Reliability Engineering and System Safety 169 (2018) 209–223

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Post-disaster infrastructure recovery: Prediction of recovery rate using historical data

A. Barabadi*, Y.Z. Ayele

Department of Engineering and Safety, UiT The Arctic University of Norway, Tromsø, Norway

ARTICLE INFO

Keywords:
Resilience
Recovery rate
Infrastructure
Robustness
Regression model
Bayesian approach

ABSTRACT

The recovery of infrastructure systems is of significant concern; in order to have effective risk management planning, an accurate prediction of the recovery time is required. A system may have different recovery paths due to the time of the accident, nature of the disruptive event, and surrounding environment, among many other factors. Hence, any model, which is employed to estimate the recovery time, should be able to quantify the effect of such influencing factors. Missing data, inappropriate assumption by analysts, and lack of applicable methodology are some practical challenges for recovery rate analysis. The purpose of this paper is to develop a methodology to address these challenges. It is based on the availability and the nature of historical data; it involves various steps, including categorizing the given data set into three groups: no or missing data set, homogeneous data set, and heterogeneous data set. Here, the Bayesian approach has been employed to handle the no or missing data set group. For the heterogeneous data set group, the proposed methodology suggested the application of covariate based models. Finally, for the homogeneous data set, the methodology employed statistical trend tests, to find the appropriate regression models. The application of the methodology is illustrated by real case studies.

1. Introduction

Natural disasters, such as floods, volcanic eruptions, tsunamis, etc., happen frequently around the world. A natural disaster in a vulnerable area may have different types of consequences including fatalities, injuries, property damage, etc. [1,2]. The resilience of the infrastructure in an area can mitigate the adverse consequences of such kinds of disruption. Resilience is often described as a function of robustness and rapidity [3]. Robustness is defined as the ability of a system to resist the initial adverse effects of a disruptive event, while the rapidity is the rate or speed at which a system is able to return to appropriate operability following the disruption [3–5]. The robustness of infrastructure is very important in preventing the impact of disruptive events; however, considering the fact that it is not possible to prevent all disasters, an acceptable recovery rate is also required to reduce the consequence of natural disasters. Moreover, critical infrastructure systems such as transportation, energy, health and communication, etc. are strongly interdependent, and a failure in a critical infrastructure system can spread by cascading effects to other areas of infrastructure or sectors and then complicate matters further [6]. Hence, in the case of a disruptive event, which affects a given critical infrastructure, it is crucial to increase the recovery rate, in order to reduce the cascading effects on the other critical infrastructures [3,5].

A schematic representation of the resilience concept in a specific infrastructure system, in terms of system function before and after an extreme/disruptive event, is depicted in Fig. 1. As Fig. 1 illustrates, an extreme event such as a flood occurs at time t_0, and, over time, system function recovers until t_R, and then after t_R, the system is fully restored. The consequences of the disruption can be reduced through ex-ante mitigation or ex-post actions. For instance, the consequences of a disruption in a power distribution system, due to flooding, can be reduced through ex-ante mitigation such as installing backup generators for critical customers.

To reduce the consequences from extreme/disruptive events and assure effective consequence/risk management planning, precise estimation of the resilience of infrastructures plays a crucial role. It helps to identify specific actions that will eliminate or mitigate consequences associated with specific problems, regardless of the cause. Over the years, a number of studies have been carried out to quantify the resilience of critical infrastructure systems; see e.g. Bruneau et al. [7], Chang et al. [3], Rose [8], MacKenzie and Barker [9]. For instance, Bruneau et al. [7] used the expected loss due to an earthquake over time to quantify resilience for various types of physical and organizational systems. MacKenzie and Barker [9] employed a Dynamic Inoperability Input-Output Model to quantify the resilience of a critical infrastructure sector. Chang et al. [3] summarized the obstacles to fostering infrastructure resilience in three main groups as: partial incentives, limited and asym-

* Corresponding author.
E-mail address: abbas.b.abadi@uit.no (A. Barabadi).

http://dx.doi.org/10.1016/j.ress.2017.08.018
Received 20 March 2017; Received in revised form 31 July 2017; Accepted 29 August 2017
Available online 31 August 2017
0951-8320/

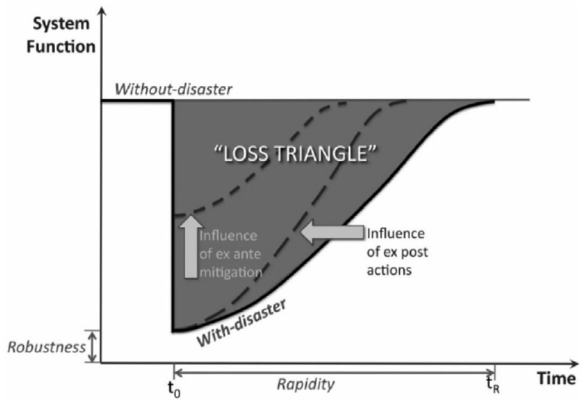

Fig. 1. Resilience concept, adapted from Ref. [3].

metric information, and lack of experience. Thereafter, they developed an approach based on expert judgments to address these challenges. In general, the main focus of the available literature related to resilience modeling is on the optimization of preparedness and resource allocation strategies under different circumstances [5].

