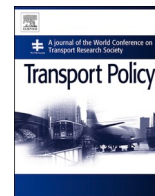




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Dynamic governance decisions on multi-modal inter-city travel during a large-scale epidemic spreading

Yunlong An^a, Xi Lin^a, Meng Li^{a,*}, Fang He^b

^a Department of Civil Engineering, Tsinghua University, Beijing, 100084, People's Republic of China

^b Department of Industrial Engineering, Tsinghua University, Beijing, 100084, People's Republic of China

ARTICLE INFO

Keywords:

COVID-19
Travel ban
Multi-modal transportation
Reinforcement learning

ABSTRACT

Within half a year, COVID-19 spreads to most countries in the world, as well as posed a great threat to the public health of human beings. The implementation of non-pharmaceutical intervention (NPI), including travel ban, proved to be an effective way for controlling the epidemic spreading, e.g., the ban of inter-city transportation stops transporting virus through passengers between cities. However, travel ban could significantly impact many industries, e.g. tourism and logistics, thus jeopardizing the regional economy. This paper focus on assisting the national or regional government to make dynamic decisions on restricting and recovering intercity multi-modal travel services. Our model can characterize impacts of inter-city traffic on the spread of the COVID-19, as well as on the regional economy. By applying a reinforcement learning approach, we develop an online optimization model to identify the modal-specific travel banning strategy that can balance the epidemic control as well as the negative impacts on regional economy. The numerical study based on a network of multiple cities in China shows that the proposed approach can generate better strategies compared with some existing methods.

1. Introduction

In 2020, the outbreak of COVID-19 became a disaster for all human. The outbreak was first detected in December 2019. Soon, the WHO began to declare the outbreak as a Public Health Emergency of International Concern (PHEIC) on January 30. By the end of July, nearly 15 million people were infected and over 500 thousand people died worldwide. More than 4 million people have been infected in the United States and 150 thousand have died. In the early stage, the Non-pharmacological interventions (NPI), including travel ban, social distancing, school closure and emergency response, seem to be the only effective way to prevent and slow down the pandemic of diseases when the vaccine is under research. For instance, public transportation increases the risk of person-to-person contact and aggravates the spread of the epidemic across cities or countries. Travel restrictions are helpful to control the risk of epidemic spreading. Wuhan adopted the travel ban on Jan.23th right after the outbreak before the Chinese New Year. Wuhan restricted the intercity traffic and closed the airports, railway stations and the highway. The travel ban had been effective until April 8, two weeks after the declaration of indigenous pandemic interruption. Many studies suggested that Wuhan's travel restrictions had a positive effect

on controlling the spread of COVID-19. Some studies indicated that the travel quarantine of Wuhan delayed the overall epidemic progression in mainland China (Chinazzi et al., 2020; Tian et al., 2020). Leung et al. (2020) found the reproduction number, denoted by R_t , decreased substantially in all selected cities and provinces since control measures were implemented, and have since remained below 1, which indicates the epidemic has been controlled. With travel restriction (no imported exposed individuals to Beijing), the number of infected individuals in seven days will decrease by 91.14% in Beijing, compared with the scenario of no travel restriction (Tang et al., 2020). With the pandemic of COVID-19 around the world, many other countries also adopted the travel ban or traffic restriction to protect people from COVID-19. For example, Australians have followed the urgings of National Cabinet and all levels of government, to limit travel and social contact, which has thus far resulted in a "flattening of the curve" (Beck and Hensher, 2020). At the same time, almost all the airline companies reduced flights, and a larger reduction in air travel through airports in a large part of the cumulative incidence area would lead to a gradual decrease in the risk flow (Nakamura and Managi, 2020). In addition to governance decisions, individual actions are also important. But individuals do not internalize the external cost of infection risks they impose on others and the health

* Corresponding author.

E-mail address: mengli@tsinghua.edu.cn (M. Li).

<https://doi.org/10.1016/j.tranpol.2021.01.008>

Received 28 December 2020; Accepted 24 January 2021

Available online 15 February 2021

0967-070X/© 2021 Elsevier Ltd. All rights reserved.

care system when making their own travel (social-activity) decisions; In order to induce individual travel decision-makers to internalize this external cost, the government actions are necessary (Oum and Wang, 2020).

However, with the achievement of epidemic control, the whole society is paying a big price for various NPIs. The complex dynamic social-economic system needs mobility, and a long time travel ban has direct and indirect impacts on many businesses, such as tourism, restaurants and cinemas, leading to more unemployment. The evidence is given by some latest studies. For example, Guan et al. (2020) shows that COVID-19 control measures affected the global supply-chain and caused losses and the complexity of global supply chains magnified losses beyond the direct effects of COVID-19. Also, other studies show differences in strictness of such measures and the rapidity with which governments have imposed and relaxed the measures have a divergent impact on public health and economic losses (Wells et al., 2020; Inoue and Todo, 2019). In fact, many cities and organizations around the world are calling for hope to resume transportation and production as soon as possible. It is a critical decision for government to figure out when to recovery traffic to restart the intercity mobility. Therefore, it is important to understand the dynamics of the spreading of epidemic, in particular to estimate the risk of transmission over time, and analyze the effect of the control policies. This paper focuses on the evaluation of travel ban policies with different levels of strictness considering the balance between epidemic control and economic development. A network-based epidemic spreading model is established, in which each node represents a city. The link between nodes represents the connection between cities, and different travel modes as well as mode choice process are considered. The paper proposes a reinforcement-learning (RL)-based multi-modal decision-making framework under uncertainty. The RL-based framework tends to find a policy that helps to make a balance between public health risk prevention and economic development. The paper also evaluates various potential epidemic control strategies from the public health intervention side, and make a comparison between RL-based approach and some traditional control policies.

The rest of the paper is organized as follows. Section 2 elaborates on the framework of the methodology, the network formulation, and the epidemic transmission model. Section 3 introduces the RL-based decision-making framework and its execution procedure. Section 4 presents a series of case studies, including three virtual city networks with different sizes, and a real city network containing 15 cities, and the experiments illustrate epidemic spreading across the cities and the effect of different travel ban policies. In Section 5, we conclude the paper and offer some policy implications.

2. Methodology

2.1. City network modeling

2.1.1. Network setting

We adopt a directed graph $\mathcal{G} \equiv (\mathcal{N}, \mathcal{E})$ to represent the network of cities, in which \mathcal{N} is the finite set of nodes and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N} \times \mathcal{M}$ is the set of directed arcs, where \mathcal{M} stands for the travel mode. For a pair of nodes n and n' , we have $(n, n', m) \in \mathcal{E}$ if and only if there is a link connecting these two nodes such that the people can move directly from city n to city n' by traffic mode m . For each node $n \in \mathcal{N}$, we can find a set of its successor nodes $\varphi(n)$, i.e. $\varphi(n) \equiv n' : (n, n', m) \in \mathcal{E}$.

2.1.2. Node representation

Each node stands for a city, and every node has a label, L_n , which owns 3 attributes. $L_n = (O_n, \mathcal{X}_n, \mathcal{P}_n)$. O_n represents the location of city n . \mathcal{X}_n is the compartment set, i.e., $\mathcal{X}_n = \{S_n, E_n, I_n, H_n, R_n, D_n\}$, and the sum of elements in \mathcal{X}_n equals to the population of city n . \mathcal{P}_n are the set of parameters in epidemic model, i.e., $\mathcal{P}_n = \{\beta_n, \sigma_n, \gamma_n\}$, and the

parameters will be explained in section 2.2.

2.1.3. Link representation

The link l in network stands for a kind of travel mode from one city to another. Each link has 4 attributes, i.e., $l_{(n,n',m)} = (d_l, t_l, c_l, C_l)$. d_l is the distance between city n and city n' . t_l, c_l, C_l are the travel time, travel cost and capacity from city n to city n' by mode m , respectively.

2.2. Epidemic modeling

The modeling of epidemic spreading is a tool that has been used to study the mechanisms by which diseases spread, to predict the future course of an outbreak, and to evaluate strategies to control an epidemic. The most popular epidemic model is the compartment model (Kermack and McKendrick, 1927), in which the population is assigned to different compartments with labels. People can transfer between compartments. The order of the labels usually shows the flow patterns between the compartments. The setting of labels is different for different diseases, common labels include $S, I,$ and R (Susceptible, Infectious, Recovered). In the study of COVID-19, the susceptible-exposed-infectious-recovered (SEIR) model (Aron and Schwartz, 1984), shown in Fig. 1, is one of the most adopted methods. The population is divided into four different compartments, susceptible (denoted by S , the people who are able to contract the disease), exposed (E , the people who have been infected by the disease but not yet infectious), infectious (I , the people who are capable of transmitting the disease), and removed (R , the people who have recovered with immunization or died). And the people progress between those four compartments. SEIR model and its improved version have widely used in COVID-19 related research (Kucharski et al., 2020; Wu et al., 2020; Chen et al., 2020). In this paper, an improved SEIR model is coupled into the network model as one integrated part. When considering the travel of infected people between cities, we also consider the risk of people being infected in the city and on transportation. However, the infection characteristics in cities and transportation are still slightly different, and we adopt a distinction in the model.

To be more in line with the actual situation, we improve the base model and develop an extended SEIHRD model, since in practice we can only find infected persons by nucleic acid testing (NAT), and there are always some patients who are not effectively identified. The new compartment H represents patients who have been admitted to the hospital. And compartment I represents the unreported infected person. Compartment R stands for the recovered person and compartment D stands for the dead. We distinguish between compartment I and compartment H , because those who have been found infected will be sent to the hospital, and they will not move to other cities, let alone infect others. Undetected infections, including hidden infections and undetected overt infections, will become the source of infection. At the same time, the cure rate and death rate are related to whether the infected person is hospitalized or not.

