



UvA-DARE (Digital Academic Repository)

Simulating city-level airborne infectious diseases

Mei, S.; Chen, B.; Zhu, Y.; Lees, M.H.; Boukhanovsky, A.V.; Sloot, P.M.A.

DOI

[10.1016/j.compenvurbsys.2014.12.002](https://doi.org/10.1016/j.compenvurbsys.2014.12.002)

Publication date

2015

Document Version

Final published version

Published in

Computers, Environment and Urban Systems

[Link to publication](#)

Citation for published version (APA):

Mei, S., Chen, B., Zhu, Y., Lees, M. H., Boukhanovsky, A. V., & Sloot, P. M. A. (2015). Simulating city-level airborne infectious diseases. *Computers, Environment and Urban Systems*, 51, 97-105. <https://doi.org/10.1016/j.compenvurbsys.2014.12.002>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.



Simulating city-level airborne infectious diseases



Shan Mei^{a,*}, Bin Chen^a, Yifan Zhu^a, M.H. Lees^b, A.V. Boukhanovsky^d, P.M.A. Sloot^{b,c,d}

^a Institute of Simulation Engineering, School of Information System and Management, National University of Defense Technology, Changsha 410073, China

^b Computational Science, University of Amsterdam, Amsterdam 1012, Netherlands

^c Nanyang Technological University, 639798 Singapore, Singapore

^d National Research University of Information Technologies, Mechanics and Optics, St. Petersburg 197101, Russia

ARTICLE INFO

Article history:

Received 18 October 2013

Received in revised form 11 December 2014

Accepted 12 December 2014

Available online 13 January 2015

Keywords:

Geographical Information System (GIS)

Multi-Agent Systems (MAS)

Infectious diseases

Epidemics

ABSTRACT

With the exponential growth in the world population and the constant increase in human mobility, the possible impact of outbreaks of epidemics on cities is increasing, especially in high-density urban areas such as public transportation and transfer points. The volume and proximity of people in these areas can lead to an observed dramatic increase in the transmission of airborne viruses and related pathogens. Due to the critical role these areas play in transmission, it is vital that we have a comprehensive understanding of the ‘transmission highways’ in these areas to predict or prevent the spreading of infectious diseases in general. The principled approach of this paper is to combine and utilize as much information as possible from relevant sources and to integrate these data in a simulated environment that allows for scenario testing and decision support. In this paper, we describe a novel approach to study the spread of airborne diseases in cities by combining traffic information with geo-spatial data, infection dynamics and spreading characteristics. The system is currently being used in an attempt to understand the outbreak of influenza in densely populated cities in China.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

City-level airborne epidemics are constant threats to healthy living. With the fast growth of the world’s population and the constant increase in human mobility, the danger of outbreaks of epidemics is rising. For example, the pandemic influenza A (H1N1), also known as Human Swine Influenza/Swine Flu, caused an international outbreak in Mexico in 2009, and it caused a serious epidemic in China. Indeed, China is highly susceptible to pandemic influenza A (H1N1) due to the large population and high residential density. According to the report by the Ministry of Health of China, the provinces in mainland China had reported 19,589 confirmed cases, 14,348 cured cases, 10 severe cases and several deaths up to 30th September 2009 (Ministry of Health of China, 2009).

In urban areas with high density such as public transportation and transfer points, where people frequently experience close proximity to one another, we observe a striking increase in the transmission of airborne viruses and related pathogens. To correctly model and simulate airborne epidemics, it is critical that the city infrastructure, which causes these hot spots of transmission, be

analyzed and captured in detail. We utilized Geographic Information Systems (GIS) to model the infrastructure of a city that is likely to be threatened by epidemic attacks. GIS facilitates storing, querying and visualizing city infrastructure including roads, regions with diverse functionality, public transportation, and other attributes. To model airborne disease spread, it is important to understand how city infrastructure is used by the inhabitants and acts as the container of infection. We addressed path routing based on city transportation to capture mobility of people and transmissions that occur in localities, especially public transit. This approach was used because in many developing countries such as China, the overly crowded public transportation system can greatly exacerbate airborne epidemics.

Based on geo-spatial information, it is essential to model a local population that dwells in a city with their spatio-temporal behavior. “There is growing recognition that the solutions to the most vexing public health problems are likely to be those that embrace the behavioral and social sciences as key players” (Mabry, Olster, Morgan, & Abrams, 2008). “Human behavior plays an important role in the spread of infectious diseases, and understanding the influence of behavior on the spread of diseases can be key to improving control efforts” (Funk, Salathé, & Jansen, 2010). Obtaining a strong understanding of the ‘transmission highways’ in urban areas regarding the transmission locations and relevant behavior of people is vital to predict and prevent the spread of infectious

* Corresponding author. Tel./fax: +86 731 84574538.

E-mail addresses: Meishan.ann@gmail.com (S. Mei), nudtc9372@gmail.com (B. Chen), stephen.zhuyifan@gmail.com (Y. Zhu), m.h.lees@uva.nl (M.H. Lees), avb_mail@mail.ru (A.V. Boukhanovsky), P.M.A.Sloot@uva.nl (P.M.A. Sloot).

diseases. Therefore, investigating the patterns that are relevant to social contacts, and consequent airborne virus transmissions, is of great importance.

Although the work on social networks and their application in epidemiological research plays an important role, it is almost impossible to directly correlate the transmission between two agents (represented as vertices or nodes) that occur along network edges (links) to people's spatio-temporal behavior. This difficulty is because social networks are substantial abstractions of realistic human social contacts, in which spatial information is retracted and replaced by topological information. The reader is referred to literature (Schneeberger et al., 2004; Sloot, Ivanov, Boukhanovsky, Van De Vijver, & Boucher, 2008; Karlsson, Jansson, Normark, & Nilsson, 2008; Mei, Vijver, Xuan, Zhu, & Sloot, 2010b; Mao & Bian, 2010; Mei, Quax, Vijver, Zhu, & Sloot, 2011) for details.

