

SYSTEM AND METHOD FOR PERFORMING STATISTICAL FAILURE MODELLING

TECHNICAL FIELD

[0001] Various aspects of this disclosure relate to systems and methods for performing statistical failure modelling.

BACKGROUND

[0002] Statistical failure analysis has been widely adopted in many industries, including power, aviation, manufacturing, public infrastructure, biomedical and public health, oil and gas, etc. Through statistical failure analysis, engineers can gain useful information on the failure properties of the equipment of interest, and hence make appropriate maintenance and replacement strategies for asset management.

[0003] Typically, to conduct a statistical failure analysis, statistical failure modelling is performed, i.e. a statistical model (e.g., a Weibull distribution) is fitted to information from life-time data of the equipment of interest. However, other factors are typically not considered such as the life-time of the equipment, which may (in many situations) incur different phases due to various external factors, such as operating conditions. Some failures may occur for different reasons other than equipment degradation (e.g., early and random failures) and degradation speed may change over the life-time of the equipment (i.e. there may be different phases of degradation in the life-time of equipment).

[0004] Conventional statistical modelling such as those based on two-parameters or three-parameters Weibull distributions, or other variants generated purely based on the life-time data do not capture phase information (i.e. information about different phases of the life-time of the equipment) in the modelling process (i.e. design and fitting of one or more respective distributions) and hence, may lead to biased or even false judgement in the failure prediction.

[0005] Accordingly, efficient approaches for performing statistical failure analysis which take multiple life-time phases into consideration are desirable.

SUMMARY

[0006] Various embodiments concern a method for performing statistical failure modelling, including; generating, for each time of a sequence of times, a respective data point from failure data of a group of devices, generating, for each of a plurality of candidate phase boundary times among the times of the sequence of times, a first fitted function by fitting a first instance of a parameterized function to the data points of times before the candidate phase boundary time and a second fitted function by fitting a second instance of the parameterized function to the data points of times after the candidate phase boundary time, determining, for each candidate phase boundary time, a difference between the first fitted function and the second fitted function, and modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold.

[0007] According to one embodiment, the method includes comparing, for each candidate phase boundary time, the determined difference with the predetermined threshold.

[0008] According to one embodiment, the method includes determining a ranking of at least a subset of the candidate phase boundary times according to the differences determined for the candidate phase boundary times, wherein a candidate phase boundary time is ranked higher than another candidate phase boundary time if its determined difference is higher than that of the other candidate phase boundary time.

[0009] According to one embodiment, the method includes determining the ranking of those candidate phase boundary times for which the determined difference is above the predetermined threshold.

[0010] According to one embodiment, the method includes selecting, if there are multiple ones of the candidate phase boundary times for which the determined difference is above the predetermined threshold, one of the candidate phase boundary times for which the determined difference is above the predetermined threshold taking into account the determined ranking and modelling failure behaviour of the group of devices depending on the selected candidate phase boundary time.

[0011] According to one embodiment, the method includes selecting the candidate phase boundary time taking into account expert knowledge.

[0012] According to one embodiment, the parameterized function is a linear function being parameterized by slope and offset.

[0013] According to one embodiment, the difference is the difference between the slopes of the first fitted function and the second fitted function.

[0014] According to one embodiment, for each of the times, the data point is a probability of failure of a device up to the time.

[0015] According to one embodiment, the method includes receiving failure data about the plurality of devices and estimating the probabilities of failure from the failure data.

[0016] According to one embodiment, the method includes estimating the probabilities of failure using a Kaplan-Meier estimator.

[0017] According to one embodiment, the method includes modelling failure behaviour of the group of devices depending on, if it exists, a candidate phase boundary time for which the determined difference is above a predetermined threshold.

[0018] According to one embodiment, modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold, using a phased bi-Weibull distribution to model failure behaviour.

[0019] According to one embodiment, a time period until the candidate phase boundary time is used as first phase and a time period after the candidate phase boundary time is used as second phase.

[0020] According to one embodiment, modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold, checking whether a phase change indicated by the candidate phase boundary is due to early failures and modelling failure behaviour of the group of devices depending on whether the phase change is due to early failures and/or checking whether the phase change is due to a failure problem of a batch of the group of devices and modelling failure behaviour of the group of devices depending on whether the phase change is due to a failure problem of a batch of the group of devices.

