

Forecasting the Transmission Trends of Respiratory Infectious Diseases with an Exposure-Risk-Based Model at the Microscopic Level

Ziwei Cui ^{1,1,1}, Ming Cai ², Yao Xiao ², Zheng Zhu ², Mofeng Yang ², and Gongbo Chen ²

¹School of Intelligent System Engineering

²Affiliation not available

October 31, 2023

Abstract

Respiratory infectious diseases (e.g., COVID-19) have brought huge damages to human society, and the accurate prediction of their transmission trends is essential for both the health system and policymakers. Most related studies concentrate on epidemic trend forecasting at the macroscopic level, which ignores the microscopic social interactions among individuals. Meanwhile, current microscopic models are still not able to sufficiently decipher the individual-based spreading process and lack valid quantitative tests. To tackle these problems, we propose an exposure-risk-based model at the microscopic level, including 4 modules: individual movement, virion-laden droplet movement, individual exposure risk estimation, and prediction of new cases. First, the front two modules reproduce the movements of individuals and the droplets of infectors' expiratory activities. Then, the outputs are fed to the third module for estimating the personal exposure risk. Accordingly, the number of new cases is predicted in the final module. Our model outperforms 4 existing macroscopic or microscopic models through the forecast of new cases of COVID-19 in the United States. Specifically, mean absolute error, root mean square error and mean absolute percentage error by our model are 2454.70, 3170.51, and 3.38% smaller than the minimum results of comparison models, respectively. In sum, the proposed model successfully describes the scenarios from a microscopic perspective and shows great potential for predicting the transmission trends with different scenarios and management policies.

Forecasting the Transmission Trends of Respiratory Infectious Diseases with an Exposure-Risk-Based Model at the Microscopic Level

Ziwei Cui^a, Ming Cai^a, Yao Xiao^{a,*}, Zheng Zhu^b, Mofeng Yang^c and Gongbo Chen^d

^aSchool of Intelligent System Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, Guangdong, China

^bCollege of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang, China

^cMaryland Transportation Institute, Department of Civil and Environmental Engineering, University of Maryland at College Park, Maryland, USA

^dGuangdong Provincial Engineering Technology Research Center of Environmental and Health Risk Assessment; Department of Occupational and Environmental Health, School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong, China

ARTICLE INFO

Keywords:

Environmental Epidemiology
Respiratory Infectious Diseases
COVID-19
Microscopic Model
Exposure Risk
Public Health

ABSTRACT

Respiratory infectious diseases (e.g., COVID-19) have brought huge damages to human society, and the accurate prediction of their transmission trends is essential for both the health system and policymakers. Most related studies focus on epidemic trend forecasting at the macroscopic level, which ignores the microscopic social interactions among individuals. Meanwhile, current microscopic models are still not able to sufficiently decipher the individual-based spreading process and lack valid quantitative tests. To tackle these problems, we propose an exposure-risk-based model at the microscopic level, including 4 modules: individual movement, virion-laden droplet movement, individual exposure risk estimation, and prediction of transmission trends. Firstly, the front two modules reproduce the movements of individuals and the droplets of infectors' expiratory activities, respectively. Then, the outputs are fed to the third module to estimate the personal exposure risk. Finally, the number of new cases is predicted in the final module. By predicting the new COVID-19 cases in the United States, the performances of our model and 4 other existing macroscopic or microscopic models are compared. Specifically, the mean absolute error, root mean square error, and mean absolute percentage error provided by the proposed model are respectively 2,454.70, 3,170.51, and 3.38% smaller than the minimum results of comparison models. The quantitative results reveal that our model can accurately predict the transmission trends from a microscopic perspective, and it can benefit the further investigation of many microscopic disease transmission factors (e.g., non-walkable areas and facility layouts).

1. Introduction

The unexpected outbreak and rapid spread of Corona Virus Disease 2019 (COVID-19) have damaged the world tremendously, and brought a profound influence on various fields such as medicine, industry, energy (Anand et al., 2021; Bontempi et al., 2020; Sharifi et al., 2021; Wang et al., 2020). At present, a growing number of countries have moved into a post-pandemic phase, i.e., the overall spread of COVID-19 has been controlled, but intermittent small-scale outbreaks still occur (Jin et al., 2021). COVID-19 is a respiratory infectious disease (RID) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Zhang et al., 2022a). RID poses a huge threat to the population and public health, and recent studies have found that the loss of life expectancy due to RID stood at 1.29 years in 2017 globally (Huang and Guo, 2021). To live with the ongoing challenges by COVID-19 or the re-emergence of other respiratory viral infections, people must raise awareness and take measures for infection prevention and control in daily life (Coccia, 2021; Zhu and Zhu, 2021).

*Corresponding author

✉ cuiwz3@mail2.sysu.edu.cn (Z. Cui); caiming@mail.sysu.edu.cn (M. Cai); xiaoyao9@mail.sysu.edu.cn (Y. Xiao); zhuzheng89@zju.edu.cn (Z. Zhu); mofeng@umd.edu (M. Yang); chengb36@mail.sysu.edu.cn (G. Chen)
ORCID(s): 0000-0002-9438-7431 (Y. Xiao)

Various factors that affect the transmission of RIDs can be grouped into 6 main categories (see Table 1): virus-related factors, population characteristics, economic factors, scene factors, environmental and geographical factors, prevention and control measures. It is important to examine the impacts of these factors on transmission, which are beneficial for disease prevention and control. For example, non-pharmaceutical interventions (NPIs) (e.g., maintaining safe social distance (Flaxman et al., 2020; Lai et al., 2020); entry limitation policy (Xiao et al., 2021)), which are in the last category, play essential roles in stopping the spread of respiratory viral infections. NPIs can efficiently respond to emerging epidemics, and they are long-term approaches that interfere with people's social behavior (Duives et al., 2021; Perra, 2021). In a public place, it is necessary to find the most critical NPIs, which can effectively reduce the number of infected cases. Thus, to help evaluate factors' impacts, a model that can accurately describe diseases' transmission process is necessary (Ren et al., 2020).

Mathematical epidemiological models are helpful to decipher the complex transmission process of epidemics (Araya, 2021; Hosseini et al., 2020), which generally includes macroscopic level and microscopic level models (Xiao et al., 2021). At the macroscopic level, researchers focus on the infection and recovery process of disease among the population (Earn, 2008), and have proposed many compartmental models and computational intelligence methods:

Table 1

There are some factors that affect the transmission of RIDs.

Categories	Examples and References
Virus-Related Factors	Strains of the Virus (Abu-Hammad et al., 2020)
Population Characteristics	Age (Abu-Hammad et al., 2020); Genetic Makeup of Populations Blood Groups (Abu-Hammad et al., 2020)
Economic Factors	Gross Domestic Product (GDP) (Zhang et al., 2021); Commercial Trade (Bontempi, 2022; Bontempi and Coccia, 2021; Bontempi et al., 2021)
Scene Factors	Facility Layouts (Xiao et al., 2021); Inter-Provincial Travels (Ahmadi et al., 2021)
Environmental and Geographical Factors	Air Pollution (Conticini et al., 2020; Lembo et al., 2021; Urrutia-Pereira et al., 2020; Wu et al., 2021); Solidwaste Generated by Infected Individuals (Al Huraimel et al., 2020); Sunspot Numbers (Nasirpour et al., 2021)
Prevention and Control Measures	NPIs, e.g., Maintaining Safe Social Distance (Flaxman et al., 2020; Lai et al., 2020), Entry Limitation Policy (Xiao et al., 2021); Vaccines (Daniels et al., 2021; Verma et al., 2021); Statins and Other Medicines (Davoudi et al., 2021); Virus Detection Technologies (Li et al., 2021; Zhang et al., 2022b)

in the former, there are susceptible–infected–removed, or recovered (SIR) model (Kermack and McKendrick, 1927), the susceptible–infected–recovered–susceptible (SIRS) model (Hethcote, 1976), and their extended modifications (Black and McKane, 2010; Cross et al., 2007; Liu et al., 1987); in the latter, there are the Long Short-Term Memory (LSTM) method (Aragão et al., 2021), the hybrid intelligent approach based on fractal theory and fuzzy logic (Castillo and Melin, 2020), the multiple ensemble neural network model with fuzzy response aggregation (Melin et al., 2020), and others. Although the macroscopic models have established the research discipline of mathematical epidemiology, most of them need continuously updated data or large amount of data to obtain optimized system parameters, and they may ignore details in modeling the social interactions among individuals. Therefore, macroscopic approaches can be insensitive in evaluating NPIs or require strong assumptions to overcome the incapability. Fortunately, microscopic models address these limitations to some extent because they focus on the disease spreading between individuals. However, current microscopic models are still not able to sufficiently decipher the individual-based spreading process and lack valid quantitative tests.

In this study, an exposure-risk-based model at the microscopic level is developed, and the principal contributions can be described as follows.

- *First*, the movements of individuals as well as the dynamics of droplets are separately modeled and coupled to calculate the personal exposure risk. An integrated transmission process of the RID at the individual level is formulated.
- *Second*, a bridge between the macroscopic epidemic transmission data and our microscopic model at individual level is built. As a result, our model can be quantitatively calibrated and validated through the macroscopic data, such as the number of new cases.
- *Third*, based on the proposed model, the influences of factors, that affect the model input values or scenarios

(e.g., non-walkable areas and facility layouts), can be quantitatively evaluated.

The rest of the paper is organized as follows. Section 2 is the review of previous studies. Section 3 introduces our model, and Section 4 manifests applications of the model based on real-world data in the COVID-19 of the United States. Section 5 presents the comparison of different models. In Section 6, discussions and future perspectives are reported. Finally, we summarize this paper and present suggested directions for future work in Section 7.

2. Literature Review

There are some microscopic models forecasting the spreading trends of RIDs. For example, previous scholars have estimated individual-level mobility due to daily activities (such as work, study, or shopping) (Eubank et al., 2004). In these models, RIDs can transmit between individuals at one specific location and spread between locations due to mobility. It should be noted that the location mentioned here is a specific area like a school or shopping mall, not a physical location of the individual. Unlike these activity-based models, some studies focus on the transmission of respiratory viruses, and they have constructed the Well-Riley models and aerosol infectious dose-response models (Bazant and Bush, 2021; Krieger et al., 2020). These models assume that infectious particles are distributed homogeneously, which results in the same infection risks among individuals regardless of their physical distance from the infector. Recent studies have considered the spatiotemporal distribution of pathogens in the environment under different transmission routes, and the personal exposure risk is determined based on the duration and distance of infectious contacts (Arav et al., 2021; Gao et al., 2021). Nevertheless, transmission is modeled as a static event between infectors and susceptible individuals in the aforementioned models. Thus, they cannot describe the epidemic spreading process with time-varying individual physical distances during individual movements.

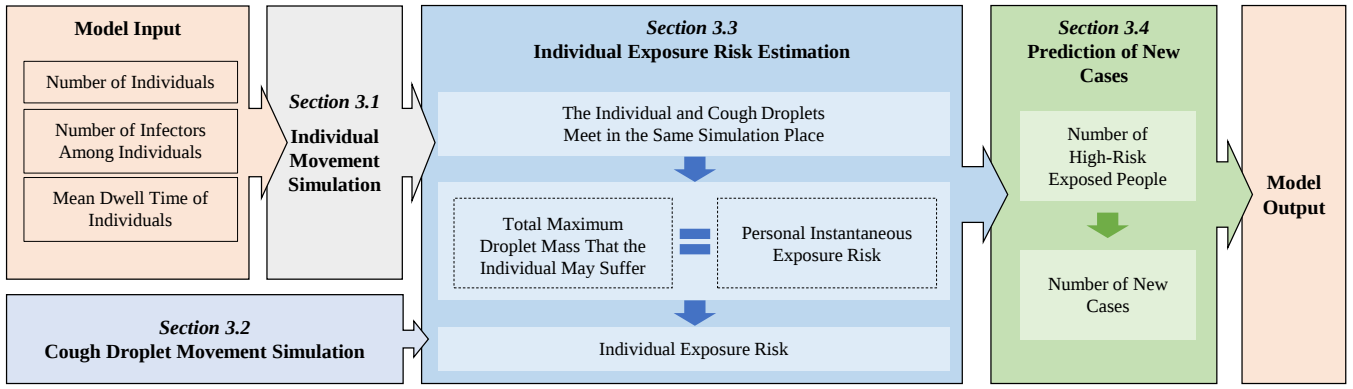


Fig. 1: The flowchart of our model.