Existing studies, including EpiSimS (Epidemic Simulation System) (Eubank et al., 2004; Barrett, Eubank, & Smith, 2005; Valle et al., 2006; Valle, Hyman, Hethcote, & Eubank, 2007) which is an extraordinary example, grounded this study. EpiSimS is a discrete-event-driven stochastic simulation model used for investigating the spread of diseases in large urban populations. "The original EpiSimS model was based on the city of Portland, Oregon, in which the simulated movement of more than 1.6 million individuals was constructed. Each individual in the simulation was instantiated according to actual demographic distributions drawn from census data, so that the synthetic population had the correct demographics, e.g. age distribution, household statistics, population density, etc." (Valle et al., 2007) However, EpiSimS does not subtly address how people travel, encounter one another and therefore spread airborne diseases at various spots. This concern is a main focus of this study.

We introduced a novel approach in this paper to study the spread of airborne disease in urban areas by combining traffic information with geo-spatial data, infection dynamics and spread characteristics. We combined and utilized as much information as possible from relevant sources and integrated these data in a simulation environment that allows for scenario testing and decision support.

The remainder of this paper is organized as follows. Section 2 addresses the modeling of a city and the synthesis of the population to support the simulations of airborne disease spread. Section 3 presents the GIS-based implementation and visualization of the simulation environment and performs a tentative experimentation. Last, the paper is summarized, and future directions are outlined.

2. Model

2.1. City modeling

We discuss city modeling from the aspects of city partitions and traffic (road and public transportation) networks.

2.1.1. Regions and sublocations

To construct a synthetic city, we break down major metropolitan areas into regions and sublocations (SLs) that reside inside each region. Regions, or land uses in some studies, are pieces of city land serving various purposes, such as agriculture, commerce, medication and education. Sublocations, affiliated with a specific region, represent a realistic-room-like space where people conduct their daily activities and have social contacts. The types of regions and sublocations considered in this study are subject to whether infections frequently occur inside particular spaces.

Each region is categorized into types of agriculture, residence, hospital, school, university and recreation, according to the main facilities that it provides to people, which is consistent with existing

work (Valle et al., 2006; Valle et al., 2007; Yang, Atkinson, & Ettema, 2008b; Zhang et al., 2012). In this study, we exclusively considered 7 types of regions – housing (HR), office (OR), school (SR), university (UR), medical (MR), recreational (RR) and transportation (TR) according to the general function assignment of city regions. This region partitioning requires GIS files to comprise clearly partitioned land pieces, and these pieces can be mapped to the aforementioned 7 region types, ignoring those regions (e.g., agricultural regions) that contribute less to the spread of diseases. For instance, industrial regions are mapped to ORs. Transportation regions, as special cases, are assumed to consist of only mobile rooms (i.e., compartments of trains, metros and buses).

The union of all regions is not necessarily equal to the entire city area. In other words, 'holes' are allowed on the city map. Whether regions completely encompass the city depends on the completeness of the fundamental data partition in GIS files, which have little impact on epidemic simulations. For example, agricultural regions can be ignored, due to the infection-hampering factors of low density of population and outdoor air conditions. Moreover, only human–human specific transmission is considered in this study, whereas animal–human transmission is excluded.

A region contains a set of sublocations of different classes. For example, a university region (UR) contains office sublocations (offices), residential sublocations (student dormitories and faculty members' homes), classroom sublocations (classrooms, labs and library space), recreation sublocations (cafeteria, clubs, shops, refectories and restaurants) and possibly hospital sublocations. Specifically, the recreational class includes shops, restaurants, cinemas, supermarkets and all other relevant places that provide services for recreation, relaxation or sales of life necessities. In this study, we classified sublocations as housing (HS), office (OS), classroom (CS), patient room (PS), recreational (RS) and transportation (TS). Table 1 lists the classifications of regions and secondary sublocations in detail.

Sublocations are virtually created with 2 dimensions inside each region. Because sublocations are usually beyond the resolution of GIS files, we generated sublocations and assigned the length and width to each of them in meters. During the course of simulations, individuals are attached to one sublocation at a time and conduct activities (working, staying at home, entertaining, etc.) inside. Accordingly, people interact with only those who stay inside the same sublocation, although visually sublocations can overlap. Additionally, each sublocation is characterized as being either indoor or outdoor, conveying different transmission probabilities of viruses inside the space. For many airborne viruses, outdoor conditions such as sunshine, heat, wind blowing and air circulation can lower the infection probability between the infected and the susceptible.

2.1.2. The road and public transportation networks

City traffic routes are modeled as a road network (RN) and a public transportation network (PTN).

The assemblage of roads in a city can be mapped to a road network (RN). Roads, as the transport infrastructure of a city, are composed of road sections and crossings. We establish the road network, denoting crossings by nodes and sections by edges, as shown in Fig. 1. Edges can be unidirectional or bidirectional, indicating that they correspond to one-way or two-way road sections, respectively. A crossing joins several road sections together. The number of sections (usually 2, 3 or 4) that a crossing links indicates the connectivity of the crossing. Accordingly, the entire city roadway can be mapped to a complex network¹ of vast nodes and edges.

¹ In the context of network theory, "a complex network is a network with non-trivial topological features that mostly do not occur in simple networks such as lattices" (Newman, 2003).

Table 1
Regions and sublocations.

Region type	Secondary sublocation (SL) classes
Housing region	Housing SLs: houses, apartments...; Office SLs: community service offices...; Recreational SLs: shops, community gardens...; Classroom SLs: children daycare...
Office region	Office SLs: offices, factory workshops...; Recreational SLs: cafeteria, sport places, restaurants...
School region (for elementary, middle and high schools)	Housing SLs: faculty members' households...; Office SLs: teachers' offices...; Classroom SLs: classrooms, labs and library space...; Recreational SLs: cafeteria, shops, refectories...
University region	Housing SLs: student dormitories, faculty members' households...; Office SLs: teachers' offices...; Classroom SLs: classrooms, labs and library space...; Recreational SLs: cafeteria, clubs, shops, refectories...
Medical region	Office SLs: doctors' offices...; Patient room SLs: medical wards...; Recreational SLs: refectories...
Recreational region	Recreational SLs: all types of retail, service, meal and shop SLs
Transportation region	Transportation SLs: compartments of trains, metros and buses, etc.