A system, as Fig. 1 illustrates, may have different recovery paths due to the structure, ex-ante mitigations, ex-post actions at the time of the accident, nature of the disruption, competence of the recovery crew involved, and surrounding environment, among many other factors [5]. Hence, any model, which is employed to estimate the recovery rate, should be able to capture the impact of the influencing factors on the recovery rate. Capturing and modeling the impact of these factors is useful in optimizing the over- and under-preparation of recovery crews and materials required for the restoration of disrupted infrastructure systems. Moreover, this information can be used to establish better collective response planning and the coordination of restoration efforts for other dependent critical infrastructures.

However, most of the restoration process literature is oriented toward lessons learned and management results that are generally unquantifiable [10–13]. Further, models applied in estimation of the recovery rate have focused on providing some restoration curves (percentage of customers with service versus time) [14–17], the application of detailed network models [18,19] and Markov processes [20–22]. In these methods, input data requirements are significant, with results typically location specific and limited in transferability to other areas without developing new models [22,23]. This is where the application of the appropriate statistical approach is realized, for modeling the stochastic behavior of recovery when predicting the recovery rate even before damage assessments are available from the field. Recently, the application of regression models with covariates, for estimating the repair rate, is becoming popular in reliability engineering [5,24,25]. In these models, all influence factors on the repair rate are modeled as covariates. Covariates are defined as all those factors, which may have an influence on the repair process. For instance, Gao et al. [25] proposed the proportional repair model (PRM) concept, for modeling the impact of Arctic operational conditions on the repair process of offshore production facilities. Further, Barabadi et al. [24] demonstrated the application of the PRM to model the effect of time-dependent and time-independent covariates on the repair rate of crushing plants in mining industries. In general, using regression modeling can enable us to utilize a set of covariates and recovery time data from similar past events to predict the recovery rate during future events. Moreover, an effective regression approach can model a range of covariates that can influence recovery durations.

However, there is a lack of literature that models the trajectory of the recovery of infrastructure systems over time considering external and internal impacts using regression models with covariates [5,23]. In addition, considering the diversity of available regression models, the limited available studies have not provided an implementation methodology to help select appropriate models for a specific data set. In one of the earliest studies, Liu et al. [23] implemented the two most common types of survival analysis models, the accelerated failure time model (AFT) and the Cox proportional hazard model (PHM). These models, used for the estimation of power outages, could be applied as a storm was approaching rather than after damage assessments had been completed. They recommended AFT over Cox PHM, largely because the model output is easier to interpret. Later, Barker and Baroud [5] suggested PHM as an effective approach to estimate the recovery of power distribution. Nateghi et al. [26] compared some statistical methods, including AFT, PHM, data mining techniques, Bayesian additive regression trees (BART), and multivariate additive regression splines, for modeling power outage durations during hurricanes and examined the predictive accuracy of these methods. They compared the out-of-sample predictive accuracy of five distinct statistical models for estimating power outage duration times caused by Hurricane Ivan in 2004. Thereafter, they concluded that BART is the most appropriate model.

However, none of these studies highlighted the fact that the nature of the data (heterogeneity, homogeneity, type of the trend between data, dependency between data, etc.) and the aim of the analysis are two important subjects for modeling covariates and recovery time data. For instance, for complete data, assuming PHM in place of the AFT or vice versa does not have a significant effect on results. However, in the presence of censored data, this assumption has a significant effect on the estimate of the relative importance of the covariates [27]. In other words, there is no single model that is able to accurately analyze different types of sets of recovery data. Moreover, covariate models such as PHM are very sensitive to the missing data and covariates; hence, in the case of the omission of covariates, the result of recovery rate prediction using covariate based models can be erroneous [28]. Morever, for an effective recovery rate estimation analysis, a comprehensive data set—which is complete, meaningful, and structurally accurate—should be readily available. The data should reflect the actual condition, which the infrastructure had at the time of the recovery and all the conditions which the ex-post crew experienced during the recovery process, starting from the time at which the work for recovery arose until the job is finished and the infrastructure returns to normal system function. However, in reality, in many cases the data collection processes are not designed for resilience analysis, especially for recovery rate analysis. Hence, conventional data collection processes failed to collect essential information and data for recovery analysis, such as information regarding the recovery time, ex-ante and ex-post measures, recovery process, logistics, the involved recovery process teams, operating environment conditions, etc.

In general, missing data and information, inappropriate assumption by analysts and lack of applicable methodology are some of the practical challenges for recovery rate analysis using regression models. The selection of a model should be based on the type of data available, the recovery process of the infrastructure system, and the objective of the analysis. Hence, the aim of this paper is to propose a methodology for suggesting an appropriate statistical model, by considering the complex nature of the data, to address the above discussed challenges. The rest of this paper is organized as follows: Section 2 discusses the basic concept and the methodology developed for recovery rate estimation using regression models and the Bayesian approach. In Section 3, the application of this methodology is demonstrated by a case study. Finally, Section 4 provides the conclusions.

2. Recovery rate estimation using regression models

The main steps for the recovery rate estimation using regression models are depicted in Fig. 2. This methodology broadly classifies the data set, based on the nature and availability of data, into three categories including: *i)* no or missing data set, *ii)* homogeneous data set, and *iii)* heterogeneous data set. The first step in the estimation of the