This is an author produced version of a paper published in "Journal of Quality in Maintenance Engineering". This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:

Al-Najjar, Basim

"Economic criteria to select a cost-effective maintenance policy"

Journal of Quality in Maintenance Engineering, 1999, Vol. 5,

Issue: 3, pp 236-248

URL: <http://dx.doi.org/10.1108/13552519910282692>

Access to the published version may require subscription.

Economic criteria to select a cost-effective maintenance policy

Basim Al-Najjar

Lund University and Växjö University, Sweden

e-mil: Basim.Al-Najjar@masda.hv.se

Abstract

The reputation of an organisation is often built through hard work on improving quality, reliability, delivery time and price. In this paper a graphical method for the selection of a cost-effective monitoring technique is suggested. This graphical method is also used to select the most cost-effective replacement vibration level, when a vibration-based maintenance is implemented, i.e. when the available data are mainly condition-based replacement. This method is based on the concept of the Total Time on Test, TTT-plot. The use of this method is explained by three examples.

Keywords: Maintenance costs, Vibration-based maintenance, Cost-effectiveness, Generalised Total Time in Test- plots, Age replacement policy.

Introduction

A competitive product or service is usually based on a balance between productivity, quality and production cost. The analysis of maintenance and quality-related events and their costs for example provides a method of assessing the overall effectiveness of the system and of determining problem areas, opportunities, savings and action priorities.

Maintenance policies can be characterised by costs, number of stoppages, time between replacements, availability, replacement levels when using vibration-based monitoring systems, (VBMS). Maintenance policy may be considered new when some of its characteristic factors are changed.

In this paper we address the problem of selecting a cost-effective monitoring technique based on observational data collected from different monitoring parameters, where units are operated until failures or condition-based replacements occurred. Here, we suppose that the condition of a component can be assessed based on the current value of a monitored parameter, e.g. the vibration level or wear rate.

In (Bergman, 1977), only one monitoring technique was considered. Here, we are mainly concerned with the comparison of many condition monitoring (CM) techniques. We will first give a brief description of Total Time on Test, TTT-plots and how to use them for the determination of age. We suggested the use of a generalisation of TTT-plots proposed by (Bergman, 1977) for the selection of a cost-effective CM technique and the most cost-effective replacement vibration level when the available data are mainly condition-based replacements. In this paper we mean by each bearing a rolling element bearing.

Total time on test, TTT-plot

Suppose that we are given n observations $t_1, \dots, t_n$ from a particular life distribution $F(\cdot)$. Let these observations be ordered due to their sizes, i.e. $t_1 \leq \dots \leq t_n$. Let T_i denote the total time generated in ages less or equal to t_i , i.e. $T_1 = nt_1$, and generally

$$T_i = \sum_{j=1}^i (n-j+1) (t_j - t_{j-1})$$

for $i=1, \dots, n$ where $t_0=0$. Also, let $u_i = T_i/T_n$.

Example 1

Assume that 8 components are observed until failures, their times to failure, $\zeta_1, \dots, \zeta_8$, and the calculated quantities are given in Table 1. The TTT-plot is obtained by plotting u_i versus i/n , see Fig.1.

i	ζ_i	T_i	u_i	I/n
1	8.7	69.6	0.267	0.125
2	11.6	89.9	0.345	0.25
3	21.3	148.1	0.568	0.375
4	26.1	172.1	0.66	0.5
5	37.4	217.3	0.834	0.625
6	38.2	219.7	0.843	0.75
7	49.8	242.9	0.93	0.875
8	67.5	260.6	1.0	1.0

Table 1. Failure times and calculated quantities of TTT-plots.

The TTT-plot gives a dimensionless view representing times to failure of the tested components. The deviation of the plot from the diagonal provides information about the deviation of the plotted data from the exponential distribution. The plot is applicable in detecting whether the failure rate function is increasing or decreasing. The TTT-plot technique is not developed here, for more details the reader is referred to (Bergman, 1977), and (Klefsjö, 1986).

Fig.1. TTT-plots.

Maintenance cost

Consider three maintenance policies: Breakdown maintenance (BDM), age-based maintenance (ABM) and condition-based maintenance (CBM). Suppose that these policies are used to maintain a component or equipment. Measurements, analysis, diagnosis, repairs are assumed to be performed by internal resources. The required assets, spare parts and experts are all provided internally.

