Pre-disaster Investment Decisions for Strengthening the Chinese Railway System under Earthquakes

Yongze Yan¹, Liu Hong¹,²*, Xiaozheng He³,⁴, Min Ouyang¹,², Srinivas Peeta³,⁵, Xueguang Chen¹,²

¹ School of Automation (Huazhong University of Science and Technology) Wuhan, Hubei 430074, China
² Key Lab. for Image Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan 430074, PR China.
³ NEXTRANS Center (Purdue University), 3000 Kent Avenue, West Lafayette, IN 47906, USA
⁴ Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180, USA
⁵ Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA

Abstract:
This study proposes a framework to determine the investment plan to strengthen a railway system which is subject to earthquake hazard. The proposed framework includes four parts: (1) Construct a two-layer (physical layer and service layer) railway network representation; (2) Generate earthquake scenarios based on historical earthquake data; (3) Formulate an investment optimization model to minimize the expected railway system service loss subjected to an investment budget constraint, where the service loss is quantified based on the affected train flow; (4) Solve the optimization model by using Genetic Algorithm. Taking the Chinese railway system (CRS) as an example, the proposed framework has been applied and the results show that the solution of the proposed framework is more responsive to the earthquake impact on railway system compared to topology-based methods. Note that the proposed framework can also be extended to identify pre-disaster investment plans for other transportation systems under natural disasters.

Keywords: Earthquakes; Railway system; Investment decision; Service loss; Genetic Algorithm

1 Introduction

Railway is one of the most important long-distance transportation modes in many
countries since the eighteenth century. In China, about 2.5 billion passengers and 3.4 billion metric tons of cargo were transported by the Chinese railway system (CRS) in 2015 ("Yearbook", 2016). However, China is a country seriously affected by earthquakes, and is suffered from 33% of the serious continental earthquakes, despite it only covers 7% of the land area in the world. Earthquakes can cause railway trains interrupted or delayed, and decrease the functionality of the CRS. Pre-disaster investment to strengthen the railway system can mitigate the impacts of the consequent unpredictable disruptions (Orcesi and Frangopol, 2011). However, the budget for pre-disaster investment is typically limited. From a pure monetary perspective, it is neither affordable nor acceptable to strengthen the entire railway system, especially for large scale ones such as the CRS whose rail mileage is more than 100,000 KM. Thus, the challenge is how to strengthen a subset of the railway components through investment under limited budget constraint in a planning context to retain railway service after disasters.

This problem belongs to the domain of disaster management, which can be analyzed as a three-stage process: pre-disaster investment to strengthen the system (Hong et al., 2015; Hong et al., 2017; Ouyang and Fang, 2017; Ouyang et al., 2017; Peeta et al., 2010; Wang et al., 2012; Wang et al., 2013), post-disaster adaptive response to minimize the system loss (Corman and D'Ariano, 2012; Corman et al., 2014; Gao et al., 2016; He and Peeta, 2014; He et al., 2015; Jespersen-Groth et al., 2009; Samà et al., 2017; Zhan et al., 2016) and post-disaster re-construction to recover the system (Yan and Shih, 2009). Pre-disaster investment in a railway system plays an essential role in system protection as it entails the need to strengthen the railway components to decrease their failure probabilities under disasters. In order to assess the effect of an investment plan, future disasters should be estimated first, and then the damage state of each affected system component under disasters is estimated respectively with the consideration of pre-disaster investment plan on weak components, the system performance after disasters can be calculated based on the post-disaster states of all components, and finally the effectiveness of an investment plan can be assessed in terms of the
improvement ratio of system performance under all these disasters.

Compared to other natural disasters, such as floods and hurricanes, earthquakes are hard to predict precisely. To estimate the seismic effect on system components, several alternative methods have been developed, including Probabilistic Seismic Hazard Analysis (PSHA), historical earthquakes based analysis, and Monte-Carlo simulation based analysis. The PSHA provides the annual rate of exceeding some level of earthquake ground shaking at a site for a range of intensity levels, which has been widely used for almost 50 years by governments and industries (Mulargia et al., 2017). Poljanšek et al. (2012) used PSHA to estimate the post-disaster state of the components of European gas and electricity networks, and studied their seismic vulnerability from a topological point of view. The historical earthquakes based analysis can also be used to generate scenarios. Tantala et al. (2008) used three different magnitude earthquakes located at a historic epicenter, namely, the M5.2 quake in NYC in 1884, to estimate the seismic effect on the buildings in New York City Metropolitan Region. Bommer et al. (2002) modified historical earthquake catalogue to estimate the seismic loss of different type of buildings in Turkish. The Monte–Carlo simulation based analysis is commonly used to generate synthetic earthquake catalogues in the commercial sector (Crowley and Bommer, 2006). Windeler et al. (2004) used Monte–Carlo simulation method to present seismic risk for residential buildings in the western United State. Among the three methods, the PSHA is useful when analyzing the seismic risk of single facility, but it cannot reflect the correlation of ground motion intensities of different components in a network, so it is unsuitable for seismic analysis of transportation system distributed in a larger area. Inspired by Crowley and Bommer (2006), and considering the large-scale feature of the CRS, the Monte–Carlo simulation based analysis is adopted in this paper.

The post-disaster states of railway components are estimated by fragility analysis under a specific earthquake scenario, and these components include railway stations (Carpinteri et al., 2016), tunnels (Yang et al., 2013), bridges (Banerjee and Chi, 2013; Jia et al., 2013; Mackie et al., 2012; Ramanathan et al., 2015; Siqueira et al., 2014) and
embankments (Li et al., 2009). Masanobu et al. (2000) examined the fragility curves of a bridge by two different approaches: the time-history analysis and the capacity spectrum method. Karim and Yamazaki (2007) developed fragility curves for isolated bridges and compared them with the ones of the non-isolated systems. Argyroudis and Pitilakis (2012) constructed the fragility curves for shallow metro tunnels in alluvial deposits. Maruyama et al. (2010) constructed the fragility curves of expressway embankments in Japan using statistical analysis. For the long railway tracks between two stations in the CRS, the ground motion intensity is varying at different sites along the track, because different sites may have different epicentral distance. To calculate the post-disaster states of the whole railway track, this paper adopts the method used for pipelines (Duenas-Osorio et al., 2007). This method divides a railway track into many small segments, the failure probability of each segment is calculated based on its fragility curve and the ground motion intensity, and the failure probability of the whole railway track is synthesized from that of all its segments. The fragility curves of railway stations and railway segments used in this paper are obtained from HAZUS (FEMA, 2004).

Based on the post-disaster states of all railway components, the system performance under disasters can be calculated. Topological metrics are used to measure railway system performance in some related works (Wang et al., 2012), such as average shortest path length (Zhou and Wu, 2010), network efficiency (Luo et al., 2014), size of the giant component (Osei-Asamoah and Lownes, 2014; Wang et al., 2013) and connectivity (Hong et al., 2017; Ouyang et al., 2015). Purely topological metrics do not consider the flow in the system, different topological metrics may obtain contradictory results when estimating the system performance (Ouyang and Fang, 2017; Zhang et al., 2015). Ouyang et al. (2014) compared the topology based model and real train flow model in railway system for the vulnerability analysis. Passenger flow can also be used to assess transportation system performance, but the related data is often hard to obtain, especially for large transportation system like the CRS. As an alternative, the affected trains are used to estimate the railway system performance in this paper.
These aforementioned studies deal with some aspects of the problem in this paper, but few of them integrate the above models for different aspects as a comprehensive framework to solve the problem. Peeta et al. (2010) proposed a stochastic program based method to make a pre-disaster investment decision for strengthening highway system under earthquake risk, but this work did not consider the earthquake scenarios and component fragilities, with the link failure probabilities given in advance. Hong et al. (2015) proposed a methodology to quantitatively assess the railway system vulnerability under floods using historical data, but they did not consider the budget constraint for pre-disaster investment. To the best of the authors’ knowledge, there is little work on system level pre-disaster investment analysis for large-scale railway systems under earthquakes with the consideration of the investment budget constraint.