Fig. 2 shows how people progress each compartment in SEIHRD model. The rate of spreading is associated with the probability of transmitting disease between the susceptible people (S) and the exposed people (E), which is controlled by the infectious rate β . Besides, the exposed people are also infectious in case of COVID-19, so we use β_E and β_I to represent the infectious rate of exposed people and infected people,

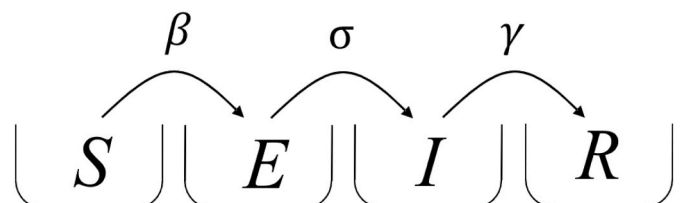


Fig. 1. Basic SEIR model diagram.

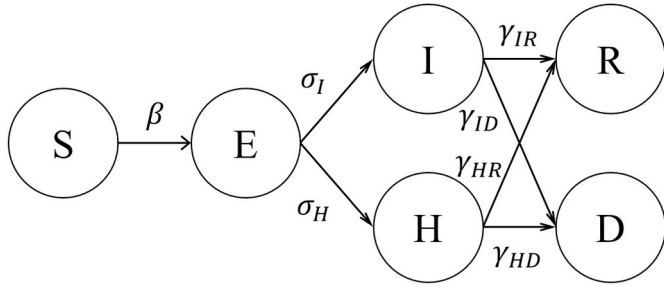


Fig. 2. SEIHRD model diagram.

respectively. The rate of exposed people (E) becoming infected person (I) and hospitalized patients (H) is defined as the incubation rate, σ_I and σ_H . At last, the cure rate and mortality rate of untreated patients and hospitalized patients are different, denoted by $\gamma_{ID}, \gamma_{IR}, \gamma_{HD}, \gamma_{HR}$. Because the period for the pandemic of COVID-19 control we focus on is only several months, the effect of birth rate or death rate on the population is ignored, and the change of population in each city is only affected by the travel between cities. $M_{ij,m}^s$ represent the susceptible people who move from city i to city j by mode m . Eqs. (1)–(6) mathematically depict the entire progress dynamics.

$$\frac{dS_i}{dt} = -\frac{\beta_E S_i E_i}{N_i} - \frac{\beta_I S_i I_i}{N_i} - \sum_j \sum_m M_{ij,m}^s + \sum_k \sum_m M_{k,i,m}^s \quad (1)$$

$$\frac{dE_i}{dt} = \frac{\beta_E S_i E_i}{N_i} + \frac{\beta_I S_i I_i}{N_i} - (\sigma_I + \sigma_H) E_i - \sum_j \sum_m M_{ij,m}^E + \sum_k \sum_m M_{k,i,m}^E \quad (2)$$

$$\frac{dI_i}{dt} = \sigma_I E_i - (\gamma_{IR} + \gamma_{ID}) I_i - \sum_j \sum_m M_{ij,m}^I + \sum_k \sum_m M_{k,i,m}^I \quad (3)$$

$$\frac{dH_i}{dt} = \sigma_H E_i - (\gamma_{HR} + \gamma_{HD}) H_i \quad (4)$$

$$\frac{dR_i}{dt} = \gamma_{IR} I_i + \gamma_{HR} H_i \quad (5)$$

$$\frac{dD_i}{dt} = \gamma_{ID} I_i + \gamma_{HD} H_i \quad (6)$$

where $N = S + E + I + H + R + D$.

The spreading of COVID-19 not only occurs in cities. Some evidence shows that COVID-19 is also very easy to spread within public transportation, and the risk of spreading is related to the capacity and the crowdedness. As a result, the SEIHRD diagram we apply in our city network can be divided into two categories, node-based SEIHRD, and link-based SEIHRD. The node-based model is adopted to describe the epidemic situation inside the city, while the link-based model is used to describe the risk of the epidemic spreading in public transportation, like the airplane, HST, coach and private car. Although there is no hospital on buses, considering that the infected person with severe symptoms will be sent for treatment after getting off the buses, we also consider the H part.

2.2.1. Time-varying infection rate

To control the pandemic of epidemic, the government implemented a series of public health intervention policies, i.e. emergency response and social distancing, which is an important factor affecting the spreading of epidemic. However, the strength of these measures is difficult to be accurately quantified. We used a truncation function to describe the change of infectious rate β . t_{free} is the time for the government to initiate the first-level response. Before taking any prevention policies, β keeps as a constant. After that, β decays exponentially with time, approaching zero. This setting is in accordance with some relevant studies on SARS

(Ni, 2009). With the assumption of the monotonically decreasing infection rate, one outbreak or independent outbreaks of the pandemic is considered in the epidemic model. For the scenario with independent outbreaks, we can apply the model at the beginning of each outbreak.

$$\beta_t = \begin{cases} \beta, & \text{if } t \leq t_{free} \\ \beta \exp(-\xi(t - t_{free})), & \text{if } t > t_{free} \end{cases} \quad (7)$$

where β is the initial infectious rate, ξ are the parameters controlling the exponential decay speed.

2.3. Travel demand modeling

For the sake of simplification, we presently assume that the travel demand between cities follows the gravity model. The intensity of travel demand between the two cities is proportional to the population of the two cities and inversely proportional to the square of the distance. Also, we have a random number η_t^{ij} , which follows a normal distribution, to model the randomness. A city will have a certain percentage of people leaving the city, which is a random number. The distribution of travel demand is related to the population of destinations. We denote W_t^{ij} as the travel demand from city i to city j in t day, and it is calculated by Eq. (8):

$$W_t^{ij} = G \eta_t^{ij} \frac{Z_i Z_j}{d^2(i,j)} \quad (8)$$

where Z_i, Z_j is the population of city i and city j , respectively. $d(i,j)$ is the distance between city i and city j . G is a scaling parameter.

2.4. Travel ban action

As one of the primary method to control the epidemic pandemic, travel ban is adopted by many governments. Define binary variables a_t^l such that $a_t^l = 1$ if link l under mode m is open at time t , and $a_t^l = 0$ otherwise. We set two scenarios to perform travel ban.

2.4.1. Scenario I

In scenario I, the government tends to perform the policies when they are needed. In other words, the government will decide both when to perform travel ban policies and when to cancel the policies as well as recover the traffic. At the same time, the opening and closing of the travel ban can only be performed once in the entire period. All links are assumed to be open at the beginning. Eqs. 9–11 shows the travel ban under Scenario I.

$$\sum_{t=1}^T (a_t^l - a_{t-1}^l) = 0 \quad (9)$$

$$\sum_{t=1}^T |a_t^l - a_{t-1}^l| = 2 \quad (10)$$

$$a_t^l \in \{0, 1\} \quad \forall t \in \mathcal{T} \quad (11)$$

2.4.2. Scenario II

In some cases, the outbreak of COVID-19 is a very urgent event, and the government should take all the available policies as soon as possible. In Scenario II, we assume all the travel is banned as soon as possible, and the government only needs to decide whether to cancel the policies and recover travel. Eqs.12–14 shows the travel ban under Scenario II.

$$\sum_{t=1}^T (a_t^l - a_{t-1}^l) = 1 \quad (12)$$

$$\sum_{t=1}^T |a_t^l - a_{t-1}^l| = 1 \quad (13)$$

$$a_t^l \in \{0, 1\} \quad \forall t \in \mathcal{T} \quad (14)$$

2.5. Mode choice and travel demand transfer

We consider multi-modal transportation between cities. Because of the travel ban, the compensation phenomenon between transportation modes is considered in our model. If a transportation mode stops operating, the traffic originally allocated to that mode will be transferred to other modes. Taking high-speed train (HST) and aircraft as an example, if the aircraft is banned, a large part of passengers will change their choice to travel on HST. The flow between two cities is associated with a vector f_t^l to characterize the distribution in different modes, where $f_t^l = \{f_t^m : m \in \mathcal{M}\}$.

Furthermore, each mode of transportation has the capacity limit. If the passengers who choose certain transportation exceed the capacity limit, their needs will not be met. We adopt an improved multinomial logit model (MNL) to describe the choice of travelers. In the traditional MNL model, travelers are characterized by their utility U_t^l . Assuming that U_t^l has a deterministic part V_t^l in terms of the travel cost c_t^l and the travel time t_t^l , and a random error ε_t distributed with a Gumbel distribution, it is possible to know that the mode choice probability follows a multinomial logit such that:

$$U_t^l = V_t^l + \varepsilon_t^l \quad (15)$$

$$V_t^l = c_t^l + \rho t_t^l \quad (16)$$

where ρ are the travelers' average time value, where we use social average hourly income. Then the mode $m \in \mathcal{M}$ choice probability, p_t^l , satisfy:

$$p_t^l = \frac{\exp\{V_t^l\}}{\sum_{m \in \mathcal{M}} \exp\{V_t^m\}} \quad (17)$$

Denote f_t^l as the mode specify flow. Considering the capacity constrain, that is, $f_t^l \leq C^l$, we propose a multiple traffic distribution method. It can also be treated as a dynamic ticketing process. Every passenger chooses a mode with the probability p_t^l . If the number of passengers choosing a certain mode reaches its capacity, the extra passengers can only choose from other modes that have not reached the capacity. At last, all travelers have successfully chosen a mode, or all modes have no remaining capacity. The mode choice is shown in **Algorithm 1**.

