MODELING OF RECOVERY RATE OF INFRASTRUCTURE SYSTEM USING THE HISTORICAL DATA

Abbas BARABADI
Associate professor, University of Tromsø, Norway
Abbas.b.abadi@uit.no

Ove T. GUDMESTAD
Professor, University of Stavanger, Norway
Ove.t.gudmestad@uis.no

Mostfa ALIYARI
North Khorasan Electrical Distribution Company, Iran

Keywords: Infrastructure, Risk Management, Recovery Rate, Resilience

ABSTRACT

The resilience of infrastructure systems is of significant concern and in the case of disruptive events such as earthquakes, it is very important to have an effective plan to improve the time to recovery of infrastructure systems. Hence, an effective statistical model needs to be used to evaluate time to recovery of infrastructure. This paper applies some statistical technique from biostatistics; covariate based statistical methods, for estimating the time to recovery as well as identification and qualification of the effect of influence factors on the recovery time. The application of the method is illustrated by a case study from an electric power outage data set.

INTRODUCTION

The resilience of infrastructure systems is of significant concern in the case of disruptive events such as earthquakes. Resilience is often described as a function of robustness and recovery rate. Robustness defines the ability of a system to resist the initial adverse effects of a disruptive event and the recovery rate shows the rate or speed at which a system is able to return to an appropriate operability following the disruption (McDaniels et al., 2008; Barker and Baroud, 2014). In order to have an effective risk management plan in the case of natural disaster, it is very important to have an accurate estimation of the recovery rate of the infrastructure as well as knowing the impact of such natural disaster on different elements of society (the consequence of the disruptive event). Historical data play an important role in the estimation of the recovery rate of a disruptive event as they reflect the conditions that the recovery crew and different components of infrastructure has experienced during the recovery process. Historical recovery data are generally non-homogeneous. This can be due to differences in disruptive events, operational and environmental conditions, requirements and available resources, recovery procedures, etc. Hence, the statistical method that is going to use in analysing such data should be able to model such influence factors. Recently, the applications of covariate based statistical models (CBSMs) in order to estimate the repair rate have been addressed in reliability engineering (Barabadi et al., 2011a; Barker and Baroud, 2014; Gao et al., 2010). In these models, all influence factors on recovery rate are modelled as covariates. For example, Gao et al. (Gao et al., 2010) developed the concept of the proportional repair model (PRM), and later Barabadi et al. (Barabadi et al., 2011a) showed the application of the PRM to model the effect of time-dependent and time-independent covariates on the repair rate of the equipment. This paper developed a methodology in order to analysis the collected historical recovery data using the covariate based statistical models (CBSMs). The application of the method will be illustrated by a case study.

RECOVERY RATE ESTIMATION USING COVARIATE BASED STATISTICAL MODELS

Figure 1 shows the main steps for the recovery rate estimation using the covariate-based models. As this figure shows, at the first stage of the recovery process of different elements of selected infrastructures the process needs to be mapped. The process map shows the design logic of the recovery process and the set of the activities that need to be carried out to recover the distributed infrastructure. The process map included boxes and arrows joined together. A process map is a powerful visual tool for understanding how a recovery process operates. It is easy to find the weaknesses and areas of risk in a recovery process with its drawn process map. It allows seeing interconnectivity within steps, across steps, and the impact of each step's reliability on the process outcome. Moreover, it will help to design a better process and to create key performance indicators to monitor and measure process improvements.

