Advances in Industrial and Manufacturing Engineering 2 (2021) 100036

Contents lists available at ScienceDirect

Advances in Industrial and Manufacturing Engineering

journal homepage: www.journals.elsevier.com/advances-in-industrial-and-manufacturing-engineering

Rapid and lean multifactorial screening methods for robust product lifetime improvement

George Besseris

Advanced Industrial and Manufacturing Systems, The University of West Attica and Kingston University, Greece

ARTICLE INFO

Keywords:
Life cycle improvement
Robust product reliability
Multifactorial product screening
Censored and accelerated industrial experiments
Lean
Leagile

ABSTRACT

Reliability enhancement is indispensable in modern operations. It aims to ensure the viability of complex functionalities in competitive products. We propose a *full-robust* screening/optimization method that promotes the rapid multi-factorial profiling of censored highly-fractionated lifetime datasets. The method intends to support operational conditions that demand quick, practical and economical experimentation. The innovative part of this proposal includes the robust split and quantification of structured lifetime information in terms of location and dispersion tendencies. To accomplish the robust data-reduction of lifetimes, maximum breakdown-point estimators are introduced to stabilize potential external-noise intrusions, which might be manifested as outliers or extremities. The novel solver provides resilience by robustifying the location (median) and dispersion (Rousseeuw-Croux Q_n) estimations. The proposed profiler fuses dichotomized and homogenized lifetime information in a distribution-free manner. The converted and consolidated lifetime dataset is non-parametrically pre-screened to ensure error balances across effects. Consequently, any strong effects that maximize the lifetime response are diagnosed as long as the error symmetry has been previously established. We discuss problems that may be encountered in comparison to other multi-factorial profilers/optimizers upon application to densely-fractionated-and-saturated experimental schemes. We comment on the lean and agile advantages of the proposed technique with respect to several traditional treatments for the difficult case that implicates small and censored survival datasets. The robust screening procedure is illustrated on an industrial-level paradigm that concerns the multi-factorial reliability improvement of a thermostat; the trial units have been subjected to conditions of censoring and use-rate acceleration.

1. Introduction

Innovation and quality are vital to elevating modern operations to peak performance (Al-Hakim and Jin, 2013; Maillard, 2015; ReVelle, 2001; Silva et al., 2014). Both strategies rely on an enterprise's data-driven capacity to rapidly generate and apply new knowledge (Bendoly et al., 2012). Operational maturity progresses through various phases that combine leagile (lean-and-agile) engineering philosophies and six-sigma quality initiatives (Balle et al., 2017; Cherrafi et al., 2016; Singh et al., 2017; Vinodh et al., 2008; Virmani et al., 2018). A core manufacturing priority is the reliability improvement of intricate processes/products. This is because the continuous minimization of process waste and product failures leads to robust operational behavior and, hence, to promising product placement in the markets. Consequently, overall brand profitability tends to grow (King and Jewett, 2010; Mackelprang et al., 2015). The Six Sigma toolbox provides several pathways to design quality and reliability into innovative products (Cudney and Agustiady, 2016; Pyzdek and Keller, 2014). However, in manufacturing, we often encounter non-normal processes because of either their inherent nature or their low operational maturity, or even both. To predict and improve the reliability performance status of non-normal processes remains a challenging subject in Six Sigma (Aldowaisan et al., 2015). Thus, the discovery of new reliability screening/optimization techniques justifiably merits further exploration such as to encompass real-life processes, which are governed by strong non-normal tendencies.

Design of Experiments (DOE) furnishes the main arsenal for carrying out reliability screening and optimization studies (Condra, 2001). Reliability improvement requires specialized DOE approaches, which harmoniously enmesh classical reliability theory (Lawless, 2003; Meeker and Escobar, 1998) with traditional experimental planning and analysis (Box et al., 2005; Taguchi et al., 2004). Maximizing the product lifetime response is the crux in reliability analysis. While investigating lifetime trends, an experimenter is often confronted with the practical need to

E-mail address: besseris@uniwa.gr.

https://doi.org/10.1016/j.aime.2021.100036
Received 5 August 2020; Received in revised form 24 December 2020; Accepted 18 February 2021

screen and optimize product lifetimes that may be much longer than the allotted duration of the improvement project. To shorten the trial time span and accelerate the data collection procedure, censored trials (Collins et al., 2013; Nelson, 1990) and use-rate acceleration (Meeker and Escobar, 1998) are designed into the sampling scheme.

In a typical manufacturing environment, a comprehensive list of factors should be initially considered for a factor-screening effort to be effective. To improve product lifetimes, voluminous experiments are usually anticipated (Bhote and Bhote, 2004). However, the exorbitant number of trials may be viewed as a formidable obstacle to expediting a reliability improvement project. The necessity for discovery is counteracted by substantial costs in wasted test materials and labor hours. Moreover, real screening/optimization studies should be directly conducted on a production line to ensure that any predictions are to be meaningful and applicable. Thus, additional costs are to be realized from interrupting the production flow in order to divert operational availability to industrial trials. Thereupon, early researchers in reliability engineering resorted to conducting fractionated trials (Hamada and Wu, 1991, 1995) in order to: 1) reduce the total volume and duration of the experiments, 2) minimize operational unavailability, 3) curtail research expenditures, and 4) accelerate decision-making. Highly efficient plans rely on 'rapid-and-dense' structured sampling, which is furnished by fractional factorial design (FFD) schemes (Box et al., 2005). FFDs play the role of standard trial recipe planners in Six Sigma (Pyzdek and Keller, 2014). A popular class of FFDs which is suitable for industrial product/process improvement projects are the Taguchi-type orthogonal arrays (OAs) (Taguchi et al., 2004). Past efforts combined Taguchi-type OAs with survival analysis to provide a basic framework for multi-factorial screening/optimization solvers (Hamada 1993, 1995). Such approaches mainly demand on previously determining a proper parametric reliability distribution, which should convincingly represent the fitted lifetime OA-dataset. Ordinarily, the Weibull and the lognormal distributions have been considered as the primary candidate reliability functions of choice (Bullington et al., 1993; Joseph and Yu, 2006; Lio et al., 2015; Lv et al., 2017; Piña-Monarrez and Ortiz-Yañez, 2015; Wang et al., 2017). Of course, there are many other families of reliability functions that should not be disregarded. This is the reason that professional statistical software packages are equipped to carry out, as a preliminary phase, a distribution identification analysis, before proceeding to any formal prediction studies. Since fractionated trials are often sped-up by keeping short the extent of replication, a distribution identification session might not always be fruitful. We may not overlook cases where samples are susceptible to distribution multiplicities, or even, on the other end, to indeterminate distributions. Moreover, it might even occur that miscellaneous types of distributions may be involved in order to describe the sample lifetime data across the various OA-recipes. The intermixing of multifarious known and even unresolvable parametric distributions might lead to "messy" datasets (Milliken and Johnson, 2004). Messy data are intriguing because of their association to real complex processes. Messy data analysis may be arduous. Ostensibly, order statistics may be more suitable in interpreting complex phenomena than standard t-distribution statistics (Ludbrook and Dudley, 1998). Robust estimators, which deliver maximum breakdown-point performance, become imperative in the treatment of messy data (Wilcox, 2010). Recently, specialized nonparametric treatments have been recommended to aid in delineating the behavior of small non-normal samples (Pett, 2015; Siebert and Siebert, 2017).

Following Taguchi's approach to robust design, it is instructed that the factor analysis be split in terms of location and dispersion measures (Taguchi et al., 2004). The suggested measures for central tendency and variability are the response mean and the signal-to-noise ratio (SNR), respectively (Taguchi et al., 2000). For the mean and the SNR estimations to be valid, the sample size of the replicated data should be adequate and the behavior of the collected observations ought to adhere to normality. Otherwise, both estimations might lead to weak or even spurious predictions (Silver, 2015). Unfortunately, in survival analysis, data normality is not guaranteed. Another aspect that complicates matters is that lifetime samples are usually expected to be available in suboptimal size - due to practical and economic constraints – as it was mentioned earlier. An approximate minimum sample size, which is encountered in reliability research, is in the vicinity of ten observations (Darby, 2010).

By saturating a selected OA scheme, maximum utilization of the experimental plan is achieved. As a result, the selected design accommodates the maximum number of controlling factors that it could handle (Taguchi et al., 2004). From an operational standpoint, it is a desirable condition, because saturation also presents an extra opportunity to minimizing further the trial expenditures. In Taguchi methods, the uncontrolled noise is captured by summarizing the replicate behavior in terms of the mean and SNR estimations (Taguchi et al., 2004). This data-reduction step converts the OA-sampled data-matrix into two component vectors (mean and SNR), each posing in an 'unreplicated' form. Regular multifactorial treatments, such as the analysis of variance (ANOVA) and the general linear regression modeling (GLM), cannot process OA-datasets in a 'saturated-and-unreplicated' state. All available degrees of freedom are strictly assigned to size the magnitudes of the effects (Box et al., 2005). Consequently, an estimation of the aleatory and epistemic uncertainties may not be attained (Briggs, 2016). The problem is exacerbated when there is also involvement of messy data (Milliken and Johnson, 1989).

In 'saturated-and-censored' OA trials, the synchronous conditions of sample 'smallness' and data 'messiness' may undermine the capability of a regular multifactorial solver to deliver robust results. Robustness cannot be relinquished in operational environments that expect high performance (Roy, 2010). A novel multifactorial reliability profiler is proposed to permit the robust lifetime screening/optimization, while simultaneously incorporating location and dispersion information. This aspect may be an advance with respect to the classical Taguchi methodology. In Taguchi methods, two separate searches for active effects are conducted – based on their influence on either the location or the dispersion tendencies of the response. Thus, two separate groups of strong effects are identified. Then, an empirical 'mending' of the two separate predictions is attempted to arrive to a joint (compromised) solution. To fortify the lifetime data-conversion process, we substitute the mean and SNR estimators with two robust estimators. The median replaces the mean as a preferred location estimator (Hoaglin et al., 2000). The median sustains the maximum attainable breakdown point value at 50% (Huber and Ronchetti, 2009). Thus, it robustly stabilizes the response location estimations. On the contrary, the mean is unguarded against underlying skewed or messy data trends, since its breakdown-point performance plummets to the minimum attainable value of 0%. Similarly, the scale estimator Q_n replaces the scale SNR measure (Rousseeuw and Croux, 1993). This action is necessary because the SNR estimator carries the minimum possible breakdown point value of 0%. Instead, the Q_n estimator retains the maximum possible breakdown point value of 50%, even for asymmetrical distributions.

To recapitulate, we attempt to process a replicated 'censored-and-saturated' lifetime OA-dataset, which is dichotomized and reduced into a form of two robust measure vectors. The median vector and the Q_n-estimator vector represent the response location and dispersion components. It is a succinct information structure as long as the two vector components are not correlated with each other. If it is verified that they are uncorrelated, the two vectors may be treated with an appropriate analyzer, which handles multiple responses. The robust multi-factorial screening and optimization of multiple responses is a challenging task because it requires specialized solvers (Gabrel et al., 2014; Kim and Lin, 2006). In dealing with 'saturated-unreplicated-and-censored' OA-datasets, the robust solver should be equipped to assess the symmetry status of the residual errors across all the examined factor-settings. Otherwise, the predicted group of strong effects may be misleading or even erroneous. The desirability analysis (DA) is a popular method that is employed to tackle multi-fractionated multi-response datasets (He et al., 2012; Jeong and Kim, 2009). However, the DA has not been extensively

Input(factorial recipes) **Output** (lifetime replicates)

$$\begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \qquad \begin{pmatrix} lt_{11} & \cdots & lt_{1r} \\ \vdots & \ddots & \vdots \\ lt_{n1} & \cdots & lt_{nr} \end{pmatrix}$$

Fig. 1. Compact OA arrangement of *n*-experimental recipes, *m*-controlling factors and *r*-replicated lifetimes.

studied in reliability improvement applications, which involve censoring and non-normal responses. In general, the DA is considered a practical - but also subjective - approach. A primary concern is that the DA uses a composite desirability score to evaluate the "goodness of the optimization" (Derringer and Suich, 1980). The concept of the desirability-score performance has not been shown to correspond to a statistical significance according to a reference law. Ergo, the DA predictions are not gaged against a probabilistic distribution. A second issue is that the DA uses ordinary regression fitting techniques to model the examined OA-datasets. As it was mentioned earlier, the condition of 'unreplication-saturation-and-censoring', if it is imposed on OA-datasets, will not permit the estimation of the statistical significance of the regression coefficients. This is also extended to imply that the main assumptions of the regression analysis will not be testable. Moreover, a quantification of the coefficient of determination is not feasible. The emphasis on the effectiveness of the multi-factorial analysis of censored lifetime data has been placed on preselecting the right parametric model (Wang et al., 2017). In the past, solutions have been suggested that track optimal survival behavior by directly profiling the split OA-data in terms of Weibull-type shape and scale parameters (Besseris, 2010). For small samples and messy datasets the constant shape-parameter assumption in Weibull regression may not be detectable (Mueller and Rigdon, 2015). Thus, a more general modeling approach may be intriguing, if it permits the data-conversion process with no prior knowledge about a particular parametric reference law. This motivates us to propose a distribution-free multi-factorial method to treat censored lifetime experiments. It is a leagile approach. It is 'lean', because it simplifies the data-processing phase by proposing a non-iterative solver. Hence, it automatically reduces computational work. It is also 'lean', because there are no binding assumptions for posterior checking, as it is required, for instance, when implementing ANOVA and GLM treatments. So, the intermediate work of examining the validity of the assumptions is eliminated. Finally, it is a 'lean' technique, because it eliminates the work for searching and establishing a parametric reliability model prior to conducting the multifactorial analysis. On the other hand, it is 'agile' because it is applicable to any survival distribution, while being adaptable and responsive to any censored OA-dataset. The new methodology is elucidated by undergoing a reliability improvement study for an industrial-level thermostat. The previously published dataset has been selected because it highlights a use-rate accelerated and censored (highly fractionated) experimental plan for small sampling that simultaneously accommodates a mix of several numerical and categorical controlling factors (Bullington et al., 1993). Moreover, the interpretation of the underlying phenomena in the non-normal saturated datasets continues to attract attention today (Lv et al., 2017; Wang and Kececioglou, 2000; Wang et al., 2017; Wu and Hamada, 2009). The solution is compared with alternative approaches and their respective predictions. We comment on advantages and disadvantages at several points during the analysis process.