Algorithm 1. Travel mode choice and demand transfer

Algorithm 1: Travel mode choice and demand transfer

Input : The travel demand set \mathcal{J} , the capacity set \mathcal{C} , the action set A_t , and link labels;

Output: The flow distribution result \mathcal{F}_t ;

```

1 for origin city  $i = 1, \dots, n$  do
2   for destination city  $j = 1, \dots, n$  do
3     if there is a link  $l$  between city  $i$  and city  $j$  then
4       for passenger  $1, \dots, J_t^{i,j}$  do
5         Check available travel mode with  $\mathcal{M}' \equiv \{m | C^l a_t^m > 0 \quad \forall m \in \mathcal{M}\}$ ;
6         if  $\mathcal{M}' \neq \emptyset$  then
7           Calculate choice probability  $p_t^l$  for  $m \in \mathcal{M}'$ ;
8           Randomly choose a mode with the probability distribution;
9         end
10      end
11    end
12  end
13 end
```

2.6. Economic growth impact modeling

The implementation of the Non-pharmaceutical Intervention (NPI) has a side effect on business and the economic system. The tool should be of interest to policymakers and others who wish to use it to understand the potential impacts of the travel ban on health and economic outcomes in their communities. Then they weigh the trade-offs between various policies and decide when and how these interventions can be relaxed. Travel ban helps to prevent the spreading of epidemic, but it also hinders economic development.

Our society and economic system are so complex that it is hard to directly calculate the system loss. So we use some approximate methods. Mobility, represented by traffic turnover, brings about the circulation of regional businesses and personnel and promotes consumption and industry development. Some evidence from existing studies also suggest this, for example, [ShunLi et al. \(2005\)](#) showed that GDP and traffic turnover volume in China have a granger co-integration relationship, and [Donzelli \(2010\)](#) found that the spreading of the traffic demand increases the rate of international tourism, generates new jobs and improves the income of the area. Taking the direct impact of mobility on regional consumption into account, we take Total Retail Sales of Consumer Goods (TRSCG), denoted by E , as an economic indicator. At the same time, we adopt traffic turnover volume, denoted by $g_t g_t^l = f_t^l d_t$, as the traffic indicator. We assume the economic impact is a convex function of traffic turnover volume. Considering a series of other economic factors, we construct an econometric model and analyze the impact of variables by regression. The economic growth contributed by traffic turnover volume can be calculated through the studying of marginal effect. The approximate economic growth function is shown in Eq. (18).

$$E_t = F(X_g, X_1, \dots, X_n) \quad (18)$$

where X_g is the traffic turnover volume, X_1, \dots, X_n are other economic factors. And the marginal economic growth contributed by traffic turnover volume, denoted by ΔE , can be calculated by Eq. (19).

$$\Delta E_t = \frac{dF(X_g, X_1, \dots, X_n)}{dX_g} \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} g_t^l \quad (19)$$

2.7. Objective function

For the policymakers, there are two goals they hope to achieve through implementing a travel ban. The first goal is to avoid the spreading of epidemic, which can be simplify treated as minimizing the number of confirmed patients, to make sure the cumulative number of infections is as few as possible. At the same time, they hope that the side effect of interventions, which can be treated as the effect on economic growth, can be as little as possible.

Hence, taking above consideration into account, the objective function can be formulated as Eq. (20):

$$\tau_H \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} H_t^n + \tau_D \sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} D_t^n - \sum_{t \in \mathcal{T}} \Delta E_t \quad (20)$$

where H_t^n is the new confirmed cases at time step t in city n . In objective function, the first two terms $\sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} H_t^n$ and $\sum_{n \in \mathcal{N}} \sum_{t \in \mathcal{T}} D_t^n$ stands for the total number of confirmed cases and deaths, respectively. The last term $\sum_{t \in \mathcal{T}} \Delta E_t$ stands for the total economic growth contributed by traffic turnover volume, which is hoped to be maximized. τ_H is the average treatment cost for a epidemic patient, and τ_D stands for death loss, which emphasizes the importance of saving lives though it is not the actual expenditure. The goal of the entire decision-making process is to minimize the above objective function.

3. Reinforcement-learning-based decision making framework

3.1. Motivation and setting

The dynamic travel ban decision problem is hard to solve directly. This is because the governors can only make policies based on previous and current epidemic data, and the actions made in any stage have an impact on the future state, which should be considered in the decision-making process. Owing to the stochasticity of traffic and epidemic pandemic, we cannot formulate the problem of determining the travel ban into a compact form. The next subsection will introduce an online framework for deciding the travel ban policy based on the reinforcement learning technique.

3.2. Online travel ban policy determination

The travel ban decision process can be treated as a Markov Decision Process (MDP). Considering the entire period for epidemic spreading, it can be divided into several days $t_1, t_2, \dots, t_n \in \mathcal{T}$, and each day is a de-

cision interval. At the start of each interval, the government decides the status of each link with different mode, close or open, and within the day, the status remains unchanged. The states in day t are defined as a vector $\mathbf{S}_t = \left(\frac{H_n^{t-1} + H_{n'}^{t-1}}{2} \mid l_{(n,n',m)} \in \mathcal{L}_E \right)$, where $\frac{H_n^{t-1} + H_{n'}^{t-1}}{2}$ stands for the average number of new hospitalized patients of the cities on both side of link $l_{(n,n',m)}$ in day $t - 1$. And we define H_n^0 as the initial number of the hospitalized patients. \mathbf{S}_t is a $|\mathcal{L}_E|$ -dimensional vector, in which $|\mathcal{L}_E|$ is the cardinality of the link set. Also, the actions in day t are defined as a vector $\mathbf{A}_t = (a_l^t \mid l \in \mathcal{L}_E)$, where $a_l^t \in \{0, 1\}$ is the status of link l in day t , and 1 means open while 0 means close. \mathcal{S} and \mathcal{A} stand for the set for \mathbf{S} , \mathbf{A} , respectively.

The policymakers decide policies based on the current number of confirmed cases. This process can be simulated as a learning process. The agent behaves in an environment according to a policy that specifies how the agent selects action at each state, in which policy π gives a map: $\mathbf{S} \rightarrow \mathbf{A}$. The goal of the agent is to find a best policy π^* to maximize the long-term expected cumulative reward at current time. To solve such a problem, a common objective is to learn an action-value function $Q_\pi(\mathbf{S}, \mathbf{A})$, where $Q_\pi(\mathbf{S}, \mathbf{A}) = \mathbb{E}_\pi \left\{ \sum_{i=0}^{+\infty} \mu^i r_{t+i} \mid \mathbf{S}_t = \mathbf{S}, \mathbf{A}_t = \mathbf{A} \right\}$, where $0 < \mu < 1$ is a discount factor. The learning process seeks to find a solution for the Bellman equation $V^*(\mathbf{S}) = \max_{\mathbf{A} \in \mathcal{A}} \left\{ \mathbb{E}_\pi \left\{ r + \mu V^*(\mathbf{S}') \right\} \right\}$, where $V(\mathbf{S}) = \max_{\mathbf{A} \in \mathcal{A}} Q(\mathbf{S}, \mathbf{A})$ and \mathbf{S}' is the next state.

The problem is defined on the network, the dimension of space for action will reach $2^{|\mathcal{L}_E|}$, which is so large that can not be solved with traditional tabular methods. We present a framework of Twin Delayed Deep Deterministic Policy Gradient (TD3) (Fujimoto et al., 2018) for online determination of travel ban policy in each day based on the above definitions. TD3 is based on Actor-Critic (AC) framework and the AC method originate from Deep Q-Learning (DQN). AC method and its improved version have been widely applied in online decision-making problems, such as taxi dispatching (Kim et al., 2020) and adaptive traffic signal control (Aslani et al., 2017). TD3 avoids overestimation and reduces function approximation error of the AC method by delaying

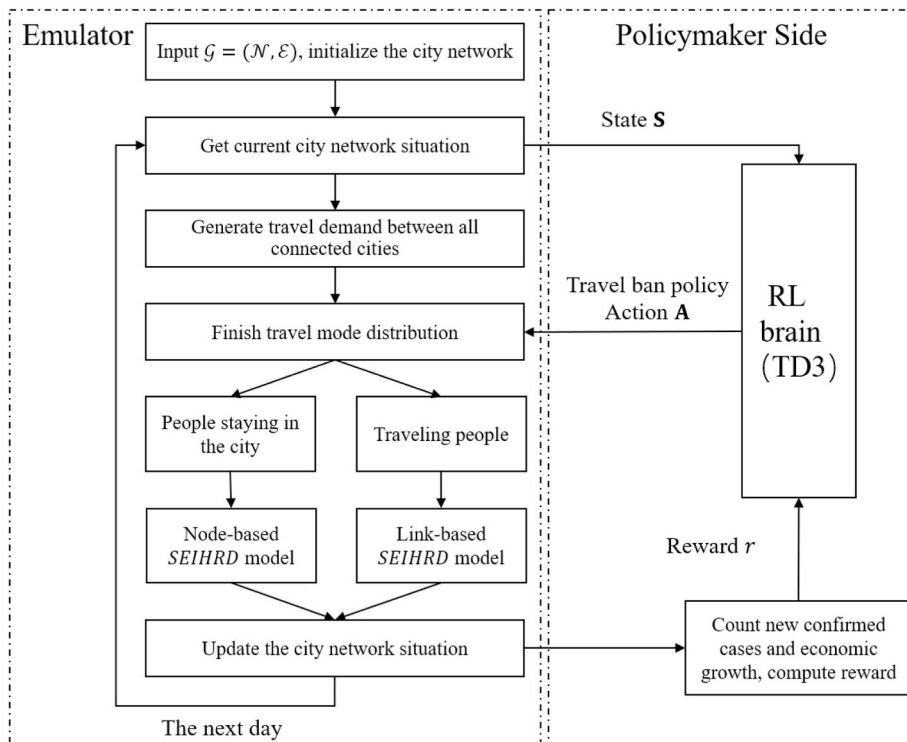


Fig. 3. The execution of RL framework.

policy updates.

