

Article

A Reinforcement Learning Model of Multiple UAVs for Transporting Emergency Relief Supplies

Daiki Hachiya ¹, Erick Mas ² and Shunichi Koshimura ^{2,*}¹ Graduate School of Engineering, Tohoku University, Aoba 468-1, Aramaki, Aoba-ku, Sendai 980-8572, Japan² International Research Institute of Disaster Science, Tohoku University, Aoba 468-1, Aramaki, Aoba-ku, Sendai 980-8572, Japan

* Correspondence: koshimura@irides.tohoku.ac.jp

Abstract: In large-scale disasters, such as earthquakes and tsunamis, quick and sufficient transportation of emergency relief supplies is required. Logistics activities conducted to quickly provide appropriate aid supplies (relief goods) to people affected by disasters are known as humanitarian logistics (HL), and play an important role in terms of saving the lives of those affected. In the previous last-mile distribution of HL, supplies are transported by trucks and helicopters, but these transport methods are sometimes not feasible. Therefore, the use of unmanned aerial vehicles (UAVs) to transport supplies is attracting attention due to their convenience regardless of the disaster conditions. However, existing transportation planning that utilizes UAVs may not meet some of the requirements for post-disaster transport of supplies. Equitable distribution of supplies among affected shelters is particularly important in a crisis situation, but it has not been a major consideration in the logistics of UAVs in the existing study. Therefore, this study proposes transportation planning by introducing three crucial performance metrics: (1) the rapidity of supplies, (2) the urgency of supplies, and (3) the equity of supply amounts. We formulated the routing problem of UAVs as the multi-objective, multi-trip, multi-item, and multi-UAV problem, and optimize the problem with Q-learning (QL), one of the reinforcement learning methods. We performed reinforcement learning for multiple cases with different rewards and quantitatively evaluated the transportation of each countermeasure by comparing them. The results suggest that the model improved the stability of the supply of emergency relief supplies to all evacuation centers when compared to other models.

Keywords: unmanned aerial vehicle (UAV); humanitarian logistics (HL); disaster resilience; emergency relief supplies; vehicle routing problem (VRP); Q-learning (QL)



Citation: Hachiya, D.; Mas, E.; Koshimura, S. A Reinforcement Learning Model of Multiple UAVs for Transporting Emergency Relief Supplies. *Appl. Sci.* **2022**, *12*, 10427. <https://doi.org/10.3390/app122010427>

Academic Editors: Diogo Ribeiro and Mauro Lo Brutto

Received: 11 August 2022

Accepted: 6 October 2022

Published: 16 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When a disaster such as an earthquake or tsunami occurs, a quick and sufficient distribution of emergency relief supply is required. The logistics activities carried out to rapidly provide adequate aid supplies (relief goods) to people affected by disasters are commonly known as humanitarian logistics (HL) [1]. The importance of HL cannot be overstated, as it can affect the death tolls in regions suffering from disasters [2]. Post-disaster HL involves multiple activities; donation soliciting, material convergence, and last-mile distribution [3]. Donation soliciting refers to the collection of materials at donor sites, whereas material convergence refers to the flow of materials from donor sites to the end sites (distribution centers) inside disaster areas [4]. Last-mile distribution, which is the final stage of the relief supply chain, refers to the delivery of materials from end sites to survivors (to individual demand locations; e.g., shelters), and many disasters have shown the challenges of Last-mile distribution [5,6]. In the Great East Japan Earthquake of 11 March 2011, many problems occurred in the delivery of relief supplies, and, as a result, hundreds of thousands of survivors were unable to receive relief supplies during the first six days of the disaster [5,6]. One of the most important problems was the lack

of sufficient transportation to meet the demands of all the shelters. Road transportation of emergency relief supplies was difficult due to a lack of fuel oil [7] and disruptions in the transportation network [8]. On the other hand, the number of helicopters that can be used in an emergency is limited, and it would be difficult to transport supplies to all shelters by helicopter alone, since helicopters are used for a variety of operations, such as transporting supplies, assessing damage [9], and rescuing and searching for victims [6]. In fact, in the disaster, the maximum number of *bosai* helicopters used was only 47 [10] for the total of 1692 shelters (refugee centers) [6] in the three hardest-hit prefectures in Japan (Iwate, Fukushima, and Miyagi). It is predicted that it will be difficult to use many helicopters to transport supplies during a disaster. Thus, there is great interest in unmanned aerial vehicles (UAVs) which can be used regardless of the damage to the roads and in large numbers as a means for transportation.

UAVs, also known as drones, are currently applied in several fields, such as aerial photography, surveying, and pesticide spraying. In addition, UAVs are used in the field of logistics due to improvements in UAV performance and related technologies, such as batteries [11]. Recently, there has been a growing interest in the field of logistics using UAVs. For instance, Amazon and Walmart have each been working on new platforms that use UAVs to deliver shipments to customers [12]. Similarly, other companies, such as DHL, Google, and Alibaba, also began developing their own UAVs [13]. UAVs have also been demonstrated to transport goods in emergency and disaster scenarios. Yakushiji et al. [14] conducted a series of drone transport experiments to demonstrate the use of drones for transporting emergency supplies. In Rwanda, Zipline transported blood, medicine, and vaccines to remote areas by drone [15]. On the other hand, most UAVs are smaller than helicopters, and the amount of items that can be transported at one time (maximum payload) is limited. Therefore, when transporting emergency relief supplies using UAVs, it is necessary to make the most effective use of the UAVs, taking into consideration their battery and maximum payload limitations.

In recent years, there have been various studies on delivery using UAVs [11,16–27]. In general, these problems are formulated as UAV routing problems (UAVRPs), which are a special kind of vehicle routing problem (VRP) [16]. A VRP attempts to find the optimal routes for one or more vehicles to deliver commodities to a set of locations [17]. In conventional UAVRPs for delivery, the objective function is set as follows total traveling/delivery time [17–20], total travel distance [21,22], total number of covered tasks [22], additional costs due to outside delivery deadlines [21,23], location priority [18], and total disutility for the delivery [24]. However, in the post-disaster HL situation, applying these objective functions is not adequate. Equitable relief supplies distribution among recipients is also a critical consideration in post-disaster HL [28]. This is due to the high stakes associated with unsatisfied and/or late-satisfied demand [1]. Although considerations of equity are featured in the land transportation literature [1,29], the authors are unaware of instances where equity was considered in relation to UAVRPs. Considering equity cost as an objective function in UAVRPs may allow for optimal transportation of goods in disaster situations.

In this study, motivated by challenges associated with the transportation of emergency relief supplies to isolated areas by tsunami, we present a model for planning the transportation of supplies by multiple UAVs in unique situations during a disaster. In this case, not only are many different types of supplies needed, but also the disparity in the amounts of supplies to be transported to each shelter must be minimized. In this study, we formulated the problem as an NP-hard UAV routing problem that determines the amount of supplies to be transported by multiple UAVs and destination shelters and then used Q-learning (QL) [30], a typical method of reinforcement learning (RL), to improve the efficiency of planning the transportation of multiple UAVs. The purpose of this study is to propose an efficient logistics planning strategy for disaster response using multiple UAVs. Our proposed method considered the following three perspectives: (i) the rapid transport of supplies, (ii) the urgency of supplies, and (iii) the equity (fairness) in the amount of supplies for each shelter. The main contributions of this study are as follows:

- In the UAVRPs, equity metric was implemented to take into account the post-disaster HL situation, demonstrating that the disparity in supplies per evacuation center can be eliminated;
- We proposed a QL algorithm for solving UAVRPs, discussed the parameter settings of QL, and outperformed meta-heuristics methods in the conditions of the previous study;
- We introduced three metrics that are considered important in the transportation of supplies in a post-disaster HL. We tested several transport strategies with different weights of the three metrics and evaluated the response of each strategy to each metric.

The remainder of this study is organized as follows: Section 2 reviews previous studies related to disaster response using UAVs and the logistics of UAVs. Section 3 gives an overview of the model constructed in this study. Section 4 describes the proposed method, QL. Section 5 describes the comparison and validation of the methods used in this study, the results, and the discussion. Section 6 gives the conclusion and future work.

2. Related Works

This study focuses on the potential of using UAVs in disaster response. A summary of the UAV routing problems for logistics is presented in the following.

2.1. Potential of Using UAVs in Disaster Response

In recent years, many studies have discussed the use of big data and information technology in disaster risk management [31,32].

Zacharie et al. [33] introduced a rapid human body detection method using image processing from a UAV camera to save lives during natural disasters such as earthquakes. They showed that irrespective of distance, a camera mounted on a UAV can clearly detect a human body. Nagasawa et al. [34] proposed a path planning method for multiple UAVs to aid in 3D building damage surveys or disaster situations. They proposed a methodology that combines a fuzzy C-means method for assigning camera location points to each UAV with a route optimization algorithm for calculating the visit order of the camera location points for each UAV by solving the Multiple Traveling Salesmen Problem (MTSP). Alhindi et al. [35] explored the potential of utilizing UAVs for crowd management during emergency evacuations. They suggested a simulation model with two UAV guidance approaches: partial guidance and full guidance. Klaine et al. [36] presented an intelligent solution based on RL to find the best position for multiple UAVs to be used as cellular hot spots in an emergency scenario. They maximized the number of users covered by the system, while in this case, the UAVs were limited by both backhaul and radio access network constraints. Chowdhury et al. [37] proposed a mixed-integer linear programming model for a Heterogeneous Fixed Fleet Drone Routing problem (HFFDRP) that minimizes the post-disaster inspection cost of a disaster-affected area.

As described above, various applications of UAVs are being considered in disaster response, and there is a strong possibility that UAVs could be applied to the transportation of emergency relief supplies during a disaster. This is because UAVs have the potential to save time and cost compared to traditional means of transportation and to enable the transport of emergency relief supplies to disconnected areas (e.g., tsunami inundation areas) [18,38].

2.2. UAV Routing Problem for Logistics

There are two types of UAVRPs: drone-only problems, which use only UAVs, and truck-drone problems [25,39–42], which combine UAVs and land transportation. In this section, we review research on drone-only problems that are primarily relevant to this study, considering the disruption of transportation infrastructure in a disaster. As the VRP is an NP-hard problem, exact algorithms are efficient only for small problem instances. Since real-world problems tend to be quite large, heuristics and metaheuristics are often more practical [43]. Hence, in the literature, many UAVRPs have been solved using heuristics

and metaheuristics. Dorling et al. [17] proposed two multi-trip VRPs for UAV delivery to minimize costs subject to a delivery time limit or minimize the overall delivery time subject to a budget constraint. They proposed a cost function that considers an energy consumption model and UAV reuse and then applied it in a simulated annealing (SA) heuristic to find suboptimal solutions to practical scenarios. Song et al. [22] proposed a mixed integer linear programming (MILP) formulation for the derivation of persistent UAV delivery schedules and developed a receding horizon task assignment (RHTA) heuristic with numerical examples for island-area delivery. They set the objective function to maximize the weighted sum of two objectives: the total number of customers covered and the total traveling distance during the delivery service.

