Applying a Genetic Algorithm for Finding the Worst Scenario during Post-Disaster Recovery of Water Distribution Networks

Shunichi Tada[a], Taro Kanno[b], and Kazuo Furuta[c]
[a] the University of Tokyo, Tokyo, Japan, tada-shunichi513@g.ecc.u-tokyo.ac.jp
[b] the University of Tokyo, Tokyo, Japan, kanno@sys.t.u-tokyo.ac.jp
[c] the University of Tokyo, Tokyo, Japan, furuta@rerc.t.u-tokyo.ac.jp

Abstract: Because water is an essential resource for numerous activities, a water distribution network (WDN) is of critical importance. Therefore, it is necessary to appropriately prepare for post-disaster restoration of WDNs. To evaluate the restoration plan of damaged pipes, we have been developing an agent-based simulation that can reproduce restoration processes of WDNs and the manner in which restoration plans affect the performance of each subsystem in a city during post-disaster periods. Various studies involving manual generation of the damage scenario have been reported; therein, the number of damaged pipes was estimated by an empirical equation considering the magnitude of the earthquake and pipe properties, while a geographical distribution of the damaged pipes was randomly selected. Our previous research revealed that the performance of the restoration plan was highly dependent on the geographical distribution of damaged pipes. Therefore, evaluating the resilience of WDNs using randomly generated scenarios is challenging; however, determining the most challenging scenario is necessary for a given number of damaged pipes. In addition, scenarios for training and exercise must be designed and prepared based on challenges in training objectives, not randomly or in an ad hoc manner. However, it is difficult to classify scenarios according to their challenges or characteristics owing to several possible damage distributions. In this study, we applied a genetic algorithm (GA) to explore the most challenging scenario of damage distribution for repair. In the GA employed, an individual represented a disaster scenario that described the geographical distribution of damaged pipes, and the population represented a set of various scenarios. We used the resilience triangle obtained from the simulation results as an objective function. Through genetic operators such as reproduction, crossover, and mutation, an individual with the largest resilience triangle, which represented the worst scenario, was obtained. We could also obtain a set of scenarios with different resilience triangles, that is, a set of scenarios ordered based on their difficulties. We conducted a test search with 100 damaged pipes out of over 4000 pipes and confirmed that our proposed method achieved conversion after 100 generations. By analyzing the population after conversion, we identified some standard pipes under challenging scenarios, which suggested that these pipes were critical for the resilience of WDNs.

1. INTRODUCTION

Along with electricity, gas, and transportation facilities, a water distribution network (WDN) is a critical lifeline of every city. Once the WDN is damaged during disasters such as earthquakes, water management organizations must consider a restoration plan to prevent prolonged disturbances in citizen activities. Because it is impossible to observe the effect of the restoration process in reality, several studies have developed simulation models to represent the behavior of WDNs and their restoration processes. For instance, Liu et al. [1] developed a detailed simulation model to evaluate the effect of WDN restoration and confirmed the manner in which three restoration strategies affected the performance of a WDN following a disaster. Their study only considered the availability of the WDN. However, modern cities are so complicated that it is often necessary to consider the manner in which the restoration process of the WDN and other city components affect each other to avoid unrealistic results. For example, a WDN needs electricity to provide water to each facility; therefore, the availability of power grids must be considered during WDN restoration. In addition, activities such as bathing or goods manufacturing often need water, indicating that WDN restoration also influences these activities. Liu et al. used a genetic algorithm (GA) for restoration planning; however, there still exists a lack of reliable information for practitioners, as the results produced by GAs are difficult to interpret.

To evaluate the effect of WDN restoration in complicated and interdependent environments, our research group has been developing a simulation that can reflect the multiple interdependencies within various city functions [2]. Moreover, to obtain interpretable simulation results, we developed an optimization method employing the heuristics provided by practitioners [3] and a method to interpret the results of GAs using the random forest algorithm. [4]. Based on our previous research, we concluded that restoration plans highly depend on the simulation scenario, particularly the damage scenario, which indicates the distribution of damaged pipes. Therefore, to ensure simulation interpretability, the characteristics and difficulties associated with the damage scenarios must be considered.