Here, the maintenance cost is broken down to its basic elements. Denote by c_1 the cost incurred by a planned action, e.g. adjusting, repair or replacement of the component, independent of which maintenance policy is involved. The costs included by c_1 , when the machine is running 24 hours daily, may be classified as:

1. Spare parts (c_{1S}) such as a bearing, lubricant and equipment.
2. Man-hour (c_{1M}), e.g. all costs incurred by repair, adjustment, cleaning and lubricant change.
3. Production losses during maintenance time (c_{1P}). Assume that the problem is identified and localised by CM system, experts or is planned in advance by ABM.

Denote by c_2 an additional cost, which is suffered only at failure and is also independent of which policy is involved. It is considered to include the costs of:

1. Consequential damage (c_{2C}) to other parts in the machine.
2. Additional production losses during the additional downtime (c_{2P}):
 - Times to localise the fault (c_{2L}) and to select the repair team (c_{2S}).
 - Wait time for equipment and spare parts arrival (c_{2W}).
 - Extra time to repair consequential damage (c_{2R}).
3. Production losses because of bad quality associated with failures (c_{2Q}).
4. Insurance premiums against the occurrence of failure-based accidents (c_{2I}).

5. Environmental damage (c_{2E}) e.g. pollution of air, water and earth and high noise level.
6. Delivery delays (c_{2D}).
7. Company interest losses due to reduction of the market share (c_{2M}).
8. Expenses of investing capital in unnecessary redundancies in spare parts, equipment, and personnel (c_{2X}) to avoid long waiting times.

Denote by S_i the capital invested to use the i th policy, $i = \text{BDM}, \text{ABM}, \text{CBM}$. Denote by $^*S_{\text{BDM}}, ^*S_{\text{ABM}}, ^*S_{\text{CBM}}$ the long run average implementing costs per unit time. S_{CBM} includes the costs of: measuring and analysis equipment, personnel salaries, training in how to interpret signals and diagnose component deterioration, charge for office and workshop for maintenance staff, administrative and miscellaneous expenses.

S_{ABM} is the sum of salaries of maintenance staff, local charges, expenses for tools, administration, staff training and miscellaneous. The capital invested to apply BDM (S_{BDM}) is considered equal to zero, because no action is taken until machine failure.

Denote by $C_i(t)$ the total expected maintenance cost per unit time when applying i th maintenance policy during $(0,t)$. Denote by $E[N_i(t)]$ the expected number of removals when i th policy is used, i.e. expected number of planned and unplanned replacements during $(0,t)$. Then, $C_{\text{BDM}}(t)$, is:

$$C_{\text{BDM}}(t) = \frac{(c_1 + c_2) \cdot E[N_{\text{BDM}}(t)]}{t} \quad (1)$$

and $C_i(t)$, may be written as:

$$C_i(t) = \frac{\{c_1 \cdot E[N_i(t)]_{\text{planned}} + (c_1 + c_2) \cdot E[N_i(t)]_{\text{failure}} + S_i\}}{t} \quad (2)$$

where C_i can be written as the average cost of one cycle divided by average cycle length (Bergman, 1977). Thus, the cost equations may be written as:

$$C_{BDM} = \frac{(c_1 + c_2)}{\mu} \quad (3)$$

$$C_{ABM} = \frac{[c_1 + c_2 P(\zeta \leq T)]}{E[\min(\zeta, T)]} + S_{ABM} \quad (4)$$

$$C_{CBM} = \frac{[c_1 + c_2 P(\zeta \leq T_x)]}{E[\min(\zeta, T_x)]} + S_{CBM} \quad (5)$$

where

$$c_1 = c_{1S} + c_{1M} + c_{1P} \quad (6)$$

$$c_2 = c_{2C} + c_{2P} + c_{2Q} + c_{2I} + c_{2E} + c_{2D} + c_{2M} + c_{2X} \quad (7)$$

T: Time to planned replacement.

ζ : Time to failure, variable.