This paper develops a general framework for pre-disaster investment problem of railway systems under earthquakes and applies the framework to the CRS. In the framework, the railway system is modeled as a two-layer network, which is useful to clearly estimate the damage states of railway components and assess the whole system loss by the affected train flow respectively; then a pre-disaster optimal investment plan problem is modeled as a binary nonlinear stochastic programming, and finally a Self-Adaptive Genetic Algorithm is used to solve it.

The remainder of the paper is organized as follows. Section 2 describes the problem, including the two-layer network representation of the railway system, generation of earthquake scenarios, assessment of railway system service loss, and effectiveness analysis of pre-disaster investment. Section 3 formalizes the optimization model and develops a solution method for the model. Section 4 applies the proposed framework to the CRS and compares the optimal investment plan with topology-based methods to investigate the tractability and effectiveness of the proposed framework. Section 5 concludes this study and discusses potential directions for future research.

2 Problem description

The Chinese railway system suffered great seismic loss in the past, each year the government (CRS belong to the Chinese Central Government) spends a large amount
of money to strengthen the railway system. To improve the effectiveness of these investments, the problem is how to find an optimal pre-disaster investment plan for the CRS to remain operational after earthquake hazards with the consideration of a limited investment budget. The optimal pre-disaster investment plan means to strengthen a subset of railway components to minimize the expected system loss under a budget constraint. To calculate the expected system loss, earthquake scenarios are needed, which are used to measure the effectiveness of different investment plans. A two layer network including a physical layer and a service layer is adopted to represent the railway system (Kurant and Thiran, 2006), through which the earthquake damage to railway components can be analyzed in the physical layer and the railway system loss can be calculated in the service layer according to the damage state of components in the physical layer. Based on the earthquake scenarios and the railway system model, the effectiveness of an investment plan can be measured by the system loss reduction ratio after investment under these earthquake scenarios. There are five parts in this section to describe the problem: introduction of notation system, two-layer network representation of the CRS, earthquake scenario generation, service loss calculation and investment effectiveness analysis.

2.1 Notation system

All the notations used in this paper are listed below to describe and understand the problem more conveniently.

- \( G = (V, E) \) The physical layer of the CRS network
- \( V = \{v_i\}_{i=1}^{n} \) The set of nodes in the physical layer
- \( E = \{e_j\}_{j=1}^{m} \) The set of edges in the CRS network
- \( Q = (V, L) \) The service layer of the CRS network
- \( L = \{l\} \) The set of links in the service layer
- \( T = \{t_i\}_{i=1}^{n} \) The set of all trains in the railway system, \( t \) denotes train
- \( F_v(t) \) The stations passed by train \( t \)
- \( F_e(t) \) The links in the service layer passed by train \( t \)
- \( F_e(l) \) The edges sequence in the physical layer mapped for the link \( l \) in the service layer
- \( \xi_v \) The post-disaster state of the node \( v \)
- \( g_k \) The k-th segment in the edge \( e \)
- \( \xi_e \) The post-disaster state of the edge \( e \)
2.2 Two-layer network representation of railway system

In this study, a railway system is modeled as a two-layer network, including a physical layer and a service layer. In the physical layer, railway stations are represented
as nodes while railway tracks between stations are represented as edges. Without loss of generality, when more than one railway track connects two stations, only one edge is used to represent these tracks in the physical layer (Wei et al., 2015). Differing from edges in the physical layer, links in the service layer represent the train services between stations. If two stations are served by a train without stops in between, there is a link connecting these two stations in the service layer. The physical layer contains the topological (connections between stations and tracks) and geographical (locations of stations and tracks) information of the railway system, while the service layer contains flow information on how trains serve passengers.

This two-layer network representation is particularly suitable for analyzing railway system performance under disasters because the system performance regarding train flow can be quantified at the service layer from the flow (passengers) perspective, while the damage on the railway tracks under disasters is assessed at the physical layer. When an earthquake affects a railway track, the corresponding edge in the physical layer may be disrupted and the trains that pass through this edge should be interrupted. The information on which trains pass through the disrupted edge in the physical layer is determined by the mapping from the train service in the service layer to the railway edges in the physical layer. The formulation of the two-layer network is described below:

The physical layer is modeled as $G = (V, E)$, where $V = \{v_1\}_{i=1}^{\vert V \vert}$ denotes the set of nodes, and $E = \{e_i\}_{i=1}^{\vert E \vert}$ denotes the set of edges in the physical layer. $V$ and $E$ can be obtained from the railway system map, for example the CRS map (Gaotie.cn, 2010). The nodes in both the physical layer and the service layer are identical. The service layer is denoted as $Q = (V, L)$, where $L = \{l_i\}_{i=1}^{\vert L \vert}$ denotes the set of links in the service layer. Denote the set of all trains in the railway system as $T = \{t_i\}_{i=1}^{\vert T \vert}$. The stations and links in the service layer passed by train $t_i$ are denoted as $F_v(t_i)$ and $F_l(t_i)$ respectively, $F_v(t_i) \subset V$, $F_l(t_i) \subset L$. $L$, $T$ can be obtained directly from the railway system time table ("Timetable", 2010). Each link $l_i$ in $L$ can be mapped to a
subset of the edges in $E$, but this information cannot be obtained directly from the map or time table of railway system, because the time table only lists the stations a train stops but not the stations it passes. Similarly, $F_v(t_i)$ and $F_l(t_i)$ also cannot be obtained from the time table directly. In this study, a link $l_i$ in the service layer is mapped to a sequence of edges in the physical layer, represented as $F_e(l_i) \subset E$, where $F_e(l_i)$ is obtained from the shortest path in the physical layer between the start and end nodes of $l_i$. Using the two-layer network and the method mentioned above, the mapping between the trains in the service layer and the stations and edges in the trains routes in the physical layer can be established, which is used to determine the affected trains when some railway components are damaged in Section 2.4.

Fig.1 and Table.1 give a simple example to show the application of the two-layer network approach, and it is also used in Section 2.4 to illustrate the calculation process for the states of affected edges under earthquake.

![Figure 1: A simple example of two-layer network for railway system](image)

**Table 1**

The objects in the railway system.

<table>
<thead>
<tr>
<th>Train set</th>
<th>{t_1, t_2, t_3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train route information in the service layer (from time table)</td>
<td>$t_1$: stations {v_1, v_2, v_3}, links {l_1, l_2}  \ $t_2$: stations {v_6, v_2, v_7}, links {l_3, l_4}</td>
</tr>
</tbody>
</table>
Table 1 shows that the station set and the edge set in the physical layer can be obtained from the railway map, while the train service information and the link set in the service layer can be obtained from the train timetable. However, the mapping between the links in the service layer and the edges in the physical layer, as well as the train route information in the physical layer, cannot be obtained directly. In this paper, through searching the shortest path between a link’s start and end stations in the physical layer, the mapping between links and edges can be established. For example, the link \( l_1 \) in the service layer can be mapped to two edges \( e_1 \) and \( e_3 \) in the physical layer. Then the train route information in the physical layer can be obtained from the train service information in the service layer, and the mapping between links and edges.

