# Field-Reliability Predictions based on Statistical System Life Cycle Models*

Lukas Felsberger[1,2], Dieter Kranzlmüller[1], and Benjamin Todd[2]

[1] Institut für Informatik, Ludwig Maximilians Universität Muenchen, Oettingenstr. 67, D-80538 München, Germany
[2] CERN, Route de Meyrin, 1211 Genève, Switzerland

**Abstract.** Reliability measures the ability of a system to provide its intended level of service. It is influenced by many factors throughout a system life-cycle. A detailed understanding of their impact often remains elusive since these factors cannot be studied independently. Formulating reliability studies as a Bayesian regression problem allows assessment of their impact simultaneously and to identify a predictive model of reliability metrics.
The proposed method is applied to currently operational particle accelerator equipment at CERN. Relevant metrics were gathered by combining data from various organizational databases. To obtain predictive models, different supervised machine learning algorithms are applied and compared in terms of their prediction error and reliability. Results show that the identified models accurately predict the mean-time-between-failure of devices – an important reliability metric for repairable systems - and reveal factors which lead to an increased dependability. These results provide valuable inputs for early development stages of highly dependable equipment for future particle accelerators.

**Keywords:** Reliability Prediction · System Life Cycle · Bayesian Learning.



CERN-ACC-2020-0017
14/09/2020

## 1 Introduction

Reliability measures the ability of a system to perform as expected during its intended lifetime. The "field-reliability" of complex repairable systems is a result of all actions during all stages of its system lifecycle. These stages are (1) conceptual design, (2) detailed design and testing, (3) manufacturing, (4) installation, (5) operation and maintenance, and (6) phase-out and disposal. At each stage an interplay of complex technical, organizational and human processes leads to a more or less desirable outcome in terms of system reliability.

An assessment of all stages and processes is not feasible since models capturing the interactions between all relevant processes in system development do not

* This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 05E12CHA). The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-99740-7_7.

exist. Therefore, most common reliability methods focus on certain stages and aspects during a system life cycle which can be modeled and understood - an overview will be given in section 2. However, such methods struggle to quantify the overall uncertainty of reliability predictions in a systematic way since contributions by disregarded relevant processes during a system life cycle are not straight-forward to include.
Instead of focusing on models for certain stages and aspects of a system we propose to learn a statistical model of the whole product life cycle to predict the observed field-reliability with machine learning techniques as depicted in figure 1. For a set of existing comparable systems with known field-reliability so-called "quantitative reliability-indicators" are gathered. Using the reliability-indicators as input variables and the field-reliability metric as target variables a statistical reliability model is learned by a supervised machine learning algorithm.
The learned model will always be an approximation of the true underlying system life-cycle process. The lost accuracy due to the statistical model and the limited granularity of the reliability-indicators can be quantified by Bayesian methods. Thereby, the overall predictive certainty can be quantified in an efficient way based on the available data.

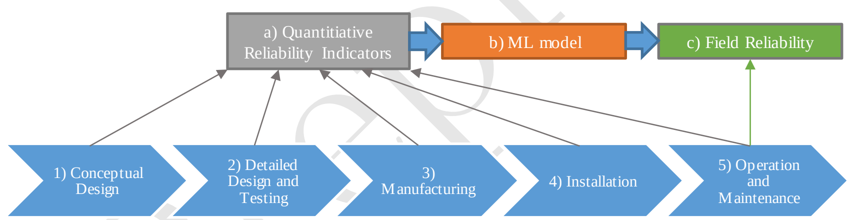


**Fig. 1.** Illustration of the proposed approach. The achieved field-reliability (c) can be seen as the result of relevant processes during the whole product life-cycle (1-5). It is not feasible to capture and model all of the relevant processes. Instead, it is proposed to learn a reduced-order statistical life-cycle model (b) with machine-learning algorithms based on quantitative reliability-indicators (a).

We will demonstrate that the learned models accurately predict reliability metrics even with a limited set of reliability-indicators (as is the case at early stages of a system's life-cycle). Compared to traditional reliability assessment methods this leads to a reduced workload for reliability predictions and to a systematic quantification of uncertainties. Furthermore by an appropriate choice of reliability-indicators and machine learning algorithms one can study the influence of each individual reliability-indicator. This information assists engineers in design decisions for highly reliable systems.

The rest of the paper is structured as follows: In section 2 related methods to reliability predictions are presented. In section 3 the methodology of our approach is explained and in section 4 it is applied in a use-case.

## 2 Literature Review

A general review of the challenges in reliability studies is given in [19]. The author of [19] concludes that the two major challenges in reliability studies are complexity and uncertainty. Reliability studies must consider technical, organizational and human factors each of which influences the field-reliability of systems. In the following paragraph a selection of reliability prediction methods to tackle these problems is given.

*Reliability Engineering Methods* Scientific literature on reliability engineering prediction methods of electronic systems is numerous. An attempt to classify and evaluate the existing methods is given in the IEEE standard 1413 [16, 5] and its successors. In this standard they have been classified as based on

- handbooks,
- stress and damage models (frequently referred to as physics-of-failure based),
- field-data.

Most methods are based on early designs of the considered system and the selected components.
A common criticism for handbook based models is that they do not consider interactions of components but only single-component faults. However, faults due to single-component failures are not dominant [15, 4, 6, 1, 11]. As a result the actual field reliabilties can deviate from the predicted ones by orders of magnitude [10]. The author of [4] argues that some methods should not be seen as "field-reliability prediction methods" and rather as part of a review process at a stage when limited information on the final design is available.
Stress- and damage models are in general more accurate than handbook based methods. However, the development of such methods requires more effort [15].
Instead of assessing the system on the component level, some approaches use a top-down approach in which the field-reliability of new systems is estimated from field-data of similar systems in operation [8, 9].

