

Competing risks for repairable systems: A data study

Helge Langseth and Bo Henry Lindqvist

*Department of Mathematical Sciences, Norwegian University of Science and
Technology, Norway*

Abstract

We analyse a dataset from the Offshore Reliability Data (OREDA) Database, looking for a model, which can be used to unveil aspects of the quality of the maintenance performed. To do so we must investigate the mathematical modelling of maintenance and repair of components that can fail due to a variety of failure mechanisms.

NOTE! UNTIL WE GET A “GO” FROM THE OREDA PROJECT THE DATA IN Table 1 SHOULD BE TREATED AS CONFIDENTIAL. FINAL ACCEPTANCE FROM OREDA IS EXPECTED IN THE NEAR FUTURE.

Key words: Competing risks, repairable systems, data analysis.

1 Introduction

In this paper we employ a model for components which fail due to one of a series of “competing” failure mechanisms, each acting independently upon the system. The components under consideration are repaired upon failure, but are also preventively maintained. The preventive maintenance (PM) is performed due to casual observation of an evolving failure. The maintenance need not be perfect; we use a version of an *imperfect repair* model to allow a flexible yet simple maintenance model. Our motivation for analysing this dataset is to estimate quantities, which describe the “goodness” of the maintenance crew; their ability to prevent failures by performing thorough maintenance at the

Email addresses: helgel@math.ntnu.no (Helge Langseth), bo@math.ntnu.no (Bo Henry Lindqvist).

correct time. Our main focus in this paper is to analyse a dataset from the OREDA database (OREDA 2002), which gives the intermediate failure times, the “winning” failure mechanism associated with each failure (i.e. the failure mechanism leading to the failure), and the maintenance activity.

2 Dataset and model implications

The dataset we want to analyse is presented in Table 1. This dataset, which is from the OREDA database (OREDA 2002), describes a single compressor system on an offshore installation. A compressor system can be broken down into several subsystems (compressor unit, lubrication system, shaft seal system, etc.); this particular dataset gives an account of the *compressor unit*. The compressor unit is again divided into several *maintainable items*. A maintainable item is defined in OREDA as “an item that contribute a part, or an assembly of parts, that is normally the lowest indenture level during maintenance”. The maintainable items making up the compressor unit are, among others, valves, internal piping and radial bearing. From analysing the database we can identify which maintainable items that lead to a particular failure of the compressor unit, but we are not always able to tell which maintainable items that are affected by the corresponding repair. We will therefore follow the OREDA Handbook and make our analysis at the subunit level, neglecting the (partial) information regarding maintainable items. The dataset gives the time of each event, the failure mechanism leading to it, and the failure’s severity.

The failures can roughly be seen as the result of two different failure mechanisms, coded as either “1” or “2” in the *FM* column of Table 1. The last digit of the FM code is used to give further details; 1.0 is a general description of the FM (default value, e.g., “General mechanical failure”) whereas codes 1.1 – 1.4 are specialisations of this failure mechanism. The same coding is employed for failure mechanism “2”. In our analysis we will focus on failure mechanism “2”, where we group together failures from failure mechanisms coded with values from 2.0 to 2.7. Failures due to other failure mechanisms are treated as external events (i.e., random censoring).

Formally, we consider a mechanical component which is set into operation at time $t = 0$. We assume that the component is (as good as) new at that time. At some random times T_1, T_2, \dots the component fails. After failure the component is immediately repaired and put back into service. The data can

Time	FM	Severity									
220	1.0	I	429	2.0	C	651	1.0	D	1109	2.0	D
233	1.0	I	460	?	?	657	1.0	D	1117	2.2	I
234	1.4	I	470	2.0	C	660	2.0	D	1197	2.0	C
240	2.6	D	474	1.0	C	666	1.0	I	1258	1.1	C
265	1.0	I	475	2.0	I	668	1.0	I	1269	2.0	C
270	1.3	I	476	2.5	I	680	2.0	D	1297	2.0	D
273	1.0	I	508	2.2	D	681	1.0	D	1309	1.0	D
279	?	C	522	2.0	I	684	2.0	D	1322	2.0	C
285	1.0	I	523	2.0	I	691	1.0	I	1346	2.0	D
287	?	C	535	2.0	D	693	1.0	D	1349	2.0	D
294	2.3	D	542	1.0	D	705	1.0	C	1359	2.7	D
295	2.0	I	570	2.0	I	717	1.0	C	1363	1.0	C
300	1.0	D	580	1.2	C	834	2.0	C	1448	2.0	C
325	2.0	C	604	2.0	D	837	1.2	C	1476	2.0	D
328	1.0	C	612	2.0	C	841	1.0	C	1481	2.0	D
333	2.0	C	613	2.0	C	843	1.0	C	1557	1.3	C
365	1.0	I	614	2.0	C	845	1.0	C	1606	1.3	I
368	2.3	D	615	2.0	D	875	1.0	C	1610	2.0	D
369	2.4	I	634	1.0	I	972	1.0	C	1642	?	D
381	1.0	I	636	?	C	1037	1.0	C	1659	2.7	D
417	2.0	I	637	1.0	D	1084	1.0	I			
418	2.1	D	638	2.0	I	1091	1.0	D			

