Acta Tropica 162 (2016) 188–195

Contents lists available at ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica

Mapping intra-urban transmission risk of dengue fever with big hourly cellphone data

Liang Mao[a,*], Ling Yin[b], Xiaoqing Song[b,c], Shujiang Mei[d]

[a] Department of Geography, University of Florida, Gainesville, FL, USA
[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[c] State Key Laboratory of Information Engineering in Surveying, Mapping, Remote and Sensing, Wuhan University, Wuhan, China
[d] Shenzhen Center for Disease Control and Prevention, Shenzhen, China

ARTICLE INFO

Article history:
Received 24 March 2016
Received in revised form 22 June 2016
Accepted 25 June 2016
Available online 27 June 2016

Keywords:
Dengue fever
Risk assessment
Cellphone tracking data
Urban area
Spatial uncertainty
GIS

ABSTRACT

Cellphone tracking has been recently integrated into risk assessment of disease transmission, because travel behavior of disease carriers can be depicted in unprecedented details. Still in its infancy, such an integration has been limited to: 1) risk assessment only at national and provincial scales, where intra-urban human movements are neglected, and 2) using irregularly logged cellphone data that miss numerous user movements. Furthermore, few risk assessments have considered positional uncertainty of cellphone data. This study proposed a new framework for mapping intra-urban disease risk with regularly logged cellphone tracking data, taking the dengue fever in Shenzhen city as an example. Hourly tracking records of 5.85 million cellphone users, combined with the random forest classification and mosquito activities, were utilized to estimate the local transmission risk of dengue fever and the importation risk through travels. Stochastic simulations were further employed to quantify the uncertainty of risk. The resultant maps suggest targeted interventions to maximally reduce dengue cases exported to other places, as well as appropriate interventions to contain risk in places that import them. Given the popularity of cellphone use in urbanized areas, this framework can be adopted by other cities to design spatio-temporally resolved programs for disease control.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Risk of large and uncontrollable disease outbreaks in urban areas is more likely than ever for many developing countries (WHO, 2016). Accelerated urbanization has concentrated a non-immune population in settings, where main factors contributing to increased virus transmission are present, e.g., high vector and population density (Knudsen and Slooff, 1992). Besides, intra-urban travels of millions of people complicate the spreading pattern of virus, making disease containment more difficult (Frank and Engelke, 2001; Stoddard et al., 2009). Such urban challenges entail an ad hoc risk assessment approach to support effective disease control, such as identifying transmission foci, focusing control efforts to high-risk areas, and managing importation risk (Tatem et al., 2014). The approach development, however, has often been impeded by the data availability at a fine urban scale, e.g., the lack of detailed travel records for a massive population.

In recent years, big cellphone tracking data have been integrated into the risk mapping of disease transmission, given its capability of depicting travel behavior of disease carriers in unprecedented details (Le Menach et al., 2011; Qi and Du, 2013; Tatem et al., 2014, 2009). Millions of cellphone tracking records, when combined with disease incidence or parasite prevalence data, not only reveal existing and potential disease risk, but also help identify 'source' and 'sink' regions of disease dispersion, thus providing new insights into control programs. Still in its infancy, such an integration in the current literature has several weaknesses that hinder its applications in an urban context. First, a majority of studies have been focused on mapping disease risk at national and provincial scales, for instance the recent work in Kenya (Wesolowski et al., 2012), Namibia (Tatem et al., 2014), and Zanzibar region of Tanzania (Le Menach et al., 2011). At these large scales, the intra-urban movements of cellphone users are often neglected. Few efforts have been devoted to mapping risk at a fine urban-scale, where intra-urban travels should be explicitly represented. Second, the cellphone tracking data previously used are often call detail records (CDR) that log the location of cellphone users only when they make telecommunication transactions, such as a phone call or text message. This irregular

* Corresponding author.
E-mail address: liangmao@ufl.edu (L. Mao).

http://dx.doi.org/10.1016/j.actatropica.2016.06.029
0001-706X/© 2016 Elsevier B.V. All rights reserved.

