Building a Data-Driven Vital Sign Indicator for an Economically Optimized Component Replacement Policy

Hyung-il Ahn[1], Ying Tat Leung[2], and Axel Hochstein[3]

[1,3] *GE Software, GE Global Research, San Ramon, CA, 94583, USA*
hyungil.ahn@ge.com
axel.hochstein@ge.com

[2] *IBM Almaden Research Center, San Jose, CA, 95120, USA*
ytl@us.ibm.com

ABSTRACT

In asset-intensive services, a well-known challenge is to maintain high availability of the physical assets while keeping the total maintenance cost low. In applications of high-value machinery such as heavy industrial equipment, a traditional approach is to perform periodic maintenance according to a runtime-based schedule. Most equipment vendors publish a maintenance schedule based on a "standard" or "average" working environment. In addition, it is a common practice that maintenance schedules from equipment vendors are highly conservative in order to reduce in-field failures which gives an adverse perception of a vendor's reputation. Therefore, such a schedule may not result in satisfactory performance as measured according to the owner's business objectives. Also, the assumption of normal operating condition may not apply in some situations. For example, stresses due to frequent overloading, continuous usage of engine at a high rate in tough environments, machine usage beyond its designed capacity can serve as good contributors to excessive wear and premature failures. In this paper we propose a novel computational framework to build a data-driven economically optimized vital sign indicator for a given component type and an economic criterion (e.g., average maintenance cost per unit runtime) by combining different sources of historical data such as total runtime hours, load carried, fuel consumed and event information from sensors. This new vital sign indicator can be viewed as a transformed time scale and used to find the optimal threshold value (or "scheduled replacement time equivalent") for a component replacement policy. Our case study was based on the collected data from 50 mining haul trucks over about 6 years in one of the largest mining service companies in the world. We present that the new vital sign indicator-based replacement policy for a critical component type largely improves on the traditional runtime-based schedule in terms of a given economic criterion, achieving a lower total maintenance cost of the enterprise.

1. INTRODUCTION

A traditional replacement policy for components in asset-intensive service business is often based on runtime hours-based fixed time interval ("scheduled replacement time") that the manufacturer of equipment recommends for scheduled maintenance. This is based on standard usage in an average situation assumed by the manufacturer. Most equipment vendors publish a maintenance schedule based on a "standard" or "average" working environment. In addition, it is a common practice that maintenance schedules from equipment vendors are highly conservative in order to reduce in-field failures which gives an adverse perception of a vendor's reputation. Therefore, such a schedule may not result in satisfactory performance as measured according to the owner's business objectives. Also, the assumption of normal operating condition may not apply in some situations. For example, stresses due to frequent overloading, continuous usage of engine at a high rate in tough environments, machine usage beyond its designed capacity can serve as good contributors to excessive wear and premature failures.

In asset-intensive services, a well-known challenge is to maintain high availability of the physical assets while keeping the total maintenance cost low (Jardine & Tsang, 2013). The optimization of replacement decision policy based on component failure predictions has been critical in the area of condition-based predictive asset management. One of the most popular approaches involves modeling a proportional hazard function (Cox PHM) with time-dependent covariates and a Weibull baseline hazard function

(Banjevic, Jardine, Makis, & Ennis, 2001)(Jardine, Banjevic, Montgomery, & Pak, 2008). In practice, the modeled hazard function using this approach is not guaranteed to be monotonically increasing, and thus, it often involves a complicated algorithm to compute the optimal policy (Wu & Ryan, 2011). Furthermore, a non-monotonic hazard function is not very intuitive and cannot be viewed as a new kind of time scale. Equipment managers would often like to have a time scale-like monotonically increasing measure for the component replacement policy. Then, they could use this new vital sign indicator measure exactly in the same way they used the runtime measure for replacement decisions.

In this paper we propose a novel computational framework to build a data-driven economically optimized vital sign indicator for a given component type and an economic criterion (e.g., average maintenance cost per unit runtime) by combining different sources of historical data such as total runtime hours, load carried, fuel consumed and event information from sensors. A vital sign indicator can provide a measure that contains useful information with respect to the "health" of a piece of a component or equipment, and can therefore support improved decision making in terms of maintenance planning and execution, as well as production maximization. This new vital sign indicator can be viewed as a transformed time scale and used to find the optimal threshold value (or "scheduled replacement time equivalent") for a component replacement policy. We provide an individualized maintenance plan for each component based on its real usage. Our approach involves classification and regression techniques for estimating a hazard rate and uses the "individualized" cumulative failure probability model for building a vital sign indicator.

Our case study was based on the collected data from 50 mining haul trucks over about 6 years in one of the largest mining service companies in the world. We present that the new vital sign indicator-based replacement policy for a critical component type largely improves on the traditional runtime-based schedule in terms of a given economic criterion, achieving a lower total maintenance cost of the enterprise.

2. COMPONENT REPLACEMENT POLICIES

2.1. Runtime-based Replacement Policy

Figure 1 shows an example of the failure probability density function with T^* (optimal scheduled replacement time) for a component type. Assuming that a company has run a scheduled replacement policy at T^*, at the time of collecting the component data for our analysis, the historical list of all components of this component type over a group of equipment include running components (at the time of data collection), schedule-replaced components, and failure-replaced components. In Figure 1 each circle represents a component in the list. All blue circles before T^* correspond to running components and their observed runtimes at the time of data collection. All blue circles after T^* correspond to schedule-replaced components. Note that companies in practice often do not keep the exact replacement schedule at T^*. All red circles before T^* correspond to in-field failure replacements. Note that running and scheduled replacement components are considered "right-censored" samples in survival analysis. That is, we know that the components survived at the time of data collection or scheduled replacement, but cannot tell when those components would actually fail in the future.