Similarly, Jiang et al. [21] established a model for UAV task assignment in logistics that is solved by an improved particle swarm optimization (PSO) algorithm. Initially, they set a time window at each node, imposed a penalty cost for delayed delivery and a cost based on the total distance traveled by all UAVs, and then minimized the cost using the PSO algorithm. Chowdhury et al. [25] proposed a continuous approximation (CA) model that determines the optimal distribution center locations, their corresponding service regions, and ordering quantities to minimize the overall distribution cost for disaster-relief operations. Li et al. [23] focused on the issue of UAV logistics in urban environments and developed an automatic delivery system to support the delivery of packages. They optimized two objectives: customer satisfaction and total completion time. A variable neighborhood search (VNS) algorithm framework is used to generate the approximate optimal solution for their problem. Rabta et al. [18] considered UAV applications in last-mile distribution in HL and presented an optimization model for the delivery of multiple packages of lightweight relief items. They set the objective to minimize the total traveling distance (or time/cost) of the UAV under payload and energy constraints. Shi et al. [19] proposed a bi-objective mixed integer programming model for the multi-trip drone location routing problem, which allows simultaneous pick-up and delivery, and shorten the time to deliver medical supplies in the right place. Then, a modified NSGA-II (Non-dominated Sorting Genetic Algorithm II) which includes double-layer coding, is designed to solve the model. Ghelichi et al. [20] presented an optimization model to design the logistics for a fleet of drones for timely delivery of medical packages to remote locations. They tackled the problem of limited payload capacity by scheduling and sequencing a set of deliveries. Gentili et al. [24] first addressed the problem of locating the platforms as well concurrently determining which platform serves which demand points and in what order, in order to minimize total disutility for product delivery. Then, the two-period problem where the platforms can be relocated, using usable road network, after the first period.

UAVRPs have been studied for heterogeneous UAVs. Chen et al. [26] dealt with the path planning problem of UAVs with different abilities in multi-region systems. Inspired by density-based clustering methods, they first designed an algorithm to classify regions into clusters and obtained approximate optimal point-to-point paths for UAVs, such that the coverage task is performed correctly and efficiently. In the other study by Chen et al. [27], they focused on the coverage path planning problem of heterogeneous UAVs, and present an ant colony system (ACS)-based algorithm to obtain good enough paths for UAVs and fully cover all regions efficiently.

These models are limited in assuming items only be transported by UAVs without considering a concept of equity. In the transportation of items at the time of a disaster, often referred to as post-disaster HL, it is necessary to consider equity in the distribution of relief supplies as an important issue [1,44]. In addition, existing heuristic approaches are not yet efficient enough to solve large-scale problems or problems in dynamic environments. Therefore, this study focuses on a typical reinforcement learning method, QL, which is a powerful approach to complex sequential decision-making problems with large or continuous state and action spaces [45]. UAVRPs studied in this study is a multi-period sequential decision problem with a large number of states and actions, which is suitable for QL. RL method can be used to model the decision-making of agents that can adapt to dynamic

environments based on learning from previous knowledge, and, thus, has potential for application to dynamic real-world environments in post-disaster situations. For example, RL have been studied in HL, such as the deployment of emergency infrastructure, the selection of rescue paths and the prediction of relief demand [46]. However, there are few studies in which reinforcement learning has been applied to UAVRPs in post-disaster HL, so this study developed a QL method for planning the transportation of supplies. This study aims to optimize transport in static scenarios where the state of the environment does not change, but the methods in this study can be extended and applied to optimization in dynamic scenarios.

3. Model Description

In this section, we describe a model that assumes the transportation of supplies during a disaster. The model is composed of an environment and agents that move in the environment.

Notations are summarized in Table 1.

Table 1. Notations description.

Notation	Description
Environment	
N_S	Number of shelters
j, k	Indices for shelters
$S = \{0, 1, 2, \dots, N_S\}$	Set of shelters including the depot ($j = 0$)
N_I	Number of item types
i	Index for item types
$I = \{1, 2, \dots, N_I\}$	Set of item types
d_{ijt}	Demand of the item i of the shelter j at time instant t
$D_{jt} = \{d_{1jt}, d_{2jt}, \dots, d_{N_Ijt}\}$	Set of demands of the shelter j at instant t
$Y_t = \{D_{1t}, D_{2t}, \dots, D_{N_S t}\}$	Set of the remaining demand of all shelters at instant t
h_{jk}	Distance between shelter j and k
p_{ij}	Penalty cost of item i
b_{ij}	Time limit of the item i of the shelter j
Agent	
N_U	Number of UAVs
l	Index for UAVs
$U = \{1, 2, \dots, N_U\}$	Set of UAVs
w_{ijklmn}	Amount of item i transport from shelter j to k as n th location in the trip m by UAV l
N_l	Number of trips of UAV l
m	Index for trip
M_{lm}	Number of location that UAV l traveled in trip m (include depot)
n	Index for number of location
C	Maximum payload of UAV
E	Maximum amount of energy
a	Acceleration
V_{max}	Maximum speed
t_{take}	Take-off time
t_{land}	Landing time
t_{serve}	Servicing time

Table 1. Cont.

Notation	Description
t_{jk}	Flight time between shelter j and k
f_{jk}	Transportation cost from shelter j to k
v	Amount of UAV payload
δ_0	Energy needed for take-off and landing for an empty UAV [18]
δ	Additional energy amount needed for take-off and landing with an additional item [18]
ρ_0	Energy to fly one distance unit for an empty UAV [18]
ρ	Additional energy amount needed to fly one distance unit with one item [18]
L_m	Destination shelter of UAV m at instant t
u_{jlm}	Time when UAV l transport items for shelter j in trip m
Algorithm	
t	instant
A_t	Action at certain instant t
S_t	Agent state at certain instant t
R_t	Reward at certain instant t
Q	Action-value function
α	Learning rate
γ	Discount factor
ϵ	Chance of choosing a random action
N_E	Number of episodes
T	Termination time

3.1. Problem Definition

In this study, we assumed a scenario in which last-mile distribution to isolated shelters due to tsunami inundation is performed. In this scenario, multi-rotor UAVs are used due to its ability to provide vertical take-off and landing (VTOL). It was assumed that sufficient quantities of relief supplies and batteries for recharging were stored at the nearest depot (distribution center) and that the demands of the isolated shelters were known in advance. To determine the UAV transport routes and the transport supplies strategy, we formulated the UAVRPs as multi-objective, multi-trip, multi-item, and multi-UAVs. Each UAV is subject to battery and payload limitations and cannot meet the demands of all shelters in a single trip; therefore, it makes multiple round trips between shelters and a depot until it completes. The case of road blockage that limits land access and the unavailability of helicopters for air supply transportation is also considered in the presented scenario. Therefore, the exclusive use of multi-UAVs is presented here. We acknowledge that the future incorporation of multi-UAVs in disaster response require specific protocols for the coordination and organization of these units together with other human-operated vehicles and aircrafts. A more detailed explanation is given below.

3.2. Environment

To estimate the transportation of supplies in the event of a disaster, we selected the depot and shelters at the Ushioe district in Kochi City, Kochi Prefecture, Japan. The map of the Ushioe district is shown in Figure 1. In the case of the occurrence of the Nankai Trough earthquake, this area is expected to be inundated by tsunamis (Figure 2) and to face land subsidence caused by the earthquake. It is assumed that long-term inundation will continue even after the tsunami recedes. Therefore, relief activities in isolated disaster areas are an important issue. According to the tsunami inundation forecast data of the Ministry of Land, Infrastructure, Transport and Tourism of Japan (MILT) [47], six of the seven designated tsunami shelters in the target area are expected to be inundated by the tsunami and flooded for a long period of time. Since there are no heliports at the shelters, it is likely that vehicles, such as trucks, and aircraft, such as helicopters, cannot transport

emergency relief supplies during inundation. In this scenario, multiple UAVs can transport the supplies in the area.

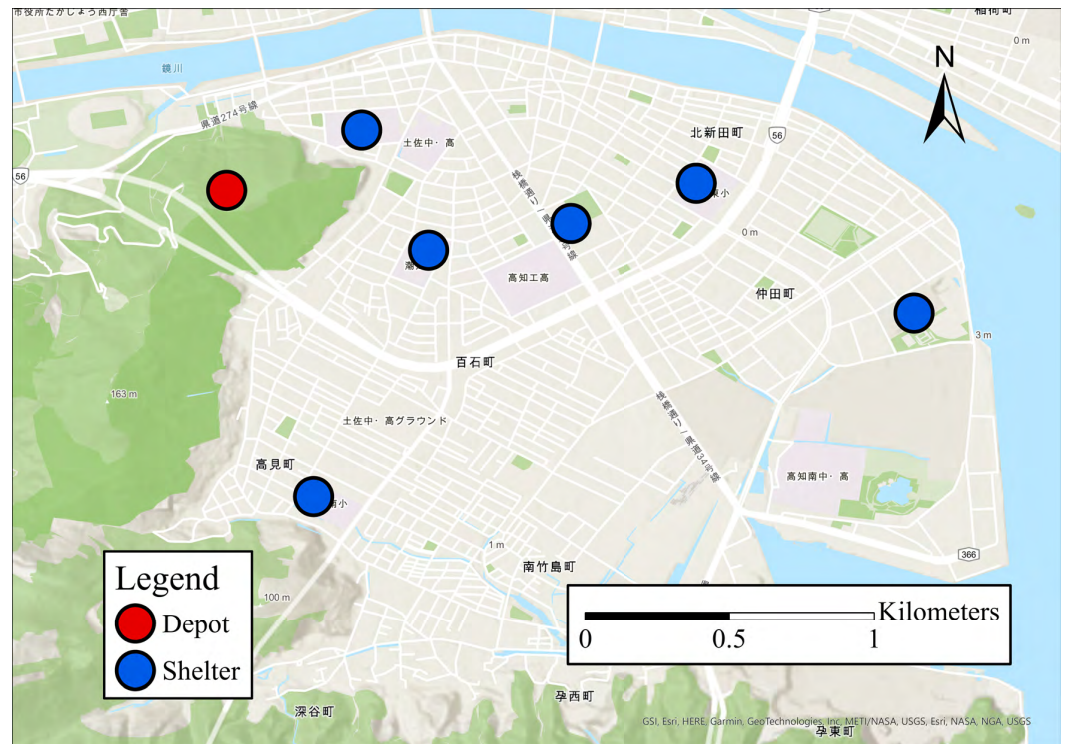


Figure 1. Depot (red) and shelters (blue) in Ushio district in Kochi City, Kochi Prefecture, Japan.