*Reliability Program Assessment* A different approach to evaluate the field-reliability of systems is taken in [13]. The likelihood of achieving the required field-reliability is estimated by a review of the design process. Each system is assigned a score depending on its design process and it is shown that this score correlates with the probability of fulfilling field-reliability requirements. Thereby organizational aspects of reliability are taken into account.

*Organizational and Human Reliability Analysis* In the review article [19] section 3.1.3 is dedicated to non-technical factors in reliability studies since its contribution to the field-reliability can be significant.

In our work we propose to infer the most relevant processes or factors in a system life-cycle from the field-reliability data of a set of systems. This allows to include organizational and human reliability factors. The method can be applied at any stage of a system life-cycle to guide engineering decisions.

## 3 Methodology

In this section the relevant definitions are made, the methods used are explained and the general methodology is described.

### 3.1 Definitions

*System reliability* It is generally defined as the ability of a system to provide its intended level of services for a specified time $t$. For repairable systems it is usually measured as availability $A$ which is defined by

$$A = \frac{MTBF}{MTBF + MTTR} \tag{1}$$

with $MTBF$ being the mean-time-between-failure and $MTTR$ being the mean-time-to-repair. The $MTBF$ is being calculated as

$$MTBF = \frac{t_{operation}}{n_{faults}} \tag{2}$$

with $t_{operation}$ being the cumulative operational time of the considered devices and $n_{faults}$ being the total number of faults within the operational time. The $MTTR$ can be evaluated by

$$MTTR = \frac{t_{inrepair}}{n_{faults}} \tag{3}$$

with $t_{inrepair}$ being the total time a system is in repair and $n_{faults}$ the total number of faults during the operational time. The un-availability $U_A$ is given by $U_A = 1 - A$. Note that a constant failure rate is assumed.

*System life-cycle* It is the overall process describing the lifetime of a system. It is a concept from systems engineering to address all stages of a product from its beginning to end. Here these stages shall be divided into (1) conceptual design, (2) detailed design and testing, (3) manufacturing, (4) installation, (5) operation and maintenance, (6) and phase-out and disposal. [3]

[3] Depending on the system under study the definitions of the stages may change. The proposed methodology is not restricted to this specific choice of stages.

*System definition* This discussion is focused on repairable electronic systems. A more precise definition will be given for the use case in section 4. [4]

### 3.2 Method

The central assumption is that the observed field-reliability is the outcome of all technical, organizational and human processes during all stages of a system's life-cycle. It is unfeasible to model all these interactions due to their complexity and non-linearity. Therefore, we restrict ourselves to learning a statistical model of the observed field-reliability of comparable systems based on reliability-indicators collected throughout the system life-cycle. Modern machine learning algorithms are capable of learning accurate predictive models of field-reliability based on the relevant reliability-indicators. The loss of information due to the limited availability of data and the intrinsic uncertainty of the problem can be assessed by using Bayesian machine learning methods.

**Life Cycle Analysis by Machine Learning** To arrive at a firm mathematical description of the proposed method let us hypothesize the existence of a deterministic model $\mathbf{F} : \mathcal{Z} \mapsto \mathcal{Y}$ to determine any field-reliability metric $\mathbf{Y} \in \mathcal{Y}$ from all relevant input variables $\mathbf{Z} \in \mathcal{Z}$ in the form of

$$\mathbf{Y} = \mathbf{F}(\mathbf{Z}). \tag{4}$$

This would be a model to quantify the contribution of all relevant process towards the field-reliability during the whole system life cycle. Since it is not possible to derive such a formula or to gather all relevant inputs for practical purposes we try to approximate the true field-reliability metrics $\mathbf{Y}$ by a reduced model

$$\mathbf{Y} \approx \mathbf{y} = f(\mathbf{x}), \tag{5}$$

with $\mathbf{x} \in \mathcal{X}, dim(\mathcal{X}) \ll dim(\mathcal{Z})$, being the set of collected reliability-indicators and $f : \mathbf{x} \mapsto \mathbf{y}, \mathbf{y} \in \mathcal{Y}$ being an approximate model. When supplied with pairs of input and output data $\mathcal{D} = \{(\mathbf{x_1}, \mathbf{Y_1}), ..., (\mathbf{x_N}, \mathbf{Y_N})\}$ a statistical learning algorithm can learn such a model by minimizing a certain loss function $l : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}$. This is essentially a regression problem and there exists a vast range of learning algorithms for this.

For our purposes we prefer learning algorithms which fulfill three additional requirements. Firstly, to quantify the uncertainty of the predictions of the reliability metrics, we want to learn probabilistic models

$$p(\mathbf{Y}|\mathbf{x}). \tag{6}$$

[4] There is no implicit restriction for the proposed method to electronic repairable systems. It can also be used for non-repairable systems and for mechanic, electric, electronic or software systems. However, the definitions of the fault metrics must be adapted.

Our method will be based on an arbitrary non-linear mapping from the reliability-indicators to features $\Phi : \mathcal{X} \mapsto \mathbb{R}^n$. Since it is of interest which features are relevant, secondly, it is beneficial if the learned models are of a parametric form

$$p(\mathbf{Y}|\mathbf{w} \cdot \Phi(\mathbf{x})), \tag{7}$$

with $\mathbf{w} \in \mathbb{R}^n$ being a weight vector indicating the relevance of each feature. Thirdly, some methods prefer models in which only a few features are relevant. These are preferred from a practical point of view since that requires a reduced data collection effort for predicting field-reliability. A general justification of such methods on philosophical grounds is given by Occam's razor [7].
Concrete algorithms fulfilling these criteria will be presented in section 4. Even though the outlined requirements are not mandatory, they greatly facilitate the data collection and model assessment process by providing direct feedback as will be discussed in the following paragraphs.