Several studies have integrated pedestrian dynamics into epidemic spreading models to tackle the issue (Kim and Quaini, 2020; Ronchi and Lovreglio, 2020). Indeed, pedestrian dynamics are suitable for describing individual decisions and actions in mass gathering scenarios. Pedestrian-based epidemic spreading models are composed of the individual movement module (e.g., the social force model (Xiao et al., 2021), the nomad model (Duives et al., 2021), a social force model coupled with an Eikonal equation (Abdul Salam et al., 2021)) and the disease transmission module (e.g., the model based on the cut-off distance (Xiao et al., 2021), the QVEmod (Duives et al., 2021), the non-local SEIS contagion model (Abdul Salam et al., 2021)). The former module simulates the general individual crowd movement and outputs time series of personal positions. Based on the outputs from the first module, the second module evaluates the disease transmission risk from the infectors to susceptible individuals.

However, previous pedestrian-based epidemic spreading models ignored critical factors for simplification, e.g., the exposure risk is a fixed term when the individual is exposed to the infection risk area (Xiao et al., 2021), and they potentially overestimate or underestimate the number of high-risk people. Besides, only limited situations are analyzed, such as the cruise (Fang et al., 2020), the supermarket (Parisi et al., 2020), and the academic building (Romero et al., 2020). Moreover, different scenarios have various scales and geometries, e.g., tables, chairs, and other furnish and decorate in a restaurant (10 m × 9 m) (Duives et al., 2021). These variable settings affect the virus transmission risks and bring more model inputs and computation costs. On the other hand, a general public place can represent all scenes, e.g., an empty room of 10 m × 9 m represents a restaurant or store of the same scale without considering the indoor geometries. Meanwhile, most outputs of existing microscopic-level models are hard to verify due to the lack of actual data, such as the personal exposure risk and the number of high-risk exposed people. Then, these models are used directly or after the qualitative analysis but lack

quantitative tests. Hence, building a model for a general public place that can be validated and applied to all scenes is necessary and practical.

3. Model

To construct the microscopic level exposure-risk-based model, we should define the exposure risk of individuals. RIDs are transmitted through viral droplets produced by respiratory activities such as breathing, talking, coughing, and sneezing. Here, the typical symptom of most RIDs, i.e., coughing, is considered in our model. Thus, we define the instantaneous personal exposure risk as the maximal mass of virion-laden droplets produced by a typical cough exposed to infectors at a certain moment. Then, the personal exposure risk can be calculated by directly summing up instantaneous personal exposure risks during the dwell time.

The components of our model are shown in Fig. 1. There are three important parameters (i.e., number of individuals C^{total} , number of infectors among individuals C^{inf} , and mean dwell time of individuals T^{dwell}) as the input. Besides, the output includes not only the result of existing pedestrian-based spreading models (i.e., number of high-risk exposed people C^{risk}), but also the key outcome (i.e., number of new cases C^{new}). In addition, personal exposure risk is obtained as the intermediate output.

3.1. Individual Movement Simulation

In this part, the movements of individuals are modeled within a general public place. *First*, we simulate the space and individuals of the reality. *Second*, the pedestrian dynamic model (i.e., the social force model) to reproduce individuals' movements is introduced. *Finally*, time series of individual positions can be estimated for Section 3.3.

Herein, the simulation space is a general place without obstacles or other non-walkable areas, but it has boundaries like an empty indoor room. Whether individuals from model inputs are infected or not, their movement modelings are the same, as shown in Fig. 2, where each person is represented as a circle with a radius r^{ped} . Thus, when we model the motion

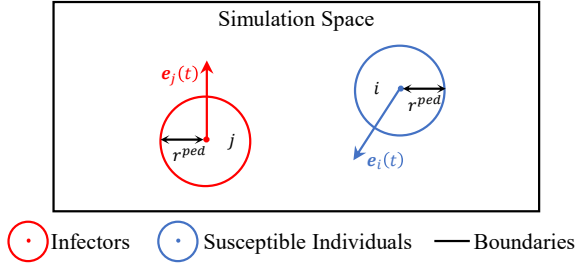


Fig. 2: The sketch map of the individuals moving in the simulation space.

of one person during the dwell time, the movements of all others can be determined in the same way.

The social force model is one of the most widely used microscopic models, and it is also the basic model to simulate pedestrian dynamics in commercial software (e.g., PTV-Vissim, MassMotion) (Bouchnita and Jebrane, 2020a,b). Hence, the social force model is applied here, and the movement of individual i (with a mass of m_i) at time t is driven by the resultant force $f_i(t)$ as

$$f_i(t) = f_i^{\text{drv}}(t) + \sum_{j \text{ near}} f_{i,j}^{\text{ped}}(t) + \sum_w f_{i,w}^{\text{obs}}(t), \quad (1)$$

where $f_i^{\text{drv}}(t)$ reflects the desire of individual i to maintain a certain walking speed v_i^0 towards a certain direction in a relaxation time τ_i ; $f_{i,j}^{\text{ped}}(t)$ is the interaction force between the objective individual i and the neighboring individual j ; $f_{i,w}^{\text{obs}}(t)$ means the instant interaction force between the objective individual i and the wall/obstacle w . More details about the social force model can be found in the literature (Chraïbi et al., 2011; Helbing and Molnar, 1995; Kretz et al., 2018).

As the motion of individual i is driven under Newton's second law with a second-order dynamics function, the velocity $v_i(t)$ and the unit direction vector $e_i(t)$ of individual i is determined in

$$\frac{dv_i(t)e_i(t)}{dt} = \frac{f_i(t)}{m_i}, \quad (2)$$

based on which the location $p_i(t)$ of individual i is finally obtained as

$$\frac{dp_i(t)}{dt} = v_i(t)e_i(t). \quad (3)$$

In addition, to simulate the crowd movements, a random walking process is set by adjusting the desired initial direction e_i^0 . Then, when individual i is close to hitting the boundary, the model will force the person to change the desired direction to any direction e_i^b that is away from the boundary.

3.2. Cough Droplet Movement Simulation

Virion-laden droplet transmission of a typical cough is modeled through Computational Fluid Dynamics (CFD)

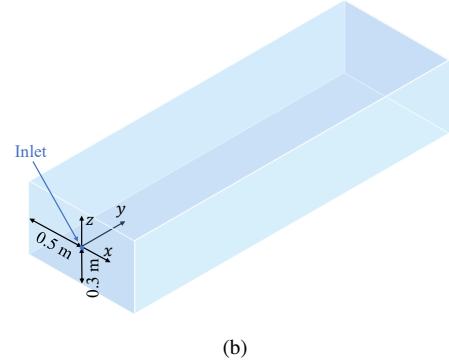
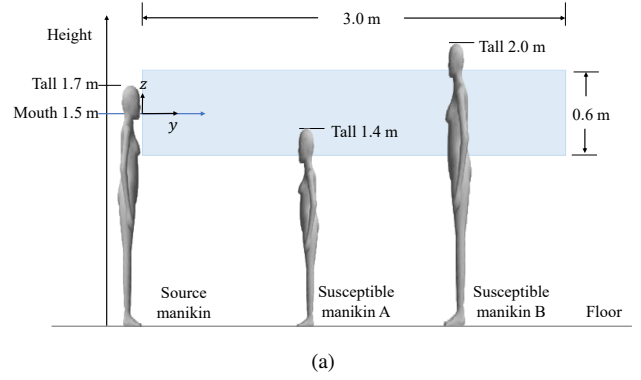


Fig. 3: Schematic diagram of (a) the computational domain with source and susceptible manikins, and (b) the numerical simulation computational domain.

simulations within a closed environment. *First*, the computational domain and grids are simulated. *Second*, we choose the appropriate methods and parameter settings from literatures for the numerical simulation. *Finally*, after completing the simulation with the commercial CFD solver ANSYS Fluent release 2020 R2, time series of positions and masses of each cough droplet can be determined for Section 3.3.

An infector is represented by a source manikin 1.70 m tall, representing an average-sized human as Fig. 3(a) (Liu et al., 2017). The mouth of the infector is 1.50 m from the ground, and the mouth outflow is roughly horizontal. For susceptible people whose height ranges from 1.40 m to 2.00 m, we take 1.20 m to 1.80 m as the breathing height area, as shown in the blue area of Fig. 3(a). The computational domain is in a cuboid shape and is illustrated by Fig. 3(b), in which “x”, “y”, and “z” axes represent the lateral, axial (streamwise), and vertical directions, respectively. The length of the computational domain in the “y” axis is set to be 3.00 m to consider the scope of a cough fully. Now, the blue area of Fig. 3(a) is the view of Fig. 3(b) along the “x” axis. The inlet is a square with an area of $3.70 \times 10^{-4} \text{ m}^2$, representing the human mouth (Borro et al., 2021). The center of the inlet is denoted by (0, 0, 0), which is also the origin of the coordinate system. The computational domain and grids are generated by using Gambit 2.4.6. The grid resolution is 0.01 m, and the total number of computational cells is approximately 1.80 million.

To accurately estimate the consequences of the coughing event, reliable modeling methods and settings of numerical simulation are important. The transmission medium of the cough is modeled as an incompressible ideal gas with constant properties calculated at ambient conditions. The flow field evolution of coughing is time-dependent, so the simulations are conducted under a transient condition. The gravitational acceleration is -9.81 m/s^2 along the “z” axis, and the energy equation is required. Since the droplet volume fraction is very low in the cough flow, the Eulerian-Lagrangian method is used in this study (Gupta et al., 2010; Zhu et al., 2006). According to previous study (Borro et al., 2021), a time step size of 0.01 s is used, with 10 sub-iterations. The total flow time of the transient simulation is 15.00 s, which is enough to investigate the dynamic characteristics of the droplets produced by coughing. More detailed settings are determined based on literature (Aliabadi et al., 2010; Bi, 2018; Chao et al., 2009; Duguid, 1946; Gao and Niu, 2006; Zhang and Li, 2012; Zhang et al., 2017).

The droplet dispersion process in the computational domain is obtained after the simulation. The results show that the drop velocity increases with a larger diameter, and the droplet can be suspended for a longer period when the diameter is less than or equal to $1.00 \times 10^{-5} \text{ m}$, as the same in (Borro et al., 2021).

3.3. Individual Exposure Risk Estimation

In this part, we *first* give the mathematical presentation of the instantaneous personal exposure risk. As mentioned above, when the person and the cough droplets meet in the same simulation place, the instantaneous exposure risk is defined as the possible maximal mass of droplets suffered. *Then*, based on the numerical simulation results of the typical cough, we count the mass of droplets vary with the time at several fixed distances planes, and analyze the data distribution to determine the model function, which is the final formula of the instantaneous exposure risk. *Finally*, the total exposure risk during the visiting time for each person is estimated.

It should be noted that, according to a few studies that analyze the dispersion of cough-generated droplets in the wake of a walking person (Li et al., 2020), the situation can be more complicated when there are more individuals. Besides, the direction of the cough is likely to be different from that of the individual’s movement, so the impact of walking on cough transmission is unpredictable. Therefore, in this model, we assume that the infector j is stationary when he/she starts the g -th cough in the position $\mathbf{p}_{j,g}^{\text{start}}$ at time $t_{j,g}^{\text{start}}$. Then, cough droplets’ movements are only determined by the physical distance from $\mathbf{p}_{j,g}^{\text{start}}$ and the time interval from $t_{j,g}^{\text{start}}$.

All positions needed here can be determined based on the crowd simulation in Section 3.1. When individual i in the position $\mathbf{p}_i(t)$ at time $t (\geq t_{j,g}^{\text{start}})$, the instant distance from the position $\mathbf{p}_{j,g}^{\text{start}}$ is $d_{i,j,g}(t)$ and the time interval from $t_{j,g}^{\text{start}}$ is $t_{j,g}^{\text{interval}} = t - t_{j,g}^{\text{start}}$. If $d_{i,j,g}(t)$ and $t_{j,g}^{\text{interval}}$ are both in

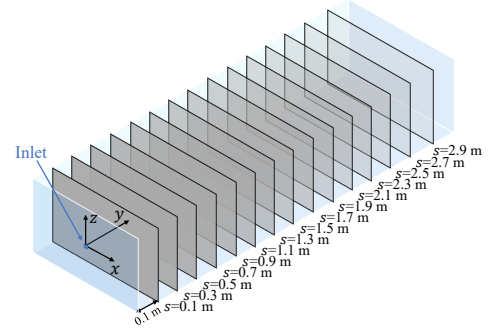


Fig. 4: 15 representative x-z planes in the three-dimensional schematic diagram of the computational domain.

the cough infectious distance D^{inf} and infectious time T^{inf} respectively, based on the definition, we have

$$E_{i,j,g}(t) = M_{j,g} \left(d_{i,j,g}(t), t_{j,g}^{\text{interval}} \right), \quad (4)$$

where $E_{i,j,g}(t)$ is the instantaneous exposure risk of the individual i exposed to individual j 's g -th cough, and $M_{j,g} \left(d_{i,j,g}(t), t_{j,g}^{\text{interval}} \right)$ is the maximal mass of droplets generated from individual j 's g -th cough when spreading to the distance $d_{i,j,g}(t)$ after $t_{j,g}^{\text{interval}}$.