2. Methodology

The objective is to screen and/or improve a product's reliability through mini trials. A 'mini' endeavor is construed to be motivated by practical and economic incentives. We contemplate the acceleration of trials by: 1) reducing the total required trial volume, 2) lifetime censoring and 3) use-rate speed-up. The reduction of the planned experimental work is accomplished by implementing structured FFDs/OAs (Box et al., 2005; Taguchi et al., 2004). We examine datasets that have been generated by the general two-level OAs under the conditions of: 1) a limited number of replicated runs (small lifetime samples), and 2) censored (Type I) lifetime observations - to a predefined cutoff value (T_o). To have a typical robust estimation of a sample, we minimally need at least five observations – to form a boxplot (Hoaglin et al., 2000). In particular to survival studies, this low limit may be about ten - as many as the deciles - as reported in high-criticality publications (Darby, 2010).

The OAs are fractional factorial matrices of the (i,j) type. The $OA_n(2^m)$ tables are quantized arrays (Besseris, 2013) of all-purpose ('ready--to-use') recipes, which allow the manipulation of as many as m controlling factors (columns) by limiting the number of required recipes (rows) to a maximum of n ($<2^m$) combinations. We require the maximum utilization of an OA trial-plan, which is achieved by saturating all available array columns. This means that the OA-scheme permits information collection from the maximum number of effects (m) it can handle, with no additional overhead in costs and time. At saturation point, it holds that $n = m + 1$; this indicates that all degrees of freedom are solely assigned to the examined effects. Consequently, no residual error can be estimated with ordinary treatments, i.e. methods based on ANOVA or GLM. We label the group of investigated factors as: $\{X_1, X_2, \ldots, X_m\}$. Their respective predetermined settings on a selected saturated (i,j)-OA, are denoted, then, as: $\{(x_{i1}, x_{i2}, \ldots, x_{im}) \in \Re \mid i = 1, 2 \ldots n\}$. Accelerated failure tests are to be replicated r times per trial recipe. The resulting r-replicated lifetime entries are symbolized as: $\{(lt_{1j}, lt_{2j} \ldots lt_{nj}) \in \Re \mid j = 1,2, \ldots, r\}$. A concise depiction of a typical input-output OA arrangement, which the factors and replicated lifetime responses are positioned on the left- and right-hand side of the design, respectively, is provided in Fig. 1. The proposed data analyzer does not require that a parametric reliability framework have been previously established. The focal assumption of the central limit theorem is naively absent from the backbone of the proposed developments. We generalize in the proposed 'framework' to encompass such cases where there might not be available a common model to describe the replicate dataset, which is generated from all executed OA-recipes.

In other words, different OA-recipes might generate samples that map to different reliability models. Furthermore, the data processing is not impeded by other complicating conditions, which may be exacerbated by the restricted sampling tactics. The analyzer is apt to resolve relationships under the more extreme conditions of: 1) an indeterminate distribution and 2) multiplicities, i.e. several distributions fitting the same sample data. Besides this agility to convert even pragmatic 'messy' datasets, the proposed processor may offer more tangible gains in comparison to ordinary multifactorial treatments. Comparing to ANOVA, the proposed analyzer is built on a distribution-free perspective to waive ANOVA's two main pre-processing assumptions: 1) the normality of residual errors, and 2) the homoscedasticity. Similarly, comparing to GLM, the proposed analyzer is relieved from the assumption-verification step and its concomitant data post-processing. Namely, the post-processing graphical work that is eliminated is: 1) the histogram of residuals, 2) the normal plot of residuals, 3) the residuals-vs-fits plot and 4) the residuals-vs-order plot. Therefore, the processor simplifies the analysis

(Censor-differenced lifetimes) (Data reduction) (Rank ordering) (Rank fusion process)

$$\begin{pmatrix} T_o - lt_{11} & \cdots & T_o - lt_{1r} \\ \vdots & \ddots & \vdots \\ T_o - lt_{n1} & \cdots & T_o - lt_{nr} \end{pmatrix} \rightarrow \begin{pmatrix} \widehat{M}_1 \\ \vdots \\ \widehat{M}_n \end{pmatrix} \begin{pmatrix} \widehat{Q}_{n1} \\ \vdots \\ \widehat{Q}_{nn} \end{pmatrix} \rightarrow \begin{pmatrix} rm_1 \\ \vdots \\ rm_n \end{pmatrix} \begin{pmatrix} rq_1 \\ \vdots \\ rq_n \end{pmatrix} \rightarrow \begin{pmatrix} rm_1^2 + rq_1^2 \\ \vdots \\ rm_n^2 + rq_n^2 \end{pmatrix} \rightarrow \begin{pmatrix} ssr_1 \\ \vdots \\ ssr_n \end{pmatrix}$$

Fig. 2. Data reduction, estimator rank ordering and rank fusion of the lifetime dataset of Fig. 1.

cycle by designing in it the modern operational stipulation of lean engineering.

In lack of a direct estimation of the residual error, early attempts, in screening unreplicated-saturated FFD-datasets, introduced new types of assumptions to resolve this anomaly. One assumption was that the number of the strong factors is small in contrast to the rest of the investigated effects (Daniel, 1959). This is called the sparsity assumption. Moreover, the 'larger' inactive group of factors was assumed to also obey normality. The separation of active and inactive effects was inspected visually on a half-normal plot. Then, a simple filtering of the strong effects was empirically determined. There was no direct quantification of the statistical significance of the outcome, though. A different approach, which maintained the sparsity assumption, promoted an estimator for a 'pseudo standard' error in order to size the variation of any strong effects (Lenth, 1989). To truncate the examined contrasts, the Lenth approach had to introduce two constants. In the formulation, the constants were fixed in all cases, even though there is no evidence of their universality. To control the error rate for each individual effect, the t-distribution was assumed to be the reference law. Subsequently, in an arbitrary fashion, one-third of the available degrees of freedom are accounted toward an estimate of the t-score. Overall, both methods (the half-normal plot and Lenth test) are the prevailing options for screening in mainstream statistics software. The above reality offered an opportunity to experiment with a novel 'lean-and-agile' solver. Thus, the proposed analyzer was designed to rid of the sparsity assumption, be constant-free, and distribution-free. To summarize 'small-and-messy' lifetime sample data, a proper robust measure that captures the central tendency of the replicate propensities is imperative.

A simple resistant location estimator – with maximum breakdown point efficiency - is the median, which is defined as follows (Hoaglin et al., 2000; Huber and Ronchetti, 2009; Milliken and Johnson, 2004):

$$\widehat{M}\mathrm{i} = \mathrm{median}\left\{ lt_{\mathrm{i1}}, lt_{\mathrm{i,2}}, \ldots, lt_{\mathrm{ir}} \right\} \quad \left| \quad \widehat{M}\mathrm{i} \in \Re \;\; \forall\, \mathrm{i} = 1, 2, \ldots, \mathrm{n} \right. \tag{1}$$

This is advantageous when fending-off against outliers and extremities in inherently skewed data; it is often the norm in survival studies. Similarly, we select the Q_n-estimator to provide a resistant dispersion measure for the replicate data; it also ensures maximum breakdown-point efficiency in the presence of asymmetric data distributions. It is defined as follows (Rousseeuw and Croux, 1993):

$$\widehat{Q}_{\mathrm{ni}} = \mathrm{d}\left\{ \left| lt_{\mathrm{io}} - lt_{\mathrm{ip}} \right|;\; 1 \le \mathrm{o} \le \mathrm{p} \le \mathrm{r} \right\}_{(\mathrm{k})} \quad \left| \quad \mathrm{d} \in \Re,\; \widehat{Q}_{\mathrm{ni}} \in \Re \;\; \forall\, \mathrm{i} = 1, 2, \ldots, \mathrm{n} \right. \tag{2}$$

The *k*th ordered value is given by:

$$k = \binom{h}{2} \text{ with } h = r \Big/ 2 + 1 \tag{3}$$

It is an attractive feature that both estimators, the median and Q_n, are conveniently computed. We proceed by taking the difference of each lifetime data entry (output in Fig. 1) from the censoring limit, T_o (Fig. 2). Using equations (1)–(3), we reduce the censor-differenced lifetime matrix (Fig. 2) to the two vector measures according to the median and the Q_n estimators (Fig. 2).

It is obvious that the *r*-replicated lifetime dataset has been condensed down to two individual single-column ("unreplicated") responses; they represent the robustified location and dispersion properties of the original lifetime dataset. To retain both vectors in the ensuing consolidation, the two vectors should not be correlated. If it is found that they are correlated, we drop any one of the two vectors in the output arrangement. Then, the problem is cast to the simpler 'unreplicated-saturated' case (Besseris, 2013). The two vectors ought to be checked for dependence with regular regression methods. Returning to the analysis procedure for the uncorrelated case, the median vector entries (Fig. 2) are ordered to the rank vector $rm_i \in \Re \mid i = 1, 2, \ldots, n$ (Fig. 2). Similarly, the $\widehat{Q}_{ni}$ vector entries are rank-transformed to the new vector $rq_i \in \Re \mid i = 1, 2, \ldots, n$. To reach to an optimal selection of the examined factors, the optimization direction aligns with the synchronous minimization of both (and $\widehat{Q}_{ni}$) estimators. This is because the minimum median estimation of the censor-differenced lifetimes is minimized at the censoring limit, i.e. to value of 0. Likewise, by minimizing $\widehat{Q}_{ni}$, we minimize the fluctuation of the lifetime values toward the censoring limit. By default, in either case, ranks ascend by starting from the entries possessing the smallest magnitude in $\{\widehat{M}_i\}$ and $\{\widehat{Q}_{ni}\}$. The lowest rank reflects the greatest proximity to the respective goal. Squaring and summing the ranks of the two ordered estimators uniformly fuses and concurrently aggregates their constituent tendencies. The sum of squared ranks creates the new variable, $ssr_i \in \Re \mid i = 1, 2, \ldots, n$ (Fig. 2). This has practical meaning because we seek to identify those effects that concurrently influence the maximization of the reliability status. The restructuring of the original input-output relationship in terms of the 'ranked-and-fused' output is shown in concise from in Fig. 3.

Input **Output**

factorial recipes ranked-and-fused response

$$\begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} \quad \begin{pmatrix} ssr_1 \\ \vdots \\ ssr_n \end{pmatrix}$$

Fig. 3. Compact OA arrangement of *n*-experimental recipes with *m*-controlling factors and ranked-and-fused $\widehat{M}_i$ and $\widehat{Q}_{ni}$ estimators (ssr_i).