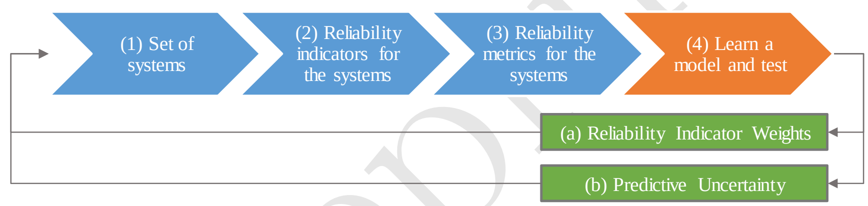


**Fig. 2.** Illustration of the iterative data collection and reliability prediction process. The choice of (1) systems, (2) reliability indicators and (3) reliability metrics influences the quality of the predictive model (4). The learning algorithm provides feed-back in the form of relevance weights for the reliability indicators (a) and uncertainty bounds for the field-reliability predictions (b).



**Data Collection and Reliability Prediction** As briefly outlined in the introduction, the proposed method is based on a statistical analysis of system life-cycles of comparable systems by a machine learning approach. To carry out the method a data-set $\mathcal{D}$ has to be compiled. To do so, several choices in terms of (1) comparable systems, (2) relevant reliability-indicators, and (3) reliability metrics have to be made. A summary of this process is illustrated in figure 3.2 and will be discussed in detail below.
As can be seen in figure 3.2 this is an iterative process. An algorithm with the desired properties as in equations 6 and 7 can systematically guide the user through this process by providing quantitative feedback on the quality of the collected data. However, when solely relying on the feedback of the algorithm one might encounter some pitfalls which can be avoided by following further guidelines given below.

*Collection of Training Systems* Since the method is based on the field-reliability of existing comparable systems, the choice of the collected systems will have an influence on the accuracy of the predictions for future systems. Three general recommendations can be given for this selection:

- Only systems which have been in use for a significant exploitation period with accurately monitored reliability metrics shall be used.
- The choice of systems for which a field-reliability model is learned shall include systems which are comparable to the system for which a field-reliability shall be predicted. In reliability studies comparable systems are similar in terms of technical, organizational and human factors throughout their life-cycle.
- The set of chosen reliability-indicators and reliability metrics has to be available for both the existing systems and those for which a prediction should be obtained.

*Collection of Reliability-Indicators* The choice of these indicators largely influences the quality of the model in terms of its accuracy and interpretability. Following statements can be made:

- Based on expert knowledge, recommendations can be given for indicators which carry important reliability information. E.g. operational conditions such as load, temperature or humidity can contribute significantly to a failure rate. Systems which are mass-produced will achieve different field reliabilities than prototypes. Products developed by highly-skilled experienced engineers are also expected to behave differently than those developed by un-skilled and unexperienced engineers [14].
- In engineering practice the collection of data is facing practical limitations due to time or other restrictions. Therefore, a natural choice is to begin to collect the indicators which require the least collection effort. For the use-case in section 4 it will be shown that accurate predictions can be obtained from a very limited set of meta-variables as reliability-indicators. Furthermore, one always needs to consider the availability of the indicators for all systems in the data-set.

*Collection of System Reliability Metrics* The choice of reliability metrics is usually given by the system under study. For our choice of system and assuming a constant failure rate[5], these are given by $MTBF$ and $MTTR$. Based on these other metrics can be derived.

*Training a Reliability Model* Having chosen a set of systems, reliability-indicators and reliability metrics one is able to compile a data-set $\mathcal{D} = \{(\mathbf{x_1}, \mathbf{Y_1}), ..., (\mathbf{x_N}, \mathbf{Y_N})\}$

[5] This assumption can be relaxed by e.g. predicting a parameterized failure rate distribution over time. Then, instead of $MTBF$ and $MTTR$ the reliability metrics are the parameters of the distribution. This requires a different data collection and can be considered for future work.

for which $\mathbf{x_i}$ and $\mathbf{Y_i}$ are the collected reliability-indicators and the field-reliability metrics for system $i$, respectively. This is the data-set based on which a reliability model shall be learned.
The general machine learning approach is to split the data-set into a training data-set $\mathcal{D}_{train}$ and a test data-set $\mathcal{D}_{test}$ [6]. Based on the training set, a learning algorithm will find a model as in equation 5 for which we can evaluate its accuracy with the test data-set $\mathcal{D}_{test}$. Based on this accuracy estimate it is neither possible to evaluate the confidence of the predictions nor to assess the relevance of the selected reliability-indicators. Assuming a learning algorithm is used which satisfies equations 6 and 7 one is able to do so.
The confidence or uncertainty of the predictions provides feedback on the amount and quality of the collected data. The weight vector $\mathbf{w}$ indicates the relevance of the features and the reliability-indicators. Depending on the complexity of the mapping $\Phi : \mathcal{X} \mapsto \mathbb{R}^n$ from the reliability-indicators to the features we can identify the most important reliability-indicators. Using this information and expert knowledge, we can iteratively refine our data-set (choice of systems and reliability-indicators) and feature mapping $\Phi$ and confirm iterative improvements by comparing the model predictions with the true field-reliabilities in the test data-set[7].