For example, the route of the train \( t_3 \) in the physical layer includes stations \( \{v_4, v_2, v_5, v_7, v_8\} \) and edges \( \{e_3, e_4, e_6, e_7\} \), where stations \( \{v_2, v_7\} \) are not included in \( t_3 \) in the service layer. According to the above route information, if any components in the route of \( t_3 \) in the physical layer is damaged under an earthquake, the train \( t_3 \) will be interrupted.

In Fig. 1, when an earthquake occurs, \( v_7, e_6 \) and \( e_7 \) are impacted, the interrupted

<table>
<thead>
<tr>
<th></th>
<th>( t_3 ): stations ( {v_4, v_5, v_8} ), links ( {l_5, l_6} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station set in the network (from map)</td>
<td>( {v_1, v_2 \ldots v_8} )</td>
</tr>
<tr>
<td>Link set in the service layer (from time table)</td>
<td>( {l_1, l_2 \ldots l_6} )</td>
</tr>
<tr>
<td>Edge set in the physical layer (from map)</td>
<td>( {e_1, e_2 \ldots e_7} )</td>
</tr>
<tr>
<td>Mapping between links in the service layer and edges in the physical layer (from the shortest path)</td>
<td>( l_1: {e_1, e_3} ); ( l_2: {e_4, e_5} ); ( l_3: {e_2, e_3} ) ( l_4: {e_4, e_6} ); ( l_5: {e_3, e_4} ); ( l_6: {e_6, e_7} )</td>
</tr>
<tr>
<td>Train route information in the physical layer (from the mapping between links and edges)</td>
<td>( t_1): stations ( {v_1, v_4, v_2, v_3} ), edges ( {e_1, e_3, e_4, e_5} ) ( t_2): stations ( {v_6, v_4, v_2, v_5, v_7} ), edges ( {e_2, e_3, e_4, e_6} ) ( t_3): stations ( {v_4, v_2, v_5, v_7, v_8} ), edges ( {e_3, e_4, e_6, e_7} )</td>
</tr>
</tbody>
</table>
train set is \{t_2, t_3\}. When a flood occurs, \( v_1 \) and \( e_1 \) are impacted, the interrupted train set is \{\( t_1 \)\}. If the railway system is affected by the earthquake and the flood simultaneously, which is a multi-sites disaster scenario, the impacted train set is the union of those affected trains under each single disaster scenario, \{\( t_1, t_2, t_3 \)\}. This means the two-layer model can also be used for multi-sites disaster analysis. Further, the data of bridges and tunnels’ location in the railway system is not currently available, so bridges and tunnels are not considered in this paper. When the related data is available, the bridges and tunnels can also be described as special nodes or edges, after obtaining their seismic fragility curves, the proposed framework is still valid.

### 2.3 Earthquake scenario generation

This study uses historical earthquake data and seismic belts to generate future earthquake scenarios. The Chinese seismic belts distribution map, including 25 seismic belts, is used to characterize the research area (Wang, 2013). The Monte–Carlo simulation based analysis is adopted to generate earthquake scenarios in these seismic belts, where the earthquake probability is estimated by Poisson (time-independent) process based on historical earthquake data (Crowley and Bommer, 2006). The historical earthquake data used in this study is obtained from Advanced National Seismic System (ANSS) catalogue (NCEDC, 2013), which includes the location, magnitude and date of each past earthquake event. In total, 12,750 historical earthquake events happened in China with magnitude greater than 4.5 on the Richter scale are used for earthquake scenarios generation, because only earthquakes with magnitude greater than 4.5 can cause damage to the railway system in the study context according to the hazard earthquake catalogue provided by the China Earthquake Data Center (2015). The 25 seismic belts in China are shown as shaded areas in Fig. 2. Because the railway network in Taiwan is isolated from that in the mainland, the Taiwan seismic belt labeled as the 25th is not discussed in this study.
This paper generates the earthquake scenario for a time of interest. This earthquake scenario may include multiple earthquakes, which are described by their seismic epicenters and their magnitudes. Let $h$ denotes a scenario and $\eta$ denotes a typical earthquake in $h$. $\eta$ is generated using the following procedures:

(i) For each year during the period, and for each seismic belt, a random number, $P_{random}$, between 0 and 1, is generated from a uniform distribution;

(ii) If $P_{random}$ is less than the annual probability $P_{min}$ of events above a chosen minimum seismic magnitude $M_{min}$ (which is set as 4.5 in this study) and larger than the annual probability of events $P_{max}$ below a chosen maximum magnitude $M_{max}$ (which is set as 8.5, as it is the largest magnitude in the historical data), then it is assumed that there will be an earthquake happened in that seismic belt in that year. Otherwise, there will be no earthquake. Here, $P_{min}$ and $P_{max}$ are obtained from the annual rate $N_{min}$ and $N_{max}$ according to the Poisson model:

\begin{align*}
N_{min} &= 10^{c_1-c_2 M_{min}} \quad (1) \\
N_{max} &= 10^{c_1-c_2 M_{max}} \quad (2) \\
P_{min} &= 1 - \exp(-N_{min}) \quad (3)
\end{align*}
\[ P_{\text{max}} = 1 - \exp (-N_{\text{max}}) \]  

(4)

where the parameters \( c1 \) and \( c2 \) are estimated for each seismic belt using least square regression based on the historical earthquake data. The estimated values of \( c1 \) and \( c2 \) in each seismic belt in China are listed in the Appendix A.

(iii) When an earthquake is generated, its magnitude \( M \) is determined by \( P_{\text{random}} \) and the recurrence relationship:

\[ N_{\text{random}} = -\ln (1 - P_{\text{random}}) \]  

(5)

\[ M = \frac{c1 - \log(N_{\text{random}})}{c2} \]  

(6)

(iv) The epicenter of each generated earthquake in a seismic belt is randomly chosen within that belt through the ArcGIS technique.

(v) Based on the magnitude and epicenter of the generated earthquake, the next step is to determine the ground-motion distribution of the earthquake. This study uses peak ground acceleration (PGA) and peak ground deformation (PGD) attenuation relationship to estimate the earthquake ground-motion at any location around the epicenter, which is then used to estimate the post-disaster state of railway stations and railway edges in the physical layer (Wilson et al., 2015).

This study uses the following equation proposed by Toro et al. (1997) to calculate the median PGA:

\[ \ln(\text{median PGA}) = 2.20 + 0.81 \times (M - 6) - 1.27 \times \ln(\sqrt{\text{Repi}^2 + 9.3^2}) + 0.11 \times \max \left[ \ln \left( \frac{\text{Repi}}{100} \right), 0 \right] - 0.0021 \sqrt{\text{Repi}^2 + 9.3^2} \]  

(7)

where \( \text{Repi} \) denotes the epicentral distance of the site under the seismic event, and \( M \) denotes the seismic event magnitude. Note that to get the final PGA in each site, it needs to further consider the inter-event and intra-event standard deviations (Crowley and Bommer, 2006), and for illustrative purpose, this study simply uses this median PGA as the final PGA for further railway damage analysis at each site. The PGD is calculated according to the work by Luo et al. (2010).
2.4 Modelling of railway system service loss

The railway system service loss under an earthquake is measured through the affected train flow in the service layer, which is determined by the post-disaster states of the affected railway components in the physical layer.

2.4.1 Post-disaster component states in the physical layer

The post-disaster states of all railway components (including stations and edges) are determined by these components’ fragility curves and the ground motion intensity at their sites obtained in Section 2.3. In this paper, the railway components’ fragility curves are obtained from HAZUS (Argyroudis et al., 2013; FEMA, 2004).