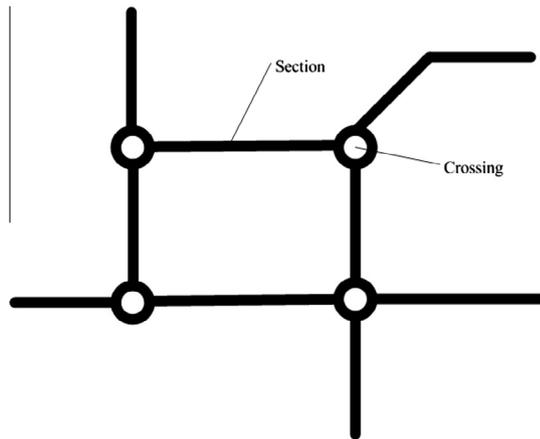


Fig. 1. Road network.

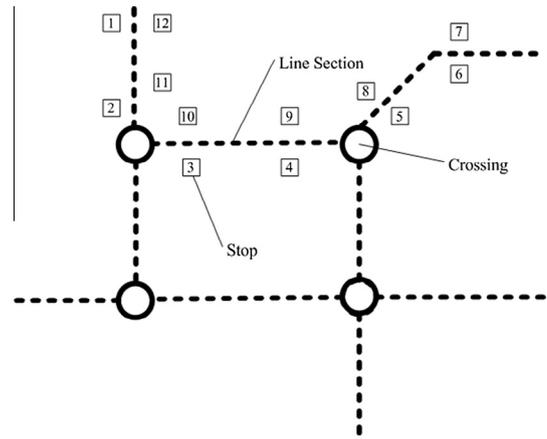


Fig. 2. Public transportation network.

The degree of each node shows the number of neighboring sections that this crossing connects with. Obviously, the degree of nodes in this road network is greater than or equal to 1.

The assemblage of public transportation routes can be mapped to a public transportation network (PTN) in terms of lines and stops. Public transit consists of many bus/tram and metro/train lines along which buses/trams and metros/trains operate frequently during day time. Buses/trams and metros/trains depart every few minutes from the starting stop of a line and move towards the destination stop. People get on or off at each stop along each line. Many lines join at a stop to create a transfer station. Every line is composed of head-tail line sections. Similar to a crossing in the roadway, a line crossing joins several line sections together. We constructed the PTN by denoting crossings by nodes and sections by edges, whereas stops are scattered along both sides of each edge, as shown in Fig. 2. The example in Fig. 2 shows the route of a bus line in one direction that passes 6 bus stops (displayed in square) No. 1–6 along three line sections, and in the other direction, it passes stops No. 7–12 along the same three line sections.

Note that the edges in both the RN and the PTN are unnecessarily associated with the boundary formation of regions. Edges can cross through regions and provide travel entries for nearby people. If an individual needs to travel from one sublocation to another, he/she needs to locate the closest travel point (either a road entry or a bus/train/metro stop) on a nearby edge of the RN or the PTN.

2.1.3. Travel routing

People travel within a city by foot, car, taxi, bike and public transportation. We discuss their travel routes by utilizing the previously constructed RN and PTN. For simplicity, we focused on the

mobility of people inside a single city and did not consider the relatively less common commutes between cities.

For travel by means other than public transportation, we considered routing in two ways. First, for computation simplicity, we assumed that people move along a straight line connecting the start-point (P_s) and the end-point (P_e) for very short distances (for example, when $Distance(P_s, P_e) \leq 3$ km). In this manner, network involvement is avoided to reduce the computation complexity of short-distance routing ($Path = L(P_s, P_e)$ where L denotes the line between the two points). Second, for cases above the short distance threshold, we assumed people move along the shortest feasible path in the RN. The fundamental theoretical achievements in the field of graph theory can help seek the shortest or feasible paths, such as by using the so-called Dijkstra algorithm for node-to-node shortest path computation. Therefore, if $Distance(P_s, P_e) > 3$ km, the travel route result will be the combination of 5 parts (see Fig. 3), i.e., $Path = L(P_s, N_{rs}) + L(N_{rs}, N_{ns}) + SP(N_{ns}, N_{ne}) + L(N_{ne}, N_{re}) + L(N_{re}, P_e)$, where N_{rs} is the nearest point on road (edges in RN) to P_s , N_{ns} is the nearest node (in RN) to N_{rs} , SP computes the shortest path (plotted in gray) between two points in RN (N_{ns} and N_{ne} in this case), N_{re} is the nearest point on road to P_e and N_{ne} is the nearest node to N_{re} . Additionally, the SP result can be substituted by other feasible paths if traffic avoidance needs to be considered.

Traveling by public transit further complicates routing. Looking for the shortest paths in PTN does not necessarily solve the routing problem because buses/metros move along predefined stop-by-stop lines instead of shortest paths. To overcome this obstacle, we designed a breadth-first algorithm as illustrated in Fig. 4. We began with P_s and P_e (i.e., the start-point and the end-point). To compute paths based on public transportation, a set of stops close to P_s , denoted by S_1 , were obtained first. We searched these stops

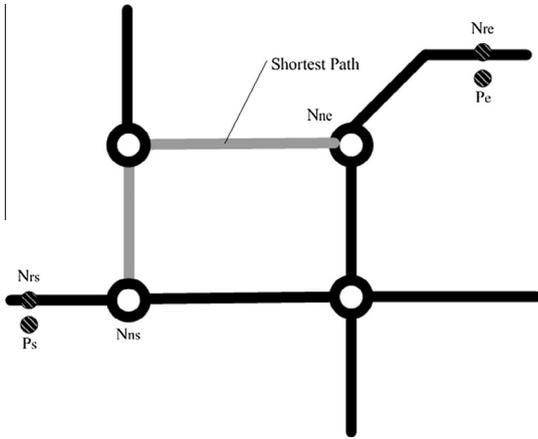


Fig. 3. Travel routing for the shortest path.

within a given distance of P_s , and the resulting circle with center point P_s and given radius is called the *extension area* of P_s . We defined the *extension operation* $Ex(stop, radius)$ to get all the stops inside the circle with center point $stop$ and given *radius* (therefore, $S_1 = Ex(P_s, radius)$). In the same way we can get E_1 , a set of close stops to P_e by utilizing the corresponding extension operation $Ex(P_e, radius)$. The two sets of S_1 and E_1 represent the stops at which people board the transport and the stops from which people alight from for the sake of traveling from location P_s to P_e .