[0021] According to one embodiment, checking whether a phase change indicated by the candidate phase boundary is due to early failures comprises checking whether the candidate phase boundary is smaller than a predetermined threshold time.

[0022] According to one embodiment, checking whether a phase change indicated by the candidate phase boundary is due to early failures is performed using expert knowledge.

[0023] According to one embodiment, checking whether a phase change indicated by the candidate phase boundary is due to a failure problem of a batch of the group of devices is performed using expert knowledge.

[0024] According to one embodiment, modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold and if the phase change is due to early failures, discarding failure data related to the early failures from the failure data and modelling failure behaviour of the group of devices using the remaining failure data.

[0025] According to one embodiment, modelling failure behaviour of the group of devices using the remaining failure data with a Weibull distribution.

[0026] According to one embodiment, modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold and if the phase change is due to a failure problem of a batch of the group of devices, discarding failure data related to the batch of the group of devices from the failure data and modelling failure behaviour of the group of devices using the remaining failure data.

[0027] According to one embodiment, the method includes using the a (standard) Weibull distribution (with two parameters but according to one embodiment also understood to include a Weibull distribution with three parameters) to model failure behaviour for a time period covering the whole sequence of times in response to the determined difference not being above the predetermined threshold for all of the candidate phase boundary times.

[0028] According to one embodiment, a method for controlling one or more devices belonging to a group of devices is provided including modelling failure behaviour for the group of devices of any one of the embodiments described above and controlling the one or more devices according to the modelled failure behaviour.

[0029] According to one embodiment, a data processing system is provided including a communication interface, a memory and a processing unit configured to perform the method of any one of the embodiments described above.

[0030] According to one embodiment, a computer program element is provided including program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the embodiments described above.

[0031] According to one embodiment, a computer-readable medium is provided including program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of the embodiments described above.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

- FIG. 1 illustrates statistical failure modelling.
- FIG. 2 shows a diagram with the cumulative distribution function of the two-parameters Weibull distribution for three different combinations of its parameters.
- FIG. 3 shows an example of a Weibull probability plot.
- FIG. 4 shows a Weibull Probability Plot, including data relating to failures in an early failure stage and data relating to failures in a wear-out stage.
- FIG. 5 shows a Weibull Probability Plot including data relating to failures in a first phase with slower aging and data relating to failures in a second phase with faster aging.
- FIG. 6 shows a flow diagram illustrating a data phasing process according to an embodiment.
- FIG. 7 and FIG. 8 show an example of a Weibull Probability Plot for a first candidate phase boundary and a second candidate phase boundary.
- FIG. 9 illustrates an example of the procedure of FIG. 6.
- FIG. 10 shows a diagram illustrating the cumulative distribution function of the phased bi-Weibull mode for the example of FIG. 9.

- FIG. 11 shows a flow diagram illustrating a data phasing process according to another embodiment.
- FIG. 12 shows a flow diagram illustrating a method for performing statistical failure modelling.
- FIG. 13 shows a data processing system according to an embodiment.

DETAILED DESCRIPTION

[0033] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the disclosure may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the disclosure. Other embodiments may be utilized and structural, and logical changes may be made without departing from the scope of the disclosure. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

[0034] Embodiments described in the context of one of the devices or methods are analogously valid for the other devices or methods. Similarly, embodiments described in the context of a device are analogously valid for a vehicle or a method, and vice-versa.

[0035] Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

[0036] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0037] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0038] In the following, embodiments will be described in detail.

[0039] FIG. 1 illustrates statistical failure modelling.

[0040] Data 102 which contains information about the time of failures of devices of equipment 101 is collected (e.g. a fleet of vehicles, a set of machines etc.). The data 102 may include failure data for devices that have failed and for example indicate a time at which devices (e.g. vehicle components) have failed as well as operation data for devices that are operating (i.e. indicate which devices are still operating). The data 102 (including both failure data as well as operation data) is also referred to as life-time data. A statistical model 103 is generated from the life-time data 102 and may for example model a failure probability per device depending on the life-time (e.g. time since start of operation) of the device.

[0041] The statistical model 103 may include a set of probability distributions P_{θ} , $\theta \in \Theta$, which depend on a set of observations $S = \{t_1, t_2, \dots\}$ (which are reflected in the data 102).