In day t , we define the *reward* as the sum of the economic growth contributed by traffic turnover volume and economic loss caused by confirmed cases and deaths in this interval, which is shown in Eq. (21). The total expected discount reward return $R = \sum_{t=0}^T \mu^t r_t$. During the training process, TD3 also uses experience replay to sample previous transitions randomly, and thereby smooths the training distribution over many past behaviors.

$$r_t = \Delta E_t - \tau_H \sum_{n \in \mathcal{I}'} H_t^n - \tau_D \sum_{n \in \mathcal{I}'} D_t^n \quad (21)$$

We follow the research in Fujimoto et al. (2018) and create two critic networks $Q_{\theta_1}, Q_{\theta_2}$ as well as an actor network π_{ϕ} as well as their *target network* $Q_{\theta_1'}, Q_{\theta_2'}, \pi_{\phi'}$. Clipped double-Q learning by two critic networks will help to avoid the overestimation of Q-value, which is shown in Eq. (22). At the same time, the parameters of the *target network* and actor network $\theta_1', \theta_2', \phi_1', \phi$ will be frozen for a fixed number of iterations while updating the critic network θ_1, θ_2 by gradient descent to address variance and enhance the stability of the algorithm. In this paper, all the networks are 3-layer fully connected network, in which the hidden layer has 256 cells.

$$y = r + \mu \min_{i=1,2} Q_{\theta_i'}(\mathbf{S}', \pi_{\phi_1}(\mathbf{S}')) \quad (22)$$

The experience tuple $(\mathbf{S}, \mathbf{A}, r_t, \mathbf{S}')$ is stored in the replay memory, and our algorithm samples uniformly at random from memory when performing updates. The update of critics Q_{θ} follows Eq. (23), and that of actors follows Eq. (24). We use Adam's algorithm (Kingma and Ba, 2014) as the optimizer. The detailed procedure for our TD3 implementation is provided in Algorithm 2.

$$\theta_i \leftarrow \min_{\theta_i} K^{-1} \sum (y - Q_{\theta_i}(\mathbf{S}, \mathbf{A}))^2 \quad (23)$$

$$\nabla_{\phi} J(\phi) = N^{-1} \sum a = \pi_{\phi}(\mathbf{S}) \nabla_{\phi} \pi_{\phi}(\mathbf{S}) \quad (24)$$

Algorithm 2. Dynamic decision of travel ban policies with TD3

Table 1
Epidemic model parameters.

Parameter	Description	Value	Source
β_I	The infectious rate contracting infected people	0.7	Li et al. (2020)
κ	The ratio of β_I and β_E	1	Wu (2020)
β_E	The infectious rate contracting latent people	$\kappa \beta_I$	Calculation
β_{Air}	The infectious rate in airplane	1.1	Author assumption
β_{HST}	The infectious rate in HST	0.9	Author assumption
β_{Coach}	The infectious rate in coach	0.7	Author assumption
β_{Car}	The infectious rate in car	0	Author assumption
D_I	Latency period	3.69	Li et al. (2020)
D_R	Infectious period	3.47	Li et al. (2020)
δ	Reporting rate	0.14	Li et al. (2020)
σ_I	The incidence of unreported patient	$(1 - \delta)/D_I$	Calculation
σ_H	The incidence of hospitalized patient	δ/D_I	Calculation
γ_{IR}	The cure rate of unreported patient	$0.85/D_R$	Partly author assumption
γ_{ID}	The fatality rate of unreported patient	$0.15/D_R$	Partly author assumption
γ_{HR}	The cure rate of hospitalized patient	$0.99/D_R$	Partly author assumption
γ_{HD}	The fatality rate of hospitalized patient	$0.01/D_R$	Partly author assumption
ξ	Infectious rate decay parameters	0.1	Curve fitting

3.3. Execution procedure

The last subsection introduces the executive procedure to synthesize the proposed RL framework with the solving framework for dynamically determining the travel ban and recovering policies. For the implementation, we first need to establish a simulation environment for the city network considering epidemic pandemic and people moving. The detailed execution procedure is illustrated in Fig. 3. As observed, there are two key components in the framework, i.e., emulator, and RL brain. During the total studying period, the city network model coupled with the SEIHRD model continues to simulate the epidemic spreading process. First, the travel demand is generated by some demand models (e.g., the gravity model shown in Eq. (8)). After applying the travel ban from

Algorithm 2: Dynamic decision of travel ban policies with TD3

- 1 Initialize critic networks $Q_{\theta_1}, Q_{\theta_2}$, and actor network π_{ϕ} with random parameters θ_1, θ_2, ϕ ;
- 2 Initialize target networks $\theta_1' \leftarrow \theta_1, \theta_2' \leftarrow \theta_2, \phi' \leftarrow \phi$;
- 3 Initialize replay buffer \mathcal{B} ;
- 4 **for** $t = 1$ **to** T **do**
- 5 Select action with exploration noise $\mathbf{A} \sim \pi_{\phi}(\mathbf{S}) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma)$;
- 6 Check the feasibility of \mathbf{A} to ensure that the restarted link will not be closed again;
- 7 Apply action and observe reward r as well as new state \mathbf{S}' ;
- 8 Store transition tuple $(\mathbf{S}, \mathbf{A}, r, \mathbf{S}')$ in \mathcal{B} ;
- 9 Sample mini-batch of K transitions $(\mathbf{S}, \mathbf{A}, r, \mathbf{S}')$ from \mathcal{B} ;
- 10 $\tilde{\mathbf{A}} \sim \pi_{\phi'}(\mathbf{S}') + \epsilon, \epsilon \sim \text{clip}(\mathcal{N}(0, \tilde{\sigma}), -c, c)$;
- 11 Check the feasibility of $\tilde{\mathbf{A}}$ to ensure that the restarted link will not be closed again;
- 12 $y \leftarrow r + \mu \min_{i=1,2} Q_{\theta_i}(\mathbf{S}', \tilde{\mathbf{A}})$;
- 13 Update critics $\theta_i \leftarrow \min_{\theta_i} K^{-1} \sum (y - Q_{\theta_i}(\mathbf{S}, \mathbf{A}))^2$;
- 14 **if** $t \bmod d$ **then**
- 15 Update ϕ by the deterministic policy gradient:
- 16 $\nabla_{\phi} J(\phi) = K^{-1} \sum \nabla_{\mathbf{A}} Q_{\theta_1}(\mathbf{S}, \mathbf{A})|_{\mathbf{A}=\pi_{\phi}(\mathbf{S})} \nabla_{\phi} \pi_{\phi}(\mathbf{S})$;
- 17 Update target networks:
- 18 $\theta_i' \leftarrow \alpha \theta_i + (1 - \alpha) \theta_i'$;
- 19 $\phi' \leftarrow \alpha \phi + (1 - \alpha) \phi'$;
- 20 **end**
- 21 **end**

Table 2
Parameter settings for TD3.

Parameter	Value
Learning rate	0.001–0.005
Discount factor	0.99
Training batch size	64
Policy delay	2
Target smoothing coefficient	0.05
Min replay buffer size	150
Max replay buffer size	5000
Training batch size	64
Step size (day)	50
Number of iterations	500

the RL brain, the network updates the link status and finishes the travel mode choice process. Then, those traveling people enter the link-based SEIHRD model while other people staying in cities enter node-based SEIHRD model. All the SEIHRD models are established separately and are independent of each other to consider the differences between cities and links. At last, the traveling people enter their target city, and the whole city network updates situation as well as calculates current reward. From the policymaker side, the RL brain receives the situation of epidemic spreading as the state, and then gives the current time travel ban policy (every link is open or not) to the city network. When one step simulation is over, the city network returns the reward.

4. Numerical examples

4.1. Test settings

In this section, we perform a variety of experiments to test the performance of our proposed travel ban decision framework for gaining some insights. We evaluate the solutions on the metrics of cumulative reward defined by Eq. (21).

4.1.1. Epidemic model settings

The parameters setting reflects the characteristics of the disease as well as has an important impact on the performance of the epidemic model. As humans' understanding of COVID-19 is constantly in progress, the current selection of parameters is only given by the existing literature or news. The epidemic model parameter setting is shown in Table 1. Asymptomatic patients can infect others, but there is no evidence supporting the infectivity of the asymptomatic patient is weaker than infectious patient (Wu, 2020), and we take $\kappa = 1$. At the same time, the cure rate and fatality rate of unreported patients and hospitalized patients (reported patients) are different. As of July 30, the country with the highest crude fatality rate (death cases/confirmed cases) was Yemen, at 28.35%, followed by France at 16.33%. So we estimate that the fatality rate of unreported patients is 15%. Without a run on medical resources, the fatality rate of hospitalized patients is much lower than unreported patients. In early March, the fatality rate of patients in other provinces in China except Hubei was 0.86%. So we estimate the fatality rate of hospitalized patients is 1%.

However, there is no study to carefully analyze the infection rate on different travel tools. We set the infection rate of different modes according to the crowdedness and capacity of different modes as well as referring to the city infection rate. We set $\beta_{Air} > \beta_{HST} > \beta_{Coach} > \beta_{Car}$. Among those three collective transits, the airplane is the most crowded (economic class). Although the capacity of HST is the highest (considering all the cabins), the probability of people moving across cabins is not high. The vast majority of coach services are for short trips; coach passengers normally arrive at the station for a short time, and the infectious risk at the station is much lower than the airport or railway station. The research from Zhang et al. (2020) also suggests that flights contribute the most to confirmed cases, followed by HST and coach. For the private car, we assume the $\beta_{Car} = 0$ because the car can take up to 5

people, and it is generally not shared with strangers for intercity traffic. Besides, the infectious rate in transportation will not decay.