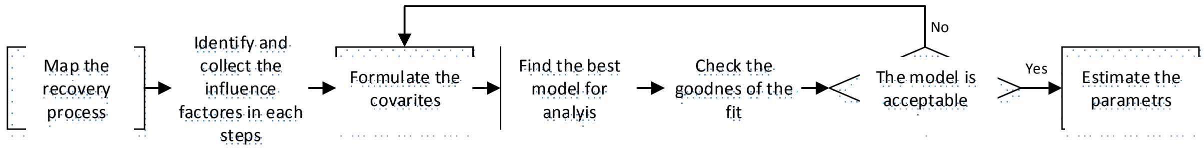

Fig.1. Recovery rate estimation using the covariate-based model

In the next stage, all influence factors in the different steps of the recovery process need to be identified. For example in the recovery process of a feeder in a power distribution system, replacing a failed power pole is an important step. Hence, all operational conditions such as ambient temperature, which may affect this replacement, need to be identified. Thereafter, the identified covariates need to be formulated. In general, covariates can be divided into two main groups: (i) categorical covariates, and (ii) continuous covariates. Categorical covariates are qualitative variables. These can be binary or have multiple categories. A binary covariate can be handled in the model by use of an indicator variable, coded such as zero or one. Continuous covariates have a defined scale, and can be quantified, which can change linearly or nonlinearly, such as power models. Moreover, each of these two groups (categorical or continuous covariates) can be time-dependent or time-independent.

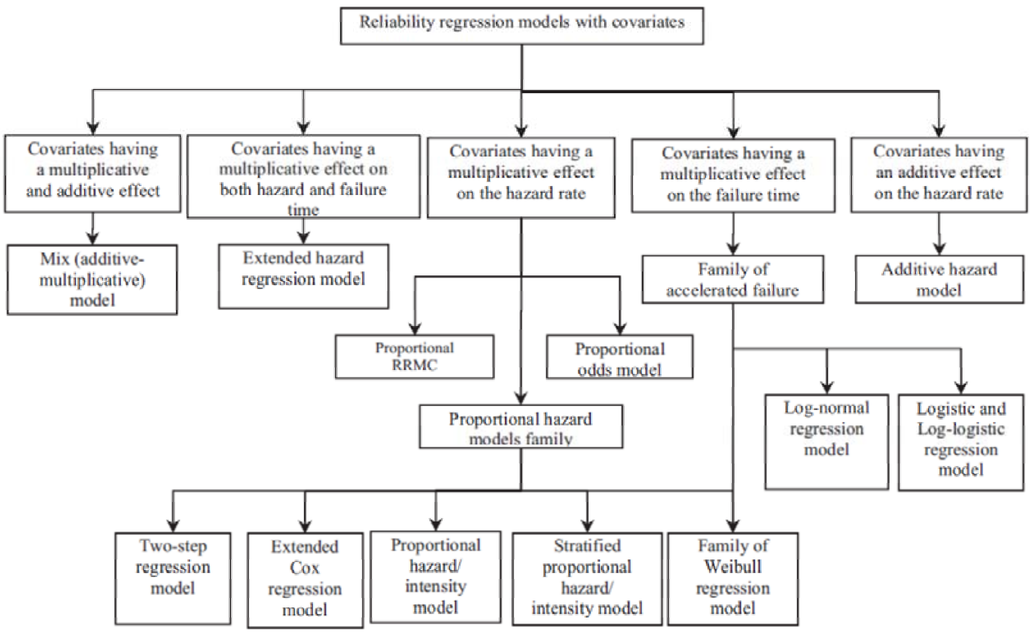

Fig. 2. The most important covariate based statistical models, CBSMs (Barabadi et al., 2014)

After formulation of the covariates the next stage is the selecting the appropriate statistical approaches. In general (based on the assumption that the covariates can affect the reliability of the item) different statistical methods have been developed and used in medical science and reliability engineering (Barabadi et al., 2011b). Figure 2 shows the most important CBSMs. İn general these models can be categorized in two main groups, namely: parametric and non-parametric methods. In the parametric method, such as the family of accelerated failure time models, the lifetime of a system is assumed to have a specific distribution such as Weibull. On the other hand, in the non-parametric method, such as the proportional hazard models family, no specified distribution is assumed for the lifetime of a system. In general, the basic theory of these non-parametric methods is to build the baseline hazard function using historical failure data and the covariate function using covariate data. The baseline hazard function is the hazard rate that an item will experience when the effect of the covariates are equal to zero. The covariate function shows how the baseline hazard model will be changed due to the effect of covariates.