The input-output OA arrangement in Fig. 3 is in the 'saturated-unreplicated' form. This condition demands a specialized tool to gauge the effect strengths of the examined controlling factors. We utilize a "block-and-profile" surrogate tool (Besseris, 2013) in order to detect uncertainty asymmetries from across factor-settings. This is a crucial step that should be completed before computing and awarding significance to the investigated effects. Briefly, we reiterate the key processing steps. First, we 'meta-dimensionalize' $\{ ssr_i \}$ to $\{ ssr_{i_1,i_2,\ldots,i_m} \}$. In this manner, transparency improves since it allows direct tracking of the setting combinations for all *m* factors. The linear model that is considered is:

$$ssr_{i_1,i_2,\ldots,i_m} = M + \sum_{j=1}^{m} D_j + \varepsilon_{i_1,i_2,\ldots,i_m} \tag{4}$$

The residual error, $\varepsilon_{i_1,i_2,\ldots,i_m}$, consists of a random error plus any other spontaneous unknowable intrusions. The indices $i_1, i_2, \ldots, i_m$ are symbolically 'binary', coded as '-' and '+', to represent the factor-setting endpoints. The grand median, *M*, the median for each factor-setting,

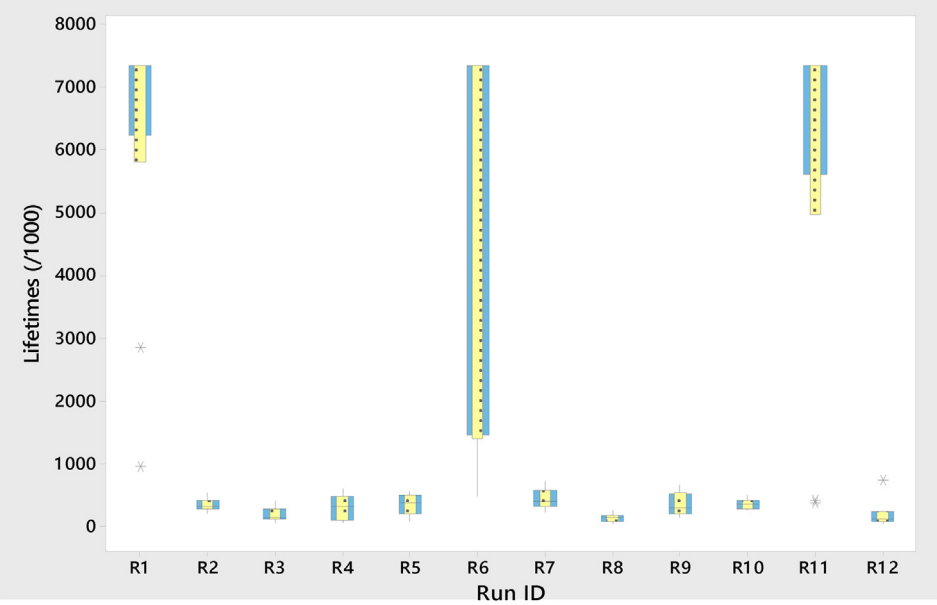


Fig. 4. Boxplot screening of the 12 thermostat-lifetime OA-datasets - censored at 7342 (k-cycles).

M_l, and their associated differences, D_l, in equation (4) are:

$$M = Med(\{\ ssr_{i_1,i_2,\ldots,i_m}\}) \text{ forall } i_1, i_2, .., i_m \tag{5}$$

$$M_l = \left\{ \begin{array}{ll} M_l^+ = Med(\{LT_{i_1,i_2,\ldots i_l,\ldots,i_m}\}) & \text{if } i_l \rightarrow + \\ M_l^- = Med(\{LT_{i_1,i_2,\ldots i_l,\ldots,i_m}\}) & \text{if } i_l \rightarrow - \end{array} \right\}$$
$$\text{for all } i_1, i_2, \ldots, i_{l-1}, i_{l+1}\ldots, i_m \tag{6}$$

$$D_l = \left\{ \begin{array}{ll} D_l^+ = M_l^+ - M & \text{if } i_l \rightarrow + \\ D_l^- = M_l^- - M & \text{if } i_l \rightarrow - \end{array} \right\} \tag{7}$$

The error asymmetry is quantified through the distribution-free error balances:

$$ssr''_{i_1,i_2,\ldots,i_m} = M + \varepsilon_{i_1,i_2,\ldots i_l,\ldots i_m} \text{ for all } i_l \text{ and } 1 \le l \le m \tag{8}$$

The error balances are rank-ordered:

$$ssr''_{i_1,i_2,\ldots,i_m} \rightarrow r'_{i_1,i_2,\ldots i_l,\ldots i_m} \text{ for all } i_l \text{ and } 1 \le l \le m \tag{9}$$

The minimum rank-sum (Wilcoxon, 1945) of error balances is simply computed for each considered factor setting (Besseris, 2013).

$$TE_l = Min \left\{ \begin{array}{ll} RSE_l^+ = \sum_{i_l} r'_{i_1,i_2,\ldots i_l,\ldots,i_m} & \text{if } i_l \rightarrow + \\ RSE_l^- = \sum_{i_l} r'_{i_1,i_2,\ldots i_l,\ldots,i_m} & \text{if } i_l \rightarrow - \end{array} \right\}$$
$$\text{for all } i_1, i_2, \ldots, i_{l-1}, i_{l+1}\ldots, i_m \tag{10}$$

The minimum rank-sum is sized on the Wilcoxon-Mann-Whitney reference scale (Mann-Whitney, 1947; Wilcoxon, 1945) to deliver the corresponding statistical significance (p-value). Then, the p-value is contrasted against a standard error rate, i.e. $\alpha = 0.05$. If the error asymmetry is statistically significant (p-value $< \alpha$), the uncertainty intrusions may interfere with and hence mar the effect strength estimation in the profiling process. Then, for that particular controlling factor, we may infer that the predicted strength size might not be reliable. Exact p-values are obtained using the Mann-Whitney test from the software package MINITAB® (v.18). The reconstructed one-factor responses are:

$$ssr'_{i_1,i_2,\ldots,i_m} = M + D_l + \varepsilon_{i_1,i_2,\ldots i_l,\ldots i_m} \text{ for all } i_l \text{ and } 1 \le l \le m \tag{11}$$

They also receive a rank-ordering to obtain:

$$ssr'_{i_1,i_2,\ldots,i_m} \rightarrow r_{i_1,i_2,\ldots i_l,\ldots i_m} \text{ for all } i_l \text{ and } 1 \le l \le m \tag{12}$$

Reorganizing the effects in terms of the minimum rank-sums, we finally obtain:

$$T_l = Min \left\{ \begin{array}{ll} RS_l^+ = \sum_{i_l} r_{i_1,i_2,\ldots i_l,\ldots i_m} & \text{if } i_l \rightarrow + \\ RS_l^- = \sum_{i_l} r_{i_1,i_2,\ldots i_l,\ldots i_m} & \text{if } i_l \rightarrow - \end{array} \right\} \text{ for all } i_1, i_2, \ldots, i_{l-1}, i_{l+1}\ldots, i_m \tag{13}$$

The statistical potency of the effects is measured against the Wilcoxon-Mann-Whitney reference distribution. The exact p-values are again computed by the software package MINITAB® (v.18), which also allows for corrections in the case of tied values. The p-value performance of the strong effects will be controlled at a false discovery rate of $\alpha = 0.05$ (Benjamini and Hochberg, 1995). In a nutshell, the methodology may be paced in three distinct stages: 1) a distribution identification screening (optional), 2) an uncertainty symmetry screening, and 3) an effect strength profiling (screening/optimization).

3. Results

3.1. Data pre-screening

The original thermostat lifetime dataset featured a twelve-run eleven-factor FFD problem (Bullington et al., 1993). Each of the twelve independent samples consisted of ten lifetime observations. We opt to pre-screen each of the twelve samples in order to identify a parametric reliability distribution that could best fit each. An immediate comment is that all twelve datasets may be discerned in two groups: 1) those datasets that include some observations which are truncated at the censoring limit ('censored'), and 2) those datasets that contain no truncated observations at all ('uncensored'). In Fig. 4, we display the twelve samples in a box plot screening (MINITAB 18). Three runs generated data, which had to be truncated at the censoring limit (7342 k-cycles). This 'two-group' discrimination is instructive because it allows releasing information with different accuracy potential. Several types of reliability distributions may be tried to fit the 'uncensored' datasets. Their goodness of fit may be objectively evaluated by ordinary tests, such as the Anderson-Darling (AD) test. On the hand, 'censored' datasets need to be treated with the adjusted Anderson-Darling (aAD) test, which relays no information on

Table 1
Goodness of fit for 14 ordinary reliability distributions (for the 9 uncensored datasets).

	Recipe Number																	
	R2		R3		R4		R5		R7		R8		R9		R10		R12	
Distribution	**AD**	**P**	**AD**	**P**	**AD**	**P**	**AD**	**P**	**AD**	**P**	**AD**	**P**	**AD**	**P**	**AD**	**P**	**AD**	**P**
Normal	0.311	0.5	0.441	0.229	0.247	0.674	0.331	0.442	0.338	0.423	0.207	0.815	0.384	0.323	0.194	0.855	1.362	<0.005
Lognormal	0.229	0.74	0.232	0.73	0.431	0.243	0.598	0.086	0.224	0.76	0.167	0.91	0.247	0.673	0.164	0.915	0.32	0.47
3-Parameter Lognormal	0.232	0.97	0.249	0.987	0.321	0.542	0.376	0.105	0.24	0.856	0.178	0.862	0.3	0.854	0.197	0.833	0.275	0.088
Exponential	2.683	<0.003	1.259	0.049	0.789	0.187	1.302	0.043	2.157	0.005	1.471	0.027	1.187	0.06	2.92	<0.003	0.715	0.235
2-Parameter Exponential	0.983	0.042 0.000	0.549	0.223 0.017	0.46	>0.250 0.066	0.863	0.067 0.049	0.641	0.15 0.000	0.456	>0.250 0.005	0.343	>0.250 0.005	0.487	>0.250 0.000	0.388	>0.250 0.03
Weibull	0.357	>0.250	0.324	>0.250	0.343	>0.250	0.47	0.227	0.333	>0.250	0.172	>0.250	0.308	>0.250	0.242	>0.250	0.637	0.083
3-Parameter Weibull	0.27	>0.50 0.238	0.292	>0.500 0.258	0.3	>0.500 1	0.306	0.467 0.476	0.258	>0.500 0.213	0.235	>0.500 0.297	0.278	>0.500 0.395	0.18	>0.500 0.267	0.304	>0.500 0.03
Smallest Extreme Value	0.589	0.11	0.644	0.08	0.284	>0.250	0.305	>0.250	0.521	0.176	0.404	>0.250	0.602	0.099	0.362	>0.250	1.747	<0.010
Largest Extreme Value	0.235	>0.25	0.301	>0.250	0.352	>0.250	0.464	0.233	0.249	>0.250	0.177	>0.250	0.297	>0.250	0.204	>0.250	0.734	0.045
Gamma	0.243	>0.250	0.281	>0.250	0.378	>0.250	0.525	0.203	0.271	>0.250	0.166	>0.250	0.281	>0.250	0.187	>0.250	0.609	0.134
3-Parameter Gamma	0.583	1	0.404	1	0.304	1	0.942	0.723	0.365	1	0.208	1	0.317	1	0.206	1	0.372	0.066
Logistic	0.269	>0.250	0.432	0.232	0.279	>0.250	0.36	>0.250	0.348	>0.250	0.209	>0.250	0.374	>0.250	0.199	>0.250	0.92	0.008
Loglogistic	0.209	>0.250	0.251	>0.250	0.443	0.219	0.562	0.093	0.251	>0.250	0.19	>0.250	0.26	>0.250	0.181	>0.250	0.255	>0.250
3-Parameter Loglogistic	0.208	0.93	0.246	0.805	0.305	0.641	0.361	0.157	0.238	0.714	0.186	0.955	0.304	0.73	0.191	0.83	0.193	0.28
***LRT P values (p-values in red font)**																		

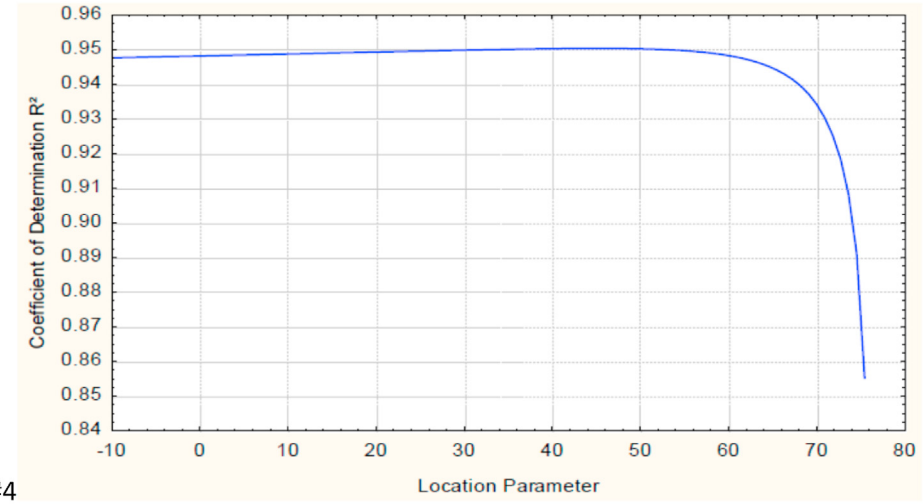
Table 2
Goodness of fit for eleven ordinary reliability distributions (three censored dataset cases).