On the assumption that all coughs are typical, to get the deterministic mathematical formula of $M_{j,g} \left(d_{i,j,g}(t), t_{j,g}^{\text{interval}} \right)$, it is sufficient to analyze the change of droplet mass during the transmission of one typical cough. Since the density of all droplets in Section 3.2 is set to the same value, the larger the diameter, the greater the droplet mass. Droplets with a large mass fall quickly, and their destinations are always close to the injector. In contrast, others with small mass exist in the computation region for a long time, and their physical distances from the injector are random. Hence, it is hard to find a suitable deterministic formula to describe the change of droplet mass with the time-varying distance. To solve this problem, we select representative x-z planes and model the arriving droplet mass varies with time for each plane. For a typical cough, the mass of droplets $M_s(t^{\text{cg}})$ passing through the x-z plane along “y” axis in distance s m at time $t^{\text{cg}} (\leq T^{\text{inf}})$ can be counted. Since the individual is represented as a circle (i.e., $r^{\text{ped}} = 0.2 \text{ m}$) (Xiao et al., 2021), the calculation domain in the simulation is divided with a length of 0.20 m along the “y” axis, and the x-z plane at the center of each segment is taken as the representative of each region. More exactly, the value of s starts from 0.10 m to 2.90 m with an interval of 0.20 m, and there are 15 representative x-z planes $\Phi = \{\phi_1, \phi_2, \dots, \phi_s, \dots, \phi_{15}\}$ as shown in Fig. 4.

Moreover, to avoid reducing the counted droplet mass caused by the setting of timestep T^{step} in pedestrian dynamics simulation, the module uses the total mass of droplets passing through ϕ_s during $\left[\frac{t^{\text{cg}}}{T^{\text{step}}} \right]$ timesteps as $M_s(t^{\text{cg}})$. As settings in the second module (Section 3.2), the simulation time step size of cough droplet transmission is 0.01 s. Therefore, the simulation time step size of pedestrian dynamics

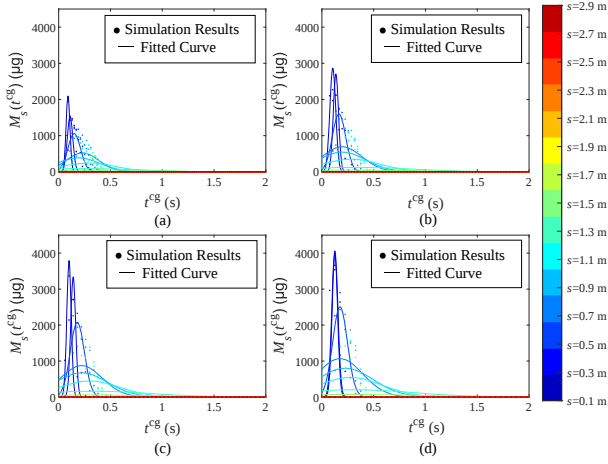


Fig. 5: $M_s(t^{cg})$ versus t^{cg} at different times: (a) $T^{\text{step}} = 0.02s$; (b) $T^{\text{step}} = 0.03s$; (c) $T^{\text{step}} = 0.04s$; (d) $T^{\text{step}} = 0.05s$.

T^{step} should be bigger than 0.01 s. When we set $T^{\text{step}} = 0.02s, 0.03s, 0.04s, 0.05s$, respectively, the statistical results show that the distribution of $M_s(t^{cg})$ varies with t^{cg} and it conforms to the Gaussian distribution (see Fig. 5). Thus, for the ϕ_s , $M_s(t^{cg})$ is obtained by

$$M_s(t^{cg}) = a_s * \exp\left(-\left(\frac{t^{cg} - b_s}{c_s}\right)^2\right), \quad (5)$$

where a_s , b_s and c_s are parameters of the fitted Gaussian distribution function for specific ϕ_s .

After determining the function of the $M_s(t^{cg})$ for a typical cough, since t^{cg} in the typical cough is the same as $t_{j,g}^{\text{interval}}$ of individual j 's g -th cough, if $d_{i,j,g}(t) \in [s - 0.1, s + 0.1]$ and $t_{j,g}^{\text{interval}} \leq T^{\text{inf}}$, the formular of $M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}})$ is represented by

$$M_{j,g}(d_{i,j,g}(t), t_{j,g}^{\text{interval}}) = a_s * \exp\left(-\left(\frac{t_{j,g}^{\text{interval}} - b_s}{c_s}\right)^2\right). \quad (6)$$

As a consequence, if $d_{i,j,g}(t) \in [s - 0.1, s + 0.1]$ and $t \leq t_{j,g}^{\text{start}} + T^{\text{inf}}$, there is

$$E_{i,j,g}(t) = a_s * \exp\left(-\left(\frac{t - t_{j,g}^{\text{start}} - b_s}{c_s}\right)^2\right). \quad (7)$$

Then, the exposure risk of individual i during the visit follows

$$E_i = \sum_{t=t_i^{\text{enter}}}^{t_i^{\text{enter}}+T_i^{\text{dwell}}} \sum_{j=1}^{J(t)} \sum_{g=1}^{J_G(t)} E_{i,j,g}(t), \quad (8)$$

where t_i^{enter} and T_i^{dwell} respectively denotes the place enter time and the dwell time of individual i , $J(t) (\leq C^{\text{inf}})$ indicates the number of infectors in the simulation place at time t , and $J_G(t)$ is the number of infectious coughs of the infector j at time t .

3.4. Prediction of Transmission Trends

In this part, we first determine the number of high-risk exposed people C^{risk} during the simulation horizon. Then, based on C^{risk} , we present a novel method to predict the number of new cases C^{new} .

Individual i whose $E_i > \alpha$ is defined as the high-risk exposed individual, where α is the cut-line of high exposure risk. When the number of total individuals for the simulation is C^{total} , the value of C^{risk} changes with the α and we have

$$C^{\text{risk}}(\alpha) = \sum_{i=1}^{C^{\text{total}}} \psi(E_i, \alpha), \text{ where } \psi(E_i, \alpha) = \begin{cases} 1, & \text{if } E_i > \alpha \\ 0, & \text{otherwise} \end{cases}. \quad (9)$$

Admittedly, C^{new} always increases with the growth of C^{risk} , i.e., there is a positive correlation between the C^{new} and the C^{risk} . From the viewpoint of mathematics, C^{new} can be set as a function F of C^{risk} quantitatively as

$$C^{\text{new}} = F(C^{\text{risk}}). \quad (10)$$

We consider the simplest relationship to determine the function $F(C^{\text{risk}})$ by assuming that $F(C^{\text{risk}})$ is a linear equation. Moreover, in an extreme case, when there are no infectious diseases, there are no high-risk exposed people and new cases, i.e., $C^{\text{new}} = C^{\text{risk}} = 0$. Based on these analyses, the function $F(C^{\text{risk}})$ passes through the origin and be defined as

$$C^{\text{new}} = F(C^{\text{risk}}) = \beta * C^{\text{risk}}(\alpha), \quad (11)$$

where β is the proportionality coefficient.

To determine the values of parameters α and β in different scenes, the actual historical numbers of new cases are needed. When C_v^{new} and C_v^{real} are the predicted and actual values of the $v (= 1, 2, \dots, V)$ -th historical observation, respectively, indexes such as mean absolute error (MAE) can be used to describe the error between C_v^{new} and C_v^{real} . Herein, we select MAE as the index, and the appropriate values of α and β can be estimated as

$$\alpha^*, \beta^* = \arg \min_{\alpha, \beta} \left(\frac{1}{V} \sum_{v=1}^V |C_v^{\text{new}} - C_v^{\text{real}}| \right). \quad (12)$$

Finally, the C^{new} in the future can be predicted with the α^* and β^* via Equations (9-11).

4. Applications in the COVID-19 of the United States

4.1. Data Sources

To estimate RID transmission based on the actual situation, we used data from all 50 states and Washington, D.C. in the United States (U.S.) during the spreading of COVID-19. From May 1st to July 25th in 2020, 79 days with complete national-level data were utilized, and statistics of public data were made based on days.

There were 3 inputs in our model, as shown in Fig. 1, and they were introduced as follows: 1) The number of

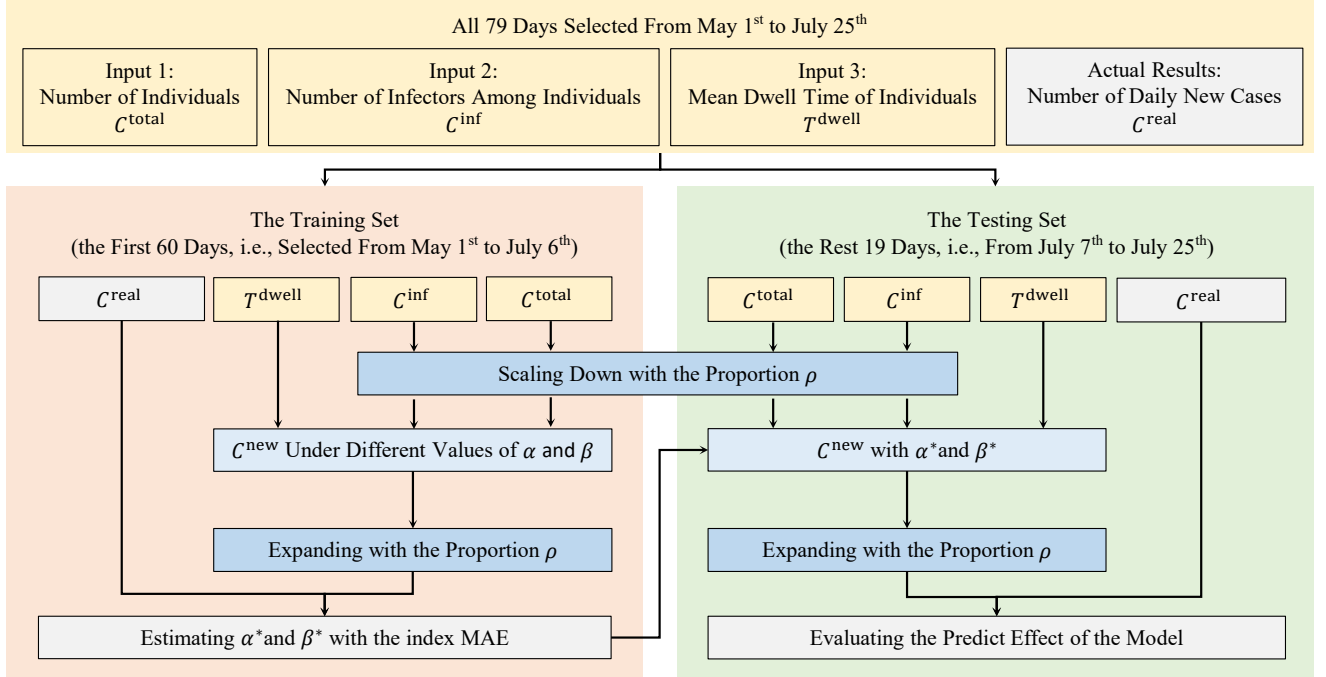


Fig. 6: Schematic diagram of data relations used in this case.

individuals C^{total} each day was approximated as the daily population traveling out of home, which was collected from the U.S. Department of Transportation's Bureau of Transportation Statistics Trips by Distance – National data product (Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland, 2021; Zhang et al., 2020). 2) The number of infectors among individuals C^{inf} was not available from public data because whether each infector had a trip in a single day was unknown. As there were some undiagnosed cases in the travel crowd, we presented a reasonable assumption, i.e., the ratio of infectors to individuals was equal to the ratio of the total number of cases to the population traveling out of home. The total number of cases in each day was found from the U.S. Centers for Disease Control and Prevention (Centers for Disease Control and Preventions (CDC), 2021a). 3) Mean dwell time T^{dwell} of individuals represented the average time spent for each person in the public space. We collected this data from the University of Maryland COVID-19 Impact Analysis Platform (Xiao et al., 2021; Zhang et al., 2020).

The study period (79 days selected from May 1st to July 25th) was divided into a training set (the first 60 days, i.e., selected from May 1st to July 6th) and a testing set (the rest 19 days, i.e., from July 7th to July 25th), as shown in Fig. 6. Besides, to reduce the computational cost, the number of individuals C^{total} and the number of infectors among individuals C^{inf} were scaled down with a proportion ρ for the simulation in both the training and testing sets. Consequently, the main outputs of the simulation, i.e., the number of high-risk exposed people C^{risk} and the number

of daily new cases C^{new} , should be expanded with the same proportion ρ after the simulation. We used the first day in the study period (May 1st, 2020) as the benchmark, then 245,469,060 individuals who had at least one trip at that day could be scaled down to 10,000 individuals for the simulation, thus the proportion $\rho = \frac{10,000}{245,469,060} = 4.07 \times 10^{-5}$. The values of α^* and β^* could be estimated in the training set, and then be adopted in the testing set to evaluate the model.