The failure probability of an affected railway station is calculated based on the PGA at the station site and its fragility curve in HAZUS, and the fragility curve is dependent on the building type of the railway station. Due to the lack of survey data on building type of all railway stations in the CRS and most of them are concrete moment frames, the fragility curve for concrete moment frame in HAZUS is used as the fragility curve of all railway stations in the CRS, where the fragility curve is modeled as a lognormal-distributed function that specifies the building failure probability under a given level of PGA value. The post-disaster state of a node (station) $v$ is denoted as $\xi_v$, where $\xi_v = 1$ means $v$ is damaged and $\xi_v = 0$ means $v$ is not damaged. The failure probability of $v$ under a given earthquake $\eta$, denoted as $P(\xi_v = 1|\eta)$, is obtained from the fragility curve in HAZUS.

The post-disaster state of a railway edge under earthquake is determined by the PGD around the edge and the earthquake resistance ability of the edge. Because the detailed information of earthquake resistance ability of each edge is not available, it is assumed that all the edges are homogeneous and their post-disaster state is only related to the PGD value. The railway edges usually have long distance, so the PGD value of different part of an edge is different. Hence, the PGD at any site along the railway edge is not appropriate to be used to calculate the post-disaster state of the whole edge. In this study,
a railway edge is divided into small segments of equal length, the failure probability of each segment is calculated individually based on the PGD value at the middle site of the segment, and the failure probability of the whole edge is the integration of its segments’ failure probabilities. Note that for an infrastructure edge, like a pipeline (Duenas-Osorio et al., 2007), with a long length, its seismic fragility is usually modeled by dividing it into many small segments, and the failure probability of each segment is characterized by a repair rate, which corresponds to the number of repairs per kilometer of brittle links, and then failure probability of the whole edge can be modeled as the integration of the failure probability of all segments. However, in HAZUS internal file, it does not provide the repair rate for railway tracks, and directly provides their fragility curves in terms of PGD, so this study simply uses the following method to calculate the failure probability of railway edges.

Let $g_k$ denotes a segment of an edge $e$, $e = \{g_k\}_{k=1}^{K_e}$, and $k$ denotes the segment index, $K_e$ is the number of segments in edge $e$. For example, in Fig. 1, the edge $e_7$ between node $v_7$ and $v_8$ is divided into 4 segments with equal length: $e_7 = \{g_1, g_2, g_3, g_4\}$. Given an earthquake $\eta$, the failure probability of the edge $e$ under $\eta$ can be expressed as:

$$P(\xi_e = 1 | \eta) = 1 - \prod_{k=1}^{K_e} (1 - P(\xi_{g_k} = 1 | \eta))$$

(8)

where $\xi_e$ denotes the post-disaster state of the edge $e$, $\xi_{g_k}$ denotes the post-disaster state of segment $g_k$, and the failure probability of $g_k$ under $\eta$, denoted as $P(\xi_{g_k} = 1 | \eta)$, which is obtained from the fragility curve in HAZUS.

The post-disaster state of the physical layer $G$ is an integration of the post-disaster states of all the nodes and edges in the physical layer, denoted as $\xi_G = \{\xi_V, \xi_E\}$, where $\xi_V = \{\xi_v | v \in V\}$ and $\xi_E = \{\xi_e | e \in E\}$.

For illustrative purpose, this study assumes that the post-disaster states of all components in the physical layer are independent. Then, the occurrence probability of $\xi_G$ under earthquake $\eta$ is calculated by:
\[ P(\xi_G|\eta) = \prod_{v \in V} (\xi_v \cdot P(\xi_v = 1|\eta) + (1 - \xi_v) \cdot P(\xi_v = 0|\eta)) \] 
\[ \prod_{e \in E} (\xi_e \cdot P(\xi_e = 1|\eta) + (1 - \xi_e) \cdot P(\xi_e = 0|\eta)) \]  
\[ P(\xi_v = 0|\eta) = 1 - P(\xi_v = 1|\eta) \]  
\[ P(\xi_e = 0|\eta) = 1 - P(\xi_e = 1|\eta) \]  

2.4.2 Service loss estimation

The railway system loss under an earthquake is measured through the affected train flow, that is the number of the affected trains. The affected trains under earthquake train are determined by the damage states of nodes and edges in the physical layer.

The train \( t \) is interrupted if at least one of the railway components it passed by is damaged in the post-disaster network. Let \( \xi_{t,\xi_G} \) denotes the post-disaster state of the train \( t \) under the network state \( \xi_G \). \( \xi_{t,\xi_G} \) is a binary variable. \( \xi_{t,\xi_G} = 0 \) means the train \( t \) is not interrupted, and \( \xi_{t,\xi_G} = 1 \) means the train \( t \) is interrupted. Then the railway system service loss \( D(\xi_G) \) under the network damage state \( \xi_G \) can be expressed as

\[ D(\xi_G) = \sum_{t \in T} \xi_{t,\xi_G} \]  
\[ \xi_{t,\xi_G} = \max(\xi_{F_v(t),\xi_G} \cup \xi_{F_i(t),\xi_G}) \]  
\[ \xi_{F_v(t),\xi_G} = \{\xi_v|v \in F_v(t), \xi_G\} \]  
\[ \xi_{F_i(t),\xi_G} = \{\xi_{l,\xi_G}|l \in F_i(t)\} \]  
\[ \xi_{l,\xi_G} = \max(\xi_e|e \in F_e(l), \xi_G) \]

where \( \xi_{F_v(t),\xi_G} \) denotes the post-disaster states of all nodes passed by the train \( t \) in the network state \( \xi_G \). \( \xi_{F_i(t),\xi_G} \) denotes the post-disaster states of all links passed by the train \( t \) in the network state \( \xi_G \). \( \xi_{l,\xi_G} \) denotes the post-disaster state of the link \( l \).
Based on the occurrence probability of the network state \( \xi_G \) under an earthquake based on Eq. (9), the railway system service loss under earthquake \( \eta \) is expressed as:

\[
D(\eta) = \sum_{\xi_G \in \Xi} D(\xi_G) * P(\xi_G | \eta)
\]

where \( \Xi \) denotes the set of all possible \( \xi_G \), \( \Xi = \{ \xi_G \} \).

The railway system service loss under the period earthquake scenario \( h \) is calculated by:

\[
D(h) = \sum_{\eta \in \mathcal{H}} D(\eta) = \sum_{\eta \in \mathcal{H}} \sum_{\xi_G \in \Xi} D(\xi_G) * P(\xi_G | \eta)
\]

The railway system service loss under a period earthquake scenario is the sum of system the railway service loss under all the earthquake events in the scenario. The railway system service loss calculation process can be summarized as the following three steps:

Step 1: Calculate the ground motion intensities (PGA, PGD) at the sites of all railway components affected by the earthquake \( \eta \).

Step 2: Estimate the failure probabilities of the railway components based on their fragility curve and the ground motion intensities calculated in step1.

Step 3: Use Monte Carlo simulation to determine the affected trains:

Step 3.1: Determine all the failed railway components under earthquake, which form the poster-disaster state of the physical layer.

Step 3.2: Determine the affected train set based on the failed railway components under earthquake, the number of trains in the set is used to measure the service loss.

### 2.5 Effectiveness analysis of investment plan

The failure probability of a railway component will decrease after being invested. In this study, a parameter \( \alpha \in (0,1) \), called maintenance degree, is defined to represent the
decrease ratio of the failure probability of an invested railway component.

An investment plan is denoted as \( y \in Y \), where \( Y \) is the set of all investment plans. \( y = (y_V, y_E) \), where \( y_V = \{y_v | v \in V\} \) denotes the investment states of stations and \( y_E = \{y_e | e \in E\} \) denotes the investment states of edges. If stations \( v \) or edge \( e \) is chosen to be strengthened, then set \( y_v = 1 \) or \( y_e = 1 \); otherwise set \( y_v = 0 \) or \( y_e = 0 \).