The most important step in historical data analysis using the CBSMs is selecting the appropriate model as if the historical data does not follow the applied model, result of analysis may be misleading. A suitable statistical approach must be selected based on the effect of covariates on the repair process (Kumar and Westberg, 1997, Barabadi et al., 2014). Figure 3 shows the systematic methodology to select the appropriate model. In order to select the appropriate CBSM for a given data set, the failure data must be grouped on the basis of the discrete value of a single covariate, or on a combination of discrete values for a set of covariates. Then a plot of the logarithm cumulative hazard rate versus time for each group (stratum) of data must be compared. The product-limit method can be used to calculate an empirical estimate of the cumulative hazard rate (Kumar and Klefsjö, 1994). If the proportional hazard model (PHM) is the appropriate model for the data, these plots should be parallel in a vertical direction. If these plots appear to be parallel in a horizontal direction, the accelerated failure time (AFT) model may be the appropriate model. If the plotted curves are not parallel, the stratified PHM may be suitable. If the plotted curves are parallel but in neither a vertical nor a horizontal direction, a mixed or Weibull regression model may be appropriate. However, for uncensored failure data, the assumption of the PHM in the place of the AFT or vice versa does not have a significant effect on the estimate of the relative importance of the covariates (Kumar and Westberg, 1997).

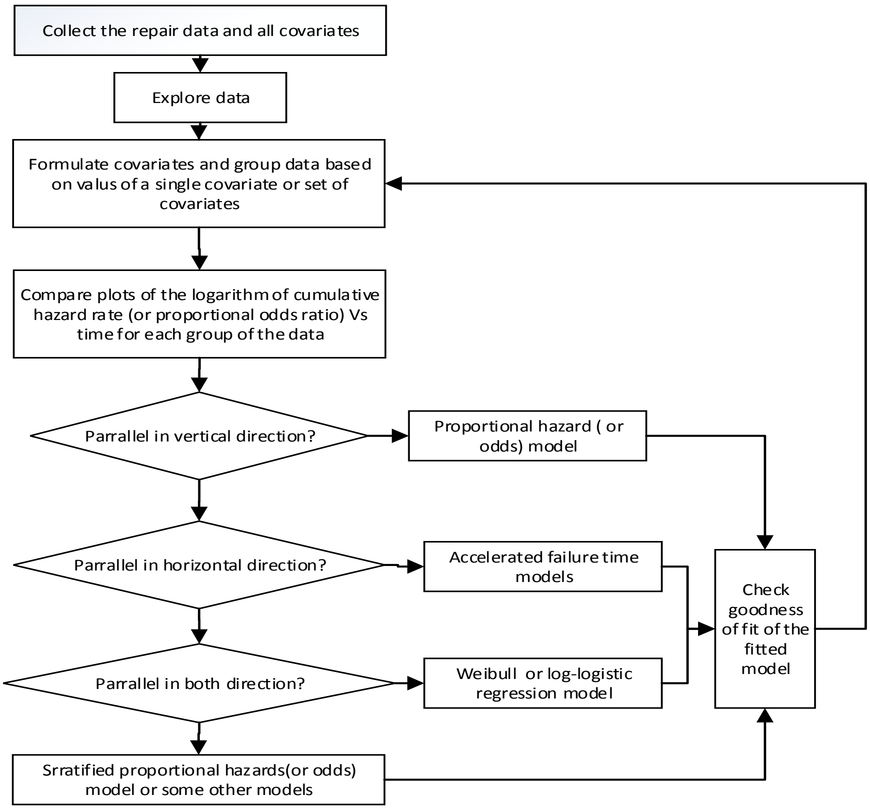

Fig.3. Guideline for selection the appropriate model(Kumar and Westberg, 1997)

After selecting the appropriate model the goodness of the fit and parameter estimation need to be carried out. The estimation of the reliability function in these methods requires an advanced statistical software package. A number of programs and packages are available for the calculation of the parameters of these methods such as SPSS, SAS and S-plus. However, some of these methods have theoretical and computational difficulties while estimating the parameters of the models. In general, models from the proportional hazards family and the accelerated failure time family appear to be suitable for practical applications (Kumar and Westberg, 1997).

CASE STUDY

The electrical power distribution network (EPDN) is the backbone of any society. The electrical power distribution is the last stage in the delivery of the power to the customer from the main transmission system. The EPDN includes all parts of electrical utility systems between bulk power sources and the customers' service-entrance equipment. The main function of the EPDN is to supply electrical power generated from large sources to customers at the desired voltage level and with a degree of appropriate reliability (Li, 1994). The EPDN is a critical infrastructure that is essential for the functioning of a society and economy. Hence, in the event of disruptive events it needs to be recovered as soon as possible. In this case study the recovery data from the North Khorasan Electrical Distribution Company (NKEDC) are used for the analysis. Due the geological location of the Bojnored, the NKEDC has experienced several disruptions due to natural disasters such as flood, hurricane and lighting. Figure 4 shows one of the failures due to flood.