In the model validation, the model's outcome (i.e., number of new cases C^{new}) was compared with the actual (i.e., number of daily new cases C^{real}), and the evaluation indexes were mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Besides, the real number of daily new cases C^{real} was obtained from the U.S. Centers for Disease Control and Prevention (Centers for Disease Control and Preventions (CDC), 2021a).

4.2. Simulation Setups

In fact, there are various indoor scenes (e.g., restaurant, cinema, and subway hub), and their specific geometry or activity may impact the transmission process, which can be captured by our microscopic model. Even so, paying too much attention to the heterogeneous settings is unnecessary, since considering all specific settings is not applicable, and a more practical way is to formulate a general public space. Accordingly, a 22m $\times$ 22m indoor room without non-walkable areas was formulated, and there was an entrance on one side of the room and an exit on the opposite side (see

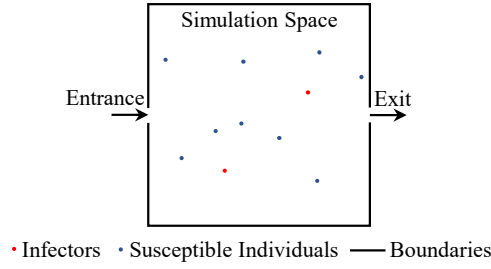


Fig. 7: The sketch map of the simulation space in this case.

Fig. 7). The red and blue dots represented the infectors and the susceptible individuals, respectively.

In the simulations, the specific individual number, the infector number and the mean dwell time were determined according the realistic data introduced in Section 4.1. Initially, no individual existed in the room, and the crowd were designed to enter the indoor room through the entrance in sequence. The individual followed a random walking pattern in the room during the dwell time and leave the room through the exit. The individual was represented by a circle (i.e., $r^{\text{ped}} = 0.2$ m), and the relaxation time and the desired speed were $\tau_i = 0.50$ s and $v_i^0 = 1.34$ m/s, respectively. The infector averagely coughed every 15 s after entering the room, and the infectious time of a cough followed a uniform distribution from 0 to 15 s. The infectious distance D^{inf} of a typical cough was set at 1.70 m since $M_s (t^{\text{cg}})$ equaled almost 0 at a greater distance according to Fig. 5(c).

4.3. Simulation Outputs

4.3.1. Individual Exposure Risk Estimation

The exposure risk of each person in a single day is obtained from the simulation. Picking 4 days (May 1st, May 17th, July 10th, July 25th) as examples, the inputs of the simulation (i.e., C^{total} , C^{inf} , and T^{dwell}) and the statistical results of simulation outputs (i.e., the number of susceptible individuals C^{sus} with $E_i = 0$ or $E_i > 0$) are shown in Table 2 and Fig. 8.

Simulation results show that C^{sus} with $E_i \in (0.00, 0.50]$ μg on May 1st is 2,722 more than that on July 25th. However, C^{sus} with $E_i > 13.00$ μg on May 1st is 1,652 less than that on July 25th. In other words, with the increase of virus transmission time, more and more individuals have higher exposure risks. This trend is consistent with existing cognition and in line with the reality, which demonstrates the validity of our model.

4.3.2. Prediction of New Cases

In order to analyze the contributions of 2 parameters (α and β) to the model results, Sobol method is adopted and the total-effect index is used as the evaluation measurement. According to the previous results of personal exposure risk, the variation range of α is between 0.50 μg and 100.00 μg . Besides, β is set to change from 1.00×10^{-5} to 2.00×10^{-3} through numerous tests. Saltelli's sampling scheme generates 6,000 samples for each training day. Thus, the

Table 2

These are examples of simulation inputs and statistical results.

Date	Inputs			Results	
	C^{total}	C^{inf}	T^{dwell} /minutes	C^{sus} with $E_i=0$	C^{sus} with $E_i>0$
05/01	10,000	39	26	1,092	8,869
05/17	10,210	55	24	724	9,431
07/10	10,210	128	25	225	9,857
07/25	9,844	169	23	138	9,537

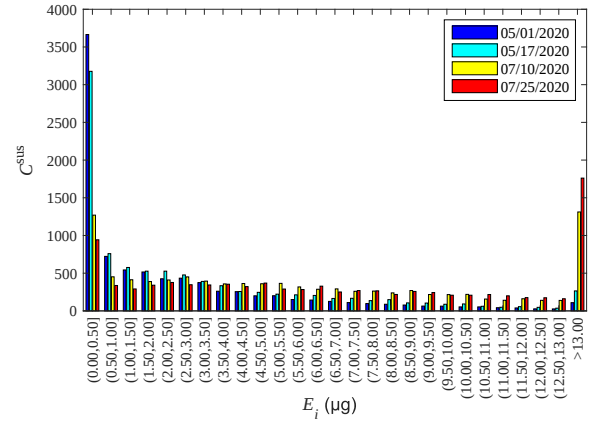


Fig. 8: The number of individuals varies in different exposure risk segments.

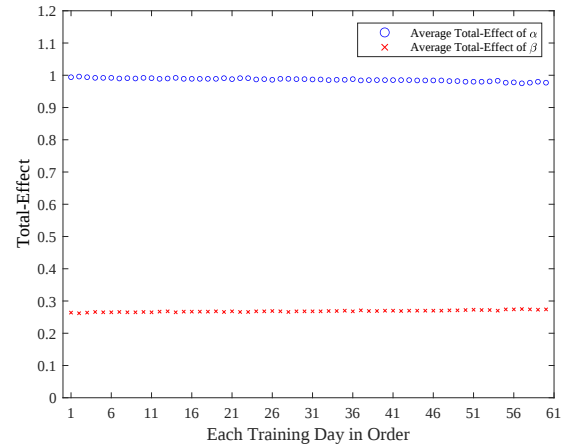


Fig. 9: The average total-effect of parameters α and β in each training day.

average total-effect of parameters α and β in each training day are shown in Fig. 9, and all days' average total-effect of α and β are 0.986 and 0.269, respectively, where the confidence interval level is 95.00%. Results illustrate that the output C^{new} of our model depends on parameters α and β . Parameter α makes more contributions as expected because it is the cut-line of high exposure risk, which is the key to determining both the C^{risk} and C^{new} .

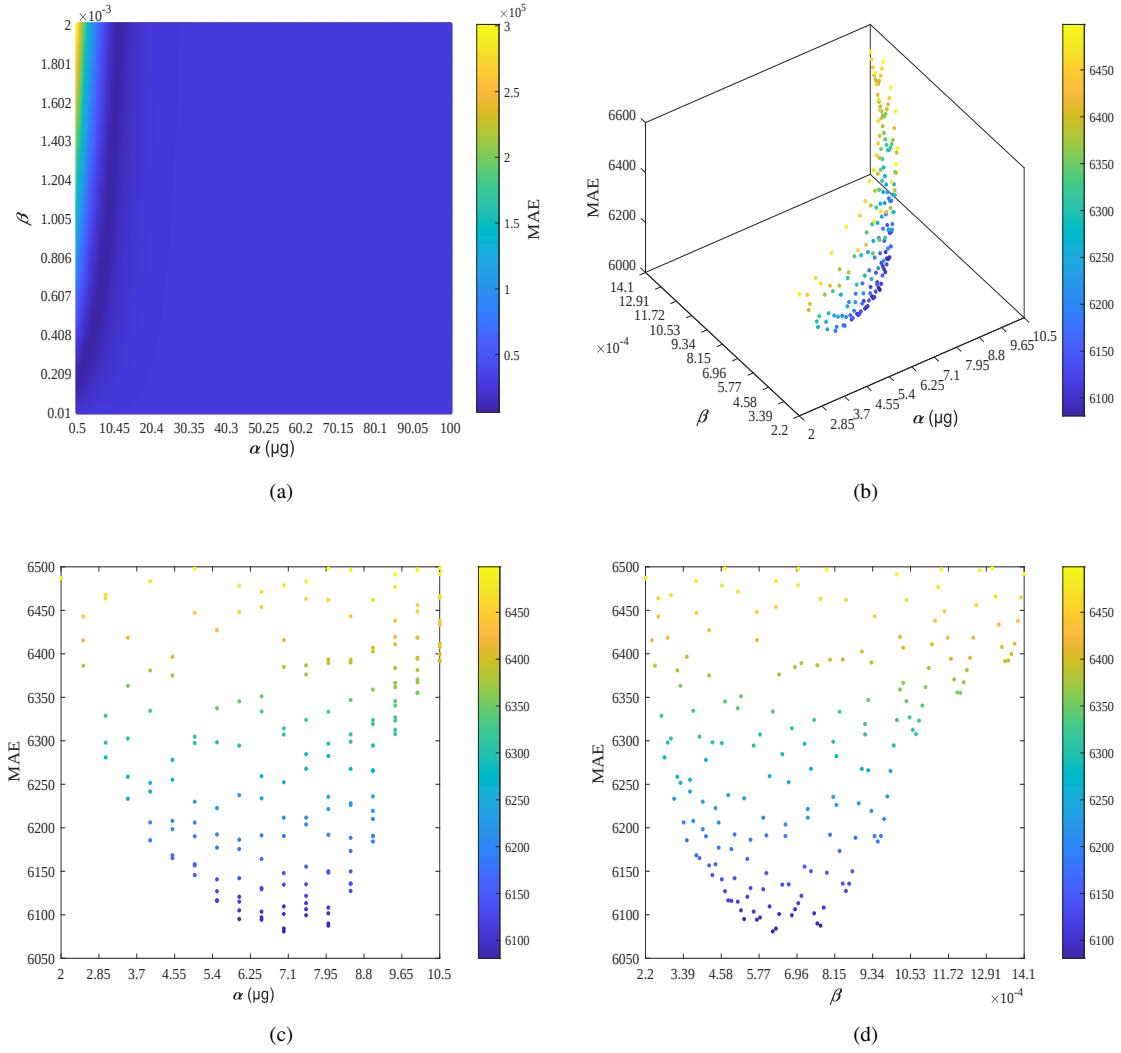


Fig. 10: MAE varies with α and β . (a) The side view of $\alpha - \beta$ when $\alpha \in [0.50, 100.00] \mu\text{g}$ and $\beta \in [1.00 \times 10^{-5}, 2.00 \times 10^{-3}]$. (b-d) The front view and 2 side views when MAE are less than 6,500.

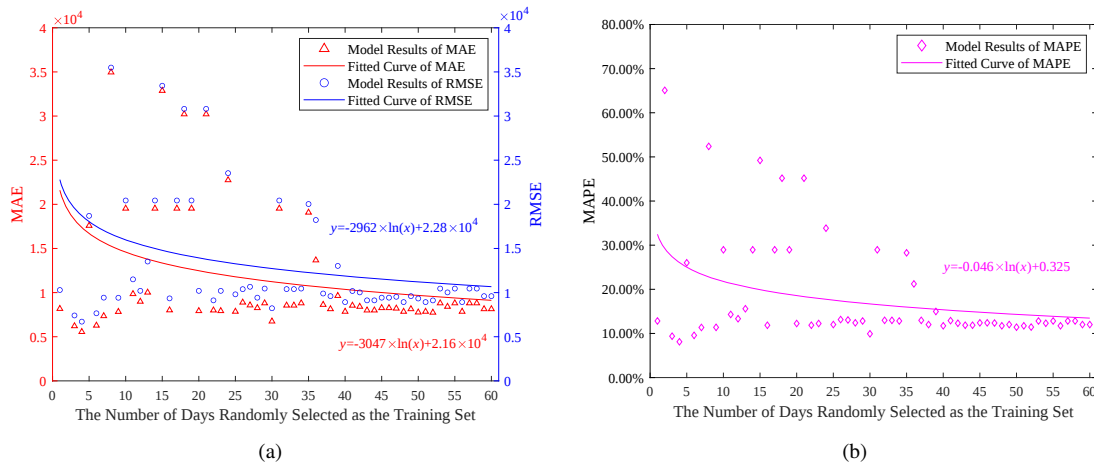


Fig. 11: (a) MAE, RMSE, and (b) MAPE change with the number of days randomly selected as the training set.

To predict the daily new cases C^{new} in the testing set, parameters α^* and β^* are estimated first. After expanding the predicted results with the proportion ρ (4.07×10^{-5} in Section 4.1), MAE varies with α and β can be seen from Fig. 10. Herein, we find $\alpha^* = 7.00 \mu\text{g}$ and $\beta^* = 6.20 \times 10^{-4}$ (corresponding MAE is 6,080.89) based on the training set according to Equation (12). Note that the dark blue curve in Fig. 10(a) indicates that β increases with the growth of α as expected.