The investment cost of railway stations and unit length of tracks may vary at different locations in the railway system. Because the detailed data is not available, it is assumed that all railway stations have the same investment cost, and a unit length of the railway track (edge) also has the same cost. This study adopts the replacement values of transportation system components in HAZUS as the investment cost of a railway station and a unit length of edge, denoted as \( c_v \) and \( c_u \) respectively. In this study, it is assumed that with the increase of the maintenance degree, the investment cost will increase correspondingly. The relationship between investment cost and maintenance degree is expressed as increasing functions \( c_v = f_{\text{invest}}(c_v, \alpha) \) and \( c_u = f_{\text{invest}}(c_u, \alpha) \).

If a node \( v \) is strengthened by the investment plan \( y \), its new failure probability under earthquake \( \eta \) is calculated by Eq. (19):

\[
P_y(\xi_v = 1 | \eta) = (1 - \alpha) \times P(\xi_v = 1 | \eta)
\]

where \( P(\xi_v = 1 | \eta) \) is obtained in Section 2.4.1.

If an edge \( e \) is strengthened by the investment plan \( y \), its new failure probability under earthquake \( \eta \) is calculated by Eq. (20) and Eq.(21):

\[
P_y(\xi_e = 1 | \eta) = 1 - \prod_{k=1}^{K_e} (1 - P_y(\xi_{g_k} = 1 | \eta))
\]

\[
P_y(\xi_{g_k} = 1 | \eta) = (1 - \alpha) \times P(\xi_{g_k} = 1 | \eta)
\]

The occurrence probability of \( \xi_G \) with investment plan \( y \) under earthquake \( \eta \) is calculated by Eq.(22):

\[
P_y(\xi_G | \eta) = \prod_{v \in V} \left( \xi_v \times P_y(\xi_v = 1 | \eta) + (1 - \xi_v) \times P_y(\xi_v = 0 | \eta) \right)
\]
\[
* \prod_{e \in E} \left( \xi_e \cdot P_y(\xi_e = 1|\eta) + (1 - \xi_e) \cdot P_y(\xi_e = 0|\eta) \right)
\]  
(22)

\[
P_y(\xi_v = 0|\eta) = 1 - P_y(\xi_v = 1|\eta)
\]  
(23)

\[
P_y(\xi_e = 0|\eta) = 1 - P_y(\xi_e = 1|\eta)
\]  
(24)

The railway system service loss under the investment plan \( y \) for the earthquake scenario \( h \) is calculated as:

\[
D_y(h) = \sum_{\eta \in h} D_y(\eta)
\]  
(25)

\[
D_y(\eta) = \sum_{\xi_G \in \mathcal{G}} D(\xi_G) \cdot P_y(\xi_G|\eta)
\]  
(26)

3 Optimization model

In this section, a pre-disaster investment optimization model is proposed to identify a subset of railway stations and edges in the physical layer to be strengthened under limited budget, to minimize the expected service loss under earthquake hazards.

3.1 Problem formalization

An investment plan is a subset of railway components which are chosen to be strengthened, and if a railway component \( v \) or \( e \) is strengthened, its failure probability will decrease. Each railway component has an investment cost, \( c_v \) for node \( v \) and \( c_u \cdot Len(e) \) for edge \( e \), and the sum of the investment cost \( b_y \) should be bounded by the total budget constraint \( B \). \( b_y \) can be calculated as below:

\[
b_y = \sum_{e \in E} c_u \cdot Len(e) \cdot y_e + \sum_{v \in V} c_v \cdot y_v
\]  
(27)

where \( Len(e) \) denotes the length of edge \( e \).

Then, the pre-disaster investment optimization model can be represented as:

\[
\min_{y \in Y} \exp_{\eta \in H} \left[ \sum_{\eta \in h} \sum_{\xi_G \in \mathcal{G}} D(\xi_G) \cdot P(\xi_G|\eta) \right]
\]  
(28)

Subjected to: (12) - (16); (19) - (27)
\[ b_y \leq B \] (29)
\[ y_v \in \{0,1\}, \quad \forall v \in V \] (30)
\[ y_e \in \{0,1\}, \quad \forall e \in E \] (31)

where the maintenance degree \( \alpha \) and the total budget \( B \) are two parameters of the model.

The objective function in the investment optimization model is to minimize the expected railway system service loss under all earthquakes scenarios. As shown by Eq. (28), the pre-disaster investment optimization model can be regarded as a binary nonlinear stochastic programming model.

Note that, for a nation-wide system as CRS, the scale of problem (Eq. (28)-(31)) is large, requiring heavy computation to solve for optimality. First, the total number of binary decision variables of the problem is large that is the sum of the total numbers of nodes and links, i.e., \(|V| + |L| = 400 + 505 = 905\) for CRS. Therefore, the search space contains \( 2^{905} \approx 2.7 \times 10^{272} \) candidate solutions. Second, problem (Eq. (28)-(31)) is NP-hard, because it can be regarded as a special mixed integer nonlinear program that is known NP-hard (Kannan and Monma, 1978). Third, the objective function (28) does not have a closed form. The computation of \( P(\xi | \eta) \) depends on the seismic belt where the earthquake event \( \eta \) occurs. No well-defined distributions could describe this random variable. Thereby, the computation of expectation relies on simulation. This precludes a fast assessment of solution quality to design efficient solution algorithm.

### 3.2 Solution method

In the literature, many solution methods have been proposed to solve the stochastic mixed integer programming (SMIP) problems. The basic concept behind these methods is to decompose the original SMIP into a two-stage (or multiple-stage) program (Laporte and Louveaux, 1993; Schultz, 1993), where the first stage is formulated as a discrete mixed integer program (MIP) and the second stage is a continuous stochastic program (SP) (Birge and Louveaux, 2011). The first-stage MIP contains the collection of integer decision variables while the second-stage SP
evaluates the consequence, commonly formulated as the expectation of the performance metric, given the decisions made in the first stage. Recent algorithmic advances in solving SMIP focus on applying decomposition schemes (Infanger, 1992; Mulvey and Ruszczyński, 1995; Rockafellar and Wets, 1991; Trapp et al., 2013) to solve the second-stage stochastic program efficiently, whereas the theoretical properties developed for MIP have been applied to the first stage program. For instance, Carøe and Tind (1998) proposed an L-shaped decomposition to decompose the original SMIP into two stages, where the cutting plane and branch and bound techniques are applied to solve the first stage MIP sub-problem and the second stage SP is assumed to be linear with stochastic right-hand-side. As noted by many studies (Sen, 2005), the existing solution methods typically rely on a special structure, e.g. an embedded mixed integer linear program, to achieve a good computational efficiency. These methods are not applicable to solve for optimality if the SMIP formulation contains discontinuous, nonsmooth, and nonconvex functions. In addition, these methods do not address the computational difficulties associated with the inherent combinatorial structure, which renders the first stage MIP intractable (Sen, 2005). Therefore, it is important to develop an efficient solution method for the nonlinear stochastic mixed integer programming developed in this study.

It is difficult to solve the problem Eq. (28)-(31) efficiently, and the solution for the problem can be represented as a set of investment decisions on the railway components. This paper adopts a general algorithm, Genetic Algorithm, as the solution algorithm.

As an evolutionary algorithm, Genetic Algorithm is inspired by population genetics, and evolution at the population level (Brownlee, 2012), which is widely used for integer and mixed integer optimization problems (Deep et al., 2009). Ferreira and Bretas (2012) used Genetic Algorithm to solve the nonlinear binary programming model for optimal allocation of sectionalizing switches in distribution systems. Hence, we propose a Genetic Algorithm to solve this pre-disaster investment optimization model.