Distribution	Recipe ID		
	R1	R6	R11
	adj AD	adj AD	adj AD
Weibull	36.85	43.12	36.86
Lognormal	36.85	43.12	36.86
Exponential	36.85	43.15	36.91
Loglogistic	36.85	43.12	36.86
3-Parameter Weibull	36.85	43.13	36.86
3-Parameter Lognormal	36.86	43.14	36.86
2-Parameter Exponential	36.86	43.13	36.86
3-Parameter Loglogistic	36.86	43.14	36.86
Smallest Extreme Value	36.85	43.14	36.86
Normal	36.85	43.14	36.86
Logistic	36.85	43.14	36.86

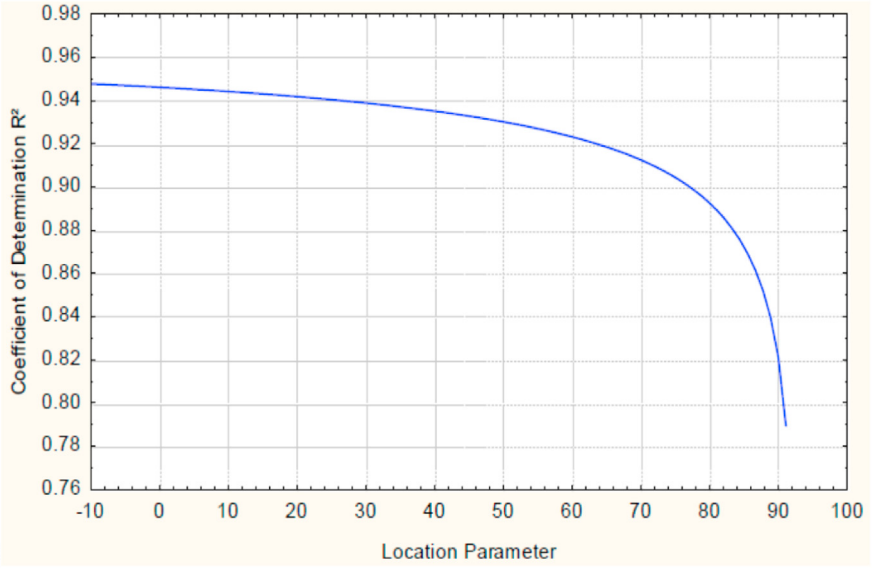
the significance of the results. In Table 1, we tabulate the goodness of fit for 14 common reliability distribution functions (MINITAB 18), for the nine 'uncensored' datasets (runs #2–5, 7–10, 12). We have also included the p-value (red font) of the likelihood-ratio test (LRT-P) to indicate any potential correction if we decide to add a third parameter in the fitted model. The primary disposition is that several candidate distributions might achieve comparable performance, leading to similarly low AD scores. For example, we may elaborate on the results regarding to the sample behavior of run #2 (R2-dataset). We notice that the goodness of fit is statistically indistinguishable among the following functions: lognormal, 3-parameter lognormal, 3-parameter Weibull, largest extreme value, gamma, logistic, log-logistic and 3-parameter log-logistic. In all eight models, their estimated AD scores are in the vicinity of 0.2. This multiplicity of equivalent models might be perplexing. In lack of specifying a parametric reference law, conventional effect-profiling attempts might result to dubious outcomes. Surprisingly, analogous trends emerge when fitting the three 'censored' datasets (runs # 1, 6 and 11). From Table 2, we notice the perfect agreement among all of the estimated aAD scores of the three datasets - for all fitted distributions. There is a conspicuous kind of bias that emerges when the number of the survived test units exceed the number of failed units in type I censored experiments. We observe that the number of censored observations dictates the fitting performance. Different trials with equal number of censored observations perform similarly regardless of: 1) the variability of the failed units (runs #1 and 11) and the type of the fitted reliability model. Moreover, the number of censored observations regulates the fitting performance between trials (runs #1 or 11 against run #6). While a sample size of 10, for each run, may be generally valid for normal and non-normal distributions (Dodson, 1994; Jantschi and Bolboaca, 2018; Marsaglia, 2004), the validity of the parametric estimations rely on the validity of the selected reference law. Finally, it is illuminating to contemplate how individual recipes drastically influence lifetime parametrization and model uncertainty. We exemplify the diverse trends of the goodness-of-fit for a typical 3-parameter Weibull model; it has been applied to three separate runs (R4-, R5-and R6-datasets). Plotting the coefficient of determination (R^2) versus the location parameter, in Fig. 5, we discern three distinctly different tendencies:

1) The variable R^2 may experience a plateau. Thus, the uncertainty range for the location estimation (Run #4) is uniformly broadened; the location estimation also includes a zero value.
2) The variable R^2 may be 'optimally' located 'anywhere' on the tracing curve (Run #5).
3) The variable R^2 may reach a peak value for a given location value that minimizes uncertainty (Run #6).

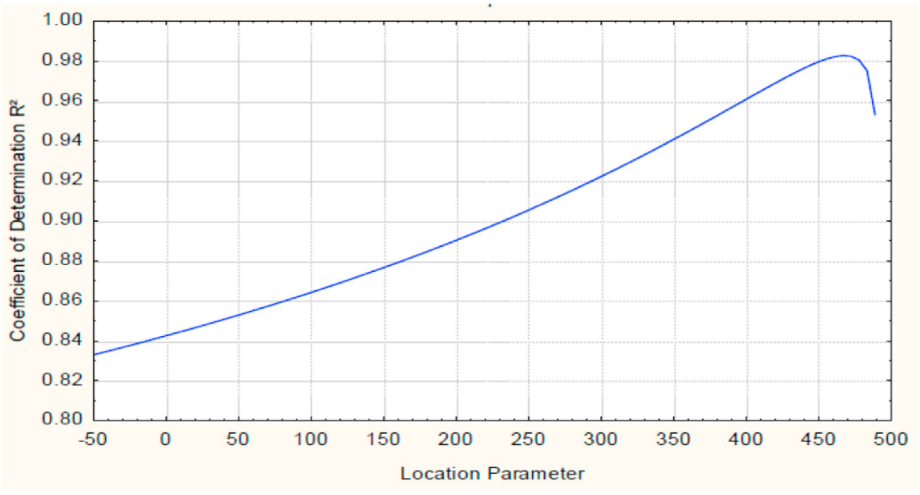
We infer that there is an exhibited variety with respect to the functional tendencies of the variable R^2, due to the location parameter. This might render the chance to identify a single parametric model rather remote. To gain more insight from the examined data, we estimate the relevant robust descriptive statistics along with the common distribution measures of symmetry and peakedness. We tabulate the descriptive statistics values (Table 3) in terms of the typical three quartiles and a robust measure of variation, the interquartile range (IQR). The common distribution shape estimators, kurtosis and skewness, have been computed, too. We notice that there are great differences among the estimated magnitudes of the medians, as well as those among the IQRs. Median lifetime values fluctuate from 117 to 7342 (k-cycles). Similarly, their respective IQR values range from 92.5 to 5888 (k-cycles). We also conclude that the skewness and peakedness dramatically vary from run to run. This may imply that complex mechanisms influence the lifetime response at different factorial combinations. Among runs, in Table 3, right-skewness dominates left-skewness by a factor of two. Furthermore, the group of the 'censored' runs favors left-skewed formations (Table 3).



Run #4



Run #5



Run #6

Fig. 5. Graphs of the coefficient of determination versus location parameter for 3-parameter Weibull model.

Table 3
Descriptive statistics of the original and log-transformed dataset.

Variable	Original Data						Transformed Data					
	Q1	Median	Q3	IQR	Skewness	Kurtosis	Q1	Median	Q3	IQR	Skewness	Kurtosis
CT1	6218	7342	7342	1124	−1.94	2.49	8.664	8.901	8.901	0.237	−2.35	5.26
CT2	293	328	420.5	127.5	0.75	1.07	5.68	5.792	6.0414	0.3614	−0.05	0.76
CT3	125	151	281.8	156.8	0.91	0.13	4.827	5.017	5.639	0.813	−0.13	−0.07
CT4	110.8	317	490	379.3	0.12	−1.32	4.707	5.748	6.194	1.487	−0.59	−1.2
CT5	215.3	390	498.5	283.3	−0.39	−1.26	5.335	5.966	6.211	0.876	−1.06	0
CT6	1454	7342	7342	5888	−1.05	−1.09	7.26	8.901	8.901	1.641	−1.31	0.14
CT7	326	409	591.8	265.8	0.65	−0.32	5.787	6.01	6.383	0.596	0.03	−0.48
CT8	86.8	141	179.3	92.5	0.57	−0.2	4.457	4.943	5.182	0.725	−0.29	−0.66
CT9	214	314	527.3	313.3	0.63	−0.77	5.343	5.749	6.258	0.914	−0.14	−0.96
CT10	296	357.5	432.3	136.3	0.45	−0.37	5.6886	5.8787	6.0687	0.3801	0.05	−0.68
CT11	5612	7342	7342	1731	−1.78	1.41	8.186	8.901	8.901	0.715	−1.78	1.42
CT12	84.5	117	238.5	154	2.56	7.08	4.423	4.762	5.473	1.05	1.03	1.3

Table 4
Differenced dataset with respect to the right-censoring limit value (7342 k-cycles).

Run #	dCT1	dCT2	dCT3	dCT4	dCT5	dCT6	dCT7	dCT8	dCT9	dCT10
1	6385	4496	0	0	0	0	0	0	0	0
2	7136	7058	7046	7037	7029	6999	6978	6922	6920	6799
3	7279	7229	7213	7204	7193	7189	7125	7070	7031	6940
4	7266	7238	7229	7108	7072	6978	6944	6861	6825	6731
5	7250	7216	7097	7092	6952	6952	6863	6855	6809	6769
6	6852	6371	5727	574	0	0	0	0	0	0
7	7110	7016	7016	6991	6970	6896	6883	6752	6745	6610
8	7286	7271	7250	7238	7216	7186	7181	7175	7126	7079
9	7200	7200	7104	7095	7032	7024	6922	6860	6679	6670
10	7083	7076	7036	7005	6995	6974	6970	6916	6891	6832
11	6961	6922	0	0	0	0	0	0	0	0
12	7286	7280	7250	7238	7229	7221	7178	7110	7084	6611

The overall standard error for skewness is 0.687. This indicates that the group of the 'uncensored' runs is more prone to symmetrical spreads. But run #12 does not conform to normality (Table 1) as the p-value for its AD test score is significant to a level of 0.01. Similarly, the kurtosis estimations display mixed-up sign tendencies. The kurtosis overall standard error is 1.334. Run # 12 appears to produce the only leptokurtic sample. Data from runs #1 and #11 deviate from normal peakedness behavior. Their corresponding AD test scores are 2.37 ($p < 0.005$) and 2.53 ($p < 0.005$). In Appendix A, we provide comparisons of the adjusted Anderson-Darling test scores (maximum likelihood estimates) along with graphical depictions for the trials (runs #2–5 and #7–10) that have not outright rejected normality in Table 1. The non-normal distribution models, which were preferably used in previous publications (Weibull, lognormal and smallest extreme value), are contrasted against the performance of a normal fit. It appears that the estimates from the non-normal candidate models justifiably compete with a normal fit, according to the adjusted Anderson-Darling test scores. The lifetime datasets are usually log-transformed in order to improve the resolution of the diagnostic pre-screening in difficult cases. In Table 3, we have repeated the same descriptive analysis for the log-transformed lifetime values. The picture is still perplexing, as the replicates from runs # 1 and 11 appear left-skewed, while the dataset that is related to run #1 switches behavior to appear leptokurtic.