Table 1

This dataset describes the service time of a single component over a period of 1659 time units taken from the OREDA data (OREDA 2002). The failures are caused by different failure mechanisms, coded in the *FM* column. The failure mechanisms can roughly be grouped into two groups: One containing codes 1.0 – 1.4; the other containing codes 2.0 – 2.7. Each group of failure mechanisms is coded “hierarchically”: For instance 1.0 is a general description of the failure mechanisms in the first group whereas codes 1.1 – 1.4 are specialisations of this failure mechanism. The same coding is employed for the other group. *Severity* describes the criticality of the failure (“C” denotes critical failures, “D” are degraded failures, “I” are incipient failures, and “?” represent missing data). For confidentiality reasons all failure times are scaled with an undisclosed factor α (that is, if the true failure time is T_i , the reported failure time is $\alpha \cdot T_i$; as this corresponds to a change of scale only, it does not change the validity of the following analysis).

now be given as an ordered sequence of points

$$(Y_i, J_i); i = 1, 2, \dots, n, \tag{1}$$

where each point represents an event. Here

$$\begin{aligned}
Y_i &= \text{inter-event time, i.e. time since previous event} \\
&\quad (\text{time since start of service if } i = 1) \\
J_i &= \begin{cases} 0 & \text{if critical failure} \\ 1 & \text{if degraded failure} \\ 2 & \text{if external event.} \end{cases} \quad (2)
\end{aligned}$$

We have to find a model for data of type (1). The basic ingredient in such a model is the hazard rate $\psi(t)$ at time t , for a component which is new at time $t = 0$. We assume that $\psi(t)$ is a continuous and integrable function on $[0, \infty)$. In practice it will be important to estimate $\psi(\cdot)$ since this information may, e.g., be used to plan future maintenance strategies.

The first thing we observe from our dataset is that when a failure is detected, the maintenance crew will repair the failure, and put the component back in operation. Thus, the dataset must be considered using theory for repairable systems (see e.g., (Ascher and Feingold 1984)). The most frequently used models for repairable systems assume either perfect repair (renewal process models) or minimal repair (nonhomogeneous Poisson-process models). In our case this picture is complicated by the fact that the analysis is performed at the subunit level. Some of the maintainable items building up the compressor unit will be repaired thoroughly after an event, whereas others may be left as they are or given only minor adjustments. Hence, neither perfect nor minimal repair models may be appropriate, and we shall here adopt the imperfect repair model presented by Brown and Proschan (1983). This will introduce one parameter, p , which is the probability of perfect repair for a preventive maintenance. This quantity is of interest since it can be used as indication of the quality of maintenance. The parameters may in practice be compared between plants and companies, and thereby unveil maintenance improvement potential.

The events can be categorised as either *i)* Critical failures, *ii)* Degraded failures *iii)* Incipient failures) or *iv)* External events (component taken out of service or some other random censoring). This information can be read off the *Severity* column in Table 1. In the OREDA database a critical failure is an event that causes immediate and complete loss of a system's capability of providing its output. A degraded failure is defined as a failure that prevents the system from providing its output within specifications, and which may develop into a critical failure in time. An incipient failure is defined as a failure that not immediately causes loss of the system's capability of providing its output, but that could develop to a critical or degraded failure in the near future if not attended to. We will not distinguish between incipient and degraded failures in our analysis, but term both severities "Degraded".

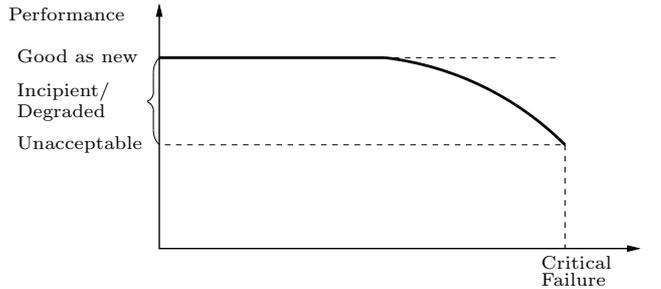


Fig. 1. Component with degrading performance

The development of a failure can be seen as depicted in Fig. 1. We assume that the component is continuously deteriorating when used, so that the performance gradually degrades until it falls outside a preset acceptable margin. As soon as the performance is unacceptable, the component experiences a critical failure. Before the component fails it may exhibit inferior performance, which will be found as a degraded failure in our dataset. This is a “signal” to the maintenance crew that a critical failure is approaching, and that the inferior component may be repaired. In our dataset these “signals” are detected through continuous condition monitoring, observed production interference, or functional testing. When the maintenance crew intervenes and repairs a component before it fails critically, the repair action is called (an unscheduled) preventive maintenance. The maintenance crew will typically strive for avoiding critical failures, because they are considered more costly than the degraded ones.

