Characteristics of Extrapolation-Optimized Response Surface Designs for Multivariate Reliability Modelling and Lifetime Prediction

Marco Arndt[a*], Martin Dazer[a]
[a] Institute of Machine Components (IMA), University of Stuttgart, Stuttgart, Germany
[*] marco.arndt@ima.uni-stuttgart.de

Abstract: For multivariate reliability assessment, accelerated life tests (ALT) must be conducted using experimental design (DOE) according to response surface methodology (RSM). This involves testing under accelerated load, as far as this is technically feasible and statistically evaluable in terms of lifetime. However, this often places the operational area of the test design within a stress range favorable for the experiment but not representative of field load levels, necessitating extrapolation with validation points for benchmark purposes. This paper addresses optimized experimental design for nonlinear multivariate lifetime assessment, considering validation points that aid extrapolation without significantly reducing efficiency. It proposes systematically relocating experimental points for validation and analyzing their lifetimes in conventional experimental designs. Initially, accelerated test points within a reliability test, structured within a central composite design (CCD), are applied to a response surface design (RSD). Maximum likelihood estimation (MLE) for a parametric failure distribution with significance analysis provides an RSM model predicting lifetimes extrapolating to field levels. A further set of test points, used to validate the RSM model, is allocated in the extrapolation direction. Numerical evaluation of the RSD through Monte Carlo simulations, using both test sets, considers various characteristics: prediction variance, differences in effect amount for coefficient estimates, statistical power, and coefficient correlation. This approach supplements conventional RSDs for life testing with ALT, balancing practicable testing duration and reasonable extrapolation quality through improved information distribution.

Keywords: design of experiments, extrapolation, ALT, multivariate lifetime testing, reliability modeling

1. INTRODUCTION

The reason why Accelerated Life Testing (ALT) is frequently applied in reliability engineering is usually quite conclusive: reliable products or technical systems typically might, under normal usage conditions, exhibit lifespans that are impractical for testing purposes This holds especially when the goal is to establish a data foundation for lifetimes or to demonstrate reliability. Typically, tests are then accelerated by increasing the stress levels, triggering failure mechanisms in a timely manner while not changing the End-of-Life (EoL) failure mode [1]. Subsequently, a specific regression model, for example using Generalized-Log-Linear Models (GLM) and Maximum Likelihood Estimation (MLE), is estimated, which, in addition to Physics of Failure (PoF), allows deriving a life-stress model based on the empirical data [2,3]. Previous work such as [4] among others, and while considering various analysis methods [5], noted that the lifetime to be modeled following these principles could also depend on two continuous influencing factors instead of only one (which also straightforwardly applies to more than two factors), making a statistical experimental design essential.
Response Surface Methodology (RSM) offers compelling advantages in experimental design (DOE) concerning the requirements of reliability engineering [6]. In particular, factorial designs are described here, such as the Central Composite Design (CCD), which has highly advantageous properties due to its orthogonal and rotatable composition. Here, several factors can be considered in an uncorrelated manner and effects can be estimated independently of each other. These are crucial properties for testing industrial applications. The resulting response surface also provides a constant prediction performance or prediction variance radially from the center of the experimental design, which is ideal for building an operational reliability model. This is especially true when it is not clear where the ideal parameter space for testing a particular technology might be located - which corresponds to the typical situation in business practice. In addition to countless optimal designs, which are tailored in various ways to a multitude of performance characteristics of the model being formed [7,8], the CCD arranges the test points in a highly efficient manner when there is no prior knowledge. Interactions and quadratic effects, which may occur especially in technical products where a PoF is not yet understood, can be captured. In summary, for technical products whose reliability equation is to be determined, it is simply not clear in advance where the optimal test conditions in the parameter space might be or what optimal design for reliability formation might be chosen. Although optimal designs have been described for