4.1.2. Economic growth model settings

We have established an econometric model to study the marginal contribution of traffic turnover to economic growth. We use Total Retail Sales of Consumer Goods (TRSCG) as the dependent variable and Public Finance Expenditure (PFE, denoted by X_F), Profits of Industrial Enterprises (PIE, denoted by X_I), Total Import and Export (TIE, denoted by X_E), total traffic turnover volume (denoted by X_g) as the explanatory variable. Mobility can well promote the development of the consumer industry. PFE reflects government intervention in society as well as the country's macroeconomic situation. PIE affects the level of disposable income. TIE reflects international trade, which affects a lot of industries in China. We collected monthly data of China from 2010 to 2019, a total of 120 observations, and regressed the model. The results indicate that X_g has a significant impact on the TRSCG ($p = 0.019$). We estimate the marginal effect of X_g , denoted by w_g , and calculate the marginal economic growth from X_g by $\Delta E_t = w_g \Delta X_g$. Besides, according to the public information,¹ the average treatment cost in China τ_H and death loss τ_D are set as 32 thousand yuan and 200 thousand yuan, respectively.

The setting of w_g, τ_H, τ_D affects the objective function. It also reflects the importance government places on economic development and epidemic control. The sensitivity analysis is conducted in section 4.2.

4.1.3. RL model settings

The setting of parameters affects the performance of RL algorithm, which is shown in Table 2. We adjust the learning rate in different scenarios for better learning performance. The total studying period is 50 days, and the number of training iterations is 500.

4.1.4. Epidemic and traffic data

We select 15 metropolises in China and collect the epidemic data (the number of confirmed cases, recovered cases, and death cases) as well as the intercity traffic data. The epidemic data is from the National Health Commission (NHC), and the traffic data can be collected from the Baidu immigration index. We collect the epidemic and traffic data from Jan.23th to Mar.13th (50 days), which involves a complete outbreak and the control process of COVID-19 in China. Facing the pandemic of COVID-19, most provincial governments initiated the first-level response and considered to perform a travel ban around Jan.23th, while millions of people were moving across the cities to return to their hometown. It was the day when Wuhan, the first city to report the COVID-19 case, cut off traffic connections with outside. About 50 days later, the COVID-19 pandemic in China had been roughly controlled. Another two weeks later, the Chinese government declared the indigenous pandemic interruption on Mar.28th.

4.2. Experiments on hypothetical city networks

4.2.1. Virtual city network construction

Firstly we perform our model in some virtual city networks. However, there may not have a direct road connection between some pairs of cities, i.e., the graph representing the network is not necessarily a complete graph. Some studies have pointed out that the terrestrial transportation network, such as railway network or highway network, is a type of small-world network (Aldrich et al., 2015; Viana & da Fontoura Costa, 2011). Also, the small-world network architecture is used to describe a real-world application for a next-generation airline network (Sawai, 2012). The small-world network is a type of mathematical graph in which most nodes are not neighbors of one another, but the neighbors

¹ Source: <https://baijiahao.baidu.com/s?id=1671090168528252190&wfr=spider&for=pc>.

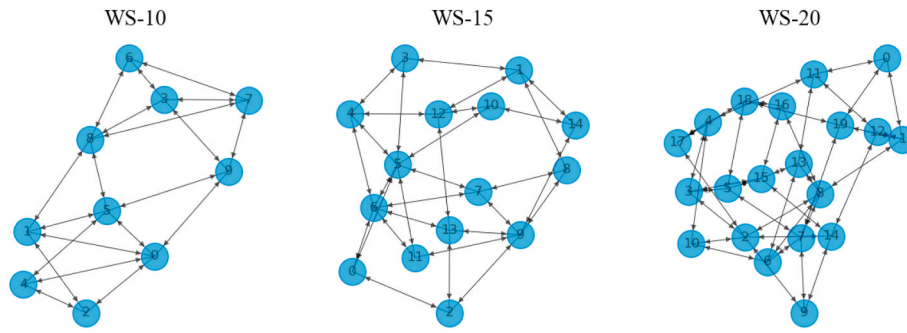


Fig. 4. The illustration of WS network.

Table 3
Information of 15 metropolises.

City name	Population (10 ⁴)	Longitude	Latitude	Initial confirmed cases	t_{free}
Beijing	2154	116.41	39.90	44	0
Shanghai	2428	121.47	31.23	20	1
Guangzhou	1531	113.26	23.13	7	0
Shenzhen	1344	114.06	22.54	15	0
Chengdu	1658	104.07	30.57	7	1
Hangzhou	1036	120.16	30.27	6	2
Chongqing	3124	106.55	29.56	54	1
Wuhan	1121	114.31	30.59	495	1
Xian	1000	108.94	34.34	2	2
Tianjin	1562	117.20	39.08	5	1
Nanjing	850	118.80	32.06	3	2
Zhengzhou	1035	113.63	34.75	3	2
Changsha	839	112.94	28.23	8	0
Shenyang	832	123.43	41.81	2	2
Qingdao	950	120.38	36.07	3	2

of any given node are likely to be neighbors of each other and most nodes can be reached from every other node by a small number of hops or steps. In the existing literature, there are a lot of kinds of small-world networks, among which the best-known family was proposed by Watts and Strogatz (1998), called Watts-Strogatz (WS) network. A WS network is defined to be a network where the typical distance between two randomly chosen nodes (the number of steps required) grows proportionally to the logarithm of the number of nodes N in the network. The WS network is constructed process can be described by Algorithm 3.

Algorithm 3. WS network construction

Algorithm 3: WS network construction

- Input** : The desired number of nodes N , the mean degree K (assumed to be an even integer), and a special parameter β , satisfying $0 \leq \beta \leq 1$ and $N \gg K \gg \ln N \gg 1$;
- Output:** An undirected graph with N nodes and $\frac{NK}{2}$ undirected edges;
- 1 Construct a regular ring lattice, a graph with N nodes each connected to K neighbors, $K/2$ on each side. That is, if the nodes are labeled $0 \dots N - 1$, there is an edge (i, j) if and only if $0 < |i - j| \bmod (N - 1 - \frac{K}{2}) \leq \frac{K}{2}$;
 - 2 **for** node $i = 0, \dots, N - 1$ **do**
 - 3 Take every edge connecting i to its $K/2$ rightmost neighbors, that is every edge $(i, j \bmod N)$ with $i < j \leq i + K/2$;
 - 4 Replace $(i, j \bmod N)$ with (i, k) with probability β where k is chosen uniformly at random from all possible nodes while avoiding self-loops ($k \neq i$) and link duplication;
 - 5 **end**
-

We use the WS network to represent the traffic connections between cities. We generate three networks with $N = 10, 15, 20, K = 4, \beta = 0.6$ and name them WS-10, WS-15, WS-20. Fig. 4 illustrates three city networks, which have 40, 60, 80 directed arcs, respectively.

Other information about the city is randomly generated within a certain range. The location of each node is generated randomly, and the distance between the two cities is calculated from their locations. The city population is a random number from 0.5 to 10 million. The travel demands between cities are generated by the gravity model introduced in section 2.3. The initial number of confirmed cases varies from 0 to 30. The time for the government to initiate the first-level response t_{free} is set as 0. The travel time and ticket price of different travel modes are generated according to the distance between the two cities.

4.2.2. Real city network construction

In order to show the model's ability to apply in the real world, we select 15 metropolises in China and construct a real city network, named Real-15. The population, location, initial confirmed cases and the time to start first-level emergency response are shown in Table 3. The start day is Jan. 23th, 2020. Considering that those cities are all metropolises, transportation is very convenient, so we assume the Real-15 is a fully-connected network.

At the same time, we collect the epidemic data (the number of confirmed cases, cure cases and death cases) and the traffic data. Although the immigration index is collected from vehicle data, it reflects the travel intensity in different cities.

Fig. 5 shows the change of migration index in the whole studying period. The migration index quickly decays to 0 in all the cities, espe-