In this study, we developed a new scenario exploration method using a GA to find the scenario that demonstrated the worst performance against a particular static restoration strategy. We aimed to generate a set of scenarios ranked according to their difficulties through this method. Furthermore, by analyzing the scenarios that exhibited the worst performance compared with others, we attempted to determine the tendency of the distribution that may cause poor performance.

2. POST-DISASTER RECOVERY SIMULATION OF THE WDN

2.1. Simulation Model

In our previous research [2], we developed an agent-based simulation model with the following three subsystems: civil lives, industries, and lifelines. This simulation model considers the multiple interdependencies between and within these three subsystems and evaluates the manner in which the recovery process of lifelines affects city functions. We implemented two models: the network and agent models to represent the function of each lifeline and citizen behavior.

Network model

Herein, lifelines are represented as multilayered networks. In these networks, links represent the facilities that distribute each lifeline resource, such as pipes for the WDN, and nodes represent the connection points of these facilities. Lifeline resource facilities, such as reservoirs, are located on the nodes.

Agent model

There exist two types of agents: facility agents that represent static facilities, such as houses, companies, and lifeline resource facilities, and mobile agents that represent citizens. The actions of mobile agents differ according to their jobs.

2.2. Evaluation Function

An evaluation function was defined to calculate the resilience triangle of the entire city system as the performance indicator to evaluate restoration processes. Because we focused on the effect of the restoration process on each subsystem, the performance was calculated as the weighted sum of the performances of each subsystem. In this study, we configured the weights as $\alpha:\beta:\gamma = 1:1:1$. Herein, regarding the performance of each subsystem, $P_{lifeline}$ represents the performance of lifelines calculated according to their availability rates during a disaster period, $P_{industy}$ represents the performance of industries calculated according to completion rates of company activities, and P_{civil} represents the performance of civil lives calculated as the average of daily activities that agents could perform each day compared with the number of activities performed during a normal period. Because the restoration proceeds gradually, the performance of these three subsystems changes daily, which is why we aggregate them at the end of the simulation. More definitions can be found in our previous research [2].

$$P_{city} = \alpha \times P_{lifeline} + \beta \times P_{industry} + \gamma \times P_{civil} \tag{1}$$

2.3. Optimization Method

In our previous study [2], we optimized the recovery plan of a WDN to improve its performance using a GA. However, the recovery plan generated by the GA was difficult to interpret, and thus, it was ineffective for use by practitioners. To generate a restoration plan that provided meaningful feedback to practitioners, we developed another optimization method that used seven heuristics provided by practitioners [3]. These seven heuristics were divided into repair prioritization and task assignment heuristics. The repair prioritization heuristics contain four heuristics representing empirical rules related to the repair prioritization of damaged pipes. Alternatively, the task assignment heuristics comprise three heuristics that represent empirical rules on the assignment of repair tasks to restoration teams.

Repair prioritization heuristics

1. Upstream heuristic
   This heuristic assigns repair priority to upstream pipes. This priority is assigned based on the distance from the reservoir.
2. Important facility heuristic
   This heuristic prioritizes pipes connected to essential facilities (e.g., evacuation centers and hospitals). This priority is assigned according to the distance from the nearest important facility.
3. Main pipe heuristic
   This heuristic prioritizes primary pipes. There exist two types of pipes: main and branch pipes. Main pipes distribute water to branch pipes throughout a city, whereas branch pipes branch off from the main pipes and distribute water directly to faucets.
4. Many houses heuristic
   This heuristic prioritizes pipes that are connected to several houses. This priority is assigned according to the number of houses connected to the pipes.