*Reliability Prediction by the Learned Model* Once a model is trained and validated we can use it to predict the field-reliability of new systems. This is done by providing the collected or estimated reliability-indicators $\mathbf{x_{j,predict}}$ for the new system $\mathbf{j}$. Having trained a model satisfying formula 6, it is possible to estimate the uncertainty of its predictions - a valuable input for further studies based on the predictions.

The iterative data collection and the reliability prediction process will be illustrated for a use-case in section 4. The feedback of the algorithms in terms of chosen systems, reliability-indicators and features will be assessed.

## 4 Use Case

In this section the proposed method is used to learn a model for the expected field-reliability of the next generation of accelerator power converters at the conceptual stage. It is based on a statistical life-cycle model learned by field-data of the current and past generations of power converters. The system of interest, the collected data and features, the used learning algorithms and the results will be discussed.

[6] Alternatively models can also be trained by cross-validation and commonly a few data-items are left out for later validation purposes.

[7] To avoid an overfitting of the test-data we recommend to shuffle the data items between training and test data-set during the iterations.

**System Definition** The considered systems are magnet power converters at the CERN particle accelerator facilities. A power converter is a device to transform electrical energy. The conversion is in terms of voltage, current and waveform. Magnet power converters control the flow of current through particle accelerator magnets. In order to achieve precise magnetic fields these converters generally need to control the output current very precisely.

### Dataset, Reliability Metrics and Reliability-Indicators

*Set of systems* At CERN there are currently more than 6000 power converters of approximately 600 different types in use. Their field-reliability is continuously tracked by a centralized computerized maintenance management system (CMMS). After removing converter types with a cumulative operational time $t_{operation}$ of less than ten years and cleaning the data, approximately 300 power converter types can be used for reliability analysis. Table 1 gives an overview of minimal and maximal characteristic attributes of power converters in the dataset.

**Table 1.** Illustration of characteristic power converter attributes of the studied dataset.

| | **Power [W]** | **Current [A]** | **Voltage [U]** | **Age [yrs]** | **MTBF [hrs]** |
|---|---|---|---|---|---|
| **Minimum** | $10^{-6}$ | $10^{-4}$ | $10^{-3}$ | 2.2 | $10^{3}$ |
| **Maximum** | $10^{8}$ | $4 \cdot 10^{4}$ | $10^{5}$ | 49.7 | $6 \cdot 10^{5}$ |



One has to note that the systems operate in various environments (laboratories, experiment tunnels, outside) and are sometimes exposed to radiation. Some converters are developed in-house and some are commercially available. Considering the vast range of converter types and operational environments one would not expect a global model to accurately predict the field-reliability. Therefore, both local- and global-models will be trained.

*Reliability indicators for the systems* The initial choice of reliability-indicators depends on

- the system development stage at which the prediction shall be carried out,
- recommendations from system experts,
- the availability of data and time or effort which can be attributed to the analysis.

For our use case a conceptual design stage shall be considered. In that case, only the rated power ($P$), rated voltage ($U$), rated current ($I$) and quantity of each power converter type shall be available. Clearly these (meta-)indicators are absolutely incapable of capturing the complex processes which occur during a

system life cycle and influence the field-reliability. However, all information and processes which are not captured shall result in an increased uncertainty in our predictions of the field-reliability.
To include expert knowledge, a power-converter classification scheme was developed with equipment experts at CERN. The following list of additionally collected reliability-indicators is based on a trade-off between the recommendations from the experts and restrictions due to the availability of data:

- Avg. Age: The average age of converters for each converter type. Depending on a reactive or preventive maintenance strategy a decreasing or constant availability as a function of the age is expected, respectively.
- Cum. Age: The cumulative age of converters for each converter type. A dependency of the availability on the cumulative age could indicate both a organizational learning curve in terms of a more efficient maintenance and a degradation with age of the converters.
- Pol 0-9: The polarity of the converter. This indicates the operating modes, technology and complexity of the converter[8].
- Acc. 1-9: The accelerator in which the converter type is used. Depending on the accelerator the converter type is exposed to different operating conditions[9] and operation modes.
- in Acc.: The number of different particle accelerators in which each power converter is used.

We can probe different indicators for their information content by appropriate Bayesian learning methods. The required learning algorithms will be introduced later in this section.

*Reliability metrics for the systems* The studied field-reliability metrics are $MTBF$, $MTTR$, Availability $A$ and Un-Availability $U_A$ as defined in section 3[10]. These are directly computed in the CMMS with the necessary variables for power converter type $i$ which are defined as follows:

- $t_{operation,i}$: Cumulative time in operation of all converters of converter type $i$. Note that commissioning and testing times are not counted towards operation time.
- $n_{faults,i}$: Cumulative number of faults of all converters of converter type $i$ during the operational time $t_{operation,i}$. Note that only internal faults of

[8] The discrete set of polarities is given by: (1) Unipolar, (2) Bipolar Switch Mechanic, (3) Bipolar I - Unipolar U - 2 Quadrants, (4) Unipolar I Bipolar U 2 Quadrants, (5) Bipolar Pulse-Width-Modulation, (6) Bipolar Relay, (7) Bipolar Electronic I/U, (8) Bipolar Anti-Parallel 4 Quadrants, (9) Bipolar I-circulation 4 Quadrants and, (0) un-specified or other Polarity.

[9] E.g. the radiation levels differ on the kind of accelerator. However, there is also different operation conditions within each of the accelerators.