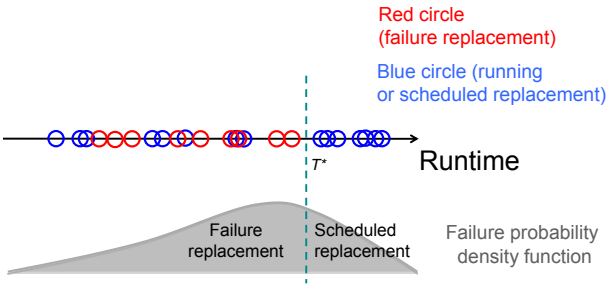


Figure 1. An example of failure probability density function with the optimal scheduled replacement time T^*

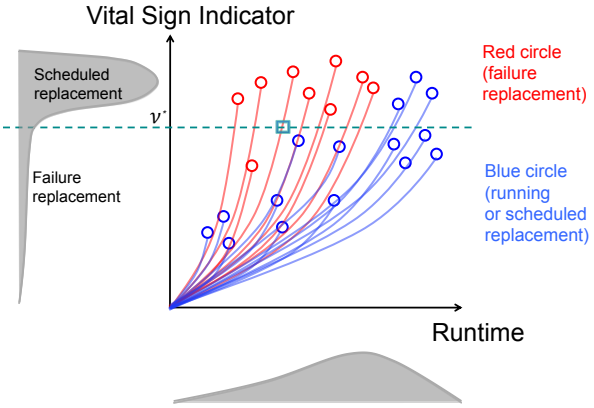


Figure 2. An example of vital sign indicator with the optimal scheduled replacement vital sign value v^*

Note in Figure 1 that the standard deviation of the failure probability density function is very large; thus, we have too many in-field failure-replaced components

2.2. Vital sign-based Replacement Policy

Now we conceptually explain the development of our new vital sign indicator model. For the historical list of all components, we also have the corresponding time-stamped logs of runtime hours (meter), total fuel consumption, total work (load) and sensor events. Imagine that for the component data and the failure probability density function shown in Figure 1, we can design a vital sign indicator

(vertical axis) in Figure 2 using some features derived from all available information. Note that the time/color of each circle in Figure 2 are exactly the same as those of the corresponding circle in Figure 1, and the color (failure replacement (red), running or scheduled replacement (blue)) of each path is based on the collected component data (i.e., the traditionally employed runtime-based replacement policy), not according to the new vital sign-based replacement policy.

Then, we propose a vital sign indicator-based scheduled replacement policy that replaces components when their vital sign value reaches a threshold value v^*. In Figure 2, the dotted line shows the threshold value. Each path in the runtime vs. vital sign indicator 2-dimensional plot corresponds to a component and shows its vital sign indicator profile over the runtime. Note that the runtime (= the value in the horizontal axis) at the intersection point between the threshold line and the path for a component indicates the actual replacement time using the policy.

Keep in mind that the failure probability density function in terms of the vital sign indicator axis depends on our model of a vital sign indicator. Intuitively, one desirable characteristic for being a good vital sign indicator is a small standard deviation in the vital sign indicator axis. This contributes to a better classification, using a constant v^*, between the failure-replaced components (above the v^* line) and the other running/schedule-replaced components (below the v^* line). In other words, if this vital sign indicator-based scheduled replacement policy had been used in the past, most of failure-replaced components in the collected data (red circles) would have been replaced on schedule (at v^*) before the actual in-field failures. However, this characteristic about the failure probability is not a sufficient condition to be a good vital sign indicator model, since the average runtime to scheduled replacements (i.e., the average of actual runtimes from intersection points at v^*) and the average runtime to failure replacements should also be large values. For this reason, we should look into the shape of vital sign paths in the runtime vs. vital sign indicator 2-dimensional plot. We will explain it using economic optimization equations below in more detail.

3. ECONOMIC OPTIMIZATION

3.1. Runtime-based Replacement Policy

Let $F(t)$ be the cumulative failure probability function at runtime t (= $\Pr(T \le t)$ where T is a random variable denoting the runtime at failure), $S(t) = 1 - F(t)$ be the survival probability function at t. When we deal with the dataset from real industry practice, it is very likely that there is no failure data after the scheduled replacement time the company has employed during the period of the dataset. Therefore, we would not make a good estimate on the exact shape of the function over the time after the current scheduled replacement time. However, in this paper we assume that the survival probability function can be estimated using a parametric Weibull fit (Fox, 2002) to the runtime and failure data.

For our economic optimization analysis, we are provided the economic and logistic parameters including

C_f = in-field failure replacement cost, which includes the part and labor cost to replace the component, the retrieval cost of equipment from the field, and lost revenue due to blocking other equipment when it fails in the field (called "circuit break"),

C_p = scheduled replacement cost, which includes the part and labor cost to replace the component,

c_d = cost per unit downtime of the equipment, including lost revenue that could have been contributed by that piece of equipment,

DT_f = down time due to an in-field failure,

DT_p = down time due to a scheduled replacement.

In general, in-field failure replacement cost and downtime are greater than scheduled replacement cost and downtime, respectively ($C_f > C_p$, $DT_f > DT_p$).

Denote by t_p the scheduled replacement time for the policy, which is our optimization target. With this scheduled replacement policy, the mean time to failure replacement that happens before t_p is denoted by t_f and estimated as:

$$t_f = \frac{1}{F(t_p)} \int_0^{t_p} t f(t) dt = t_p - \frac{\int_0^{t_p} F(t) dt}{F(t_p)}$$

A new component lifetime cycle starts at the installation time of a component. The component may be replaced due to an in-field failure or a scheduled replacement finishing its lifetime cycle.

For a runtime-based replacement policy, we choose t_p to minimize the *average maintenance cost per unit runtime*.

average total time per cycle

$$= (t_f + DT_f)F(t_p) + (t_p + DT_p)(1 - F(t_p))$$

average run time per cycle $= t_f F(t_p) + t_p(1 - F(t_p))$

average maintenance cost per unit runtime

$$= \frac{average\ maintenance\ cost\ per\ cycle}{average\ runtime\ per\ cycle}$$

$$= \frac{(average\ failure\ replacement\ cost\ per\ cycle + average\ scheduled\ replacement\ cost\ per\ cycle)}{average\ runtime\ per\ cycle}$$

$$= \frac{(C_f + c_d DT_f)F(t_p) + (C_p + c_d DT_p)(1 - F(t_p))}{t_f F(t_p) + t_p(1 - F(t_p))}$$

$$= (C_f - C_p + c_d(DT_f - DT_p))\frac{F(t_p)}{X} + (C_p + c_d DT_p)\frac{1}{X}$$

where X = *average run time per cycle*

$$= t_f F(t_p) + t_p(1 - F(t_p))$$

As t_p (= the scheduled replacement time) is set to a higher value, there is more chance of in-field failure replacements, that is, $F(t_p)$ (= the total probability of in-field failure replacements) becomes larger (See Figure 3).

