

Information Sciences

Policy and newly confirmed cases universally shape the human mobility during COVID-19

Kehan Li^{1,2,#}, Chao Li^{1,#}, Yinfeng Xiang¹, Fengxiang He², Shibo He¹, Jiming Chen^{1,*}, Yi Fang³, Dacheng Tao² & Youxian Sun¹

¹College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China;

²JD Explore Academy, JD.com, Beijing 100176, China;

³Westlake Institute for Data Intelligence, Hangzhou 310012, China

#Contributed equally to this work.

*Corresponding author (email: cjm@zju.edu.cn)

Received: 7 October 2021; Revised: 21 January 2022; Accepted: 8 February 2022; Published online: 12 May 2022

Abstract: Understanding how human mobility pattern changes during the COVID-19 is of great importance in controlling the transmission of the pandemic. This pattern seems unpredictable due to the complex social contexts, individual behaviors, and limited data. We analyze the human mobility data of over 10 million smart devices in three major cities in China from January 2020 to March 2021. We find that the human mobility across multi-waves of epidemics presents a surprisingly similar pattern in these three cities, despite their significant gaps in geographic environments and epidemic intensities. In particular, we reveal that the COVID-19 policies and statistics (i.e., confirmed cases) dominate human mobility during the pandemic. Thus, we propose a universal conditional generative adversarial network based framework to estimate human mobility, integrating COVID-19 statistics and policies via a gating fusion module. Extensive numerical experiments demonstrate that our model is generalizable for estimating human mobility dynamics accurately across three cities with multi-waves of COVID-19. Beyond, our model also allows policymakers to better evaluate the potential influences of various policies on human mobility and mitigate the unprecedented and disruptive pandemic.

Keywords: COVID-19, human mobility, generative adversarial network

INTRODUCTION

Backgrounds. The outbreak of unprecedented global pandemic coronavirus disease 2019 (COVID-19) has spread to over 200 countries and continued to ravage the world, leading to over 200 million infections and 4 million deaths [1] around the world. Modern cities with high population density and frequent commuting are natural hotbeds for virus transmission. What's worse, the threats from COVID-19 are far from disappearing. Nowadays, a new wave of COVID-19 Delta variant [2], appearing to cause more severe illness and be highly infectious, has wreaked havoc across the world again. The daily confirmed cases soar to 200,000 in the U.S. even after over 188 million vaccination [3]. These horrific phenomena remind us to rethink whether herd immunity can protect us considering the rapid virus mutations.

On the other hand, non-pharmaceutical interventions (NPIs) [4] have proven their effectiveness in controlling the transmission of COVID-19 in many countries [4]. For example, Wuhan, the first city attacked by

COVID-19 in China, suffered a drastic increase of tens of thousand confirmed cases, nearly overwhelming the local health care system [5]. The Chinese government successfully contained the epidemic in Wuhan and pushed the confirmed cases to zero (i.e., Zero-COVID strategy [6]) by strict lockdown and mobility restrictions, which prevent the other areas in China from the invasion of the epidemic.

Despite the superiority of mobility restrictions in cutting down the spread of COVID-19 and decreasing infection risks shown in Figure 1B, plenty of studies also discuss the negative effects of mobility restrictions on the economics [7–9], especially the brick-and-mortar business [10]. Consequently, the governments always face the dilemma in adopting strict mobility restrictions for public health [11] and economic recession, which is criticized by the media on either too late or early, too strict or loose policies [12]. Furthermore, there also have been a lot of previous studies pointing out that the human mobility level can forecast the trends of COVID-19 in China, Italy, and the U.K. [13–15] and potentially affect the economic [16, 17]. Therefore, it is imperative to model and estimate human mobility response under different mobility restriction policies, enabling the government to better control the epidemic transmission and adapt interventions during the pandemic. However, accurately predicting human mobility during COVID-19 is challenging for several reasons. First, the human mobility depends on complicated multi-scale dynamics, which is associated with varied sociodemographic characters [18], geographical conditions [19], available transportation and travel options [20]. Second, long-term epidemic-related prevention and control policies may have thoroughly reshaped the human mobility patterns and are still ongoing [21–23], leading to inconsistency compared with the normal history data. Third, the data availability also limits the related research on human mobility considering the COVID-19 itself is a Black Swan event.

Observations and hypotheses. We utilize the anonymous human mobility data conforming to the privacy standards uploaded by the smart devices, ranging from January 1, 2020 to March 31, 2021 of three cities, Beijing (BJ), Dalian (DL), and Shijiazhuang (SJZ), of China. We can thereby comprehensively analyze and compare the human mobility dynamics of multi-waves COVID-19 across cities. Specifically, we aggregate the human mobility (i.e., the sum of regional visits) of Beijing per week, which is shown in Figure 1A, which has undergone two waves of epidemics during 2020. The deeper colors (e.g., bright regions) of the heatmap indicate higher human mobility levels. Intuitively, there are very likely to exist similar patterns from decline to revival for each wave of COVID-19, considering the Chinese government carries out the Zero-COVID strategy from the beginning to the end. Our dataset is valuable for understanding how the Zero-COVID strategy affected human mobility dynamics and eliminated the COVID-19 in China.

Nevertheless, estimating the human mobility changes for the next wave of COVID-19 across cities is still challenging due to the significantly distinct patterns in geographical terrain situations and epidemic intensity (see Supplementary Material, extended Figure 1), let alone the very limited mobility data available. The latest research reveals strong correlations between COVID-19 transmission and human mobility changes [24]. In other words, it is very probable that COVID-19 related factors potentially influence the shift of human mobility levels. To better answer this question, we dive into the underlying logic that drives the change of human mobility depicted in Figure 1C, where the human mobility, policy, and COVID-19 statistics are tightly coupled. Specifically, the COVID-19 statistics largely reshape human mobility by directly leading to strict policies on social distancing or implicitly influencing people's willingness to travel. Then, the reshaped human mobility affects both epidemic control and economic status. There are three frequently used COVID-19 statistics, i.e., newly confirmed cases, cured cases, and death cases. To identify the principal