Task assignment heuristics

1. Short duration heuristic
   This heuristic prioritizes tasks with the shortest restoration durations. The duration is calculated by considering the sum of the travel time, preparation time, and labor time.
2. Close distance heuristic
   This heuristic prioritizes tasks that are closer to the location of each restoration team.
3. Area intensive restoration heuristic
   This heuristic prioritizes tasks that are located near other pipes that need to be repaired. This is evaluated based on the distance from the nearest pipe to be repaired.

The heuristics within each heuristic type are independent. Therefore, the order of application of each of the four or three heuristics must be considered. A total of 144 (= 4! × 3!) patterns of the application order can be obtained, and the difference in the order represents the difference in the restoration strategy.

2.4. Scenario

The simulation scenario was divided into two parts: the damage and support scenarios. The damage scenario represented the distribution of damaged pipes, and the support scenario represented the situation of resources used during restoration, including the properties of stockpiles and supply plans. In our previous research [3], we assumed some of these scenarios according to the knowledge provided by practitioners. For example, the number of damaged pipes was estimated using an earthquake damage estimation formula [5], and the support scenarios were assumed based on a real mitigation plan. However, the distribution of damaged pipes under the estimated number was uncertain; therefore, they were randomly determined. Because most heuristics use geographical factors, the results of our previous research [3] indicated that the appropriate order of heuristics depends on the scenario, particularly the damage scenario.

3. SCENARIO EXPLORATION USING A GA

In this study, we applied a GA to explore the damage scenario that demonstrated the worst performance. The implementation of the GA is explained in the following three subsections.

3.1. Genotype and Phenotype

Figure 1 illustrates a gene representing the ID of a damaged pipe and a chromosome representing the set of damaged pipes. The chromosome length represents the number of damaged pipes; therefore, it was determined and fixed according to the earthquake damage estimation formula [5]. One damaged WDN is generated and used in the simulation using one chromosome.

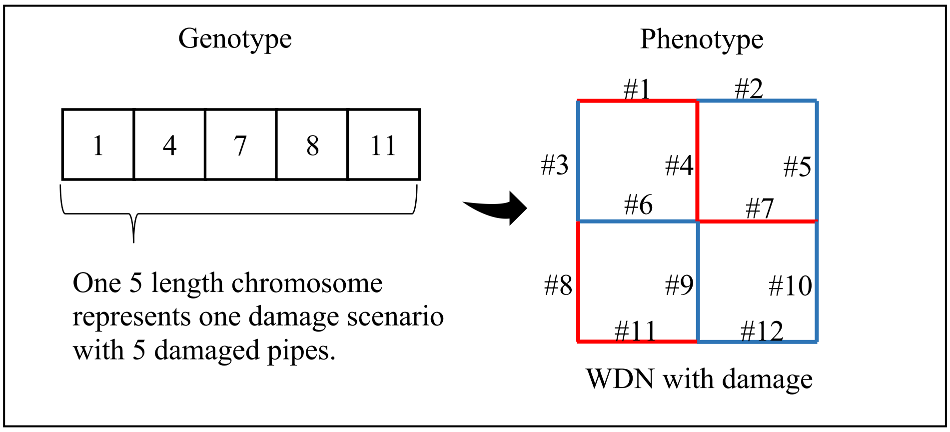


Figure 1: Example of Genotype and Phenotype

3.2. Fitness Function

Various studies have employed GAs to locate individuals with the highest fitness. In this research, we defined the fitness function as follows (Equation (2)) to identify the individual that demonstrates the lowest performance P_{city}.

$$fitness = 1 - P_{city} \tag{2}$$

3.3. Genetic Operators

Typically, chromosomes are updated using the following three operators: reproduction, crossover, and mutation. In this study, the following three operations were implemented:

Reproduction

Reproduction is an operator that allows an individual with good fitness to survive in the next generation. We implemented two selection methods: elitism and roulette wheel selection to select individuals for survival. Through elitism selection, N_{elite} highest individuals were selected, while $N_{roulette}$ individuals were selected according to their fitness. The number of selected individuals (N_{elite} and $N_{roulette}$) and the probability p_i that individual i is selected in the roulette wheel selection were calculated based on the following equations using the total population N, elitism selection rate r_{elite}, roulette wheel selection rate $r_{roulette}$, and the fitness of each individual f_i.