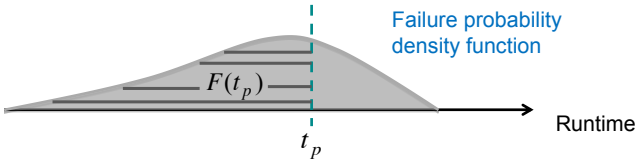


Figure 3. The trade-off between the average runtime per cycle and $F(t_p)$ (= total in-field failure probability)

Since $DT_f > DT_p$ and $C_f > C_p$ in general, the optimization goal of minimizing the average maintenance cost per unit runtime is achieved by increasing *average runtime per cycle* ($X = t_f F(t_p) + t_p(1 - F(t_p))$) and decreasing in-field failure probability per cycle $F(t_p)$. Note that there is a trade-off between decreasing $F(t_p)$ and increasing the *average runtime per cycle*. In general, decreasing $F(t_p)$ that would involve fewer failure replacements can be obtained by decreasing t_p, but this then reduces the *average run time per cycle*. Note that $t_f < t_p$ in general. Also, note that as $F(t_p)$ becomes smaller, t_p becomes more weighted in the estimate of average run time per cycle. Given $F(t), C_f, C_p, c_d, DT_f$ and DT_p, the average maintenance cost per unit runtime is a function of t_p, which is denoted as g.

$$g(t_p) = \frac{(C_f + c_d DT_f)F(t_p) + (C_p + c_d DT_p)(1 - F(t_p))}{t_f F(t_p) + t_p(1 - F(t_p))}$$

It is important to note that the cumulative failure probability function $F(t)$ is fixed and can be estimated using the failure data for the component type we analyze. Note also that t_f depends on $F(t)$. Then, the optimized time threshold for the scheduled replacement policy is $t_p^* = \arg\max_{t_p} g(t_p)$.

3.2. Vital Sign-based Replacement Policy

Let v be vital sign indicator. $\hat{F}(v)$ be the cumulative failure probability function at vital sign v (= $\Pr(V \le v)$ where V is a random variable denoting the vital sign at failure), $\hat{S}(v) = 1 - \hat{F}(v)$ be the survival probability function at v. Note that we estimate this survival probability function by a local regression (loess) on the Kaplan-Meir (KM) estimate (Therneau, 2000) using the vital sign and failure data.

Denote by v_p the vital sign threshold value for scheduled replacements for the vital sign-based scheduled replacement policy, which is our optimization target. Then, $\hat{F}(v_p)$ is the total expected probability of failure replacements, and $1 - \hat{F}(v_p)$ is the total expected probability of scheduled replacements. With this scheduled replacement policy, the expected time to scheduled replacement at v_p is denoted by $\hat{t}_p$. Also, the expected time to failure replacement is denoted by $\hat{t}_f$. In this paper we estimate $\hat{t}_p$ and $\hat{t}_f$ under reasonable assumptions.

Let $Comp[v \ge v_p]$ denote the set of all components whose vital sign value reaches v_p in the dataset, whereas $Comp[v < v_p]$ denotes the set of all components whose vital sign value $v < v_p$ for all time t in the dataset.

Let $P[v \ge v_p]$ denote the *actual* ratio of the number of components in $Comp[v \ge v_p]$ to the total number of components *in the dataset*. The actual ratio $P[v \ge v_p]$ is equal to or smaller than $1 - \hat{F}(v_p)$ (= total expected probability of scheduled replacement), since the total expected probability takes right-censored components (running at the time of data collection) into account. There are running components that would fail with $v > v_p$. We assume that those components contribute to scheduled replacements corresponding to the difference between the expected probability and the actual ratio (= $1 - \hat{F}(v_p) - P[v \ge v_p]$) and that they are schedule-replaced at v_p with the cumulative probability function of the replacement time, $\hat{F}_{v<v_p}(t) = 1 - \hat{S}_{v<v_p}(t)$ where $\hat{S}_{v<v_p}(t)$ is the survival probability function estimated using a Weibull fit to the runtime and failure data of $Comp[v < v_p]$. In other words, we assume that $\hat{F}_{v<v_p}(t)$ estimated using $Comp[v < v_p]$ is uniformly applied to all the range of $v < v_p$. Thus, the mean scheduled replacement time over those components corresponding to $1 - \hat{F}(v_p) - P[v \ge v_p]$ is the same as the mean failure time over $Comp[v < v_p]$, which is denoted by r and estimated as $r = \int_0^\infty \hat{S}_{v<v_p}(t)dt$. Thus,

$\hat{t}_f$ = expected time to failure replacement $= \int_0^\infty \hat{S}_{v<v_p}(t)dt$.

$\hat{t}_p$ = expected time to scheduled replacement

$= \{ P[v \geq v_p]\, E[t|v = v_p \text{ for } Comp[v \geq v_p]] +$

$(1-\hat{F}(v_p) - P[v \geq v_p])\, r \} / (1-\hat{F}(v_p))$.

Note that $E[t|v = v_p \text{ for } Comp[v \geq v_p]]$ is the average of scheduled replacement times at $v = v_p$ over $Comp[v \geq v_p]$.

Alternatively, we may assume that components in $Comp[v < v_p]$ that would fail after t_c contribute to scheduled replacements for the difference $(= 1-\hat{F}(v_p) - P[v \geq v_p])$, whereas components in $Comp[v < v_p]$ that would fail before t_c are failure-replaced. Also, we can estimate t_c from the constraint $\hat{F}(v_p) = \hat{F}_{v<v_p}(t_c)\ (1 - P[v \geq v_p])$. That is, the total expected probability of failure replacements over all components $(= \hat{F}(v_p))$ should be the same as the actual ratio of the number of components in $Comp[v < v_p]$ to the number of total components in the dataset $(= 1 - P[v \geq v_p])$ multiplied by the total expected probability of failure replacements before t_c over $Comp[v < v_p]$ $(= \hat{F}_{v<v_p}(t_c))$. Thus,

$\hat{t}_f$ = expected time to failure replacement

$$= t_c - \frac{\int_0^{t_c} \hat{F}_{v<v_p}(t)dt}{\hat{F}_{v<v_p}(t_c)} .$$

Then, the mean scheduled replacement time over those components corresponding to $1-\hat{F}(v_p) - P[v \geq v_p]$ is denoted by r and estimated as

$$r = \{ \int_0^\infty \hat{S}_{v<v_p}(t)dt - \hat{t}_f\ \hat{F}_{v<v_p}(t_c) \} / (1-\hat{F}_{v<v_p}(t_c)).$$

$\hat{t}_p$ = expected time to scheduled replacement

$= \{ P[v \geq v_p]\, E[t|v = v_p \text{ for } Comp[v \geq v_p]] +$

$(1-\hat{F}(v_p) - P[v \geq v_p])\, r \} / (1-\hat{F}(v_p))$.

For this vital sign-based replacement policy, we choose v_p to minimize the *average maintenance cost per unit runtime*.