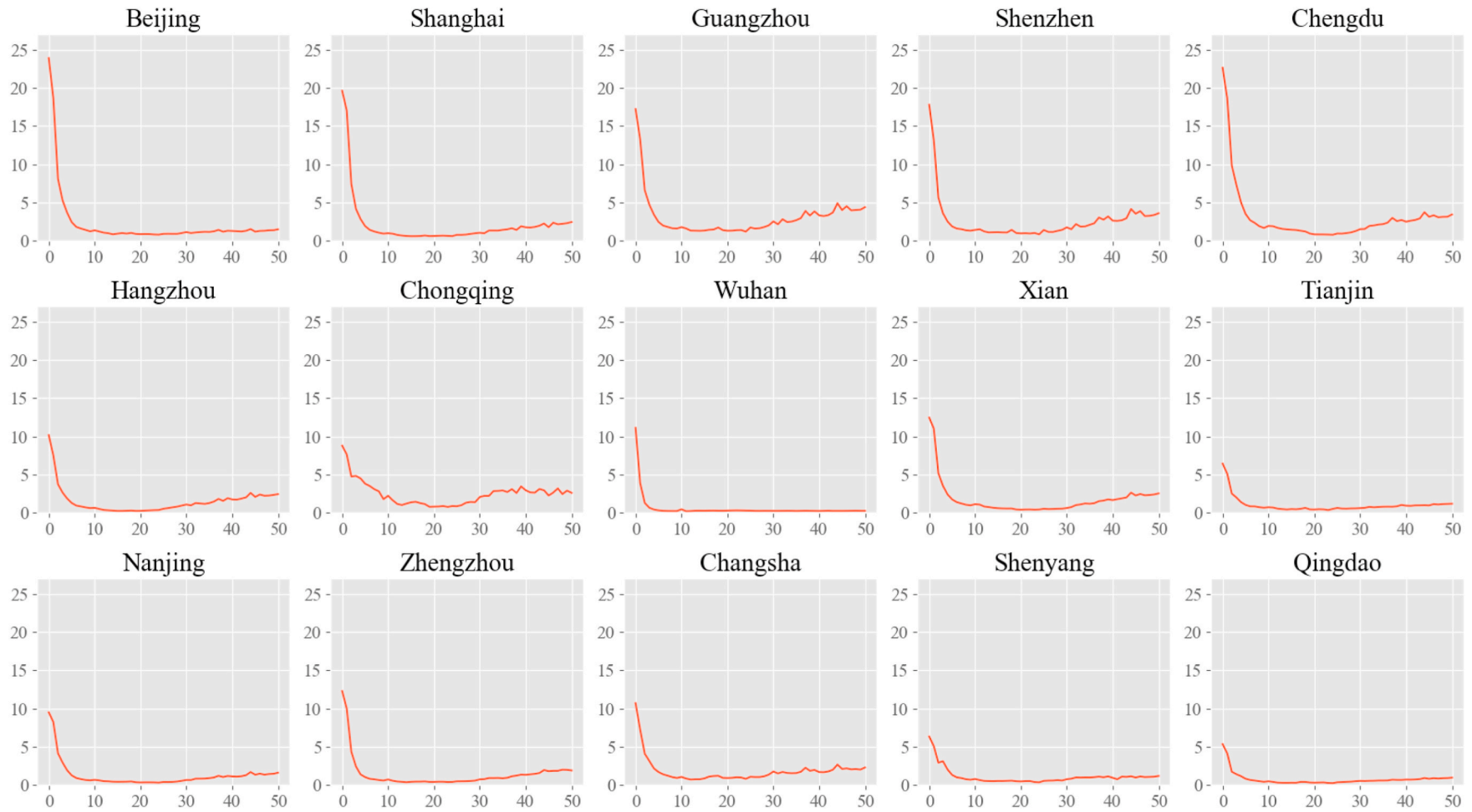


Fig. 5. Migration index of 15 cities.

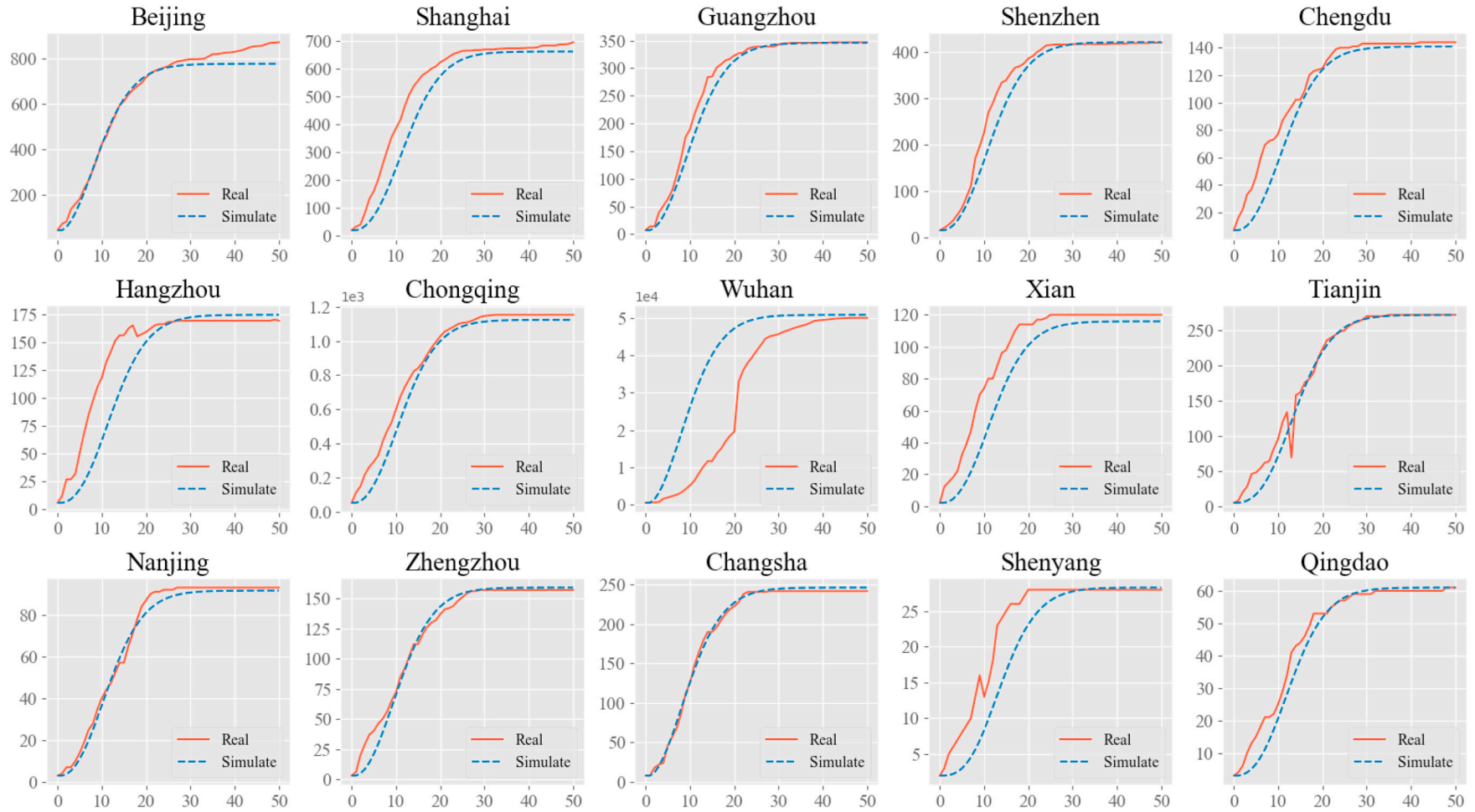


Fig. 6. The calibration result of 15 cities.

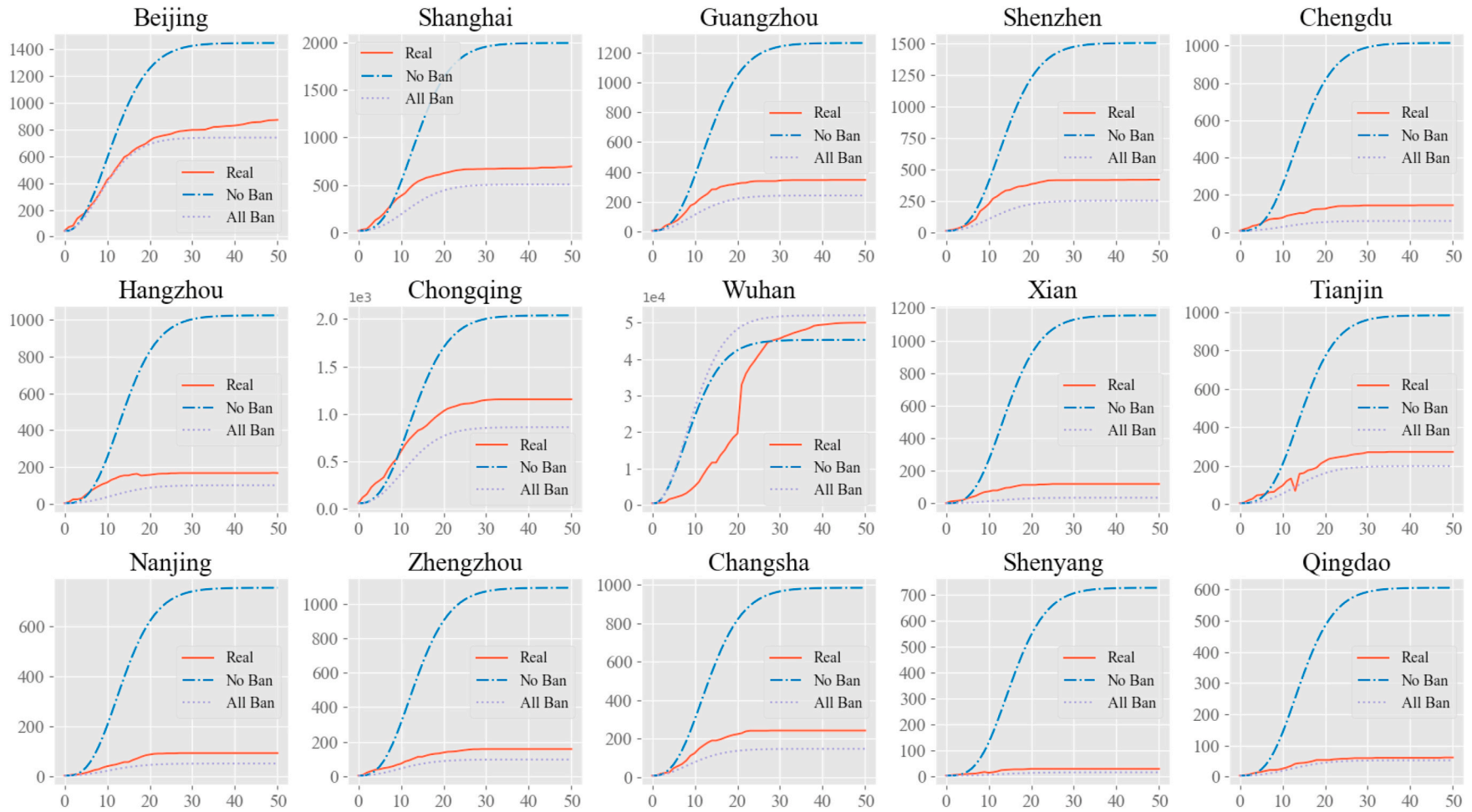


Fig. 7. The comparison of benchmark policies in 15 cities.