A set of shelters $S = \{0, 1, 2, \dots, N_S\}$ are set up in the area, where ($j = 0$) is the depot and the set of $\{1, 2, \dots, N_S\}$ represents shelters where emergency relief supplies are delivered. In this study, one shelter, which is not inundated by the tsunami, is set as depot ($j = 0$) for the transportation of supplies, and a logistics model is developed for transporting supplies from the depot to the other six shelters ($N_S = 6$).

In the event of a disaster, there will be shortages of various types of supplies, such as medicine, medical equipment, sanitary materials, food, water, and clothes. However, the priority and amount of each type of supply will differ. Therefore, it is important to consider the order of transportation and the amount to be transported according to the priority of the supplies in the transportation plan for emergency relief supplies. In this study, we focus on the duration of the “three-day crucial rescue period”, which is the first 72 h after a disaster [48], and plan to transport several kinds of lightweight and urgent emergency relief supplies (e.g., medicine, medical equipment, and sanitary materials).

As a mathematical formulation, $I = \{1, 2, \dots, N_I\}$ represents the set of item types, $D_{jt} = \{d_{1jt}, d_{2jt}, \dots, d_{ijt}, \dots, d_{N_Ijt}\}$ represents the set of each shelter’s demand at instant t , and i is the index for item types. We assume that the demand of each item is more than one unit ($d_{ijt(t=0)} \geq 1 \quad \forall i \in I, \forall j \in S$).

In addition, to consider the priority of the supplies, we set the priority rate as the weight p_{ij} and the time limit b_{ij} for the item i of the shelter j . Here, the “time limit” of an item is the “expiration time” at which the item is no longer usable (e.g., perishable food). In this study, supplies with high urgency were given a higher priority (higher penalty cost when not delivered) and a shorter transportation time limit.

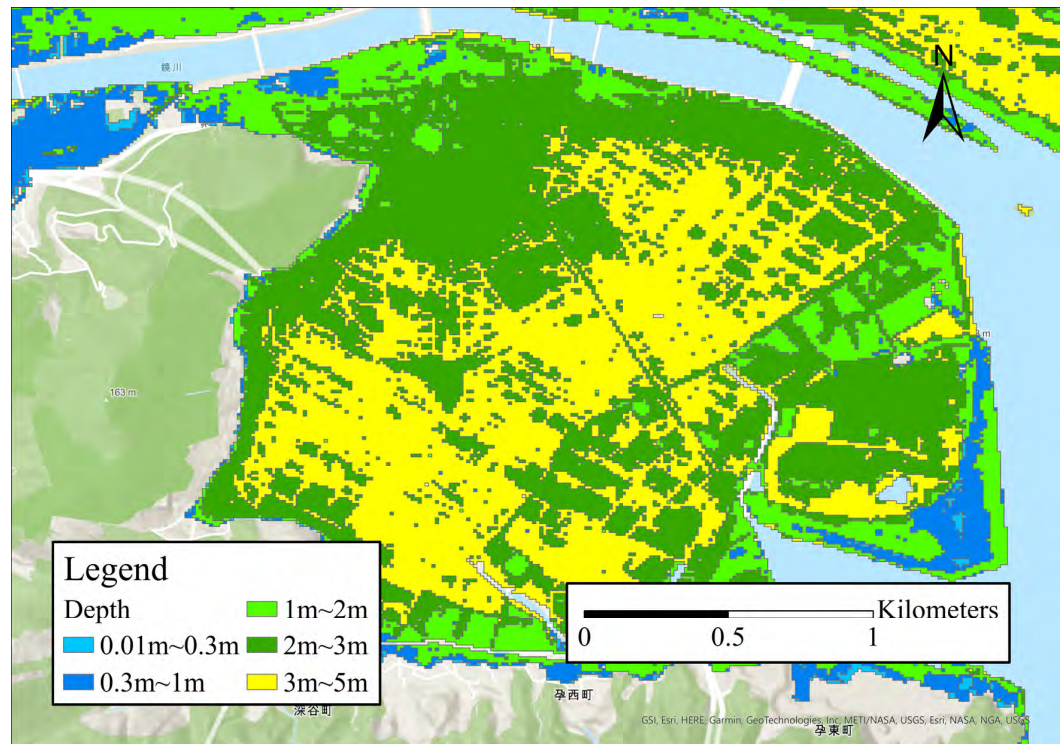


Figure 2. Estimated tsunami inundation depth for the Nankai Trough earthquake in Ushioe district in Kochi City, Kochi Prefecture, Japan. Data from [47].

3.3. Agent

The agent is assumed to be a UAV that transports emergency relief supplies. A UAV loads the supplies at depot ($j = 0$) and delivers them to shelters. A set of UAVs $\mathbf{U} = \{1, 2, \dots, N_U\}$ is used to transport them. N_U represents the number of UAVs.

We assume that UAVs are subject to the following conditions:

- All UAVs are homogeneous; thus, they have the same maximum payload C and maximum amount of energy E ;
- At the start of the transport, all UAVs are assumed to be at the depot;
- The batteries are fully charged when the UAV takes off from the depot;
- Since there are sufficient batteries for the UAV in the depot, the battery charging time is not considered;
- Each UAV has its batteries replaced only at the depot.

Each UAV consumes a battery depending on its payload and traveling distance. The energy consumption function (Equation (1)) describes the amount of energy used by the UAV to travel from shelter j to k with payload v , referring to the study of Rabta et al. [18].

$$R(v) = \delta_0 + \delta v + h_{jk}(\rho_0 + \rho v) \tag{1}$$

$$v = \sum_n \sum_{i=1}^{N_I} w_{ijklmn} \tag{2}$$

Here, h_{jk} is the distance between shelters j and k , and δ_0 is the energy for take-off and landing without supplies. δ is the additional energy amount needed for take-off and landing with an additional item, ρ_0 is the energy to fly one distance unit for an empty UAV, and ρ is the additional energy amount needed to fly one distance unit with one item. In Equation (2), w_{ijklmn} denotes the amount of item i transported from shelter j to k as the n th location in trip m by UAV l , and v denotes the total amount of items transported after the n th location in trip m by UAV l .

From the above assumptions, the UAV's maximum payload and maximum amount of energy constraints can be represented as follows:

$$\sum_{n=1}^{M_m} \sum_{i=1}^{N_l} w_{ijklmn} \leq C \quad \forall l, \forall m \tag{3}$$

$$\sum_{n=1}^{M_m} R(v) \leq E \quad \forall l, \forall m \tag{4}$$

The transportation cost f_{jk} , which is equal to the consumption time, from shelter j to k is defined as Equation (5), referring to the study of Nagasawa et al. [34].

$$f_{jk} = t_{take} + t_{land} + t_{serve} + t_{jk} \tag{5}$$

$$t_{jk} = \begin{cases} \frac{h_{jk}}{V_{max}} + \frac{V_{max}}{a} & \left(h_{jk} > \frac{V_{max}^2}{a} \right) \\ 2\sqrt{\frac{h_{jk}}{a}} & \left(h_{jk} < \frac{V_{max}^2}{a} \right) \end{cases} \tag{6}$$

where t_{take} is the take-off time, t_{land} is the landing time and t_{serve} is the servicing time (e.g., time to change batteries in the depot or unload supplies in the shelter). t_{jk} denotes the flight time between shelters j and k . Since each UAV is presumed to accelerate to its maximum speed with uniform acceleration, t_{jk} can be calculated as Equation (6), where V_{max} is the maximum speed of a UAV, and a represents the uniform acceleration of a UAV. In addition, one trip m is defined as the period from the time the UAV leaves the depot to the time it transports supplies to one or more shelters and returns to the depot.

3.4. Cost Functions

This study aims to plan optimal transportation by multiple UAVs from three perspectives: rapidity, urgency, and equity. Following those objectives, the proposed UAVRP aims to minimize the total cost, including the following three costs:

- Flight Time Cost (FC): The cost based on the total flight time of all UAVs. This corresponds to rapidity;
- Priority Cost (PC): The cost based on quick transportation of high-priority supplies. This corresponds to urgency;
- Equity Cost (EC): The cost based on equitable transportation to all shelters. This corresponds to equity.

First, FC is described as follows:

$$Z_1 = \sum_{l=1}^{N_U} \sum_{m=1}^{N_l} \sum_{n=1}^{M_m} \sum_{j=0}^{N_S} \sum_{k=0}^{N_S} f_{jk} \cdot x_{jklmn} \tag{7}$$

$$x_{jklmn} = \begin{cases} 1 & \text{(UAV } l \text{ transport from } j \text{ to } k \text{ as } n\text{th location in the trip } m) \\ 0 & \text{(otherwise)} \end{cases} \tag{8}$$

The cost function Z_1 in Equation (7) is the total flight time for all UAVs and all trips. f_{jk} is the time taken for transportation from shelter j to k , which is represented in Equation (5), and x_{jklmn} is a decision variable, which is represented in Equation (8).

Second, PC is described as follows:

$$Z_2 = \sum_{l=1}^{N_U} \sum_{m=1}^{N_l} \sum_{n=1}^{M_{lm}} \sum_{j=0}^{N_S} \sum_{k=0}^{N_S} \sum_{i=1}^{N_I} P_{ijt}(w_{ijklmn}) \cdot x_{ijklmn} \tag{9}$$

$$P_{ijt}(w_{ijklmn}) = p_{ij} \cdot w_{ijklmn} \cdot \max\{(u_{ijlm} - b_{ij}), 0\} \tag{10}$$

$$u_{ijlm} = \sum_{m=1}^m \sum_{n=1}^{M_{lm}} \sum_{j=0}^{N_S} \sum_{k=0}^{N_S} f_{jk} \cdot x_{ijklmn} \tag{11}$$

The cost function Z_2 in Equation (9) is equal to the total penalty cost of items, which depends on the transportation time and urgency of items. w_{ijklmn} represents the amount of item i transported to shelter j ($j \neq 0$) as the n th location in trip m , p_{ij} represents the penalty cost of item i , and u_{ijlm} represents the time when UAV l transports items for shelter j in trip m , which is described in Equation (11). Equation (10) represents the penalty cost when the UAV transports item i for shelter j later than the time limit b_{ij} . On the other hand, there is no penalty cost when the UAV transports item i for shelter j within the time limit b_{ij} .