$$N_{elite} = r_{elite} \times N \tag{3}$$
$$N_{roulette} = r_{roulette} \times N \tag{4}$$
$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \tag{5}$$

Crossover

Crossover is an operator that generates two individuals in the next generation using two individuals in the current generation. We implemented a uniform crossover that exchanged the gene at each position with a 50% probability between two randomly selected chromosomes. To avoid duplication of genes within one chromosome, exchange was not performed if one or both genes were shared by two chromosomes (see Figure 2). The number of individuals generated by the crossover $N_{crossover}$ was calculated based on the following equation using the crossover rate $r_{crossover}$.

$$N_{crossover} = r_{crossover} \times N \tag{6}$$

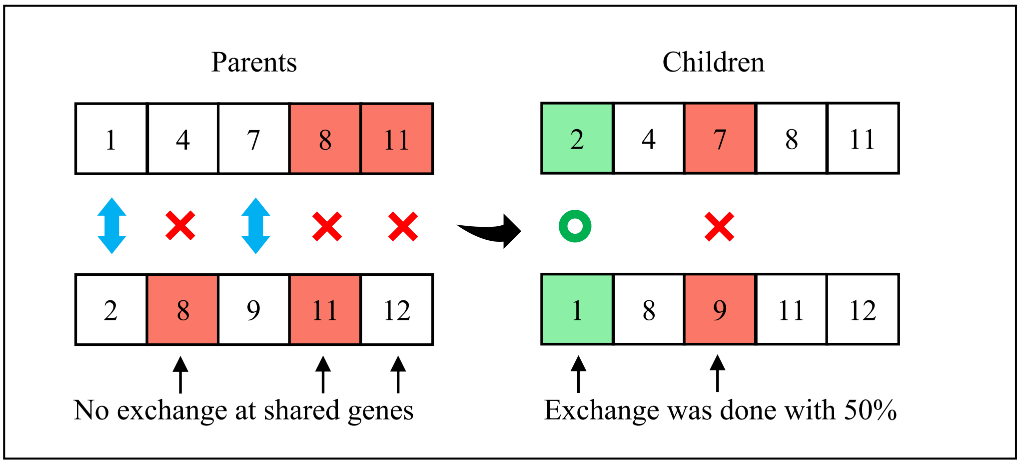


Figure 2: Crossover

Mutation

Mutation is an operator that maintains diversity among individuals by randomly altering a gene at random positions. Within individuals in the next generation generated by reproduction and crossover, each individual is judged based on whether or not the mutation will be used with a probability $r_{mutation}$. For the selected individuals, the gene in each position is judged based on whether or not it will be replaced with another ID with a probability $r_{mutation}$. A new gene is selected from the IDs that are not contained in the current chromosome to avoid duplication of genes within one chromosome.

After these operators are used, the population of individuals in the next generation must be similar to that in the current generation. To satisfy this, the three rates (r_{elite}, $r_{roulette}$, $r_{crossover}$) must satisfy the following equation:

$$r_{elite} + r_{roulette} + r_{crossover} = 1 \tag{7}$$

4. RESULTS AND DISCUSSION

4.1. Simulation Settings

For the simulation, we developed a virtual city model with the attributes listed in Table 1. Table 2 lists the parameters used to explore scenarios with the GA.

Table 1: Attributes of the simulation model

Attributes	Values
Number of pipes	4610
Number of reservoirs	1
Number of repair teams	13
Number of citizens	6990
- Non-worker citizens	3064
- Industry worker citizens	3768
- Lifeline worker citizens	20
- Repair worker citizens	105
- Truck worker citizens	33
Simulation duration	30 days

Table 2: Parameters of the GA

Parameters	Values
Length of chromosome	100
Generation	100
Population	100
Rate of elite selection: r_{elite}	0.2
Rate of roulette wheel selection: $r_{roulette}$	0.5
Rate of crossover: $r_{crossover}$	0.3
Rate of mutation: $r_{mutation}$	0.1

For the optimization method, we used the heuristics detailed in Section 2.3 following a static order. In this study, we fixed the order to "1, 2, 3, 4" for the repair priority heuristics and "1, 2, 3" for the task assignment.