parameter estimation using a GLM or the underlying distributions like Poisson and logistic regression models, there is no final consensus on the arrangement of test points in the context of the Weibull distribution [9-11]. For an GLM optimal design, the unknown and estimated design parameters, in particular the Weibull shape and scale, are required to be known in advance [11]. Rather, there is an emphasis on finding and applying model generation algorithms that could iteratively propose designs under the constraints of factor aliasing, etc. [12]. Taking this into account, in general, CCDs are still used to be selected and placed in a presumably ideally accelerated parameter space after screening tests or according to expert knowledge and constructed according to the criteria of orthogonality and star point distances α_D for axial runs [13]. If unexpected issues arise that complicate data collection due to feasibility or effort, we have already attempted to capture general effects in [14-16]. We also found that, even in the case of presumed inconsistencies, advantages may arise through tolerable effects on statistical power and effect estimation concerning cost and efficiency, and an accelerated experimental design may be tuned regarding prediction variance from the accelerated field [17].
However, this leaves room for further improvement: to what extent does the acceleration, the experimental design selection or the regression modeling now affect backtracking - in other words, the performance in prediction within that fraction of the parameter space that is directly relevant: the field levels. Typically, the widths of confidence intervals (CI) increase during extrapolation due to prediction variance from the experimental design as well as variance in model parameter and effect estimation based on empirical data. Aleatory uncertainties from lifespan measurement, epistemic uncertainties from the factorial design, and statistical uncertainties due to a limited sample size and presumably one-time experimental plan replication affect the sharpness of RSD extrapolation. For industrial applications, the question remains how to address this, on account of installing validation points towards the extrapolation field or whether these can be integrated into the RSD.
Without claiming to create a universal algorithm for generating an innovative, optimal experimental design, this paper starts from a very practical situation in experimental strategies. Therefore, a CCD is considered for the reasons mentioned above, with a provided, exemplary system behavior taken from industrial practice. Randomly accelerated lifespans over two influencing factors and five levels are simulated and tested with an orthogonal two-factor CCD. Subsequently, parameter estimation is carried out using a GLM and MLE. The same CCD is then equipped with validation points without affecting its efficiency level: based on known prior works [14-17], some CCD runs that might be suitable for tuning are omitted and rearranged for extrapolation and validation purposes. Finally, an effect estimation is numerically conducted by Monte-Carlo Analysis (MC) and compared with the previous results. To elucidate the background of this approach, RSM, optimality definitions, and parameter estimation methods are discussed first, followed by the methodology and results of this work.

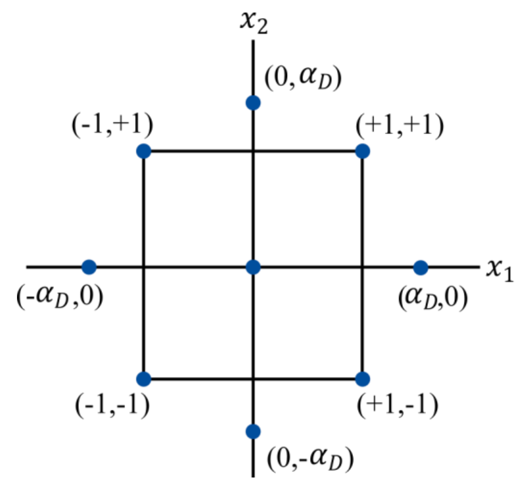
2. MULTIVARIATE ACCELERATED LIFETIME TESTING

For an experimental design to be created that can be used to estimate multivariate lifetime or a reliability model, the framework conditions that apply must be clarified first. As part of the study and representative of practical applications in process, energy or automotive engineering, we have opted for an ALT design in which complete EoL data is considered. Basically, a consistent failure mode is assumed for all constant stress levels under consideration in a CCD and the acceleration is performed across both $k = 2$ variables of increased stress. Deviations from this, such as censoring, changing or competing failure modes with a resulting variation in the parameterization of the failure distribution and unbalanced acceleration, can of course be equally taken into account with the procedures presented - they simply have to be modified. For this reason, the following must define how the RSM is assessed.

2.1. Response Surface Methodology for Lifetimes

To assess the size and significance of effects on the lifetime as the system response y from experimental observations and ultimately incorporate them into a fit for RSM, the regression coefficients β_j must be estimated as $\hat{\beta}_j$ from EoL observations through MLE. The relevant difference to normally distributed data in this evaluation is the Weibull distribution for lifetimes, which is usually the basis and generally invalidates a coefficient estimate using Ordinay Least Squares (OLS) [5]. Since all regression models are linear as long as the regression parameters are linear, second-order models with interactions can also be described in this manner. Consequently, this also includes RSM with curvatures [6].

Figure 1: $k = 2$ CCD



Run #	x_1	x_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1
5	-α_D	0
6	+α_D	0
7	0	-α_D
8	0	+α_D
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0

For a two-factor CCD considering $n_F = 4$ factorial runs, axial distances $\alpha_D = \sqrt[4]{n_F}$, c.f. Figure 1, and a second-order model this means the system response is determined by

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \quad (1)$$

where quadric terms may be substituted by

$$x_3 = x_1^2,\ x_4 = x_2^2,\ x_5 = x_1 x_2 \text{ and } \beta_3 = \beta_{11},\ \beta_4 = \beta_{22},\ \beta_5 = \beta_{12} \quad (2)$$

while adjusting the number of coefficients to $j = 1, \dots, k$ [6]. When dealing with $i = 1, \dots, n > k$ observations, the regression estimates $\hat{\beta}_j$ can be derived from multiple linear regression analysis for data with normally and independently distributed residual data using the matrix notation of the system response as follows:

$$\mathrm{y} = \mathrm{X}\beta + \varepsilon, \quad (3)$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_o \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \quad (4)$$

