General Repair Models for Maintained Systems

A.Syamsundar ¹, K. Muralidharan ² and V.N.A. Naikan ³

¹Visakhapatnam Steel Plant, Visakhapatnam, ²Department of Statistics, Faculty of Science, Maharajah Sayajirao University of Baroda, ³Department of Industrial Engineering and Management, Indian Institute of Technology, Kharagpur, India

ABSTRACT

Any system when put on operation deteriorates over a period of time and fails on crossing a threshold. Maintenance is performed either on failure or for mitigating the effects of this deterioration. In general the system is either partially maintained or renewed to bring it back to its operating state. The general behaviour of the failure process of a maintained system is modelled by means of stochastic processes called general repair processes. In these processes the effect of the maintenance carried out on the system is captured by means of a variable called the maintenance indicator which indicates the degree to which the system has been maintained. Such a maintenance indicator acting on the system’s failure age or intensity post maintenance generalizes the modelling of the failure processes. The general repair process models based on the type of the maintenance indicator can be broadly classified into two categories based on age or intensity loss or recovery and two sub-categories based on whether these take place additively or multiplicatively. This paper focuses on such models, their inference, their uses, the physical reality they depict and their shortcomings when modelling the failure processes of maintained systems.

Key words : Industrial maintenance, Maintained system, Point process, Intensity process, General repair.

1. Introduction

Any system when put on operation deteriorates over a period of time and fails on crossing a threshold. Maintenance is performed either on failure or for mitigating the effects of this deterioration. Maintenance can consist of simple cleaning, lubrication or bolt tightening to replacement of a damaged or failed component, sub assembly, assembly or sub system to replacing the entire system. The behaviour of the failure process of a maintained system is modelled by means of stochastic processes.
When the performance of the system deteriorates or it breaks down and the whole system is replaced every time this happens, the maintenance is termed as maximal. Such a system’s failure process is modelled by a renewal process. When the system is repaired as little as possible in putting it back to functional use, thus leaving it in the same condition as it was before maintenance, the maintenance is termed as minimal. Such a system’s failure process is modelled with a non homogeneous Poisson process.

Both these are extreme cases and do not reflect the general behaviour of the failure processes of maintained systems, wherein the system is partially maintained or renewed. These failure processes are modelled with general repair processes. In these processes the effect of the repair or maintenance carried out on the system is captured by means of a variable which indicates the degree to which the system has been maintained. Such a degree of maintenance acting on the system’s failure age / intensity, post maintenance, generalises the maximal and minimal repair processes. The effect of the maintenance action taken represented by a maintenance indicator is used to distinguish the various conditional intensity processes used to describe the general failure processes.

Pham and Wang [1] is the earliest paper that depicts and provides a classification of imperfect or general maintenance models. The models are classified into eight categories, each category being based on the method of treatment of imperfect maintenance. The methods of treatment are \((p,q)\) rule, \((p(t),q(t))\) rule, improvement factor, virtual age, shock, \((\alpha,\beta)\) rule, multiple \((p,q)\) rule and others. The models are treated independently and there is no attempt to link up the various models. Guo et al. [2] combine all virtual age models into one category based on the degree of repair and index the models based on the types of their degrees of repair and corresponding virtual ages. Lugtigheid et al. [3] is the first paper that defines and classifies maintenance indicators into three categories and provides methods to incorporate these into intensity process models for repairable systems. Lindqvist [4] and Syamsundar and Naikan [5] depict various imperfect repair models, provide methods of inference and their uses but do not classify them.

The present paper depicts and classifies all general repair models based on the effectiveness of the maintenance actions taken as represented by a maintenance indicator. It studies the methods of their inference, and the physical reality that they depict when modelling the failure processes of maintained systems. The nature of the repair and maintenance indicator is presented in section 2 and the general repair
models built using this indicator along with their properties are presented in section 3. The methods of inference and application of these models to maintained systems is presented in section 4. The physical reality depicted by these models and their shortcomings are presented in section 5. The conclusion is presented in section 6.

2. **Maintenance Indicator**

A maintenance indicator indicates the system state by means of its age or failure intensity and conveys the effectiveness of the maintenance actions taken on the system. Such a maintenance indicator incorporated into the stochastic process of failures of the system will provide a tractable method of arriving at a general process model of the failures or maintenance events of a system.

2.1. **Maintenance Indicator Categories**

Maintenance indicators can be categorised in two ways based on the methods chosen to represent the system state, based on the effect of maintenance actions on the system age or the system intensity. These can be further sub-categorised on whether the effects are additive or multiplicative

2.1.1. **Maintenance Indicators Affecting System Age**

Maintenance actions on the systems modify the system’s age. Without maintenance action or with a minimal maintenance action, a system ages in line with its chronological age. Whenever a maintenance action is performed it affects the age of the system and the maintenance indicator is used to reflect the change in the system age from what it was before the maintenance action.

The age of the system after maintenance called its virtual age, given by $\nu$, is different from its chronological age. The maintenance action serves to reset the virtual age. This virtual age is a function of its chronological age and possibly the history of the failure process i.e., depending on its past failures. This is given by:

$$\nu(t | H_{i-}) = \nu(t; N_i, T_1, T_2, ..., T_{N_i})$$

When the maintenance indicator acts on the system age additively its virtual age is given by:

$$\nu(t | H_{i-}) = t - \sum_{i=t}^{N_i} S(i, T_1, T_2, ..., T_i)$$
When the maintenance indicator acts on the system age multiplicatively its virtual age is given by; 
\[ \nu(t | H_{t-1}) = t \times \sum_{i=1}^{N_{t-1}} S(i, T_1, T_2, ..., T_i) \]

### 2.1.2. Maintenance Indicators Affecting System’s Failure Intensity

Maintenance actions on the system modify the system’s failure intensity. Whenever a maintenance action is carried out on the system it affects the system failure intensity and the maintenance indicator reflects the change in the system intensity either individually or cumulatively from what it was before the maintenance actions.