[10] Note that due to space limitations only the results for the $MTBF$ will be presented. Models for other metrics can be requested from the authors.

the system which required an external action to alleviate the problem are included. Internal faults which are automatically resolved or are very short and faults due to external reasons are not included. This ensures that a model for the reliability of the considered systems itself is learned and not of its surroundings.
- $t_{inrepair,i}$: Cumulative time in repair of all converters of converter type $i$ during the operational time $t_{operation,i}$. The repair time starts by a request from the system operators to the system experts and ends when the problem was resolved and the system can continue to operate.

**Algorithms** By formulating the reliability prediction problem as a supervised machine learning problem in principle we can choose from a range of existing learning algorithms to generate the desired statistical model for predictive purposes. Since the uncertainty in the field-reliability predictions shall be quantified (i.e. a model as presented in equation 6), the choice of algorithms is narrowed down. Furthermore, sparse parametric models (as in equation 7) are preferred since they potentially require fewer reliability-indicators to be collected and - more importantly - since they allow an estimation of the relevance of the choice of reliability-indicators and the generated features.

A summary of the chosen algorithms is given in table 2. Note that the scikit-learn python implementations of the algorithms were used [17]. A detailed description of each algorithm can be found on their website and in their user-guide [3]. Since the algorithms are standard implementations, only their parametrizations are be given below:



**Table 2.** Summary of learning algorithms.

| | **UQ (6)** | **Feature Weights (7)** | **Sparsity** | **Global/Local** |
|---|---|---|---|---|
| **ARD** | yes | yes | yes | Global |
| **BAR** | yes | yes | balanced | Global |
| **GP** | yes | no | no | Local |
| **ENCV** | no | yes | yes | Global |
| **SVR** | no | only for linear kernel | no | Local |

- ARD - Automatic Relevance Determination Regression: Sparse Bayesian regression technique as described in [2] - Chapter 7.2.1. The implementation is taken from [3] - Chapter 1.1.10.2.
- BAR - Bayesian Ridge Regression: A Bayesian regression method as introduced in [12]. It is similar to the ARD Regression but fewer parameters have to be determined form the data. The implementation is taken from [3] - Chapter 1.1.10.1.
- GP - Gaussian Process Regression. A kernel-trick based Bayesian Regression technique. The implementation is described in [18] - Algorithm 2.1 and was

taken from [3] - Chapter 1.7.1. The kernel is based on a combination of a radial-basis-function kernel[11] and a white-kernel[12].
- ENCV: Elastic Net Regression with hyper-parameter optimization by cross-validation. The implementation is taken from [3] - Chapter 1.1.5 - which includes a description of the algorithm.
- SVR - Support Vector Machine Regression: A kernel-trick based regression method. A description is given in [3] - Chapter 1.4.2. Linear basis functions are used with manually optimized hyper-parameters[13].

### 4.1 Results and Interpretation

**Iterative Training Procedure** In this section it will be demonstrated that the user is assisted in choosing the reliability-indicators, the set of feature functions and the selected systems in the data set by a learning algorithm with the desired properties as in equation 6 and 7. Instead of exercising the iterative process the final trained model which was obtained after multiple iterations will be presented. Based on that, it will be shown how variations of the above named choices affect the performance of the predictive model and how that is indicated by the learning algorithm.
Unless further specified, all shown results are based on models trained by the ARD algorithm. Throughout the training process all introduced algorithms were used to train the models and compared in terms of their performance. During this process, the ARD algorithm has shown to be of highest practical value and will therefore be regarded as preferred choice. Its prediction accuracy was consistently among the best[14], it learns a sparse parametric model which leads to a decreased data collection effort, it assigns relevance weights to the feature functions and and it quantifies uncertainties of both the field-reliability predictions and the feature function weights. Nevertheless, results obtained by models trained by other algorithms will be stated as well for reference.

*Reference Model* The chosen systems in the data-set and the reliability-indicators were already introduced earlier in this section. All considered systems are power converters which are operated at CERN. A large set of converter types was chosen to ensure that the converters for which the field-reliability shall be predicted are similar to the converters contained in the data-set. The chosen reliability-indicators are in this case largely predetermined by the problem statement and the availability of trustworthy data. The benefit of formulating the field-reliability prediction problem as supervised learning problem is that captured

[11] Parameters: length_scale=10.0, length_scale_bounds=(1e-2, 1e3)
[12] Parameters: noise_level=1e-5, noise_level_bounds=(1e-10, 1e+1)
[13] $C = 10, \gamma = 0.002$
[14] The prediction error was measured for a test data-set by the mean-squared-error (MSE). For each of the problems a table of the achieved MSE by different algorithms will be given.

relevant indicators are selected, the captured irrelevant indicators are ruled out and the non-captured relevant indicators contribute to an increased predictive uncertainty. This is automatically taken care of by the learning algorithm.
Based on the reliability-indicators, following features were generated for the final model:

- Based on the numeric indicators $\mathbf{x_{num}}$ linear features and logarithmic features were chosen - $\Phi(\mathbf{x_{num}}) = [\mathbf{x_{num}}, \log(\mathbf{x_{num}})]^T$.
- The categorical indicators $\mathbf{x_{cat}}$ were split into binary features, whereas the number of binary variables corresponds to the number of categories per categorical variable.

Combining all features a feature vector of 34 dimensions was obtained. Of the 281 different converter types in the data-set, 168 were used to train a model and 113 were used to test the model. The training data set contains converters which are at least 18 years old and the test set contains converters which are younger than 18 years. This is to confirm that the learned model can in principle be extrapolated to future converters [15]. To train the learning algorithms the features were re-scaled by removing their mean and a scaling to unit variance. Furthermore, the logarithms of the reliability metrics were taken instead of their nominal value.
The result of the prediction is shown in figure 3. The green line depicts the mean of the predictive distribution and the green shaded area the 95% confidence intervals. The blue dots mark the actual observed field reliabilites which are almost always contained. Note that the different converter types on the horizontal axis were ordered by the mean of the predictive distribution for illustration purposes.