[0042] An example of a probability distribution which may be fit to information contained in the failure data (included in the life-time data 102) as part of the statistical model 103 is a two-parameter Weibull distribution which has the probability density function (pdf)

$$f(t; \beta, \eta) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

and the cumulative distribution function (cdf)

$$F(t; \beta, \eta) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

which may transformed into the linear form

$$\ln(-\ln(1 - F(t))) = \beta \ln(t) - \beta \ln(\eta)$$

wherein the left hand side is also referred to as y and $\ln(t)$ is also referred to as x (see in particular the Weibull Probability Plots described below where x corresponds to the horizontal axis and y to the vertical axis).

[0043] A controller 104 (e.g. a data processing device such as a server computer controlling the equipment, in particular in response to user commands) and/or human operator (e.g. engineer) may then use the statistical model 103 for controlling the equipment group, e.g. deactivate a device (which has a high risk of failure, e.g. above a tolerable threshold), i.e. take the device out of service, perform maintenance of the device, replace the device (e.g. a machine component), change operating parameters of the device (to reduce risk of failure or impact of failure), supplement backup devices etc.

[0044] FIG. 2 shows a diagram 200 with the cumulative distribution function of the two-parameters Weibull distribution for three different combinations of the (two) parameters β and η . The variable t for example refers to the time, e.g. a time since starting operation of a device of the equipment group 101 and $F(t)$ for example is the probability that the device has failed since the beginning of its operation until time t .

[0045] FIG. 3 shows an example of a Weibull probability plot 300.

[0046] A Weibull probability plot is based on the above linear form of the cdf of the Weibull distribution and is a plot that determines if the data fits the Weibull distribution: if the data (indicated as dots) fit the Weibull distribution, the points should lie on a straight line (i.e. should fit a straight line according to the above linear form of the cdf of the Weibull distribution). The points deviating from the straight line indicates that the data do not fit the Weibull distribution.

[0047] Here, each data point (also referred to as failure data point) is associated with a time t (wherein $\ln(t)$ is the position on the horizontal axis of the plot where it is indicated) and specifies a probability of failure until that time, i.e. the complementary probability to the “survival probability” which may be estimated by Kaplan-Meier (KM).

[0048] A Weibull distribution allows efficient modelling of the aging of a device. However, in practice, failures may occur for different reasons, such as normal degradation, early failures, random failures, etc. In particular, failures may occur due to different levels of degradation (i.e. different aging rates) for different life-time phases of the respective device or component. Accordingly, it may not be possible to fit a Weibull distribution such that the observed data fits a line in a Weibull Probability Plot. Examples for this are illustrated in FIG. 4 and FIG 5.

[0049] FIG. 4 shows a Weibull Probability Plot 400 including data relating to failures in an early failure stage (or phase) 401 and data relating to failures in a wear-out stage (or phase) 402.

[0050] FIG. 5 shows a Weibull Probability Plot 500 including data relating to failures in a first phase 501 with slower aging and data relating to failures in a second phase 502 with faster aging.

[0051] A modelling using a probability distribution like a 2-parameters or 3-parameters Weibull distribution or other variants based on the life-time data does not capture the phase information. So, there is no phase information available in classical statistical failure analysis

using such an approach. This may result in the statistical model 103 having biased or even false judgement in failure prediction.

[0052] Therefore, as described in the following in more detail, according to various embodiments, a comprehensive process, also referred to as data phasing process, is provided for identifying if there is an indicative timing where the failure pattern changes (i.e. whether there is a phase change, i.e. degradation behaviour changes like in the examples of FIG. 4 and FIG. 5). The process includes procedures to choose a suitable statistical model 103 to handle data 102 with and without phase change. This allows obtaining a statistical model 103 that can lead to a better accuracy in the statistical failure analysis.

[0053] FIG. 6 shows a flow diagram 600 illustrating a data phasing process according to an embodiment.

[0054] The data phasing process gets the life-time data 102 (including failure data and operation data) as input and includes two parts: in a first part 610, an automated phasing algorithm 601 identifies where there is a phase difference throughout the life-time of the respective device and also suggests the possible phase boundary choices (if any).

[0055] The automated phasing algorithm 601 fits a respective model (referred to as partial model in the following) for each phase. In the present embodiment, each partial model is a (2-parameter) Weibull distribution.

[0056] Table 1 gives an example of the automated phasing algorithm 601 which is described in more detail in the following.