Average maintenance cost per unit runtime

$$= \frac{\textit{average maintenance cost per cycle}}{\textit{average runtime per cycle}}$$

$$= \frac{(C_f + c_d DT_f)\hat{F}(v_p) + (C_p + c_d DT_p)(1-\hat{F}(v_p))}{\hat{t}_f \hat{F}(v_p) + \hat{t}_p (1-\hat{F}(v_p))}$$

$$= (C_f - C_p + c_d(DT_f - DT_p))\frac{\hat{F}(v_p)}{\hat{X}} + (C_p + c_d DT_p)\frac{1}{\hat{X}}$$

where $\hat{X}$ = *average run time per cycle*

$$= \hat{t}_f \hat{F}(v_p) + \hat{t}_p (1-\hat{F}(v_p))$$

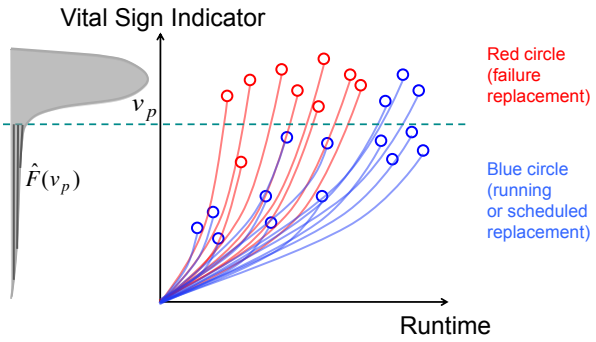


Figure 4. Vital-sign indicator functions steeply increasing around v_p: no strong trade-off between the average runtime per cycle and $\hat{F}(v_p)$ (= total in-field failure probability)

As in the analysis of the runtime-based policy, the optimization goal of minimizing *average maintenance cost per unit work* is achieved by increasing *average run time per cycle* $(= \hat{t}_f \hat{F}(v_p) + \hat{t}_p (1-\hat{F}(v_p)))$ and decreasing in-field failure probability per cycle $\hat{F}(v_p)$. However, in contrast to the runtime-based policy, with vital-sign indicator functions steeply increasing around v_p, there is no strong trade-off between decreasing $\hat{F}(v_p)$ and increasing the *average run time per cycle*. In other words, decreasing $\hat{F}(v_p)$ that would involve fewer failure replacements can be obtained by decreasing v_p but this does not necessarily lead to a large decrease of $\hat{t}_p$ (= the average of scheduled replacement times at v_p) when the vital-sign indicator functions are steeply increasing around v_p (compared with slowly increasing shaped functions). More importantly, considering the definitions of $\hat{t}_p$ (involving the term $[t|v = v_p \text{ for } Comp[v \geq v_p]]$) and $\hat{t}_f$ (involving $\hat{S}_{v<v_p}(t)$ or $\hat{F}_{v<v_p}(t)$), if decreasing v_p would allow failures that happen *later in time* to be schedule-replaced, this would tend to increase both $\hat{t}_p$ and $\hat{t}_f$, as well as decreasing $\hat{F}(v_p)$; thus, this helps the optimization

goal. Also, if decreasing v_p would allow failures that happen *earlier in time* to be schedule-replaced, this would tend to decrease $\hat{t}_p$ but still tend to increase $\hat{t}_f$ and decrease $\hat{F}(v_p)$. Note that in contrast to the runtime-based policy, $\hat{t}_f$ is not necessarily smaller than than $\hat{t}_p$ for a vital sign-based policy. That is, decreasing $\hat{t}_p$ does not lead to decreasing $\hat{t}_f$. The values of $\hat{t}_p$ and $\hat{t}_f$ at the optimization of v_p rely on the complete distribution and paths in the runtime vs. vital sign indicator 2-dimensional plot.

It is critical to note that the shape of cumulative failure probability function $\hat{F}(v')$ for any candidate threshold v' can be changed according to our modeling parameters to design a vital sign indicator. Note also that $\hat{t}_p$ and $\hat{t}_f$ for any candidate threshold v' (i.e., functions of v') depend on the designed vital sign indicator.

Given $C_f, C_p, c_d, DT_f, DT_p$, $\hat{F}(v')$, $\hat{t}_p(v')$ and $\hat{t}_f(v')$ for a designed vital sign indicator, the average maintenance cost per unit runtime is a function of v_p, which is denoted as $\hat{g}$.

$$\hat{g}(v_p \mid \hat{F}(v'),\ \hat{t}_p(v'),\ \hat{t}_f(v')) =$$

$$\frac{(C_f + c_d DT_f)\hat{F}(v_p) + (C_p + c_d DT_p)(1 - \hat{F}(v_p))}{\hat{t}_f(v_p)\hat{F}(v_p) + \hat{t}_p(v_p)(1 - \hat{F}(v_p))}$$

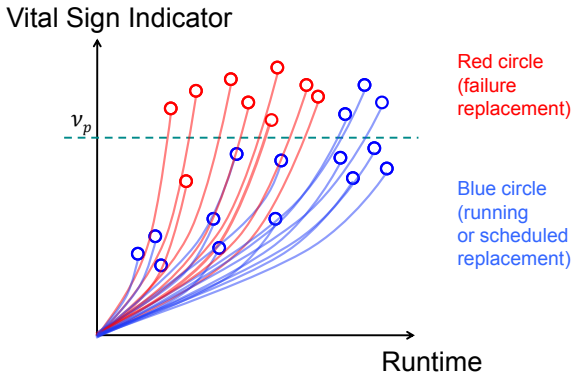
Thus, the value of $\hat{g}$ at v_p is determined by our design of the vital sign indicator, which is what the paths of vital sign over time look like.

Then, the optimized vital sign threshold value for the scheduled replacement policy using this vital sign indicator is $v_p^* = \arg\max_{v_p} \hat{g}(v_p \mid \hat{F}(v'), \hat{t}_p(v'),\ \hat{t}_f(v'))$.

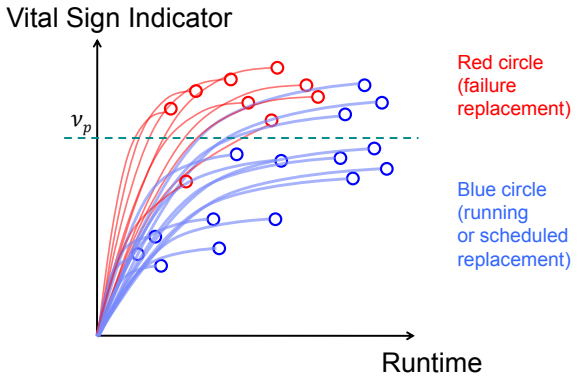
We compare the runtime-based component replacement policy with the new designed vital sign-based replacement policy in terms of the average maintenance cost per unit runtime. That is, we compare $g(t_p^*)$ with $\hat{g}(v_p^* \mid \hat{F}(v'), \hat{t}_p(v'),\ \hat{t}_f(v'))$.