Finally, we define the (EC). In previous studies, various cost functions have been defined to describe equity [44,49]. In this study, we use the EC of Huang et al. [29], which is described as follows:

$$Z_3 = \sum_{t=1}^T \sum_{j=1}^{N_S} g(r_{jt}) \tag{12}$$

$$r_{jt} = \frac{\sum_{i=1}^{N_I} d_{ijt}}{\sum_{i=1}^{N_I} d_{ijt}(t=0)} \tag{13}$$

The cost function Z_3 in Equation (12) indicates the “disutility-weighted arrival time”. The disutility function (12) encourages UAVs to not necessarily satisfy a shelter’s entire demand but rather to save supplies to serve another shelter. Equation (13) represents the rate of remaining demand of shelter j at instant t . The following piecewise linear disutility function is used in our calculations:

$$g(r) = \begin{cases} \frac{4r}{13} & (r < 0.25) \\ \frac{8r-1}{13} & (0.25 \leq r < 0.5) \\ \frac{16r-5}{13} & (0.5 \leq r < 0.75) \\ \frac{24r-11}{13} & (0.75 \leq r) \end{cases} \tag{14}$$

where r is the rate of remaining demand. The above function can be represented as in Figure 3.

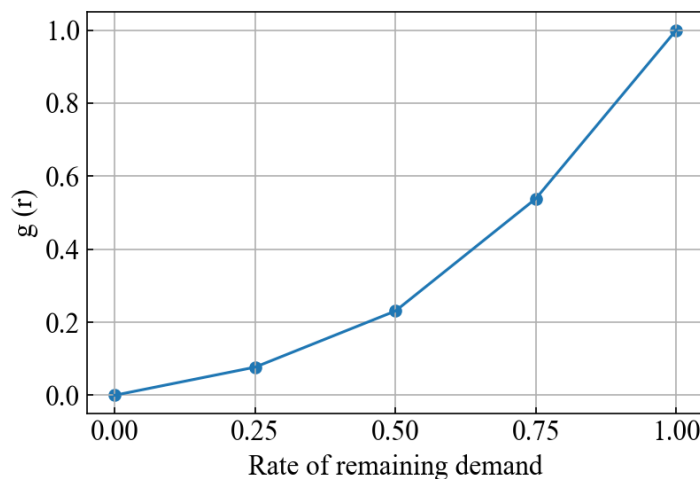


Figure 3. The disutility function for unsatisfied demand.

These three cost functions are converted to a single objective function using the weighted sum method. The objective function of this problem is described as follows:

$$\text{Minimize} \quad \lambda_1 Z_1 + \lambda_2 Z_2 + \lambda_3 Z_3 \tag{15}$$

where λ_1, λ_2 and λ_3 are the weights of the three costs.

4. Proposed Method

4.1. Q-Learning

RL is a framework for an agent to learn appropriate strategies by trial and error in an environment by obtaining rewards from that environment. RL is a machine learning method in which an agent learns by itself to determine which action to choose to maximize the total reward. Figure 4 shows the interaction between the agent and environment in RL. When the agent observes the state S_t of the environment at a certain instant t , the agent chooses an action A_t among all possible actions. The chosen action will have repercussions in the environment, and, consequently, it will influence the state at the instant $t + 1, S_{t+1}$. Whether the new state S_{t+1} is positive or negative to the agent’s main objective, it is quantified with a reward r_{t+1} . The RL approach searches for the policy that gives the highest long-term reward R_t . In this study, we use QL [30], which is a typical RL method.

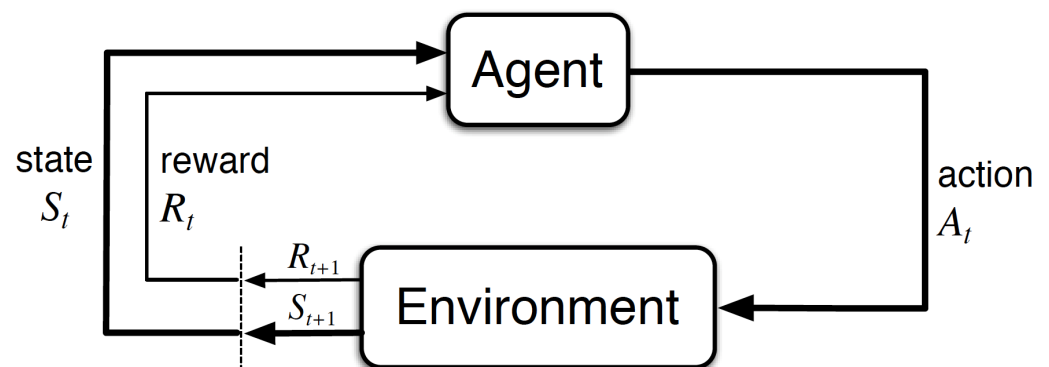


Figure 4. The agent from a reinforcement learning point of view. Image from [50].

In QL, the agent has an action-value matrix, Q matrix, which represents the value of being in a specific state S_t while choosing an action A_t at instant t . By trying different actions in different states (exploration) but also by choosing the best possible action that gives the highest reward value based on its past experience (exploitation), QL is shown to converge for any type of policy being followed [50]. The agent updates a value of the Q matrix, which is represented as $Q(state, action)$, for each action. It is defined as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{16}$$

where $Q(s_t, a_t)$ is the current Q matrix, α is the learning rate ($0 \leq \alpha \leq 1$) and γ is the discount factor ($0 \leq \gamma \leq 1$). r_{t+1} is the expected reward at instant t , and $\max_a Q(s_{t+1}, a)$ is the maximum action value among all possible actions in state s_{t+1} . In this study, the initial value of the Q matrix was set to 0 for all actions in all states of the environment.

4.2. State and Action

In this study, the agents share a single Q matrix, which is updated sequentially by each agent after each transportation. The state of the environment in the Q matrix and the action of the agent corresponding to a UAV were defined as follows:

- States: A state at instant t, S_t , represents the set of the remaining shelter demand at instant t . S_t is denoted by $S_t = Y_t = \{D_{1t}, D_{2t}, \dots, D_{st}\}$, where D_{jt} is the set of demands of shelter j at instant t .

- **Actions:** Selection of the following two elements among all possible actions: (i) the shelter/depot to transport/return and (ii) the amount of items that the UAV transports. Note that the types of supplies are selected in order of urgency among the supplies demanded by the destination shelter. An action at instant t , A_t , is denoted by $A_t = \{L_m, w_{ijklmn}\}$.

Now consider the following example, where $N_U = 1$. The initial state $S_{t=0}$ is denoted as $S_{t=0} = \{\{2,2\}, \{1,1\}\}$, which means that there are 2 shelters for transportation and the initial demand for Shelter 1 and Shelter 2 are $D_{1t(t=0)} = \{2,2\}$ and $D_{2t(t=0)} = \{1,1\}$, respectively. We assume that the initial location of UAV 1 is depot ($j = 0$). In the case of the action of the UAV, 1 is $A_{t=0} = \{1,3\}$, which means that UAV 1 transports 3 units of supplies to Shelter 1, the new state at instant $t = 1$, $S_{t=1}$, represented as $S_{t=1} = \{\{0,1\}, \{1,1\}\}$. Subsequently, in the case of the action of the UAV 1 is $A_{t=1} = \{2,1\}$, which means that UAV 1 transports 1 unit of supplies to Shelter 2, the new state at instant $t = 2$, $S_{t=2}$, represented as $S_{t=2} = \{\{0,1\}, \{0,1\}\}$. By repeating this process, the action of the agent continues until the current episode is completed.

In each state S_t , the UAV (agent) determines action A_t according to the flow, as shown in Figure 5. Here, we assume that there are six shelters, as shown in Figure 1, and that the maximum payload of the UAV is 5 ($C = 5$). First, the agent generates “All_action_list” for each state. This is the set of combinations of destinations and amounts of transport: When the agent is at the depot, there are 30 combinations (6 destinations \times 5 amounts), and when the agent is at a certain shelter, there are 26 combinations (5 destinations except the current location \times 5 amounts + return to the depot). Second, the agent determines an action A_t based on the ϵ -greedy policy (for the policy, see the next section). Finally, if the constraints of the maximum payload (Equation (3)) and maximum amount of energy (Equation (4)) are satisfied, action A_t is chosen; otherwise, the current action is deleted from “All_action_list”, and the action is chosen again.

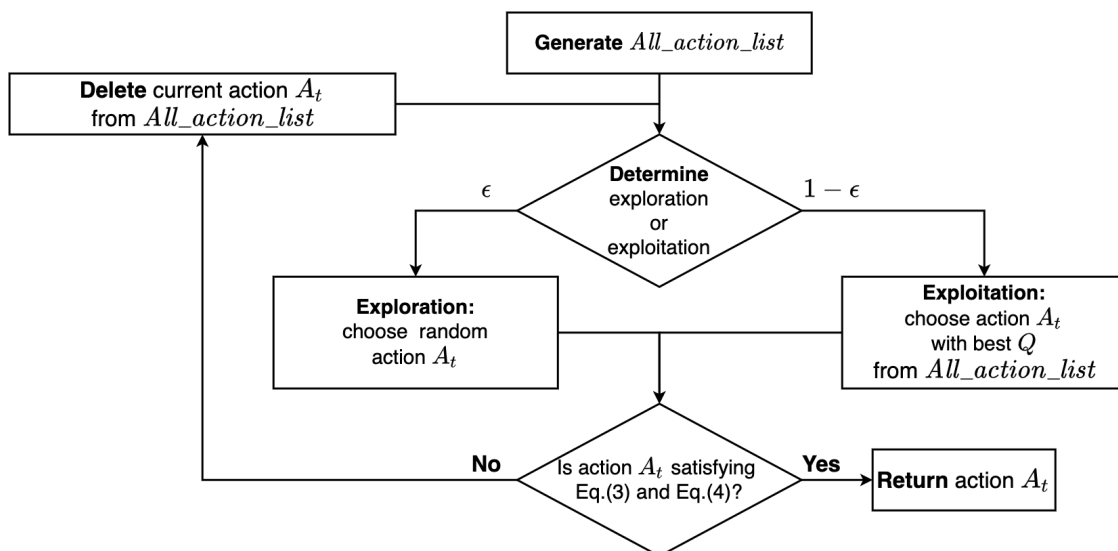


Figure 5. The flow of UAV action determination at each action instance of each episode.

The termination criteria are the conditions when the demands for all shelters are zero. It is described as follows:

$$d_{ijt} = 0 \quad \forall i, \forall j \tag{17}$$

The termination time ($t = T$) was defined as the time when the above conditions were met and the UAV returned to the depot.

4.3. Policy

In QL, since the ratio of exploitation to exploration has a significant impact on the learning results, it is necessary to adjust it well. We adopted the ϵ -greedy policy [50] as a decision strategy for actions. In the ϵ -greedy policy, parameter $\epsilon (0 \leq \epsilon \leq 1)$ is set to control the degree of exploration: the agent explores an action selected at random with probability ϵ and exploits the action that is the highest value with probability $(1 - \epsilon)$. We use Equation (18) in such a way that the parameter ϵ decays according to the episodes, referring to the study of Yu et al. [46].

$$\epsilon = \frac{0.5}{1 + e^{\left(\frac{10 \times (episode - 0.4 \times N_E)}{N_E}\right)}} \tag{18}$$

where N_E is the total number of episodes and *episode* denotes the number of the current episode. Equation (18) guarantees a high probability of exploration in the early training episodes and a low probability of exploitation in the late training episodes. Figure 6 shows the value of parameter ϵ at each episode.