Our model must take into account the relation between preventive maintenance and critical failures. As it is assumed that the component gives some kind of “signal”, which will alert the maintenance crew, it is not reasonable to model the (potential) times for preventive maintenance and critical failures as stochastically independent. We shall therefore adopt the random signs censoring of Cooke (1996). This will eventually introduce a single new parameter q , with interpretation as the probability that a critical failure is avoided by a preceding unscheduled preventive maintenance.

The dataset we analyse in this paper stems from a single component, but the analysis can easily be extended to a more general by assuming that all components fail independently of each other.

3 Basic ingredients of the model

In this section we describe and discuss the two main building blocks of the model we will use. In Section 3.1 we consider the concept of imperfect repair, as defined by Brown and Proschan (1983). Then in Section 3.2 we introduce our basic model for the relation between preventive and corrective maintenance.

3.1 Imperfect repair

The repairable systems model we propose to use for this dataset is motivated by the imperfect repair model of Brown and Proschan (1983), which we term the *BP model* in the following. We therefore start with a review of this model. As usual, we define T_i to be the consecutive *event times*, and we use Y_i for the *inter-event times* (time between events). For simplicity of notation we assume that the component is observed from time $t = 0$, and with the definition $T_0 = 0$, we have $Y_i = T_i - T_{i-1}$ for $i = 1, 2, \dots, n$ where n is the number of events (see Figure 2). We use $\psi(t)$ for the hazard rate for a component of “age” t , and

$$\lambda(t | \mathcal{F}_{t^-}) = \lim_{\Delta t \downarrow 0} \frac{P(\text{Event in } [t, t + \Delta t) | \mathcal{F}_{t^-})}{\Delta t}$$

for the conditional intensity given \mathcal{F}_{t^-} , the *history* of the counting process up to time t , see (Andersen et al. 1992). Furthermore, $N(t)$ is the number of events in $(0, t]$ and $N(t^-)$ is the number of events in $(0, t)$. We will assume that failures are repaired immediately, and we disregard the time to repair.

This notation enables us to repeat some of the most standard repair models: *Perfect repair* is modelled by $\lambda(t | \mathcal{F}_{t^-}) = \psi(t - T_{N(t^-)})$ where $t - T_{N(t^-)}$ is the time since the last event, i.e., age is measured by the inter-event times (\mathbf{Y} in Figure 2); *minimal repair* is given by $\lambda(t | \mathcal{F}_{t^-}) = \psi(t)$, that is, the age is equal to the calendar time (\mathbf{T} in Figure 2); *imperfect repair* can be modelled by $\lambda(t | \mathcal{F}_{t^-}) = \psi(\Xi_{N(t^-)} + t - T_{N(t^-)})$ where $\Xi_{N(t^-)} \in [0, T_{N(t^-)}]$ is the *effective age* immediately after the last repair. In the BP model, $\Xi_{N(t^-)}$ is defined indirectly by letting a failed component be given perfect repair with probability p ; with probability $1 - p$ it is treated with minimal repair. To determine Ξ_i ($i > 1$) we therefore need both \mathbf{T} (or \mathbf{Y}) and the maintenance history.

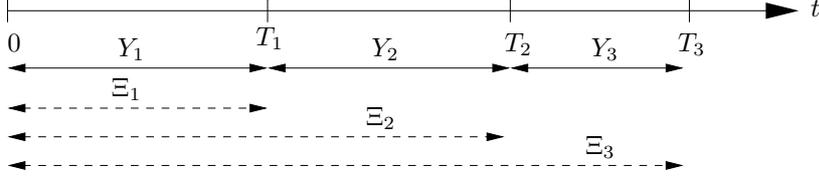


Fig. 2. We have three time dimensions to measure the age of a component: Age w.r.t. calendar time (\mathbf{T}), age w.r.t. inter-event times (\mathbf{Y}), and effective age ($\mathbf{\Xi}$). The value of Ξ_i , $i > 1$, depends upon both inter-event times as well as maintenance history. This is indicated by a dotted line for the Ξ_i 's.

