A Probabilistic Approach to Quantify Flood Risk of Chinese Railway Networks

Kai LIU\textsuperscript{1,2}, Ming WANG\textsuperscript{1,3,*} and Wei-hua ZHU\textsuperscript{1}

\textsuperscript{1}Academy of Disaster Reduction and Emergency Management, Ministry of Civil Affairs and Ministry of Education, 100875 Beijing, China
\textsuperscript{2}Key laboratory of Environmental Change and Natural Disaster, Beijing Normal University, 100875 Beijing, China
\textsuperscript{3}State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, 100875 Beijing, China

*Corresponding author

Keywords: Risk assessment, Random forest, Network efficiency, Chinese railway.

Abstract. This paper provides a probabilistic approach for the risk assessment of infrastructure networks subject to floods. The proposed risk analysis method integrates both infrastructure components’ failure probability and the expected infrastructure network efficiency or capacity loss due to component failure. The failure probability of Chinese railway system caused by floods is quantitatively analyzed using Random Forest model and historical disaster events. The expected impact of a component failure on the whole network is evaluated by calculating the efficiency loss due to the deletion of railway links. The flood risk mapping and risk curves of the national railway networks are generated, which can guide infrastructure managers and planners for decisive action of investment in preventative measures to reduce risk.

Introduction

The infrastructure systems are facing great threats posed by natural hazards. The impact of localized, specific infrastructure component failure can propagate through the network and have severe consequences that supersede the local scale. It is therefore important to perform risk analysis of infrastructure system to identify the relative criticality of the network elements likely to be affected. This study establishes a probabilistic approach for risk analysis of infrastructure networks exposed to natural hazards considering both the failure probability of system components and the consequences of failure on the whole network.

Methodology

Failure Probability of Infrastructure Components

In this study, we apply the RF method, introduced by Leo Breiman in 2001\textsuperscript{1}, to calculate the failure probability of railway network component using a historical disaster dataset. The RF algorithm generates a number of random binary classification trees grown using a subset of the original samples via a bootstrap process. Each node of the classification tree is split by the best variable among a set of variables that are randomly selected according to certain criteria. Then, the algorithm combines all trees’ prediction decisions and generates the final prediction by a ‘forest’, voted by all trees. The failure probability $P_i$ of the $i^{th}$ component were determined by the voting results of all trees generated by the RF model as given in equation (1):

$$P_i = \frac{1}{N_p} \sum_{b=1}^{N_p} f_b(V_b, D_b)$$  \hspace{1cm} (1)
Where $f_b$ is the regression tree; $V_b$ and $D_b$ are the collected variables and historical state, respectively. criterion and trains a decision or regression tree $f_b$, $b = 1, \ldots, N_T$ on $V_b, D_b$.

**Failure impact of Infrastructure Components**

In this study, the impact of the component failure is evaluated by network efficiency loss. The network efficiency index $E$ is defined as a composite index between length of the shortest path and flow volume, modifying the distance efficiency index proposed by Newman:

$$E = \sum_i \sum_j \frac{q_{ij}}{L_{ij}}$$

(2)

where $E$ is the average efficiency of an un-degraded network, and $L_{ij}$ is the length of the shortest path between two connected nodes $i$ and $j$; note that when a pair of nodes $i$ and $j$ become disconnected, then $L_{ij} = \infty$; $q_{ij}$ represents flow volume across the shortest path between nodes $i$ and $j$.

Network efficiency upon interruption of edge $(u,v)$ that connects nodes $u$ and $v$ can be evaluated by the following equation:

$$E^{(- (u,v))} = \sum_i \sum_j (1 - \alpha \delta_{ij}) \frac{q_{ij}}{L_{ij}^{(- (u,v))}}$$

where $\alpha$ represents the ratio of flow detours flows when linkages are interrupted: 0 and 1 indicate no redistribution and complete redistribution of interrupted flows, respectively.

**Risk Assessment of Infrastructure Networks**

The infrastructure networks risk is assessed as the probable network efficiency loss due to service interruption, which is calculated by multiplying the probabilities of component failures by their associated consequences. The risk $R_i$ caused by the $i$th link with respect to a hazard $h$ is given by:

$$R_i = P_i \times ERI_i$$

(4)

**Application**

**The Chinese Railway Network**

As of 2016, the Chinese railway system had approximately 2,933 passenger stations. The network topology of the Chinese railway network was constructed according to the following rules: train stations were represented by nodes, and any two stations with a direct flow connection were linked by an edge. Originating, terminating, and transiting were all considered when building the network. Multiple stations in the same city were combined into one node. The final extracted railway network consisted of 529 nodes and 678 edges (Figure 1). In total, 7,115 daily passenger train trips are operated across the network. The daily numbers of passenger trains are assigned to each edge as the flow values.
Failure Probability of Railway network Components