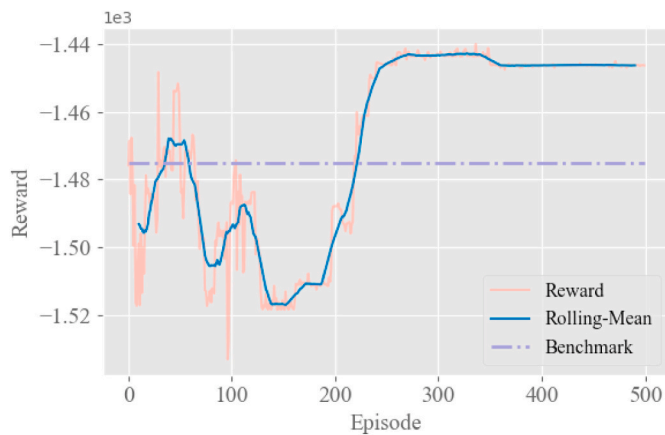


Fig. 8. Convergence of TD3 framework online for travel ban policy decision.

cially in Wuhan because of the travel ban. And a lot of cities tried to recover intercity traffic in early March. Some southeastern cities, like Guangzhou, Shenzhen, and Chengdu, are recovering particularly fast. The Baidu map can also provide the share rate of migration index on different destination cities. Based on the migration index, sharing rate and standard volume, we can estimate the daily traffic demand between all cities during the whole period.

4.2.3. Epidemic model calibration

According to the real epidemic data and traffic data, we can calibrate our epidemic model for every city. The comparison between real data and simulation results is shown in Fig. 6. The curve of cumulative confirmed cases fits well in most cities. We note that the curves produced by the SEIHRD model is smooth, but the real curves may not be as smooth due to some incidents, such as statistical errors, changes in diagnostic criteria, etc.

4.2.4. Experiments result analyses

In three virtual city networks, the travel demand is generated by the gravity model shown in Eq. (8). In the real city model, the travel demand is calculated by the average travel demand from Jan.17th to Jan.19th. It is multiplied by a random number to reflect the randomness of demand.

We design two benchmark policies for comparison, i.e., all-ban policy and no-ban policy. Under the all-ban policy, all the links are close in the whole period, and the traffic turnover volume is zero. Conversely, all the links keep open under the no-ban policy and the traffic turnover volume is maximized. Fig. 7 illustrates the comparison between all-ban policy and no-ban policy in Real-15. Because very strict

Table 4
Tabular comparison of learned policies and benchmark.

Network	Policy	Obj(10 ⁶)	Confirmed cases(10 ³)	Turnover(10 ⁹)	Improvement
WS-10	All ban	-1475.12	38.61	0.00	\
	No ban	-1478.66	39.70	19.90	
	RL-SCO1	-1436.21	37.79	4.38	2.64%
	RL-SCO2	-1439.98	37.80	2.52	2.38%
WS-15	All ban	-2052.70	53.29	0.00	\
	No ban	-2216.61	60.00	30.73	
	RL-SCO1	-2002.33	52.41	6.19	2.45%
	RL-SCO2	-1984.13	52.35	11.82	3.34%
WS-20	All ban	-2652.22	69.13	0.00	\
	No ban	-2708.81	73.67	30.07	
	RL-SCO1	-2507.47	66.36	10.78	5.46%
	RL-SCO2	-2510.63	66.17	8.28	5.34%
Real-15	All ban	-2311.43	54.67	0.00	\
	No ban	-2295.67	61.16	89.89	
	Real	-2284.22	54.75	10.38	
	RL-SCO1	-2254.24	55.10	25.65	1.31%
	RL-SCO2	-2254.25	56.55	44.36	1.31%

Table 5
Sensitivity analysis for parameters in objective function.

	$\frac{Obj_{Allban} - Obj_{Noban}}{ Obj_{Noban} }$	w_g				
		-60%	-30%	+0%	+50%	+100%
τ_H, τ_D	-60%	7.39%	4.51%	1.44%	-4.14%	-10.39%
	-30%	8.97%	7.39%	5.77%	2.93%	-0.09%
	+0%	9.58%	8.50%	7.39%	5.49%	3.51%
	+50%	10.05%	9.34%	8.62%	7.39%	6.13%
	+100%	10.28%	9.76%	9.22%	8.32%	7.39%

travel restrictions were implemented in real life, the number of confirmed cases under the all-ban policy is similar to the actual situation. However, if the no-ban policy is implemented, the number of infected people in cities other than Wuhan will rise sharply. Therefore, the travel ban can well prevent the spreading of the epidemic among different cities.

The RL framework is implemented in city network WS-10, WS-15, WS-20 and Real-15. The learning performance in WS-10 is shown in Fig. 8. The TD3 method converges well after 300 epochs. We set the higher objective function value of all-ban and no-ban policies as the benchmark, and we can find that the RL framework can find a better policy outperforming the benchmark with obvious improvements.

Table 4 shows the comparison result of travel ban policies produced by the RL framework and benchmark policies. As mentioned above, we introduce two travel ban scenarios. In scenario I, the government decides both the close time and recovery time. In scenario II, the government only decides the recovery time. RL-SCO1 and RL-SCO2 stand for the RL strategies under scenario I and scenario II, respectively. It can be found that all-ban policy is slightly better than no-ban policy, and the policies applied in actual situations in Real-15 is better than no-ban and all-ban policies. However, the policies given by the RL framework beats all the benchmarks in all networks. In WS-10, the policy from RL-SCO1 achieves a lower number of confirmed cases while ensuring some mobility. In WS-15, RL-SCO2 gives the best policy, traffic turnover volume is guaranteed with slightly increasing confirmed cases. In WS-20, RL-SCO1 guarantees traffic turnover volume at the expense of a small increase in confirmed cases. In Real-15, RL-SCO1 and RL-SCO2 achieve nearly the same objective function value, though the policies are different. In short, the RL framework helps to come up with a better travel ban policy, which can increase the value of the objective function by 1.31%–5.46%.

The setting of parameters in the objective function has a great effect on the decision. The values of w_g, τ_H, τ_D reflect the importance placing on epidemic control and economic development, also affect our evaluation of the policy. Table 5 shows the sensitivity analysis in WS-15 under All-

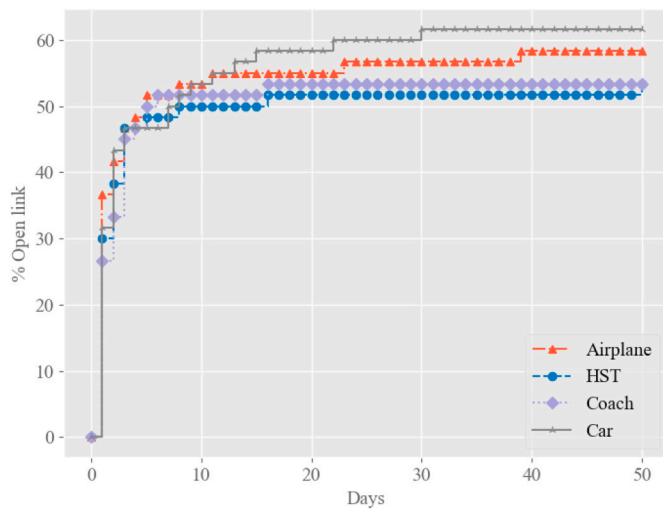


Fig. 9. Analysis of learned travel ban policy.

ban and No-ban policies. The values of w_g, τ_H, τ_D will not affect the simulation result under two benchmark policies, but change the value of the objective function. We changed these values from a decrease of 60% to an increase of 100%. The number in the table represents the improvement of the All-ban policy compared to the No-ban policy. The upper right part of the table represents the case that economic operations are very important while the epidemic prevention and control are relatively not urgent, and in that case, No-ban policy is better. On the contrary, the case in the lower left part of the table puts epidemic prevention first. In this case, it would be better to adopt the All-ban policy.

By analyzing the learned policy, we can also understand its characteristics. Fig. 9 illustrates the learned best policy in WS-15 under scenario II. The x-axis is time and the y-axis is the percent of open links in each day. The policy makes the open rate increase with time. However, we can observe that the open rate of the private car keeps the highest while that of HST is the lowest. The result suggests that we can develop different policies for different travel modes, for example, consider removing highway restrictions first.

At the same time, we find some characteristics of the early open link. Fig. 10 illustrates the bar-plot of link demand in the middle period in WS-10. Some links with low demand are recovered first. The result suggests that we can consider recovering low demand traffic links, which can guarantee the necessary economic connections without too much risk of the epidemic spreading.

4.2.5. Lessons learned from post COVID-19 world

The pandemic of COVID-19 has greatly affected the economic development and people’s daily lives across the world. Public transport increases the risk of people-to-people contact and aggravates the spread of COVID-19 across cities. It is important for the government to take effective policy intervention to reduce epidemic damage. Meanwhile, the travel restriction also disrupts a wide range of industries (e.g., retail sales, tourism, entertainment, etc.), causing the recession of the national economy. For example, China’s GDP in the first quarter fell by 5.3% compared to the same period last year. Therefore, it is very important to develop effective policy-making criteria and implement them prudently for balancing the above sides.