Parameters:	
$A = \text{Minimum Data Requirement } (2A \leq N + 1)$	
$B = \text{Number of Suggested Results } (B \leq N - 2A + 2)$	
$\delta = \text{Slope Difference Threshold}$	
Operations:	
	Represent the input failure data according to a Weibull probability plot (WPP) where the probability of failure is estimated by Kaplan-Meier (KM). Let this include data points numbered 1, 2, 3, ..., N
	for $k = A, A + 1, \dots, N - A + 1$, do
	Fit straight lines 1 & 2 using data points 1, 2, ..., k and $k, k + 1, \dots, N$
	Record their slopes $s_1(k)$ and $s_2(k)$ and compute slope difference $\Delta s(k) = s_2(k) - s_1(k)$ at k
	end for

	For those k such that $ \Delta s(k) > \delta$, record the B largest slope difference results: $\Delta s(k_1), \dots, \Delta s(k_B)$ (and if such k exist, set flag= TRUE); otherwise (if no such k exist), report “No significant phase difference could be found” (and set flag= FALSE)
Output:	
if flag= TRUE : suggested phase boundary: $k_1, \dots, k_B$ with corresponding slope differences $\Delta s(k_1), \dots, \Delta s(k_B)$	
if flag= FALSE : indication that no significant phase difference could be found	

Table 1

[0057] As indicated in Table 1, the automated phasing algorithm 601 takes three parameters: minimum data requirement A , number of suggested results B , and slope difference threshold δ . The minimum data requirement A sets the minimum data required to build a partial model. The number of suggested results B sets how many number of phase boundary choices should be output at most. The slope difference threshold δ is a parameter used to identify if there are differences between two partial models.

[0058] First, the input failure data (included in the life-time data 102) is represented according to a Weibull probability plot (WPP), where the corresponding data points are probability of failure estimated by the Kaplan-Meier (KM) estimator, for example. This may or may not include explicitly plotting the Weibull probability plot.

[0059] FIG. 7 and FIG. 8 show an example of a WPP 700, 800 (for a case with two phases) for a first candidate phase boundary 701 (FIG. 7) and a second candidate phase boundary 801 (FIG. 8). Let there be N dots in the WPP 700 representing N distinctive (failure) data points.

[0060] The following steps (i) and (ii) are repeated for $k = A, A + 1, \dots, N - A + 1$:

(i) let k denote the data point index of a potential phase boundary. Fit (as illustrated in FIG. 7) two straight lines using the data points $1, 2, \dots, k$, and $k, k+1, \dots, N$, respectively. In other words, a linear regression is performed (using e.g. minimum squared error). The lines are examples of two instances of a parameterized function (here a linear function with the parameters slope and offset).

(ii) record the slopes of the two lines (denoted by $s_1(k)$ and $s_2(k)$), and compute their slope difference ($\Delta s(k) = s_2(k) - s_1(k)$).

[0061] Comparing these slope differences with the slope difference threshold δ , phase boundaries are only taken from those k for which the corresponding slope difference has a

larger absolute value than the slope difference threshold (i.e., $|\Delta s(k)| > \delta$). From those, the indices with the B largest slope differences (e.g. in descending order): $\Delta s(k_1), \Delta s(k_2), \dots, \Delta s(k_B)$ (if any) are taken.

[0062] According to the result of the automated phasing algorithm 601, the first part of the data phasing process outputs suggested phase boundary choices: $k_1, k_2, \dots, k_B$ with corresponding slope differences: $\Delta s(k_1), \Delta s(k_2), \dots, \Delta s(k_B)$ in 604 if any have been found by the automated phasing algorithm 601 and otherwise outputs the conclusion that “no significant phase difference could be found” in 602.

[0063] If no significant phase could be found (indicated by the output of 602) the second part 611 of the process performs the standard Weibull modelling procedure in 603, i.e. fits a single Weibull probability distribution (e.g. according to the pdf f and cdf F given above) to the data 102 using maximum likelihood estimation.

[0064] If $\{t_i\}_{i=1}^{n_f}$ denote the failure times (i.e., for each i , a time at which a respective device has failed) and $\{t_j\}_{j=1}^{n_s}$ the life-times of devices that are still operating (i.e., for each j , the life-time of a device (e.g. from the start of its operation) that is still operating), the likelihood function of the data, denoted by $L(\beta, \eta | \text{data})$ can be obtained by

$$L(\beta, \eta | \text{data}) = \prod_{i=1}^{n_f} f(t_i; \beta, \eta) \times \prod_{j=1}^{n_s} (1 - F(t_j; \beta, \eta))$$

wherein the first product represents the devices which have failed and the second product represents the devices which have not yet failed.