The search operation continues iteratively. We denoted the resulting set of stops from the i th search by S_i and the radius adopted for the i th search by $radius_i$. To conduct an onward search according to predefined line information, we defined the operation of obtaining directly reachable stops of a given stop as $DR(stop)$, which produces a set of stops immediately next to the given stop in all connecting lines. As shown in Fig. 4, for each stop in S_1 (three in total except P_s), we can get its $DR(stop)$ set. As an illustration, in Fig. 4, each stop in S_1 is assumed to have only one directly reachable stop. Therefore, the union of the stops inside the three extension areas (three big circles with dashed perimeter) leads to $S_2 = \bigcup_{stop_2 \in DRS} Ex(stop_2, radius_2)$ where $DRS = \bigcup_{stop_1 \in S_1} DR(stop_1)$. Therefore, The search results starting from both P_s and P_e are given in Eqs. 1 and 2, respectively.

$$S_i = \begin{cases} Ex(P_s, radius_1) & i = 1 \\ \bigcup_{stop_1 \in S_{i-1}} \bigcup_{stop_2 \in DR(stop_1)} Ex(stop_2, radius_i) & i \geq 2 \end{cases} \quad (1)$$

$$E_i = \begin{cases} Ex(P_e, radius_1) & i = 1 \\ \bigcup_{stop_1 \in E_{i-1}} \bigcup_{stop_2 \in DR(stop_1)} Ex(stop_2, radius_i) & i \geq 2 \end{cases} \quad (2)$$

The termination condition for the search was achieved when $\exists i, j, S_i \cap E_j \neq \emptyset$. With minimum transits and less time consumption as two key conditions, the result is the optimal routing from P_s to P_e , involving all the stops on the way and necessary transits. Please note that the radius parameter for each i th search, $radius_i$, is tunable. For example, we can set $radius_1 = 1$ km to search for stops within a radius of 1 km from P_s or P_e and 0.05 km for the rest, implying that transits between lines can only occur if the distance between 2 nearby stops is <0.05 km. If the routing fails, which means that the algorithm results in $S_i \cap E_j = \emptyset, \forall i, j$, people need to resort to other traveling means (e.g., taxi). Arguably, in a maturely developed city, the possibility of routing failure is low and bearable for our simulations.

2.2. The synthetic population

In this study, each person is represented as an agent with individual attributes and behavior based on the Multi-Agent Systems theory which has proven to be suitable for modeling epidemics (Epstein & Axtell, 1996; Reynolds & Dixon, 2000; Heylighen, 2001; Koopman, 2006; Auchincloss & Diez Roux, 2008). To synthesize the population in a given city and investigate how people transmit airborne viruses and respond to epidemics, we needed to outline the attributes of people with respect to epidemiological and sociological characteristics and model people's daily behavior, the occurrence of contacts and subsequent infections.

2.2.1. Attributes and classification of people

The selection of attributes of people is determined by the virological, medical, sociological and demographical data that is available to support our models. We considered age, gender, susceptibility to epidemics, immunity to particular viruses, social status, infection status (susceptible, infected, infectious, treated or cured), Housing SL and Office SL. The Housing SL indicates the place where a person rests, especially at night, including homes

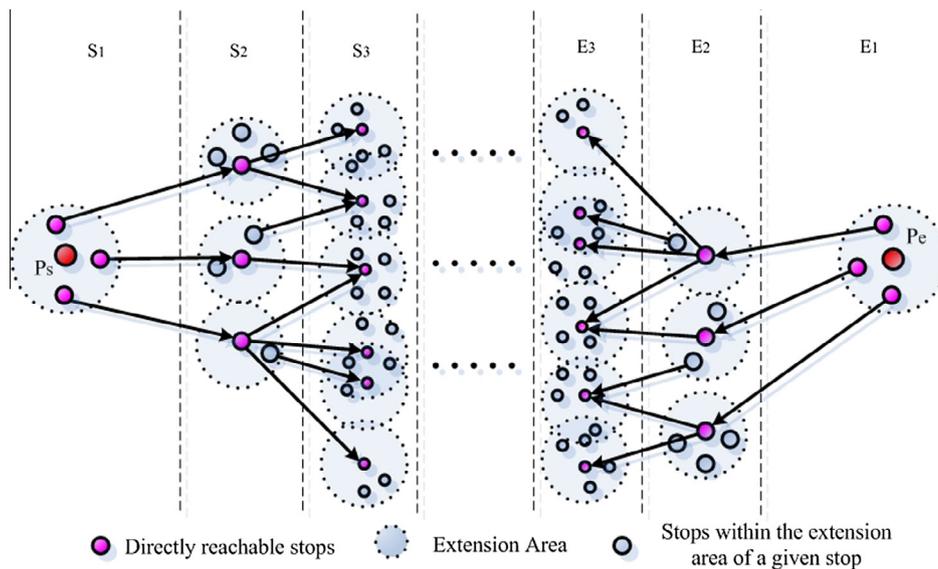


Fig. 4. Travel routing by public transportation.

and dormitories; in contrast, the Office SL indicates the place where a person spends time working (for adults) or studying (for students) during day time. Please note that study activities of students are also regarded as work activities. Initially, assigning values to these attributes for each person depends on the actual statistical distribution or rules deduced from available data. For example, the age distribution can be obtained from national census; the distribution of the distance between one's Housing SL and Office SL for people living in a given city can be estimated based on questionnaires. Assigning one's Housing SL and Office SL was performed by the following procedure: (1) each person is initially attached to a Housing SL, considering household formation rules regarding age, gender, and family size; (2) a distance value between the Housing SL and the Office SL is drawn from a particular distribution; (3) an Office SL is randomly selected on the city map at approximately a distance of the above value away from the Housing SL. Once set, some attributes are kept constant throughout simulations, and others can change during the simulation. For instance, the infection status of a person can be set susceptible (healthy) at the beginning and then change in response to particular events such as infections.

All individuals are classified according to their age structure and lifestyles, as given in Table 2. From here onward, the modeling of people's activities (in Section 2.2.2) is applied to these 5 classifications because it is believed that daily activity patterns are related to individuals' socioeconomic characteristics such as household role, lifestyle and life cycle (Kulkarni & McNally, 2000; Yang et al., 2008b).