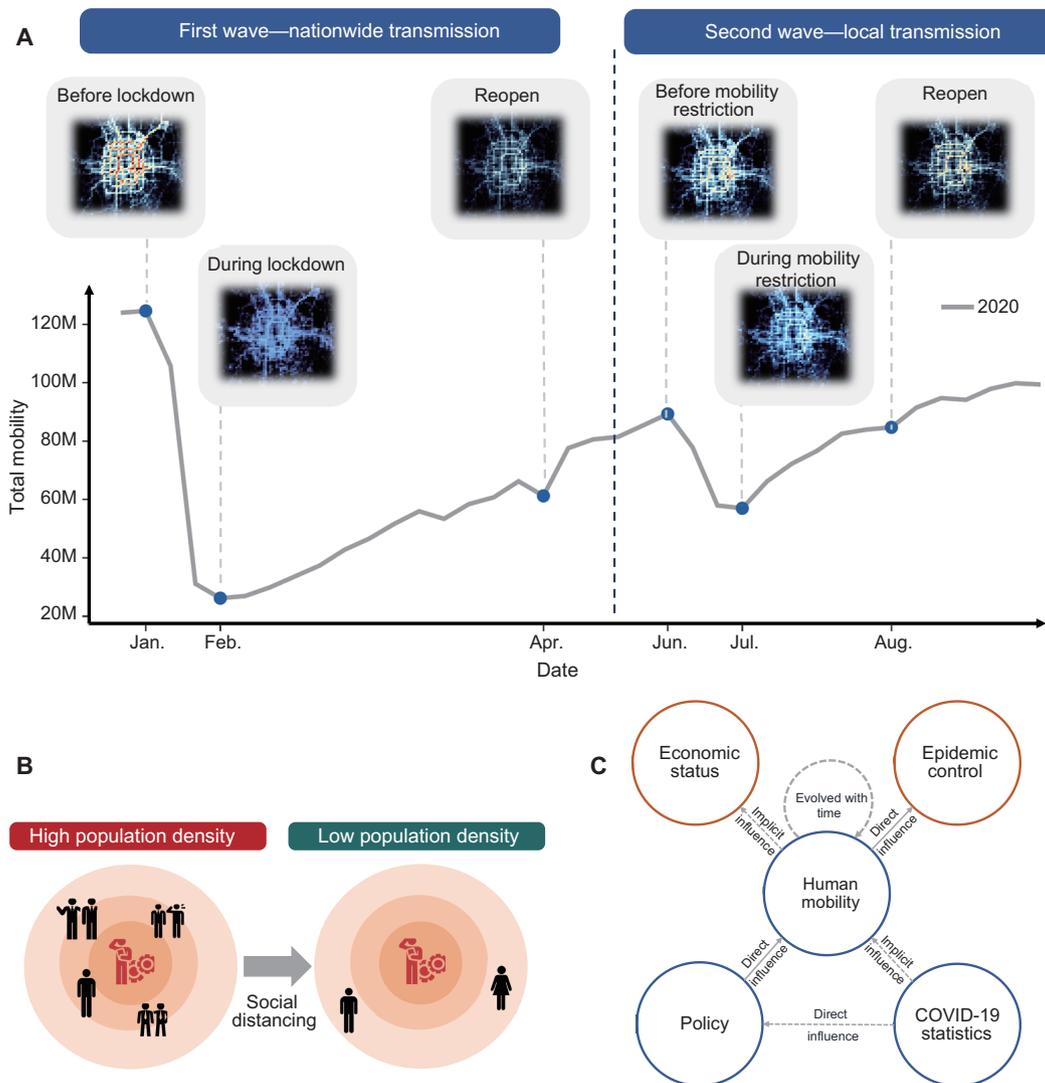


Figure 1 Problem overview. (A) Human mobility changes before, during, and after the lockdown or mobility restriction policies in Beijing, China. The bright regions represent a higher human mobility level compared with the dark regions. (B) Effects of mobility restriction and social distancing in mitigating COVID-19 transmission. (C) Relations among COVID-19 statistics, policy, and human mobility.

factors influencing human mobility, we normalize the human mobility level based on the first week of 2020 to remove the impacts from city attributes (e.g., scale and layouts). Then we plot corresponding trends along with COVID-19 statistics in Figure 2. We notice that the human mobility dynamics is highly correlated with the newly confirmed cases and mobility restriction policies [25] during the pandemic (see Supplementary Material, extended Figure 2). At the same time, the curves of dead and cured patients are delayed with fewer influences on human mobility. This phenomenon is also consistent with the actual situations, e.g., the burst increase of confirmed cases usually points to the beginning of strict lockdown or mobility restrictions, resulting in a sudden drop in human mobility level. Also, suppose there are no new confirmed cases for several consecutive weeks. In that case, human mobility will revive due to fewer mobility restrictions and people’s more willingness to travel in the city [26].

Related work. There exist plenty of studies investigating the interplay among policy, human mobility,

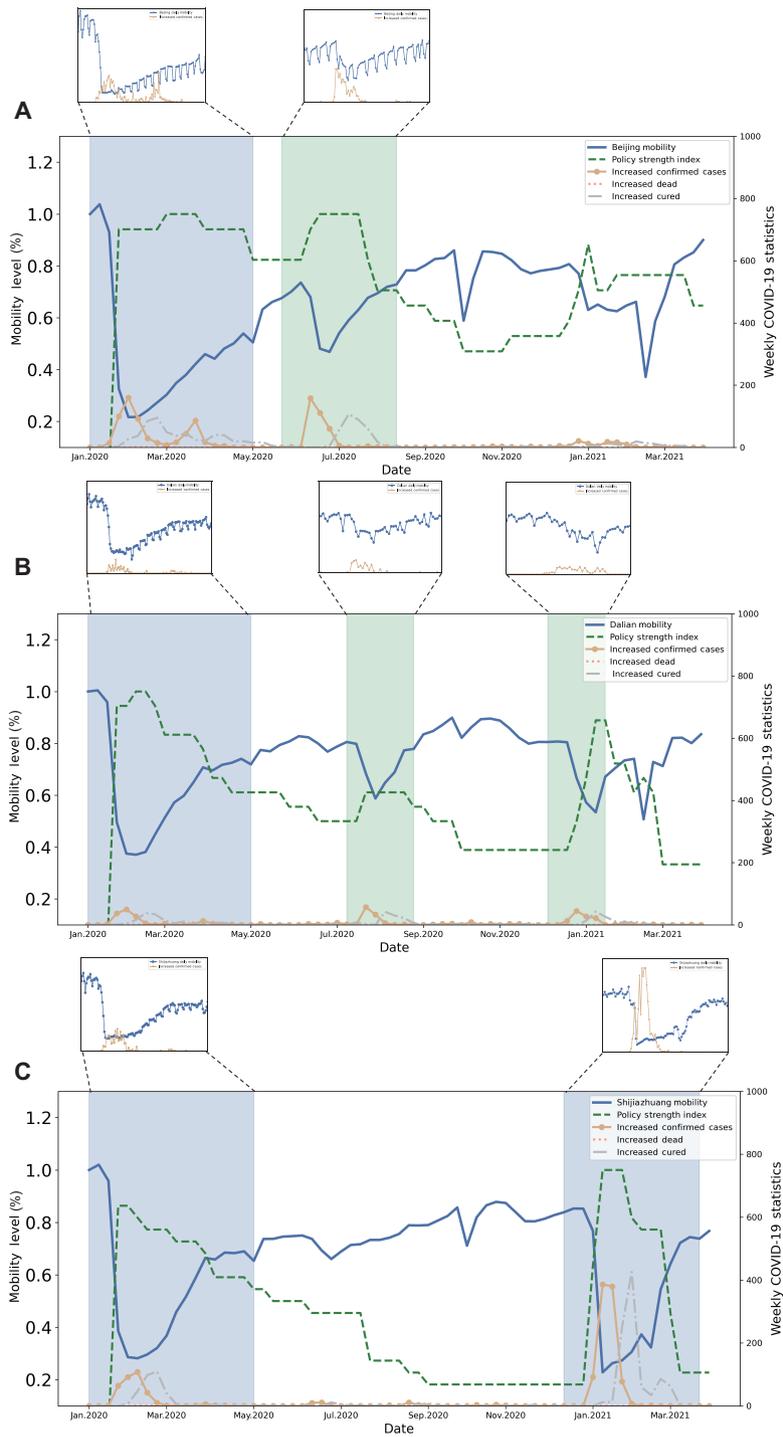


Figure 2 Interplay between human mobility level and confirmed cases across multiple cities. (A) Beijing mobility level, along with two waves of COVID-19, the first wave from January 19, 2020 to April 30, 2020, and the second wave from June 11, 2020 to August 6, 2020. (B) Dalian mobility level, along with three waves of COVID-19, the first wave from January 19, 2020 to April 30, 2020, the second wave from July 22, 2020 to August 30, 2020, and the third wave from December 16, 2020 to January 24, 2021. (C) Shijiazhuang mobility level, along with two waves of COVID-19, the first wave from January 19, 2020 to April 30, 2020, and the second wave from January 12, 2021 to February 10, 2021.

and epidemic dynamics [24,27,28]. Chinazzi *et al.* [27] investigated the effects brought by travel restriction and revealed that travel quarantines strongly delayed the nationwide COVID-19 transmission. Lai *et al.* [29] claimed that a combination of non-pharmaceutical interventions can achieve the most potent effects in controlling COVID-19. Quantified analysis of the human mobility data in the U.S. also reveals the impact of social distancing in decreasing substantial case growth [30]. Furthermore, authors in [24] proposed a dynamic mobility network to simulate the spread of COVID-19. A nationwide simulation is conducted to measure the burden brought by the COVID-19 in 2020 [31] based on a model-inference approach. These studies focused on analyzing the effects of policies on human mobility or statistically modeling the dynamics. Nevertheless, the long-term co-evolution patterns between policies and human mobility and their generalization ability over time and varied contexts are less studied.