Then, with parameters α^* and β^* , the prediction C^{new} in the testing set can be calculated and expanded with the proportion ρ . As a result, our model achieves a good prediction effect under different evaluation indexes: MAE is 8,140.00, RMSE is 9,598.17, and MAPE is 12.01%.

For some machine learning models, the number of training samples with only 60 days is small, which cannot reflect the optimal performance of the model (Zhang and Jiang, 2021). To verify whether 60 days are enough for training our model, we randomly select several days (e.g., 1 day, 2 days, ..., 59 days) from 60 days as new training sets, and analyze the changes of prediction effects in the same testing set (i.e., the rest 19 days). We estimate α^* and β^* in the same way as before for different new training sets, and the evaluation results on the testing set are shown in Fig. 11. The values of MAE, RMSE, and MAPE decrease with the increase of the number of days in training sets, and each of the indexes converges to a stable value within 60 days. Hence, 60 days are sufficient to be the training set, based on which the parameters obtained are reliable, and the corresponding prediction results can represent the best effect of our model.

5. Comparison with Other Models

In this section, following the study region and period in Section 4, we predict transmission trends via our model and several existing models from macroscopic or microscopic levels and then compare their performances.

5.1. Microscopic-Level Models

Since the proposed model focuses on the general place and determines the personal exposure risk at the microscopic level, comparison methods should have the same central points, thus Xiao's model (Xiao et al., 2021) and Hernández-Orallo's model (Hernández-Orallo and Armero-Martínez, 2021) are adopted. Besides, as there is no direct prediction of daily new cases in these two compared models, we follow Section 3.4 to make the forecast after determining the personal exposure risk by these two models, respectively.

The input data fields required by the two models are the same as our proposed model, and they have the same training and testing sets as Section 4. For the contact approach and cough approach of Xiao's model, the cut-off distance of the exposure is 1.00 m and 2.50 m, respectively, as the same settings in (Xiao et al., 2021). In Hernández-Orallo's model, the contact cut-off distance is also 1.00 m, and the parameter for adjusting the model is set at the same value 1/30 as in (Hernández-Orallo and Armero-Martínez, 2021).

The simulation place is an indoor room, as shown in Fig. 7, which has low air renewal, high temperatures, and low solar radiation, so the value of the quality of the medium in Hernández-Orallo's model is $7/9 = (1+1/3+1)/3$. Based on the training set, we follow Equation (12) to find the α^* and β^* for each model: $\alpha^* = 10.00 \text{ s}$ and $\beta^* = 1.19 \times 10^{-4}$ in Xiao's Model (corresponding MAE is 7,205.67); $\alpha^* = 0.60 \text{ MEMs}$ and $\beta^* = 1.41 \times 10^{-4}$ in Hernández-Orallo's Model (corresponding MAE is 7,172.78). Hence, based on Equations (9-11), the prediction results of the testing set by the two microscopic-level models are reported in Fig. 12 and Table 3.

5.2. Macroscopic-Level Models

The typical SIR model (Kermack and McKendrick, 1927; Maier and Brockmann, 2020) and grey model (Luo et al., 2021; Zhang and Jiang, 2021) are adopted as comparison macroscopic-level models. The SIR model is a traditional infectious disease model, and it works with the assumption that the population N^{PL} in the study region is uniform and homogeneously mixed. In the SIR model, the population is divided into three classes, namely, N^{sus} : susceptible, C^{inf} : infected, and N^{rem} : removed (by recovery and death) (Maier and Brockmann, 2020). The time-varying number of cases in each class is governed by the infectivity rate σ^i and the removal rate σ^r (Zamiri et al., 2014). Thus, when we know the initial N^{sus} , C^{inf} , N^{rem} , and the constant values of σ^i and σ^r , we can predict the number of infectors in the future, based on which the number of daily new cases C^{new} are calculated. The grey model, favored for its "simple model, strong adaptability, and easy parameter changes", has been widely used in the field of infectious diseases (Zhang and Jiang, 2021). Unlike the neural network methods that need a substantial number of datasets for training system parameters, the grey model can predict the trend of the COVID-19 well with limited information. Specifically, only the historical daily new infected cases are needed to forecast the C^{new} in the following days.

In Section 4, only 60 days (selected from May 1st to July 6th) are used as the training set due to the lack of data. Fortunately, data required by the SIR model and grey model in 67 days (from May 1st to July 6th) are available from the public datasets. Thus, the time-series data from May 1st to July 6th at the national level of the U.S. are adopted as the training set here. The testing set is the same as Section 4, i.e., 19 days from July 7th to July 25th. Besides, the parameters of the SIR model and grey model are set based on the national dataset. In SIR model, the population size of the U.S. is $N^{\text{PL}} = 327,167,434$, which comes from the U.S. Department of Transportation's Bureau of Transportation Statistics Trips by Distance – National data product (Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland, 2021; Zhang et al., 2020). The cumulative number of infected cases C^{inf} is obtained from the U.S. Centers for Disease Control and Prevention (Centers for Disease Control and Preventions (CDC), 2021a). The number of

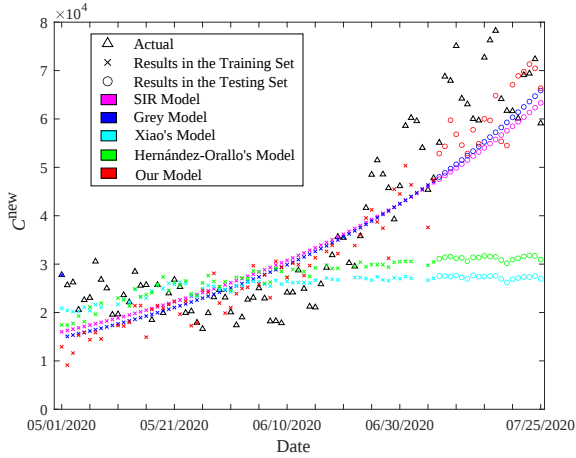


Fig. 12: The actual values and different model results of C^{new} change with the date.

Table 3

These are MAE, RMSE, and MAPE between actual and predicted results by different models.

Model	MAE	RMSE	MAPE
SIR Model	11,589.07	13,638.09	16.84%
Grey Model	10,594.70	12,768.68	15.39%
Xiao's Model	38,965.99	39,468.87	58.46%
Hernández-Orallo's Model	34,989.95	35,537.92	52.41%
Our Model	8,140.00	9,598.17	12.01%

NOTE. MAE, mean absolute error; RMSE, root mean square error; MAPE, mean absolute percentage error.

removed cases (by recovery and death) N^{rem} is provided by the record COVID-19 DATA HUB¹ (Guidotti and Ardia, 2020). The number of susceptible individuals is getting as $N^{\text{sus}} = N^{\text{PL}} - C^{\text{inf}} - N^{\text{rem}}$. After fitting the training set with the ordinary least squares, we get parameters that minimize the sum of squares of errors: $\sigma^i = 0.024$ and $\sigma^r = 0.007$. Besides, the basic reproduction number $R_0 = \frac{\sigma^i}{\sigma^r} = 3.43$ is reasonable with existing research (Centers for Disease Control and Preventions (CDC), 2021b), and it is in line with the situation when no epidemic prevention and control policies were implemented in the U.S.. For the grey model, the cumulative number of infected cases C^{inf} is adopted from the U.S. Centers for Disease Control and Prevention (Centers for Disease Control and Preventions (CDC), 2021a), and there are no other parameters to be set. Thus, the prediction results of the testing set by the two macroscopic-level models are reported in Fig. 12 and Table 3.

5.3. Comparison of Results

The time series numbers of daily new cases C^{new} in the testing set estimated by the proposed model, macroscopic models, microscopic models, and obtained via the real-world

¹See websites <https://covid19datahub.io> for more details about COVID-19 DATA HUB.

dataset are shown in Fig. 12 and Table 3. Based on these, the proposed model achieves the best prediction performance when compared with the four existing models.

The SIR model and grey model at the macroscopic level cannot determine the personal exposure risk and the number of high-risk exposed people, but these can be estimated by microscopic-level models such as ours. Meanwhile, even though time series data are used for training the SIR model and grey model, the number of samples in the training set may still be too small to guarantee the best performance. However, based on the analyses of Fig. 11, the number of samples, in this case, is enough to show the good performance of our proposed model, thus our model is more suitable for small sample data. We admit that these two models are classical and traditional, and now there are many improved ones based on them, which may have better prediction ability than our model.

The personal exposure risk is defined differently in various microscopic level models, e.g., Xiao's model, Hernández-Orallo's model, and ours. In Xiao's model, the exposure risk in the influence area is set to a constant value that cannot change with the physical distances. Hernández-Orallo's model has proposed the exposure risk based on the distance and time between individuals and the quality of the medium. Our model considers the cough droplet dispersion processes and the time-varying individuals' physical distances, and determines the exposure risk as the total maximum droplet mass that the individual may suffer. Results in Fig. 12 and Table 3 show that our model's definition of exposure risk is more appropriate than the other 2 models to forecast the transmission trends of RIDs.

6. Discussions and Future Perspectives

In the construction of the proposed model, to forecast the number of new cases, the assumption that $F(C^{\text{risk}})$ is a linear equation has been illustrated, and other functional forms (e.g., exponential function, power function) can be further studied.

In the simulation setups (Section 4), due to the lack of real-world scene and human activities data, a general indoor space without non-walkable areas and a random walking pattern of individuals are formulated for the reproduction of microscopic epidemic transmission, and by this means, the simulation results of our model can be compared with the macroscopic epidemic transmission data. Fortunately, the potential drawbacks due to ideal assumptions on the model comparison results is minimized since the involved models here are all applied in the same scene. It should be noted that our model can be used to simulate scenes with different obstacles, activities, and other settings, once we have the specific scene setting data.

The proposed model can predict transmission trends under different model inputs or scenes. Hence, for factors as Table 1 shows, if they affect the model inputs or scenes, the model outputs change accordingly, then the transmission trends caused by factors can be quantitatively explored. For example, GDP is an economic indicator (Zhang et al.,

2021). Generally, there are more people and more advanced transportation systems in areas with high-level GDP, where people have more activities in the public space. Therefore, for the inputs of our model, the number of individuals C^{total} and the mean dwell time T^{dwell} in high-level GDP areas will be larger than that in regions with low-level GDP, resulting in different prediction results based on the proposed model. Similarly, in the scene with a large number of visits, the NPI named entry limitation policy (Xiao et al., 2021), which means the maximum number of visitors inside the indoor place is limited, can change the number of individuals C^{total} definitely, and its contributions can be captured by our model. However, we must admit that the influence of several factors (e.g., sunspot numbers (Nasirpour et al., 2021)) cannot be considered in our model because they have limited effect on the model inputs or scenarios. Besides, for several prevention and control measures that block the spread of cough droplets directly near the mouth, such as facemask wearing, the transmission of droplets is different from a typical cough droplet transmission. Since the virion-laden droplet transmission of a typical cough is set as a deterministic model in Section 3, it is necessary to improve the model to evaluate these NPIs' contributions. Moreover, the direct and indirect effects of influence factors on the prediction results can be further determined with methods such as structural equation modeling (Ma and Bennett, 2021). In fact, policy-makers and political communicators should consider not only the influences of factors, but also the social acceptance and effects of different rules, and these can be further studied (Bontempi, 2021; Gollwitzer et al., 2021; Smith et al., 2021).

7. Conclusions

In this paper, we developed a microscopic level exposure-risk-based model to predict the transmission trends of RIDs in a general public place. Specifically, to determine the personal exposure risk, our model coupled the motions of individuals and the time-varying cough droplet dispersion process. Then, the number of new cases was predicted based on the assumption that it was suitable for a linear function with the number of high-risk exposed people, and the prediction could be tested quantitatively when compared with actual values. Based on COVID-19 data collected in the U.S., we constructed real-world simulations and compared the prediction effect of our model with those of several existing models. Our model achieves superior performance in forecasting the transmission trends of RIDs, and brings guiding significance to control and prevention.

Fundings

This work was supported by the National Natural Science Foundation of China (Grant No. 72101276), the Shenzhen Science and Technology Program (Grant No. GXWD 20200830165201001) and the Fundamental Research Funds for the Central Universities (Grant No. 2242021R10037).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank and acknowledge our partners and data sources in this effort: 1) the daily population traveling out of home and the population size of the U.S. from the U.S. Department of Transportation's Bureau of Transportation Statistics Trips by Distance – National data product; 2) the total number of cases in each day, the cumulative number of infected cases, and the number of daily new infected cases from the U.S. Centers for Disease Control and Prevention; 3) the mean dwell time of individuals from the University of Maryland COVID-19 Impact Analysis Platform; 4) the number of removed cases (by recovery and death) from the dataset named COVID-19 DATA HUB in the URL <https://covid19datahub.io>.