2.2.2. Daily agenda

People conduct diverse daily activities in sublocations. The daily activities consist of working, staying at home, relaxing at various recreational places, staying in a hospital after getting infected and traveling between activities, etc. People engage in each activity in a specific class of sublocation (the classification is given in Table 1). Specifically, people work in Office SLs, stay home in Housing SLs, shop/do sport/enjoy entertainment in Recreational SLs, get treatment in Patient Room SLs, study in Classroom SLs and travel in Transportation SLs. The types of activities, the corresponding places where these activities occur and the involved people classes are listed in Table 3, based on research reported in (Valle et al., 2006; Yang et al., 2008b). Additionally, please refer to Table one in (Valle et al., 2007) for the distribution of average duration by activity type. For instance, the average duration of Home activity is 12 h 24 min with a standard deviation of 5 h 8 min, and the average duration of work activity is 3 h 4 min with a standard deviation of 2 h 29 min.

The generation of the daily agenda for an individual is subject to daily activity patterns, depending on the classification of this individual. As for most people whose main activities are working, we generated their daily agenda according to the research in (Bhat & Singh, 2000; Roorda, Miller, & Habib, 2008; Yang, Chen, Wang, &

Wan, 2008a). Six activity patterns are adopted as shown in Table 4 where "*" stands for possible activities other than Home, Work and Travel. Please note that "T" (Travel) is skipped between each pair of neighboring activities. For instance, "HTWTH" stands for the complete form of "HWH". The percentages are taken from (Yang et al., 2008a) based on a survey accomplished in a China city Changsha. In simulations, a Recreation or Medical Care activity will be generated to substitute "*".

After an individual has finished an activity, he/she moves by either public or personal transportation from the current location to another sublocation where the next activity is going to take place. The travel routing, based on either RN or PTN, is computed according to Section 2.1.3. The required travel time is estimated, considering the travel means, the start-point and the end-point. Furthermore, individuals' knowledge of the global epidemic situation, such as the alert phases issued by WHO, and their own infection status, can influence their behavior which is then reflected in the generation of daily agenda. For example, if a person is aware of the severe prevalence of an epidemic or the diagnosis of his/her own infection, he/she likely prolongs his/her stay at home, decreases work time and avoids crowded places such as recreational SLs. Therefore, the parameters for generating activities can be tuned to adapt to different situations.

2.2.3. Infections due to contacts

People encounter and have contacts with others in conducting activities or traveling; thus, they may become infected with airborne viruses when epidemics are prevalent. "The probability that a susceptible person becomes infected during a visit to a sublocation depends on: how many infectious persons co-occupy the room, how long each contact lasts, the type of activity, and the infectiousness category of the infectious person (Eubank et al., 2004)". According to (Valle et al., 2007), σ is the mean number of transmission events per hour of contact between fully infectious and fully susceptible people. In this study for simplicity, we assumed that each infectious person has the same infectiousness regardless of age group classification (Table 2). For events that occur randomly in time, the number of occurrences in a period of time of length t obeys a Poisson probability law with parameter σt (Valle et al., 2007). Then, we obtain the probability of one or more occurrence in time interval t is $1 - e^{-\sigma t}$. If an infectious individual i and a susceptible individual j stay within a distance threshold D^* (tunable for epidemics) from one another in the same sublocation for a particular period of time T_{ij} (recordable in simulations), infection can occur with a probability of $P_{ij} = 1 - e^{-\sigma T_{ij}}$. According to (Chowell, Nishiura, & Bettencourt, 2007; Valle et al., 2007), σ can be estimated based on knowledge of past epidemics. For simplicity, each individual is assumed to wander around inside a sublocation during the stay, and his/her accurate coordinates are obtainable. Therefore, the distance between the two individuals who stay in the same sublocation can be measured to assess whether it is less than D^* .

Table 2
Age structure for people classification.

Classification	Description
Children under 3 years	For children under 3 years, it is simply assumed that they do not have independent activities and remain in their households
Children between 3 and 18 years	Their activity patterns are assumed to be simple: go to daycare or school at school hours and stay inside households at all other times
Adults between 18 and 60 years except college students	They go to work at working places during day time and stay inside households during night. They visit recreational places intermittently
College students between 18 and 25	They go to colleges or universities during the day time and stay inside dormitories at night. They visit recreational places very often
Adults over 60 years	They stay in their households during the day time and stay inside at night. They visit recreational places less often

Table 3
Activity types.

Activity types	Places	Involved people classes
Work (W)	Office SLs, Classroom SLs (for students), Recreational SLs (for salesmen and waiters etc.), Patient room SLs (for doctors and nurses)	All but children under 3 years and adults over 60 years
Home (H)	Housing SLs	All 5 classes
Medical care (M)	Patient room SLs (for patients)	All 5 classes
Recreation (R)	Recreational SLs (for consumers)	The latter 3 classes
Travel (T)	Transportation SLs	The latter 4 classes

Table 4
Activity patterns.

Activity patterns	Percentages (%)
HWH	53.4
HWH*H	10.3
HW*WH	2.7
HWHWH	27.1
HWHWH*H	6.5

2.2.4. Disease progression

Disease progression can be simply described by the following stages: (1) incubation with assumed non-infectiousness, (2) symptomatic period with infectiousness and (3) recovery/death (Mei, Sloom, Quax, Zhu, & Wang, 2010a). For airborne diseases such as pandemic influenza A (H1N1), a susceptible individual can avoid getting infected after vaccination. Therefore, individuals can become immunized by either natural immunization (recovery from previous infection) or vaccination. 3 parameters – $D_{incubation}$, $D_{symptomatic}$ and $D_{vaccination}$ – were introduced to our model, indicating the duration of the incubation stage, the symptomatic stage and the duration that vaccination needs to stimulate immunity, respectively. For example, we can set $D_{incubation} = 1 \sim 2$, $D_{symptomatic} = 1 \sim 7$ and $D_{vaccination} = 7 \sim 21$ days for influenza A (H1N1) (Mei et al., 2010b).

3. Case study of the pandemic influenza A (H1N1) in Changsha City

Based on available GIS data from Changsha city (Capital of Hunan Province, in the middle south of China), we implemented the model and described a simulation environment that can further allow for scenario testing and decision support.

Although the pandemic influenza H1N1 has not broken out in Changsha city – a medium-size city in China, it is vital to establish the capability to simulate a hypothetical airborne-disease outbreak in the city and examine the critical role of public transit and transfer points in disease spread. Accordingly, the model could provide a platform to examine various “What-if” scenarios, and the results can give implications to design effective policy interventions.