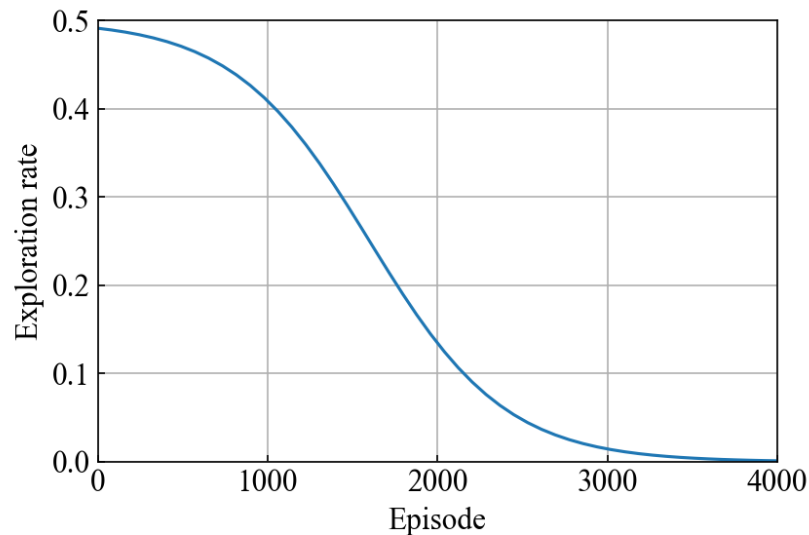


Figure 6. The ϵ -greedy exploration function when $N_E = 4000$.

4.4. Reward

The proposed UAV routing problem aims to minimize the total costs, including the flight time cost (FC), the priority cost (PC), and the equity cost (EC). Based on the previous section, FC , PC , and EC are represented in Equations (7), (9), and (12). Notably, the objective of this study is to minimize the total costs, whereas the objective of QL is to maximize the total rewards. Therefore, the objective function is adjusted to be negative, such that the objective of this study translates into maximizing the negative value of the total costs. The reward function is represented as follows:

$$R_t = \begin{cases} -(\lambda_1 Z_1 + \lambda_2 Z_2 + \lambda_3 Z_3) & (t = T) \\ 0 & (\text{otherwise}) \end{cases} \tag{19}$$

where T is the time when the episode terminates.

Finally, Figure 7 shows the flowchart of the QL process so far.

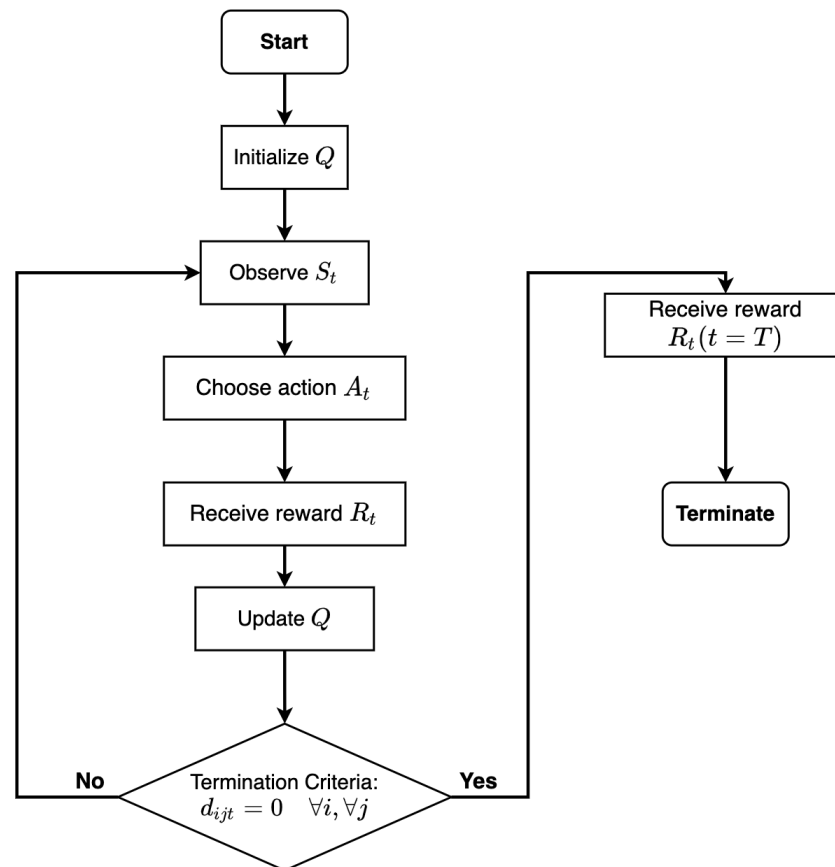


Figure 7. QL algorithm procedure at each episode.

5. Numerical Experiments

5.1. Simulation Settings

To examine the effectiveness of QL, a simulation environment was created and performed in Python. All of the experiments were implemented on a laptop PC with quad Intel Core 2.4 GHz CPU, 8GB of RAM. In this study, simulations were performed in three scenarios in which the demand for shelter supplies was changed to small, medium, and large amounts. The simulation parameters for each scenario are shown in Table 2. It is noted that in all scenarios there are three types of supplies (Items A, B, and C), with A being the most urgent and C being the least urgent. UAV parameters are shown in Table 3.

Table 2. The set parameters of each scenario.

Parameter	Scenario 1	Scenario 2	Scenario 3
Number of UAVs N_U	3	3	3
Number of item types N_I	3	3	3
Initial demand of each shelter $\{d_{1jt}, d_{2jt}, d_{3jt}\}, \forall j, t = 0$	{2, 2, 2}	{4, 4, 4}	{8, 8, 8}
Priority rate of each item $\{p_{1j}, p_{2j}, p_{3j}\}, \forall j$	{2, 1, 0.5}	{2, 1, 0.5}	{2, 1, 0.5}
Time limit of each item $\{b_{1j}, b_{2j}, b_{3j}\}, \forall j$	{400, 800, 1200}	{800, 1600, 2400}	{1800, 3600, 5400}
Number of episodes N_E	8000	24,000	24,000

Table 3. UAV parameters.

Parameter	Value
V_{max}	10 m/s
a	1 m/s ²
C	5 kg
E	275 kJ
δ_0	900 J
δ	300 J/kg
ρ_0	3 J/m
ρ	1 J/m · kg
t_{take}	30 s
t_{land}	30 s
t_{serve}	60 s

5.2. Parameter Settings

The rate of exploration (ϵ) has a significant impact on the speed of convergence in QL. To compare the exploration rate and convergence speed, the following three cases of exploration rates are used to compare the objective values for each episode: $\epsilon = 0.1$, $\epsilon = 0.5$, and $\epsilon = \text{Equation (18)}$. Here, the weights of the three costs are defined as $\lambda_1 = \lambda_2 = \lambda_3 = 1$, the initial demands of each shelter are defined as $\{d_{1j}, d_{2j}, d_{3j}\}(t = 0) = \{1, 1, 1\}$, and the other parameters are defined as in Scenario 1 in Table 2. In that scenario, the following learning rate α and discount factor γ are used, based on the study by Sutton et al. [50]: $\alpha = 0.9, \gamma = 0.1$. The convergence results are shown in Figure 8. In Figure 8, the average value of the 10 training trials is plotted for each exploration rate.

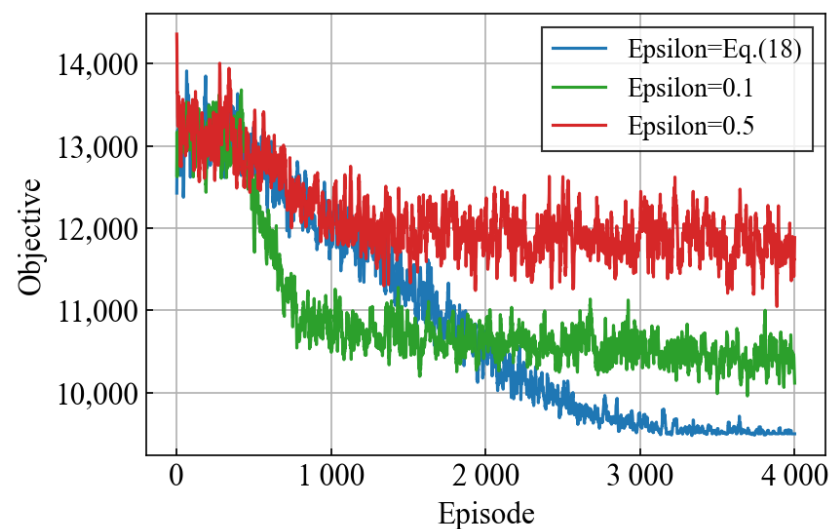


Figure 8. The learning curve obtained with QL considering different exploration rates in the test scenario.

As shown in Figure 8, when $\epsilon = 0.1$, it converged faster during fewer episodes than at other instances. However, as the number of episodes approaches the set maximum number, it can be seen that the values converge to the optimum objective value more certainly at $\epsilon = \text{Equation (18)}$ than at $\epsilon = 0.1$ and that the objective value also is more stable at $\epsilon = \text{Equation (18)}$. Therefore, in later experiments in our study, we use $\epsilon = \text{Equation (18)}$ as the exploration rate.

5.3. Comparison of Methods

To evaluate the performance of QL in this particular problem, we compared the performance of the QL and heuristic algorithms, the genetic algorithm (GA) and particle swarm optimization (PSO), in the model studied by Jiang et al. [21]. The reason for selecting the model of Jiang et al. [21] is based on the similarity in the parameterization used in our study. Other candidates for comparison could have been the models proposed by Shi et al. [19] and Ghelichi et al. [20], however, these lack some features considered in our study (i.e., the role of delivery and pickup, selection of charging stations).

In their model, 3 UAVs transport items from 1 depot to 10 demand points, as shown in Figure 9, and the following two constraints are considered:

- Maximum payload of UAV;
- Time window for transportation at each point.

The objective is to minimize the total transportation distance, taking into account the above two constraints. We performed QL on the same problem setting. The set parameters are shown in Table 4. It should be noted that the number of iterations is different between QL and GA and PSO since QL generates one transportation plan in each iteration, while the GA and PSO generate 40 delivery plans in each iteration.

Table 4. The set parameters of the GA, PSO, and QL.