The recorded dataset consisted of 357 rail service disruptions caused by floods, landslides, and debris flows from 1981 to 1998. The railway lines were meshed onto a grid scale of 1 km × 1 km grid cells, and a hazard sample dataset was created according to historical rainfall-induced hazards that physically impacted the railway system in 1980–1998. The samples recording hazards were assigned a value of 1; otherwise, a value of 0 was applied. Four categories of variables that potentially influence the probability of rainfall-induced hazards were considered to generate the RF model; the first category includes four meteorologically related variables: precipitation, rainband, climate zones, and vegetation zones. The second category includes five geomorphologically related variables: elevation, slope, curvature, landform, and land cover. The third category includes two geologically related variables: geology and soil texture. The fourth category includes two hydrologically related variables: watershed and distance from the railway line to the nearest river. The RF algorithm was implemented in the Corelearn package, which was programmed in the R environment. Each grid cell covering rail segments was considered as one sample. The hazard sample dataset was randomly divided into a subset of 60% and 40% for training and validation, respectively. The ratio between hazard and non-hazard data in the training model was fixed at 1:1, a commonly used strategy for imbalanced data samples. In total, 9802 hazard and 9802 non-hazard samples were used as training samples, and 6536 hazard and 31,950 non-hazard samples were used as testing samples.

The prediction accuracy of the RF model was 98.13% for the testing data, implying excellent prediction ability of the model. According to the RF results, the failure map of the national railway system subjected to rainfall-induced hazards was produced using ArcView™ GIS (Figure 2).
Railway Network Efficiency Evaluation and Risk Assessment

In this study, the daily passenger flows were assumed to be proportional to train numbers. Fig. 3a illustrates railway line criticality according to percentage efficiency loss without considering train redistribution from the disrupted lines. The most critical lines are identified as Beijing–Shanghai, Beijing–Harbin, Beijing–Guangzhou, and Shanghai–Zhuozhou. Failure of the most critical link (i.e., Changzhou–Wuxi, located on the Beijing–Shanghai line), was predicted to cause 4.1% efficiency loss to the whole network. The risk to the Chinese railway network caused by a specific link’s failure probability is assessed. Higher-risk links, including Beijing–Shanghai, Beijing–Shenyang, and Jiaozuo–Lanzhou, have higher failure probabilities associated with rainfall-induced hazards combined with high criticality, placing them in the higher-risk category. In contrast, lines with higher criticality but a lower failure probability may exhibit lower overall risk, such as Beijing–Wuhan and Beijing–Jiu long. Because of their reduced influence on whole-network efficiency, lines in Northwest and Southwest China generally have moderate risk levels, although they have the highest failure probabilities. The results indicate the importance of combining component failure probability analysis and the whole-network consequences of failure when performing risk analysis.

Figure 3. a) Criticality map and b) Risk map of the national railway network subject to floods.

Figure 4 presents the risk curves of the Chinese railway network. The exceedence probability (EP) curves denote the probability of exceeding a certain impact given the occurrence of the failure of a railway link. The flood risk curve of China shows that once a railway link fails, there are about 12% probability that more than 0.1% of the network efficiency will be reduced. The risk curves derived from different regions make the results comparable. Results show that the flood risk of the railway networks has a obvious geographic variation. Given a rail link fails in a region, the probability of having efficiency reduction to the whole Chinese railway networks is relatively high in Northeast China, although it is unlikely to have relative high impact (for example, higher than 1% efficiency loss). The risk curve in east China shows that there is a small chance (extreme case) that one link failure may lead to more than 3% efficiency loss in the whole networks. The central and south China have relatively lower risk comparing to other regions. This can be due to, on one hand, the lower probability of flood occurrence (Figure 2), on the other hand, the denser rail networks that make detours more convenient.
Conclusion

In this study, we present a probabilistic approach for evaluation of infrastructure criticality and risk assessment of infrastructure networks subjected to natural hazards. The proposed risk analysis method integrates both infrastructure components’ failure probability and the expected infrastructure network efficiency or capacity loss due to component failure. This method facilitates identification of high-risk network links with both relatively high susceptibility to natural hazards and substantial disruption-related impacts on the whole network.

Acknowledgement

This research was financially supported by the National Science Foundation of China (41401601), Beijing Natural Science Foundation (L150005), and Youth Scholars Program of Beijing Normal University (2014NT2).

References