To simulate the pandemic influenza A (H1N1), a Message Passing Interface (MPI) based parallel simulation was used to implement the computational experiments. As shown in Fig. 5, the regions in the city, such as hospitals and schools, are grouped and then distributed to different computational nodes. With the help of MPI, agents move among regions in various Local Processes (LPs) in the runtime.

Fig. 6 displays the visualization of the city. Roads are shown as lines, bus/metro stops as squares, buses/metros as stars, and persons as circles. All these objects are stored, queried and manipulated efficiently based on GIS. When simulations are running, stars and circles are moving on the 2-dimensional map, representing that buses/metros operate along lines and people conduct daily activities according to their own agenda.

Based on the aforementioned models and implementation, 6,000,000 agents are simulated in four computational nodes with 48 CPU cores and 128G memory, connected by kilomega networks. It costs approximately 4 h to simulate a 1-month spread of pandemic influenza A (H1N1), if using timesteps of 10 min. We used a baseline infectiousness to depict the probability per hour that a susceptible person will get infected if there is a symptomatic infectious person in the same sublocation at the same time, and we let this baseline infectiousness be 0.00285 transmissions per hour (Valle et al., 2006; Stroud et al., 2006). Therefore, we set $\sigma = 0.00285$ and the distance threshold $D^* = 2$ m (see Section 2.2.3).

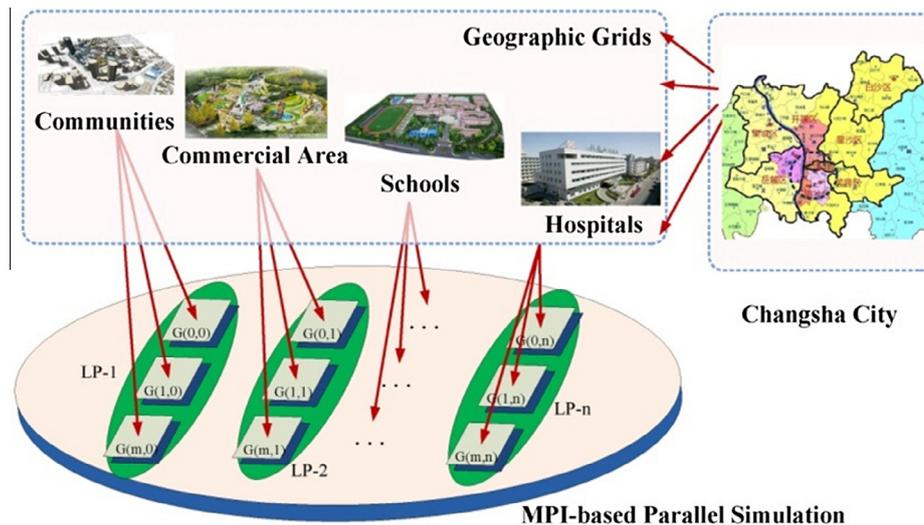


Fig. 5. MPI-based parallel simulation.

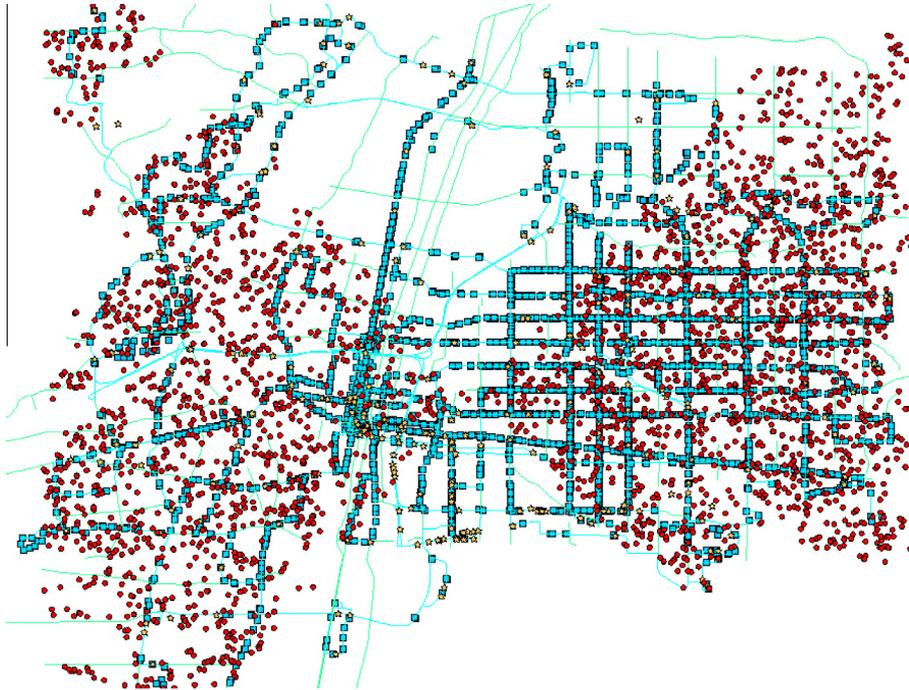


Fig. 6. City visualization.

Table 5
Initial settings of the experiments.

Parameters	Value
Agent count	6,000,000
Initial infected agents	40
Vaccination ratio	10%
Isolation duration	7 days
Close-office duration	7 days
Interventions activated time	4th/5th day

We designed three interventions to control the epidemics: Vaccination Ratio, Isolation and Office Closure. The initial settings of the experiments are listed in Table 5. 10% of the randomly chosen agents are vaccinated and become immune to the disease at particular time points. The duration of Isolation is seven days, which means that the isolated agents will stay at home for seven days

if they have had previous contact with any other infected agent. 5% of the Offices were randomly chosen to be closed for the next seven days when the Office Closure intervention is activated, so that agents who are working at these offices will stay home during the closure period. The results of the experiments are shown in Fig. 7. Each simulation result is an average of 30 realizations.