The maintenance action serves to reset the failure intensity. This failure intensity is a function of its chronological age and possibly the history of the failure process i.e., depending on its past failures. This is given by;
\[ \lambda(t | H_{t-1}) = \lambda(t; N_t, T_1, T_2, ..., T_{N_t}) \]

When the maintenance indicator acts on the system’s failure intensity additively its failure intensity is given by; 
\[ \lambda(t | H_{t-1}) = \lambda(t) - \sum_{i=1}^{N_{t-1}} S(i, T_1, T_2, ..., T_i) \]

When the maintenance indicator acts on the system’s failure intensity multiplicatively its failure intensity is given by;
\[ \lambda(t | H_{t-1}) = \lambda(t) \times \sum_{i=1}^{N_{t-1}} S(i, T_1, T_2, ..., T_i) \]

### 2.2. Maintenance Indicators that Vary Over Time

The maintenance indicators have generally been chosen to be constant i.e., the effect of maintenance actions on the system is averaged out.

Maintenance indicators can also be generated to take into account the changing effects of maintenance actions on the system. Percy and Alkali [6], Kahle and Love [7], and Lugtigheid et al. [8] have proposed such type of maintenance indicators.

Percy and Alkali [6] have developed maintenance indicators which vary with age. They state that the maintenance indicators given by \( \rho \) can be taken as random variables
(stochastic indicators), specified functions of \( i \) or \( t \) i.e., deterministic indicators or combined indicators having both the stochastic as well as the deterministic components.

The maintenance indicator taken as independent exponential random variables is given by;

\[
\rho_i | \lambda \sim \text{Ex}(\lambda)
\]

These can also be gamma, lognormal or Weibull random variables. Deterministic maintenance indicators as increasing and decreasing functions of time may be given by;

\[
\rho_i \text{ or } (1 - \rho_i) = \frac{t_i}{t_i + \lambda}
\]

Other deterministic indicators can be functions with or without extended formulations. These are given by;

\[
\rho_i | \lambda, i \sim \exp(-i\lambda) \quad \text{and} \quad \rho_i | \lambda = \theta \exp(-i\lambda) + \phi, \quad \text{where} \quad \lambda = \lambda, \theta, \phi
\]

Combined indicators are given by;

\[
\rho_i | \lambda, t_i \sim \text{Ex}(\lambda_i) \quad \text{where} \quad \lambda_i = \frac{t_i}{t_i + \lambda}
\]

Kahle and Love [9] have developed a maintenance indicator that depends on the maintenance time. The indicator is given by;

\[
\rho_i = 1 - \Phi(\log(r_i) - 2.4) \quad \text{where} \quad \Phi \text{ is the distribution function of the standard normal distribution, } r_i \text{ the repair time after } i \text{th failure and } 2.4 \text{ the estimated mean value of the log repair times}
\]

For this function the maintenance indicator nearly equals one for small repair times indicating minimal repair and nearly equals zero for large repair times indicating maximal repair.

Lugtigheid et al. [8] have developed maintenance indicators that are a weighted average of the sub-system operating ages. This is given by;

\[
\rho(t) = \sum_{j=1}^{p} \omega_j \nu_j(t),
\]

where \( \omega_j \) is the weight of the sub-system \( j \) with \( \sum_{j=1}^{p} \omega_j = 1 \)
\( v_j(t) = t - T_{N_{j-1}} + e_{j} \) is the accumulated operating time of the \( j \)th sub-system

\( e_{j} \) is the accumulated operating time of the \( j \)th sub-system used for replacement during maintenance.

3. General Repair Models and Their Properties

All the maintenance indicators devised above are used to generate general repair process models for maintained systems. The general repair process models, postulated over the past thirty five years can be broadly categorised into age based and intensity based models and sub-categorised based on whether they are modified additively or multiplicatively.

3.1. Age Based Additive General Repair Models

The majority of the models the age based general repair models fall under the sub-category where the ages of the system are modified additively.

Kijima and Sumita [10] and Kijima [11] introduced a model based on age loss/recovery as Kijima model I also called the G-Renewal Process (GRP). In this model the system age of only the previous failure epoch i.e., the time between the previous two failures is improved. The conditional intensity of Kijima model I is given by;

\[
\lambda(t \mid H_{i-} ) = \lambda(t - T_{N_{i-}} + v_{N_{i-}} ) \quad \text{where} \quad v_{N_{i-}} = \rho \sum_{i=1}^{N_{i-}} (t_i - t_{i-1})
\]

Brown et al. [12] and Kijima [11] introduced BMS or Kijima model II, in which the total age of the system is improved. The conditional intensity of Kijima model II is given by;

\[
\lambda(t \mid H_{i-} ) = \lambda(t - T_{N_{i-}} + v_{N_{i-}} ) \quad \text{where} \quad v_{N_{i-}} = \sum_{i=1}^{N_{i-}} \rho^{N_{i-}+1-i} (t_i - t_{i-1})
\]

Finkelstein [13] extended Kijima’s model II by letting the failure intensity of the system be a function of the function of the system’s virtual age. The conditional intensity of the model is given by;

\[
\lambda(t \mid H_{i-} ) = \lambda(\phi(t - T_{N_{i-}} + v_{N_{i-}} ))
\]

Kijima’s model I reflects a reality where the maintenance action only reduces the damage that has occurred during the previous failure epoch while Kijima’s model II
reflects the reality where the maintenance action reduces the cumulative damage of all the previous failures. Dagpunar [14] observed that in Kijima model I as the failure intensity increases to infinity the failure times tend to zero i.e, the system will tend to fail immediately on maintenance. He suggested that this is not an appropriate model for most maintained systems. Jacopino et al. [15] observed that the mean cumulative intensity function of the Kijima model I can be adequately approximated by an exponential function while that of Kijima model II can be approximated by a linear function for high number of maintenance actions. At low number of maintenance actions, less than 10, no difference was found between both the models.