Parameter	Value		
	GA	PSO	QL
Population number	40	40	-
Iteration	100	100	4000
Calculation times	50	50	50

The minimum transport distances obtained by each method are shown in Table 5, which is the minimum value of all trials. It is noted that this study does not compare computation times due to differences in computer performance. Figure 9 shows the shortest route obtained by PSO and QL. From these results, it can be observed that the QL method is useful for the UAV transportation planning problem since it can find a shorter route than the other methods. The reason why QL obtained a better solution than other methods is considered to be the large number of evaluations in QL. In QL, the state of each environment is evaluated at the end of each action, which may have resulted in earlier convergence than methods that evaluate plans after all actions have been completed, such as PSO and GA. For this reason, it can be said that the QL method designed in this study is useful to improve the efficiency of the item transportation route.

Table 5. Comparison of minimum transportation costs of each method.

	Algorithm	Minimum Distance [km]
Jiang et al. [21]	PSO	350.30
Jiang et al. [21]	GA	379.04
our study	QL	322.44

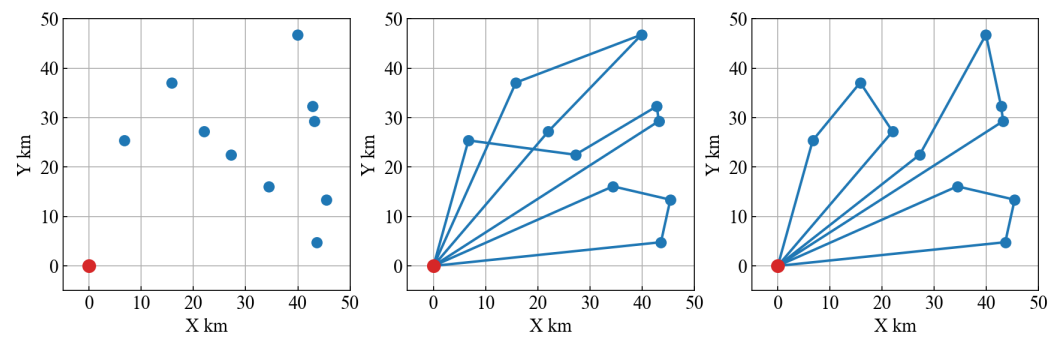


Figure 9. The model of Jiang et al. [21]. (left) Location coordinates of 1 depot (red) and 10 demand points (blue). (center) The optimal route obtained by PSO of Jiang et al. [21]. (right) The optimal route obtained by QL of our study.

5.4. Performance Comparison

In this study, we evaluated the transport of items for each case against our model by conducting a comparative examination of the transport of five cases with different rates of three objective assignments in QL. The cases are listed in Table 6. In Table 6, Case 1 minimizes the total flight cost based on the study of Dorling et al. [17]. Case 2 minimizes the sum of the total flight cost and the penalty cost of items based on the study of Jiang et al. [21]. Case 3 minimizes the equity cost based on the study of Huang et al. [29]. Case 4 and Case 5 minimize the sum of three costs and two costs other than the total flight cost, as introduced in our study. QL was performed 10 times for each case in each scenario and analyzed from three perspectives (rapidity, urgency, and equity). For each trial, the best performance (lowest objective value) transportation plan is selected for comparison.

Table 6. Rate of reward for each case.

	Case	Rate		
		Rapidity (λ_1)	Urgency (λ_2)	Equity (λ_3)
Dorling et al. [17]	Case 1	1	0	0
Jiang et al. [21]	Case 2	0.5	0.5	0
Huang et al. [29]	Case 3	0	0	1
Our study	Case 4	0.33	0.33	0.33
Our study	Case 5	0	0.5	0.5

Figure 10 shows the total flight time of all UAVs in each case of each scenario. Numerical values denote the average of 10 trials, and error bars denote 95% confidence intervals. As shown in Figure 10, for all scenarios, Case 1 is seen to have the shortest flight time, while for the other cases, the length of flight time is shown to vary from scenario to scenario.

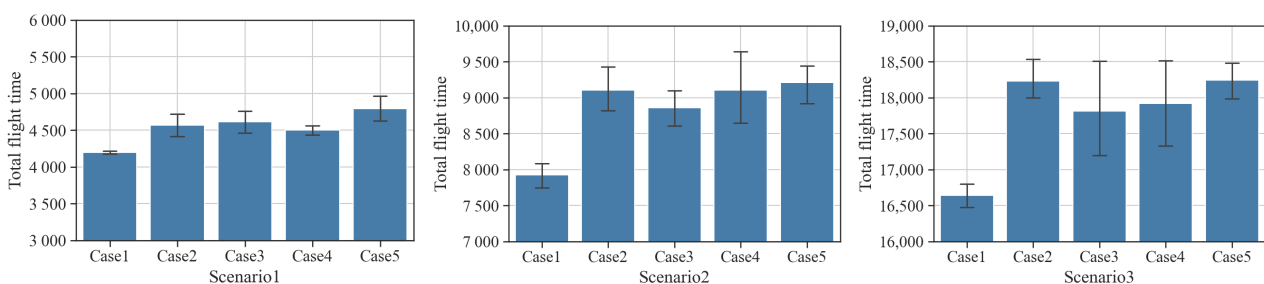


Figure 10. Total flight time of the UAV in each scenario. The values are averages of 10 training trials, and the error bars represent 95% confidence intervals.

Figure 11 shows the penalty cost for each item in each scenario. The penalty cost depends on the ends of the transport time and the urgency of the item. In all scenarios, Case 1 has by far the highest penalty cost for urgent items (especially Item A). Therefore, the transportation strategy of Case 1 is not suitable for the transportation of emergency relief supplies in the event of a disaster that requires rapid transportation of supplies and has a wide range of supply urgency. Case 2, Case 3, Case 4, and Case 5 have similar values for all items in Scenario 1 and Scenario 2. On the other hand, as the size of the instance increases, the penalty cost of Case 3 increases relative to the other cases. Figure 12 shows the total service level of each shelter in each scenario. Case 3 has the lowest cost, and Case 1 has the highest cost in all scenarios. Figure 12, Figure 10, Figure 11, Case 2, Case 4, and Case 5 have similar trends. This could be because the initial demand for all shelters is the same. To minimize the objective function of urgency Z_2 , a transportation strategy that delivers items with high urgency to all shelters quickly and then delivers items with low urgency to all shelters is effective. Similarly, the objective function equity Z_3 can be decreased by transporting a small number of items to each shelter quickly.

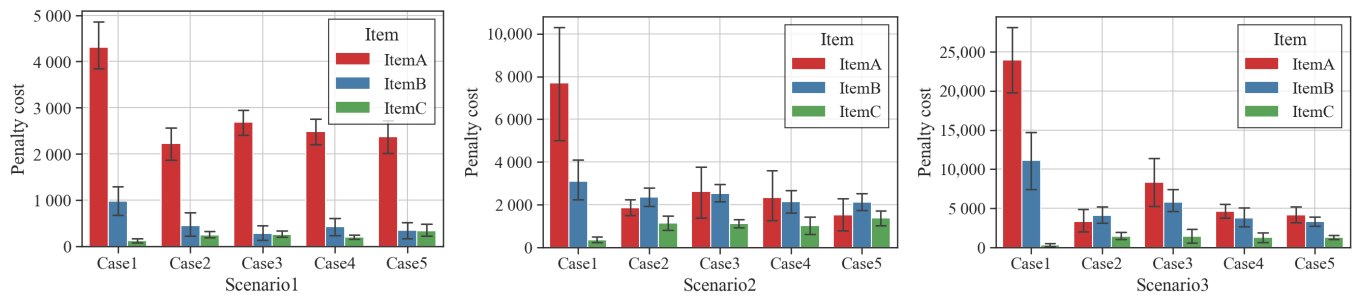


Figure 11. Penalty cost for each item in each scenario. The values are averages of 10 training trials, and the error bars represent 95% confidence intervals.

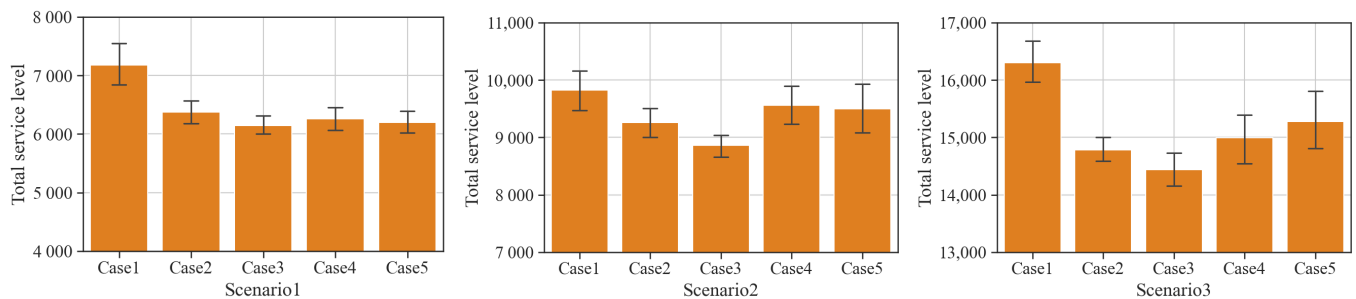


Figure 12. Total service level of each shelter in each scenario. The values are averages of 10 training trials, and the error bars represent 95% confidence intervals.

Figures 13 and 14 show the route and resource allocation to minimize the objective value of 10 trials in Scenario 1 of Case 1 and Case 4, respectively. Figure 13 shows that in each trip, each UAV transports a large number of items to a single shelter. On the other hand, Figure 14 shows that each UAV is transporting supplies to multiple shelters on a single trip. In Figure 14, in terms of the amount of supply for each shelter, it can be confirmed that the supply is widely distributed to many shelters and that QL enables learning in accordance with the objectives of the study.

Figure 15 shows the minimum objective values for each number of UAVs in Scenario 1. In Figure 15, the average value of the 10 trials is plotted for each cost. In this case, the weights of the cost functions were set to $\lambda_1 = \lambda_2 = \lambda_3 = 1$, and numerical values denote the average of 10 times. When the number of UAVs is changed from 1 to 2, a 65% decrease in total cost is observed, and when the number of UAVs is changed from 2 to 3, a 55% decrease in total cost is observed. Thereafter, total costs continue to decrease, although the rate of decrease in total cost per unit decreases. In the case of UAVs larger

than 4, little change is observed in the rapidity cost and the urgency cost, and only the equity cost continuously decreases. In addition, when the number of UAVs is larger than 4, the urgency cost converges to almost zero. It can be concluded that the transportation of all items has been completed within the time limit for each item type. The result suggests how many UAVs are required to meet the time limit for the shelters. In the case that the time limit cannot be met, we suggest that fair transportation during a disaster could be achieved by considering the equity cost, a metric to maximize satisfaction bias by reducing unfair supply allocation.