Here, **y** represents the vector of observed responses (loadcycles, revolutions, lifetime, etc.), **X** denotes the matrix of predictor variables (influencing stresses), **β** is the vector of regression coefficients to be estimated (effects), and **ε** stands for the vector of random errors (measurement errors, randomness). A fundamental assumption when using lifetime being parameterized with a distribution from the exponential family including location and scale parameters such as the two-parameter Weibull distribution is that it can be incorporated into the analytical lifetime or reliability estimation via GLM as

$$\ln(T) = y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j. \quad (5)$$

Taking into account the characteristic lifetime T as the Weibull scale parameter and b as the Weibull shape parameter, the associated GLM can be derived by applying the General Log-Linear relationship to the Weibull distribution which prepares the term for MLE as follows:

$$f(t, X) = b \cdot t^{b-1} \cdot e^{-b\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_j\right)} \cdot e^{-t^b e^{-b\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_j\right)}}, \ \ t > 0\,. \quad (6)$$

Further evidence why this is superior to a non-linear OLS or other transformation approaches, especially for coefficient and variance estimation, can be found in [6,18]. Finally, the GLM suitable for the MLE results from summing up all $t_i \sim \mathcal{W}(T; b)$ EoL observations correspondingly on the basis of Eq. 6 to

$$\Lambda = \ln\left(L(t_i; b, \beta_0, \dots, \beta_k)\right) = \sum_{i=1}^{n}\left[\ln(b) - b \cdot \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_j\right) + (b-1)\cdot \ln(t_i) - t_i^b \cdot e^{-b\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_j\right)}\right]. \quad (7)$$

At this point, note that aforementioned cases such as censored (suspension) data or varying life-stress models (Arrhenius, Power, etc.) are also covered in [2,3,5] with respective GLM equations. Once the metrics of the two-parameter Weibull distribution and the model coefficients have been estimated in $\hat{\theta}_i = \left[\hat{b}\ \ \hat{\beta}_0\ \ \hat{\beta}_1 \dots\right]'$ using a suitable optimization algorithm such as the Newton-Raphson Method or Patternsearch to maximize the likelihood in Eq.7 [3,18], they can be tested for significance with respect to the model equation and then checked against a confidence level α, c.f. Section 2.3.

2.2. Optimality Criteria

As mentioned in the introduction within Section 1, on the one hand optimal designs were already considered for the underlying, assumed parameterizations, so that they sought to provide the best possible framework to the model in Eq.6 [9,10]. On the other hand, there are optimal designs that minimize the CI of the regression coefficients $\hat{\beta}_j$ on average or across the volume or through the variance (etc.) due to the arrangement of the test runs [6,7,13]. Although a basic CCD is preferred for ALT for the reasons mentioned above in this work, optimal designs still yield utilizable metrices that evaluate optimality and can be used for comparison in a single measure also for other test point arrangements. While the CI for T or, correspondingly, for the resulting reliability equation could actually be approximated rather straightforward, which is also a sound approach particularly for a CCD with at least $n > 15$ EoL test runs by default [3], their multidimensional evaluation within the numerical MC approach presented later is nevertheless almost impossible. Therefore, if a new test design layout is to be evaluated on the basis of a CCD using optimality criteria for lifetime prediction and extrapolation, it is reasonable to utilize optimality criteria for comparison.

A-criterion:

The A-criterion measures the width of the CI of $\boldsymbol{\beta}$ on average. A-optimality is achieved by minimizing it [6,7]. Analytically, it is calculated via the sum of the main diagonal elements (or the trace tr) of the inverted information matrix $\mathbf{M} = \mathbf{X}'\mathbf{X}$:

$$A = tr(\mathbf{M}^{-1}). \quad (8)$$

D-criterion:

D-optimality and variations thereof are the most frequently used optimality criteria. It aims to minimize the volume of CI for $\boldsymbol{\beta}$. This is achieved by maximizing the D-criterion or minimizing its inverse. Put simply, the maximum determinant leads to maximum information [6,7,13]. The D-criterion therefore results from:

$$D = |\mathbf{M}|. \quad (9)$$

G-criterion:

G-optimality addresses the prediction variance of an experimental design. Despite statisticians and experiment designers still seeming to focus too less on optimized prediction test plan with good prediction properties [6], these are particularly important in industrial, technical applications. Former may partly originate from the fact that the scaled prediction variance (SPV) $v(\mathbf{x}_0)$ must be taken into account, which, unlike $|\mathbf{M}|$ or $tr(\mathbf{M}^{-1})$, cannot be summarized in a simple metric but depends on each location $\mathbf{x}_0^{\boldsymbol{T}} = [1, x_{01}, \dots, x_{0k}]$ in the parameter space (or region of interest) R according to

$$v(\mathbf{x}_0) = \frac{N\mathrm{Var}[\hat{y}(\mathbf{x}_0)]}{\sigma^2} = \mathbf{x}_0^T \left(\frac{\boldsymbol{M}}{N}\right)^{-1} \mathbf{x}_0 = N\mathbf{x}_0^T (\boldsymbol{M})^{-1}\mathbf{x}_0 \quad (10)$$

considering N test runs [6,13]. In a G-optimal design, the largest prediction variance to be expected within the parameter space is minimized. Often the worst or largest prediction variance occurs at the edge of the parameter space under consideration. The G-criterion thus results in

$$G = \max_{\mathbf{x}_0 \in R} \upsilon(\mathbf{x}_0). \tag{11}$$

I-criterion:
In I-optimality, the average $\upsilon(\mathbf{x}_0)$ (Integrated Variance) within a design space R is sought to be minimized [6]. The average is determined using the integral of the SPV over the volume of the region of interest. Therefore, the I-criterion results from

$$I = \frac{1}{\int_R d\mathbf{x}} \cdot \int_R \mathbf{x}_0^T (\boldsymbol{M})^{-1} \mathbf{x}_0 \, d\mathbf{x}. \tag{12}$$

