mean of 30 datapoints ($=120/4$). The 22 censored points are dispersed within the 120 total datapoints. We usually accept 30 points as an adequate amount of data to at least estimate a mean. Any influence of censored values is only going to exaggerate the high setting of one factor with the low setting of the other in order to exhibit the “crossing lines” pattern. It is not a terminal decision that indeed those interactions are strong at this stage. This will be checked statistically in the next (screening) phase (Table 1 & Figure 2) in order to be

deemed appropriate to enter the reliability modeling phase and finally be checked on the respective reliability distributions, which include normal and non-normal options.

The proposed approach introduced a triple-layer screening using: (1) a graphical interaction screening, (2) a statistical factorial screening, and (3) a distribution analysis of the screened results. It was shown that even though the normality of residuals was not attained, however, the Pareto graphing in concert with the Lenth test offered a sharp separation of active and inactive factorial/interaction effects as it was shown in Figure 2. The proposed approach was conceived for situations where available FFD data for censored observations cannot receive augmented interaction effect analysis in the life-data regression framework given the fact that involving all possible two-way interactions in the regression model would be prone to overfitting; the particular case study would expect 55 interaction terms in the structural model.

In this work, the location estimator was maintained to be the mean of the sampled data points as it was intended in all past works that investigated the industrial thermostat problem [14,16,18,19,22]. The location estimator is also consistent with the ANOVA results that were obtained in the original research (Tables 4 and 7 in ref. [14]), as well as the performed regression analysis outcomes (Tables 5–7 in ref. [14]), which included the sampled censored values. There might be a limitation in using censored observations in the overall estimation scheme. However, multi-parameter Kaplan–Meier comparison methods were not available in the implemented software application to contrast simultaneously the candidate interactions.

Violations of model assumptions may be critical if tested effects have narrow margins of magnitude between them and with respect to the threshold limit on the Pareto graph/Lenth test screening plot (Figure 2). However, in the illustrating example, strong and weak effects have been grouped distinctly in the two classes such as the strong group of effects to proceed to the final screening requirement using multi-distributional life-data regression. On the other hand, contested effects around the threshold limit may be included in the final screening that follows, even if they are under the limit but in proximity. Violations on the Pareto/Lenth test screening is expected because the reliability reference law is still at large at this phase. However, the fitness of the proper distribution is examined in the next (third) phase. Therefore, all three phases should be performed in order to succeed in including in the model contributions from interactions. This is an empirical resolution, since the alternative to handle all 55 interactions at the same time is not attainable with some other method. Even if there was, an overfitting issue would loom around the modeling structure. This also aids to appreciate the difficulty and uniqueness of the problem.

The practical implications of attempting to predict a lifespan structural model comprising both main effects and interactions are expected to be a more complete endeavor. Experimenting only with main effects in reliability engineering, nonadditive insights are bound to be omitted and the synergistic discovery is impossible to achieve, leading to imprecise screening/optimization of a process or product. Very strong effects appear to be manifested in all five distribution screenings, which means no matter of the specification of each reliability model, the presence of a particular active predictor will be detected irrespective of the specialized probabilistic mechanism that is set to detect it. On the other hand, effects that do not emerge as active in specific distributional screenings, for example like the CI interaction in the exponential and normal distribution screenings, might mean that it may not also influence a constant failure rate mechanism (exponential trend) or a symmetrical failure model with a mean time-to-failure possessing a low variation (normal trend).

5. Conclusions

In conclusion, this study constitutes an informative endeavor toward further decoding and understanding the problem of how to improve product life extension. A comprehensive case study on improving thermostat end-functionality with specific product/process interventions was considered, eventually aiming to generate beneficial modifications in order to prolong a product's lifespan. In the present work, the use of appropriate DOE models was successfully implemented, enabling an effective statistical analysis of the candidate interactions among the factors/interactions that could affect the thermostat's lifetime. This analysis led to the identification of the most significant candidate interactions, which were subsequently examined further with the aid of additional analytical tools. In contrast to past research, this work views the factorial controls on a system basis since it allows potential interactions to surface between different factors.