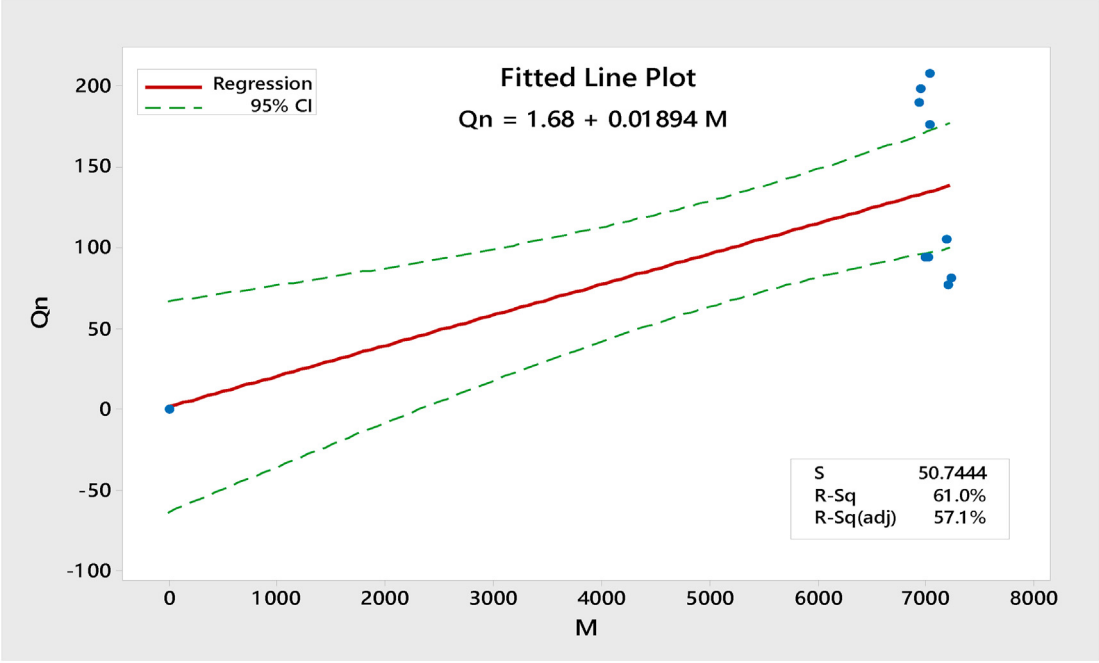
The Anderson-Darling test scores for runs # 1 (2.32), #6 (1.62), and #11 (2.51) are statistically significant at a level of 0.01. This outcome also supports the argument that the three 'censored' runs may not be modeled by a transformed normal distribution, and a more intricate modeling should be pursued. From Fig. 1, we notice that comparing among different goodness-of-fit performances (based on the regular Anderson-Darling scores), the normal distribution could be actually nominated to describe eight specific trials (R2-R5, R7-10). However, Q-Q plot (STATISTICA 7.0) comparisons demonstrate that some non-normal distributions compete or outcompete the normal distribution fits by delivering narrower 95%-confidence bands in most cases, i.e. Weibull distribution: trials # 2, 8, and 10, beta distribution: trials # 4, and 5, and gamma distribution: trials # 3, 7 and 9. Even so, the attempted normal/non-normal fits are questionable because often not all points can be restricted in the expected 95% confidence bands, thus further accentuating the overall messy behavior of the various trials.

Table 5
Median and Q_n data reduction, their associated ranked quantities, and their sum of squared ranks (SSR).

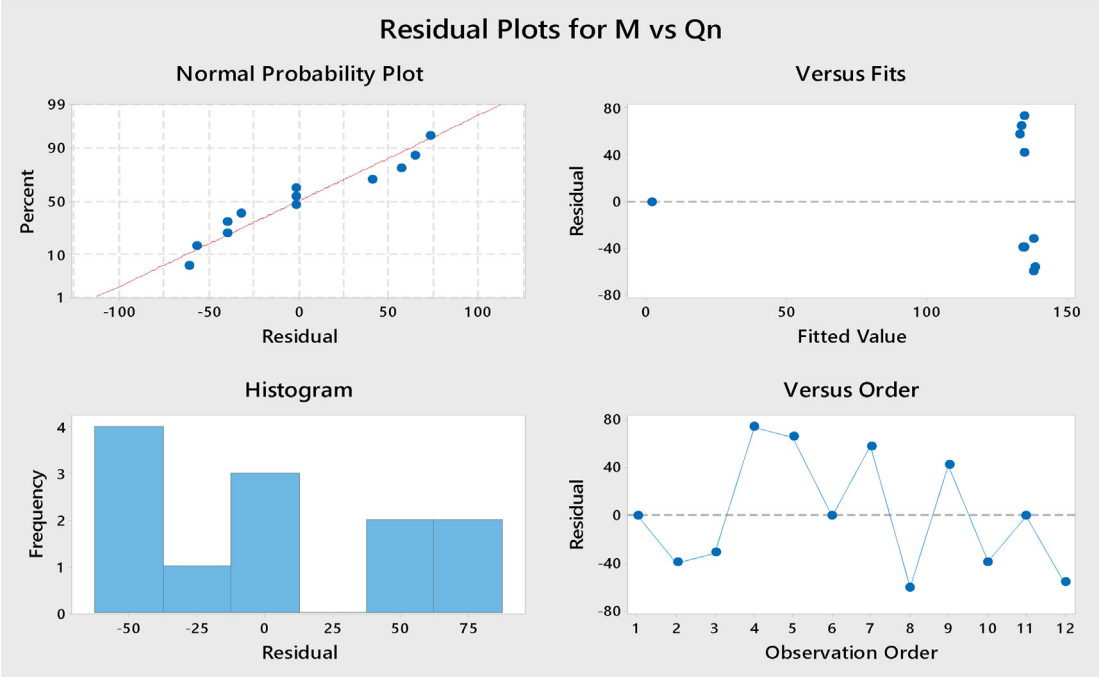
Run #	M	rM	Q_n	rQ_n	SSR
1	0	2	0	2	8
2	7014	7	94.29	6.5	91.25
3	7191	10	105.47	8	164
4	7025	8	207.75	12	208
5	6952	5	198.16	11	146
6	0	2	0	2	8
7	6933	4	190.17	10	116
8	7201	11	76.71	4	137
9	7028	9	175.79	9	162
10	6984.5	6	94.29	6.5	78.25
11	0	2	0	2	8
12	7225	12	81.5	5	169

3.2. Robust multifactorial profiling

The original lifetime dataset (Bullington et al., 1993) has been differenced with respect to the right-censoring limit of 7342 k-cycles. The resulting values for the twelve runs are listed in Table 4. In Table 5, we list the median (M) and Q_n values for each run. In Fig. 6A, the linear regression of M vs Q_n demonstrates that this fitting may not be informative, as several points lie outside the 95% confidence band. The moderate goodness-of-fit performance, according to the estimation of the adjusted coefficient of determination, adj R^2, at a value of 57.1%, may not be profitably assessed. The residual plots (Fig. 6B) also seem to



A.



B.

Fig. 6. Linear regression (A) and residual plots (B) of Q_n versus median (M).

support the fact that some non-random errors may not be ruled out. It is a 'messy' portrayal that, nevertheless, does not attest to that a correlation exists between the M and Q_n vectors. This justifies the decision to proceed to the profiling process by preserving both responses in the analysis. Next, we rank order both vector estimators. The ranked values of the median (rM) and Q_n (rQ_n) along with their consolidation into a sum of squared ranks (SSR) are listed in Table 5.

In Table 6, we tabulate the median value of the SSR response for each individual factor-setting. Alongside, we provide each setting's relative strength which is calculated with respect to the grand median SSR value of 126.5. A practical way to (subjectively) compare the strength of the effects is to depict them in a main effects plot (Fig. 7). It is obvious that the effect E is the predominant influencer. It is likely that factor H could also be active, but it would need further examination. For highly fractionated datasets, an ordinary main effects plot is not equipped to provide estimation for the statistical significance of the examined effects.

To confront this inadequacy, in Table 7, we peruse the error asymmetries across all factor levels using the proposed methodology. At a first

Table 6
Median SSR response table for the twelve factors and their relative strength with respect to the grand median (126.5).

Factor	Level	Median SSR	Relative Strength
A	1	118.6	−7.9
	2	126.5	0
B	1	126.5	0
	2	112.1	−14.4
C	1	126.6	0.1
	2	126.5	0
D	1	114.1	−12.4
	2	139	12.5
E	1	43.1	−83.4
	2	163	36.5
F	1	140	13.5
	2	114.1	−12.4
G	1	112.1	−14.4
	2	126.5	0
H	1	72.5	−54
	2	131	4.5
J	1	126.5	0
	2	126.6	0.1
K	1	107.6	−18.9
	2	131	4.5
L	1	131	4.5
	2	114.1	−12.4

glance, we observe that only the L-effect may be tagged as significantly prone to asymmetry of errors ($p < 0.05$). However, by controlling for false discovery ($\alpha = 0.05$), we find that the highest critical value is 0.0045 (=0.05/11) for the effect that exhibits the strongest asymmetrical uncertainty. Hence, nonparametric asymmetry may not actually be claimed for any of the investigated factors. This is because that even the smallest significance value of the L-effect (p-value = 0.0087) has exceeded the critical value of 0.0045. Coincidentally, the direct comparison (Table 7) between the two levels of the L-effect results in a weak outcome ($p > 0.10$). The nominated strong factors are E, F and H. By controlling the false discovery rate at a level of 0.05, the factor E emerges as the predominant influence, which concurrently maximizes the lifetime of the thermostats; it simultaneously minimizes both the median and the Q_n differences. This is because the factor E has the smallest p-value (=0.00217), which is less than the cutoff point of 0.0045 (=0.05/11). The second smaller p-value, 0.0043, is identified to factor H, which is

Table 7
Error checking and factor significance using the consolidated SSR dataset through the new approach.

Factor	error check		Factor Significance	
	U test	p-value (exact)	*U* test	p-value (exact)
A	16.5	0.818182	14	0.588745
B	15.0	0.699134	12	0.393939
C	16.5	0.818182	18	1.000000
D	10.0	0.240260	18	1.000000
E	17.0	0.937229	0	0.002165
F	14.0	0.588745	4	0.025974
G	10.5	0.240260	7	0.093074
H	16.0	0.818182	1	0.004329
J	13.5	0.484848	12	0.393939
K	14.5	0.588745	12.5	0.393939
L	2.5	0.008658	9	0.179654

also below the cutoff point of 0.0091 (=2*0.05/11); it becomes the second stronger effect in the hierarchy. The third smaller p-value, 0.026, is identified to factor F which is above the cutoff point of 0.014 (=3*0.05/11). Hence, the factor F is inactive and none of the remaining factors may be deemed significant. From Table 6, the optimal settings that coincide with the minimization of the SSR response identically occur at 'setting 1' for both factors E and H.

4. Discussion

To gain new insights from using this new approach, it is imperative to view and compare the data analysis outcomes from different angles, which other multi-factorial methods could potentially provide. We should point that one overall computational advantage of the proposed method is that besides being a non-iterative approach, it also does not require a Gram-Schmidt orthogonalization phase. A log-transformation preprocessing of the non-normal data is not a binding action for the proposed tool to be effective. A convenient – but subjective and primitive - way to probe the hierarchy and the relative strength of many factors is by separately assessing the two robust estimators through the use of the main effects plots (MINITAB 18) – one for the median response, M (Fig. 8A), and the other for the scale response, Q_n (Fig. 8B). It is immediately revealed that both effects, E and H, are the leading influencers for both lifetime components. Additionally, both of their slopes are co-