2.3. Significance Analysis

Using the model optimization parameters $\hat{\theta}_i$ of Section 2.1, which are suitable for maximizing the log-likelihood function according to Eq.7, the likelihood ratio test (LRT) can be used for hypothesis testing. In order to prove the significance of an effect derived from the presumed difference of scale parameters T, consider a stochastic system response variable $y \sim \mathcal{W}(T, b)$ which is meant to hold with constant failure-probability parameterization by the shape parameter b for each combination of factor level changes. For the assessment, the following statement is used:

$$\mathrm{H}_0 \colon \theta_i = 0; \;\; \mathrm{H}_1 \colon \theta_i \neq 0. \tag{13}$$

With this approach, two wrong decisions may be likely due to chance: H_0 is rejected although it is true (type-I-error); or H_0 is not rejected although it is false (type-II-error). The probabilities of these errors result in α for the type-I and in β for the type-II error. Consequently, the power of a test can be calculated numerically using Monte Carlo. This takes place via the proportion of cases of iterations in which an effect is assessed as significant by the p-value. It thus results in the probability of correctly recognizing an existing effect as such and defines

$$\text{power} = 1 - \beta. \tag{14}$$

In industrial experimentation, power of around $80\ \%$ and more is generally considered as satisfactory [13]. Therefore, the LRT serves as the statistical procedure to assess the significance of model parameters in GLMs, that is, it performs a significance assessment in the same way as ANOVA does for normally distributed data (or normally distributed residuals). It is based on the comparison of the likelihoods of two models: a comprehensive (complete) model that includes the parameter or parameter group of interest ($L(\hat{\theta}_i)$)and a reduced model that excludes these parameters ($L(\hat{\theta}_{-i})$). The basis of the likelihood ratio test is the ratio of the maximum likelihoods of both models. Mathematically, the test value LR is defined as [6,12]:

$$LR = -2 \ln \frac{L(\hat{\theta}_{-i})}{L(\hat{\theta}_i)}. \tag{15}$$

Assuming the null hypothesis that $df = 1$ excluded parameter in the reduced model has no significant influence on the target variable, the test value LR conforms to a χ^2_{df} distribution with df degrees of freedom. The p-value is thus determined by

$$p \;=\; P\left(\chi^2_{df} \geq \mathrm{LR}\right). \tag{16}$$

If $\alpha = 0.05 > p$ applies, the parameter considered can be assessed as significant to the model equation.

3. EXTRAPOLATION-OPTIMIZED RESPONSE SURFACE DESIGN

Various fundamentals and boundary conditions were presented to this point, which offer tools and motivation for accelerating and optimizing test designs with regard to extrapolation in ALT. From the field of algorithm development for optimal test designs, the reasons why these have not yet been fully researched and presented are outlined. Simultaneously, a reference is made to [14-17], in which it has already been shown that manipulations of a CCD forming an RSM and its orthogonality mainly cause two circumstances: marginally disadvantageous or usable beneficial impairments in efficiency by omitting or shifting individual test points, quite tolerable effects on the coefficient estimation and performance guarantee by power with simultaneously optimized extrapolation properties.

To summarize:

- CCDs offer fundamentally researched advantages in forming conventional RSD using ALT [6,7];
- Optimal RSD does not exist regarding Weibull-GLM and extrapolation in prediction [10,12];
- Effects of omitting or rearranging test points in a CCD offer clear advantages: run 6, run 8 and, for instance, run 13 (c.f. Figure 1) could be rearranged in the extrapolation direction for better prediction variances [17].

It is precisely these changes that are analyzed below. In connection with this, the adaptation is evaluated through the parameter estimation within a Weibull-GLM as well as the characteristics and differences in the model parameter estimation, the power and the optimality criteria. First, exemplary model parameterizations of an effective area and the simulation approach are presented for the study.