[0065] The estimates $\hat{\beta}, \hat{\eta}$ of the shape and scale parameters β, η can be obtained by solving the following optimization problem:

$$\hat{\beta}, \hat{\eta} \in \arg \max_{\beta, \eta > 0} \log L(\beta, \eta | \text{data}),$$

(e.g. by setting the derivatives of the log term with respect to β and η to zero) which completes the standard Weibull modelling procedure of 603.

[0066] If the automated phasing algorithm 601 suggests phase boundary choices (output of 604), this implies that the respective class of devices could have experienced phase change at the time suggested by the algorithm with ranks (e.g. according to the descending order of slope differences) based on the statistics of the data. These choices are evaluated in 605 (e.g.

according to engineering judgment and expert knowledge) with regard to the class of devices under analysis. With the consideration of statistical analysis and/or expert knowledge, a phase boundary (denoted by t_0) is confirmed (in a simple case, the one with the highest slope difference may be used) and is used in 606 to fit a more complicated probability distribution to the data 102 than the two-parameter Weibull distribution described above, namely a “phased bi-Weibull distribution”, whose probability density function (pdf) and cumulative distribution function (cdf) are respectively defined as

$$f(t; \beta_1, \eta_1, \beta_2, \eta_2) = \begin{cases} \frac{\beta_1}{\eta_1} \left(\frac{t}{\eta_1}\right)^{\beta_1-1} e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} & t < t_0, \\ \left[\frac{\beta_1}{\eta_1} \left(\frac{t}{\eta_1}\right)^{\beta_1-1} + \frac{\beta_2}{\eta_2} \left(\frac{t-t_0}{\eta_2}\right)^{\beta_2-1}\right] e^{-\left[\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t-t_0}{\eta_2}\right)^{\beta_2}\right]} & t \geq t_0. \end{cases}$$

$$F(t; \beta_1, \eta_1, \beta_2, \eta_2) = \begin{cases} 1 - e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} & t < t_0, \\ 1 - e^{-\left[\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{t-t_0}{\eta_2}\right)^{\beta_2}\right]} & t \geq t_0. \end{cases}$$

[0067] As can be seen, before the phase boundary t_0 , the cdf is the same as the one of the two-parameter Weibull distribution, specified by the first scale parameter η_1 and shape parameter β_1 . After the phase boundary t_0 , there is another term adding to the power of exponential term, introducing the second scale parameter η_2 and shape parameter β_2 .

[0068] These four parameters may be estimated by maximum likelihood estimation (MLE) of the phased bi-Weibull distribution. For convenience of presentation, the pdf and cdf of phased bi-Weibull distribution are respectively denoted by

$$f(t; \beta_1, \eta_1, \beta_2, \eta_2) = \begin{cases} f_1 & t < t_0, \\ f_2 & t \geq t_0. \end{cases}, \quad F(t; \beta_1, \eta_1, \beta_2, \eta_2) = \begin{cases} F_1 & t < t_0, \\ F_2 & t \geq t_0. \end{cases}$$

[0069] Further, the data can be classified into the following types:

- (failure data) let $\{t_i\}_{i=1}^{n_{f_2}}$ be the failure times smaller than t_0 , and $\{t_n\}_{n=1}^{n_{f_2}}$ be the failure times greater than or equal to t_0 (of devices that have failed)
- (operating data) let $\{t_j\}_{j=1}^{n_{s_1}}$ be the life-times smaller than t_0 , and $\{t_m\}_{m=1}^{n_{s_2}}$ be the life-times greater than or equal to t_0 (of devices that are still operating).

[0070] Then, the likelihood function of the data, denoted by $L(\beta_1, \eta_1, \beta_2, \eta_2|\text{data})$, can be obtained by

$$L(\beta_1, \eta_1, \beta_2, \eta_2|\text{data}) = \prod_{i=1}^{n_{f_1}} f_1(t_i; \beta_1, \eta_1) \times \prod_{n=1}^{n_{f_2}} f_2(t_n; \beta_1, \eta_1, \beta_2, \eta_2) \\ \times \prod_{j=1}^{n_{s_1}} (1 - F_1(t_j; \beta_1, \eta_1)) \times \prod_{m=1}^{n_{s_2}} (1 - F_2(t_m; \beta_1, \eta_1, \beta_2, \eta_2))$$

[0071] The estimates of the two shape and two scale parameters $\hat{\beta}_1, \hat{\eta}_1, \hat{\beta}_2, \hat{\eta}_2$ can be obtained by solving the optimization problem

$$\hat{\beta}_1, \hat{\eta}_1, \hat{\beta}_2, \hat{\eta}_2 = \arg \max_{\beta_1, \eta_1, \beta_2, \eta_2 > 0} \log L(\beta_1, \eta_1, \beta_2, \eta_2|\text{data}),$$

which completes the phased bi-Weibull modelling.