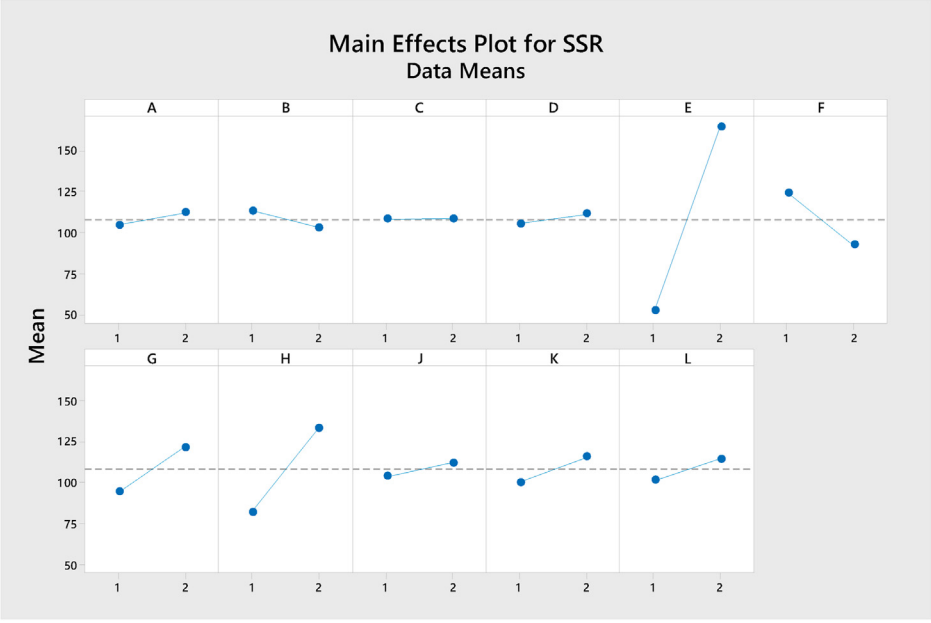
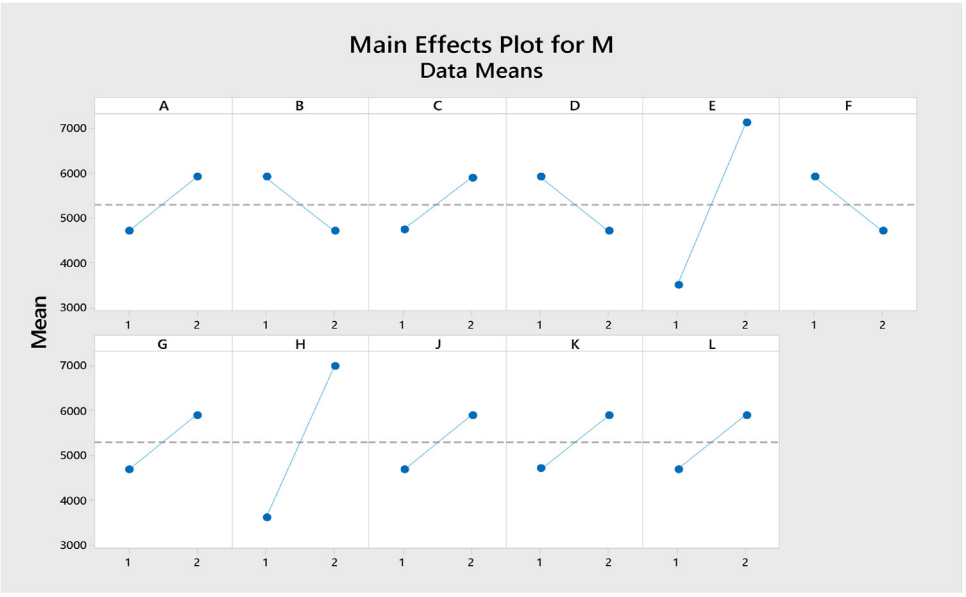
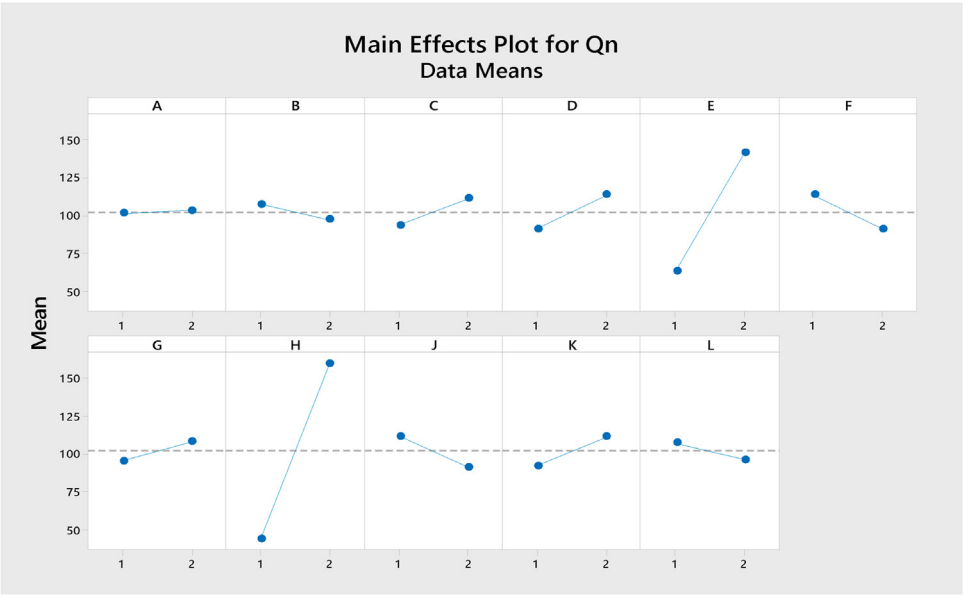


Fig. 7. Main effects plot for SSR.



A.



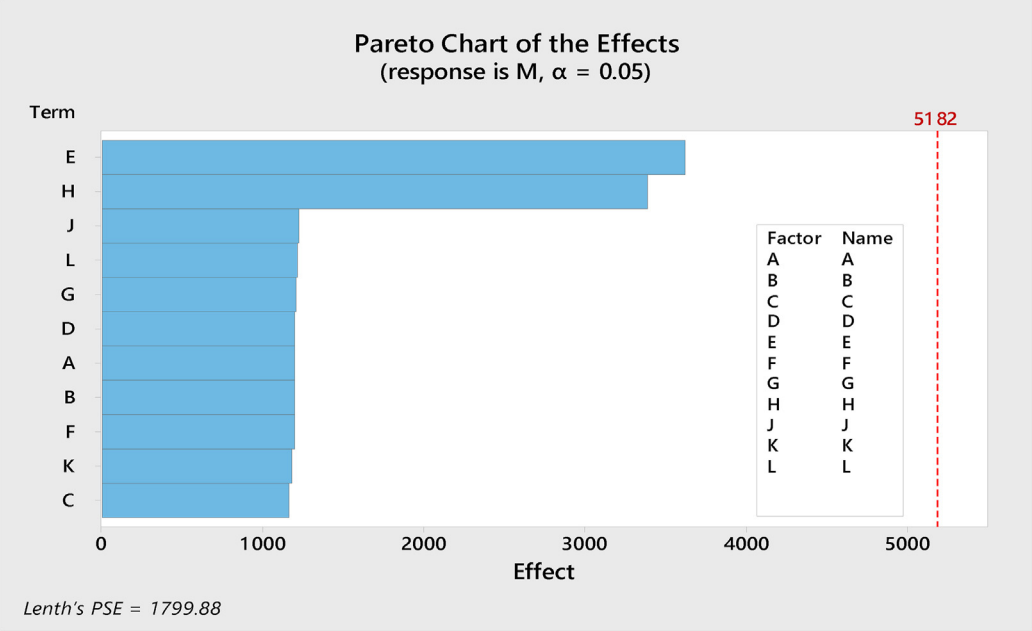
B.

Fig. 8. Main effects plot for: A) the median response (M), and B) the variation response (Q_n).

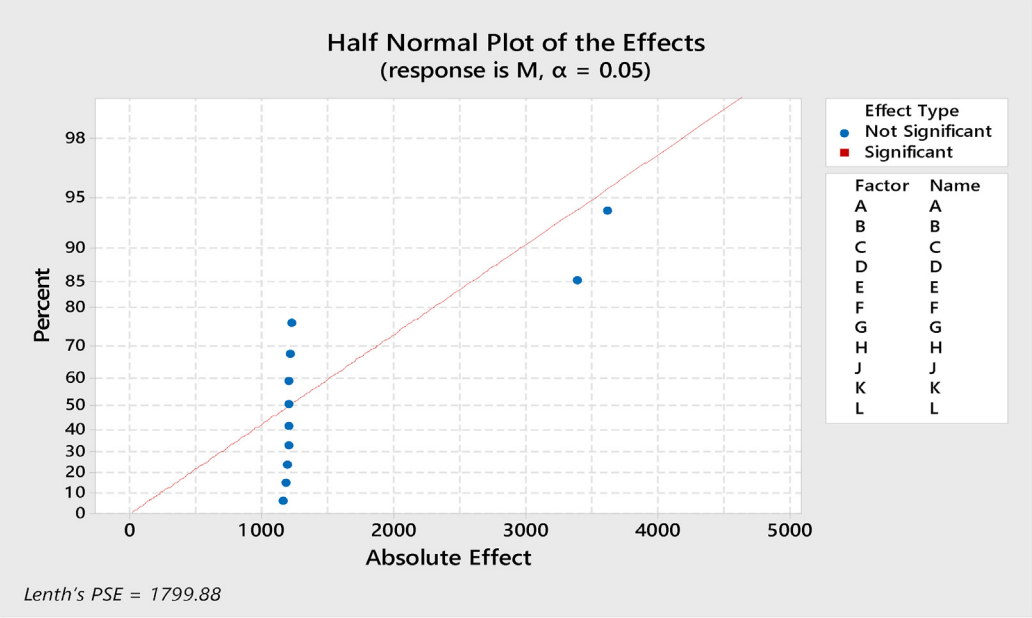
directionally elevating. Hence, it is the setting that is labelled as '1', for both controlling factors, which individually minimize the M and Q_n quantities. This outcome agrees with our findings in the preceding section. Of course, the advantage of our method is that the final diagnosis was concurrent and quantified with a statistical significance value.

Next, we compare our results with two widely used screening techniques (MINITAB 18), such as: 1) the Lenth test (1989), which portrays effects on a Pareto graph (Fig. 9A), and 2) the half-normal plot (Daniel, 1959), which is assorted with the Lenth-test critical values (Fig. 9B).

No significant effect may be identified to the median screening - at a comparable error rate of 0.05. Repeating the same procedure for the Q_n estimator (Fig. 10), both plots indicate that effects E and H are now statistically significant at a level of 0.05. Therefore, combining the available information from the two graphical screenings, one might infer that the same two effects impel the overall maximization of the thermostat lifetime by tightening its variability. It would be instructive to repeat the same graphical screening by directly using as a variable the new unified non-parametric response, SSR. The profiled effects are shown in Fig. 11 in terms of the Lenth-test/Pareto-chart and the half-normal plot. In agreement with the two previous (individual) screenings, the effects E and H appear to be active at an error rate of 0.05. The proposed terminal solution, we arrived at in the preceding section, also agrees with the solution that was achieved utilizing the Lenth-test and the half-normal plot on the new 'two-in-one' response, SSR. The clear advantage of our method is that it speeds up the diagnostics generation process while being totally constant-free, sparsity-free and distribution-free. Oppositely, the estimation of the critical cutoff point for both, the Pareto chart and the half-normal plot, are dependent on Lenth's pseudo-error. Ostensibly, the pseudo-error is regulated by two fixed constants, which are necessary to pace the trimming of the calculated regression



A.

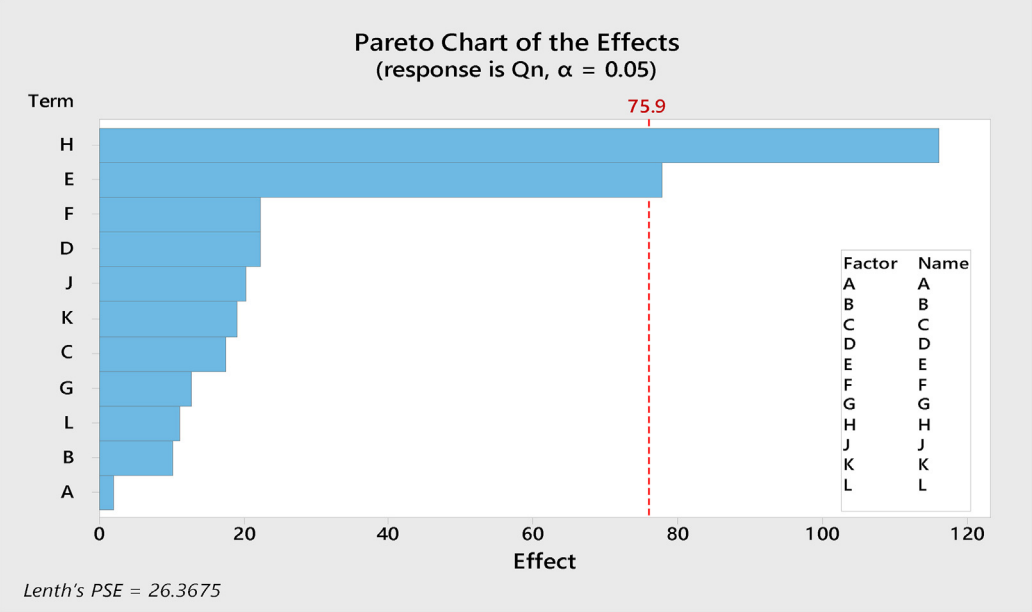


B.