3.1. Study Approach

In order to illustrate the outlined potentials and their effects when a CCD is applied within RSM, a two-dimensional model is implemented as a baseline in this study: a quadratic response surface with interaction through x_1 and x_2. This results in a synthetically generated second-order model based on Eq. (1). To provide a comparable dataset and for simplicity, a default system model is specified with a predetermined value for each model parameter θ_i. They compose the (physical) multivariate lifetime response in hours depending on b as the dispersion paramter and T as the expectation composition derived from the GLM. In addition, another normally dispersed measure is superimposed on β_j. Since the regression coefficients of the GLM determine the effects of system influences x_1 and x_2 on t_i, variance is intended to simulate random deviations caused, for example, by pure randomness, generally varying system behavior or measurement errors by means of industrial experimentation. This results in the model compositions according to Table 1, which generates the following system behavior based on Eq. (5) including scatter:

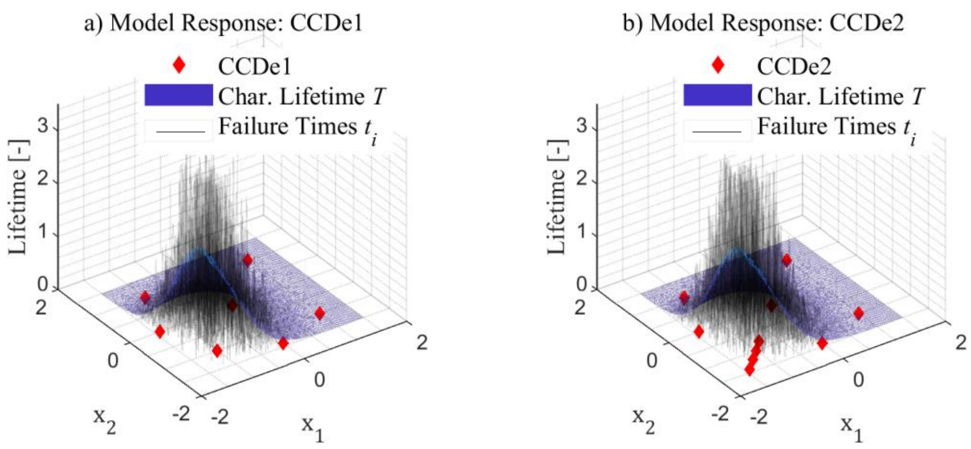
$$T = \exp(\beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2), \qquad \beta_j \sim \mathcal{N}(\beta_j, \sigma). \tag{17}$$

Additionally, to broaden the basis of evaluation, both a more distinct failure mode with moderate variance $b = 2.5$ is considered in Example 1 on the one hand, rather representative for a precise fatigue failure, and, on the other hand a second failure behavior is covered in Example 2 occupying more space of time with correspondingly dominant dispersion (lower failure rate, $b = 1.4, \sigma = 0.2$), compare Table 1. The default model shows, defined in normalized terms, that factor x_1 is twice as dominant as x_2, and the interaction $x_1 x_2$ has a positive effect on lifetime. The quadratic effects are equivalent.

Table 1: Parameterizations of the Synthetic Model

#	b	β_o	β_1	β_2	β_{12}	β_{11}	β_{22}	σ
Example 1	2.5	10	-2	-1	1	-1	-1	0.2
Example 2	1.4	10	-2	-1	1	-1	-1	1.0

Figure 2: Response Surfaces with the Considered Test Designs



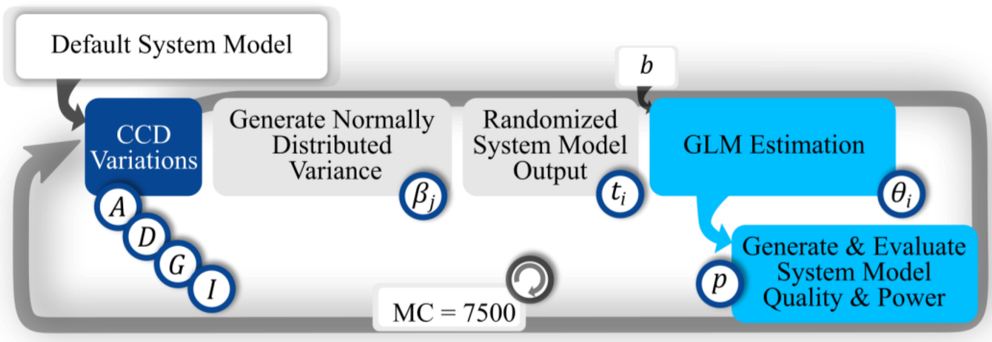
This model is finally fed into the simulation setup with MC. For each MC iteration, a selected experimental design is applied to the failure behavior of Eq.6 and Table 1 with randomly drawn t_i [h] for each stress level. Figure 2 exemplifies the response surface with superimposed noise from Weibull-distributed failure times over the considered parameter space and the corresponding edited designs. Three designs are considered: a default CCD as benchmark and two modifications of it according to the introduction of Section 3.
The CCDs with orthogonality deviations investigated in the studies with normal distributed data from [14,17] are adopted here: the CCDe1, c.f. Figure 2-a), with only 4 central points (run 9 to run 12) and omission of the northern (run 8) and eastern (run 6) star points within the more accelerated region, as it was shown here that potential experimental savings in EoL with tolerable impairments can pay off. And secondly, the CCDe2 with the same setup as in the first manipulation but 3 repositioned runs within the extrapolation field level region on the southwest directed radii $\sqrt{0.8^2+0.8^2} \approx 1.31$, $\sqrt{1.2^2+1.2^2} \approx 1.69$ and $\sqrt{\alpha_D^2+\alpha_D^2} = 2$, see Table 2