References

- Abdul Salam, P.S., Bock, W., Klar, A., Tiwari, S., 2021. Disease contagion models coupled to crowd motion and mesh-free simulation. *Mathematical Models and Methods in Applied Sciences*, 1–19.
- Abu-Hammad, O., Alnazzawi, A., Borzangy, S.S., Abu-Hammad, A., Fayad, M., Saadaleddin, S., Abu-Hammad, S., Dar-Odeh, N., 2020. Factors influencing global variations in covid-19 cases and fatalities; a review 8, 216.
- Ahmadi, M., Sharifi, A., Khalili, S., 2021. Presentation of a developed sub-epidemic model for estimation of the covid-19 pandemic and assessment of travel-related risks in iran. *Environmental science and pollution research* 28, 14521–14529.
- Al Huraimel, K., Alhosani, M., Kunhabdulla, S., Stietiya, M.H., 2020. Sars-cov-2 in the environment: Modes of transmission, early detection and potential role of pollutions. *Science of the Total Environment* 744, 140946.
- Aliabadi, A.A., Rogak, S.N., Green, S.I., Bartlett, K.H., 2010. CFD simulation of human coughs and sneezes: a study in droplet dispersion, heat, and mass transfer, in: *ASME International Mechanical Engineering Congress and Exposition*, pp. 1051–1060.
- Anand, U., Cabrerios, C., Mal, J., Ballesteros Jr, F., Sillanpää, M., Tripathi, V., Bontempi, E., 2021. Novel coronavirus disease 2019 (covid-19) pandemic: from transmission to control with an interdisciplinary vision. *Environmental research* 197, 111126.
- Aragão, D.P., Oliveira, E.V., Bezerra, A.A., Dos Santos, D.H., da Silva Junior, A.G., Pereira, I.G., Piscitelli, P., Miani, A., Distante, C., Cuno, J.S., et al., 2021. Multivariate data driven prediction of covid-19 dynamics: Towards new results with temperature, humidity and air quality data. *Environmental Research*, 112348.
- Arav, Y., Klausner, Z., Fattal, E., 2021. Theoretical investigation of pre-symptomatic sars-cov-2 person-to-person transmission in households. *Scientific reports* 11, 1–12.
- Araya, F., 2021. Modeling the spread of covid-19 on construction workers: An agent-based approach. *Safety Science* 133, 105022.
- Bazant, M.Z., Bush, J.W., 2021. A guideline to limit indoor airborne transmission of covid-19. *Proceedings of the National Academy of Sciences* 118.
- Bi, R., 2018. A numerical investigation of human cough jet development and droplet dispersion. Master's thesis. Dept. Mech.& Matls.Eng., University of Western Ontario, London, Canada.

- Black, A.J., McKane, A.J., 2010. Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough. *Journal of the royal society interface* 7, 1219–1227.
- Bontempi, E., 2021. The europe second wave of covid-19 infection and the italy “strange” situation. *Environmental Research* 193, 110476.
- Bontempi, E., 2022. A global assessment of covid-19 diffusion based on a single indicator: Some considerations about air pollution and covid-19 spread. *Environmental Research* 204, 112098.
- Bontempi, E., Coccia, M., 2021. International trade as critical parameter of covid-19 spread that outclasses demographic, economic, environmental, and pollution factors. *Environmental Research* 201, 111514.
- Bontempi, E., Coccia, M., Vergalli, S., Zanoletti, A., 2021. Can commercial trade represent the main indicator of the covid-19 diffusion due to human-to-human interactions? a comparative analysis between italy, france, and spain. *Environmental research* 201, 111529.
- Bontempi, E., Vergalli, S., Squazzoni, F., 2020. Understanding covid-19 diffusion requires an interdisciplinary, multi-dimensional approach. *Environmental research* 188, 109814.
- Borro, L., Mazzei, L., Raponi, M., Piscitelli, P., Miani, A., Secinaro, A., 2021. The role of air conditioning in the diffusion of sars-cov-2 in indoor environments: A first computational fluid dynamic model, based on investigations performed at the vatican state children’s hospital. *Environmental research* 193, 110343.
- Bouchnita, A., Jebrane, A., 2020a. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. *Chaos, Solitons & Fractals* 138, 109941.
- Bouchnita, A., Jebrane, A., 2020b. A multi-scale model quantifies the impact of limited movement of the population and mandatory wearing of face masks in containing the covid-19 epidemic in morocco. *Mathematical Modelling of Natural Phenomena* 15, 31.
- Castillo, O., Melin, P., 2020. Forecasting of covid-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic. *Chaos, Solitons & Fractals* 140, 110242.
- Centers for Disease Control and Preventions (CDC), 2021a. Covid-19 case surveillance public data access, summary, and limitations. Accessed on: June 22nd, 2021. [Online]. Available: <https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36>.
- Centers for Disease Control and Preventions (CDC), 2021b. Covid-19 pandemic planning scenarios. Accessed on: October 20th, 2021. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>.
- Chao, C.Y.H., Wan, M.P., Morawska, L., Johnson, G.R., Ristovski, Z., Hargreaves, M., Mengersen, K., Corbett, S., Li, Y., Xie, X., et al., 2009. Characterization of expiration air jets and droplet size distributions immediately at the mouth opening. *Journal of aerosol science* 40, 122–133.
- Chraïbi, M., Kemloh, U., Schadschneider, A., Seyfried, A., 2011. Force-based models of pedestrian dynamics. *Networks & Heterogeneous Media* 6, 425.
- Coccia, M., 2021. The impact of first and second wave of the covid-19 pandemic in society: comparative analysis to support control measures to cope with negative effects of future infectious diseases. *Environmental Research* 197, 111099.
- Coticini, E., Frediani, B., Caro, D., 2020. Can atmospheric pollution be considered a co-factor in extremely high level of sars-cov-2 lethality in northern italy? *Environmental pollution* 261, 114465.
- Cross, P.C., Johnson, P.L., Lloyd-Smith, J.O., Getz, W.M., 2007. Utility of r_0 as a predictor of disease invasion in structured populations. *Journal of the Royal Society Interface* 4, 315–324.
- Daniels, S., Wei, H., Han, Y., Catt, H., Denning, D.W., Hall, I., Regan, M., Verma, A., Whitfield, C.A., van Tongeren, M., 2021. Risk factors associated with respiratory infectious disease-related presenteeism: a rapid review. *medRxiv*.
- Davoudi, A., Ahmadi, M., Sharifi, A., Hassantabar, R., Najafi, N., Tayebi, A., Kargari, H.A., Ahmadi, F., Rabiee, M., 2021. Studying the effect of taking statins before infection in the severity reduction of covid-19 with machine learning. *BioMed Research International* 2021.
- Duguid, J., 1946. The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. *Epidemiology & Infection* 44, 471–479.
- Duives, D., Chang, Y., Sparnaaij, M., Wouda, B., Boschma, D., Liu, Y., Yuan, Y., Daamen, W., de Jong, M.C., Teberg, C., et al., 2021. The multi-dimensional challenges of controlling sars-cov-2 transmission in indoor spaces: Insights from the linkage of a microscopic pedestrian simulation and virus transmission models. *medRxiv*.
- Earn, D.J., 2008. A light introduction to modelling recurrent epidemics, in: *Mathematical epidemiology*. Springer, pp. 3–17.
- Eubank, S., Guclu, H., Kumar, V.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 180–184.
- Fang, Z., Huang, Z., Li, X., Zhang, J., Lv, W., Zhuang, L., Xu, X., Huang, N., 2020. How many infections of covid-19 there will be in the “diamond princess”-predicted by a virus transmission model based on the simulation of crowd flow. *arXiv preprint arXiv:2002.10616*.
- Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W., et al., 2020. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature* 584, 257–261.
- Gao, C.X., Li, Y., Wei, J., Cotton, S., Hamilton, M., Wang, L., Cowling, B.J., 2021. Multi-route respiratory infection: when a transmission route may dominate. *Science of the Total Environment* 752, 141856.
- Gao, N., Niu, J., 2006. Transient CFD simulation of the respiration process and inter-person exposure assessment. *Building and environment* 41, 1214–1222.
- Gollwitzer, M., Platzer, C., Zwarg, C., G“oritz, A.S., 2021. Public acceptance of covid-19 lockdown scenarios. *International Journal of Psychology* 56, 551–565.
- Guidotti, E., Ardia, D., 2020. Covid-19 data hub. *Journal of Open Source Software* 5, 2376.
- Gupta, J.K., Lin, C.H., Chen, Q., 2010. Characterizing exhaled airflow from breathing and talking. *Indoor air* 20, 31–39.
- Helbing, D., Molnar, P., 1995. Social force model for pedestrian dynamics. *Physical review E* 51, 4282.
- Hernández-Orallo, E., Armero-Martínez, A., 2021. How human mobility models can help to deal with covid-19. *Electronics* 10, 33.
- Hethcote, H.W., 1976. Qualitative analyses of communicable disease models. *Mathematical biosciences* 28, 335–356.
- Hosseini, E., Ghafoor, K.Z., Sadiq, A.S., Guizani, M., Emrouznejad, A., 2020. Covid-19 optimizer algorithm, modeling and controlling of coronavirus distribution process. *IEEE Journal of Biomedical and Health Informatics* 24, 2765–2775.
- Huang, G., Guo, F., 2021. Loss of life expectancy due to respiratory infectious diseases: Findings from the global burden of disease study in 195 countries and territories 1990–2017. *Journal of Population Research* 1443–2447.
- Jin, S., He, Y., Yang, K., Gan, Q., Huang, W., Wang, X., Meng, C., Wang, H., 2021. The resumption of sports medicine during the covid-19 post-epidemic period: Experiences from wuhan, people’s republic of china. *JBJS* 103, 10–14.
- Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115, 700–721.
- Kim, D., Quaini, A., 2020. Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment. *Mathematical Models and Methods in Applied Sciences* 30, 1893–1915.
- Kretz, T., Lohmiller, J., Sukennik, P., 2018. Some indications on how to calibrate the social force model of pedestrian dynamics. *Transportation research record* 2672, 228–238.
- Kriegel, M., Buchholz, U., Gastmeier, P., Bischoff, P., Abdelgawad, I., Hartmann, A., 2020. Predicted infection risk for aerosol transmission of sars-cov-2. *MedRxiv*.
- Lai, S., Ruktanonchai, N.W., Zhou, L., Prosper, O., Luo, W., Floyd, J.R., Wesolowski, A., Santillana, M., Zhang, C., Du, X., et al., 2020. Effect of non-pharmaceutical interventions to contain covid-19 in china. *nature* 585, 410–413.