Our solutions. As current deep learning-based human mobility prediction models focus on deterministic outputs based on rich history observations [32,33], they cannot generate probabilistic estimations of human mobility. Furthermore, considering the practical demands of policymakers, generative models, being able to learn the distribution of collected data and simulate multiple potential human mobility conditions under different policies, are more useful for human mobility estimation during COVID-19. For example, the government can formulate phased reopen plans based on the simulated human mobility response for the COVID-19 pandemic. In this paper, we show a conditional generative adversarial network [34], usually used for image generation, can be adapted to model the complex dynamics among newly confirmed cases, policies, and human mobility. Based on Figure 2, we find that almost all the cities experience an abrupt decline to a relatively slow revival phase, resonating with the COVID-19 related policies and confirmed cases. Surprisingly, the history of human mobility provides less information to estimate the unprecedented and fast-changing human mobility during COVID-19 (see Supplementary Material, extended Figure 3). Thus, instead of modeling the long-term dependencies, we focus on predicting the human mobility transition between adjacent time slots shaped by the potential COVID-19 statistics, policies, and the latest dynamic of human mobility. To be specific, we propose a policy-human mobility interplay network (PHMIN) to estimate the human mobility changes based on conditional generative adversarial network. Here, the conditions are mainly derived from the COVID-19 statistics and policies, potentially influencing human mobility. Compared with traditional prediction methods, our model can flexibly learn wide-ranging human mobility dynamics and accurately extend to multi-waves of epidemics across cities.

MATERIALS AND METHODS

Dataset and preprocessing

In this study, one year and three months anonymized human mobility data from three cities are utilized, ranging from January 1, 2020 to March 31, 2021, including Beijing (BJ), Shijiazhuang (SJZ), and Dalian (DL), China. This de-identified dataset is collected in a crowdsourced way from over 10 million smart devices opt-in uploading location reports, processed by the Westlake Institute for Data Intelligence for research purposes conforming to the Personal Information Security Specification of China (2019). Considering the privacy issues, we further aggregate the data into gridded heatmaps to represent the regional human mobility level.

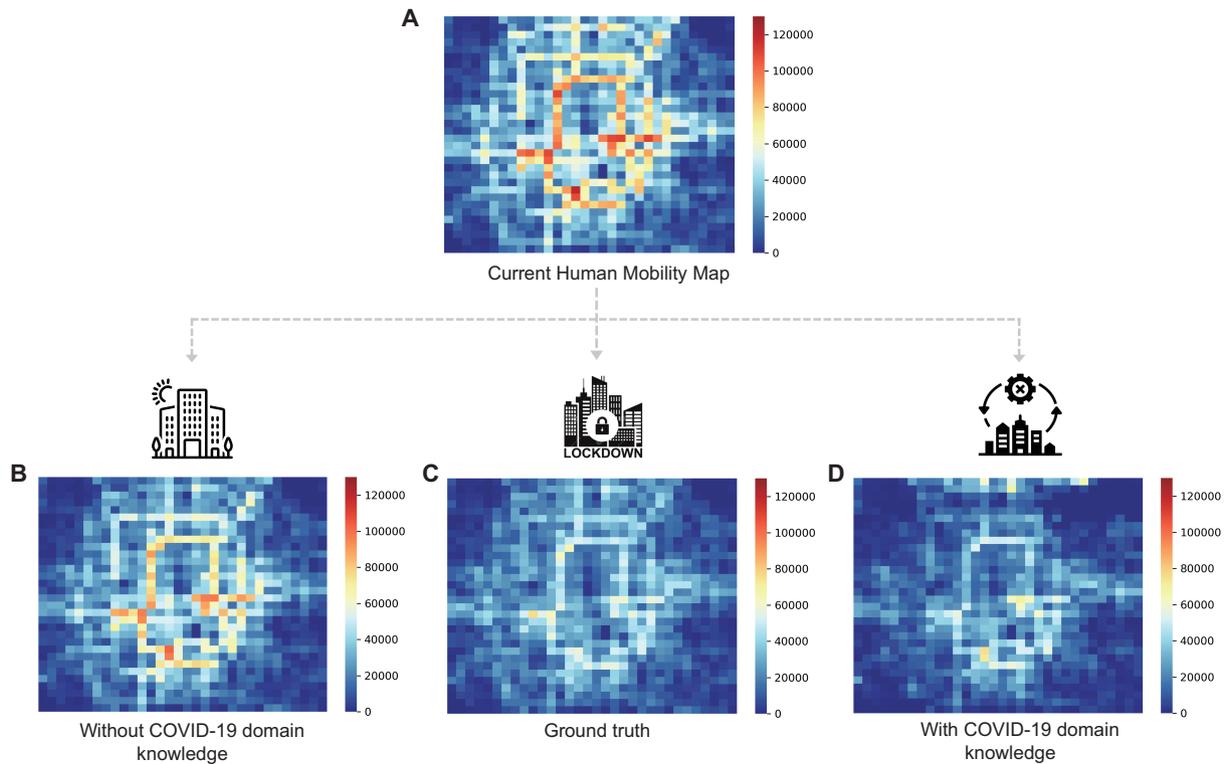


Figure 3 The role of domain knowledge including COVID-19 policies and statistics in BJ1. (A) BJ1 before lockdown; (B) estimated BJ1 after lockdown without domain knowledge; (C) real BJ1 after lockdown; (D) estimated BJ1 after lockdown with domain knowledge.

Moreover, we specify the duration and intensity of each wave of COVID-19 for different cities in Figure 2 according to the official statements. The intensity of the epidemic is divided into two levels, depending on the degree of mobility restrictions, marked with blue and green parts in Figure 2, respectively. The major difference between high and low-intensity policies is whether the mobility curves reveal the weekly periods. High-intensity policies such as strict lockdown usually lead to a constant low mobility level. In contrast, median policies just result in a small drop without covering up weekly periods (i.e., normal life rhythm). From Figure 2, we can observe the similarities and variations across cities, which are highly correlated with the epidemiological situations. Before formally formulating our problem, we provide several critical definitions.

- **Region:** The city is divided into $H \times W$ grids with equal size (900 m \times 600 m according to the Geohash6 standards [35]). Each grid represents a region of the city.
- **Human mobility level:** We measure the regional human mobility level m via the total visits of PoIs (point of interests), including shopping malls, residential areas, and companies. Human mobility level can explain most local transmissions during COVID-19 based on existing research [24], which reveals that infections are more likely to happen in the PoIs with higher human mobility levels (i.e., population density). Therefore, each region is associated with a series of human mobility levels across the study period.
- **Human mobility map:** To explore the overall human mobility of the city, we further define the human mobility map as $\mathbf{M} \in \mathbb{R}^{H \times W}$ with each entry $m_{i,j}$ representing the human mobility level of the corresponding region.
- **COVID-19 statistics:** Based on the results in Figure 2, we identify the most direct and informative

index, confirmed cases \mathbf{C} as the representative statistics.

- COVID-19 policies: Many countries have taken various policies to control the COVID-19 pandemic, especially on restricting human mobility. Here we refer to the Oxford COVID-19 Government Response Tracker project [25], which collected the strength of daily policies in China with several indicators, including containment and closure policies, economic policies, and health system policies. Specifically, we focus on the policies tightly related to human mobility, such as school and workplace closing, restrictions on gatherings and internal movement, stay at home requirements, denoted as \mathbf{P} .