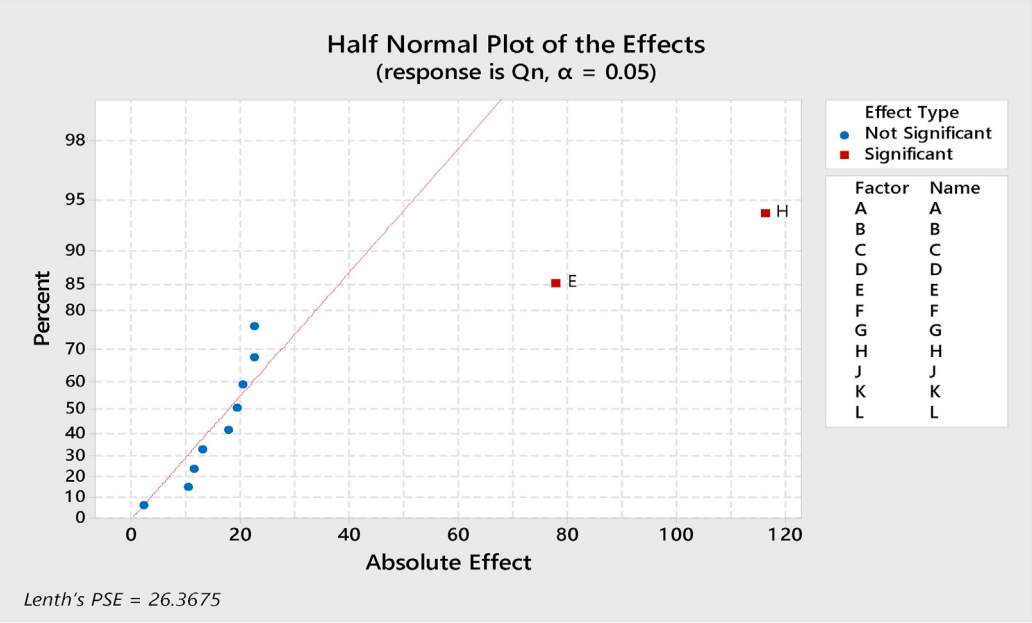
Fig. 9. Screening the effects on the median response with: A) the Pareto chart, B) Half normal plot.

coefficients. The two required Lenth constants have not been universally proven to be valid for all types of FFD datasets. Hence, they might impose some unknown degree of subjectiveness on the solver.

Finally, the desirability analysis is applied on the dichotomized lifetime response, which is represented as independent variable vectors, M and Q_n (Fig. 12). Again the results seem to confirm that effects E and H drive the composite desirability score to a value of 1.0 - in agreement with our predictions. Moreover, it is discerned from Fig. 12 that the optimal settings are E1 and H1, respectively. This outcome also agrees with our recommendations. Nevertheless, we stress the fact that our proposed solution is directly interpretable in congruence to the prevailing concept of the statistical significance. On the other hand, a perfect score in composite desirability, i.e. a value of 1, is not meaningful in terms of an affirmed probabilistic reference law. There is no rule to map a composite desirability score to a p-value estimate. Furthermore, to fit the two derived desirability functions, the response optimizer (MINITAB 18) must be fed by proper shape weights, which are picked by a trial-and-error process. The function shape weights tend to range across two orders of magnitude, i.e. from values of 0.1–10. The desirability analysis requires professional software support for the calculations because of the two sequential model fitting stages. The first stage determines an empirical multivariate model - to predict the component response - using regression analysis. Since the FFD-dataset is in the saturated-unreplicated form, the regression coefficients cannot be t-tested in order to quantify



A.

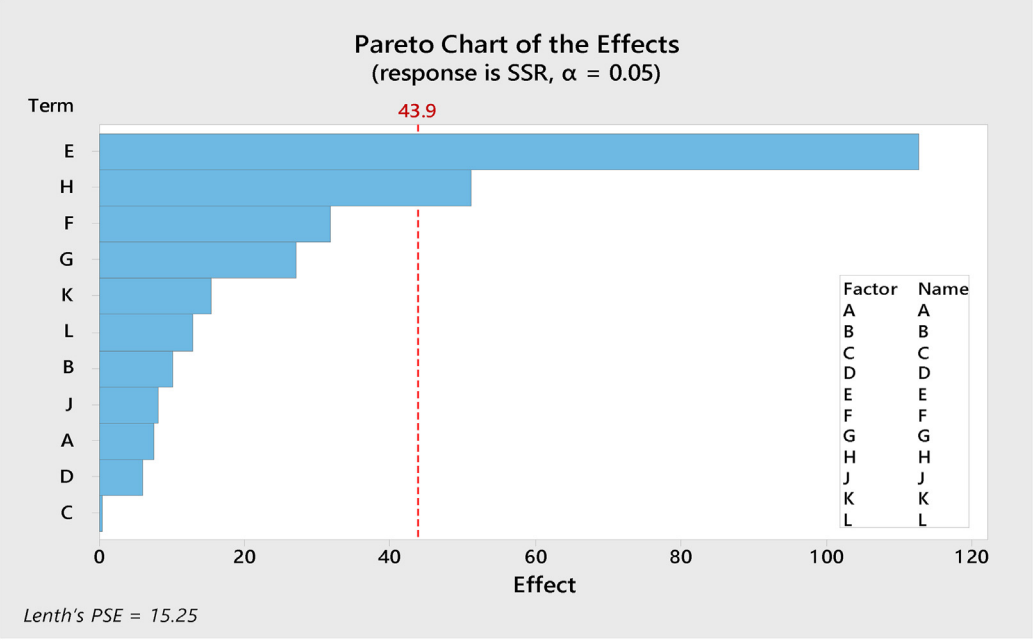


B.

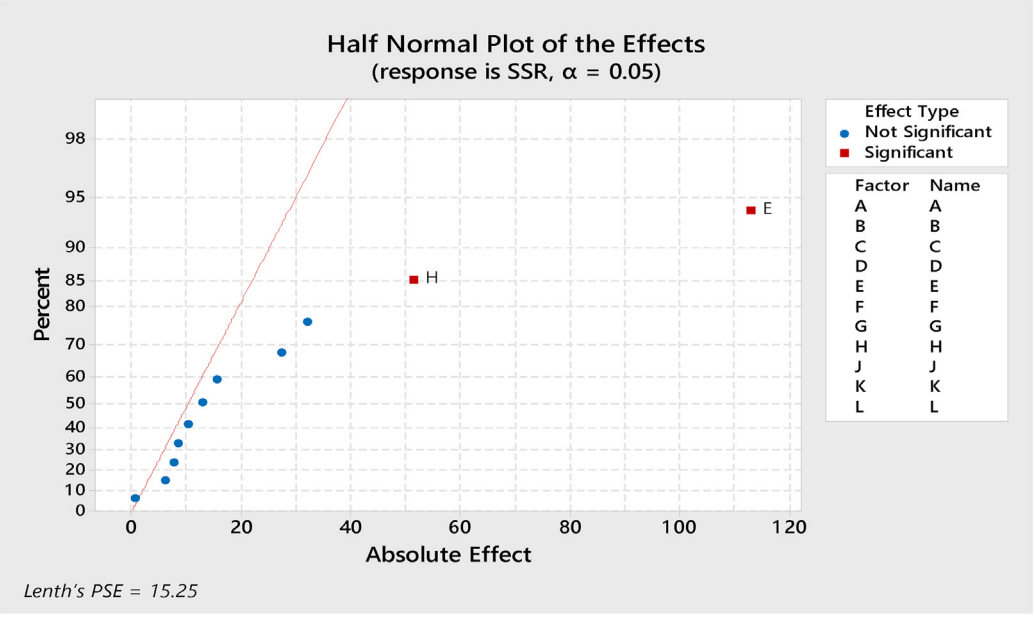
Fig. 10. Screening the effects on the Q_n response with: A) the Pareto chart, B) Half normal plot.

significance; the coefficient of determination is inestimable. Thus, the (quasi-fitted) empirical model comprises of incomplete information, which must be dispatched to the next data processing phase. The second phase searches for a solution, which maximizes the composite desirability score – always based on the available empirical model. Computationally, then, our method offers simplicity. Besides being 'agile-and-lean', during the full data conversion cycle, it aims to ensure a rapid completion of the solution cycle. Overall, our method disagrees on the total number of discovered strong effects with respect to the original solution (Bullington et al., 1993), which was found that all factors were important (Table 8). However, our solution exactly agrees on the number and type of effects with both methods: 1) the Berk and Picard (1991) method and 2) the Kececioglou method; both were employed and commented in the original publication (Bullington et al., 1993). It also agrees with the lognormal method of Wu and Hamada (2009). It is worthwhile to consider the value of the proposed method from a more obscure angle in agile decision-making. It nonparametrically resolved a difficult non-normal highly-fractionated FFD-dataset, which was initially diagnosed to be of: 1) the lognormal-type (Bullington et al., 1993; Wu and Hamada, 2009), 2) the Weibull-type (Wang and Kececioglou, 2000), and 3) the smallest-extreme-value-type (Lv et al., 2017; Wang et al., 2017).

It becomes more transparent now that the proposed technique uses the "less is more" aspect of lean thinking and the simplicity and resilience of the agile mentality – foundations well imbued in Occam 's razor. The new approach is lean because it requires less data (less cost and time)



A.



B.

Fig. 11. Screening the effects on the SSR response with: A) the Pareto chart, B) Half normal plot.

through its highly fractionated design. It is accomplished by coalescing a mini factorial-recipe design with censored lifetime sampling. The technique is computationally lean because not only it uses no iterative solvers, but it also does not require a Gram-Schmidt orthogonalization phase. Furthermore, since it does not involve regression coefficients, it is computationally simpler, i.e. more agile, because it produces no residuals. Residual analysis requires inspecting for independence of errors (autocorrelation effect) which might be a risky endeavor for small datasets. Work reduction is realized as the data analysis does not demand graphical inspection of the essential assumptions in residual analysis assumptions through: (a) the normal-probability plot, (b) the data histogram (c) the 'residuals-versus-fitted-values' plot, and (d) the 'residuals-versus-observation-order' plot. It is also lean on the tactical level because the extensive distribution identification search that was witnessed in the elucidated example is not a prerequisite for the new factor profiler to operate. The selection of maximum breakdown-point estimators to evaluate location and dispersion lifetime properties in tandem with the distribution-free (Wilcoxon-Mann-Whitney) comparison of effects offers the resilience for an agile data treatment. The multifactorial screening becomes now constant-free and not dependent on the sparsity assumption. Hence, it is an agile approach because it is simpler and responsive.

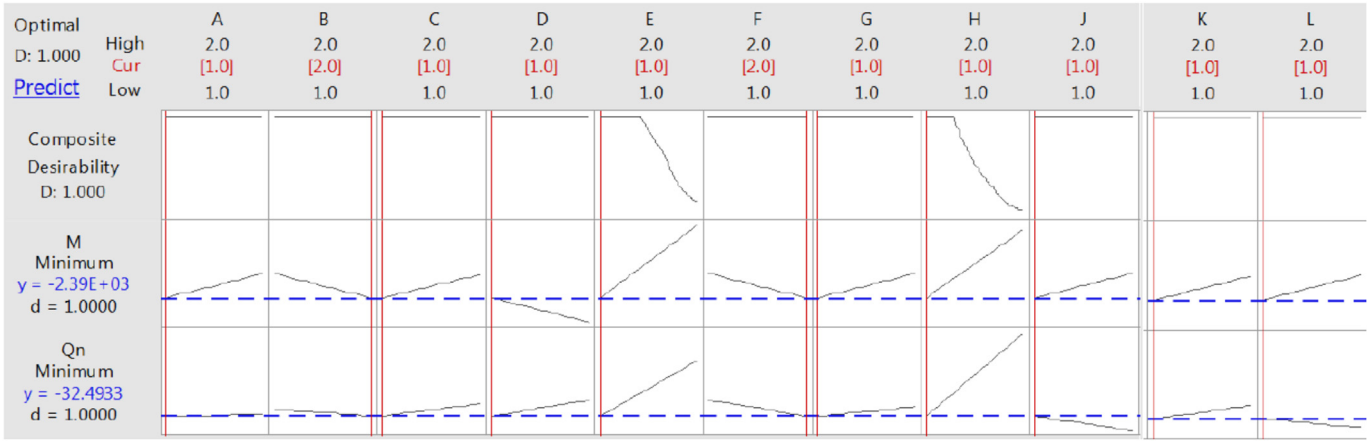


Fig. 12. Desirability analysis for the robust measures, M and Q_n.

Table 8
Main effects predictions in previous research and new results.