- Lembo, R., Landoni, G., Cianfanelli, L., Frontera, A., 2021. Air pollutants and sars-cov-2 in 33 european countries. *Acta Bio Medica: Atenei Parmensis* 92.
- Li, X., Dong, Z.Q., Yu, P., Wang, L.P., Niu, X.D., Yamaguchi, H., Li, D.C., 2021. Effect of self-assembly on fluorescence in magnetic multiphase flows and its application on the novel detection for covid-19. *Physics of Fluids* 33, 042004.
- Li, Z., Wang, H., Zhang, X., Wu, T., Yang, X., 2020. Effects of space sizes on the dispersion of cough-generated droplets from a walking person. *Physics of Fluids* 32, 121705.
- Liu, L., Li, Y., Nielsen, P.V., Wei, J., Jensen, R.L., 2017. Short-range airborne transmission of expiratory droplets between two people. *Indoor air* 27, 452–462.
- Liu, W.m., Hethcote, H.W., Levin, S.A., 1987. Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of mathematical biology* 25, 359–380.
- Luo, X., Duan, H., Xu, K., 2021. A novel grey model based on traditional richards model and its application in covid-19. *Chaos, Solitons & Fractals* 142, 110480.
- Ma, Y., Bennett, D., 2021. The relationship between higher education students' perceived employability, academic engagement and stress among students in china. *Education+ Training*.
- Maier, B.F., Brockmann, D., 2020. Effective containment explains subexponential growth in recent confirmed covid-19 cases in china. *Science* 368, 742–746.
- Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland, 2021. Trips by distance – national. Accessed on: June 27th, 2021. [Online]. Available: <https://data.bts.gov/Research-and-Statistics/Trips-by-Distance-National/6ced-86in>.
- Melin, P., Monica, J.C., Sanchez, D., Castillo, O., 2020. Multiple ensemble neural network models with fuzzy response aggregation for predicting covid-19 time series: the case of mexico, in: *Healthcare, Multidisciplinary Digital Publishing Institute*. p. 181.
- Nasirpour, M.H., Sharifi, A., Ahmadi, M., Jafarzadeh Ghouschi, S., 2021. Revealing the relationship between solar activity and covid-19 and forecasting of possible future viruses using multi-step autoregression (msar). *Environmental Science and Pollution Research* 28, 38074–38084.
- Parisi, D.R., Patterson, G.A., Pagni, L., Osimani, A., Bacigalupo, T., Godfrid, J., Bergagna, F.M., Brizi, M.R., Momesso, P., Gomez, F.L., et al., 2020. Social distance characterization by means of pedestrian simulation. *arXiv preprint arXiv:2009.04019*.
- Perra, N., 2021. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Physics Reports* 913, 1–52.
- Ren, J., Yan, Y., Zhao, H., Ma, P., Zabalza, J., Hussain, Z., Luo, S., Dai, Q., Zhao, S., Sheikh, A., et al., 2020. A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for covid-19. *IEEE Journal Of Biomedical and Health Informatics* 24, 3551–3563.
- Romero, V., Stone, W.D., Ford, J.D., 2020. Covid-19 indoor exposure levels: An analysis of foot traffic scenarios within an academic building. *Transportation Research Interdisciplinary Perspectives* 7, 100185.
- Ronchi, E., Lovreglio, R., 2020. Exposed: An occupant exposure model for confined spaces to retrofit crowd models during a pandemic. *Safety Science* 130, 104834.
- Sharifi, A., Ahmadi, M., Ala, A., 2021. The impact of artificial intelligence and digital style on industry and energy post-covid-19 pandemic. *Environmental Science and Pollution Research* 28, 46964–46984.
- Smith, T.O., Belderson, P., Dainty, J.R., Birt, L., Durrant, K., Chipping, J.R., Tsigarides, J., Yates, M., Naughton, F., Werry, S., et al., 2021. Impact of covid-19 pandemic social restriction measures on people with rheumatic and musculoskeletal diseases in the uk: a mixed-methods study. *BMJ open* 11, e048772.
- Urrutia-Pereira, M., Mello-da Silva, C., Solé, D., 2020. Covid-19 and air pollution: A dangerous association? *Allergologia et immunopathologia* 48, 496–499.
- Verma, A., Bhattacharya, P., Zuhair, M., Tanwar, S., Kumar, N., 2021. Vachochain: Blockchain-based 5g-assisted uav vaccine distribution scheme for future pandemics. *IEEE Journal of Biomedical and Health Informatics*, 1–1.
- Wang, L., Didelot, X., Yang, J., Wong, G., Shi, Y., Liu, W., Gao, G.F., Bi, Y., 2020. Inference of person-to-person transmission of covid-19 reveals hidden super-spreading events during the early outbreak phase. *Nature communications* 11, 1–6.
- Wu, X., Liu, Z., Yin, L., Zheng, W., Song, L., Tian, J., Yang, B., Liu, S., 2021. A haze prediction model in chengdu based on lstm. *Atmosphere* 12, 1479.
- Xiao, Y., Yang, M., Zhu, Z., Yang, H., Zhang, L., Ghader, S., 2021. Modeling indoor-level non-pharmaceutical interventions during the covid-19 pandemic: a pedestrian dynamics-based microscopic simulation approach. *Transport Policy* 109, 12–23.
- Zamiri, A., Yazdi, H.S., Goli, S.A., 2014. Temporal and spatial monitoring and prediction of epidemic outbreaks. *IEEE journal of biomedical and health informatics* 19, 735–744.
- Zhang, J., Jiang, Z., 2021. A new grey quadratic polynomial model and its application in the covid-19 in china. *Scientific Reports* 11, 1–27.
- Zhang, L., Darzi, A., Ghader, S., Pack, M.L., Xiong, C., Yang, M., Sun, Q., Kabiri, A., Hu, S., 2020. Interactive covid-19 mobility impact and social distancing analysis platform. *Transportation Research Record*, 03611981211043813.
- Zhang, L., Li, Y., 2012. Dispersion of coughed droplets in a fully-occupied high-speed rail cabin. *Building and Environment* 47, 58–66.
- Zhang, Y., Aycok, L., Chen, X., 2021. Levels of economic development and the spread of coronavirus disease 2019 (covid-19) in 50 u.s. states and territories and 28 european countries: an association analysis of aggregated data. *Global Health Journal (Amsterdam, Netherlands)* 5, 24–30.
- Zhang, Y., Feng, G., Kang, Z., Bi, Y., Cai, Y., 2017. Numerical simulation of coughed droplets in conference room. *Procedia Engineering* 205, 302–308.
- Zhang, Z., Cui, F., Cao, C., Wang, Q., Zou, Q., 2022a. Single-cell rna analysis reveals the potential risk of organ-specific cell types vulnerable to sars-cov-2 infections. *Computers in biology and medicine* 140, 105092.
- Zhang, Z., Ma, P., Ahmed, R., Wang, J., Akin, D., Soto, F., Liu, B.F., Li, P., Demirci, U., 2022b. Advanced point-of-care testing technologies for human acute respiratory virus detection. *Advanced Materials* 34, 2103646.
- Zhu, C.C., Zhu, J., 2021. Dynamic analysis of a delayed covid-19 epidemic with home quarantine in temporal-spatial heterogeneous via global exponential attractor method. *Chaos, Solitons & Fractals* 143, 110546.
- Zhu, S., Kato, S., Yang, J.H., 2006. Study on transport characteristics of saliva droplets produced by coughing in a calm indoor environment. *Building and environment* 41, 1691–1702.

1. Parameter Settings for the Simulation of a Typical Cough

The transmission medium of the cough is modeled as an incompressible ideal gas with constant properties calculated at ambient conditions. The flow field evolution of coughing is time-dependent, so the simulations are conducted under a transient condition. Since the droplet volume fraction is very low in the cough flow, the Eulerian-Lagrangian method is used in this paper: 1) The continuous phase (the flow field) is modeled by the Eulerian method, which is to solve the Navier-Stokes equations; 2) The discrete phase (the droplet) is modeled by the Lagrangian method, i.e., by tracking the motion of the droplets; 3) The discrete phase can exchange mass, momentum, and energy with the continuous phase, thus fully considering the influence of the flow field on particle diffusion.

The renormalization group (RNG) k - ε model with scalable wall functions is selected as the turbulence model, and the discrete phase model is used for the droplet diffusion (Aliabadi et al., 2010; Gao and Niu, 2006; Zhang and Li, 2012; Zhang et al., 2017). Meanwhile, the transport model between components is used. In the discrete phase model, unsteady particle tracking is selected, and the particle time step size is 1×10^{-3} s. For tracking parameters, the max number of steps is 50,000, and the step length factor is 5. In physical methods, two-way turbulence coupling, stochastic collision, break-up, and coalescence are picked.

Due to the differences in measurement methods and the physiological variability of different volunteers, the size distribution of the droplets discharged from a cough is not unique (Chao et al., 2009). We use the droplet size distribution from (Chao et al., 2009; Duguid, 1946), and the population is modeled by fitting the data with a Rosin-Rammler distribution. The droplet group is injected in the time range of 4.20×10^{-2} s to 1.36×10^{-1} s with a temperature of 35 °C. The injection speed is zero, and all droplets are driven initially by the flow field from the mouth (Borro et al., 2021). We pick the spherical drag model and consider the effect of secondary break-up through a Taylor Analogy Breakup model. Then, the Discrete Random Walk model is used to consider the effects of the turbulent flow on the droplets (Bi, 2018). Moreover, each droplet consists of multi-components (Aliabadi et al., 2010): 1) 94% volume fraction of the pure water which can be evaporated during the transmission, and 2) 6% volume fraction of the pure water but non-evaporative, which represents the mucus and physiological electrolytes that carry infectious agents. The boundary condition of the inlet is defined to be a velocity-inlet, and the velocity direction is set as normal to the boundary. The fluid injected from the inlet consists of air and water vapor at 35 °C, with the mole fraction 93.80% and 6.20%, respectively (Bi, 2018). The velocity imposed at the inlet is a function of time during a cough, and the flow field has the peak velocity v_{ip} (around 22.06 m/s) occurring at 6.60×10^{-2} s (Chao et al., 2009). The “trap” boundary condition is applied for all surfaces, i.e., the droplet will stay at the surface when it collides with the surface and then eliminate from the domain (Borro et al., 2021).

The pressure-based solver is employed as it is suitable for incompressible flows. The Semi-Implicit Method for Pressure-Linked Equations (SIMPLE) algorithm is carried out as the pressure-velocity coupling method. We select first-order upwind for turbulence-related variables and second-order for other variables (such as pressure). The bounded second-order implicit formulation is the transient formulation, since it can bring a higher accuracy and better stability than others.

In the numerical simulation of a typical cough, when all droplets are integrated on the y-z plane, the positions of droplets at eight different times on the y-z plane ($x = 0$) are shown in Fig. 1. Different colors of droplets represent different diameters, which can be determined with the legend.

2. Sensitivity Analysis of α and β

Sensitivity analysis can identify the importance of input parameters in determining the value of an assigned output variable. There are mainly two approaches to analyzing sensitivity: Local Sensitivity Analysis, and Global Sensitivity Analysis. The former only focuses on the impact of one parameter on the calculation results, ignoring the interaction between parameters. However, the latter considers the influence of the whole parameter space on the results, and it can be further divided into qualitative (or screening) methods and quantitative methods. Qualitative sensitivity measures explore the relative importance of input variables affecting the output variable (e.g., Morris method), and quantitative analysis methods can obtain the quantified contribution of input parameters to the output parameter (e.g., Sobol method).

In general, when there are multiple parameters, qualitative measures are first used to conduct qualitative analysis on the parameters. Then, based on the results, parameters with high sensitivity are selected for further quantitative analysis. Here, only 2 parameters (α and β) need to be considered in the sensitivity analysis process, so the quantitative analysis methods are used directly. The widely used Sobol method is adopted here, which is a variance-based measure (Sobol, 2001). It can deal with nonlinear responses, and measure the effect of interactions in non-additive systems.

The total-effect index is used here as the evaluation identification, which measures the contribution to the output variance of the input variable, including all variance caused by its interactions, of any order, with any other input variables. The higher the value of total-effect, the higher contribution of the input parameter. Note that the sum of each parameters' total-effect will be greater than or equal to 1.00.

According to the previous results of personal exposure risk, the variation range of α is between 0.50 μg and 100.00 μg . Besides, β is set to change from 1.00×10^{-5} to 2.00×10^{-3} through numerous tests. Saltelli's sampling scheme generates

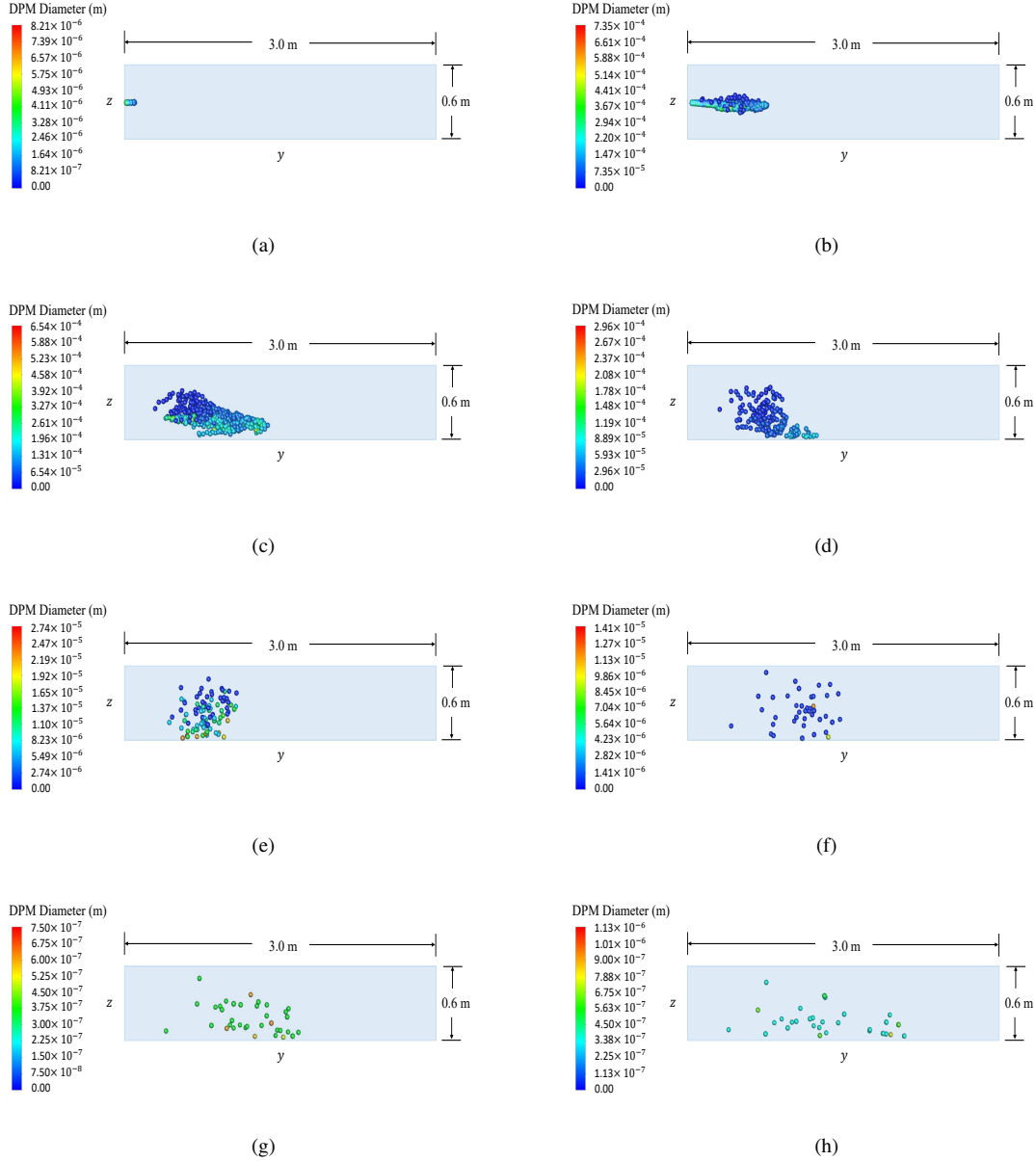


Fig. 1: Evolutions of the cough droplets' diameters in the computational domain on the y-z plane at different times: (a) 0.05 s; (b) 0.15 s; (c) 0.30 s; (d) 0.50 s; (e) 1.00 s; (f) 5.00 s; (g) 10.00 s; (h) 15.00 s.