T_x : Time to replacement defined on the condition assessed by monitoring parameter value, $x(t)$, i.e. the instant when $x(t)$ first reaches a predetermined level, where T_x is such that for $t \geq 0$, the event $\{T_x \geq t\}$ is determined by $\{x(s), 0 \leq s \leq t\}$ independent of $\{\zeta, x(s), s \geq t\}$ because the replacement would be performed only when it is necessary.

$$\mu = E(\zeta).$$

Partial local optimisation of (4) and (5) can be achieved through optimising only the first term of the equation by using, for example for (5), the iterative method suggested by (Bergman, 1978). The rule for making an optimal choice between these three strategies BDM, ABM and CBM is: Select the policy which yields the least C_i .

Implementing CM techniques for detecting the machine condition effectively yields that the number of failures is almost zero because the component is almost always replaced before a failure. Thus, it is important to realise that there exist some extra costs due to the occurrence

of more failures when using ABM. These extra costs (C_{extra}) are not easily observable in the cost equations above and can be summarised by:

1. Extra capital investment in spare parts and equipment to reduce waiting time.
2. Extra costs for larger store, more personnel and larger floor space for 1 above.
3. Higher insurance premiums.
4. Losses due to loss of company reputation and market share.
5. ABM leads in many cases to reduction of component life and increase in the number of stoppages.
6. Extra expenses for failure-based environmental damage.
7. Extra production losses due to bad quality associated with extra failures.

Thus, at the selection of a cost-effective maintenance policy these costs should be considered if c_2 is considered equal for the compared policies. The role is then: Select the policy which yields the least C_{itotal} ,

$$C_{\text{itotal}} = C_i + C_{\text{extra}}$$

Age Replacement

Consider the failure times $\zeta_1, \dots, \zeta_n$, which are of unknown distribution. Then, the empirical distribution function (F_n) may be defined as

$$F_n(t) = [1/n] * [\text{number of } \zeta_i \text{ such that } \zeta_i \leq t],$$

Where, n represents the total observed failures. In order to estimate C , we replace $P(\zeta \leq T)$ by its estimator $F_n(t)$. Then ${}^*C_{\text{ABM}}$ is

$${}^*C_{\text{ABM}} = \frac{c_1 + c_2 \left(\frac{i}{n} \right)}{\frac{1}{n} T_n u_i} + {}^*S_{\text{ABM}} \quad (8)$$

The optimum age to replacement is that which minimises (8). It is proved by (Ingram and Scheaffer, 1976), that the time interval minimising (8) may be found among $\zeta_1, \dots, \zeta_n$. Thus, to estimate the optimal age replacement interval it is enough to find the index i_0 minimising $^*C_{ABM}$. In case i_0 is equal to n , the replacement occurs at failure.

The index i_0 may also be estimated graphically from the TTT-plot. To determine i_0 , draw the line through $(-\frac{c_1}{c_2}, 0)$ which touches the plot and has the largest slope. If this line passes through $(i_0/n, T_{i_0}/T_n)$ our estimator is the optimum replacement interval equals t_{i_0} , see Fig.1. The estimated replacement interval is close to the optimum if the number of observations is large enough (Bergman, 1977).

A generalised TTT-plots to compare maintenance policies

CM parameter value is not always increasing in operating time. Shock Pulse Measurements (SPM) may decrease when contact areas in a bearing become smoother by rubbing action, see Fig.2.

Figure 2. Typical parameter values in time shows $S_j(x)$ of these components.

Define

$$x_j = \sup_{t \leq \zeta_j} x_j(t)$$

That is, x_j is the largest value of the monitoring parameter observed during monitoring time t , where $t \leq \zeta_j$, $i=1, \dots, n$. Now, let us order the indices so that $x_{(i)}$ corresponds to that x_j which is the i th in size. Then, the ordered indices are $x_{(1)} \leq \dots \leq x_{(n)}$. Define $S_j(x)$ as the total time on test generated by the j th component before its parameter value exceeds the level $x_{(i)}$ for the first time, i.e.

$$S_j(x) = \inf \{t; x_j(t) \geq x\},$$

let

$$S(x) = \sum_{j=1}^n S_j(x)$$

Now, define $T_i = S(x_{(i)})$

which is the total time accumulated by all components while their parameter values are less or equal to $x_{(i)}$. TTT-plots can be obtained by the same way, illustrated in Example 1. The ratio i/n is actually an estimate of the probability that the failure occurs before the parameter value has exceeded $x_{(i)}$, if no planned replacement is done. While T_i/n is an estimate of the expected time to replacement, if no failure has occurred. Thus, the estimate of C_{CBM} when policy π is used is

$$C_{(CBM) \pi} = \frac{c_1 + c_2 \left(\frac{i_{o\pi}}{n} \right)}{\frac{1}{n} T_{i_{o\pi}}} + {}^*S_{(ABM) \pi} \quad (9)$$

By analogy, $i_{o,\pi}$ minimising $C_{(CBM) \pi}$ can be estimated graphically.