Because the large scale of the CRS increases the complexity of the model, the model solution speed should also be considered. Some researchers use new sampling
technique such as Opposition-Based Learning (OBL), Opposite-Center Learning (OCL) for convergence speed-up of meta-heuristic optimization algorithms (Dhahri and Alimi, 2010; Tizhoosh, 2005; Xu et al., 2015). Shomali et al. (2017) used Monte-Carlo parallel simulation to speed up computations. At the same time, Self-Adaptive Genetic Algorithm is widely used in optimization (Apolinar and Rodriguez, 2017; Zhou et al., 2015), it can overcome the premature convergence to the local optimum in traditional Genetic Algorithm (Mahmoodabadi and Nemati, 2016). Lu et al. (2015) used Self-Adaptive Genetic Algorithm to estimate Jiles–Atherton (JA) model parameters. To speed up the model solving and avoid premature convergence to the local optimum, parallel simulation and Self-Adaptive Genetic Algorithm are used to solve the proposed optimization model in this paper.

Self-Adaptive Genetic Algorithm is realized by dynamically adjusting the crossover and mutation operator parameters. The crossover probability and mutation probability are self-adjusted based on the local best fitness value of the last generation and the global best fitness value of all generations. The crossover probability and mutation probability are determined as follow:

\[
P_{ga}(t + 1) = \begin{cases} 
\max(P_{ga}(t) \times SC_{re}, P_{ga}^{min}) \quad & SC_{re} \in (1,2) \quad \text{if} \quad f_{local}(t) > f_{global} \\
\min(P_{ga}(t) \times SC_{in}, P_{ga}^{max}) \quad & SC_{in} \in (0,1) \quad \text{if} \quad f_{local}(t) < f_{global}
\end{cases}
\]

Because the adjust strategies of crossover probability and mutation probability are analogy, they are denoted as \(P_{ga}\) uniformly. \(P_{ga}(t + 1)\) and \(P_{ga}(t)\) denote the crossover probability or mutation probability of the next generation and the last generation respectively. \(f_{local}(t)\) denotes the best fitness value of the last generation, \(f_{global}\) denotes the global best fitness value, and these two variables decide how to adjust \(P_{ga}\) used in the next generation. \(SC_{re}\) and \(SC_{in}\) denote the reducing and increasing scale factor respectively, which determine the scale of the crossover and mutation probability adjustment. \(P_{ga}^{min}\) and \(P_{ga}^{max}\) are two preset parameters, denote the minimum and maximum crossover or mutation probability, and these two preset parameters provide the value range of the corresponding probability. In the solving
process, the most time-consuming part is pre-disaster investment effect estimation for each investment plan in the population, the dynamically changing $P_{ga}$ can better optimize the population size to reduce the solving time and avoid premature convergence.

4 Numerical experiment

In this section, the specific data of the CRS is used in a numerical experiment to demonstrate the application of the proposed framework. We also illustrate the quality of the experiment results by comparing the pre-investment plan obtained from the proposed framework and those from topology-based methods.

4.1 Experiment setup

As only the data of passenger trains is available at present stage, this paper only considers the passenger trains in the CRS. The CRS has 4,197 passenger trains which served for 400 railway stations (2010). There are 400 nodes and 505 edges in the physical layer (Ouyang et al., 2017), shown in Fig. 3 (a), and 400 nodes and 834 links in the service layer, as shown in Fig. 3 (b). From the Fig. 3 (b), the link densities in the eastern and southern parts of China are higher than those of other parts, because these areas usually have high GDP and population.
The rail segment length in the physical layer is set as 10 km. There are 6126 segments in the physical network. Due to limited available seismic hazard data in China, the parameters of fragility curves are obtained from HAZUS Technical Manual (2004).

In the experiment, the two functions, which are used to represent the relationships between investment cost and maintenance degree of railway components, $c_v = f_{\text{invest}}(c_{v0}, \alpha)$ and $c_u = f_{\text{invest}}(c_{u0}, \alpha)$ are simplified as $c_v = c_{v0} \ast (1 + \alpha)$ and $c_u = c_{u0} \ast (1 + \alpha)$ respectively. The maintenance degree $\alpha$ is set as \{0.1, 0.2, 0.3, \}.
0.4), and the budget is set as \{\$1 million, \$3 million, \$5 million, \$7 million, \$9 million\}. The investment cost of a railway station and one-unit length of edge are set as \$2000 and \$1500 respectively according to the HAZUS internal files.

The parameters of Genetic Algorithm are set as follows:

The number of generations \( N_{GA} \) is set as 20 and the number of populations \( N_{pop} \) is set as 100 according to the convergence performance of Genetic Algorithm, which will be described in Section 4.2. \( N_{gene} \) is set as 905 to keep consistent with the number of stations and edges in the physical layer. The evolution parameters of Genetic Algorithm are self-adaptive with the convergence of the service loss, which described in Section 3.2. The number of earthquake scenario \( N_{EQ} \) is set as 1000. The period \( T_{int} \) is set as 1 year because the train timetable of the CRS changes rapidly.

The most time-consuming computation in the experiment is to obtain railway components’ failure probabilities, including three steps: distance calculation, PGA or PGD calculation and failure probability calculation. In the experiment, for one investment plan in one generation of the Genetic Algorithm, the above three steps need to be run about \( 9.8 \times 10^{10} \) times. Other operations, like identifying the affected trains, also need to be run about \( 3.8 \times 10^{9} \) times. There are hundreds of investment plans in one generation, so the computation cost is very large.

Note that, the tunnels and bridges along the railway also can be modeled as special nodes or edges. The complexity of the problem will be much larger for practical engineering application. The length of segment in the physical layer is an issue that deserves further discussion, this issue also is a common problem for all line-shaped components under earthquake (Duenas-Osorio et al., 2007). The selection of the current value is a balance of calculation accuracy and simulation time. The smaller segment will increase the resolving time of the problem.

### 4.2 Experiment results

To validate the algorithm, the effect of the optimal investment solution of the
The proposed framework is compared with that of the enumeration approach through a simple example. Considering the heavy computation load of the enumeration approach, in this example, the maintenance degree is set as 0.4 and the budget is set as $5600 so that at most two components can be strengthened, which makes the enumeration approach feasible.

The convergence of the proposed solution method is shown in Fig. 4, and the final service loss is 14.96 trains. The minimal service loss obtained from enumeration approach is 14.92 trains. The service loss decrease ratio of the investment plan obtained through the proposed solution method is 13.8%, while that of the enumeration approach is 14.1%. It implies that the result obtained using our method can reach 97.9% of the global optimal result in this example. Note that, in this validation experiment, the enumeration approach program runs for over two days on a 28-core computer, while our solution method requires less than one hour on the same computer.
Fig. 5. The railway components (150 stations and 199 edges) chosen to be invested with budget constraint $5$ million and maintenance degree 0.4.