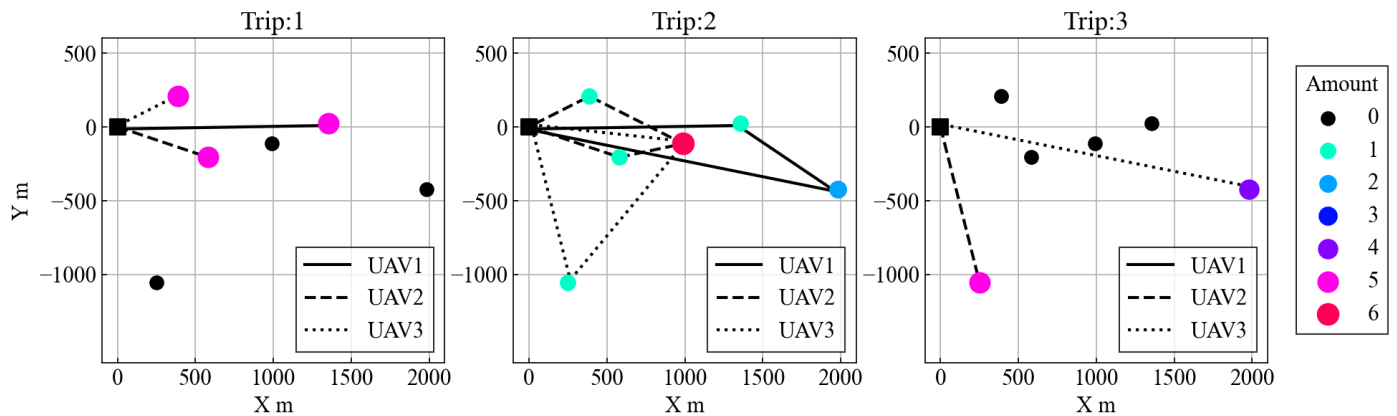


Figure 13. Route and resource allocation to minimize the objective value of Case 1 in Scenario 1. The lines represent the route of each UAV, and the color and size of the dots represent the amount of supplies supplied to each shelter.

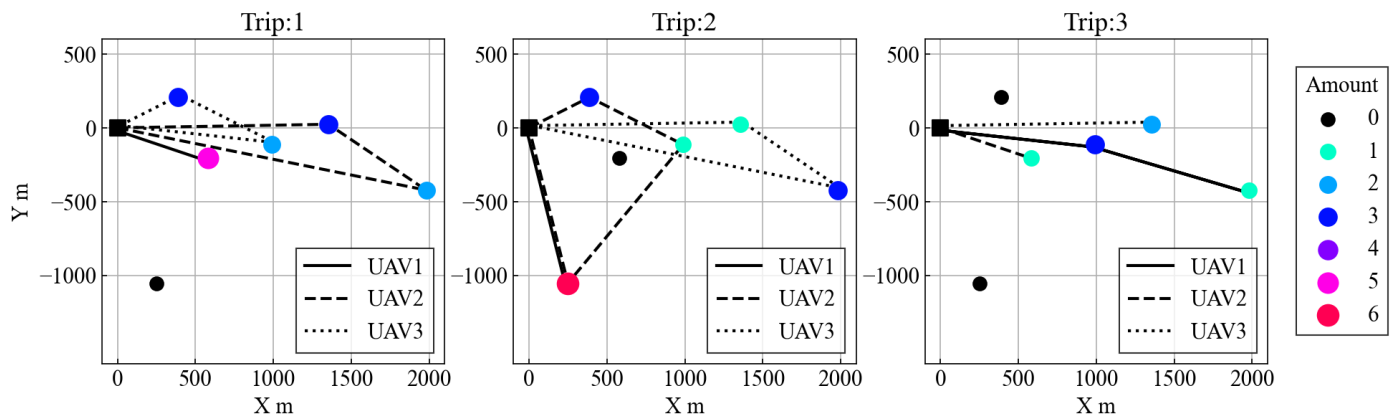


Figure 14. Route and resource allocation to minimize the objective value of Case 4 in Scenario 1. The lines represent the route of each UAV, and the color and size of the dots represent the amount of supplies supplied to each shelter.

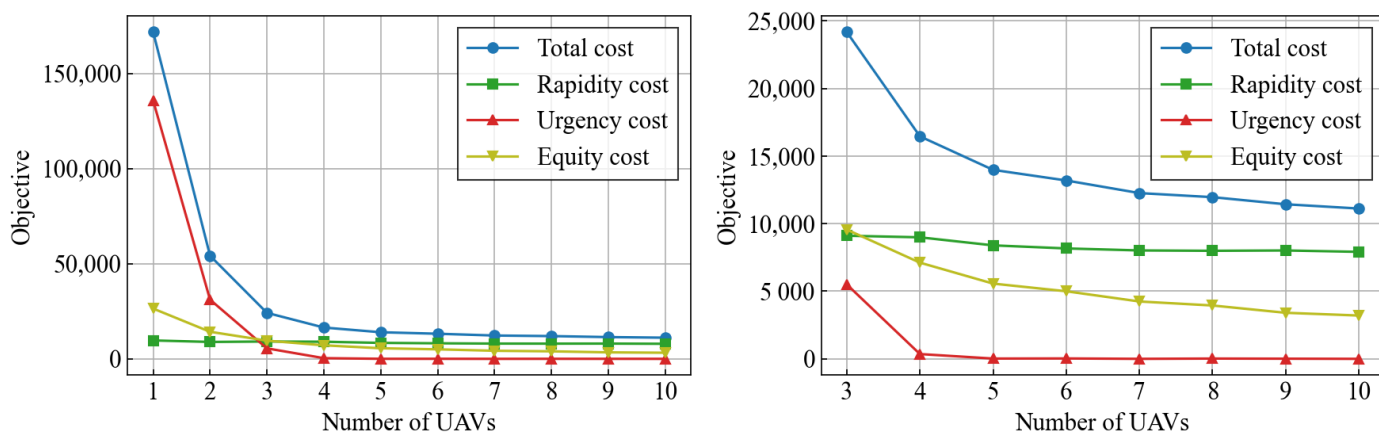


Figure 15. Minimum objective values for each number of UAVs in Scenario 1. (left) N_U is from 1 to 10. (right) N_U is from 3 to 10.

6. Conclusions and Future Work

Supplying isolated shelters with sufficient supplies quickly and adequately during a disaster saves many lives and provides victims with a sense of security. Soon after a disaster occurs, land transportation may be difficult due to damaged roads and traffic congestion, and UAVs offer a promising potential solution to this problem. However, the limited payload of a UAV is insufficient for a single transport to meet the various material demands of shelters, so considering the sequence of multi-UAV destinations and the types of materials to be transported is an important issue in a crisis situation.

In this study, we developed a model for transporting emergency relief supplies by multiple UAVs using QL. To evaluate the performance of QL in the UAVRPs, we compared the transportation routes of the previous model. As a result, we were able to develop route planning with a shorter transportation route than the conventional method, which confirms the performance of our method in improving the efficiency of transportation distance. In addition, we confirmed that it is possible to transport high-priority items quickly and to eliminate the disparity in supply among shelters. We were able to quantitatively evaluate the transportation time and the percentage of high-priority items that could be transported within a time limit for each of the disaster response measures. This result may be used to reduce the deviation of the supply to each shelter when the demand for supplies is so large that the transportation of supplies cannot be completed within a certain time limit. In particular, in a disaster such as a tsunami-related disaster, where damage occurs over a wide area, a large number of people may be displaced and the demand for shelters may increase massively at the same time. Since the equity-oriented transportation in this study can prevent disparities in supplies at each shelter, it can be applied as a transportation strategy for supplies, such as medicines and blood, that require small quantities but need to be supplied quickly.

For the practical application of this study, it is necessary to grasp the needs of each evacuation center for supplies in advance. This could be solved by utilizing UAVs equipped with communication capabilities as emergency communication networks [36] and by forecasting the demand for supplies [51]. With this information, effective last-mile delivery may be achieved by utilizing this study’s method, taking into account the number of UAVs available, payload limitations, and battery limitations.

There is room for improvement in the model. In HL such as disaster response, transportation must be planned in as little time as possible due to the uncertainty of demand for supplies. However, QL, like other heuristic methods, requires time for calculation because it must determine various parameters and then perform optimization. It is necessary to extend the UAV transportation planning problem to consider highly uncertain situations, such as the urgency of demand and supplies and the number of UAVs available.

Planning the transportation of items by multiple UAVs using the model of this study will enable the stable and rapid transportation of items to isolated disaster areas such as tsunami-flooded areas in the event of an actual disaster. The ultimate goal of this research is to develop a multiple UAV planning tool to optimize the allocation of UAVs and support decision making for disaster relief and supply transportation.