The curves in Fig. 7 show the count of infected and recovered agents changing with simulation time. We performed two sets of experiments with different intervention activation days. The maximum total infected is approximately 6,800 if the interventions are activated on the 5th day. However, if the interventions are activated one day earlier, the maximum total infected is greatly decreased to 3,100 agents. Moreover, approximately 19% patients get infected at public transit and transfer points (data not shown in Fig. 7). The two curves that plot the count of recovered agents go up much faster during the period 15th–25th days because more infected agents generate more recovered agents. Additionally, the

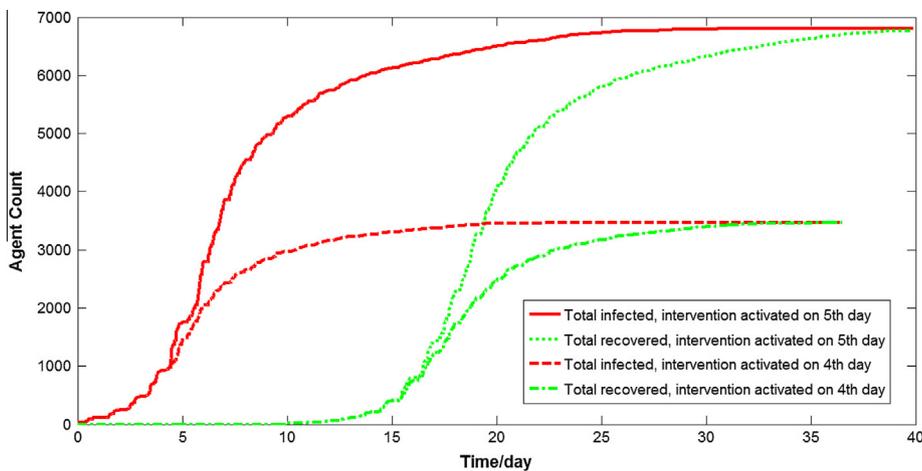


Fig. 7. Total infected and recovered agents changing with time (simulation results only).

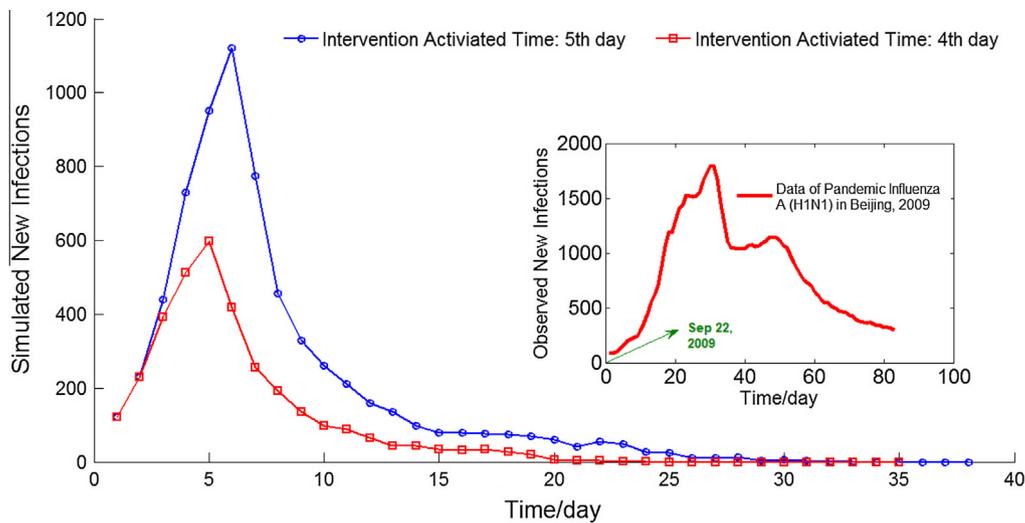


Fig. 8. Simulated new infections in Changsha and observed new infections in Beijing in 2009.

epidemic period is shortened if the interventions are activated earlier (e.g., in this case, the end day of the spread (until all agents recover) if the interventions are activated on the 4th day is 4 days earlier than if the interventions are activated on the 5th day).

To further validate our model, we compared the simulated results in Changsha with the historical data of the outbreak of the pandemic influenza A (H1N1) in Beijing in 2009, as shown in Fig. 8. The two curves with legends show the simulated daily new infections with different intervention activation days (on the 5th day and the 4th day, respectively). The insert shows the observed new infections since September 22, 2009 in Beijing with data from (Xiao-li et al., 2010). Both the simulated and observed new infections shortly reach a peak value and then go down modestly due to preventions. Specifically, the Changsha simulations reach the peaks of 590 and 1,100 on the 5th and 6th day corresponding to the 4th-day activation and the 5th-day activation experiments, and the peak value of the Beijing case is 1,785 on the 30th day. The implementation of a series of measures to prevent and control the pandemic influenza A (H1N1) reduced the cumulative number of laboratory-confirmed cases of pandemic influenza A (H1N1) (Xiao-li et al., 2010). This finding is also consistent with our simulation results. Although Beijing has approximately three times the population of Changsha and the curves in Fig. 8 are plotted at different time scales, the similarity between the curves of the two cases of Changsha and Beijing reveals that our model can capture the underlying principles of airborne disease spread and help predict the impact of control interventions.

4. Conclusion

We have developed a novel system that integrates the most relevant geo-spatial and dynamic information required to assess the potential outbreak of airborne diseases in an urban environment. We combine GIS data with traffic and mobility patterns as well as knowledge on behavioral aspects. The information is represented in a dynamic multi-agent simulation system. Agents travel and conduct activities according to their assigned daily agenda and therefore interact in different sublocations based on our innovative multi-stop travel routing (mix of pedestrian, car, taxi, bike and public transport modes) between any two points in the city. Our simulation records and analyzes the transmission occurring in transportation sublocations, as well as other places, to better understand how seriously airborne diseases spread at public transport and transfer points. This research is superior to other existing

work and can be especially useful to apply to densely populated cities in China.

The system allows for interactively exploring various alternative scenarios to support decision making and prevention, prediction and recovery of an outbreak. In the future, we will focus on calibration and validation of the system based on a more sophisticated empirical study compared to historical epidemiological data.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by China National Scientific Fund (No. 71373282 & No. 91324014), ViroLab (Sloot et al., 2009) and the European DynaNets (www.dynanets.org) grant (EU Grant Agreement Number 233847). The Research was also partly sponsored by a grant from the 'Leading Scientist Program' of the Government of the Russian Federation, under contract 11.G34.31.0019. Boukhanovski and Sloot acknowledge supported by Russian Scientific Foundation proposal #14-21-00137.