For a specific city, given historical and current human mobility maps $\{\mathbf{M}_{t-1}, \mathbf{M}_t\}$ and corresponding COVID-19 statistics $\{\mathbf{C}_{t-1}, \mathbf{C}_t, \mathbf{C}_{t+1}\}$ (plenty of articles have achieved the accurate real-time forecasts on newly confirmed cases [36–38]), and policies $\{\mathbf{P}_{t-1}, \mathbf{P}_t, \mathbf{P}_{t+1}\}$ (can be made by policymakers in advance), we try to estimate the future human mobility map \mathbf{M}_{t+1} .

The workflow of PHMIN

As shown in Figure 4, PHMIN is designed based on the conditional generative adversarial network, which mainly consists of two components: (1) a generator including a transformer-based encoder and domain knowledge aware decoder (i.e., COVID-19 statistics and policies); (2) a discriminator to distinguish the human mobility dynamics during the COVID-19 and improve the model training.

Generator. The human mobility maps are essentially highly correlated on both spatial and temporal domains. However, there is neither long-term data nor stable patterns to learn from during the COVID-19 progression in our problem. Therefore, we model the human mobility changes in an incremental way shown in Figure 4A. The input of the generator is the calculated difference between two continuous human mobility maps $\Delta\mathbf{M}_{t-1} = \mathbf{M}_t - \mathbf{M}_{t-1}$, indicating the latest temporal trends on human mobility. On the other hand, the spatial features are more complicated due to the multi-scale dynamics we have mentioned in the Introduction. Therefore, we first introduce the convolution layer and transformer encoder inspired by [39] to extract the local and long-range correlation between regions, respectively. Specifically, the transformer encoder is composed of a multi-head self-attention module followed by an MLP layer, where both parts are equipped with layer-normalization and skip-connections. The feature maps $\mathbf{f}_0 \in \mathbb{R}^{H \times W \times C}$ derived from the convolution layer are directly flattened into length- $H \times W$ sequence of C -dimension embedded tokens. Then transformer encoder can recursively calculate the correlations between each pair of tokens, which has stronger representation capability without locally inductive bias. The processed 1D sequence of tokens is reshaped back into the 2D feature map $\mathbf{f}_1 \in \mathbb{R}^{H \times W \times C}$, serving as the output of encoder in the generator. The encoded feature is then concatenated with the COVID-19 condition feature map to jointly estimate the $\Delta\widehat{\mathbf{M}}_t$ via the decoder.

Domain knowledge adaptation. Another critical component in PHMIN is the domain knowledge adaptation module, which aims to fuse COVID-19 related statistics and policies as conditions. However, the conditions here are not as explicit as the class labels when generating images. For example, when going through a wave of epidemics completely from decline to revival, we will encounter a pair of similar human mobility heatmaps with opposite trends. This phenomenon enables us to provide conditions with direction information to guide the evolution of human mobility. Similar to the generator’s input, the first-order difference of confirmed cases and policies can also provide rich information on the trends. For example, a

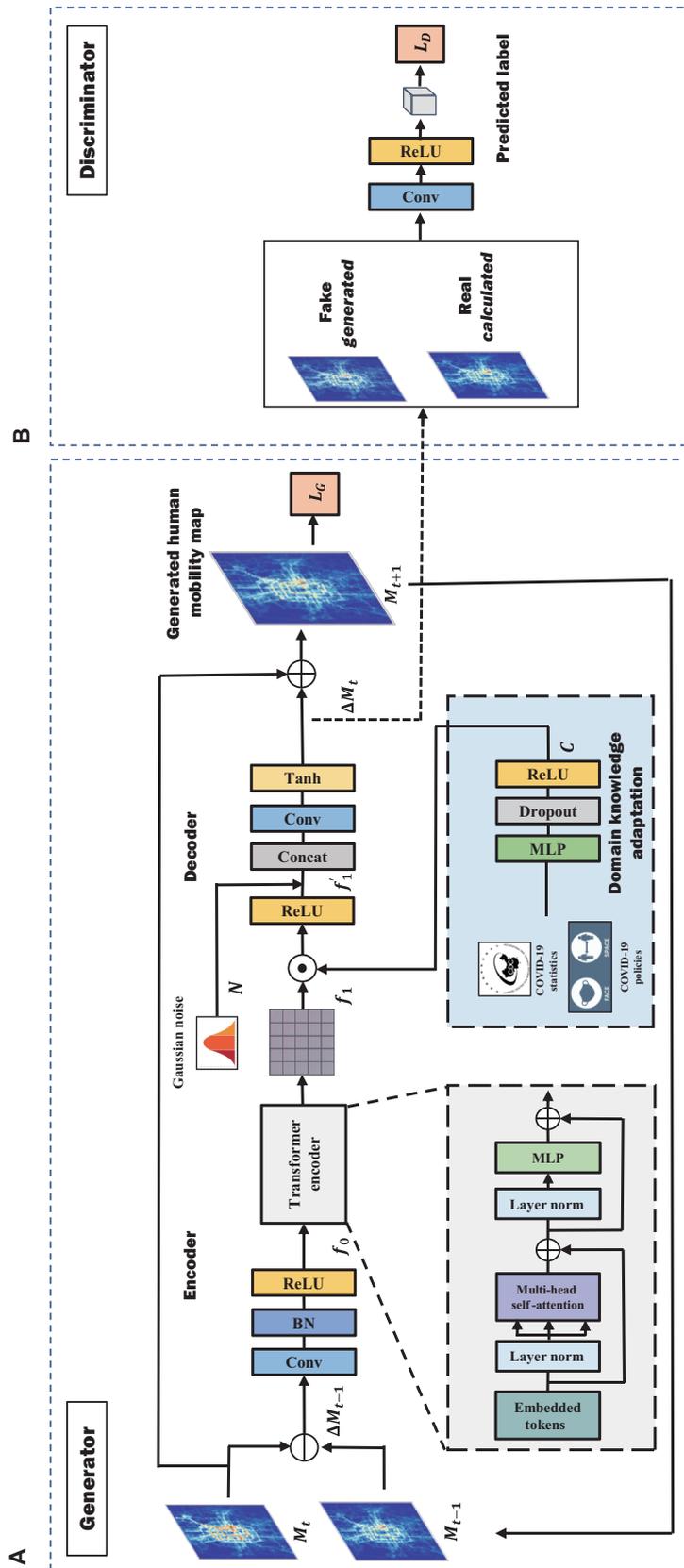


Figure 4 The framework of PHMIN. The framework mainly consists of two parts. (A) The generator for human mobility estimation. The original input contains the human mobility map of the last two time steps. The domain knowledge from COVID-19 statistics and policies is embedded and adapted to guide the estimation of future human mobility. (B) The discriminator is implemented to distinguish whether the human mobility dynamics is consistent with the period of COVID-19 or not.

sudden increase in the newly confirmed cases usually leads to more strict mobility restrictions. By contrast, a decline in the number of confirmed cases is likely to cause people to cease their vigilance along with more frequent mobility. When the confirmed cases return to zero and last for two weeks, the mobile restrictions will be removed in the corresponding city of China [40]. Therefore, the changes of newly confirmed cases or policies are more valuable, shaping people’s confidence towards the future and implicitly impacting people’s mobility intention. Furthermore, considering the different scales and population bases of cities, directly feeding the absolute value of the difference is not applicable. We introduce a dense network to turn it into latent vectors $\mathbf{C} \in \mathbb{R}^{H \times W \times c_0}$. Then we adopt a gating fusion as the activation function of spatial-temporal features, which uses \mathbf{C} to activate features f_1 as $f'_1 = \sigma(\mathbf{C} \otimes f_1)$. Considering the randomness and unknown factors during this process, we further add the noise vector $\mathbf{N} \in \mathbb{R}^{H \times W \times n_0}$ concatenated with the latent vector. This part serves as the domain knowledge driving the evolution of human mobility. Finally, the features derived from recent human mobility changes, COVID-19 statistics, and policies jointly estimate the future human mobility $\Delta \mathbf{M}_t$. A skip connection is constructed across the generator to produce the estimation for next time step $\mathbf{M}_{t+1} = \mathbf{M}_t + \widehat{\Delta \mathbf{M}}_t$. In this way, the complex dynamics of human mobility is modeled progressively.