The \((p,q)\) rule model of Beichelt [16] and Brown and Proschan [12] and the \((p(t),q(t))\) rule model of Beichelt and Fisher [17] and Block et al.,[18] where \(p,q\) represent the probabilities of minimal and maximal repairs are special cases of Kijima model II. The conditional intensities of the models are given by:

\[
\lambda(t \mid H_{r-}) = \lambda(t - T_{N_{r-}} + \nu_{N_{r-}})
\]

where

\[
\nu_{N_{r-}} = \sum_{i=1}^{N_{r-}} \prod_{j=1}^{N_{r-}} (1 - Z_j) (t_i - t_{i-1})
\]

B-P and BBS models depict a physical reality wherein, a minimal repair is performed a number of times on the system at each failure till a state is reached where the entire system is replaced. These models depict a system for which sequences of superficial repairs are interspersed with periodic overhauls. Such a possibility exists when a catastrophic failure leads to a replacement or major overhaul but other lesser failures lead to a minimal repair. Whitaker and Samaniego [19] also state that these models can also be applied to the case when the type of repair is dictated by external factors like availability of replacement rather than the age or condition of the system.

Dagpunar [20] proposed a model III which is more general than the models I and II in that the current virtual age is a functional of the earlier virtual age and the current age since the last repair. The conditional intensity of the model is given by:

\[
\lambda(t \mid H_{r-}) = \lambda(t - T_{N_{r-}} + \nu_{N_{r-}}) \quad \text{where} \quad \nu_{N_{r-}} = \phi(\nu_{N_{r-1}} + t - T_{N_{r-}})
\]

Dagpunar [14] observed that in Kijima model I as the failure intensity increases to infinity the failure times tend to zero i.e, the system will tend to fail immediately on maintenance. He suggested that this is not an appropriate model for most maintained systems. He suggested a more general structure for the virtual age as a function of the previous virtual age and time since the last maintenance action. He indicated that Kijima model II is a special case of this model and called it model III.
Dagpunar [20] proposed a more general structure for the general repair model structure with the virtual age as a general function of the previous virtual age and time since the last maintenance action. The conditional intensity of the model is given by:

\[ \nu_{N_r} = \phi(\nu_{N_{r-1}}, t - T_{N_r}) \]

Guo and Love [21] provide a general repair model which has a linear function version of the virtual age function of Dagpunar [20]. The conditional intensity of the model is given by:

\[ \nu_{N_r} = (1 - \omega)(\nu_{N_{r-1}}) + (1 - \rho)(t - T_{N_r}) \]

Dorado et al. [22] proposed a unified model (DHS), which contains as special cases all the above models, with a conditional intensity given by:

\[ \lambda(t \mid H_r) = \Theta_{N_r} \lambda(\Theta_{N_r} (t - T_{N_r}) + \nu_{N_r}) \]

For \( \Theta_{N_r} = 1 \), the model reduces to a virtual age model.

For \( \Theta_{N_r} = 1, \nu_{N_r} = 0 \), the model reduces to a maximal repair model.

For \( \Theta_{N_r} = 1, \nu_{N_r} = T_{N_r} \), the model reduces to a minimal repair model.

Malik [23] introduced the proportional age reduction (PAR) or improvement factor model. He stated that maintenance causes a system’s age to improve from what it was before maintenance. Kijima, Morimura and Suzuki [24] also proposed the same model. The conditional intensity of the model is given by:

\[ \lambda(t \mid H_r) = \lambda(t - \rho T_{N_r}) \]

where \( \rho = 0 \) leads to minimal repair and \( \rho = 1 \) leads to maximal repair.

Stadje & Zuckerman [25] introduced a fixed improvement factor model wherein the fixed improvement factor reduces the chronological age of the system by a fixed amount after each maintenance action. The conditional intensity of the model is given by:

\[ \lambda(t \mid H_r) = \lambda(t - T_{N_r} + \nu_{N_r} - \rho) \]

where \( \rho = 0 \) and \( \nu_{N_r} = T_{N_r} \) leads to minimal repair and \( \rho = \nu_{N_r} \) leads to maximal repair.

Doyen and Gaudoin [26] proposed the ARA (Arithmetic reduction of age) models with different memories, generalizing Malik’s improvement factor model. The size of the memory represents the number of repairs whose deterioration can be mitigated by the current repair. The conditional intensity of the model is given by:

\[ \lambda(t \mid H_r) = \lambda \left( t - \rho \sum_{j=0}^{\min(m-1,N_{r-1})} (1 - \rho)^j T_j \right) \]
The PAR model indicated a proportional reduction of the chronological age. This was generalised by in the ARA model wherein in addition to the maintenance factor depending on the previous repair epoch, all repair epochs, or it could depend on \( m \) previous epochs with \( 2 \leq m \leq N_{r} - 1 \). This afforded more flexibility to the age models with the maintenance action reducing the damage accumulated in the previous \( m \) failure epochs.