In post COVID-19 world, on the one hand, the pandemic of COVID-19 continues one after another almost in every corner of the world. In countries such as Spain, France and Switzerland,² strict travel restriction

² <https://www.forbes.com/sites/alexledsom/2020/12/13/december-eu-travel-restrictions-by-country-christmas-quarantine-and-covid-19-test-requirements/?sh=7711c7de3f15>.

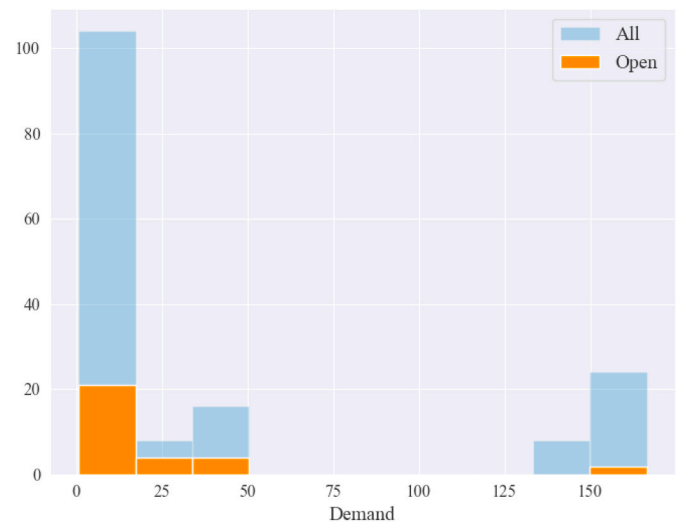


Fig. 10. Demand distribution of the recovered links.

policies are being implemented. From SARS(2003), MERS(2015) to COVID-19(2019), mankind is still facing threats from many other infectious diseases (e.g. Ebola, Zika, H7N9). With the availability of big data nowadays, data-driven decision-making becomes crucial for governments. In the early stages, epidemiologists can obtain some key parameters of the infectious disease as well as some characteristics from the confirmed cases, including infectious rate β . With the key information, the model can simulate inter-city spreading and provide useful suggestions for travel managers. At the same time, with the proposed methodology and numerical experiments, we can learn that:

- Cutting off all the traffic connections is essential in the early stages of the severe spread of the epidemic.
- Highway can be considered the first to recover because traveling with private cars owns the least cross-infection probability.
- HST, airplane, and coach should be carefully considered reopen because the clustering risk still exists; it may be infected not only during the trip but also at the stations with potentially huge crowd.
- Some traffic links with relatively low demands can be recovered first, as they can guarantee necessary economic connections without too much risk of the epidemic spreading.

5. Conclusions

This paper develops a methodological framework for multi-modal dynamic travel ban policy determination under the pandemic of epidemic to balance epidemic control and regional economic development. Coupled with SEIHRD epidemic model, we develop a city network model containing multi-modal traffic between cities. By applying node-based and link-based epidemic models, the city network model takes the risk of infection in cities and transportation into account. Base on MNL, an mode choice model is established to allocate demand under capacity constraints. At the same time, mobility plays an important role in the social-economic system as well as has a direct or indirect effect on many industries. We establish a model to estimate the marginal economic growth contributed by traffic turnover volume in order to demonstrate the trade-off between the control of epidemic spreading and economic loss.

For supporting the dynamic decision-making, we develop an RL-based framework for online multi-modal travel ban decision. We adopt TD3 as the RL algorithm to reduce variance and avoid over-estimation. We construct three virtual city networks using WS network, WS-10, WS-15, WS-20. as well as a real city network Real-15. The calibration results indicate that our model fits well with realistic data.

We adopt all-ban and no-ban policies as the benchmarks. The policies given by the RL framework outperform any benchmark and increase by 1.31%–5.46% in four city networks in terms of the objective function, which considers the losses caused by the spread of the epidemic and the economic growth contributed by traffic turnover volume. At the same time, the learning result indicates that differentiating policies for different intercity travel modes can be considered.

This study introduces the RL framework in network-level epidemic modeling and considers the online multi-modal travel ban decision problem. Our study only provides a preliminary sketch; several extensions can be made in future studies. First, some additional policies considering the feature of the complex network can be studied. Second, the economic growth model can be further enhanced for mining the microscopic mechanisms to achieve better estimation on actual cases. Third, our numerical example only considers the realistic cases in China, but now the balance between epidemic control and economic maintenance is an urgent issue faced by most of governments around the world. Lastly, the current-adopted epidemic model only considers one-wave spreading, and how to extend it to incorporate multi-wave cases remains a valuable and challenging topic to be investigated in the future.

Acknowledgement

This research is supported by grants from National Key Research and Development Program of China (2018YFB1601600).

References

- Aldrich, P.R., El-Zabet, J., Hassan, S., Briguglio, J., Aliaj, E., Radcliffe, M., Mirza, T., Comar, T., Nadolski, J., Huebner, C.D., 2015. Monte Carlo tests of small-world architecture for coarse-grained networks of the United States railroad and highway transportation systems. *Phys. Stat. Mech. Appl.* 438, 32–39.
- Aron, J.L., Schwartz, I.B., 1984. Seasonality and period-doubling bifurcations in an epidemic model. *J. Theor. Biol.* 110, 665–679.
- Aslani, M., Mesgari, M.S., Wiering, M., 2017. Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events. *Transport. Res. C Emerg. Technol.* 85, 732–752.
- Beck, M.J., Hensher, D.A., 2020. Insights into the impact of covid-19 on household travel and activities in Australia the early days under restrictions. *Transport Pol.* 96, 76–93.
- Chen, T.-M., Rui, J., Wang, Q.-P., Zhao, Z.-Y., Cui, J.-A., Yin, L., 2020. A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infect. Dis. Poverty* 9, 1–8.
- Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M.E., Longini, I.M., Vespignani, A., 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* 368, 395–400.
- Donzelli, M., 2010. The effect of low-cost air transportation on the local economy: evidence from southern Italy. *J. Air Transport. Manag.* 16, 121–126.
- Fujimoto, S., Van Hoof, H., Meger, D., 2018. Addressing Function Approximation Error in Actor-Critic Methods arXiv preprint arXiv:1802.09477.
- Guan, D., Wang, D., Hallegatte, S., Davis, S.J., Huo, J., Li, S., Bai, Y., Lei, T., Xue, Q., Coffman, D., et al., 2020. Global supply-chain effects of covid-19 control measures. *Nat. Human Behav.* 1–11.
- Inoue, H., Todo, Y., 2019. Firm-level propagation of shocks through supply-chain networks. *Nature Sustain.* 2, 841–847.
- Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. - Ser. A Contain. Pap. a Math. Phys. Character* 115, 700–721.
- Kim, B., Kim, J., Huh, S., You, S., Yang, I., 2020. Multi-objective predictive taxi dispatch via network flow optimization. *IEEE Access* 8, 21437–21452.
- Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization arXiv preprint arXiv:1412.6980.
- Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R.M., Sun, F., Jit, M., Munday, J.D., et al., 2020. Early Dynamics of Transmission and Control of Covid-19: a Mathematical Modelling Study. *The lancet infectious diseases.*
- Leung, K., Wu, J.T., Liu, D., Leung, G.M., 2020. First-wave covid-19 transmissibility and severity in China outside hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet* 395 (10233), 1382–1393.
- Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J., 2020. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science* 368, 489–493.
- Nakamura, H., Managi, S., 2020. Airport risk of importation and exportation of the covid-19 pandemic. *Transport Pol.* 96, 40–47.
- Ni, S., 2009. Research on Modeling of Infectious Disease Spreading Based on Complex Network Theory. Ph.D. thesis. Tsinghua University.
- Oum, T.H., Wang, K., 2020. Socially optimal lockdown and travel restrictions for fighting communicable virus including covid-19. *Transport Pol.* 96, 94–100.
- Sawai, H., 2012. Reorganizing a new generation airline network based on an ant-colony optimization-inspired small-world network. In: *2012 IEEE Congress on Evolutionary Computation. IEEE*, pp. 1–8.
- ShunLi, W., LinLiu, J., PingXu, G., 2005. Co-integration Analysis of the Impact of the Gdp Net Growth on the Traffic Turnover Volume Net Growth. *Central South Highway Engineering*, p. 51.
- Tang, B., Wang, X., Li, Q., Bragazzi, N.L., Tang, S., Xiao, Y., Wu, J., 2020. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *J. Clin. Med.* 9, 462.
- Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M.U.G., Li, B., Cai, J., Xu, B., Yang, Q., Wang, B., Yang, P., Cui, Y., Song, Y., Zheng, P., Wang, Q., Bjornstad, O.N., Yang, R., Grenfell, B.T., Pybus, O.G., Dye, C., 2020. An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in China. *Science* 368, 638–642.
- Viana, M.P., da Fontoura Costa, L., 2011. Fast long-range connections in transportation networks. *Phys. Lett.* 375, 1626–1629.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of “small-world”networks. *Nature* 393, 440–442.
- Wells, C.R., Sah, P., Moghadas, S.M., Pandey, A., Shoukat, A., Wang, Y., Wang, Z., Meyers, L.A., Singer, B.H., Galvani, A.P., 2020. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl. Acad. Sci. Unit. States Am.* 117, 7504–7509.
- Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, China: a modelling study. *Lancet* 395, 689–697.
- Wu, Z., 2020. Contribution of asymptomatic and pre-symptomatic cases of covid-19 in spreading virus and targeted control strategies. *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi* 41, 801–805.
- Zhang, Y., Zhang, A., Wang, J., 2020. Exploring the roles of high-speed train, air and coach services in the spread of covid-19 in China. *Transport Pol.* 94, 34–42.