It is easy to see that for $i \geq 1$ we have under the BP model that $\Xi_i = 0$ with probability p and $\Xi_i = \Xi_{i-1} + Y_i$ with probability $1 - p$; $\Xi_0 = 0$ by definition. We can also express Ξ_i by using the inter-event times:

$$\Xi_i = \begin{cases} 0 & \text{with probability } p \\ Y_i & \text{with probability } p \cdot (1 - p) \\ Y_{i-1} + Y_i & \text{with probability } p \cdot (1 - p)^2 \\ \vdots & \\ \sum_{j=2}^i Y_j & \text{with probability } p \cdot (1 - p)^{i-2} \\ \sum_{j=1}^i Y_j & \text{with probability } (1 - p)^{i-1} \end{cases} . \quad (3)$$

For simplicity of notation we follow (Kijima 1989) and introduce the random variable D_i , used to denote the outcome of the repair after the i 'th event; $D_i = 0$ if it was a perfect repair and $D_i = 1$ if it was minimal. The BP model with parameter p corresponds to assuming that all D_i are *i.i.d.* with $P(D_i = 0) = p$, $P(D_i = 1) = 1 - p$, $i = 1, \dots, n$; we will assume that D_i may be unobserved. It follows that

$$\Xi_i = \sum_{j=1}^i \left(\prod_{k=j}^i D_k \right) Y_j . \quad (4)$$

Let $\mathbf{h} = \{D_1, D_2, \dots, D_n\}$ be a repair history; that is, a set consisting of realisation of all maintenance actions telling if it was perfect or minimal. Let \mathcal{H} be the set of possible repair histories. Note that $|\mathcal{H}| = 2^{n-1}$. (Strictly speaking, there are 2^n possible repair histories, but as the last maintenance action has no effect on the observed inter-event times we have no interest in the repair at the last event, and it suffices to consider 2^{n-1} histories.)

Notice that we can express the conditional intensity given the history as $\lambda(t | \mathcal{F}_{t-}) = \psi(\Xi_{N(t-)} + t - T_{N(t-)})$. Consequently, to make the BP model oper-

ational, we only need the parameter p and the distribution function $f_{\Xi}(\xi) = \psi(\xi) \exp(-\Psi(\xi))$; $\Psi(\xi)$ is the accumulated hazard, $\Psi(\xi) = \int_{u=0}^{\xi} \psi(u) du$. Having defined those, the distribution function of the inter-event time Y_i is to be calculated as

$$f_{Y_i}(y_i | y_1, \dots, y_{i-1}) = \sum_{j=1}^i f_{\Xi} \left(\sum_{k=j}^i y_k \right) \cdot P \left(\prod_{k=j}^{i-1} D_k - \prod_{k=j-1}^{i-1} D_k = 1 \right), \quad (5)$$

where $D_0 = 0$ (the component is assumed to be as good as new at $t = 0$) and we define $\prod_{k=i}^{i-1} D_k = 1$. It is enlightening to recognise Equation (5) as a mixture distribution, where the mixture weights, $P \left(\prod_{k=j}^{i-1} D_k - \prod_{k=j-1}^{i-1} D_k = 1 \right)$, can be interpreted as the probability that the j 'th repair was the *last* perfect repair before time T_i .

One of our goals will be to estimate the BP-parameter p . It is a quantity of some interest, because it can be used as an indication of the quality of the performed maintenance. The parameter can be compared between plants and companies, and thereby unveil maintenance improvement potential (because it measures maintenance quality independent of the failure processes).

3.2 Modelling preventive maintenance versus critical failures

Recall from Section 2 that PM interventions are reported as degraded failures. Degraded failures censor critical failures, and the two types of failure may be highly correlated. We model the interaction between PM and critical failures as a competing risks problem. The “true” underlying competing risks model is not identifiable from a competing risks dataset, in particular can any dataset of this type be explained by a model of independent risks (Tsiatis 1975). Bunea and Bedford (2002) investigate the result of this model uncertainty on maintenance optimisation, and conclude that the effect of making wrong model assumptions can be substantial. Input from domain experts and careful analysis of the information actually available in the dataset is therefore of major importance when deciding how to model the relationship between preventive maintenance and critical failures. In our case, the degraded failures are defined as a step towards critical failures (consider again Fig. 1). Maintenance personnel who find a component in a degraded state will (typically) repair it to avoid a critical failure, because critical failures usually lead to higher costs than repairing the degraded failure. Hence, it is reasonable to assume the competing risks to be positively correlated.

A number of possible ways to model interaction between degraded and critical failures are discussed by Cooke (1996). We adopt one of these, called random signs censoring. In the notation introduced in Section 2 we consider here the case when we observe pairs (Y_i, J_i) where the Y_i are inter-event times whereas the J_i are indicators of failure type (critical or degraded). For a typical pair (Y, J) we let Y be the minimum of the potential critical failure time X and the potential degraded failure time Z , while $J = I(Z < X)$ is the indicator of the event $\{Z < X\}$ (assuming that $P(Z = X) = 0$ and that there are no external events). Thus we have a competing risks problem. However, while X and Z would traditionally be treated as independent, random signs censoring makes them dependent in a special way.

The basic assumption of random signs censoring is that the event of successful preventive maintenance, $\{Z < X\}$, is stochastically independent of the potential critical failure time X . In other words, the conditional probability $q(x) = P(Z < X | X = x)$ does not depend on the value of x .