Veber et al. [27] provided a novel improvement on Kijima model I. They replaced the baseline distribution with a weighted sum of component distributions to obtain a better approximation to the failure process of a maintained system. The conditional intensity of the model is given by:

\[
\lambda(t \mid H_{r}^{-}) = \sum_{j=1}^{m} w_j \hat{\lambda}(t_j - T_{N_{r}} + \nu_{j_{N_{r}}}) \quad \text{where} \quad \nu_{j_{N_{r}}} = \rho \sum_{i=1}^{N_{r}} (t_{j_i} - t_{j_i-1})
\]

The properties of the models remain the same as Kijima model I.

### 3.2. Age Based Multiplicative General Repair Models

The next sub-category of age based general repair models are generated with ages that are modified multiplicatively.

Lam [28] a Geometric process and Wang and Pham [29] proposed a Quasi-Renewal (Q-R) process. In both these processes a factor is introduced which when acting multiplicatively on a random variable, produces another random variable forming a renewal process. The factor acts as a geometric ratio of the process and indicates the trend of system deterioration over a number of failures and repairs. The conditional intensity of the models is given by:

\[
\lambda(t \mid H_{r}^{-}) = \rho^{N_{r}} \lambda(\rho^{N_{r}} (t - T_{N_{r}}))
\]

For \( \nu_{N_{r}} = 0 \), the DHS model reduces to the geometric process or quasi-renewal model.

These models were proposed to model the failure processes of systems where the deterioration of the system was more. The inter failure times \( (t - T_{N_{r}}) \) form a quasi-renewal or geometric process if for any parameter \( \rho > 0 \), \( \rho^{N_{r}} (t - T_{N_{r}}) \) forms a
renewal process. These processes with a ratio $0 < \rho < 1$ will be applicable for a failure process with successive inter-failure times which are stochastically increasing and with $\rho > 1$ will be applicable for a failure process with successive inter-failure times which are stochastically decreasing. They will approximate a renewal process i.e., a failure process exhibiting maximal repair if $\rho = 1$.

Finkelstein [30] proposed a model similar to the geometric process model, terming it as geometrical particular deteriorating renewal process (GPDRP). In this model a scaling factor is applied to the inter-failure times. The conditional intensity of the model is given by:

$$\lambda(t \mid H_{i-1}) = \rho^{N_i} \lambda\left(\rho^{N_i} (t - T_{N_i})\right)$$

This process has properties similar to the geometric process. The process with a ratio $0 < \rho < 1$ is applicable for a failure process with successive inter-failure times which are stochastically increasing and with $\rho > 1$ is applicable for a failure process with successive inter-failure times which are stochastically decreasing. It approximates a renewal process i.e., a failure process exhibiting maximal repair if $\rho = 1$.

Finkelstein [31] defined a more general model, the non-ideal repair process model. In this model the time is perturbed by another process $W(t - T_{N_i})$. The conditional intensity of the process is given by:

$$\lambda(t \mid H_{i-1}) = w_{N_i}^{N_i} (t - T_{N_i}) \lambda(W_{N_i}^{N_i} (t - T_{N_i}))$$

The intensities are independent of each other and reflect a series of deteriorating intensities. These cannot be used to model preventive maintenance. The maintenance serves to bring the system back into operation but is not able to arrest the increasing deterioration of the system.

Kaminskiy and Krivtsov [32] have proposed a G1-Renewal Process (G1RP) which is similar to the geometric and quasi-renewal process in that maintenance indicator acts multiplicatively on the age of the failure process. Its conditional intensity function is given by:

$$\lambda(t \mid H_{i-1}) = \frac{1}{(1 + \rho)^{N_i}} \lambda\left(\frac{t - T_{N_i}}{(1 + \rho)^{N_i}}\right)$$

For $\rho = 0$ the system is set back to the age of a new system. For $\rho < 0$ the system is a deteriorating one and for $\rho > 0$ the system is an improving one.
Rangan and Thyagarajan [33] have proposed a model wherein a dampening factor acts on the virtual age to modify it multiplicatively. They also assumed that the system of initial age $t_0$ is put into operation. The conditional intensity function of the model is given by;

$$\lambda(t \mid H_{\tau}) = \lambda(\nu_{N_{\tau}} \exp(-\rho(t - T_{N_{\tau}}))) = \lambda(t_0 \exp(-\rho t))$$

where; $\nu_{N_{\tau}} = \nu_{N_{\tau-1}} \exp(-\rho(T_{N_{\tau}} - T_{N_{\tau-1}})) = t_0 \exp(-\rho T_{N_{\tau}})$

For $\rho = 0$ the system is set back to the initial age $t_0$. For $\rho = \infty$ the system is set back to the age of a new system. For $\rho < 0$ the system is a deteriorating one and for $\rho > 0$ the system is an improving one.

Wu and Clements-Croome [34] have proposed a novel general repair process model, the extended Poisson process model (EPP) on lines similar to the geometric process. The conditional intensity function of the model is given by;

$$\lambda(t \mid H_{\tau}) = (\alpha a^{N_{\tau}} + \beta b^{N_{\tau}})\lambda$$

where the distribution function is exponential. For $a = 1, b = 1$ the process becomes a HPP. For either of the $\alpha a^{N_{\tau}}$ or $\beta b^{N_{\tau}}$ being equal to zero while the other is not, the process becomes a GP. In other combinations the model can represent the failure patterns of the bathtub curve or other complicated patterns.

### 3.3. Intensity Based Additive General Repair Models

The majority of the models developed under the category of intensity loss or recovery fall under the first sub-category where the intensities of the system are modified additively.