The feature weights are illustrated in figure 4. Note that the estimates by all considered algorithms [16] are given. This is to show that irrespective of the chosen learning algorithm a similar model is learned. The ARD and the BAR model assign distributions to the weights whereas the other models are based on point estimates. For all models the most pronounced feature is the logarithm of the quantity of converters per type (log(Quantity)).

In table 3 a) the mean squared error (MSE) between the (mean of) the predicted values and the true observed field values is shown. A similar performance is observed for the different algorithms.

Based on this final model we will investigate the influence of changing the set of systems, the set of reliability-indicators and the form of the feature functions with respect to the performance of the learned predictive model.

[15] Note that for the iterative training procedure this splitting by age was not always used.
[16] Except for the GP algorithm which does not assign these weights.

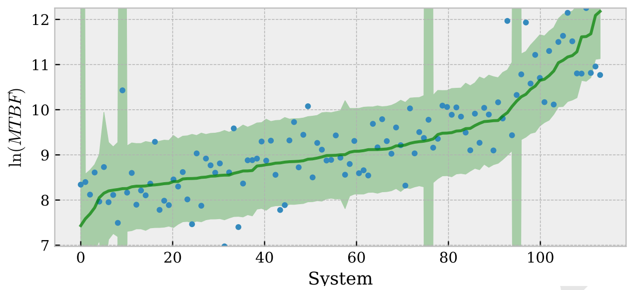


**Fig. 3.** Prediction of the $\log(MTBF)$ with the final model for the test data-set. The green line depicts the mean of the predictive distribution and the green shaded area the 95% confidence intervals. The blue dots mark the actual observed field reliabilites which are almost always contained. Note that the different converter types were ordered by the mean of the predictive distribution for illustration purposes.

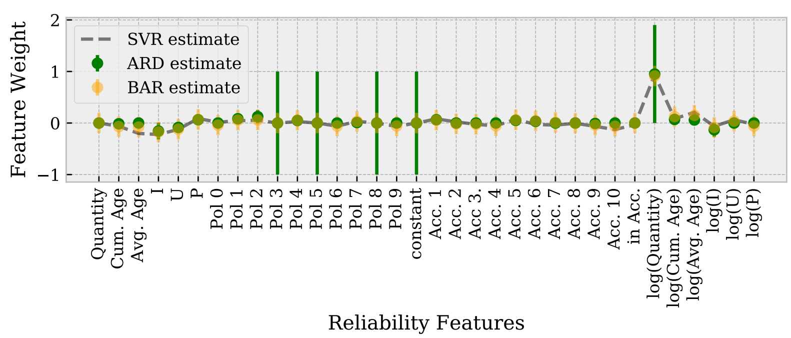


**Fig. 4.** Estimated feature weights for the prediction of the $MTBF$. This is to show that irrespective of the chosen algorithm a similar model is learned. For all models the most pronounced feature is the logarithm of the quantity of converters per type (log(Quantity)).

*Choice of Systems* In order to illustrate the influence of the choice of systems in the data-set we will subsample the training data items. Instead of 168 converter types only 42 types are used to learn a predictive model. The predictions of this model are shown in figure 5. Note that the 95% confidence interval significantly grows in comparison with figure 3 and that the predictive distributions become noisy for some estimates. The BAR learning algorithm has shown to be more robust for fewer data items to learn from. The MSE for the test set is shown in table 3 b).

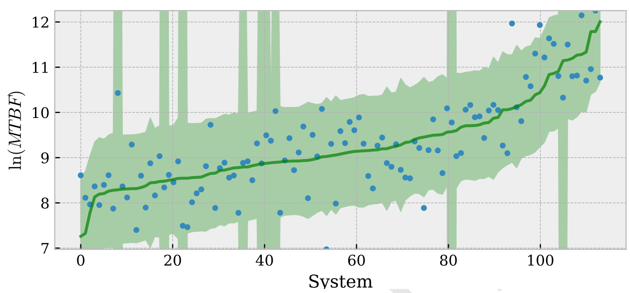


**Fig. 5.** Prediction of the $\log(MTBF)$ with the final model for the test data-set using only 42 instead of 168 training data items. The green line depicts the mean of the predictive distribution and the green shaded are the 95% confidence interval. The blue dots mark the actual observed field reliabilities which are almost always contained. Note that the 95% confidence interval significantly grows in comparison with figure 3.

This illustrates that a reduced amount of systems as training data reduces the confidence of the reliability predictions. The accuracy as measured by the MSE decreases slightly. This is consistent with the expected behaviour.

*Choice of Reliability-Indicators* in this section the influence of the choice of reliability-indicators shall be studied. It is expected that their choice largely influences the quality of the predictions. Since we already know from the training model which features and, hence, which indicators are relevant, we will remove the most important reliability-indicator - the quantity of converters per type - to test if the algorithms are still able to predict the field-reliability with a certain confidence.

Figure 6 shows the predictions of the model as trained by the ARD algorithm for the test data-set. The 95% confidence interval is significantly larger as in figure 3. The removal of a single feature leads to very uncertain predictions. This form of direct feedback greatly facilitates the choice of reliability features.