If $\hat{g}(v_p^* \mid \hat{F}(v'), \hat{t}_p(v'),\ \hat{t}_f(v')) > g(t_p^*)$, this means that the designed vital-sign based replacement policy is more beneficial in terms of the economic criterion.

4. BUILDING A VITAL SIGN INDICATOR BASED ON CLASSIFICATION AND REGRESSION



(a) Convex-shaped vital sign indicator model



(b) Concave-shaped vital sign indicator model

Figure 5. Comparing convex-shaped and concave-shaped vital sign indicator models

In Figure 5(a) and (b), we compare two hypothetical vital sign indicator models (convex-shaped and concave-shaped) when the failure probability density functions in the vital sign indicator axis are the same, although this would hardly happen in practice. For the same vital sign threshold value v_p, the convex shape in Figure 5 (a) would have a greater average runtime to scheduled replacement ($\hat{t}_p$ = the average of runtimes from all intersection points) than the concave shape in Figure 5 (b). The convex paths would predict the upcoming failures near the actual failure times, whereas the concave paths would predict the upcoming failures too early. The concave paths would have a smaller average runtime due to too early replacements. Thus, in general, the convex-shaped vital sign indicator model would be more desirable than the concave-shaped one. This is also why we should look into the complete vital sign paths, not just examining the shape of failure probability density function or $\hat{F}(v_p)$.

Before explaining our vital sign indicator model, we first introduce the notion of "*individualized* cumulative failure

probability function". For each individual component, let us consider a hypothetical population of components that share the same history of covariates as that component has. Then, we can define a cumulative distribution function of the failure time for the population. We call it the individualized cumulative failure probability function for the component. In addition, the individualized cumulative failure probability function $F_j(t)$ of component j has the following relationship with the individualized cumulative hazard function $H_j(t)$:

$F_j(t) = 1 - S_j(t) = 1 - \exp(-H_j(t))$ where $S_j(t)$ is the individualized survival probability function.

In this paper we model the vital sign indicator using the individualized cumulative failure probability function. That is, the vital sign indicator for a component is the same as its individualized cumulative failure probability over runtime.

In the runtime-based policy we select the best scheduled replacement time so that the cumulative failure probability $F(t_p)$ optimizes the economic criterion. In contrast, in the vital sign-based policy for scheduled replacements, we apply a selected vital sign threshold value to the individualized cumulative failure probability functions $F_j(t)$ of components. This is the same as applying a common threshold to the individualized cumulative hazard functions $H_j(t)$. Note that this individualization in cumulative failure probability (or cumulative hazard) is critical to allow each component to have its own transformed time scale for the replacement policy.

The individualized cumulative hazard $H_j(t)$ assesses the total amount of accumulated risk that the component j has faced from the beginning of time until the present, while the (instantaneous) hazard rate assesses the risk that a component which has not yet had the failure so will experience it within a unit of runtime (Singer & Willett, 2003). Compared to using the hazard rate in designing a scheduled replacement policy, applying the individualized cumulative hazard $H_j(t)$ has some advantages. First, in contrast to the hazard rate, the individualized cumulative hazard may capture the accumulated wear and tear over the component runtime. Second, the individualized cumulative hazard is always increasing, whereas the hazard rate may be fluctuating up and down over the runtime. Note that the characteristic of monotonically increasing is necessary because the vital sign indicator is conceptualized as a transformed time scale. In addition, people usually think that the accumulated wear and tear is always increasing over the runtime, that is, the quality of a component becomes worse with runtime.

Considering that our dataset includes daily-interval samples, we define the daily hazard $h_j(d)$ on date d for component j by the total hazard during the daily runtime. That is, daily hazard = hazard rate × daily runtime. Then, we can estimate the individualized cumulative hazard by summing up all daily hazards until the present time t: $H_j(t) = \sum_{\text{all } d \text{ in } \{d:Meter(j,d)\le t\}} h_j(d)$ where $Meter(j,d)$ is the accumulated runtime hours over days up to and including date d.

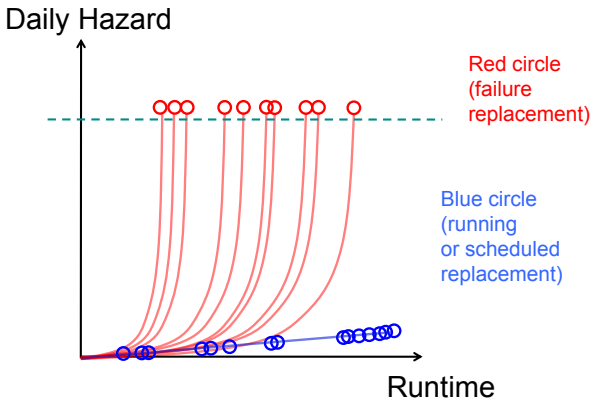


Figure 6. An example of the "designed" daily hazard as a regression target variable

It is important to note that the "estimated" daily hazard depends on our selection of covariates and the model. Also, daily hazard estimates from a desirable model would predict its failure near the date of actual failure time. Wrong predictions or too early predictions of failures would lead to the reduction of average runtime. Thus, it will be better to find the covariates and model that enable the daily hazard estimates to be convex-shaped and very close to the maximum value (= 1) near the date of actual failure time (e.g., Figure 6). In practice, however, we do not require the daily hazard estimates to be necessarily convex-shaped, because it may not be possible with our selected features and modeling choice. We only want the individualized cumulative hazards to satisfy some desired characteristics (monotonically increasing, high values of $\hat{t}_p$ and $\hat{t}_f$, high vital sign values on the failure times) for the economic criterion. Thus, we set up our problem of designing a vital sign indicator model as a regression task where the regression target variable is the "designed" daily hazard $\tilde{h}_j(d)$ we specify on any date d for component j as follows:

- If the component was failure-replaced, $\tilde{h}_j(d) = (Meter(j,d)/Meter(j,T_F(j)))^{\alpha}$ where $Meter(j,d)$ is the total runtime hours up to and including date d, $T_F(j)$ is the finally observed date (or the replaced date), and $\alpha \ge 1$.

- If the component was schedule-replaced or actively running, $\tilde{h}_j(d) = \beta(Meter(j,d)/M_{max})^{\alpha}$ where $M_{max} = max_i[Meter(i,T_F(i))]$ = the maximum total runtime hours over all components in the dataset, and β ($\ll 1$) is a small positive number close to 0 (e.g., $\beta = 0.1$).

That is, shown as in Figure 6, the first equation satisfies the condition that failure-replaced components have the maximum value (= 1) near the date of actual failure time. Also, the second equation allows the running/schedule-replaced components to have low values over their runtimes.