Fig.4. (A) Repair in NKEDC after a flood, (B) Failed power pole due to the flood

Using the available historical data and the expert opinions the defined covariates in this case study are: length of the feeder, number of recovery groups, time of the event (night or day), event type (flood, hurricane and fire) and the date (season) of the event. Here we used the date of the event, which reflect the ambient operational conditions such as temperature.

The event type, date of the event and time of the event are categorical covariates and they need to be formulated. The formulation of the covariates is shown in Table 1. Table 2 shows an example of the data and their associated covariates.

Table 1: Formulation of the covariates

Time of the event		Event type			Date of the event			
Day	Night	Flood	Fire	Hurricane	Spring	Summer	Fall	Winter
1	0	1	2	3	1	2	3	4

Table 2: An example of collected data and their associated covariates

Time to recovery (TTR), hours	Date of event	Time of event	Length of the feeder (km)	Number of recovery groups	Event type
1 010	1	1	50	3	Flood
166	1	1	60	1	Flood
533	2	0	50	3	Flood
292	1	1	50	2	Hurricane
270	2	0	60	2	Hurricane
997	4	1	50	3	Hurricane
563	4	1	60	2	Hurricane
55	2	0	50	1	Fire
160	2	1	60	1	Fire
41	2	0	60	1	Fire

In the current case study, most plots indicated that the curves showing the logarithm of the cumulative hazard rate for different strata of covariates were roughly parallel vertically. Hence, the PHM or PRM models can be used for this data set. As mentioned the PRM are developed based on PHM to predict the maintainability of equipment. In the PRM, the repair rate of a component is the product of a baseline repair rate $\mu_0(t)$, and a functional term $\psi(z\beta)$, which describes how the repair rate changes as a function of influential covariates. The PRM is described as follows (Barabadi et al., 2011a):

$$\mu(t,z) = \mu_0(t)\psi(z\beta) \tag{1}$$

where z is a row vector consisting of the covariates, and β is a column vector consisting of the regression parameters. The baseline repair rate is the repair rate under the standard conditions, $z=0$, and requires $\psi(z\beta)$ $=1$, when there is no influence of covariates on the repair time. The shape of the baseline repair rate and the regression coefficients for the covariates may be estimated from historical data or by using input from experts. Different parameterization forms of $\psi(z\beta)$ can be used, such as the log linear form, $\psi(z,\beta) = e^{\beta^t z}$, the linear form, $\psi(z,\beta) = 1 + \beta^t z$, and the logistic, $\psi(z,\beta) = \log(1 + \beta^t z)$.

Table3: The results of the analysis

Steps	Risk factor	β	Standard error	Wald	Degrees of freedom	Statistical significance	Exp(β)
Step 1	Number of the recovery group (Z_1)	-3.207	0.881	13.245	1	0.000	0.040
	Type of event (Z_2)	1.104	0.422	6.847	1	0.009	3.015
	Length of the feeder (Z_3)	-0.144	0.071	4.135	1	0.042	0.866
	Time of the event (Z_4)	1.556	0.837	3.458	1	0.063	4.739
	Date of the event (Z_5)	0.122	0.287	0.180	1	0.671	1.130
Step 2	Number of recovery group (Z_1)	-3.147	0.868	13.144	1	0.000	0.043
	Type of the event (Z_2)	1.120	0.417	7.205	1	0.007	3.065
	Length of the feeder (Z_3)	-0.135	0.067	4.065	1	0.044	0.874
	Time of event (Z_4)	1.577	0.837	3.551	1	0.060	4.841

The analysis of data was done using the software SPSS with the backward stepwise method [3, 9]. At the fires stage, the regression coefficient β was estimated and the significance of each β was tested by calculating the Wald statistics and its p-value. In this study, a p-value of 5% is considered as the upper limit to check the significance of covariates. The results of the analysis are shown in Table 3. In this table, Exp(β) is the repair ratio. This ratio indicates the expected changes in the recovery rate when its categories change, or, for continuous covariates, it predicates the change in the repair rate for each unit increase in the covariate.