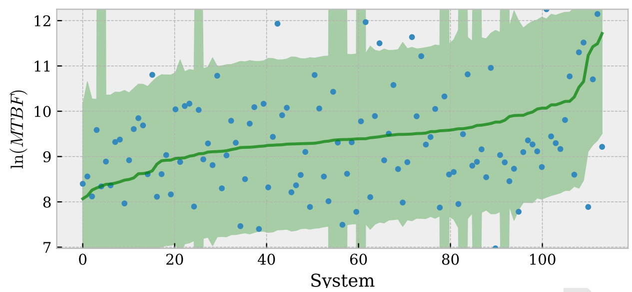


**Fig. 6.** Prediction of the $\log(MTBF)$ with the final model for the test data-set using all reliability-indicators except for the quantities of converters per type. The green line depicts the mean of the predictive distribution and the green shaded area the 95% confidence intervals. The blue dots mark the actual observed field reliabilities which are almost always contained. Note that the 95% confidence interval significantly grows in comparison with figure 3.



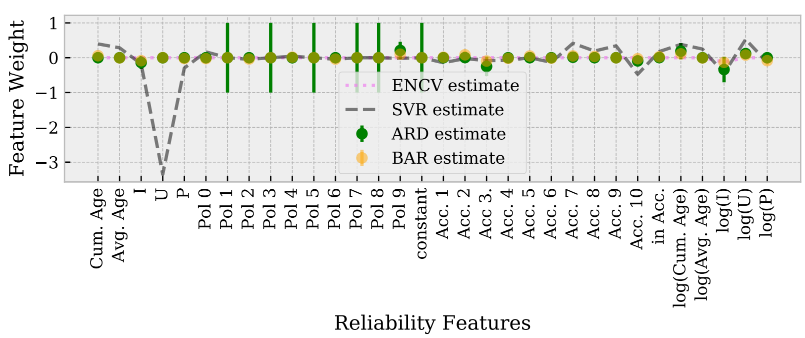


**Fig. 7.** Estimated feature weights for the prediction of the $MTBF$. Note that for the ENCV algorithm model no feature is activated.

Table 3 c) shows the obtained MSE by the different algorithms. Compared to the reference model 3 a) the errors increased. This illustrates that the choice of features is very significant.

*Choice of Feature Mapping* There is in principle no restriction for the mapping of the reliability-indicators to input features. Here, the results based on a modified mapping will be shown. It is similar to the reference model, except that second order interactions of the numeric variables are accounted for. This means that

$$\Phi(\mathbf{x_{num}}) = \left[\mathbf{x_{num}}, \log(\mathbf{x_{num}}), [\mathbf{x_{num}}, \log(\mathbf{x_{num}})] \cdot [\mathbf{x_{num}}, \log(\mathbf{x_{num}})]^T\right]^T . \quad (8)$$

One would expect that based on these more complex features a more accurate model can be learned which comes at the cost of the interpretability of the individual feature weights.

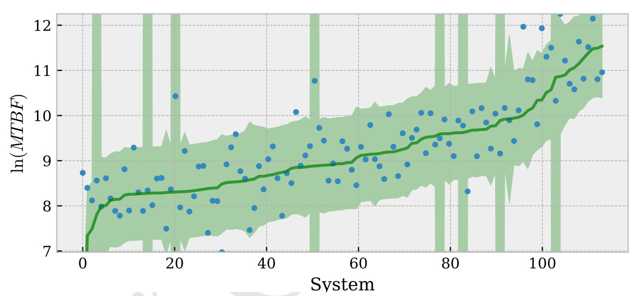


**Fig. 8.** Prediction of the $\log(MTBF)$ with the final model for the test data-set using all reliability-indicators except for the quantities of converters per type. The green line depicts the mean of the predictive distribution and the green shaded are the 95% confidence interval. The blue dots mark the actual observed field reliabilities which are almost always contained. Note that the 95% confidence interval significantly grows in comparison with figure 3.

Figure 8 shows the predictions of the model trained by the ARD algorithm for the test data-set. The 95% confidence interval is of a similar size as in figure 3. The addition of non-linear feature functions does not improve the predictive accuracy in this case.

Table 3 d) shows the MSE obtained with the different algorithms. Compared to the reference model errors in table 3 a) the errors are of the same order. One can conclude that for this kind of problem the addition of non-linear features is not necessary.

**Table 3.** The mean squared error between the (mean of) the predicted values and the true observed field values of the test data-set for different modifications of the reference model - a) MSE for the reference model, b) MSE for a reduced set of systems, c) MSE for a reduced set of reliability-indicators, and d) MSE for non-linear numeric features.

| | ARD | BAR | GP | ENCV | SVR |
|---|---|---|---|---|---|
| **a)** | **0.4007** | 0.4332 | 0.4644 | 0.753 | 0.4262 |
| **b)** | **0.4654** | 0.6175 | 0.6174 | 0.7877 | 9.233 |
| **c)** | 1.3291 | 1.1594 | 1.1517 | **1.1490** | 24.041 |
| **d)** | 0.6703 | **0.4394** | 0.4447 | 0.6333 | 14.073 |

In the last section it was demonstrated that an algorithm fulfilling the initially stated requirements (equations 6 and 7) supports the user to find an accurate and reliable model for the prediction of field-reliablities based on easy-to-collect data.

**Prediction** In the preceding section a validated statistical life-cycle model was learned from field-data. Using this model, the expected field-reliability for future power converters can be predicted at a conceptual design stage. The power converters will be used for a hypothetical new accelerator which will be build and operated at CERN. This is relevant, since we learn a model based on power converters whose life cycle is similar in terms of technical, organizational and human factors [17].
Using the data-set which was collected one can learn a predictive model based on reliability indicators which are available at early design stages by using e.g the ARD or BAR algorithm. The learned model can be used to predict the mean and the 95% confidence interval of the expected field-reliability metrics. Since it is already known that the quantity of converters per type is highly relevant, recommendations[18] for the overall powering strategy for a new accelerator can already be given at the conceptual design stage.