4.2. Results

In this study, we performed scenario exploration several times using the settings detailed in the previous section. The graphs presented on the next page illustrate a part of the results obtained. We plotted all the individuals and their generation (horizontal axis) and performance (vertical axis). The blue plot represents the individual with the worst performance for each generation. In addition, the average performance in each generation is indicated by a red line.

Figures 3 and 4 indicate that the average and minimum performances gradually decrease as the generation progresses. The worst performance explored in the first result is not as bad (0.9225); alternatively, the worst performance explored in the second result is worse (0.7327). Regarding the search space of this exploration, there were 8.5×10^{207} possible combinations, while both results explored 3070 combinations. In addition, 4568 pipes appeared in the first result, and 4551 pipes appeared in the second result out of 4610 pipes.

Figure 3: Result 1 - without worse performance

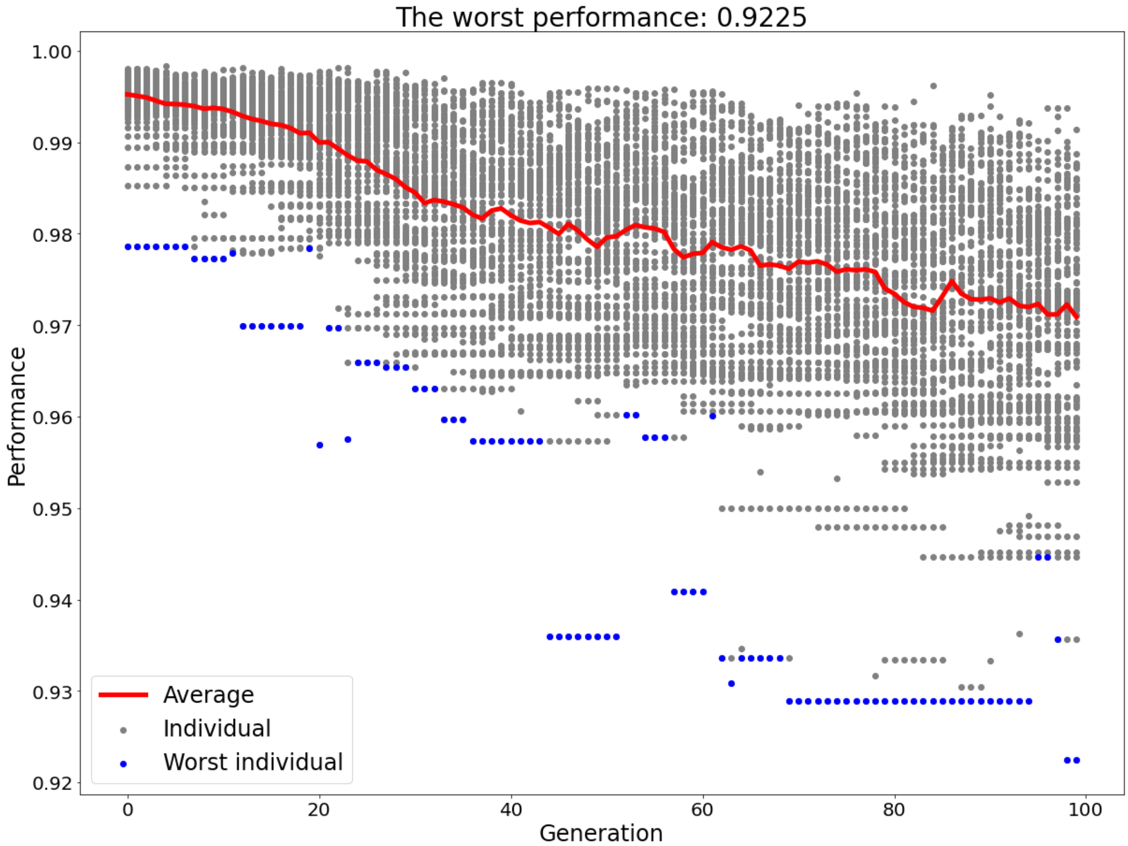
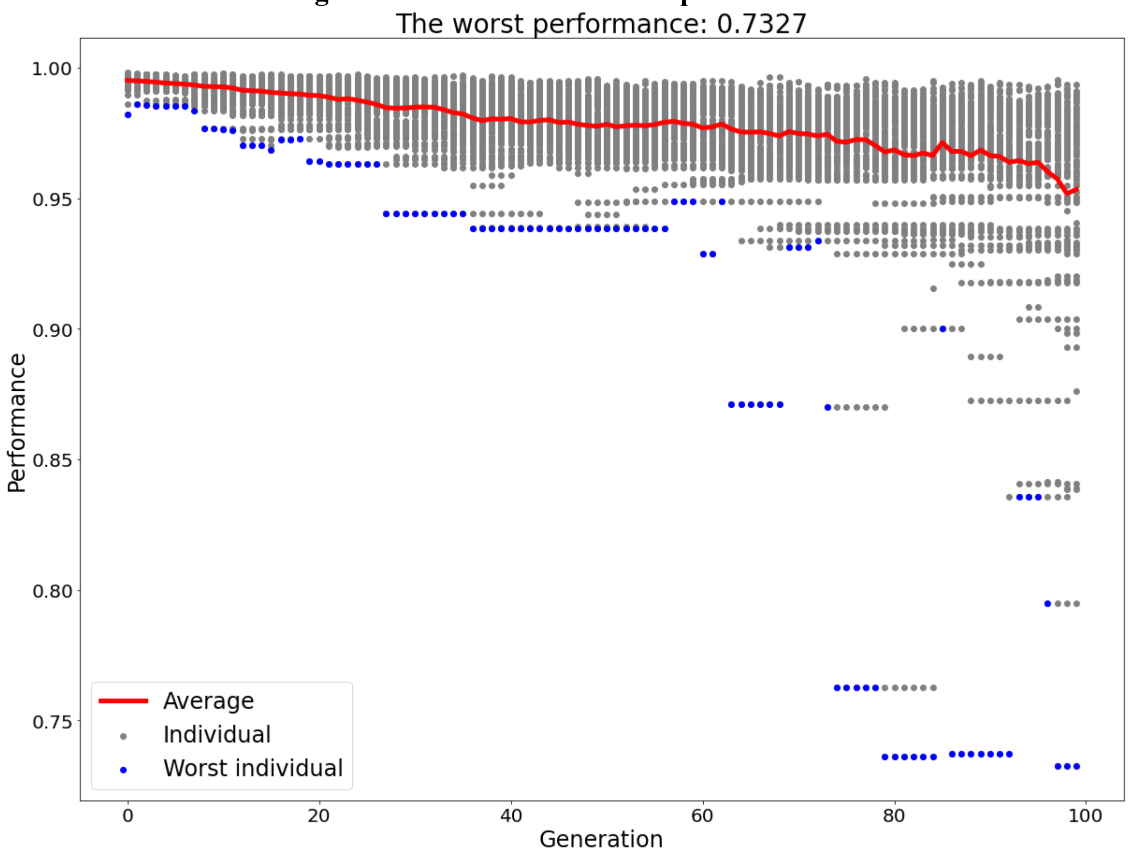


Figure 4: Result 2 - with worse performance



4.3. Discussion

Analysis of appeared pipes in worst individuals

In the second result, presented in Figure 4, there exist five different individuals with a performance lower than 0.8. As shown in Table 3, there are 11 pipes shared among more than three scenarios.