[0072] FIG. 9 illustrates an example Weibull Probability plot 900 of the above procedure with

- Minimum data setting: 11 (113/10)
- Phase segregation threshold for slope difference: 3
- The suggested phase boundary choices (age/years): [40 39 38 37 35]
- The corresponding slope differences: [4.01821122 3.85238367 3.62627997 3.3467357 3.10693089]

[0073] The determined values of the parameters of the (standard) 2-parameter Weibull model for the indicated data points are

$$\hat{\eta} = 53.78, \hat{\beta} = 3.59$$

and the parameters for the determined values of the phased bi-Weibull model are

$$k_1 = 40, \hat{\eta}_1 = 78.25, \hat{\beta}_1 = 2.14, \hat{\eta}_2 = 15.12, \hat{\beta}_2 = 1.95.$$

[0074] As can be seen, the phased bi-Weibull provides a better fitting than 2-parameter Weibull when the data incurs a phase change, and implies that the equipment may around incur an accelerated degradation at around age 40.

[0075] FIG. 10 shows a diagram 1000 illustrating the cumulative distribution function of the phased bi-Weibull mode (with the parameter values as given above) for the example of FIG. 9. As can be seen, the change of remaining useful life as a result of phase change can be spotted.

[0076] FIG. 11 shows a flow diagram 1100 illustrating a data phasing process according to another embodiment.

[0077] FIG. 11 includes operations 1101 to 1106 (of two parts of the data phasing 1110 and 1111) as described with reference to FIG. 6 except for that there are additional checks 1107 and 1108 which, if they are positive, lead to usage of standard Weibull modelling of 1103 with a first modification 1109 of the input data (for the standard Weibull modelling 1103) or to a restart of the data phasing process (i.e. return to 1101) with a second modification 1112 of the input data (for the automated phasing algorithm of 1101), respectively.

[0078] Specifically, when with consideration of statistical analysis and/or expert knowledge, a phase boundary (denoted by t_0) has been confirmed in 1105 an evaluation is made in 1107 to decide whether the phase change (represented by the phase boundary t_0) is due to early failures: if the identified phase boundary t_0 is smaller than a pre-defined value (threshold) t_m , it is decided that the phase change is due to early failure. As a result, data points that correspond to failure data that is less than t_0 years old are disregarded in 1109 and the standard 2-parameter Weibull modelling is applied in 1103 to the remaining data points (i.e. to the data set modified in that manner). This is motivated by the fact that early failures which show a different pattern are likely due to human errors or defects in products and do not belong to the scope of degradation relevant modelling and hence, only the second phase (after the detected phase boundary t_0) is of modelling interest.

[0079] If the identified phase boundary t_0 is greater than or equal to the predetermined value t_m , investigations on the phase change are conducted by engineering judgement and expert knowledge in 1108 to decide whether the phase change (represented by the phase boundary t_0) is due to a batch problem (i.e., failures from the same batch of devices among the equipment 101 which for example comprises devices of multiple batches). If it is decided that the phase change is due to a batch problem, life-time data (both failure data and operating data) of this batch is discarded from the life-time data in 1107, and the entire data phasing process is re-run for the life-time data modified in this manner to check if there is still a phase change and perform, depending on whether there is still a phase change, a modelling of 1103 or 1106. If both decisions 1107 and 1108 are negative, i.e. it is decided that the identified phase change neither due to early failure nor due to a batch problem, the

identified phase boundary t_0 is used in 1106 to fit a phased bi-Weibull distribution to the life-time data as in 606.

[0080] In summary, according to various embodiments, a method is provided as illustrated in FIG. 12.

[0081] FIG. 12 shows a flow diagram 1200 illustrating a method for performing statistical failure modelling.

[0082] In 1201, for each time of a sequence of times, a respective data point from failure data of a group of devices is generated.