To be conservative, one could define x_j as the level of the monitoring parameter at failure. Also, define $S_j(x_{(i)})$ as the total time on test generated by the j th component before its parameter value exceeds the level of the i th failure, $x_{(i)}$. Also, we assume that the CM

equipment does not fail and have no effect on the condition of the components. Now, denote by n_i the number of failures occurring before the parameter value exceeds the level $x_{(i)}$. The use of the generalised, GTTT-plots, for the k policies is illustrated in example 2.

Example 2

Assume that eight components are monitored by means of four monitoring parameters until respective failure times $\zeta_1, \dots, \zeta_8$. Let the techniques used to monitor these components be the age replacement policy and CBM using CM techniques π , θ and ϕ .

The CM technique ϕ is supposed to use a new parameter whose relation with the deterioration process under consideration is not well understood. Note that the number of components used in this example may not be enough to estimate C with high precision.

Select six levels, $x_{(0)}, \dots, x_{(5)}$, for each CM parameter, so that the first level is arbitrary while the other five levels represent the parameter values at failures of the components.

Figure 3. GTTT-plots of age, π , θ and ϕ policies.

Graphically, the optimum i_o/n for age replacement policy and CBM policies π and θ , is the same and equal to $1/8$ and the corresponding values of T_{i_o}/T_n are 0.4, 0.9 and 0.95, respectively, see Fig.3.

Assume that $S_\pi=1.22$, $S_\theta= 2.4$ and $S_{ABM}= 0.4$ SEK/time unit, also assume that $c_1=1000$ and $c_2=10000$ SEK. Then, by applying (9), C for these policies are 2.1, 3.20 and 2.4 SEK/time unit, respectively. Trivially, the cost-effective policy is that which uses CM technique π .

It is seen that the failure rate increases dramatically when $x_\pi(t)$ or $x_\theta(t)$ increases. When n is large, it is easy to verify that the failure rate is approximately equal to zero before $x_\pi(t)$ or $x_\theta(t)$ exceeds the level $x_{(0)}$. Both CM techniques π and θ have large explanatory power while the technique ϕ does not. The age based failure rate increases in the component age.

On the other hand, if the information supplied by a monitoring parameter is non- or weakly correlated with the deterioration process under consideration, then its TTT-plot fluctuates about the diagonal. This means that the failure rate when using such a technique ϕ should be approximately constant. Thus, the TTT-plot corresponding to a technique should reveal a very weak relation between the monitored parameter and actual component condition, see Fig.3.

Selection of a cost-effective maintenance policy

In general, defects can not be limited always to only one part of a rolling element bearing such as inner or outer race after a period of damage initiation. If the damage is started at one part, e.g. at the inner race it may spread gradually to the other parts. Thus, the evaluation of the bearing condition would not be reliable if it were based on monitoring the defect vibration frequency of only one part of the bearing.

The root mean square (RMS) of the essential frequencies generated by bearing defects is suggested as a measure of the bearing condition, call it Bearing Defect Vibration Energy (BDVE). The essential frequencies may be defined as: Bearing defect frequencies and their

multiples which exceeded a predetermined level which can be assessed based on the machine vibration history.

In this paper, the condition of a bearing is evaluated from the current value of BDVE. In this application, GTTT-plots are used to select the most cost-effectiveness vibration replacement level when planned replacement data are mostly available.

In vibration-based maintenance, the replacement, in general, is performed as soon as a predetermined level is exceeded. Thus, according to the traditional failure definition not enough failure data are available at the industries implementing VBMS. The residual of the operating time of a component can be estimated by, e.g. using the graphical method suggested in (Al-Najjar, 1996I). This means that with better data coverage and quality it is possible to make use of as much as possible of a component mean effective life.