Discriminator. Figure 4B illustrates the detailed structure of the discriminator. The inputs to the discriminator are human mobility changes rather than generated fluctuating human mobility maps, which makes our learning process more accessible and stable. Specifically, there are two types of inputs for the discriminator, i.e., generated human mobility changes with label fake and calculated real human mobility changes with label real. The discriminator is trained to predict the label of the human mobility map, indicating whether the generated human mobility map is consistent with real distribution.

Optimization. Our model is trained in an adversarial way, where the generator is trained to simulate the realistic human mobility response, and the discriminator aims to judge the correctness of the generated samples. The adversarial process essentially plays a min-max two-player game with a unique optimal solution when the generator can produce human mobility maps, effectively confusing the discriminator. Therefore, our loss function contains an estimation loss from the generator and an adversarial loss from the discriminator. Given the input human mobility maps $\mathbf{M}_t, \mathbf{M}_{t-1}$, we can get the pixel-wise mean square error (MSE) loss between estimated human mobility map $\widehat{\mathbf{M}}_{t+1} = G(\mathbf{M}_t, \mathbf{M}_{t-1}, \mathbf{C}, \mathbf{P}, \mathbf{N})$ and the ground truth \mathbf{M}_{t+1} . Furthermore, some suburbs and natural terrains (e.g., rivers, seas, and mountains) are desolate and uninhabited, whose samples are more like noise than real human mobility patterns. We further introduce a mask matrix $\mathbf{K} \in \mathbb{R}^{H \times W}$ to reduce the negative effects brought by noisy samples, whose elements are binary with one indicating regions with available data and zero for uncertain regions. Thus, we can calculate the masked loss for the generator:

$$L_G = \frac{1}{\|\mathbf{K}\|_{1,1}} \|\widehat{\mathbf{M}}_{t+1} - \mathbf{M}_{t+1}\|_{F,1}^2, \tag{1}$$

where \odot implies the element-wise multiplication for matrix. $\|\cdot\|_{1,1}$ and $\|\cdot\|_F$ stand for the $L_{1,1}$ and Frobenius norm, respectively. The mask matrix \mathbf{K} provides supplementary information associated with the real-world sampled data distribution. Many studies [41] have pointed out that adding a discriminator to distinguish the generated samples from the real can effectively improve the learning process by exerting a regularization as follows:

$$L_D = \mathbf{E}_{\mathbf{M} \sim p_{\text{data}}} \log D(\mathbf{M}_{t+1}, \mathbf{M}_t) + \mathbf{E}_{\mathbf{N} \sim p_N} \log(1 - D(G(\mathbf{M}_t, \mathbf{M}_{t-1}, \mathbf{C}, \mathbf{N}))). \tag{2}$$

Finally, combining the losses from both the generator and discriminator, the adversarial loss can be written as

$$\min_G \max_D L_G + \lambda L_D, \quad (3)$$

where the parameter λ is the weight that controls the trade-off between two objectives. We seek the saddle point by alternatively updating the parameters of the generator and the discriminator with Adam optimizer [42]. Besides, the discriminator classifier will not work during the test phase once the model is well-trained.

Experimental settings

The whole dataset is divided into seven sub-datasets based on the specific city-wave combination, i.e., BJ-1, BJ-2, DL-1, DL-2, DL-3, SJZ-1, SJZ-2. Naturally, we can take a pair of sub-datasets as the training and test set (i.e., source and target dataset from the point of model generalization), where the time of the training set should be earlier or simultaneous compared with the test set time considering the practical constraints.

Metrics. Considering both absolute and relative deviations of the estimation, two metrics are taken for comprehensive model evaluations, including mean absolute error (MAE) and mean absolute percentage error (MAPE). Moreover, the mask matrix \mathbf{K} is also introduced to handle the data missing and uncertainty. Therefore, the metrics are modified as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{z} \sum_{i=1}^z \frac{1}{\|\mathbf{K}\|_{1,1}} \times \|\widehat{\mathbf{M}}_i^f - \mathbf{M}_i^f\| \odot \mathbf{K} \|_{1,1}, \\ \text{MAPE} &= \frac{1}{z} \sum_{i=1}^z \frac{1}{\|\mathbf{K}\|_{1,1}} \times \|\widehat{\mathbf{M}}_i^f - \mathbf{M}_i^f\| \oslash \mathbf{M}_i^f \odot \mathbf{K}_i^f \|_{1,1}, \end{aligned} \quad (4)$$

where z is the number of samples. $\|\cdot\|_{1,1}$ and $\|\cdot\|_F$ imply the $L_{1,1}$ and Frobenius norm, respectively. \odot and \oslash stand for the element-wise multiplication and division for the matrix. Generally speaking, smaller values of metrics suggest a better performance.

Baselines. We first introduce three baseline methods for comparison as follows:

- Historical average (HA): A heuristic method simply taking the average of the same time interval. In this article, we calculate the average of the latest two time steps.
- ARIMA [43]: A classical autoregressive integrated moving average model for the time series forecast.
- Conditional GAN (cGAN) [34]: A GAN-based model directly combines the condition with noise without considering the difference between time steps.
- DeepST [33]: A deep learning based spatial-temporal prediction model.

Model variants. To explore the effects of designed modules, we further conduct experiments on the following methods:

- PHMIN: The proposed model.
- PHMIN-rm: Remove the proposed mask matrix from our model.
- PHMIN-rpc: Remove both the policies and COVID-19 statistics conditions from our model.
- PHMIN-rp: Remove both the policies conditions from our model.

RESULTS AND DISCUSSION

Multiple experimental settings indicate the generalization ability of our model

Considering there are three cities with seven waves of COVID-19 and two types of intensity, we divide the experiments into three groups to analyze the results hierarchically under (1) inter-city similar intensity; (2) inter-city different intensity; (3) intra-city similar/different intensity. These experiments emphasize a fundamental question: whether our model is generalizable in different contexts. We will answer this question in detail in the following. The bold results in the following tables indicate the best performance under different experimental settings.

Inter-city similar intensity. Based on the results shown in Table 1, we have several observations as follows.

- Generally speaking, our model achieves the best performance compared with baselines, which proves the excellent generalization ability of our model across cities. Specifically, the relatively simple HA and ARIMA models show better performance than DeepST and GAN, to some extent implying that the most valuable information in our problem is associated with the latest situations on human mobility. Furthermore, the conditional GAN slightly outperforms the DeepST due to its advantage in modeling unknown factors of human mobility in a generative manner. DeepST, on the other hand, can provide a deterministic prediction based on rich history observations, which is not suitable for our scenario.

- Given the same test city, the model derived from different training cities results in diverse performance. When testing on DL1, SJZ1, and SJZ2, the model trained based on the BJ1 shows better results than other cities. This can be attributed to the best data quality from BJ, which has more available samples with a high population density.