Table 2: Experimental Design Setups (× = Omitted / Repositioned Runs)

Design	Run	1	2	3	4	5	6	7	8	9	10	11	12	13
CCDe1	x_1	-1	+1	-1	+1	-α_D	×	0	×	0	0	0	0	×
	x_2	-1	-1	+1	+1	0	×	-α_D	×	0	0	0	0	×
CCDe2	x_1	-1	+1	-1	+1	-α_D	-0.8	0	-1.2	0	0	0	0	-α_D
	x_2	-1	-1	+1	+1	0	-0.8	-α_D	-1.2	0	0	0	0	-α_D

and Figure 2-b). A positioning of this proportion from the center of the parameter space may occur in acceleration benchmarks for industrial experiments and is exemplary for the present simulation study. This, however, preserves the efficiency in terms of the total number of runs to be carried out, even without adjusting for the presumably longer test time at low stress levels.

Figure 3: Procedure of the Applied Simulation Approach



Eventually, at each run the failure time of the synthetic system is transferred to the parameter estimation process from Section 2. Here, b is considered as prior knowledge, since the error mechanism is given by the shape anyway and the coefficient estimate is not unnecessarily falsified depending on a Weibull parameter estimate. As an outcome, the estimated set of model coefficients and the p-values are thus saved for each iteration whereas the replication number of the design corresponds to 1. Finally, this results in generic statements about overall properties of the estimated coefficients, specifying the respective median coefficients and the power as the proportion of MC iterations for which $p < a$ holds. The MC simulation is carried out with a volume of 7,500 iterations, compare Figure 3. This ensures sufficient reproducibility and a tolerable variance $< 2\,\%$ in the power calculation as it reaches numerical saturation - the scatter does not decrease further for increasing numbers of iterations.

3.2. Results

In order to derive generic findings from the procedure presented in Section 3.1, the results of the parameter estimate for both Examples (see Table 1) are first presented and then analyzed. The performance in the form

Table 3: Model Estimate Median Results and Power (shading highlights remarkable deviations)

Origin	Estimate	β_0	β_1	β_2	β_{12}	β_{11}	β_{22}
Example 1 ($b = 2.5; \sigma = 0.2$):							
MC	β_j [-]	+9.999	-1.999	-0.998	+0.999	-0.992	-0.998
CCD	$\hat{\beta}_j$ [-]	9.977	-1.998	-1.001	+0.992	-1.047	-1.057
	Power [%]	100	100	99.56	96.69	99.93	99.95
CCDe1	$\hat{\beta}_j$ [-]	+9.978	-2.022	-1.028	+1.024	-1.084	-1.089
	Power [%]	100	99.99	99.21	97.17	99.03	99.01
CCDe2	$\hat{\beta}_j$ [-]	+10.019	-2.018	-1.019	+1.070	-1.078	-1.078
	Power [%]	100	99.99	99.76	99.76	99.57	99.59
Example 2 ($b = 1.4; \sigma = 1.0$):							
MC	β_j [-]	+9.994	-1.991	-0.992	+1.000	-1.008	-1.015
CCD	$\hat{\beta}_j$ [-]	+10.595	-1.912	-0.953	+0.948	-1.580	-1.607
	Power [%]	100	98.13	82.19	71.49	91.65	92.05
CCDe1	$\hat{\beta}_j$ [-]	+11.532	-1.997	-1.008	+1.055	-1.924	-1.974
	Power [%]	100	91.99	74.39	68.61	85.39	85.81
CCDe2	$\hat{\beta}_j$ [-]	+11.949	-1.133	-0.248	+2.065	-1.503	-1.640
	Power [%]	100	87.37	77.84	87.76	87.37	89.12

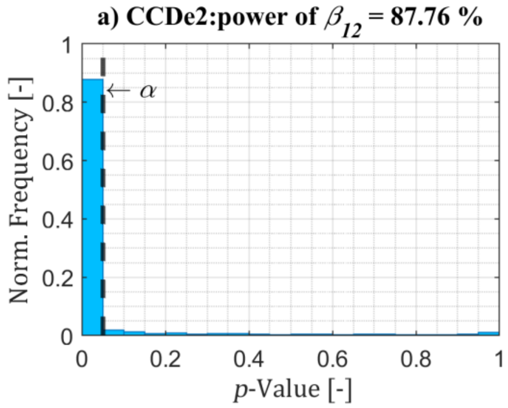
of the optimality metrics for confidence intervals in the coefficient estimation is presented using the experimental design variants from Table 2, c.f. Table 4. Here, CCDe1 and CCDe2 show increasing but reasonable average values for the widths of the CI (*A*), whereby direction-dependent optimizations as shown in [17] cannot be detected at this point. The volume (*D*) of the CI coefficients remains minimal. The maximum SPV (*G*) decreases comparatively. The average SPV (*I*) is constant. Further, the assessed model parameters, which are estimated for each MC iteration on the grounds of randomized lifetimes, are estimated by GLM solving and recorded at the median, see Table 3. In this way, associated p-values are also approximated to the extent of the MC iterations. The resulting power as a probabilistic proportion of the MC for the cases in which an estimated coefficient is correctly recognized as such is also evaluated by the respective deviation. This directly correlates the ability to predict (Optimality Criteria), the performance in effect estimation (Model Parameters) and the probability of successful test performance (Power), see Table 3 and Table 4.