Fig. 5 illustrates the locations of railway components to be invested based on the optimal solution when the investment budget is $5$ million and the maintenance degree is 0.4. As illustrated in the figure, many invested edges are located close to several big cities of China, such as Beijing, Wuhan, and Zhengzhou. Western China is a region with low railway mileage and low population density. However, it has many earthquake belts. Once an earthquake occurs, all trains passing through this region are affected. Hence, the optimization model also suggests several stations and edges in the western China for investment. In the southeast of China, many railway stations and edges that do not belong to any seismic belt are also identified for investment in the optimal plan. This is because the number of trains passing through in this area is large.
Fig. 6. Convergence of the solution method under different maintenance degrees
Fig. 6 illustrates the convergence process of the CRS service loss under different maintenance degrees. Fig. 7 illustrates the convergence process of the CRS service loss under different budgets. The expected service losses change very slowly after 15
generations and become stable after 20 generations in most cases. The final service loss decreases with the increase of the budget and the maintenance degree. Increasing budget cannot decrease the railway system service loss significantly when maintenance degree is small. For example, when the maintenance degree is 0.1, the final service loss is 15.2 trains when the budget is $1 million, while the service loss is 14.1 trains when the budget is $9 million. The decrease ratio for the service loss is only 7.2% by more $8 million budget. If the maintenance degree is 0.4, the decrease ratio of the service loss by increasing the budget from $1 million to $9 million is about 28.8%. If the budget is small (such as $1 million), the service loss is 15.2 trains when the maintenance degree is 0.1, and 14.9 trains when the maintenance degree is 0.4. The decreased service loss by increasing the maintenance degree is only 0.3 trains (decrease ratio is about 1.9%). If the budget is $9 million, the decreased service loss by increasing the maintenance degree from 0.1 to 0.4 is 2.5 trains (decrease ratio is about 18%). This shows that, any marginal increase of one of the two parameters (the maintenance degree and the budget), cannot make the railway system service loss decrease significantly if the value of the other parameter is low. A good investment strategy should consider both these aspects.

4.3 Comparison with topology-based investment plans
Topological metrics are usually used to identify critical components in network science (Wang and Wu, 2010). In this section, the optimal investment plans obtained from the proposed framework are compared with two topology-based plans: degree-based and betweenness-based. Because the investment cost of stations is fairly low compared to that of edges, this study only considers the degree and betweenness of edges for illustrative purposes. Here, the degree of an edge $e$ is defined as the total number of edges connecting to it in the physical layer. The betweenness of an edge $e$ is defined as the proportion of shortest paths that pass through it in the physical layer (Su and Wang, 2009). Higher degree or betweenness edges are chosen in priority for the corresponding topology-based investment plans. The degree-based investment plan, shown in Fig. 8(a), is distributed in southeastern China; most edges are located in several big cities. The betweenness-based investment plan is shown in Fig. 8(b); most of the invested edges are located in or near the south-north and east-west railway backbones of the CRS. It is found that many of the invested edges based on network topology are located in the southeast and northeast parts of China. This is because the southeast part is the richest area in China, with many big cities deployed.
with dense railway components, and the northeast part is the old industrial region and has dense railway components. The betweenness-based investment plans are more widely and evenly distributed than the degree-based investment plans.

The expected service loss reduction ratios of the investment plans suggested by different methods are summarized in Table 2, with different budgets and maintenance degrees. Table 2 shows that the investment plans based on the proposed framework are always better than others, and the betweenness-based method is better than the degree-based method most of the time. For example, when the budget is $7 million and the maintenance degree is 0.4, the expected service loss reduction ratios based on degree and betweenness methods are 16.00% and 12.80%, respectively, while the proposed framework makes a reduction of 30.50%. The average service loss reduction ratios are listed in the last column of Table 2. The service loss reduction ratio based on the proposed framework is better than those based on degree and betweenness methods at different maintenance degree. The integration of train flow and earthquake information enables the proposed methodology to be more efficient in service loss reduction than a purely topology-based method. The difference in decreased service loss ratio between the proposed method and topology-based methods will reduce with the increase of the investment budget and the maintenance degree. This is because the railway edges selected based on topology-based methods usually have heavy train flows, and more edges with heavy flow will be selected for investment as the budget increases.

**Table 2**

Expected service loss reduction ratio compared between the proposed framework and topology-based methods

<table>
<thead>
<tr>
<th>α</th>
<th>Budget ($)</th>
<th>$1 million</th>
<th>$3 million</th>
<th>$5 million</th>
<th>$7 million</th>
<th>$9 million</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Degree</td>
<td>5.30%</td>
<td>5.40%</td>
<td>6.80%</td>
<td>8.10%</td>
<td>7.50%</td>
<td>6.62%</td>
</tr>
<tr>
<td></td>
<td>Betweenness</td>
<td>3.50%</td>
<td>3.60%</td>
<td>6.20%</td>
<td>7.00%</td>
<td>8.00%</td>
<td>5.66%</td>
</tr>
<tr>
<td></td>
<td>Optimization</td>
<td>14.60%</td>
<td>16.10%</td>
<td>18.80%</td>
<td>17.00%</td>
<td>21.30%</td>
<td>17.56%</td>
</tr>
<tr>
<td>0.1</td>
<td>Degree</td>
<td>7.30%</td>
<td>7.80%</td>
<td>9.60%</td>
<td>10.80%</td>
<td>13.00%</td>
<td>9.70%</td>
</tr>
<tr>
<td></td>
<td>Betweenness</td>
<td>4.60%</td>
<td>5.00%</td>
<td>9.40%</td>
<td>10.40%</td>
<td>12.70%</td>
<td>8.42%</td>
</tr>
<tr>
<td>0.2</td>
<td>Degree</td>
<td>5.30%</td>
<td>5.40%</td>
<td>6.80%</td>
<td>8.10%</td>
<td>7.50%</td>
<td>6.62%</td>
</tr>
<tr>
<td></td>
<td>Betweenness</td>
<td>3.50%</td>
<td>3.60%</td>
<td>6.20%</td>
<td>7.00%</td>
<td>8.00%</td>
<td>5.66%</td>
</tr>
</tbody>
</table>
### 5 Conclusions and future work

This study proposes a general framework to determine the optimal pre-disaster investment plan for strengthening railway system under earthquakes. The proposed framework selects an optimum set of stations and edges for investment to minimize the expected service loss for the railway system with the consideration of limited budget. The expected service loss is assessed based on earthquake scenarios whose locations and magnitudes are generated using historical earthquake data. The pre-disaster optimal investment plan problem is modeled as a binary nonlinear stochastic programming, and a Self-Adaptive Genetic Algorithm is used to solve it. The specific data of the CRS is used in a numerical experiment to demonstrate the practical applicability of the framework. The numerical analyses show that the proposed framework can identify investment plans that are more sensitive to the earthquake impact on the railway system compared to plans suggested by the topology-based methods. Any marginal increase of the budget or the maintenance degree will not improve the railway system performance; both aspects have a significant influence on the performance of the invested railway system. Based on the numerical analyses, investment prioritization of the CRS components is determined.

The two-layer network model in the framework can be used to analyze other types of transportation systems, such as highway system, the physical layer separates the detailed characteristic of disaster from the service layer, and the service layer separates the passenger flow from the detailed route in the physical layer. The two-layer network based framework can support the independent multi-sites disaster analysis. For interdependent multi-sites disaster scenarios, whether our framework still works
depends on the coupling effect model, which can be regarded as a direction for future research.

Based on the available historical data, this study associates the post-disaster states of railway components with the earthquake magnitude and their distance to the earthquake center. If more specific information related to the railway components are available, such as design standards, geological information and so on, the post-disaster states of railway components can be assessed more accurately. Similarly, if the passenger flow data in the railway system is available, the service loss can be calculated by the number of affected passengers and their waiting time, which is more realistic than the number of affected trains when considering the railway system service quality.

The two important components in railway systems, tunnels and bridges are not considered in current work. They can be modeled as special nodes or edges and integrated into the proposed framework if related data can be obtained. Furthermore, the segment length of the edges in the physical layer is an important parameter of the service loss assessment model, sensitivity analysis of the segment length can be conducted in future work. Although the railway system in China has been used in this paper for illustrative purposes, the framework can also be applied to analyze vulnerability of railway systems in other countries, if the datasets related to two-layer network representation and earthquake scenario generation (such as the railway system properties and the historical earthquake data) are available.