Author Contributions: Conceptualization and Overall Research Design, S.K.; methodology, D.H. and E.M.; validation D.H.; visualization D.H. and E.M.; writing—original draft, D.H. and E.M.; writing—review and editing, D.H., E.M. and S.K.; supervision, E.M. and S.K.; funding acquisition, E.M. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partly funded by the Japan Society for the Promotion of Science (JSPS) Kakenhi Program (21H05001); JST Japan-US Collaborative Research Program (JPMJSC2119); Co-creation Center for Disaster Resilience, Tohoku University; the Core Research Cluster of Disaster Science at Tohoku University (Designated National University); and Tough Cyberphysical AI Research Center, Tohoku University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Balcik, B.; Beamon, B.; Smilowitz, K. Last mile distribution in humanitarian relief. *J. Intell. Transp. Syst.* **2008**, *12*, 51–63. [CrossRef]
2. Dubey, R.; Luo, Z.; Gunasekaran, A.; Akter, S.; Hazen, B.T.; Douglas, M.A. Big data and predictive analytics in humanitarian supply chains. *Int. J. Logist. Manag.* **2018**, *29*, 485–512. [CrossRef]
3. Suzuki, Y. Impact of material convergence on last-mile distribution in humanitarian logistics. *Int. J. Prod. Econ.* **2020**, *223*, 107515. [CrossRef]
4. Jaller, M. *Resource Allocation Problems during Disasters: The Cases of Points of Distribution Planning and Material Convergence Handling*; Rensselaer Polytechnic Institute: Troy, NY, USA, 2011.
5. Das, R. Disaster preparedness for better response: Logistics perspectives. *Int. J. Disaster Risk Reduct.* **2018**, *31*, 153–159. [CrossRef]
6. Holgun-Veras, J.; Taniguchi, E.; Jaller, M.; Aros-Vera, F.; Ferreira, F.; Thompson, R.G. The Tohoku disasters: Chief lessons concerning the post disaster humanitarian logistics response and policy implications. *Transp. Res. Part A Policy Pract.* **2014**, *69*, 86–104. [CrossRef]
7. Shibata, Y.; Uchida, N.; Shiratori, N. Analysis of and proposal for a disaster information network from experience of the Great East Japan Earthquake. *IEEE Commun. Mag.* **2014**, *52*, 44–50. [CrossRef]
8. Sato, T.; Suzuki, K. Impact of Transportation Network Disruptions caused by the Great East Japan Earthquake on Distribution of Goods and Regional Economy. *J. JSCE* **2013**, *1*, 507–515. [CrossRef]
9. Koshimura, S.; Hayashi, S.; Gokon, H. The impact of the 2011 Tohoku earthquake tsunami disaster and implications to the reconstruction. *Soils. Found.* **2014**, *54*, 560–572. [CrossRef]
10. Nakachi, H.; Maki, N.; Hayashi, H.; Kobayashi, K. A Proposal of the Effective System to Utilize Helicopters During the Giant Earthquake Disaster of the Nankai Trough Based on the Study of the Great East Japan Earthquake. *J. JSNDS* **2014**, *33*, 101–114. (In Japanese)
11. Kellermann, R.; Fisher, L. Drones for parcel and passenger transportation: A literature review. *Transport. Res. Interdiscip. Persp.* **2020**, *4*, 100088. [CrossRef]
12. Al-Turjman, F.; Alturjman, S. 5G/IoT-enabled UAVs for multimedia delivery in industry-oriented applications. *Multimed. Tools. Appl.* **2020**, *79*, 8627–8648. [CrossRef]
13. Cheng, C.; Adulyasak, Y.; Rousseau, L.M. Drone routing with energy function: Formulation and exact algorithm. *Transp. Res. Part B Methodol.* **2020**, *139*, 364–387. [CrossRef]
14. Yakushiji, K.; Fujita, H.; Murata, M.; Hiroi, N.; Hamabe, Y.; Yakushiji, F. Short-Range Transportation Using Unmanned Aerial Vehicles (UAVs) during Disasters in Japan. *Drones* **2020**, *4*, 68. [CrossRef]
15. Magdalena, P.; Lora, K. Zipline's New Drone Can Deliver Medical Supplies at 79 Miles per Hour. CNBS. Available online: <https://www.cnbc.com/2018/04/02/zipline-new-zip-2-drone-delivers-supplies-at-79-mph.html> (accessed on 14 September 2022).
16. Thibbotuwawa, A.; Bocewicz, G.; Nielsen, P.; Banaszak, Z. Unmanned Aerial Vehicle Routing Problems: A Literature Review. *Appl. Sci.* **2020**, *10*, 4504. [CrossRef]
17. Dorling, K.; Heinrichs, J.; Messier, G.G.; Magierowski, S. Vehicle Routing Problems for Drone Delivery. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *47*, 70–85. [CrossRef]

18. Rabta, B.; Wankmüller, C.; Reiner, G. A drone fleet model for last-mile distribution in disaster relief operations. *Int. J. Disaster Risk Reduct.* **2018**, *28*, 107–112. [[CrossRef](#)]
19. Shi, Y.; Lin, Y.; Li, B.; Yi Man Li, R. A bi-objective optimization model for the medical supplies' simultaneous pickup and delivery with drones. *Comput. Ind. Eng.* **2022**, *171*, 108389. [[CrossRef](#)]
20. Ghelichi, Z.; Gentili, M.; Mirchandani, P.B. Logistics for a fleet of drones for medical item delivery: A case study for Louisville, KY. *Comput. Oper. Res.* **2021**, *135*, 105443. [[CrossRef](#)]
21. Jiang, X.; Zhou, Q.; Ye, Y. Method of Task Assignment for UAV Based on Particle Swarm Optimization in logistics. In Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Hong Kong, China, 25–27 March 2017; pp. 113–117.
22. Song, B.D. Persistent UAV delivery logistics: MILP formulation and efficient heuristic. *Comput. Ind. Eng.* **2018**, *120*, 418–428. [[CrossRef](#)]
23. Li, Y.; Yuan, X.; Zhu, J.; Huang, H.; Wu, M. Multiobjective Scheduling of Logistics UAVs Based on Variable Neighborhood Search. *Appl. Sci.* **2020**, *10*, 3575. [[CrossRef](#)]
24. Gentili, M.; Mirchandani, P.B.; Agnetis, A.; Ghelichi, Z. Locating Platforms and Scheduling a Fleet of Drones for Emergency Delivery of Perishable Items. *Comput. Ind. Eng.* **2022**, *168*, 108057. [[CrossRef](#)]
25. Chowdhury, S.; Emelogu, A.; Marufuzzaman, M.; Nurre, S.G.; Bian, L. Drones for disaster response and relief operations: A continuous approximation model. *Int. J. Prod. Econ.* **2017**, *188*, 167–184. [[CrossRef](#)]
26. Chen, J.; Du, C.; Zhang, Y.; Han, P.; Wei, W. A Clustering-Based Coverage Path Planning Method for Autonomous Heterogeneous UAVs. *IEEE Trans. Intell. Transp. Syst.* **2021**, 1–11. [[CrossRef](#)]
27. Chen, J.; Ling, F.; Zhang, Y.; You, T.; Liu, Y.; Du, X. Coverage path planning of heterogeneous unmanned aerial vehicles based on ant colony system. *Swarm Evol. Comput.* **2022**, *69*, 101005. [[CrossRef](#)]
28. Beamon, B.M.; Balcik, B. Performance measurement in humanitarian relief chains. *Int. J. Public Sect. Manag.* **2008**, *21*, 4–25. [[CrossRef](#)]
29. Huang, M.; Smilowitz, K.; Balcik, B. Models for relief routing: Equity, efficiency and efficacy. *Transp. Res. Part E Logist. Transp. Rev.* **2012**, *48*, 2–18. [[CrossRef](#)]
30. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [[CrossRef](#)]
31. Sakurai, M.; Murayama, Y. Information technologies and disaster management-Benefits and issues-. *Prog. Disaster Sci.* **2019**, *2*, 100012. [[CrossRef](#)]
32. Yu, M.; Yang, C.; Li, Y. Big Data in Natural Disaster Management: A Review. *Geosciences* **2018**, *8*, 165. [[CrossRef](#)]
33. Zacharie, M.; Fuji, S.; Minori, S. Rapid Human Body Detection in Disaster Sites Using Image Processing from Unmanned Aerial Vehicle (UAV) Cameras. In Proceedings of the 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Bangkok, Thailand, 21–24 October 2018.
34. Nagasawa, R.; Mas, E.; Moya, L.; Koshimura, S. Model-based analysis of multi-UAV path planning for surveying postdisaster building damage. *Sci. Rep.* **2021**, *11*, 18588. [[CrossRef](#)]
35. Alhindi, A.; Alyami, D.; Alsubki, A.; Almousa, R.; Al Nabhan, N.; Al Islam, A.B.M.A.; Kurdi, H. Emergency Planning for UAV-Controlled Crowd Evacuations. *Appl. Sci.* **2021**, *11*, 9009. [[CrossRef](#)]
36. Klaine, P.V.; Nadas, J.P.B.; Souza, R.D.; Imran, M.A. Distributed Drone Base Station Positioning for Emergency Cellular Networks Using Reinforcement Learning. *Cogn. Comput.* **2018**, *10*, 790–804. [[CrossRef](#)] [[PubMed](#)]
37. Chowdhury, S.; Shahvari, O.; Marufuzzaman, M.; Li, X.; Bian, L. Drone routing and optimization for post-disaster inspection. *Comput. Ind. Eng.* **2021**, *159*, 107495. [[CrossRef](#)]
38. Tiwari, A.; Dixit, A. Unmanned aerial vehicle and geospatial technology pushing the limits of development. *Am. J. Eng. Res.* **2015**, *4*, 16–21.
39. Murray, C.C.; Chu, A.G. The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. *Trans. Res. Emerg. Technol.* **2015**, *54*, 86–109. [[CrossRef](#)]
40. Jeong, H.Y.; Song, B.D.; Lee, S. Truck-drone hybrid delivery routing: Payload-energy dependency and No-Fly zones. *Int. J. Prod. Econ.* **2019**, *214*, 220–233. [[CrossRef](#)]
41. Kuo, R.; Lu, S.-H.; Lai, P.-Y.; Mara, S.T.W. Vehicle routing problem with drones considering time windows. *Expert Syst. Appl.* **2022**, *191*, 116264. [[CrossRef](#)]
42. Gu, Q.; Fan, T.; Pan, F.; Zhang, C. A vehicle-UAV operation scheme for instant delivery. *Comput. Ind. Eng.* **2020**, *149*, 106809. [[CrossRef](#)]
43. Braekers, K.; Ramaekers, K.; Van Nieuwenhuysse, I. The vehicle routing problem: State of the art classification and review. *Comput. Ind. Eng.* **2016**, *99*, 300–313. [[CrossRef](#)]
44. Gutjahr, W.J.; Fischer, S. Equity and deprivation costs in humanitarian logistics. *Eur. J. Oper. Res.* **2018**, *270*, 185–197. [[CrossRef](#)]
45. Jiang, Z.B.; Gu, J.J.; Fan, W.; Liu, W.; Zhu, B.Q. Q-learning approach to coordinated optimization of passenger inflow control with train skip-stopping on an urban rail transit line. *Comput. Ind. Eng.* **2019**, *127*, 1131–1142. [[CrossRef](#)]
46. Yu, L.; Zhang, C.; Jiang, J.; Yang, H.; Shang, H. Reinforcement learning approach for resource allocation in humanitarian logistics. *Expert Syst. Appl.* **2021**, *173*, 114663. [[CrossRef](#)]
47. Ministry of Land, Infrastructure, Transport and Tourism of Japan. National Land Numerical Information. Available online: <https://nlftp.mlit.go.jp/ksj/index.html> (accessed on 8 June 2022).

48. Sheu, J.B. An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transport. Res. E-Log.* **2007**, *43*, 687–709. [[CrossRef](#)]
49. Lin, Y.; Batta, R.; Rogerson, P.; Blatt, A.; Flanigan, M. A logistics model for emergency supply of critical items in the aftermath of a disaster. *Socio-Econ. Plann. Sci.* **2011**, *45*, 132–145. [[CrossRef](#)]
50. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*, 2nd ed.; The MIT Press: Cambridge, MA, USA, 2018; pp. 25–42.
51. Fang, J.; Hou, H.; Bi, Z.M.; Jin, D.; Han, L.; Yang, J.; Dai, S. Data fusion in forecasting medical demands based on spectrum of post-earthquake diseases. *J. Ind. Inf. Integr.* **2021**, *24*, 100235. [[CrossRef](#)]