[0083] In 1202, for each of a plurality of candidate phase boundary times among the times of the sequence of times, a first fitted function by fitting a first instance of a parameterized function to the data points of times before the candidate phase boundary time and a second fitted function by fitting a second instance of the parameterized function to the data points of times after the candidate phase boundary time, are generated. A candidate phase boundary times or simply candidate boundary time is a possible time separating two different phases of failure (e.g. degradation) behaviour.

[0084] In 1203, for each candidate phase boundary time, a difference between the first fitted function and the second fitted function is determined.

[0085] In 1204 failure behaviour of the group of devices is modelled depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold.

[0086] According to various embodiments, in other words, a method for performing statistical failure analysis (by means of statistical failure modelling) and in particular a data analytic process in the course of statistical failure modelling, e.g. for a fleet of is provided which allows treating failures with the consideration of different phases. According to various embodiments, a process for identifying the indicative timing where the failure behaviour changes (i.e., a phase change occurs) is performed a suitable statistical model to handle the data with different phases is built. Thus, a systematic approach for identifying whether there are any significant phase differences throughout the life-time of the equipment of interest is provided. The process may include making suggestions on how to segregate failure data (or information derived from it like the data points described above which describe failure probabilities up to respective times) into different phases and develop a suitable statistical model to handle the failure data (or the information (e.g. data points))

derived from it) with different phases. So, a systematic approach for the data phasing is provided that can improve the statistical modelling process, and hence lead to a better accuracy in the statistical failure analysis.

[0087] According to various embodiments, for example, a process for data phasing, i.e. to identify and suggest how to segregate the data into different phases, is provided, including a statistical modelling process to handle the data with different phases if there exists a phase difference. As described with reference to FIG. 6, this process may include an "automated phasing" algorithm to identify if there is a phase difference, and suggest the possible phase boundary choices if any, as well as a statistical modelling process to handle the data if a phase boundary is identified and suggested by the automated phasing algorithm. The algorithm may include obtaining a Weibull probability plot (WPP) for the data whose probability of failure is for example estimated by the Kaplan-Meier (KM) estimator, fitting two straight lines to the segregated data points by least-square estimation (LSE); obtaining candidates of a phase boundary between two phases (of different degradation behaviours) based on the comparison of the obtained slope differences of the two fitted lines with a slope difference threshold parameter; and obtaining the phase boundary choices from the candidates based on their absolute values in a descending order.

[0088] The process may further include fitting a standard Weibull model (i.e. distribution) if there is no phase difference (i.e. there is a single phase, i.e. a single degradation behaviour which may be fit to a Weibull distribution) or the phase change is due to early failures which can be discarded, fitting a phased bi-Weibull model (i.e. distribution) to handle the data with two phases if there is a phase difference which is not due to early failures (and, according to one embodiment, not due to a batch problem (i.e. a failure problem of a batch)), and obtaining the model (i.e. distribution) parameters based on maximum likelihood estimation (MLE).

[0089] The method of FIG. 12 is for example carried out by a data processing system (e.g. a computer or multiple computers) as illustrated in FIG. 13.

[0090] FIG. 13 shows a data processing system 1300 according to an embodiment.

[0091] The data processing system 1300 includes a communication interface 1301 (e.g. configured to receive failure data, e.g. via user input and/or from sensors). The server computer 1300 further includes a processing unit 1302 and a memory 1303. The memory 1303 may be used by the processing unit 1302 to store, for example, data to be processed,

such as the failure data and the derived data points. The data processing system is configured to perform the method of FIG. 12.

[0092] The methods described herein may be performed and the various processing or computation units and the devices and computing entities described herein may be implemented by one or more circuits. In an embodiment, a "circuit" may be understood as any kind of a logic implementing entity, which may be hardware, software, firmware, or any combination thereof. Thus, in an embodiment, a "circuit" may be a hard-wired logic circuit or a programmable logic circuit such as a programmable processor, e.g. a microprocessor. A "circuit" may also be software being implemented or executed by a processor, e.g. any kind of computer program, e.g. a computer program using a virtual machine code. Any other kind of implementation of the respective functions which are described herein may also be understood as a "circuit" in accordance with an alternative embodiment.