A major advantage of this approach is its applicability to any product for which it is desirable to regulate a group of controlling factors, suspected to lead to premature product failure. However, to achieve robust operational performance under real-life conditions, including data-imposing phenomena such as heteroscedasticity, censoring, and computational solver convergence, non-parametric methods may be tested for gaining further insight when encountering such technical limitations in reliability improvement projects.

The innovation of this study lies not only in the identification of the most critical candidate interactions, in contrast to the examination of isolated factors in previous research, but also in the subsequent analysis through a screening model focused exclusively on the selected most significant candidate interactions. A more precise determination of the most important interactions was achieved through the combined use of Pareto charts, Lenth test, and the lifetime response analysis of lifetime data using different regression models based on five different reliability distributions commonly accessed in statistical software packages. The determination of the strength of these candidate interactions, combined with a review of conclusions from previous studies, triggered further analysis using distributional screening of the effects. It was found that modeling the original lifetime dataset according to the Weibull, lognormal and exponential distributions provided similar screening results at comparable performances while maximizing the likelihood estimation. This work definitely demonstrates the multiplicity of reliability distributions which may offer similar profiling outcomes.

Author Contributions: Conceptualization, H.C.S. and G.B.; Methodology, G.B.; Validation, H.C.S., P.T. and G.B.; Formal analysis, H.C.S. and G.B.; Investigation, G.B.; Resources, H.C.S.; Data curation, H.C.S.; Writing—original draft, H.C.S.; Writing—review and editing, G.B.; Supervision, P.T. and G.B.; Project administration, P.T. and G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jose, S.A.; Tonner, A.; Feliciano, M.; Roy, T.; Shackleford, A.; Menezes, P.L. Smart manufacturing for high-performance materials: Advances, challenges, and future directions. *Materials* **2025**, *18*, 2255. [Google Scholar] [CrossRef] [PubMed]
2. Sahoo, S.; Lo, C.-Y. Smart manufacturing powered by recent technological advancements: A review. *J. Manuf. Syst.* **2022**, *64*, 236–250. [Google Scholar] [CrossRef]
3. Buttazzo, G.C. *Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications*; Springer: New York, NY, USA, 2011. [Google Scholar]
4. Box, G.E.P.; Hunter, W.G.; Hunter, J.S. *Statistics for Experimenters—Design, Innovation, and Discovery*; Wiley: New York, NY, USA, 2005. [Google Scholar]
5. Taguchi, G.; Chowdhury, S.; Wu, Y. *Quality Engineering Handbook*; Wiley: Hoboken, NJ, USA, 2004. [Google Scholar]
6. Jensen, P.B.; Laursen, L.N.; Haase, L.M. Barriers to product longevity: A review of business, product development and user perspectives. *J. Clean. Prod.* **2021**, *313*, 127951. [Google Scholar] [CrossRef]
7. den Hollander, M.C.; Bakker, C.A.; Hultink, E.J. Product design in a circular economy: Development of a typology of key concepts and terms. *J. Ind. Ecol.* **2017**, *21*, 517–525. [Google Scholar] [CrossRef]
8. Lawless, J.F. *Statistical Models for Lifetime Data*; John Wiley & Sons: Hoboken, NJ, USA, 2003. [Google Scholar]