We build vital sign indicator models by performing regression tasks with differently designed daily hazard setups (different α and β values), and find the best vital sign indicator model in terms of the economic optimization criterion estimate by leave-one-component-out cross-validations. We will describe it below in detail.

Provided that we have the list of past replaced components (failure or scheduled replacements) and current running ones for a component type over a group of equipment as well as the corresponding time-stamped logs of runtime hours (meter), total fuel consumption, total work (load) and sensor events, we propose a framework of building a vital sign indicator for the component type using regression.

Suppose that there are totally J components that were past replaced or are actively running for the target component type. For component j (=1, …, J), the start date of service is $T_S(j)$, and the final date of observation is $T_F(j)$. Note that the final date of observation is defined as the replaced date for past components or the last observed date for running components. For this task, the overall dataset includes all points $x(j,d)$ over component j (=1, …, J) and date d (=$T_S(j)$, …,$T_F(j)$).

Input data:
From the start date of service of component j,

- *Meter(j,d)* = accumulated runtime hours over days up to and including date d
- *Fuel(j,d)* = accumulated fuel consumption over days up to and including date d
- *Load(j,d)* = accumulated number of loads (total work) over days up to and including date d
- *EventCount(j,d)* = accumulated number of relevant sensor events for the target component type over days up to and including date d

Note that $Meter(j, T_S(j)) = 0$, $Fuel(j, T_S(j)) = 0$, $Load(j, T_S(j)) = 0$, and $EventCount(j, T_S(j)) = 0$. Here we assume that the relevant sensor event types for the component type are selected using the significance test in a univariate Cox proportional hazard model for each event type (Hastie, Tibshirani, Friedman, & Franklin, 2005)(Bair, Hastie, Paul, & Tibshirani, 2006). But other techniques including frequent sequence mining (Zaki, 2001) on component failure and event data can be exploited for the same purpose.

Given the parameters such as

N_{smooth} = positive integer for a smoothing filter,

N_{fuel} = positive real threshold value for counting the number of dates with high daily fuel rate,

N_{load} = positive real threshold value for counting the number of dates with high daily load rate,

we compute intermediate variables as follows. Note that these intermediate variables are used to calculate features. Also, the purpose of N_{fuel} and N_{load} is to count outliers. Although we present this simple rule-based outlier detection here, our framework allows other sophisticated anomaly detection algorithms to be applied for more effective feature generation.

Intermediate variables:

- *DailyMeter*(j,d) = daily meter hours on date d
  $= Meter(j,d) – Meter(j,d-1)$
- *DailyFuel*(j,d) = daily fuel consumption on date d
  $= Fuel(j,d) – Fuel(j,d-1)$
- *DailyLoad*(j,d) = daily number of loads on date d
  $= Load(j,d) – Load(j,d-1)$
- *SmoothedDailyMeter(j,d)* = average daily meter hours over past N_{smooth} days on date d
- *SmoothedDailyFuel(j,d)* = average daily fuel consumption over past N_{smooth} days on date d
- *SmoothedDailyLoad(j,d)* = average number of loads over past N_{smooth} days on date d
- *DailyFuelRate(j,d) = SmoothedDailyFuel(j,d) / SmoothedDailyMeter(j,d)*
- *DailyLoadRate(j,d) = SmoothedDailyLoad(j,d) / SmoothedDailyMeter(j,d)*
- *HighFuelRateCount(j,d)* = accumulated count of days in which the daily fuel rate > N_{fuel} over days up to and including date d
- *HighLoadRateCount(j,d)* = accumulated count of days in which the daily load rate > N_{load} over days up to and including date d

Before doing the regression task, we perform a classification task to estimate the probability of having the component failure within next M runtime hours from each date. This estimated failure probability can be used as a key predictor variable in the later regression task. We observed that this failure probability improved fitting to the designed daily hazard in the regression task.

For the classification task, we now explain how to compute features and assign labels to model the predicted failure probability.

Features for the classification task:

- *HighFuelRateCountPerMeter(j,d) = HighFuelRateCount(j,d) / Meter(j,d)*
- *HighLoadRateCountPerMeter(j,d) = HighLoadRateCount(j,d) / Meter(j,d)*
- *TotalFuelRate(j,d) = Fuel(j,d) / Meter(j,d)*
- *TotalLoadRate(j,d) = Load(j,d) / Meter(j,d)*
- *TotalEventRate(j,d) = EventCount(j,d) / Meter(j,d)*

Label assignment for the classification task:
We assign the classification label *L(j,d)* to each point *x(j,d)* that corresponds to date *d* for component *j*. Note that *x(j,d)* is a multi-dimensional vector of classification features. Among all historical data of component replacements, there are two types of replacement on the final date of observation: scheduled replacement and in-field failure replacement. The goal of the classification task is to estimate the failure probability within the next *M* runtime hours from each date *d*. With binary classification labels of *Failure* and *No Failure* classes,

- For a point *x(j,d)* on a failure-replaced component *j*, when *Meter(j, d)* is within *M* meter hours of the failure replacement (that is, *Meter(j,d) > Meter(j, T_F(j)) - M*), classification label *L(j,d)* is assigned *Failure* class. Otherwise, classification label *L(j,d)* is assigned *No Failure* class.
- For any point *x(j,d)* on a schedule-replaced component *j*, classification label *L(j,d)* is assigned *No Failure* class.
- For any point *x(j,d)* on running component *j*, classification label *L(j,d)* is assigned *No Failure* class.

To measure the performance of our model, we propose and use leave-one-component-out cross validation. That is, for each run corresponding to a component *j* (= 1,…, *J*), we split the overall dataset into the *test dataset* of all points from component *j* and the *training dataset* of all points from all *J-1* remaining components *k* (≠ *j*), build a vital sign indicator model based on the training dataset only and compute the vital sign indicator values on all points in the test dataset. In more detail, we have *J* runs in total, and in each run corresponding to a component *j* we perform the steps below.

Initial Parameters: α and β (designing daily hazards), N_{smooth}, N_{fuel}, N_{load} (computing features), *M* (modeling failure probability)

Step 1. Divide the overall dataset into the test dataset of all points from one component *j* and the training dataset of all points from remaining components.

Step 2. Using only the *training dataset*, perform the classification to build a binary classifier (e.g., applying Support Vector Classification (Cristianini & Shawe-Taylor, 2000)) to compute the failure probability $P_{failure}(j, d)$ (= probability of being *Failure* class) on each point. This estimated probability can be viewed as the failure probability within the next *M* runtime hours from date *d*.

Step 3. Design the target variable for the regression task. This regression target variable $\tilde{h}_k(d)$ for any component *k* (≠ *j*) in the training dataset should have the desired characteristic of the daily hazard such as being monotonically increasing, convex-shaped, and the maximum value on failure.