A condition-based renewal time is almost a failure time. It is not a censoring in the usual random sense. It is a failure time with a bit missing (Sherwin, 1995) and (Bergman and Klävsjö, 1995). In order to use GTTT-plots when VBMS is applied we assume that the replacements are performed just before failures, e.g. when Total Quality Maintenance, TQMain, is used (Al-Najjar, 1996II).

In this application, the Total Time on Test may be understood as Total Time in Operation because we are using CBM data instead of failure or test data. Suppose that the vibration levels of n identical components are mounted at different locations, but with same operating conditions, in the machine which are monitored until respective replacement times or at unplanned but before failure replacement (UPBFR) τ_i , $i = 1, \dots, n$. UPBFR are performed at unplanned but before failure stoppages to prevent the occurrence of failures (Al-Najjar, 1997).

The TTT-plots can be obtained in the usual way. The ratio i/n is an estimate of the probability that the planned replacement occurs before the parameter value has exceeded $x_{(i)}$, if no UPBFR is done. While T_i/n is an estimate of the expected time to a planned replacement,

if no UPBFR is done. Thus, C when policy π is used can be estimated by (9). The index $i_0 \pi$ minimising $C \pi$ can be estimated graphically as well.

We consider a specific type of rolling element bearings can be used in many locations of a paper machine. Assume that the machine is only monitored by vibration and the replacements are performed at a predetermined level (x_p). From everyday experience, the probability that these replacements occur precisely at that level is very low. The replacement at a level higher than x_p may happen due to faster increase in the vibration level than anticipated. When the next planned stoppage is not close enough the replacement is performed at a level lower than x_p to avoid failures.

Sometimes, components may be replaced at a level higher than x_p without exposing operating safety, machine function, productivity and product quality to a real risk. Elongation of the life length of a component is important to reduce stoppages and production losses.

Denote by n_i the number of replacements done before the parameter value exceeds the level $x_{(i)}$. The use of GTTT-plots, for k different replacement vibration levels is illustrated by the following example:

Example 3

The data used in this example are not real but reasonable and based on the author's practical experience within paper mill industry. Consider 8 identical replaced rolling element bearings in the database of a paper machine. The vibration was measured at the bearing house once per week. The vibration measurements, i.e. trend of BDVE, historical comments, mounting and replacement times are assumed to be available at the database. Assume also that two vibration-based maintenance policies using the same VBMS need to be compared to identify the most cost-effective replacement policy when two different predetermined vibration levels are possible.

Fig.4. The plots of the vibration levels of the 8 bearings in Example 3.

Fig.5. The GTTT-plots of the replacement policies π_1 and π_2 in Example 3.

Assume that these bearings have been replaced at five different vibration levels which are $X_1 = 2.3$, $X_2 = 2.5$, $X_3 = 3.2$, $X_4 = 3.7$ and $X_5 = 4.5$ mm/s. Let the predetermined level be $x_{p1} = 2.5$ mm/s when policy π_1 is adopted, see Fig. 4.

Now, assume that these bearings are not replaced at the above-mentioned levels but according to another replacement policy, π_2 . This means that another order of levels with a

new predetermined level $x_{p2} = 3.7$ mm/s is used, so that the maximum allowable level should not exceed 4.5 mm/s. The new replacement levels are then $X_{1'} = 3.7$, $X_{2'} = 4.0$, $X_{3'} = 4.3$, $X_{4'} = 4.5$ mm/s. According to policy π_1 bearings number 2, and {4 & 6} and {1 & 3} and 5, and {7 & 8} are replaced at the levels 2.3, 2.5, 3.2, 3.7, and 4.5 mm/s, respectively.