Chan and Shaw [35] proposed two models based on intensity loss/recovery, one with a constant reduction and the second with a proportional reduction. The conditional intensity of the model with constant reduction is given by;

$$\lambda(t \mid H_{\tau}) = \lambda(t) - \rho$$

The conditional intensity of the model with proportional reduction is given by;
\[ \lambda(t \mid H_{r-}) = \lambda(t) - \rho \lambda(T_{N_{r-}}) \]

For \( \rho = 0 \) the system is under minimal repair. For \( \rho < 0 \) the system is a deteriorating one and for \( \rho > 0 \) the system is an improving one.

Doyen and Gaudoin [26] proposed arithmetic reduction of intensity (ARI) model with different memories, generalising Chan and Shaw [35] model with proportional reduction of intensity. The size of the memory represents the number of repairs whose deterioration can be mitigated by the current repair. The conditional intensity of the model is given by;
\[ \lambda(t \mid H_{r-}) = \lambda(t) - \rho \sum_{j=0}^{\min(m-1, N_{r-1})} (1-\rho)^j \lambda(T_{N_{r-j}}) \]

Chan and Shaw models (1993) were developed to indicate a constant and a proportional reduction of the failure intensity on carrying out a maintenance action on the systems. Similar models PIM and DPP were developed wherein scaling factors acting on the intensity function. These were generalised in ARI models wherein in addition to the maintenance factor depending on the intensity of the previous repair epoch or the intensity functions of all repair epochs, it could also depend on the intensity function of \( m \) previous epochs with \( 1 < m < N_{r-1} \). This afforded more flexibility to the intensity models with the maintenance action reducing the damage accumulated in the previous \( m \) failure epochs.

3.4. Intensity Based Multiplicative General Repair Models

The next sub-category of models under this category has models with intensities that are modified multiplicatively.

Percy et al. [36] introduced proportional intensities models wherein they adopted different scaling factors for corrective and preventive maintenance intensities. They also provided for scaling factors which varied with time. The conditional intensity function of the model is given by;
\[ \lambda(t \mid H_{r-}) = \lambda(t) \sum_{i=1}^{N_{r-1}} \rho_i \]  For \( \rho = 1 \) the system is under minimal repair. For \( \rho < 1 \) the system is an improving one and for \( \rho > 1 \) the system is a deteriorating one.

Calabria and Pulcini [37] proposed a model based on scaling of the intensity and called their model discontinuous point process (DPP) model. This is the same as the Percy et
al. [36] model with constant scaling factor. The conditional intensity function of the model is given by; 
\[ \lambda(t \mid H_{t^-}) = \rho^N \lambda(t) \]
For \( \rho = 1 \) the system is under minimal repair. For \( \rho < 1 \) the system is an improving one and for \( \rho > 1 \) the system is a deteriorating one.

Lawless and Thyagarajah [38] have proposed a model wherein the intensity is modified multiplicatively by a factor based on the time since last repair. The conditional intensity function of their model is given by;

\[ \lambda(t \mid H_{t^-}) = \lambda(t) \exp(\rho(t - T_{N_{t^-}})) \]

where;

\[ \lambda(t) = \exp(\alpha + \beta t) \]
For \( \rho = 0 \) the system is under minimal repair with a log linear intensity. For \( \beta = 0 \) the model reduces to a renewal process with times between failures distributed according to a Gumbel distribution.

Calabria and Pulcini [37] have proposed different forms of the Lawless-Thyagarajah model with a power law – Weibull renewal (PL-WR) and a log linear – Weibull renewal (LL-WR) model. The conditional intensity of the PL-WR model is given by;

\[ \lambda(t \mid H_{t^-}) = \lambda(t) (u(t - T_{N_{t^-}}))^{\rho-1} \]

where \( \lambda(t) = \alpha \beta t^{\beta-1} \) and \( (u(t - T_{N_{t^-}})) = \gamma(t - T_{N_{t^-}})^{\delta} \)
For \( \rho = 1 \) the process reduces to minimal repair. For \( \rho > 1 \) the system is an improving one and for \( \rho < 1 \) the system is a deteriorating one. For \( \beta = 1 \) the process reduces to a Weibull renewal process. For \( \beta = 1, \rho = 1 \) the process reduces to a HPP.

The conditional intensity of the LL-WR model is given by;

\[ \lambda(t \mid H_{t^-}) = \rho^N \lambda(t) (u(t - T_{N_{t^-}}))^{\rho-1} \]

where \( \lambda(t) = \exp(\alpha + \beta t) \) and \( (u(t - T_{N_{t^-}})) = \gamma(t - T_{N_{t^-}})^{\delta} \)
For \( \rho = 1 \) the process reduces to minimal repair. For \( \beta = 0 \) the process reduces to a Weibull renewal process. For \( \beta = 0, \rho = 1 \) the process reduces to a HPP.
Guo et al. [39] have proposed a model wherein a dampening factor acts on the intensity to modify it multiplicatively. The conditional intensity function of the model is given by;

$$\lambda(t \mid H_{\tau^{-}}) = \lambda(t) \exp(\rho N_{\tau^{-}})$$

For $\rho = 0$ the system is under minimal repair. For $\rho < 0$ the system is a deteriorating one and for $\rho > 0$ the system is an improving one.

Babykina et al. [40] have proposed a model to link the conditional intensity function to the number of previous failures of the system in order to take care of worse than old repair case also. The conditional intensity function of the model is given by;

$$\lambda(t \mid H_{\tau^{-}}) = \lambda(t)(1 + \rho N_{\tau^{-}})$$

For $\rho = 0$ the system is under minimal repair. For $\rho > 0$ the system’s failure intensity is augmented by the number of previous failures and the system is a deteriorating one.