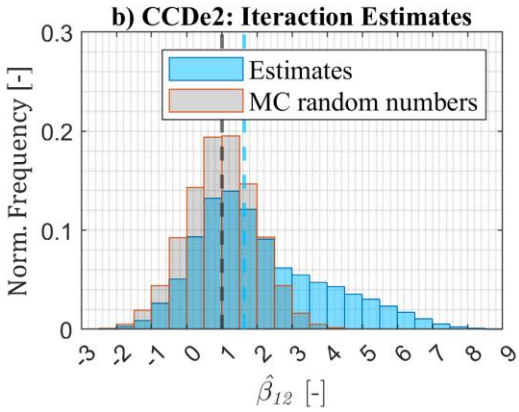
Table 4: Optimality Criteria Metrics

	A	D	G	I
CCD	0.875	3.8e-6	0.600	6
CCDe1	1.371	3.5e-5	0.467	6
CCDe2	1.204	6.9e-6	0.405	6

From this, the following findings emerge from the present study. In general, the implementation and derivation of generic effects through interventions in the design or orthogonality of established experimental designs such as the CCD is quite viable and provides corresponding statements on estimation quality and probability of success, see Table 3 and c.f. [14-17]. For both examples under consideration, the randomized coefficient's mean of the MC as well as the model parameter estimates with power values for the respective design from Table 2 are collected for all model parameters. Remarkable deviations are shaded in color. The influence of the Weibull distribution and coefficient dispersion considered here is present and clearly derivable. While in Example 1 there are no relevant deviations or anomalies in the estimation of the model parameters (max. 8.9 % for β_{22} in CCDe1) or the power values (consistently > 96 %), the situation is different in Example 2. The parameterization considering $b = 1.4$ and $\sigma = 1.0$ causes the estimate of the quadratic effects to deviate by at least around 50 % in each case. For the interaction $\beta_{12} = +2.065$, the CCDe2 is off by more than 100 %. The power for β_{12} reaches a value of less than 80 % in the case of the original CCD and the "impaired" design CCDe1. The power for $\hat{\beta}_2$ also drops below 80 % for the deviant influence in the case of the manipulated designs. This results in a worthwhile look at the histograms of the estimation patterns, c.f. Figure 4.

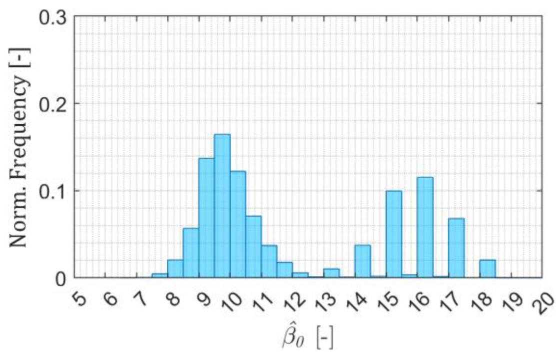
Figure 4: Result Analysis $\boldsymbol{\beta_{12}}$





If p-values are not smaller than the significance level, examplified using $\hat{\beta}_{12}$, they follow a uniform distribution. There is therefore a consistently low risk of incorrectly not recognizing an existing effect as such, see Figure 4 a). If the effect is recognized, see Figure 4 b), it deviates strongly from the specification. Here, for example, the random MC effects with normally distributed scatter can be seen compared to the histogram of the estimated values. A distribution of the estimates indicated right skewness, which is caused by the scatter and lifetime distribution. It is therefore worth taking a look at Figure 5, which visualizes the distribution of the estimated intercept $\hat{\beta}_0$ of the same design ad example in a histogram. Here it can be seen that on the basis of the likelihood from Equation 7, the estimator identifies two local optima in the limits of the variance. Based on the expression of the likelihood equation, this in turn naturally determines the estimation of the other model parameters. In other words, the response surface can be fitted variably at an angle.

Figure 5: CCDe2: Intersection Estimates



3.3. General Findings

In order to briefly summarize the results shown in Section 3.2., the following observations can first be drawn in general terms from the study:

- the paper completes missing links in design adaptations proposed in [17] with respect to SPV regarding their impact on generic deviations in power and effect estimation within a Weibull-GLM;
- the overall result regarding the effects shows that the power due to the adjustments remains/is in a tolerable to good range, the coefficient estimate may have offsets. This is due to the fact that increasing scattering favors several most-likely local optima for the response surface parameterization and GLM fit, c.f. Figure 5, and causes the median to deviate to a greater extent;
- the distribution of non-significant effect analyses is usually evenly distributed, the distribution of the estimated parameters is skewed due to the lifetime distribution and variance;
- as a result, there is a need to further check the prediction capability – e.g. with conformation statistics;
- the intended adjustment of the CCD layout causes the median total test duration to change by -5.64 % (CCDe1) and +123.92 % (CCDe2) due to the relocated "validation runs".