- Geographical similarity potentially influences the effects of model generalization across cities. The model trained based on the DL1 usually leads to worse performance. Take the results on DL1-BJ1, DL1-SJZ1, and DL1-SJZ2, for example, their results are close to ARIMA (with better MAEs while worse MAPEs). In fact, DL is surrounded by the ocean leading to irregular urban areas, while the BJ and SJZ are very regular with checkerboard layouts. The gaps in spatial distribution are likely to make the model generalization

Table 1 Inter-city similar intensity experiments

Model setting	HA		ARIMA		cGAN		DeepST		PHMIN	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
BJ1-DL1	1260	0.111	1013	0.086	2639	0.181	3773	0.255	629	0.058
SJZ1-DL1	1260	0.111	1013	0.086	1782	0.170	2617	0.187	648	0.062
DL1-BJ1	10305	0.369	6558	0.177	38316	1.337	32221	1.015	5803	0.222
SJZ1-BJ1	10305	0.369	6558	0.177	22947	0.752	15752	0.501	5412	0.177
BJ1-SJZ1	1824	0.201	1138	0.115	2413	0.238	3704	0.309	842	0.091
DL1-SJZ1	1824	0.201	1138	0.115	3368	0.497	3745	0.442	876	0.128
DL1-SJZ2	2221	0.368	1599	0.263	3760	0.805	3622	0.578	1349	0.273
BJ1-SJZ2	2221	0.368	1599	0.263	3012	0.371	3368	0.343	1876	0.253
BJ2-DL2	1136	0.063	1090	0.064	1397	0.111	1623	0.146	554	0.038
BJ2-DL3	1366	0.093	1012	0.077	1439	0.125	1924	0.158	947	0.062

harder.

Inter-city different intensity. To investigate the performance between epidemic waves with different intensity, i.e., high to median and median to high, we further conduct experiments shown in Table 2 demonstrating the following.

- In most instances, our model still performs better than the baselines. However, for the same test city, the overall performance is worse than the model obtained from the similar intensity in Table 1. This illustrates that compared with the diverse urban contexts, the intensity of the epidemic is a more critical indicator affecting the generalization ability of our model.

- The model trained from DL still performs poorly on other cities (e.g., DL1-BJ2 and DL2-SJZ2), which again illustrates the importance of geographical similarity. Additionally, the double gaps in both geography and intensity lead to even worse results than the baseline ARIMA under DL2-SJZ2.

Intra-city similar/different intensity. Intuitively, the intra-city experiments, i.e., multi-waves COVID-19 within the same city, should share more common patterns. However, the results in Table 3 reveal more insights on this process.

- Our model outperforms the baselines in all intra-city experiments. Besides, three groups of experiments, including SJZ1-SJZ2, DL2-DL3, and BJ1-BJ2, reach their best performance respectively compared with results in Tables 1 and 2, which implies the intra-city gaps might be smaller than previous conditions.

- The exception occurs in DL1-DL2 and DL1-DL3, whose results are even worse than BJ2-DL2 and BJ2-DL3, respectively. This can be attributed to BJ2, DL2, and DL3, although in different cities, they share similar intensity. Comparing results between DL2-DL3 and DL1-DL3, we deduce that our model generalizes better between similar intensity within the same city.

Table 2 Inter-city different intensity experiments

Model setting	HA		ARIMA		cGAN		DeepST		PHMIN	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
BJ1-DL2	1136	0.064	1090	0.064	3923	0.235	5131	0.301	777	0.048
SJZ1-DL2	1136	0.064	1090	0.064	2374	0.157	3707	0.225	805	0.051
BJ1-DL3	1366	0.093	1012	0.077	3478	0.205	5619	0.309	981	0.066
SJZ1-DL3	1366	0.093	1012	0.077	1958	0.151	5724	0.305	1052	0.070
DL1-BJ2	7373	0.126	9411	0.125	20808	0.437	21841	0.392	4503	0.086
SJZ1-BJ2	7373	0.126	9411	0.125	15599	0.283	16008	0.287	4678	0.087
BJ2-SJZ2	2221	0.369	1599	0.263	3423	0.631	2628	0.408	1566	0.242
DL2-SJZ2	2221	0.369	1599	0.263	5332	1.057	3926	0.596	2046	0.348

Table 3 Intra-city similar/different intensity experiments

Model setting	HA		ARIMA		cGAN		DeepST		PHMIN	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
SJZ1-SJZ2	2221	0.369	1599	0.263	2442	0.395	2110	0.304	1457	0.232
DL2-DL3	1366	0.093	1012	0.077	2066	0.189	3533	0.201	695	0.051
BJ1-BJ2	7373	0.126	9411	0.125	20047	0.324	30132	0.496	3799	0.065
DL1-DL2	1136	0.064	1090	0.064	2140	0.135	2706	0.157	714	0.042
DL1-DL3	1366	0.093	1012	0.077	1676	0.133	1752	0.140	1050	0.067

Ablation study on model variants

In this part, we aim to investigate the effects of modules via ablation studies. Specifically, we focus on the domain knowledge adaptation module by gradually removing the designed modules from the original models, i.e., mask matrix, knowledge from policy, or confirmed cases. We aggregate the experiments under different model variants shown in Table 4, from which we have several observations as follows.

- The model PHMIN-rcp shows worse performance than PHMIN-rp and PHMIN, which illustrates that the change of policies and confirmed cases provide useful prior knowledge to enhance the performance. More specifically, for the settings with similar intensity (i.e., from BJ1-DL1 to BJ2-DL3, SJZ1-SJZ2, and DL2-DL3), PHMIN-rp slightly outperforms PHMIN in some experiments. Removing the mask matrix leads to worse results under all the circumstances, which illustrates its effectiveness in mitigating the negative effects brought by regions with low-quality samples.
- The best results under each setting are achieved by either PHMIN-rp or PHMIN, implying the policy information is not always beneficial. Especially in the first wave of COVID-19, all the cities implement stringent policies, resulting in a comparable performance without policy information. Nevertheless, when it comes to different intensities, PHMIN shows an overwhelming advantage over PHMIN-rp in almost all

Table 4 Aggregated experiments on model variants

Model setting	ARIMA		PHMIN-rm		PHMIN-rcp		PHMIN-rp		PHMIN	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
BJ1-DL1	1013	0.086	899	0.076	864	0.073	523	0.050	629	0.058
SJZ1-DL1	1013	0.086	875	0.075	844	0.071	572	0.053	648	0.062
DL1-BJ1	6558	0.177	7903	0.246	7009	0.219	5776	0.213	5803	0.222
SJZ1-BJ1	6558	0.177	9005	0.279	7049	0.223	5487	0.196	5412	0.177
BJ1-SJZ1	1138	0.115	1385	0.143	1166	0.124	790	0.087	842	0.091
DL1-SJZ1	1138	0.115	1268	0.131	1159	0.119	938	0.117	876	0.128
DL1-SJZ2	1599	0.263	1719	0.271	1652	0.276	1236	0.225	1439	0.273
BJ1-SJZ2	1599	0.263	1714	0.289	1744	0.278	1080	0.155	1876	0.253
BJ2-DL2	1090	0.064	1181	0.064	825	0.049	689	0.043	554	0.038
BJ2-DL3	1012	0.077	1019	0.066	1021	0.068	857	0.060	947	0.062
BJ1-DL2	1090	0.064	894	0.052	897	0.053	710	0.045	777	0.048
SJZ1-DL2	1090	0.064	890	0.052	900	0.052	853	0.056	805	0.051
BJ1-DL3	1012	0.077	1204	0.077	1074	0.072	1024	0.073	981	0.066
SJZ1-DL3	1012	0.077	1101	0.071	1171	0.074	1267	0.087	1052	0.070
DL1-BJ2	9411	0.125	4670	0.069	5252	0.093	9346	0.177	4349	0.066
SJZ1-BJ2	9411	0.125	5908	0.098	5077	0.090	6526	0.118	4678	0.087
BJ2-SJZ2	1599	0.263	1637	0.246	1542	0.255	2019	0.334	1566	0.242
DL2-SJZ2	1599	0.263	2481	0.346	1598	0.268	2007	0.331	2046	0.348
SJZ1-SJZ2	1599	0.263	1969	0.295	1590	0.248	1083	0.155	1457	0.232
DL2-DL3	1012	0.077	1038	0.067	1022	0.068	783	0.061	695	0.051
BJ1-BJ2	9411	0.125	8546	0.138	5574	0.098	5843	0.103	3799	0.065
DL1-DL2	1090	0.064	959	0.053	824	0.048	887	0.050	714	0.042
DL1-DL3	1012	0.077	1014	0.067	1036	0.068	1179	0.077	1050	0.067

the cases (i.e., from BJ1-DL2 to DL2-SJZ2, from BJ1-BJ2 to DL1-DL3). These results can be explained by the pattern of confirmed cases that are very distinguished with different intensities, making the model generalization solely depending on confirmed cases (i.e., PHMIN-rp) unreliable. On the other side, policies are more universal across multi-waves of COVID-19, considering that the Chinese government will revert the policy to a high or very high level even just local transmissions are identified.