sampling strategy could miss numerous user movements/locations when no transactions happen. The representativeness of CDR data is under question (Zhao et al., 2016). Third, the cellphone tracking technology only registers users to currently working cell towers, whose coverage ranges from meters to kilometers. The exact location of cellphone users is often unknown and can be anywhere within the coverage areas (Chen et al., 2014; Isaacman et al., 2011). Positional uncertainty is thus inevitable and could finally propagate into disease risk maps. Such uncertainty might not be substantial for risk assessments at the national or provincial scale, but could have significant effects at an urban scale. Little attention however has been paid to quantifying this uncertainty in the disease risk assessment.

To address these limitations, this study proposed a new framework for mapping intra-urban disease risk with big hourly cellphone data, taking the dengue fever in Shenzhen city, China, as a case study. In 2014, the Guangdong province in South China was swept by an unprepared dengue fever epidemic, which started with 423 cases in August and reached 42,358 cases by the end of October (Jin et al., 2015). As the second largest city in the province (Fig. 1a), the Shenzhen city had also experienced an abrupt increase of dengue fever, and is now scaling up its dengue control program to fight future outbreaks. Understanding hotspots of dengue transmission and movement of disease carriers is critical to the control program.

2. Methods

2.1. Case-based risk mapping for local acquisition

The transmission risk of dengue fever include local acquisition and wide dispersion through travels. The case-based risk map reflects the likelihood of acquiring dengue fever virus from local existing cases. De-identified data for dengue fever cases from February 2013 to December 2014 were collected by the Shenzhen Disease Prevention and Control Center (Fig. 1b). With their reported home addresses, a total of 489 cases were georeferenced onto the street map as points. A 100 m grid risk map was then created following the random forest classification procedures outlined by Cohen et al. (2013), as described below.

2.1.1. Risk predictive variables

Gridded maps of predictive variables were collected and preprocessed to describe weather, topography, land cover, land use, and population (Table 1). The spatial resolution was set to 100 m to confine the local transmission, given that the flying range of *Aedes albopictus* is observed between 50 and 100 m (Jing et al., 1998). Rainfall and temperature strongly impact the survival and activities of dengue virus carrying mosquitoes (Hii et al., 2012; Wu et al., 2009). Elevation, topographic wetness index, the distance to water bodies, and the proximity to urban green spaces have been demonstrated to influence the risk through their effects on suitability for mosquito breeding and survival (Cohen et al., 2008; Rahman et al., 2006). The population density, workplace density, and road density indicate where mosquitoes and humans may cluster (Bhatt et al., 2013).

Values for each predictive variable were extracted for the point locations of reported dengue cases, as well as 'background' points across the study region. The background points was distributed randomly but proportionately to population density using the Geospatial Modelling Environment v0.6 (Beyer, 2012; Anderson et al., 2006; Phillips and Dudík, 2008). The population-weighted sampling resulted in 7500 background points across the study area and ensured that the territory sampled by the background points was comparable to the locations from which local cases arose (Cohen et al., 2013; Phillips et al., 2009). These background points were assumed to represent absence locations in the subsequent modeling, although they may not indicate the true absence of infection, but instead characterize the environment where people live (Cohen et al., 2013).

2.1.2. Risk mapping with random forest classification

The 'random forest' is a regression tree classification method being recently used to model the likelihood of infection or disease outbreaks. The random forest approach randomly subsets (or bootstraps) the input data to generate many different sample sets. Each sample set is used in a machine learning process to grow a classification tree, i.e., a series of rules to partition the sample data into a set of groups that are as homogenous as possible with respect to the outcome (Liaw and Wiener, 2002). For example, one such rule might differ some dengue case locations from others by a rainfall threshold, while another rule might further split the data based on the population density within a specific range. The input data are repeatedly classified according to many different classification trees, and the final prediction/classification is made by averaging across all of the individual trees (Boström, 2007).