Strong Main Effects Identification	Distribution Type	Reference
All 11 effects (A-K)	Lognormal	Bullington et al. (1993)
E, H	Lognormal with the Kececioglou method	
E, H	Berk and Picard method(1991)	
All 11 effects (A-K)	Weibull log-linear	Wang and Kececioglou(2000)
E, H	Lognormal	Wu and Hamada (2009)
Three bootstrap methods: PB: C, E, G, H BCPB: B, C, D, E, F, G, H, I BCa: B, C, D, E, F, G, H, I	Smallest extreme value (at the location parameter)	Wang et al. (2017)
A, C, E, G, H, I, J, K	Smallest extreme value (with random effects)	Lv et al. (2017)
E, H	Distribution free (Wilcoxon-Mann-Whitney)	Proposed method

5. Conclusions

Highly fractionated survival datasets, which also contain censored observations, pose several challenges to analyzers. Robust multifactorial profilers might prove beneficial in determining the impact of the studied effects in terms of location and dispersion contributions. We propose a method for screening and optimizing small censored lifetime datasets, which are programmed by dense factorial arrays. To improve complex product/process performance, the technique orchestrates the statistical assessment of the effect hierarchy. The overall concept endorses simplicity in decision-making. It heeds to the realistic needs in operations for speedy and economic discovery with minimal exploitation of the available resources (lean thinking). From a DOE perspective, two essential features were addressed: 1) the adoption of saturated fractional factorial schemes for maximum effect engagement, and 2) the necessity to work with limited data. The data characteristics that are flexibly handled are: 1) small sampling, 2) censoring (partial/full), 3) use-rate acceleration, 4) indeterminate distribution or multiplicities, and 5) experimental scheme saturation. The versatility of the proposed profiler rests on furnishing dichotomized information by splitting and reformulating the lifetime dataset in terms of robust location (median estimator) and robust dispersion (Q_n-estimator) measures. Rank operations facilitated the smooth fusion of location and dispersion information in order to seamlessly achieve the synchronous effect profiling. Advantages that are attributed to the proposed multi-factorial profiler include: 1) the statistical quantification of uncertainty asymmetry between individual factor levels, 2) the naïve resolution of unreplicated-saturated information vectors, 3) the determination of 'constant-free' effect hierarchy, 4) the determination of distribution-free effect hierarchy and 5) the individualization of the effect-strength significance.

The new methodology was tested on a published dataset that highlighted the screening and maximization of lifetime performance for a real thermostat product. Only twelve experimental recipes were necessitated to configure the profile of as many as eleven controlling factors. Each recipe was restricted to generate a small output (10 replicates). The paradigm was proved to be intriguing because of its distinctive 'messiness'; it could not be shown that it obeyed a particular parametric reference law. Instead, distribution multiplicity was evident. The censored data contributed to the overall problem complexity by neutralizing the conversion capabilities of regular solvers. No correlation was found between the median- and Q_n-estimator vectors. Thus, they were fused, analyzed and the resulting profiling indicated that only two controlling factors were statistically strong (E and H). Peripheral analysis with other combinations of methods showed that the final result is reliable. The false discovery rate was controlled at a level of $\alpha = 0.05$, which is deemed superb for this level of difficulty. Future work could explore the concurrent robust screening and optimization of multiple reliability characteristics of a product.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the Editor in Chief and the reviewers for their critical comments that led to the improvement of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aime.2021.100036.

References

Al-Hakim, L., Jin, C., 2013. Quality Innovation: Knowledge, Theory and Practices, first ed. IGI Global.
Aldowaisan, T., Nourelfath, M., Hassan, J., 2015. Six Sigma performance for non-normal processes. Eur. J. Oper. Res. 247, 968–977.

Balle, M., Jones, D., Chaize, J., 2017. The Lean Strategy: Using Lean to Create Competitive Advantage, Unleash Innovation, and Deliver Sustainable Growth, first ed. McGraw-Hill, New York.
Bendoly, E., Bharadwaj, A., Bharadwaj, S., 2012. Complementary drivers of new product development performance: cross-functional coordination, information system capability and intelligence quality. Prod. Oper. Manag. 21 (4), 653–667.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. 57, 289–300.
Berk, K.N., Picard, R.R., 1991. Significance tests for saturated orthogonal arrays. J. Qual. Technol. 23, 79–89.
Besseris, G.J., 2010. A methodology for product reliability enhancement via saturated unreplicated fractional factorial designs. Reliab. Eng. Syst. Saf. 95, 742–749.
Besseris, G.J., 2013. A distribution-free multi-factorial profiler for harvesting information from high-density screenings. PLoS One 8, e73275.
Bhote, K.R., Bhote, A.K., 2004. World Class Reliability: Using Multiple Environment Overstress Tests to Make it Happen. Amacom, New York.
Box, G.E.P., Hunter, W.G., Hunter, J.S., 2005. Statistics for Experimenters: Design, Innovation, and Discovery, second ed. Wiley, New York.
Briggs, W., 2016. Uncertainty: the Soul of Modeling, Probability and Statistics, first ed. Springer, Switzerland.
Bullington, R.G., Lovin, S.G., Miller, D.M., Woodall, W.H., 1993. Improvement of an industrial thermostat using designed experiments. J. Qual. Technol. 25, 262–270.
Cherrafi, A., Elfezazi, S., Chiarini, A., Mokhlis, A., Benhida, K., 2016. The integration of lean manufacturing, Six Sigma and sustainability: a literature review and future research directions for developing a specific model. J. Clean. Prod. 139, 828–846.
Collins, D.H., Freels, J.K., Huzurbazar, A.V., Warr, R.L., Weaver, B.P., 2013. Accelerated test methods for reliability prediction. J. Qual. Technol. 45, 244–259.
Condra, L., 2001. Reliability Improvement with Design of Experiments. CRC Press, New York.
Cudney, E.A., Agustiady, T.K., 2016. Design for Six Sigma: A Practical Approach through Innovation, first ed. CRC Press, Boca Raton.
Daniel, C., 1959. Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1, 311–341.
Darby, J.L., 2010. Sample Sizes for Confidence Limits for Reliability. Sandia National Laboratories, New Mexico & California. Report SAND2010-0550.
Derringer, G., Suich, R., 1980. Simultaneous optimization of several response variables. J. Qual. Technol. 12, 214–219.
Dodson, B., 1993. Weibull Analysis. American Society for Quality Publications, Milwaukee.
Gabrel, V., Murat, C., Thiele, A., 2014. Recent advances in robust optimization: an overview. Eur. J. Oper. Res. 235 (3), 471–483.
Hamada, M., Wu, C.F.J., 1991. Analysis of censored data from highly fractionated experiments. Technometrics 33, 25–38.
Hamada, M., 1993. Reliability improvement via Taguchi's robust design. Qual. Reliab. Eng. Int. 9, 7–13.
Hamada, M., 1995. Using statistically designed experiments to improve reliability and to achieve robust reliability. IEEE Trans. Reliab. 44, 206–215.
Hamada, M., Wu, C.F.J., 1995. Analysis of censored data from fractionated experiments: a Bayesian approach. J. Am. Stat. Assoc. 90, 467–477.
He, Z., Zhu, P.-F., Park, S.-H., 2012. A robust desirability function method for multi-response surface optimization considering model uncertainty. Eur. J. Oper. Res. 221, 241–247.
Hoaglin, D.C., Mosteller, F., Tukey, J.W., 2000. Understanding Robust and Exploratory Data Analysis. Wiley-Interscience, Hoboken.
Huber, P.J., Ronchetti, E.M., 2009. Robust Statistics. Wiley, Hoboken.
Jantschi, L., Bolboaca, S.O., 2018. Computation of probability associated with Anderson-Darling statistic. Mathematics 6 (88), 1–16.
Jeong, I.-J., Kim, K.-J., 2009. An interactive desirability function method to multiresponse optimization. Eur. J. Oper. Res. 195 (2), 412–426.
Joseph, V.R., Yu, I.T., 2006. Reliability improvement experiments with degradation data. IEEE Trans. Reliab. 55, 149–157.
Kim, K., Lin, D., 2006. Optimization of multiple responses considering both location and dispersion effects. Eur. J. Oper. Res. 169 (1), 133–145.
King, J.P., Jewett, W.S., 2010. Robustness Development and Reliability Growth: Value Adding Strategies for New Products and Processes, first ed. Prentice-Hall, New York.
Lawless, J.F., 2003. Statistical Models and Methods for Lifetime Data, second ed. Wiley, New Jersey.
Lenth, R.V., 1989. Quick and easy analysis of unreplicated factorials. Technometrics 31, 469–473.
Lio, Y.L., Lu, J.C., Ruan, L., 2015. Robust parameter design for quality and reliability issues based on accelerated degradation measurements. IEEE Trans. Reliab. 64, 949–959.
Ludbrook, J., Dudley, H., 1998. Why permutation tests are superior to *t* and *F* tests in biomedical research. Am. Statistician 52, 127–132.
Lv, S., Niu, Z., Cui, Q., He, Z., Wang, G., 2017. Reliability improvement through designed experiments with random effects. Comput. Ind. Eng. 112, 231–237.
Mackelprang, A.W., Habermann, M., Swink, M., 2015. How firm innovativeness and unexpected product reliability failures affect profitability. J. Oper. Manag. 38 (1), 71–86.
Maillard, P., 2015. Competitive Quality and Innovation, first ed. Wiley, New Jersey.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18, 50–60.
Marsaglia, G., 2004. Evaluating the Anderson-Darling distribution. J. Stat. Software 9 (2), 730–737.
Meeker, W.Q., Escobar, L.A., 1998. Statistical Methods for Reliability Data. Wiley, New York.
Milliken, G.A., Johnson, D.E., 1989. Analysis of Messy Data, vol. II. Chapman and Hall/CRC, Boca Raton. Nonreplicated Experiments.
Milliken, G.A., Johnson, D.E., 2004. Analysis of Messy Data, vol. I. Chapman and Hall/CRC, Boca Raton. Designed Experiments.
Mueller, G., Rigdon, S.E., 2015. The constant shape parameter assumption in Weibull regression. Qual. Eng. 27 (3), 374–392.
Nelson, W.B., 1990. Accelerated Testing: Statistical Models, Test Plans and Data Analysis. Wiley, New Jersey.
Pett, M.A., 2015. Nonparametric Statistics for Health Care Research: Statistics for Small Samples and Unusual Distributions, second ed. SAGE, New York.
Piña-Monarrez, M.R., Ortiz-Yañez, J.F., 2015. Weibull and lognormal Taguchi analysis using multiple linear regression. Reliab. Eng. Syst. Saf. 144, 244–253.
Pyzdek, T., Keller, P.A., 2014. The Six Sigma Handbook, fourth ed. McGraw-Hill, New York.
ReVelle, J.B., 2001. Manufacturing Handbook of Best Practices: an Innovation, Productivity and Quality Focus, first ed. CRC Press, Boca-Ratton.
Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. J. Am. Stat. Assoc. 88, 1273–1283.
Roy, B., 2010. Robustness in operational research and decision aiding: a multi-faceted issue. Eur. J. Oper. Res. 200 (3), 629–638.
Siebert, C.F., Siebert, D.C., 2017. Data Analysis with Small Samples and Non-normal Data: Nonparametrics and Other Strategies. Oxford University Press, London.
Silva, G.M., Gomes, P.J., Lages, L.F., Pereira, Z.L., 2014. The role of TQM in strategic product innovation. Int. J. Oper. Prod. Manag. 34 (10), 1307–1337.
Silver, N., 2015. The Signal and the Noise: Why So Many Predictions Fail-But Some Don't, first ed. Penguin, New York.
Singh, S., Ramakrishna, S., Gupta, M.K., 2017. Towards zero waste manufacturing: a multidisciplinary review. J. Clean. Prod. 168, 1230–1243.
Taguchi, G., Chowdhury, S., Taguchi, S., 2000. Robust Engineering: Learn How to Boost Quality while Reducing Costs and Time to Market. McGraw-Hill, New York.
Taguchi, G., Chowdhury, S., Wu, Y., 2004. Quality Engineering Handbook. Wiley-Interscience, Hoboken.
Vinodh, S., Sundararj, G., Devadasan, S.R., Rajanayagam, D., Murugesh, R., 2008. Infusing agility, innovation and quality in products. Int. J. Serv. Oper. Manag. 4 (2), 243–263.
Virmani, N., Saha, R., Sahai, R., 2018. Leagile manufacturing: a review paper. Int. J. Prod. Qual. Manag. 23 (3), 385–421.
Wang, W., Kececioglou, D.B., 2000. Fitting the Weibull log-linear model to accelerated life-test data. IEEE Trans. Reliab. 49 (2), 217–223.
Wang, G., He, Z., Xue, L., Cui, Q., Lv, S., Zhou, P., 2017. Bootstrap analysis of designed experiments for reliability improvement with a non-constant scale parameter. Reliab. Eng. Syst. Saf. 160, 114–121.
Wilcox, R.R., 2010. Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy. Springer, London.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics 1, 80–83.
Wu, C.F.J., Hamada, M.S., 2009. Experiments: Planning, Analysis, and Optimization, second ed. John Wiley & Sons, Hoboken.