[0093] While the disclosure has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A method for performing statistical failure modelling, comprising:
generating, for each time of a sequence of times, a respective data point from failure data of a group of devices;
generating, for each of a plurality of candidate phase boundary times among the times of the sequence of times, a first fitted function by fitting a first instance of a parameterized function to the data points of times before the candidate phase boundary time and a second fitted function by fitting a second instance of the parameterized function to the data points of times after the candidate phase boundary time;
determining, for each candidate phase boundary time, a difference between the first fitted function and the second fitted function, and
modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold.
2. The method of claim 1, comprising comparing, for each candidate phase boundary time, the determined difference with the predetermined threshold.
3. The method of claim 1 or 2, comprising determining a ranking of at least a subset of the candidate phase boundary times according to the differences determined for the candidate phase boundary times, wherein a candidate phase boundary time is ranked higher than another candidate phase boundary time if its determined difference is higher than that of the other candidate phase boundary time.
4. The method of claim 3, comprising determining the ranking of those candidate phase boundary times for which the determined difference is above the predetermined threshold.
5. The method of claim 4, comprising selecting, if there are multiple ones of the candidate phase boundary times for which the determined difference is above the

predetermined threshold, one of the candidate phase boundary times for which the determined difference is above the predetermined threshold taking into account the determined ranking and modelling failure behaviour of the group of devices depending on the selected candidate phase boundary time.

6. The method claim 5, comprising selecting the candidate phase boundary time taking into account expert knowledge.
7. The method of any one of claims 1 to 6, wherein the parameterized function is a linear function being parameterized by slope and offset.
8. The method of claim 7, wherein the difference is the difference between the slopes of the first fitted function and the second fitted function.
9. The method of any one of claims 1 to 8, wherein, for each of the times, the data point is a probability of failure of a device up to the time.
10. The method of claim 9, comprising receiving failure data about the plurality of devices and estimating the probabilities of failure from the failure data.
11. The method of claim 10, comprising estimating the probabilities of failure using a Kaplan-Meier estimator.
12. The method of any one of claims 1 to 11, comprising modelling failure behaviour of the group of devices depending on, if it exists, a candidate phase boundary time for which the determined difference is above a predetermined threshold.
13. The method of any one of claims 1 to 12, wherein modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above

a predetermined threshold, using a phased bi-Weibull distribution to model failure behaviour.

14. The method of claim 13, wherein a time period until the candidate phase boundary time is used as first phase and a time period after the candidate phase boundary time is used as second phase.
15. The method of any one of claims 1 to 14, wherein modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold, checking whether a phase change indicated by the candidate phase boundary is due to early failures and modelling failure behaviour of the group of devices depending on whether the phase change is due to early failures and/or checking whether the phase change is due to a failure problem of a batch of the group of devices and modelling failure behaviour of the group of devices depending on whether the phase change is due to a failure problem of a batch of the group of devices.
16. The method of claim 15, wherein checking whether a phase change indicated by the candidate phase boundary is due to early failures comprises checking whether the candidate phase boundary is smaller than a predetermined threshold time.
17. The method of claim 15 or 16, wherein checking whether a phase change indicated by the candidate phase boundary is due to early failures is performed using expert knowledge.
18. The method of any one of claims 15 to 17, wherein checking whether a phase change indicated by the candidate phase boundary is due to a failure problem of a batch of the group of devices is performed using expert knowledge.

19. The method of any one of claims 1 to 18, wherein modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold and if the phase change is due to early failures, discarding failure data related to the early failures from the failure data and modelling failure behaviour of the group of devices using the remaining failure data.
20. The method of claim 19, comprising modelling failure behaviour of the group of devices using the remaining failure data with a Weibull distribution.
21. The method of any one of claims 1 to 20, wherein modelling failure behaviour of the group of devices depending on whether there is a candidate phase boundary time for which the determined difference is above a predetermined threshold comprises: if there is a candidate phase boundary time for which the determined difference is above a predetermined threshold and if the phase change is due to a failure problem of a batch of the group of devices, discarding failure data related to the batch of the group of devices from the failure data and modelling failure behaviour of the group of devices using the remaining failure data.
22. The method of any one of claims 1 to 21, comprising using a Weibull distribution to model failure behaviour for a time period covering the whole sequence of times in response to the determined difference not being above the predetermined threshold for all of the candidate phase boundary times.
23. Method for controlling one or more devices belonging to a group of devices comprising modelling failure behaviour for the group of devices according to any one of claims 1 to 22 and controlling the one or more devices according to the modelled failure behaviour.
24. A data processing system comprising a communication interface, a memory and a processing unit configured to perform the method of any one of claims 1 to 23.

25. A computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 23.
26. A computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 23.