- We further visualize the human mobility changes with/without domain knowledge in BJ1 to obtain more insights. Figure 5A is the human mobility map before lockdown policy in the first wave of COVID-19 in Beijing. After adopting the policy, the human mobility map of BJ turns into Figure 5C, which is close to the estimation of our model in Figure 5D. If there is no COVID-19 related knowledge, the estimated human mobility will hold a relatively high level like Figure 5A. These results prove the effectiveness of COVID-19 knowledge adaptation.

Analysis of single-step and multi-step estimation

The experiments above illustrate the overall performances of our model, generalizing across cities and multiple waves of epidemics. In practice, we also care about how accurately our model can approximate the real data during COVID-19, i.e., prediction of each time step and the general trends from decline to revival. Therefore, we further conduct multiple experiments and visualize the generated results for three cities in Figure 5. We average the mobility of generated results to better reflect the tendency during each wave of COVID-19, where the shaded regions denote the variance over sampled parameters and stochastic realizations for each time step. Furthermore, our model can also recurrently generate multi-step estimations by taking the output of the model as the input for predicting the next time step. In practice, we are interested in whether multi-step estimation can predict long-term trends while mitigating accumulated errors. The multi-step estimations are also shown in Figure 5, from which we notice the following.

- In general, the results of single-step estimations are very close to the real trends of human mobility patterns, which accurately predict the drop and revival point during a wave of the epidemic. Multi-step estimations can still depict the trends of COVID-19, although they perform worse than single-step estimations on accuracy due to the accumulated errors. This proves that our model can handle the scenario of multi-step prediction.

- The errors of each time step are not uniform, where the decline stage of each wave is usually associated with higher errors than the revival stage. This can be attributed to fewer samples during the decline stage compared with relatively longer revival stages, making it difficult to learn stable patterns.

- There are still some drawbacks of the generated results. The transitions at the bottom of some curves are not as smooth as the real-world process, such as the BJ1 and BJ2 in Figures 5A and 5B. This might be that the transition stage in BJ is short and drastic with a large population base, resulting in a larger momentum of mobility change compared with other cities.

Case study on human mobility in the U.S.

To further verify the effectiveness of the proposed model, we also conduct additional experiments on the U.S. datasets. We obtain the human mobility datasets from [44], where we select two major cities on the east coast of the U.S., i.e., Boston (BOS) and Philadelphia (PA). Specifically, Philadelphia is an inland city

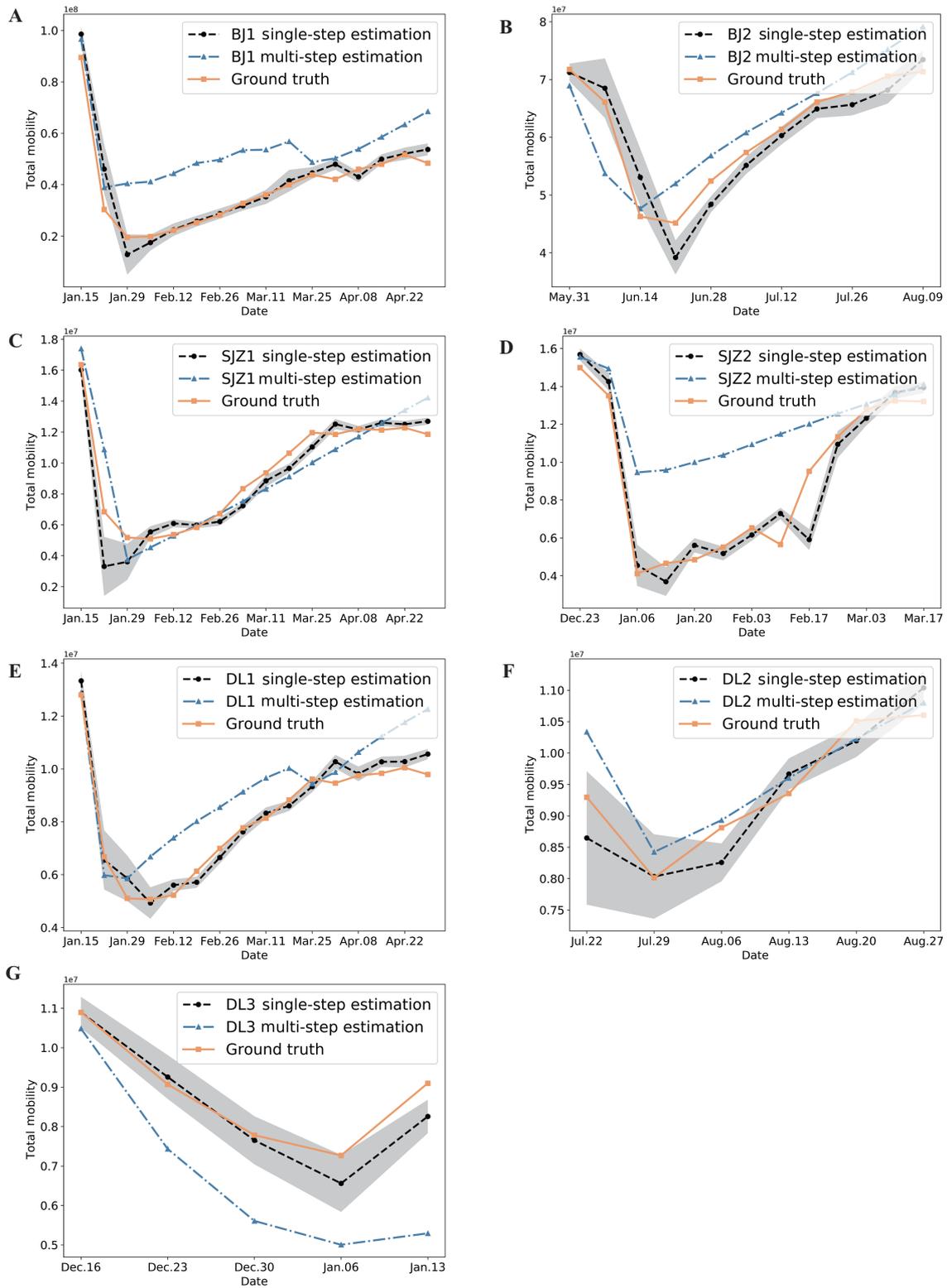


Figure 5 Estimated human mobility with single/multi-step settings across multiple cities. (A) BJ1, (B) BJ2, (C) SJZ1, (D) SJZ2, (E) DL1, (F) DL2, (G) DL3.

Table 5 Experiments on the U.S. cities

Model setting	HA		ARIMA		cGAN		DeepST		PHMIN	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
BOS-PA	2292	0.137	5226	0.303	19616	0.943	16626	0.892	2108	0.125
PA-BOS	1461	0.107	3166	0.203	4704	0.270	32835	1.545	1126	0.091

similar to Beijing or Shijiazhuang in China, while Boston is a seaside city similar to Dalian, China. The period ranges from March 2, 2020 to July 5, 2020. Moreover, considering that the crowd distribution in the U.S. is relatively sparser than in China, the aggregated human mobility map in the U.S. is associated with a bigger size (about 2.7 km×1.8 km). We further investigate the inter-city human mobility estimation between the two cities in Table 5. The results reveal that our model still outperforms all the baselines for the human mobility datasets from the U.S., which can be attributed to the similar human mobility patterns in the early stage of the COVID-19. The naive heuristic method HA also achieves a better performance than other baselines but is not as good as ours, which illustrates that the historical average is a good indicator for human mobility during COVID-19 with seemingly unusual patterns.