Table 3: Five lowest performances and shared pipes

	#1	#2	#3	#4	#5
Performance	0.7327	0.7363	0.7371	0.7626	0.7948
Pipe #873	o	o	o	o	o
Pipe #1595	o	o	o	o	o
Pipe #838	o	o	o	x	o
Pipe #961	o	x	o	o	o
Pipe #1979	o	o	o	x	o
Pipe #2691	o	o	o	o	x
Pipe #2748	x	o	o	o	o
Pipe #3115	o	o	x	o	o
Pipe #3329	o	x	o	o	o
Pipe #3332	o	o	x	o	o
Pipe #3633	o	x	o	o	o

By analyzing the 11 pipes listed in Table 3, we identified some features of these pipes.

- Six of these pipes (#838, #873, #961, #1595, #1979, and #2748) were main pipes that required heavy machines for repair.
- Pipes #838 and #873 were the only pipes that provided water to the area where evacuation sites were concentrated.
- Pipes #838 and #873 were far from the original locations of the repair squads.
- Pipes #961 and #1595 were closest to the reservoir.

In these damage distributions, pipe #838 and pipe #873 were considered the most significant among all damaged pipes because the absence of these pipes led to water outages in areas, where five evacuation sites were concentrated and approximately 700 citizens resided. However, the employed order of heuristics prioritized other pipes, such as pipe #961 and pipe #1595, because these pipes were closer to the reservoir and the original location of the repair squads. These factors resulted in the worst performance. It is possible that by altering the order of the heuristics, for example, using the critical facility heuristic and area-intensive restoration heuristic in the first place, we could improve the performance in these damage distributions.

Problems in the exploration method

Based on the two results obtained, the exploration method appropriately identified the damage distribution that demonstrated poor performance; however, we can conclude that 100 generations and 100 population were insufficient because the worst performance explored was unstable. According to the high selection rates and low mutation rate, the number of explored combinations was smaller than the total number of individuals, 10000.

5. CONCLUSION

In this study, we implemented a GA to explore the damage distributions that demonstrated poor performance. After several explorations, we discovered five different damage distributions with performances below 0.8. Through scenario analysis, we concluded that the reason for the worse performance was the order of heuristics used for optimization.

However, several issues still remain unresolved. The first issue concerns tuning the parameters used in the GA. As noted in the discussion section, the value of both generation and population and the rate of each genetic operator must be improved for more efficient exploration. The second issue concerns further analysis of the worse scenarios explored in this research. We restricted our analysis to worse scenarios; however, analyses between better and worse scenarios must be conducted to observe the tendency of the worse scenario. The third issue concerns the optimization method. In this study, we fixed the order of heuristics to represent a static restoration strategy. However, it is unrealistic to assume that practitioners only hold a single restoration strategy; rather, they employ several restoration strategies case-by-case. Thus, in conclusion, we must improve our exploration method to cope with several restoration strategies to address this.

Acknowledgements

This work was partly supported by JST-Mirai Program Grant No. JP20348594, Japan.

References

[1] W. Liu, Z. Song, M. Ouyang, and J. Li. "*Recovery-based seismic resilience enhancement strategies of water distribution networks*", Reliability Engineering & System Safety, Vol. 203, 107088, (2020).
[2] T. Kanno, S. Koike, T. Suzuki, and K. Furuta. "*Human-centered modeling framework of multiple interdependency in urban systems for simulation of post-disaster recovery processes*", Cognition, Technology & Work, Vol. 21, pp. 301-316, (2019).
[3] K. Wakayama, T. Kanno, Y. Kawase, H. Takahashi, and K. Furuta. "*Comparison of the Post-Disaster Recovery of Water Supply System by GA Optimization and Heuristics*", Instructions for Preparing Paper for ESREL 2020 PSAM 15, pp. 240-246, (2020).
[4] S. Tada and T. Kanno. "*Random Forest for Genetic Algorithm Optimization: Making it Quick and Interpretable*", Asia Simulation Conference 2021, pp. 201-208, (2021).
[5] Japan Water Works Association. "*Study of Estimation of Earthquake Damages of Water Pipes*", JWWA, 1998.