Cooke (1996) does not fully describe random signs censoring, as the conditional density function $f(z | X = x, Z < X)$ can be chosen arbitrarily. Lindqvist et al. (2004) develop a framework called the *Repair Alert* (RA) model. It is a special case of random signs censoring, where the existence of a continuous function $G(t)$ such that

$$P(Z \leq z | X = x, Z < X) = \frac{G(z)}{G(x)}, \quad 0 \leq z \leq x \quad (6)$$

is assumed. Lindqvist et al. (2004) show that for any pair of sub-survival functions compatible with random signs censoring there exists exactly one repair alert model (under some regularity condition).

The repair alert model introduces the new parameters $G(\cdot)$ and q . Langseth and Lindqvist (2003) propose a special parametrisation of the RA model called the Intensity Proportional Repair Alert (IPRA) model. This model is defined as follows:

- Let X have hazard rate function $\omega(\cdot)$ and cumulative hazard $\Omega(\cdot)$.
- $\{Z < X\}$ and X are stochastically independent.
- Conditionally, given $Z < X$ and $X = x$, the distribution of the intervention time Z satisfies

$$P(Z \leq z | X = x, Z < X) = \frac{\Omega(z)}{\Omega(x)}, \quad 0 \leq z \leq x. \quad (7)$$

Thus, the IPRA model is a repair alert model where $G(t) = \Omega(t)$. The IPRA model therefore assumes the conditional density of Z proportional to the intensity of the underlying failure process. This seems like a coarse but somewhat reasonable description of the behaviour of a competent maintenance crew. The assumptions above determine the distribution for $Y = \min(X, Z)$ as

$$f_Y(y) = (1 - q) \omega(y) \exp(-\Omega(y)) + q \omega(y) \text{Ie}(\Omega(y)) \quad (8)$$

where $\text{Ie}(t) = \int_t^\infty \exp(-u)/u \, du$ is known as the exponential integral (Abramowitz and Stegun 1965).

Using the IPRA model, we define the distribution of Y , the time to the next event for a component as good as new, from the hazard rate of the failure process, $\omega(t)$, and the parameter q . In particular, this defines $\psi(t)$, the intensity of events. Finally, we can use this to calculate the conditional intensity given the history, which is given as $\lambda(t | \mathcal{F}_{t-}) = \psi(\Xi_{N(t-)} + t - T_{N(t-)})$ under the BP-model. Langseth and Lindqvist (2003) prove identifiability in this combined model.

4 The data analysis

In this section we will generate a formal test for the applicability of the combined model for the present dataset. We will assume that the underlying distribution of *failures* (the distribution of X , when using the notation of Section 3.2) follows the log-normal distribution.

4.1 Estimation of $\phi(t)$ under imperfect repair

One of the fundamental statistics when working with competing risks models is the probability of repair beyond t , $P(Z < X | Z \geq t, X \geq t)$, denoted $\phi(t)$ by Cooke (1996). One reason for this quantity to be of such importance is that many classes of models give a unique footprint in terms of the set of possible functions $\phi(t)$ they can give rise to. For example, the repair alert models are characterised as the set of models for which $\phi(t) < \phi(0)$ for all $t > 0$ (recall that $\phi(0) = q$). Similarly, it is simple algebraic manipulations to see that the footprint of the IPRA model can be expressed as

$$\phi(t)/q = \frac{\Omega(t)\text{Ie}(\Omega(t)) - \exp(-\Omega(t))}{q\Omega(t)\text{Ie}(\Omega(t)) - \exp(-\Omega(t))}. \quad (9)$$

To test if the IPRA model fits our dataset well we would like to compare the observed $\phi(\cdot)$ to the functional relationship prescribed by Equation (9). However, to do so we must be able to calculate $\phi(\cdot)$ also when the data stems from a repairable system.

Under perfect repair, we use the relationship $\phi(t) = \frac{P(Z < X, Z \geq t)}{P(Z < X, Z \geq t) + P(Z > X, X \geq t)}$ to generate the consistent estimator $\hat{\phi}(t) = \{\#z > t\} / (\{\#z > t\} + \{\#x > t\})$; we use $\{\#z > t\}$ to denote the number of observed events where $\{Z < X, Z \geq t\}$ and similarly $\{\#x > t\}$ represents the number of observed events where $\{Z > X, X \geq t\}$. When we turn to the imperfect repair, things are not that simple: We do not know the effective age of the component as it fails, and are therefore not able to calculate $\{\#x > t\}$ and $\{\#z > t\}$ directly. Note, however, that we are able to calculate these numbers when we condition on a particular repair history, and the correct way of estimating $\phi(t)$ under the BP model would be to sum over all possible repair histories

$$\hat{\phi}(t|p) = \sum_{\mathbf{h} \in \mathcal{H}} P(\mathbf{h}|p) \cdot \frac{\{\#z > t|\mathbf{h}\}}{\{\#z > t|\mathbf{h}\} + \{\#x > t|\mathbf{h}\}},$$

where $\phi(t)$ now must be calculated as a function of p and therefore is denoted $\phi(t|p)$. However, these calculations come at the computational cost of order $\mathcal{O}(2^{n+m})$. In our case, this corresponds to summing approximately 10^{25} terms, a computationally prohibitive task.