References

- Auchincloss, A. H., & Diez Roux, A. V. (2008). A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *American Journal of Epidemiology*, 168(1), 1–8.
- Barrett, C., Eubank, S., & Smith, J. (2005). If smallpox strikes portland. *Scientific American*, 292, 54–61.
- Bhat, C. R., & Singh, S. K. (2000). A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A*, 34(1), 1–22.
- Chowell, G., Nishiura, H., & Bettencourt, L. (2007). Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of The Royal Society Interface*, 4, 155–166.
- Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. The MIT Press.
- Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M., Srinivasan, A., Totoczkal, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429, 180–184.
- Funk, S., Salathé, M., & Jansen, V. A. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: A review. *The Royal Society*, 7(50), 1247–1256.
- Heylighen, F. (2001). The science of self-organization and adaptivity. In L. D. Kiel (Ed.), *Knowledge Management, Organizational Intelligence and Learning, and Complexity*. Oxford: Encyclopedia of Life Support Systems (EOLSS) Publishers. <<http://www.eolss.net>>.

- Karlsson, D., Jansson, A., Normark, B. H., & Nilsson, P. (2008). An individual-based network model to evaluate interventions for controlling pneumococcal transmission. *BMC Infectious Diseases* 8.
- Koopman, J. (2006). Models of infectious disease agent study. *American Journal of Epidemiology*, 163(Suppl.).
- Kulkarni, A. A., & McNally, M. G. (2000). A microsimulation of daily activity patterns. Tech. Rep. UCI-ITS-AS-WP-00-7. <<http://www.its.uci.edu>>.
- Mabry, P. L., Olster, D. H., Morgan, G. D., & Abrams, D. B. (2008). Interdisciplinarity and systems science to improve population health: A view from the nih office of behavioral and social sciences research. *American Journal of Preventive Medicine*, 35(2, Suppl. 1), S211–S224, 0749–3797. DOI: 10.1016/j.amepre.2008.05.018.
- Mao, L., & Bian, L. (2010). Spatial–temporal transmission of influenza and its health risks in an urbanized area. *Computers, Environment and Urban Systems*, 34(3).
- Mei, S., Quax, R., Vijver, D. v. d., Zhu, Y., & Sloot, P. (2011). Increasing risk behaviour can outweigh the benefits of antiretroviral drug treatment on the HIV incidence among men-having-sex-with-men in Amsterdam. *BMC Infectious Diseases*, 11, 118.
- Mei, S., Sloot, P., Quax, R., Zhu, Y., & Wang, W. (2010a). Complex agent networks explaining the HIV epidemic among homosexual men in Amsterdam. *Mathematics and Computers in Simulation*, 80(5), 1018–1030.
- Mei, S., Vijver, D. v. d., Xuan, L., Zhu, Y., & Sloot, P. (2010b). Quantitatively evaluating interventions in the influenza A (H1N1) epidemic on China campus grounded on individual-based simulations. In: *The international conference on computational science 2010 (ICCS 2010)* (pp. 1669–1676). Amsterdam, The Netherlands.
- Ministry of Health of China. (2009). Report on the prevention and control information of influenza A [H1N1] by Ministry of Health of China. URL<<http://www.moh.gov.cn>>.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167–256.
- Reynolds, W. N., & Dixon, D. S. (2000). A general framework for representing behavior in agent based modeling. In: *Complex systems and policy analysis: new tools for a new millennium*. Arlington VA.
- Roorda, M. J., Miller, E. J., & Habib, K. M. N. (2008). Validation of tasha: A 24-h activity scheduling microsimulation model. *Transportation Research Part A*, 42(2), 360–375.
- Schneeberger, A., Nat, R., Mercer, C. H., Gregson, S. A. J., Ferguson, N. M., Nyamukapa, C. A., et al. (2004). Scale-free networks and sexually transmitted diseases: A description of observed patterns of sexual contacts in Britain and Zimbabwe. *Sexually Transmitted Diseases*, 31(6), 380–387.
- Sloot, P. M. A., Coveney, P. V., Ertaylan, G., Müller, V., Boucher, C. A., & Bubak, M. (2009). HIV decision support: From molecule to man. *Philosophical Transactions of Royal Society of London A*, 367(1898), 2691–2703.
- Sloot, P. M. A., Ivanov, S. V., Boukhanovsky, A. V., Van De Vijver, D. A. M. C., & Boucher, C. A. B. (2008). Stochastic simulation of hiv population dynamics through complex network modelling. *International Journal of Computer Mathematics*, 85(8), 1175–1187.
- Stroud, P. D., Sydoriak, S. J., Riese, J. M., Smith, J. P., Mniszewski, S. M., & Romero, P. R. (2006). Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. *Mathematical Biosciences*, 203(2), 301–318.
- Valle, S. D., Hyman, J. M., Hethcote, H. W., & Eubank, S. G. (2007). Mixing patterns between age groups in social networks. *Social Networks*, 29(4), 539–554.
- Valle, S. D., Kubicek, D., Mniszewski, S., Riese, J., Romero, P., Smith, J., Stroud, P., et al. (2006). EpiSimS Los Angeles case study. Tech. Rep., Los Alamos National Laboratory.
- Xiao-li, W., Peng, Y., Zhi-dong, C., Da-jun, Z., Jiang, W., Yi, Z., et al. (2010). Quantitative evaluation on the effectiveness of prevention and control measures against pandemic influenza a (hini) in Beijing, 2009. *Chinese Journal of Epidemiology*, 31(12).
- Yang, M., Chen, X., Wang, W., & Wan, T. (2008a). Commuters' trip generation model based on activity patterns. *Journal of Southeast University (Natural Science Edition)*, 38(3), 525–530.
- Yang, Y., Atkinson, P., & Ettema, D. (2008b). Individual space-time activity-based modelling of infectious disease transmission within a city. *Journal of the Royal Society Interface*, 5, 759–772.
- Zhang, T., Fu, X., Ma, S., Xiao, G., Wong, L., Kwok, C. K., et al. (2012). Evaluating temporal factors in combined interventions of workforce shift and school closure for mitigating the spread of influenza. *PLoS ONE*, 7(3). doi:10.1371/journal.pone.0032203.