9. Klion, J. Application of failure distributions to reliability. In *Practical Electronic Reliability Engineering*; Springer: Boston, MA, USA, 1992. [Google Scholar] [CrossRef]
10. Hackenhaar, I.C.; Moraga, G.; Thomassen, G.; Taelman, S.E.; Dewulf, J.; Bachmann, T.M. A comprehensive framework covering life cycle sustainability assessment, resource circularity and criticality. *Sustain. Prod. Consum.* **2024**, *45*, 509–524. [Google Scholar] [CrossRef]
11. Caroni, C. Regression models for lifetime data: An overview. *Stats* **2022**, *5*, 1294–1304. [Google Scholar] [CrossRef]
12. Nelson, W.B. *Applied Life Data Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 1982. [Google Scholar]
13. Condra, L. *Reliability Improvement with Design of Experiments*; CRC Press: New York, NY, USA, 2001. [Google Scholar]
14. Bullington, R.G.; Lovin, S.; Miller, D.M.; Woodall, W.H. Improvement of an industrial thermostat using designed experiments. *J. Qual. Technol.* **1993**, *25*, 262–270. [Google Scholar] [CrossRef]
15. Plackett, R.L.; Burman, J.P. The design of optimum multifactorial experiments. *Biometrika* **1946**, *33*, 305–325. [Google Scholar] [CrossRef]
16. Wu, C.F.J.; Hamada, M. *Experiments: Planning, Analysis, and Parameter Design Optimization*; Wiley: New York, NY, 2009. [Google Scholar]
17. Besseris, G.J. A methodology for product reliability enhancement via saturated–unreplicated fractional factorial designs. *Reliab. Eng. Syst. Saf.* **2010**, *95*, 742–749. [Google Scholar] [CrossRef]
18. Wang, G.; He, Z.; Xue, L.; Cui, Q.; Lv, S.; Zhou, P. Bootstrap Analysis of Designed Experiments for Reliability Improvement with a Non-Constant Scale Parameter. *Reliab. Eng. Syst. Saf.* **2017**, *160*, 114–121. [Google Scholar] [CrossRef]
19. Lv, S.; Niu, Z.; Cui, Q.; He, Z.; Wang, G. Reliability improvement through designed experiments with random effects. *Comput. Ind. Eng.* **2017**, *112*, 231–237. [Google Scholar] [CrossRef]
20. Besseris, G.; Tsarouhas, P. A new rapid and robust multi-factorial screening for low-percentile use-rate accelerated reliability analysis. *Int. J. Qual. Reliab. Manag.* **2022**, *39*, 316–342. [Google Scholar] [CrossRef]
21. Minitab, L.L.C. *Minitab, version 18*; Minitab, LLC: State College, PA, USA, 2017. [Google Scholar]
22. Wang, G.; Shao, M.; Cui, Q.; Lv, S.; He, Z.; Vining, G. Dual-response optimization for lifetime improvement experimental designs. *Comput. Ind. Eng.* **2021**, *158*, 107437. [Google Scholar] [CrossRef]
23. Lv, S.; Zhao, Y.; Li, S.; Wang, G.; Wang, X. Robust parameter design for constrained randomization lifetime improvement experiments. *J. Manag. Sci. Eng.* **2025**, *10*, 126–141. [Google Scholar] [CrossRef]
24. Besseris, G.J. A distribution-free method for full-factorial or fractionated reliability and yield trials. *Reliab. Eng. Syst. Saf.* **2026**, *265*, 111468. [Google Scholar] [CrossRef]
25. Wang, G.; Jiang, S.; Lv, S.; Zhao, Y.; Wang, X. A posterior preference articulation approach to reliability improvement experiments considering loss costs of manufacturer and customer. *Reliab. Eng. Syst. Saf.* **2026**, *266*, 111638. [Google Scholar] [CrossRef]
26. Montgomery, D.C. *Design and Analysis of Experiments*; Wiley: New York, NY, USA, 2020. [Google Scholar]

-
27. Rigdon, S.E.; Pan, R.; Montgomery, D.C.; Freeman, L.J. *Design of Experiments for Reliability Achievement*; Wiley: New York, NY, USA, 2022. [Google Scholar]
 28. Meeker, W.Q.; Escobar, L.A. *Statistical Methods for Reliability Data*; Wiley: New York, NY, USA, 1998. [Google Scholar]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.