Step 4. Using only the *training dataset*, build the regression model (e.g., applying Support Vector Regression (Scholkopf & Smola, 2002) to target daily hazard $\tilde{h}_k(d)$ with feature variables such as *Meter(k,d)*, *Fuel(k,d)*, *Load(k, d)*, *EventCount(k,d)* and $P_{failure}(j, d)$.

Step 5. Apply the built regression model to obtain the estimated daily hazard $h_j(d)$ for each point *x(j,d)* on component *j* in the *testing dataset*.

Step 6. Compute the individualized cumulative hazard on component *j*, $H_j(t) = \sum_{\text{all } d \text{ in } \{d:Meter(j,d)\le t\}} h_j(d)$.

Step 7. Compute the individualized cumulative failure probability on component *j*, $F_j(t) = 1 - \exp(-H_j(t))$.

After all *J* runs in leave-one-component-out cross validations, we can obtain the vital sign indicator values over all components. Given these values, we perform an optimization task to obtain the optimal threshold value for the replacement policy in terms of the economic optimization criterion such as the average maintenance cost per unit runtime. Note that in a threshold-based replacement policy, a component should be replaced when the vital sign indicator value reaches a threshold value. Optionally, we may use this estimated optimal threshold value to normalize the vital sign indicator. Then, a component should be replaced when its vital sign is 100% of wear.

In general, the parameter selections (α, β, N_{smooth}, N_{fuel}, N_{load}, *M*) influence the ultimate model. Thus, we need to find the optimal parameters to obtain the best vital sign indicator model in terms of our optimization criterion.

5. CASE STUDY

Our proposed framework of building the vital sign indicator and optimizing the economical profit was tested with one of the largest mining service companies in the world. The collected data includes the logs of daily fuel consumption, daily number of loads moved, daily meter hours, sensor event data, and component replacement history on 50 mining haul trucks over the period from January 1st 2007 to November 11th 2012. Each truck is equipped with a set of sensors triggering events on a variety of vital machine conditions. Note that the estimated overall cost of downtime for one of these haul trucks amounts to about 1.5 million USD per day. Therefore, the financial impact of reducing the downtime is very large. This is because not only is the scheduled maintenance cost high, the total cost due to unscheduled in-field failure is even higher. When one piece of equipment breaks down, in addition to stopping its own production, it may block other equipment from producing. The goal of our vital sign indicator is to optimize the tradeoff between scheduled replacement cost and unscheduled failure cost, to achieve a lower total maintenance cost of the enterprise.

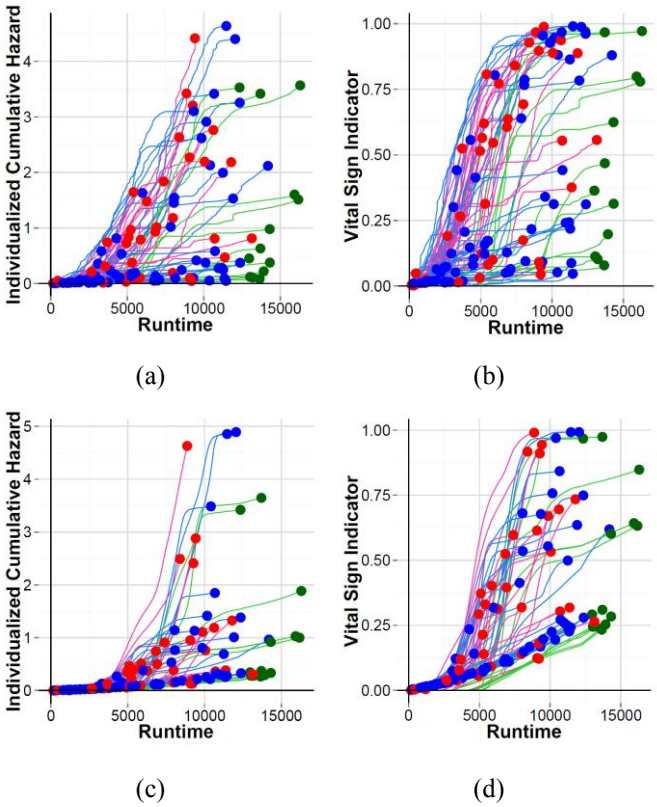


Figure 7. The individualized cumulative hazard and the vital sign indicator, (a) and (b) from SVC+SVR model, (c) and (d) from SVC+Cox model. Red = Failure replacements, Green = Scheduled replacements, and Blue = Running at the time of data collection

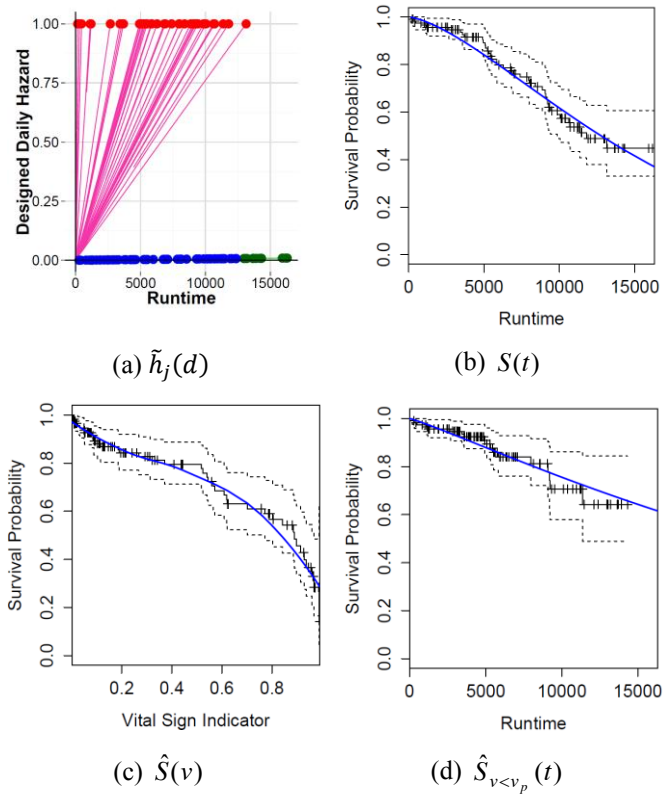


Figure 8. (a) Designed daily hazard ($\alpha = 1, \beta = 0.1$), (b) Survival probability in runtime (KM, Weibull), (c) Survival probability in vital sign indicator (KM, loess), (d) Survival probability for $Comp[v < v_p]$ (KM, loess)

In this section we present our application and results focused on one specific component type (called "X1"). To use our framework explained in the steps above, we should choose a pair of classification and regression algorithms. In general we can apply any algorithms for this purpose, but here we mainly present our results using Support Vector Classification (SVC) and Support Vector Regression (SVR). We found out that these algorithms using kernel tricks worked better than other basic algorithms including linear/quadratic discriminant analysis, generalized linear models and Cox PH regression. Also, we compared vital sign indicator models obtained using different parameter settings of α, β (designing daily hazards), N_{smooth}, N_{fuel}, N_{load} (computing features) and M (modeling failure probability) in terms of our optimization criterion. Here we show the result with the RBF kernel and the best setting of $\alpha = 1, \beta = 0.1$, $N_{smooth} = 60$, $N_{fuel} = 190$, $N_{load} = 3.0$, $M = 4890$ in our application.