4. CONCLUSION

The present work numerically approximates the effects of proposed design adjustments on effect estimation and power with the intention of extrapolation optimization - based on the most pragmatic plan selection, the CCD. According to the domain of reliability engineering, lifetime distributed data is used for which two lifetime models are examplarily specified. Firstly, it is concluded that the effects of design adjustments on the coefficients are generally moderate in the case of lifetime-distributed data via GLM, meaning that a directional improvement in the prediction variance by adjusting the designs toward field level is justifiable. On the other hand, it can be seen that the effects on the estimated model coefficients can differ for varying failure parameterizations due to aleatoric and epistemic uncertainties. From this, potential for further parameter studies regarding a general Weibul-approach can be derived directly.

References

[1] B. Bertsche, "*Reliability in Automotive and Mechanical Engineering: Determination of Component and System Reliability*", (1st ed.), Springer, 2008, Berlin.

[2] M. Modarres, M. Amiri and C. Jackson, "*Probabilistic Physics of Failure Approach to Reliability*", (1st ed.), Wiley & Sons Ltd., 2017, Hoboken NJ.

[3] G. Yang, "*Life Cycle Reliability Engineering*", (1st ed.), John Wiley & Sons Inc., 2007, Hoboken NJ.

[4] L.A. Escobar and W.Q. Meeker, "*Planning Accelerated Life Tests with Two or More Experimental Factors*", Technometrics, vol. 37, No. 4, pp.411-427, 1995.

[5] W.Q. Meeker, L.A. Escobar and F.G. Pascual, "*Statistical Methods for Reliability Data*", (2nd ed.), John Wiley & Sons Inc., 2022, Hoboken NJ.

[6] R.H. Myers, D.C. Montgomery and C. M. Anderson-Cook, "*Response Surface Methodology: Process and Product Optimization Using Designed Experiments*", (4th ed.), John Wiley & Sons Inc., 2016, Hoboken NJ.

[7] P. Goos and B. Jones, "*Optimal Design of Experiments: A Case Study Approach*", (1st ed.), John Wiley & Sons Inc., 2011, Hoboken NJ.

[8] A.C. Atkinson, A.N. Donev and R.D. Tobias, "*Optimum Experimental Designs, with SAS*", (1st ed.), Oxford University Press Inc., 2007, New York.

[9] A.I. Khuri, "*Response Surface Methodology and Related Topics*", (1st ed.), World Scientific Publishing Co. Pte. Ltd., 2006, Singapore.

[10] K. Hinkelmann, "*Design and Analysis of Experiments - Volume 3: Special Designs and Applications*", (1st ed.), John Wiley & Sons Inc., 2008, Hoboken NJ.

[11] R.H. Myers, D.C. Montgomery, G.G.Vining and T.J. Robinson, "*Generalized Linear Models: with Applications in Engineering and the Sciences*", (2nd ed.), John Wiley & Sons Inc, 2010, Hoboken NJ.

[12] C.J. Wu and M. Hamada, "*Experiments: Planning, Analysis, and Optimization*", (2nd ed.), John Wiley & Sons Inc, 2009, Hoboken NJ.

[13] D.C. Montgomery, "*Design and Analysis of Experiments*", (10th ed.), John Wiley & Sons Inc, 2020, Hoboken NJ.

[14] M. Arndt, P. Mell and M. Dazer, "*Generic effects of deviations from test design orthogonality on test power and regression modelling of central-composite designs*", Proceedings of the Probabilistic Safety Assessment and Management Conference (PSAM16), 2022, Honolulu HI.

[15] P. Mell, M. Arndt and M. Dazer, "*Non-orthogonality in test design: practical relevance of the theoretical concept in terms of regression quality and test plan efficiency*", Proceedings of the Probabilistic Safety Assessment and Management Conference (PSAM16), 2022, Honolulu HI.

[16] M. Arndt and M. Dazer, "*Analysis of Efficiency in Response Surface Designs Considering Orthogonality Deviations and Cost Models*", Proceedings of the 33rd European Safety and Reliability Conference (ESREL), 2023, Southampton.

[17] M. Arndt and M. Dazer, "*Investigating Practical Orthogonality Deviations in RSM for Improved Predictive Reliability Modeling*", Proceedings of Annual Reliability and Maintainability Symposium (RAMS), 2024, Albuquerque NM.

[18] A. Kremer and B. Bertsche, "*A New Approach for Parametrizing Multidimensional Lifetime Models*", Proceedings of the 29th European Safety and Reliability Conference (ESREL), 2019, Hannover.

[19] B. Jones, K. Allen-Moyer and P. Goos, "*A-optimal versus D-optimal design of screening experiments*", Journal of Quality Technology, vol. 53, No. 4, pp.411-427, 2021.