We therefore need an approximation to $\hat{\phi}(t|p)$. Our first approach is based on sampling. In particular, we sample N repair histories \mathbf{h}_i , and define

$$\hat{\phi}^N(t|p) = \frac{1}{N} \sum_{i=1}^N \frac{\{\#z > t|\mathbf{h}_i\}}{\{\#z > t|\mathbf{h}_i\} + \{\#x > t|\mathbf{h}_i\}}.$$

Obviously, $\hat{\phi}^N(t|p)$ converges to $\hat{\phi}(t|p)$ a.s. as $N \rightarrow \infty$. We can therefore approximate $\phi(t|p)$ at the cost $\mathcal{O}(N(n+m))$. In practice, however, N must be chosen quite large to offer reasonable results, and this motivates a different approximation:

$$\hat{\phi}^S(t|p) = \frac{\sum_{j=1}^n P(z_j + \Xi_{j-1} > t|\mathbf{x}, \mathbf{z})}{\sum_{j=1}^n P(z_j + \Xi_{j-1} > t|\mathbf{x}, \mathbf{z}) + \sum_{i=1}^m P(x_i + \Xi_{i-1} > t|\mathbf{x}, \mathbf{z})} \quad (10)$$

where $P(z_j + \Xi_{j-1} > t|\mathbf{x}, \mathbf{z})$ and $P(x_i + \Xi_{i-1} > t|\mathbf{x}, \mathbf{z})$ can be calculated using Equation (3). It is a consequence of Slutsky's theorem that $\hat{\phi}^S(t|p)$ is

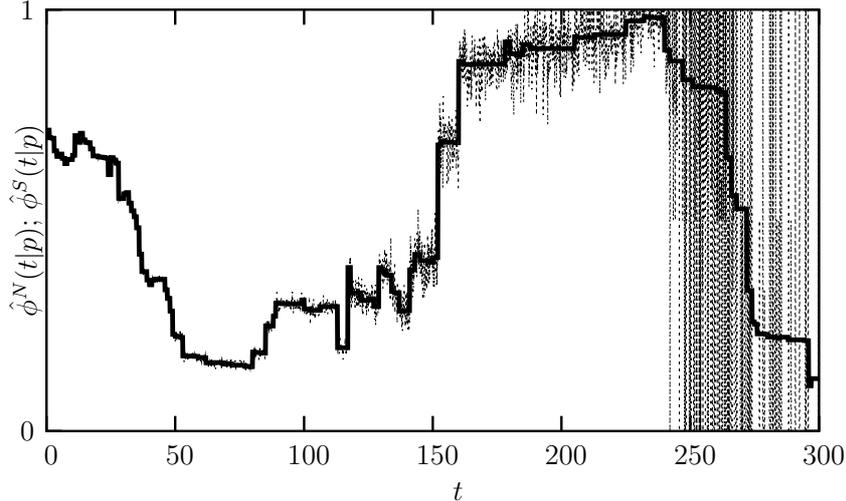


Fig. 3. The estimates $\hat{\phi}^N(t|p)$ (1000 samples per point; thin line) and $\hat{\phi}^S(t|p)$ (thick line) for our dataset. The maximum likelihood estimate $\hat{p} = .74$ was used in the calculations.

a consistent estimator of $\phi(t|p)$. It can be calculated in only $\mathcal{O}((n+m)^2)$ operations. The applicability of the two approximations for our dataset can be seen in Fig. 3. They give similar results in this case ($n = 12$, $m = 30$, $N = 1000$), although $\hat{\phi}^S(t|p)$ appears smoother for larger values of t . In the following we use $\hat{\phi}^S(t|p)$ as our approximation of $\hat{\phi}(t|p)$ if not stating otherwise.

4.2 Testing the applicability of IPRA

To use the IPRA model we make two assumptions:

- (1) The Repair Alert model must be applicable; $\phi(t|p) < \phi(0|p)$ for all $t > 0$.
- (2) Among the RA models, the IPRA model is to be chosen; $G(t) = \Omega(t)$.

It is natural to consider a two-stage test: First test if the “true” (underlying) $\phi(t|p)$ is dominated at zero, thereafter test if IPRA is a good candidate among the RA models. However, there does not, to the best of our knowledge, exist any test for the applicability of RA under imperfect repair. Dewan et al. (2002) show how U -statistics can be used to test various hypothesis regarding $\phi(t)$, but it is not obvious how the test statistic behaves when the model is integrated into the repairable systems framework. Therefore, we will not proceed along these lines here, but rather propose to use parametric bootstrap (see, e.g., (Efron and Tibshirani 1993)) to test the IPRA model directly.