CONCLUSION

Human behaviors, especially spatial-temporal mobility patterns, have become the critical determinant of COVID-19 emergence. With a detailed analysis of multiple epidemics in three cities, we identify the vital indicators shaping human mobility, i.e., COVID-19 policies and statistics. Thus, we propose the model PHMIN to estimate human mobility via modeling the influence of COVID-19 policies and statistics. Our model shows excellent generalization ability in both inter- and intra-city experiments. The results reveal that the changes in human mobility share universal patterns during epidemics. Furthermore, our model can also recurrently produce multi-step estimations, allowing the governments to evaluate the influence of different policies. They are helpful in managing and controlling mobility during a pandemic or other disaster events.

Acknowledgements

The authors thank Baosheng Yu, Yibing Zhan, and our reviewers for valuable discussions and feedbacks.

Funding

This work was supported in part by the Key Research and Development Program of Zhejiang Province (2021C03037) and the National Natural Science Foundation of China (72061147006).

Author contributions

J.C. and S.H. conceived the study. K.L., Y.X., and C.L. performed the computational analysis. D.T. and F.H. analyzed the results. Y.F. provided and analyzed the human mobility data. K.L. wrote the manuscript with C.L., S.H., and F.H. All authors reviewed and revised the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information

The supporting information is available online at <https://doi.org/10.1360/nso/20220003>. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

- 1 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; **20**: 533–534.
- 2 Kupferschmidt K, Wadman M. Delta variant triggers new phase in the pandemic. *Science* 2021; **372**: 1375–1376.
- 3 Del Rio C, Malani PN, Omer SB. Confronting the Delta variant of SARS-CoV-2, summer 2021. *JAMA* 2021; **326**: 1001.
- 4 Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Phys Rep* 2021; **913**: 1–52.
- 5 Xu S, Li Y. Beware of the second wave of COVID-19. *Lancet* 2020; **395**: 1321–1322.
- 6 Patel J, Sridhar D. We should learn from the Asia-Pacific responses to COVID-19. *Lancet Regional Health-Western Pac* 2020; **5**: 100062.
- 7 Haug N, Geyrhofer L, Londei A, *et al.* Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav* 2020; **4**: 1303–1312.
- 8 Li Y, Campbell H, Kulkarni D, *et al.* The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: A modelling study across 131 countries. *Lancet Infect Dis* 2021; **21**: 193–202.
- 9 Bonaccorsi G, Pierri F, Cinelli M, *et al.* Economic and social consequences of human mobility restrictions under COVID-19. *Proc Natl Acad Sci USA* 2020; **117**: 15530–15535.
- 10 Goolsbee A, Syverson C. Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *J Public Economics* 2020; **193**: 104311.
- 11 Yan L, Zhang HT, Goncalves J, *et al.* An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell* 2020; **2**: 283–288.
- 12 Cairney P. The UK government’s COVID-19 policy: What does “guided by the science” mean in practice? *Front Political Sci* 2021; **3**: 11.
- 13 Kraemer MUG, Yang CH, Gutierrez B, *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020; **368**: 493–497.
- 14 Hadjidemetriou GM, Sasidharan M, Kouyialis G, *et al.* The impact of government measures and human mobility trend on COVID-19 related deaths in the UK. *Transpation Res Interdisc Perspect* 2020; **6**: 100167.
- 15 Gatto M, Bertuzzo E, Mari L, *et al.* Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc Natl Acad Sci USA* 2020; **117**: 10484–10491.
- 16 Chetty R, Friedman JN, Hendren N, *et al.* The economic impacts of COVID-19: Evidence from a new public database built using private sector data. Technical Report, National Bureau of Economic Research, 2020
- 17 Martin A, Markhvida M, Hallegatte S, *et al.* Socio-economic impacts of COVID-19 on household consumption and poverty. *Econ Dis Cli Cha* 2020; **4**: 453–479.
- 18 Phillips NE, Levy BL, Sampson RJ, *et al.* The social integration of American cities: Network measures of connectedness based on everyday mobility across neighborhoods. *Sociol Method Res* 2021; **50**: 1110–1149
- 19 Thomas LJ, Huang P, Yin F, *et al.* Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. *Proc Natl Acad Sci USA* 2020; **117**: 24180–24187.
- 20 Okraszewska R, Romanowska A, Wołek M, *et al.* Integration of a multilevel transport system model into sustainable urban mobility planning. *Sustainability* 2018; **10**: 479.
- 21 Klein B, LaRock T, McCabe S, *et al.* Reshaping a nation: Mobility, commuting, and contact patterns during the COVID-19 outbreak. *Northeastern University-Network Science Institute Report* 2020

- 22 Schlosser F, Maier BF, Jack O, *et al.* COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc Natl Acad Sci USA* 2020; **117**: 32883–32890.
- 23 Deng H, Gao J, Wang Q. Network percolation reveals adaptive bridges of the mobility network response to COVID-19. ArXiv:2104.11405
- 24 Chang S, Pierson E, Koh PW, *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 2021; **589**: 82–87.
- 25 Hale T, Angrist N, Goldszmidt R, *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* 2021; **5**: 529–538.
- 26 Han ZY, Fu HH, Xu FL, *et al.* Who will survive and revive undergoing the epidemic: Analyses about poi visit behavior in Wuhan via check-in records. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021. **5**: 1–20
- 27 Chinazzi M, Davis JT, Ajelli M, *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 2020; **368**: 395–400.
- 28 Gozzi N, Tizzoni M, Chinazzi M, *et al.* Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nat Commun* 2021; **12**: 2429.
- 29 Lai S, Ruktanonchai NW, Zhou L, *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* 2020; **585**: 410–413.
- 30 Wellenius GA, Vispute S, Espinosa V, *et al.* Impacts of social distancing policies on mobility and COVID-19 case growth in the US. *Nat Commun* 2021; **12**: 3118.
- 31 Pei S, Yamana TK, Kandula S, *et al.* Burden and characteristics of COVID-19 in the United States during 2020. *Nature* 2021; **598**: 338–341.
- 32 Xie P, Li T, Liu J, *et al.* Urban flow prediction from spatiotemporal data using machine learning: A survey. *Inf Fusion* 2020; **59**: 1–12.
- 33 Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2017. 1655–1661
- 34 Mirza M, Osindero S. Conditional generative adversarial nets. ArXiv:1411.1784
- 35 Fox A, Eichelberger C, Hughes J, *et al.* Spatio-temporal indexing in non-relational distributed databases. In: *Proceedings of 2013 IEEE International Conference on Big Data*, 2013. 291–299
- 36 Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos Solitons Fractals* 2020; **135**: 109850.
- 37 Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for COVID-19 (SARS-CoV-2) pandemic: A review. *Chaos Solitons Fractals* 2020; **139**: 110059.
- 38 Cramer E Y, Ray E L, Lopez V K, *et al.* Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *medRxiv* 2021, doi: 10.1101/2021.02.03.21250974
- 39 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017. 6000–6010
- 40 Kupferschmidt K, Cohen J. Can China’s COVID-19 strategy work elsewhere? *Science* 2020; **367**: 1061–1062.
- 41 He F, Tao D. Recent advances in deep learning theory. ArXiv:2012.10931
- 42 Kingma DP, Ba J. Adam: A method for stochastic optimization. ArXiv:1412.6980
- 43 Box GE, Jenkins GM, Reinsel GC, *et al.* Time Series Analysis: Forecasting and Control. Hoboken: John Wiley & Sons, 2015
- 44 Kang Y, Gao S, Liang Y, *et al.* Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. *Sci Data* 2020; **7**: 390.