6,000 samples for each training day. Based on the training set, all days' average total-effect of α and β are 0.986 and 0.269, respectively, where the confidence interval level is 95.00%. Results illustrate that the output C^{new} of our model depends on parameters α and β . Parameter α makes more contributions as expected because it is the cut-line of high exposure risk, which is the key to determining both the C^{risk} and C^{new} .

3. MAE Varies with α and β in Comparison Microscopic-level Models

In Xiao's model and Hernández-Orallo's model, MAE varies with α and β can be seen from Fig. 2 and Fig. 3. Following Equation (12) in the manuscript, there are $\alpha^* = 10.00\text{s}$ and $\beta^* = 1.19 \times 10^{-4}$ in Xiao's Model (corresponding MAE is 7,205.67), and $\alpha^* = 0.60$ MEMs and $\beta^* = 1.41 \times 10^{-4}$ in Hernández-Orallo's Model (corresponding MAE is 7,172.78). Then, based on Equations (9-11) in the manuscript, the prediction results of the testing set by the two microscopic-level models are obtained.

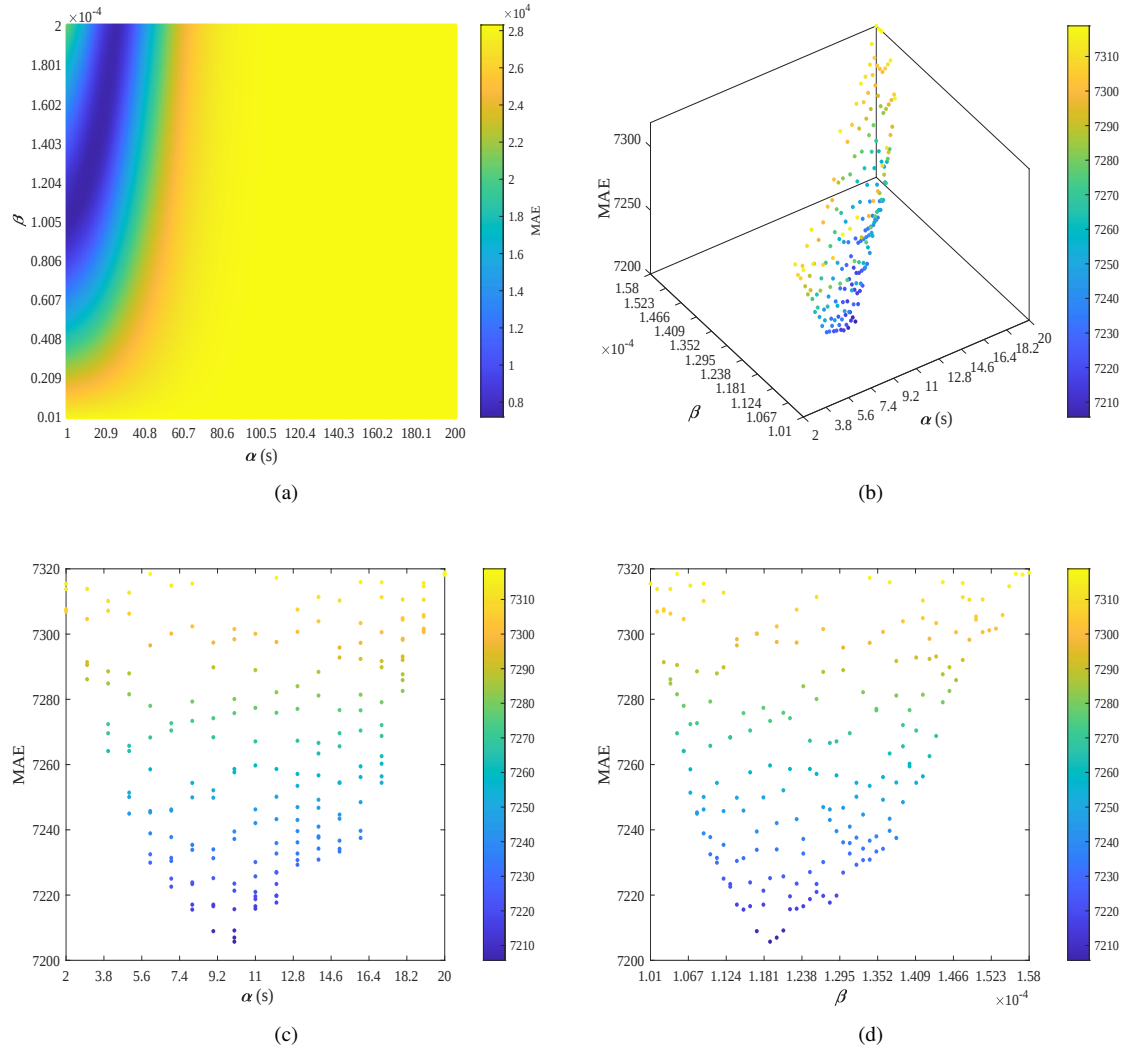


Fig. 2: MAE varies with α and β in Xiao's model: (a) the side view of $\alpha - \beta$ when $\alpha \in [1.00, 200.00]$ s and $\beta \in [1 \times 10^{-6}, 2 \times 10^{-4}]$; (b-d) the front view and 2 side views when MAE are less than 7,320.

References

- Aliabadi, A.A., Rogak, S.N., Green, S.I., Bartlett, K.H., 2010. CFD simulation of human coughs and sneezes: a study in droplet dispersion, heat, and mass transfer, in: ASME International Mechanical Engineering Congress and Exposition, pp. 1051–1060.
- Bi, R., 2018. A numerical investigation of human cough jet development and droplet dispersion. Master's thesis. Dept. Mech.& Matls.Eng., University of Western Ontario, London, Canada.
- Borro, L., Mazzei, L., Raponi, M., Piscitelli, P., Miani, A., Secinaro, A., 2021. The role of air conditioning in the diffusion of sars-cov-2 in indoor environments: A first computational fluid dynamic model, based on investigations performed at the vatican state children's hospital. Environmental research 193, 110343.
- Chao, C.Y.H., Wan, M.P., Morawska, L., Johnson, G.R., Ristovski, Z., Hargreaves, M., Mengersen, K., Corbett, S., Li, Y., Xie, X., et al., 2009. Characterization of expiration air jets and droplet size distributions immediately at the mouth opening. Journal of aerosol science 40, 122–133.
- Duguid, J., 1946. The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. Epidemiology & Infection 44, 471–479.
- Gao, N., Niu, J., 2006. Transient CFD simulation of the respiration process and inter-person exposure assessment. Building and environment 41, 1214–1222.
- Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and computers in simulation 55, 271–280.
- Zhang, L., Li, Y., 2012. Dispersion of coughed droplets in a fully-occupied high-speed rail cabin. Building and Environment 47, 58–66.
- Zhang, Y., Feng, G., Kang, Z., Bi, Y., Cai, Y., 2017. Numerical simulation of coughed droplets in conference room. Procedia Engineering 205, 302–308.

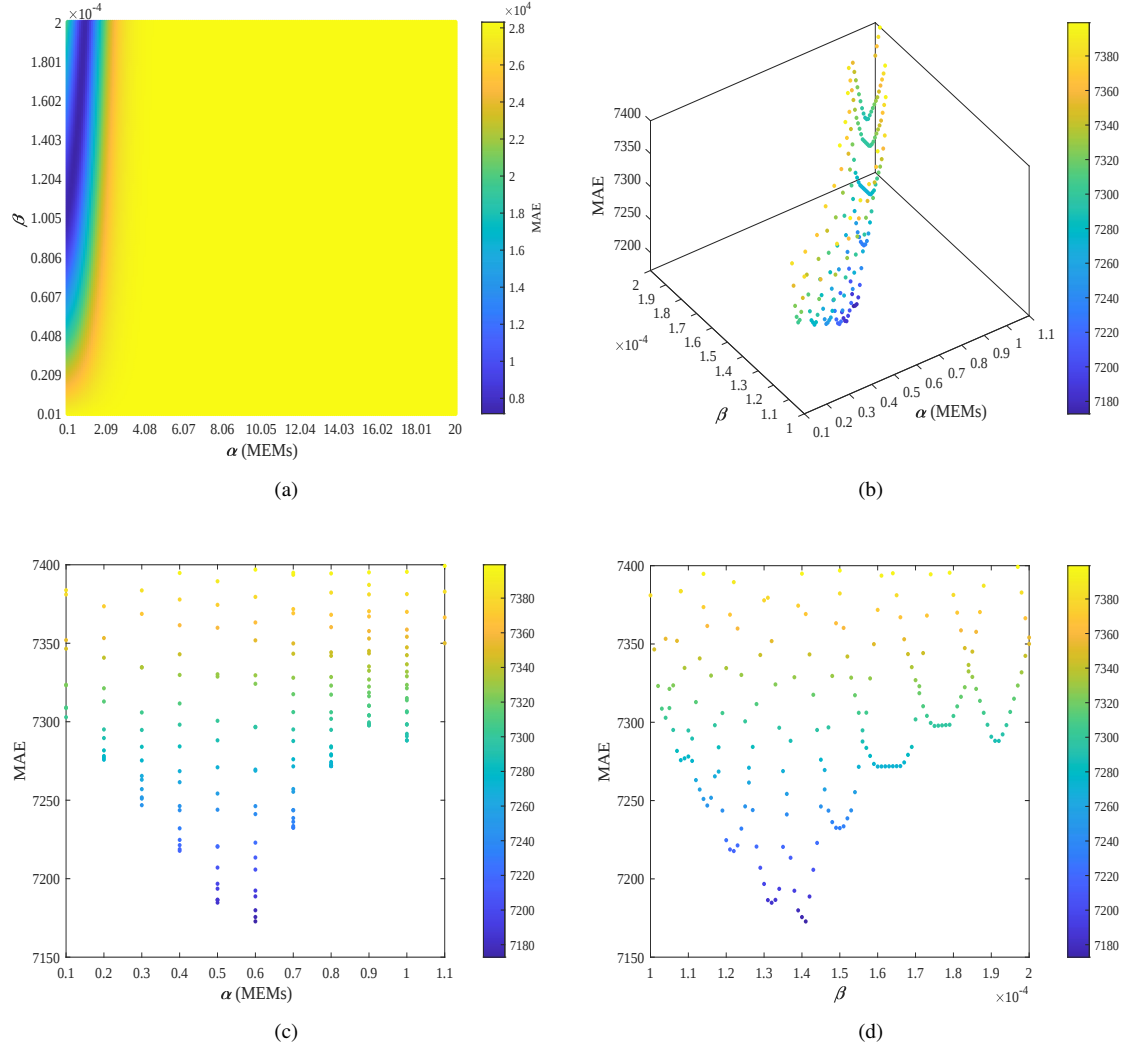


Fig. 3: MAE varies with α and β in Hernández-Orallo's model: (a) the side view of $\alpha - \beta$ when $\alpha \in [0.10, 20.00]$ MEMs and $\beta \in [1 \times 10^{-6}, 2 \times 10^{-4}]$; (b-d) the front view and 2 side views MAE are less than 7,400.