Table 1. Comparison between the traditional runtime-based policy and the vital sign indicator-based policy

	Runtime-based policy	Vital sign-based policy
Threshold	$t_p = 16500$	$v_p = 0.50$
Total failure probability	$F(t_p) = 0.63$	$\hat{F}(v_p) = 0.21$
Expected time to scheduled replacement	$t_p = 16500$	$\hat{t}_p = 14848$
Expected time to failure replacement	$t_f = 8708$	$\hat{t}_f = 7311$
Avg runtime per cycle	11592	13201
Avg failure replacement cost per unit runtime	\$30.6	\$9.3
Avg scheduled replacement cost per unit runtime	\$15.6	\$27.9
Avg maintenance cost per unit runtime	\$45.6	\$37.2

The economic and logistic parameters for the target component type are as follows: C_f = failure replacement cost = \$443600, C_p = scheduled replacement cost = \$374400, c_d = cost per unit downtime of the equipment = \$2000, DT_f= down time due to an in-field failure = 64.8 hrs, DT_p = down time due to a scheduled replacement = 48 hrs. Note that $(C_f + c_d DT_f)/(C_p + c_d DT_p) = 1.22$.

Figure 7(a) and (b) show the individualized cumulative hazard and the vital sign indicator, respectively, for the model based on SVC and SVR. In the figures, each line corresponds to a component. The color of the line and corresponding end point indicates whether the component had a failure replacement at the end (red), were running at the time of data collection (blue, right-censored) or had a scheduled replacement at the end (green, right-censored).

Figure 8(a) shows the designed daily hazard. The optimized vital sign threshold was 0.50. Based on two different approaches explained to estimate $\hat{t}_p$ and $\hat{t}_f$, we obtained almost similar values of the criterion (\$37.1 and \$37.2). Figure 8(b),(c) and (d) show survival probabilities such as $S(t)$, $\hat{S}(v)$ and $\hat{S}_{v<v_p}(t)$. Considering that the cumulative failure probability corresponds to 1 − survival probability (that is, $F(t) = 1 - S(t)$, $\hat{F}(v) = 1 - \hat{S}(v)$), note that $F(t_p) = 0.63 > \hat{F}(v_p) = 0.21$. This significant reduction in total expected failure probability is a necessary condition for being a good vital sign indicator. Also, comparing $\hat{S}_{v<v_p}(t)$ and $S(t)$ in Figure 8(b) and (d), we find that the expected lifetime of $Comp[v < v_p]$ alone is significantly longer than that of all components in the dataset.

Table 1 compares the runtime-based and vital-sign based replacement policy in terms of the average maintenance cost per unit runtime. There is about 20% cost reduction with the vital-sign based policy, compared to the runtime-based policy. The new vital-sign based policy with vital sign threshold = 0.5 has some false failure predictions so involves higher average scheduled replacement cost per unit runtime than the runtime-based policy (\$27.9 > \$15.6), but the vital-sign based policy has significantly smaller average failure replacement cost per unit runtime (\$9.3 << \$30.6) and thus, overall it is better than the runtime-based policy.

We tested Cox PH regression in combination with SVC in our framework. In fact we compared several Cox PH regression models using differently selected features as time-dependent covariates. Then, we observed that the Cox PH regression simply using the SVC-estimated failure probability as the only one time-dependent covariate worked best among them. Figure 7(c) and (d) show the individualized cumulative hazard and the vital sign indicator from this model. But, this still performed a bit worse (\$38.0) than the SVR-based model (\$37.2). Note that while Cox PH regression considers only the covariate values at sampled failure times (i.e., maximizing the partial likelihood), SVR can consider covariate values at all times (i.e., maximizing the fit to the complete paths of the designed target daily hazards).

6. CONCLUSION AND DISCUSSION

We compared our vital sign indicator-based policy with a traditional runtime-based policy in terms of the average maintenance cost per unit runtime. When the failure replacement cost of a component is extremely high, it is critical to reduce the total number of in-field failures by following the recommended option for decreasing the total expected probability of failures. We modeled our vital sign indicator based on "individualized" cumulative failure probability function for each component. This new indicator as a transformed time scale allows us to have an individualized maintenance plan for each component based on its real usage. Our case study demonstrates that the new vital sign indicator-based replacement policy can obtain greater economic value in terms of the average maintenance cost per unit runtime.

Future work will include a remaining useful lifetime (RUL) model based on this vital-sign indicator. This will involve the estimation of paths in the runtime vs. vital sign indicator 2-dimensional plot. Another future direction is to incorporate a constrained regression to make vital sign indicators suitably convex-shaped, eventually leading to lower optimal costs.

REFERENCES

Bair, E., Hastie, T., Paul, D., & Tibshirani, R. (2006). Prediction by supervised principal components. *Journal of the American Statistical Association*, *101*(473).

Banjevic, D., Jardine, A., Makis, V., & Ennis, M. (2001). A control-limit policy and software for condition-based maintenance optimization. *INFOR-OTTAWA-*, *39*(1), 32–50.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Fox, J. (2002). Cox proportional-hazards regression for survival data. *See Also*.

Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, *27*(2), 83–85.

Jardine, A., Banjevic, D., Montgomery, N., & Pak, A. (2008). Repairable system reliability: Recent developments in CBM optimization. *International Journal of Performability Engineering*, *4*(3), 205.

Jardine, A. K., & Tsang, A. H. (2013). *Maintenance, replacement, and reliability: theory and applications*. CRC press.

Scholkopf, B., & Smola, A. (2002). *Learning with kernels*. MIT press Cambridge.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Therneau, T. M. (2000). *Modeling survival data: extending the Cox model*. Springer.

Wu, X., & Ryan, S. M. (2011). Optimal replacement in the proportional hazards model with semi-markovian covariate process and continuous monitoring. *Reliability, IEEE Transactions on*, *60*(3), 580–589.

Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*, *42*(1-2), 31–60.