4. Inference and Application of the General Repair Models to Maintained Systems

To apply these models to maintained systems, parameters of the models are to be estimated from the data set of failures of maintained systems and goodness of fit of the model to the data set assessed. Methods of estimating the model parameters, assessing goodness of fit and applications to maintained systems are presented in the following sections.

4.1. Estimation Of Model Parameters

The parameters of the models can be estimated non-parametrically, which is ideal as no distributional assumptions are necessary. However non-parametric procedures have a drawback in that they do not permit prediction of the future properties of the maintained system. Non parametric estimation for the Kijima model was developed by Dorado et al. [22], BP model by Whitaker and Samaniego [19], BBS model by Hollander et al. [41] and the geometric process (quasi-renewal process) model by Lam et al. [42].

The most common and widely used parametric method of estimating the parameters of a process from the data emanating from a maintained system is the method of maximum likelihood. Estimation of the parameters is obtained by maximizing the log
likelihood function. In most of the situations it may not be possible to obtain the analytical solutions of the likelihood function which may be complex. In such cases numerical methods like Newton-Raphson iteration, Quasi-Newton Methods or the simplex procedure of Nelder and Mead have been used to obtain the estimates. Estimates of variance of the parameters can be obtained by taking the inverse of the observed information matrix and using these, the interval estimates of the parameters can be obtained.

Jack [43] has obtained maximum likelihood estimates of the Kijima models I and II using the Nelder Mead simplex procedure. Yanez et al. [44] have obtained estimates of the g-renewal process or Kijima model I using a numerical algorithm based on Monte Carlo simulation. Mettas and Zhao [45] have obtained the estimates of the Kijima model II using a form of the Newton search method. Love and Guo [46] obtained estimates for PAR model using assumed values of the maintenance indicator and only estimated the values of the other parameters using Newton-Raphson method. Calabria and Pulcini [37] obtained estimates of the PL-WR and the LL-WR models. They also obtained the confidence intervals of the parameters using the asymptotic log-normal distribution for PL-WR model and the asymptotic normal distribution for the LL-WR model.

An inherently different approach to compute the maximum likelihood function numerically is via the Expectation-Maximisation or EM Algorithm. It consists basically of two steps, the E-Step which calculates the expected log likelihood and the M-Step which finds its maximum which are repeated iteratively. Thus one difficult maximisation is broken up into a sequence of easier maximisations converging to the maximum likelihood estimator. This is mostly used with missing data.

Lim [47] carried out inference for the BP model using an EM algorithm for the case where the mode of repair is unknown for cases other than the exponential distribution being the lifetime distribution. Veber et al. [27] used the EM algorithm to obtain the estimates of the weights, the parameters of the component distributions and the maintenance effectiveness, assuming the number of component distributions.

Parametric estimation can also be carried out using the method of least squares or the method of moments. Wu and Clements-Croome [34] use the method of least squares to estimate the parameters of the EPP. Dagpunar [1998] followed a different approach in estimating the expected number of failures for the general repair model type III. Integral equations have been generated for the expected number of failures and the solution obtained using numerical methods.
4.2. Hypothesis Testing

Hypothesis testing of general repair models presents a problem. As of now no proper goodness of fit tests are available to test for general repair models against alternative maximal or minimal repair models or to distinguish between different general repair models. The maximum value of log likelihood can be taken as a criterion for testing goodness of fit of one general repair model versus the other. However the same may not provide a good criterion versus maximal or minimal repair models. In the absence of any better test the maximum likelihood value can be taken as a criterion for checking goodness of fit. When the number of parameters is large and varies from model to model the Akaike Information Criterion (AIC) is used to distinguish between the models. Graphical methods can also be used to check goodness of fit of models.

Guo and Love [21] and Love and Guo [46] obtain the standardised residuals of the fitted models and plot these against the expected residuals from a HPP. The model with the set of standardised residuals closest to the expected residuals from a HPP is taken as the model with the better fit. Calabria and Pulcini [37] have used the likelihood ratio test and the Wald test to test for DPP models against minimal repair models which are nested models. Both the test statistics are approximately distributed as $\chi^2(1)$. However to check for DPPP against DPLP general repair models they arrived at the model with the better fit by looking at the maximum of the likelihood values. Lam et al. [42] used the mean squared error between the estimated values and the actual values of the failure times to check for the model with the better fit. The model with the least mean squared error was taken as the model with the better fit to the data set when comparing minimal repair and geometric process general repair models. Percy and Alkali [48] based their assessment of the fit of the models using the maximum likelihood values. The model with the highest value was taken as the model with the better fit. Veber et al. [27] obtained the model with the best fit by plotting AIC against the number of component distributions. The model with the lowest AIC was taken as model with the better fit.

4.3. Application To Maintained Systems

The general repair process models have been applied to the failure data of various maintained systems. Guo and Love [21] applied the Kijima Type I model to roller mill failure data extracted from a local cement plant to obtain the estimates of the parameters with Weibull distribution for the baseline. They also estimated the parameters using a set of assumed maintenance indicators based on the judgement of the maintenance personnel on the effectiveness of repair. Yanez et al. [44] applied the
g-renewal model or Kijima model I to the data sets of failures of a compressor and U.S.S. Halfbeak No. 3 main propulsion motor to obtain the expected number of failures. Mettas and Zhao [45] applied the general repair models of Kijima I and II to the failure times of eight systems along with a Weibull Renewal and NHPP. Based on the least likelihood estimate they found that for both a single system no. 8 as well as for multiple systems 1 to 6 the General Repair Model Kijima II provided the best fit. Kahle and Love [7] applied the maintenance indicator based on repair times to the failure data of a hydroelectric turbine unit within the British Columbia Hydro-Electric Power Generation System.