The second group of levels, i.e. when policy π_2 is adopted, is achieved through extrapolating the levels of the first group, dash lines, and for easiness we assumed that the vibration increment is linear in time, see Fig.4. Thus, the bearing number 6', 4', {1', 2', 3'} and {5', 7', 8'} are replaced at the levels 3.7, 4.0, 4.3 and 4.5 mm/s, respectively, see the same figure. Linear extrapolation needs more justification than just convenience. The bearings are assumed to be mounted in the machine at the same time. GTTT-plots are given in Fig.5. *S_1 and *S_2 are the invested capital per unit time for using policy π_1 and π_2 respectively. Let $^*S_{\pi1} = 5.8$ SEK /hour, $^*S_{\pi2} = 11.6$ SEK /hour, $c_1 = 5000$ SEK /hour, and $c_2 = 50\,000$ SEK /hour. Then, $C_{\pi1} = 679.8$ and $C_{\pi2} = 590.6$ SEK/hour. This results in a saving equal to 89.3 SEK/hour or about 783 000 SEK /year for only increasing the predetermined level from 2.5 to 3.7 mm for 8 bearings, i.e. about 50%.

The most cost-effective policy is that which has the tangent line with the larger angle, see Fig.5. The plot in Fig.5 is started from the origin in order to cover the cases when T_i is approximately equal to zero, e.g. when the i th bearing, of n identical bearings, is replaced after very short operating time because of wrong installation.

When examining the cost equation expressed in (8) and (9) reveals that the saving increases when c_2 , c_1 increase. Thus, the economic importance of implementing VBMS increases with the invested capital per machine and the failure consequential costs.

Cost-Effectiveness

The cost-effectiveness (C_e) of each maintenance improvement may be examined by using the proportion of the difference between $(C)_b$ before and that after the improvement $(C)_a$, to the $(C)_b$, i.e.

$$C_e = 1 - \frac{(C)_a}{(C)_b}$$

At the beginning $C_e \geq 0$, i.e. $(C)_b \geq (C)_a$, due to extra expenses because of the learning period. But, beyond this period C_e should be greater than zero, i.e. $(C)_b > (C)_a$ in order to consider the improvement as a cost-effective action. Thus, C_e can be considered as a measure of the cost-effectiveness of maintenance improvements.

Conclusions

When using GTTT-plots, we can determine the optimum replacement interval and distinguish the cost-effective maintenance policy when some policies are applicable. Besides, it might be used as an indicator to discover weak-correlated or non-correlated monitoring parameters. In Example 2 we had a very clear type of relation between failure and CM measurements.

The cost of using CM techniques is reducing which makes applying these techniques appreciably cheaper than 15 years ago. Using a continuous vibration monitoring system reduces man-hour cost and increases the precision of assessing the machine condition.

Using GTTT-plots provides the possibility to assess the probability of performing a planned replacement before an UPBFR becomes evident during planning to the next replacement. This probability together with the absolute value of the monitored parameter and its trend increases the probability to avoid failures or UPBFRs.

An accurate selection of a cost-effective maintenance policy should be based on better data and accurately assessed c_1 , c_2 and S_i . C_e provides possibilities to measure and monitor the

economic progress after each development in order to define the cost-effectiveness of these developments.

References

- Al-Najjar, B. 1996. On the effectiveness of vibration-based programs. Report 9581, ISSN 1400-1942, ISRN HV/MASDA/SE/R/--9581--SE, Växjö, Sweden, April 1996I.
- Al-Najjar, B. Total Quality Maintenance: An approach for continuous reduction in costs of quality products. *Journal of Quality in Maintenance Engineering*, 2-20, Vol 2, Number 3, 1996II.
- Al-Najjar, B. Condition-based maintenance: Selection and improvement of a cost-effective vibration-based policy in rolling element bearings. Doctoral thesis, ISSN 0280-722X, ISRN LUTMDN/TMIO—1006—SE, ISBN 91-628-2545-X, Lund University, Inst. of Industrial Engineering, Sweden, 1997.
- Bergman, B. (1977). Some graphical methods for maintenance planning. *Annual Reliability and Maintainability Symposium*.
- Bergman, B. 1978. Optimal replacement under a general failure model. *Adv. Appl. Prob.* 10, 431-451.
- Bergman, B. and Klefsjö, B. 1995. *Quality from customer needs to customer satisfaction*. Studentlitteratur, Lund, Sweden.
- Ingram, C.R., and Scheaffer, R.L. (1976). On consistent estimation of age replacement intervals. *Technometrics*, 18, 213-219.
- Klefsjö, B. (1986). TTT-Transform- A useful tool when analysing different reliability problems. *Reliability Engineering*, 15, 231-241.
- Sherwin, D. J. 1995. Data analysis for practical reliability and maintenance studies: The need for a comprehensive methodology.