**Acknowledgments**

This work is jointly supported by National Natural Science Foundation of China (61572212, 51208223, 61433006 and 60903174), and the U.S. Department of Transportation through the NEXTRANS Center, the USDOT Region 5 University Transportation Center. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.
Appendix A

Table 3
Earthquake occurrence probability parameter of different seismic belts

<table>
<thead>
<tr>
<th>Seismic belts</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Dong Bei Dai</td>
<td>-1.07804</td>
<td>4.169624</td>
</tr>
<tr>
<td>2: Xi Ma La Ya Di De Zhen Dai</td>
<td>-1.87083</td>
<td>9.887203</td>
</tr>
<tr>
<td>3: Huang He Xia You Dai</td>
<td>-0.67837</td>
<td>0.565909</td>
</tr>
<tr>
<td>4: Tan Cheng-Lu Jiang Dai</td>
<td>-0.90374</td>
<td>3.601677</td>
</tr>
<tr>
<td>5: Hai He (He Bei) Ping Yuan Dai</td>
<td>-1.24061</td>
<td>5.711599</td>
</tr>
<tr>
<td>6: JinZhong (Shan Xi) Dai</td>
<td>-0.82564</td>
<td>3.371167</td>
</tr>
<tr>
<td>8: Yan Shan Dai</td>
<td>-1.09662</td>
<td>5.343719</td>
</tr>
<tr>
<td>8: Wei He Ping Yuan Dai</td>
<td>-0.58713</td>
<td>1.082769</td>
</tr>
<tr>
<td>9: Dong Nan Yan Hai Di De Zhen Dai</td>
<td>-0.76911</td>
<td>2.08686</td>
</tr>
<tr>
<td>10: He Lan Shan (Yin Chuan) Dai</td>
<td>-0.89244</td>
<td>3.416319</td>
</tr>
<tr>
<td>11: Liu Pan Shan Dai</td>
<td>-0.9231</td>
<td>2.924806</td>
</tr>
<tr>
<td>12: Lan Zhou-Tian Shui Dai</td>
<td>-0.84558</td>
<td>3.331743</td>
</tr>
<tr>
<td>13: Wu Dou Du-Ma Bian Dai</td>
<td>-1.28747</td>
<td>6.163912</td>
</tr>
<tr>
<td>14: An Ning He Gu Dai</td>
<td>-1.19875</td>
<td>5.234354</td>
</tr>
<tr>
<td>15: Dian Dong Dai</td>
<td>-1.1916</td>
<td>5.822203</td>
</tr>
<tr>
<td>16: A Er Tai Shan Dai</td>
<td>-2.94563</td>
<td>12.33866</td>
</tr>
<tr>
<td>17: Bei Tian Shan Dai</td>
<td>-1.36577</td>
<td>6.157666</td>
</tr>
<tr>
<td>18: Nan Tian Shan Dai</td>
<td>-1.79845</td>
<td>10.14306</td>
</tr>
<tr>
<td>19: Ta Li Mu Nan Yuan Dai</td>
<td>-2.81596</td>
<td>13.86435</td>
</tr>
<tr>
<td>20: He Xi Zou Lang Dai</td>
<td>-1.44448</td>
<td>6.642546</td>
</tr>
<tr>
<td>21: Xi CangZhong Bu Dai</td>
<td>-2.48027</td>
<td>12.46584</td>
</tr>
<tr>
<td>22: Kang Ding-GanZi Dai</td>
<td>-1.33261</td>
<td>5.851162</td>
</tr>
<tr>
<td>23: Jin Sha Jiang-Yuan Jiang Dai</td>
<td>-1.31102</td>
<td>5.715108</td>
</tr>
<tr>
<td>24: Nu Jiang-Lan Cang Jiang Dai</td>
<td>-1.71049</td>
<td>8.751554</td>
</tr>
</tbody>
</table>
Appendix B

Solution procedure implementation

According to the Genetic Algorithm based solution procedure, the solution method for the investment plan can be summarized in Fig. 9.

The steps in Fig. 9 are described as follows:

**Step 1** Define $N_{GA}$ as the preset number of generations, which is used as stop criteria...
for GA, and set Genetic Algorithm iteration $i_{GA} = 1$.

**Step 2** Generate initial population of investment plan $Y_{i_{GA}} = \Omega\{y\}$.

**Step 2.1** Define $N_{pop}$ as the preset number of populations. Define $i_{pop}$ as the population individuals index, and set $i_{pop} = 1$. Initial $\Omega\{y\} = \emptyset$

**Step 2.2** Generate a $N_{gene}$ (the same as number of physical layer edges) bits 0-1 vector as investment plan: $y_i$

**Step 2.3** Calculate the total length of investment plan $y_i$: $b_y = \sum_{e \in E} c_e * \text{Len}(e) + \sum_{v \in V} c_v * y_v$

**Step 2.4** If $b_y$ exceeds budget $B$, go to step 2.2. If not, go to step 2.5.

**Step 2.5** Set $i_{pop} = i_{pop} + 1$ and set $\Omega\{y\} = \Omega\{y\} \cup y_i$.

**Step 2.6** If $i_{pop}$ reaches $N_{pop}$, go to step 3. If not, go to step 2.2.

**Step 3** Calculate CRS service loss for each investment plan in the population, define $i_{chr}$ as individual index, and set $i_{chr} = 1$.

**Step 3.1** Select $i_{chr}$-th individual in the population as current investment plan $y$.

**Step 3.2** Define $N_{EQ}$ as the number of earthquake scenario, then the earthquake scenario set $H = \{h\}^{N_{EQ}}$.

**Step 3.3** Generate earthquake scenario set $H$. Each $h$ in $H$ is an earthquake scenario for $T_{int}$ (period of time interested). Each earthquake $\eta$ includes location and magnitude, where $\eta \in h$.

**Step 3.4** Calculate the CRS service loss for $\eta$ by Eq. (26).

**Step 3.5** Assess the CRS service loss for $h$ by Eq. (25).

**Step 3.6** Calculate the expected CRS service loss under all the earthquake scenarios generated when $y$ is implemented: $D_y = \text{Exp}_{h \in H}[D_y(h)]$.

**Step 3.7** If $i_{chr} < |\Omega\{y\}|$, set $i_{chr} = i_{chr} + 1$, and go to step 3.1. Else, go to step 4.

**Step 4** Find the investment plan whose corresponding service losses is the lowest in the current population: $y_{local}$. Compare its corresponding service loss $D_{local}$
with the global lowest loss $D_{global}$. Use the lower value among them as the new $D_{global}$. Use the corresponding investment plan as the new global optimized investment plan $y_{global}$.

**Step 5** If $i_{GA} < N_{GA}$, go to Step 6. Otherwise, go to Step 9.

**Step 6** Sort the investment plans by their corresponding service losses, select first $N_{set}$ (a preset number used as size of seeds of next generation population) investment plan as seeds of next generation population: $Y_{seed}$.

**Step 7** Implement mutation operator and crossover operator on $Y_{seed}$. Check whether the newly generated investment plan satisfies the investment budget constraint. If an investment plan is infeasible, then it is deleted from the solution set. Add the feasible investment plans to the next generation population $Y_{i_{GA}+1}$.

**Step 8** Set $i_{GA} = i_{GA} + 1$. Go to step 3.1.

**Step 9** End.

**References**


He, X., Peeta, S., 2016. A marginal utility day-to-day traffic evolution model based on


NCEDC, 2013. ANSS Composite Earthquake Catalog. UC Regents.


Ouyang, M., Zhao, L., Hong, L., Pan, Z., 2014. Comparisons of complex network based