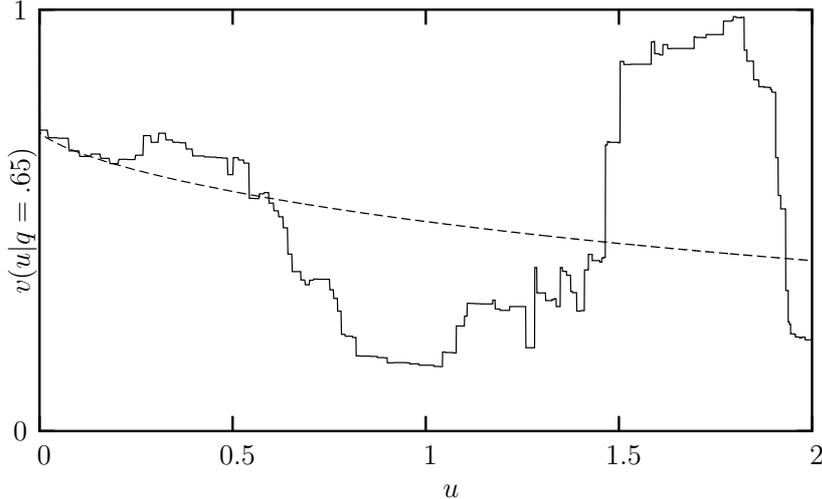


Fig. 4. The plot of the observed footprint (averaged over all repair histories; $p = \hat{p} = .74$) as compared with the theoretical one. We observe reasonable fit for smaller t , say $\Omega(t) \leq 1.5$, but poorer fit for larger values of t .

We want to use the footprint in Equation (9) as our starting point for this test. By substituting $u = \Omega(t)$, we express the IPRA footprint as

$$v(u|q) = \frac{u \text{Ie}(u) - \exp(-u)}{q u \text{Ie}(u) - \exp(-u)}. \quad (11)$$

Note that this footprint only implicitly depends on the cumulative hazard $\Omega(t)$. Fig. 4 shows $v(u|q)$ ($\Omega(t)$ and q fitted by maximum likelihood parameters) together with its observed values. We get a reasonable fit for small values of u , whereas the fit is rather poor for $u > 1.5$. However, the number of data points in this part of the model is rather small and the poor fit might therefore be expected. (When the system is as good as new there is a probability of $\exp(-1.5) = .78$ of failure beyond that time, and even some of these events will be censored by earlier maintenance.)

To formally test how applicable the combined model is for this dataset, we define Δ as the area between the theoretical footprint and the observed values in Fig. 4. A large value of Δ will suggest that the combination of IPRA and BP should be rejected. We use parametric bootstrap to approximate Δ 's distribution. The approximated density function, called $\hat{f}^{(B)}(\delta)$ in Fig. 5, can be used to make formal inference regarding the model.¹ More precisely, we

¹ On the technical side we note that since the estimates of $v(u|q)$ becomes less certain as u grows large we decided to only look at the area between the theoretical and empirical functions for $0 \leq u \leq -\ln(.1)$, that is we consider the time interval for which 90% of the failures would occur had there been no maintenance.

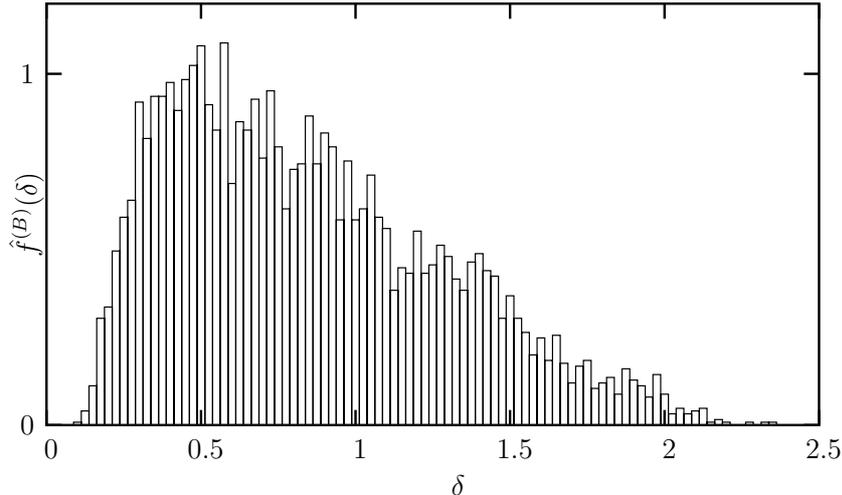


Fig. 5. $\hat{f}^{(B)}(\delta)$ is the bootstrap approximation of the distribution function of Δ . Δ is the area between footprint $v(u|p)$ and observed values.

test the hypothesis:

H_0 : The data stems from a IPRA + BP model.

H_1 : The data does not stem from a IPRA + BP model.

The dataset gives $\Delta = .65$, and from the bootstrap distribution we find that $P^{(B)}(\Delta > .65 | H_0) = 0.6$. Hence, we accept the hypothesis that IPRA combined with BP defines an appropriate class of models for this dataset.