Whitaker and Samaniego [19] applied the BP model to the failure data of the air-conditioning system of Boeing 720 airplane 7914 by assigning reasonable values for the maintenance type. Lindqvist, Langseth and Stove [49] applied their repair alert model to the failure times of ARC-1 VHF communication transmitter-receivers of a single commercial airline using a BP model and a power law function for the cumulative repair alert function. They concluded that the PM carried out was optimum for this system as the data censor the failures in a near optimal way as regards the repair alert function.

Shin et al. [50] applied the PAR model with power law baseline intensity to the data set of failures of a central cooler system of a nuclear power plant. Syamsundar and Naikan [51] have applied the ARA₁ and ARI₁ imperfect repair models to the failure times of the AMC Ambassador Cars 3 and 5 along with acceleration of the failure times using time averaged cumulative mileages.

Calabria and Pulcini [37] applied the DPP model with power law and log linear baseline intensities to the failure times of a 180 T rear dump truck. Lam et al. [42] applied the geometric process model to the data sets of failures of air conditioning equipment of Boeing 720 aircrafts nos. 3, 6 and 7, electronic computer, main propulsion engine no. 3 of U.S.S. Halfbeak and main propulsion diesel engine no. 4 of U.S.S. Grampus. Percy and Alkali [48] applied the PIM models with log linear and power law baseline intensities to the failure data of a main pump A in a petroleum industry, and ascertained the PIM model with the log linear baseline as the model with the better fit. They also found that the PM scaling factor for this model was less than one indicating PM has a beneficial effect in reducing the failure intensity of the pump. Calabria and Pulcini [52] applied the PLWR and LLWR models to failure data sets AMC Ambassador Car no. 3, aircraft generator, air-conditioning equipment of Boeing 720 airplane no. 6.
Veber et al. [27] applied the GRP with Weibull mixture distributions for the baseline to the failures data set of the main propulsion motor no. 3 of U. S. S. Halfbeak and obtained the model with a 3 component mixture distribution as the model with the better fit.

The more general repair process models of Finkelstein [30], Dagpunar [14],[20], Guo and Love [53], Dorado et al.[22] and ARA\textsubscript{m} and ARI\textsubscript{m} models of Doyen and Gaudoin [26] are yet to be seriously applied to the failure data of maintained systems.

5. Physical Reality Depicted by the Models and Their Shortcomings

The general repair models are to be used to take maintenance decisions on a regular basis. The usefulness of these models can be judged only by the extent to which they fulfil the needs of the maintenance personnel. For this to happen answers to a number of questions are to be obtained. How to select a model based on the physical realities of the maintained system? How well do the model parameters reflect the state of the systems and the effects of maintenance on the systems? What can you say and not say about the nature of the maintenance based on these models? How can the models be used to help the maintenance personnel in making objective decisions about the maintenance actions to be taken?

The following sections address these important questions and suggest some criteria for making rational decisions for selection of appropriate models for specific applications. This will help the maintenance engineers in planning their maintenance activities more effectively.

5.1. Selection of the Models

In almost all the papers the data sets of failures of maintained systems are used more as an illustration of the application of a general repair process model rather than as a real case study of maintained system. Even though a large number of models are available in the literature, no methodology has been published yet to decide the most appropriate model for a given situation.

Some pointers can be obtained on the type of model that will fit a given data set of failures based on the model properties. The BP and BBS models are generated on the assumption of several minor maintenance actions followed by a major maintenance action or replacement and can be applied in such situations. The BP model is developed on the assumption that the times between major maintenance actions or replacements are homogeneous and the BBS model that the times between major
maintenance actions or replacements reduce with age. This can be used to distinguish between the models.

When only one component or sub-system is changed on failure, the failure intensity may reset only the damage caused during the time since the previous failure and GRP, ARA$_1$, ARI$_1$ or Chan and Shaw [35] models can be applied to such situations. When more than one component or sub-system is replaced then the damage caused during a few past failures can be restored and ARA$_m$ and ARI$_m$ models may be used. If major maintenance activities are carried out on the system then ARA$_\infty$ and ARI$_\infty$ models may be used. When the system is found to be continuously deteriorating even after maintenance actions are taken on the system then GP, QRP, GPDPRP or G1RP models can be seen as more appropriate. Babykina et al. [40] model is designed to take care of systems where the system state after maintenance becomes worse than before.

A similar structure based on physical reality is not apparent in the case of all models. Given the plethora of general repair process models, a number of models become equally plausible candidates for modelling the general failure behaviour of the systems. When no physical basis is available one can only apply all the models to the failure times generated by the system and obtain the appropriate model by assessing the better fit to the data set based on the higher of the maximum log-likelihood values. This provides only a statistical fit to the data set of failures of the system rather than provide any physical basis for modelling the system. Sometimes more than one model may be found as suitable candidates for modelling the system based on the closeness of the goodness of fit.

5.2. Understanding the Failure Process, State of the System and Nature of Maintenance Actions

The estimated values of the parameters and the properties of the fitted models can be used to understand the behaviour of the failure process. The shape parameters of the models give an indication of the rapidity of deterioration of the system. The higher the shape parameter, the higher is the deterioration of the system. The maintenance indicator provides an estimate of the averaged effectiveness of the maintenance actions. The closer the maintenance indicator is to the maximal repair situation the better the quality of the maintenance actions. The intensity of failures of the system can be used to compare the failure processes of different systems. Higher intensities indicate more number of failures and the need for better maintenance.