If Exp(β) is less than 1.0, the direction of the effect is toward reducing the repair rate. According to the result of the analysis, the following covariates: Number of recovery group (Z_1), Type of event (Z_2), Length of the feeder (Z_3) and Time of event (Z_3) have a significant effect on the recovery process of the PDNW. The maintainability of the feeder on the NKEDC, which shows that the ability of the feeder under given conditions of use, to be restored to a state in which it can perform its required function as:

$$\mu(t,z) = \mu_0(t)\exp(-3.147z_1 + 1.12z_2 - 0.135z_3 + 1.577z_4) \qquad (2)$$

The maintainability of the feeder is shown in Figure 5. For example this figure shows the maintainability of the feeder after 400 hours will be equal to 87 percent.

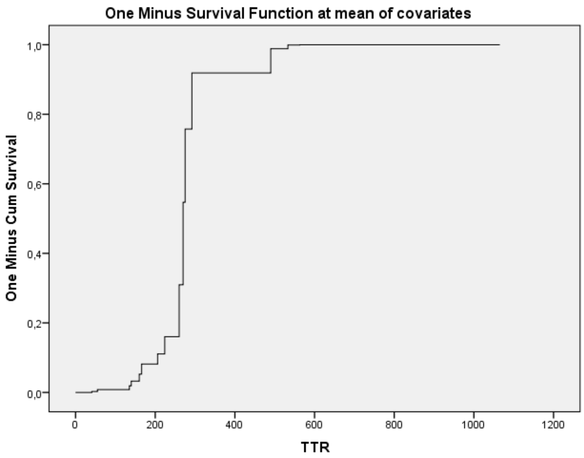

Figure 5: The maintainability of the feeder

CONCLUSIONS

The operational environments may have a significant influence on the recovery rate of infrastructure systems. Therefore, it is important to understand how they affect the recovery process of infrastructure systems. İn some cases ignoring the operational environments' effects on the recovery process may lead to wrong results in the contingency plan. Hence, a suitable statistical approach needs to be selected to analyze the effect of the covariates. This paper presents a methodology for estimating the recovery rate of infrastructure systems. In this methodology the effects of operational conditions are modeled as the covariates. Thereafter, a case study is used to demonstrate how to apply the methodology for a power distribution network. The result of the case study shows that it is necessary to consider the effect of covariates on the recovery rate of power distribution network. However, the collected data system needs to be improved in order to collect all influence factors on the recovery process of a power distribution network.

REFERENCES

Barabadi A, Barabady J and Markeset T (2011a) Maintainability analysis considering time-dependent and time-independent covariates, *Reliability Engineering & System Safety,* 96**,** 210-217

Barabadi A, Barabady J and Markeset T (2011b) A methodology for throughput capacity analysis of a production facility considering environment condition, *Reliability Engineering & System Safety,* 96**,** 1637-1646

Barabadi A, Barabady J and Markeset T (2014) Application of reliability models with covariates in spare part prediction and optimization–A case study, *Reliability Engineering & System Safety,* 123**,** 1-7

Barker K and Baroud H (2014) Proportional hazards models of infrastructure system recovery, *Reliability Engineering & System Safety,* 124**,** 201-206

Gao X, Barabady J and Markeset T (2010) An approach for prediction of petroleum production facility performance considering Arctic influence factors, *Reliability Engineering & System Safety,* 95**,** 837-846

Kumar D and Klefsjö B (1994) Proportional hazards model: a review, *Reliability Engineering & System Safety,* 44**,** 177-188

Kumar D and Westberg U (1997) Some Reliability Models for Analyzing the Effect of Operating Conditions, *International Journal of Reliability, Quality and Safety Engineering,* 4**,** 133-148

Li W (1994) *Reliability assessment of electrical power systems using Monte Carlo methods*, Springer

McDaniels T, Chang S, Cole D, Mikawoz J and Longstaff H (2008) Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation, *Global Environmental Change,* 18**,** 310-318