4.3 Do we need a repairable systems model

To this end we have fitted a IPRA model for the time between events, and added a BP model on top of that to model the repair of the system. The test in the last section accepted the hypothesis that this combined model generated the dataset in Table 1. The maximum likelihood parametrisation of the model was to use $\hat{p} = .74$ in the BP model, combined with the IPRA-model where $\hat{q} = .65$, and the underlying failure process was assumed to follow the Log Normal distributions; we used the maximum likelihood estimates for the parameters: $\hat{\mu} = 3.7$ and $\hat{\tau}^2 = 1.8^2$.

To test whether the repairable system part of the model is required, we use the likelihood ratio test. The full model as outlined in Section 3 ($\hat{p} = .74$) obtains a log likelihood of -212.1 . The IPRA as renewal process (all repairs are perfect; $p = 1$) gives log likelihood of -214.4 , which is significantly poorer than the full

model at level 0.03. IPRA as a nonhomogeneous Poisson Process (all repairs are minimal; $p = 0$) gives log likelihood of -258.9 , which is significantly poorer than the full model at all reasonable levels. We therefore conclude that the repairable systems model is indeed required (at level .03); this is in accordance with expert knowledge, which also points towards perfect and minimal repair models being unacceptable.

5 Conclusions

In this paper we have analysed a dataset from the OREDA database (OREDA 2002) describing a compressor system. We fitted a competing risks model and coupled that with the imperfect repair. Inference was done to verify that this model was capable of describing the dataset well.

We stress that due to non-identifiability of competing risks problems we cannot really infer that IPRA combined with the BP model is the *correct* model for the present dataset. We can only conclude that the data does not reject this combined model, and that if we were to choose a building-block among the RA models, then IPRA can be a reasonable alternative (this is also strengthened by the fact that IPRA is inexpensive w.r.t. the number of required parameter). To be advocating this model we would need to discuss the assumptions of RA with domain experts.

Leaving non-identifiability issues aside, the main goal of our analysis has been to find parameters that describe the quality of the performed maintenance. We have fitted maximum likelihood parameters, and found that the maintenance crew's "thoroughness" is quantified by the parameter p ($\hat{p} = .74$) and their "eagerness" by q ($\hat{q} = .65$). It is our belief that these numbers, together with information about the annual maintenance expenditure and the number of critical failures that occur per year, characterise the efficiency of the maintenance at a level that can be compared between different maintenance crews and even between different installations.

Acknowledgements

We thank the OREDA project represented by the Steering Committee Chairman Runar Østebø and Project Manager Terje Dammen for making the data

in Table 1 available to us. Furthermore, we thank the participants at the Workshop on Analysis of Competing Risks – Statistical and Probabilistic Approach (Delft, Holland) for interesting discussions.

References

- Abramowitz, M. and I. A. Stegun (1965). *Handbook of Mathematical Functions*. New York: Dover Publications.
- Andersen, P., Ø. Borgan, R. Gill, and N. Keiding (1992). *Statistical models based on counting processes*. New York: Springer-Verlag.
- Ascher, H. and H. Feingold (1984). *Repairable Systems Reliability – Modeling, Inference, Misconceptions and Their Causes*. New York: Marcel Dekker, Inc.
- Brown, M. and F. Proschan (1983). Imperfect repair. *Journal of Applied Probability* 20, 851–859.
- Bunea, C. and T. Bedford (2002). The effect of model uncertainty on maintenance optimization. *IEEE Transactions on Reliability* 51(4), 486–493.
- Cooke, R. M. (1996). The design of reliability data bases, Part I and Part II. *Reliability Engineering and System Safety* 52, 137–146 and 209–223.
- Dewan, I., J. V. Deshpande, and S. B. Kulathinal (2002). On testing dependence between time to failure and cause of failure via conditional probabilities. Technical Report isid/ms/2002/17, Statistics&Mathematics Unit, Indian Statistical Institute, Delhi Centre.
- Efron, B. and R. J. Tibshirani (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability* 26, 89–102.
- Langseth, H. and B. H. Lindqvist (2003). A maintenance model for components exposed to several failure modes and imperfect repair. In K. Doksum and B. H. Lindqvist (Eds.), *Mathematical and Statistical Methods in Reliability, Quality, Reliability and Engineering Statistics*, Chapter 27, pp. 415–430. Singapore: World Scientific Publishing Co.
- Lindqvist, B. H., H. Langseth, and B. Støve (2004). Modeling of dependence between critical failure and preventive maintenance: The repair alert model. Under review for inclusion in this special issue. Available at <http://www.math.ntnu.no/~helgel/ra.pdf>.
- OREDA (2002). *Offshore Reliability Data Handbook* (4th ed.). Distributed by Det Norske Veritas, P.O. Box 300, N-1322 Høvik, Norway. See also <http://www.oreda.com/>.
- Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences USA* 72, 20–22.