Maintenance indicators representative of the varying quality of maintenance have been developed. Guo and Love [21] have provided maintenance indicators for the various
maintenance actions based on the judgement of the maintenance personnel on the effectiveness of the maintenance actions. Their judgement was based on the time taken for repair and / or the amount of money spent on the replacement of parts. Kahle and Love [7] have provided a maintenance indicator to be a function of maintenance times. The more the time spent on maintenance the indicator is closer to the maximal repair condition. Lugtigheid et al. [8] based their maintenance indicators on a weighted average of the component ages bringing the maintenance actions closer to reality. This provides a better estimate of the system state and hence better indicates the effectiveness of maintenance actions on the system.

The type of model, which provides the better fit to the failure data of the maintained system, and its properties can provide some information on the failure behaviour and maintenance actions on the system. GRP, PAR, ARA$_1$ or ARI$_1$ models providing a better fit indicate that the maintenance actions in general are not effective in reducing the cumulative deterioration of the system while the exact opposite is the case with Kijima model II, ARA$_\infty$ and ARI$_\infty$ models wherein the cumulative deterioration of the system is addressed. ARA$_m$ and ARI$_m$ processes model cases in between these two extremes and the maintenance actions help to mitigate the deterioration which has set in during the previous $m$ failures. The sub-category of multiplicative models provide for the cases with a lesser improvement as compared to the sub-category of additive models.

5.3. Use of the Models for Taking Maintenance Actions

The above models are useful only when they satisfy the requirements of industrial maintenance. These models can help provide objective maintenance decisions by being able to estimate the quantities of interest to the maintenance managers / engineers.

Kahle [54] proposed a maintenance policy that is failure rate optimal by carrying out maintenance actions at equal intervals of times dependent on the state of the system. Lindqvist et al. [49] applied their repair alert model to the failure times of ARC-1 VHF communication transmitter-receivers of a single commercial airline. They concluded that the PM carried out was optimum for this system as the data censor the failures in a near optimal way as regards the repair alert function i.e., the crew were alert enough to carry out the PM actions in time to prevent damaging breakdowns. Percy and Alkali (2006) applied the PIM models with log linear and power law baseline intensities to the failure data of a main pump in a petroleum industry and found that the PM scaling
factor for these models was less than one indicating PM has a beneficial effect in reducing the failure intensity of the pump.

General repair process models have been studied or used to optimise the maintenance actions based mostly on cost or reliability. Wang [55] presented a survey of various models used for optimising the timing of PM actions based on cost or reliability. Jiang et al. [56] studied a maintenance model with GRP process and two types of replacement: failure and preventive to find a repair or replacement policy to minimise the long run expected average cost per unit time. They developed such a policy but did not apply it to any maintained system. Wu and Clements-Croome [34] have proposed two optimum maintenance policies using EPP for both preventive and corrective maintenance times.

5.4. Shortcomings of the Models

The general repair models have been developed one after the other by generalising existing models without recourse to physical reality i.e., without an application in mind and often based more on satisfying mathematical appropriateness rather than explaining physical phenomena. Hence given a data set of failures it is not easy to select the appropriate models for the data set. If the appropriate model is not chosen then all the estimates based on this model will prove to be not true.

A constant maintenance indicator averages out the effect of maintenance actions on the system and will not be able to indicate the exact state of the system at any given time. Indicators based on maintenance times or cost or weighted average of sub-system ages provide better indicators of system state but need additional information other than system failure times. These are to be incorporated into the general repair models to obtain the estimates of the other parameters.

To set up a maintenance policy the preventive maintenance carried out on the systems is also to be taken into consideration, which when added to corrective maintenance can only provide a better planning of maintenance actions. This can be used to optimise the maintenance based on cost or reliability. However no optimisation can be successful if considered in isolation for that particular piece of equipment. It has to be incorporated into the shop schedule along with the maintenance schedules of other equipment as well as their operating regimes. Hence the modelling needs to take care of these aspects also.

A variety of maintenance actions can be performed on the system. Routine maintenance actions like cleaning, adjustments, bolt tightening, oil change, changing
of small parts or consumables like filters, gaskets etc. are carried out. In case these actions are not carried out it may lead to the deterioration of the system and lead to quicker failures. These actions do not form part of the general repair models and are generally ignored in the modelling process.

The general repair models do not provide any information on the nature of the maintenance action to be carried out or the parts, components or sub-systems required to be changed other than indicating that the actions should depend on the state of the systems. They cannot also provide an indication of the necessary resources required for maintenance.

The basic assumption in all the general repair models is that the maintenance activity does not change the baseline distribution or the intensity as the case may be. This means that the probabilistic or parametric structure of the systems’ failure process does not change but only the structure is shifted horizontally. In reality the distribution of time to failure is bound to change for each failure epoch even with a small maintenance improvement. These changes are again not reflected by the models.

General repair models can only describe a failure process intensity of a system which is monotonously increasing or decreasing with the operating time. They cannot appropriately model the failure process of a system when a non-monotonic or complicated trend in the failure data is observed. General repair models do not take into account times taken for maintenance actions hence cannot provide estimates of maintainability or availability of the systems.

Unless all these aspects are taken into consideration modelling of a maintained system will be only partial and will not lead to correct decisions on maintenance of systems.

6. Conclusion

Though the general repair models depict a picture of reality closer to the nature of maintenance activity being carried out there still remains a lot to be done. Most of the models are motivated by statistical reasons rather than providing a picture of physical reality seen on the ground. The general repair process models fail to address a number of issues required for the maintenance of industrial systems and hence will be able to provide only partial answers to maintenance problems. There is a need for better modelling of the failure-maintenance processes of maintained systems before these can be deployed to take actions in an industrial environment on a regular basis.
References