**Discussion** One of the major insights created by applying the methods to the use-case is that the field-reliability is strongly dependent on the quantity of converters per converter type. It would be interesting to study this dependence for other use cases. The explanation for such a characteristic could be that throughout the life cycle reliability enhancing methods and processes are used for devices which are produced in higher quantities. E.g. the production of larger quantities can be optimized and leads to a decreased variability in quality and a higher field-reliability.
The method would be capable to learn a statistical model for the whole life-cycle

[17] If the new converters would be operated at another institute, the model might not be valid.

[18] E.g. the recommendation to reduce the number of power converter types and, hence, increasing the number of converters per type.

of systems. However, to do so more reliability-indicators should be gathered than was done in this work. this can be considered for future work. However, the purpose of this limited set of reliability-indicators was to illustrate that even with very high-level data a good model can be trained. It has to be pointed out that the approach is empirical and that causal relationships have to be identified or confirmed by further studies or experts.

## 5 Conclusion and Outlook

An approach was presented to predict the field-reliability of complex electronic systems at an early development stage based on learning a statistical life-cycle model from similar operational systems. It was demonstrated that the field-reliability can be predicted accurately based on very few reliability-indicators. Compared to existing methods this implies a reduced data collection effort and an integrated quantification of predictive uncertainty based on the granularity of the available information and the implicit randomness of the investigated process. The results of such a study uncover reliability relevant factors which lead to improved system designs at very early stages of design.

Sparse Bayesian Regression methods are the key to efficiently learn an accurate model in an iterative process. Bayesian methods provide feedback on the selection of data-items and their reliability-indicators. The confidence in field-reliability predictions is automatically quantified with respect to the available data and the randomness inherent in the problem.

Future research can focus on more detailed data-sets in terms of reliability-indicators. Based on that, further relevant processes for the field-reliability of systems may be uncovered.

## References

1. Barnard, R.: What is wrong with reliability engineering? In: INCOSE International Symposium. vol. 18, pp. 357–365. Wiley Online Library (2008)
2. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)
3. Blondel, M., Brucher, M., Buitinck, L., Cournapeau, D., Dawe, N., Dubourg, V., Duchesnay, E., la Tour, T.D., ander Fabisch, A., Fritsch, V., Ghosh, S., Gollonet, A.S., Gorgolewski, C.F., andre Gramfort, A., Grisel, O., Grobler, J., Halchenko, Y., Holt, B., Joly, A., Jones, T.R., Kastner, K., Kumar, M., Layton, R., Li, W., Losi, P., Louppe, G., Metzen, J.H., Michel, V., Millman, J., andreas Müller, Niculae, V., Nothman, J., andre Passos, A., Pedregosa, F., Prettenhofer, P., Bertr, Thirion, anderPlas, J.V., Varoquaux, N., Varoquaux, G., Weiss, R.: Scikit-learn user guide (2018), http://scikit-learn.org/stable/user_guide.html#
4. Denson, W.: The history of reliability prediction. IEEE Transactions on reliability **47**(3), SP321–SP328 (1998)
5. Elerath, J.G., Pecht, M.: Ieee 1413: A standard for reliability predictions. IEEE Transactions on Reliability **61**(1), 125–129 (2012)

6. Foucher, B., Boullie, J., Meslet, B., Das, D.: A review of reliability prediction methods for electronic devices. Microelectronics reliability **42**(8), 1155–1162 (2002)
7. Gauch, H.G.: Scientific method in practice. Cambridge University Press (2003)
8. Gullo, L.: In-service reliability assessment and top-down approach provides alternative reliability prediction method. In: Reliability and Maintainability Symposium, 1999. Proceedings. Annual. pp. 365–377. IEEE (1999)
9. Johnson, B.G., Gullo, L.: Improvements in reliability assessment and prediction methodology. In: Reliability and Maintainability Symposium, 2000. Proceedings. Annual. pp. 181–187. IEEE (2000)
10. Jones, J., Hayes, J.: A comparison of electronic-reliability prediction models. IEEE Transactions on reliability **48**(2), 127–134 (1999)
11. Leonard, C.T., Pecht, M.: How failure prediction methodology affects electronic equipment design. Quality and Reliability Engineering International **6**(4), 243–249 (1990)
12. MacKay, D.J.: Bayesian interpolation. Neural computation **4**(3), 415–447 (1992)
13. Miller, R., Green, J., Herrmann, D., Heer, D.: Assess your program for probability of success using the reliability scorecard tool. In: Reliability and Maintainability, 2004 Annual Symposium-RAMS. pp. 641–646. IEEE (2004)
14. O'Connor, P., Kleyner, A.: Practical reliability engineering. John Wiley & Sons (2012)
15. PANDIAN, G.P., Diganta, D., Chuan, L., Enrico, Z., PECHT, M.: A critique of reliability prediction techniques for avionics applications. Chinese Journal of Aeronautics (2017)
16. Pecht, M.G., Das, D., Ramakrishnan, A.: The ieee standards on reliability program and reliability prediction methods for electronic equipment. Microelectronics Reliability **42**(9-11), 1259–1266 (2002)
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
18. Williams, C.K., Rasmussen, C.E.: Gaussian processes for machine learning. the MIT Press **2**(3), 4 (2006)
19. Zio, E.: Reliability engineering: Old problems and new challenges. Reliability Engineering & System Safety **94**(2), 125–141 (2009)