To assess the accuracy of model predictions, 65% of dengue cases were selected at random for model training, and the other 35% were used for model testing. All of the above predictive variables were included into the random forest classification, implemented by a Matlab package (Jaiantilal, 2009), to create a model that predicts the presence of dengue fever case at a particular cell location. The prediction accuracy was examined by the area under the curve (AUC) on a receiver operating characteristic (ROC) graph. The

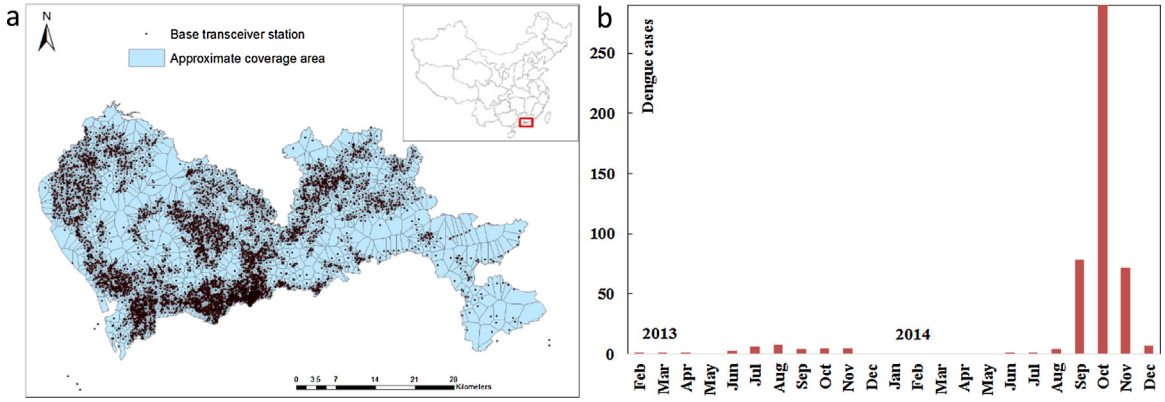

Fig. 1. a) Geographic location and spatial scope of the Shenzhen municipality, with 5941 cellphone towers (base stations) and their approximate coverage areas; b) The number of confirmed dengue cases in Shenzhen city by month in 2013–14.

Table 1
Predictive variables, data sources, and preprocessing.

Category	Variable	Data source	Data description	Preprocessing
Weather	Temperature	Shenzhen Meteorological Bureau	Monthly observations between June and December 2014 at 23 weather stations	Spatial interpolation to 100 m grids of mean, minimum, and maximum temperature
	Precipitation	Shenzhen Meteorological Bureau	Monthly observations between June and December 2014 at 23 weather stations	Spatial interpolation to 100 m grids of mean, minimum, and maximum precipitation
Topography	Elevation	NASA CGIAR-CSI GeoPortal	SRTM 90 m Digital Elevation Database	Resampling to 100 m grid
	Wetness index	Derived from the elevation data	A measure of a location' ability to accumulate and hold water.	Natural logarithm of the ratio between the flow accumulation and the tangent of slope
Population	Population density	The WorldPop project	100 m population count grid	Clipping the grid with Shenzhen city boundary
Land cover	Distance to urban green space	Urban planning, land &resources commissions of Shenzhen municipality	Shenzhen points of interest, including parks and conservation areas.	Applying Euclidean distance function in ArcGIS to generate a 100 m grid
	Distance to Water	Urban planning, land &resources commissions of Shenzhen municipality	Shenzhen points of interest, such as lakes and rivers.	Applying the Euclidean distance function in ArcGIS to generate a 100 m grid
Land use	Road density	Transport commission of Shenzhen municipality	Shenzhen road network	Applying the line density function in ArcGIS to generate a 100 m grid
	Workplace density	Urban planning, land &Resources commissions of Shenzhen municipality	Shenzhen points of interest, including schools, health facilities, restaurants, shops, hotels, commercial buildings, government agencies, etc.	Applying the point density function in ArcGIS to generate a 100 m grid

derived model was then applied to the 100 m gridded datasets of all included predictive variables, generating a predicted risk map across the study area.

2.2. Dengue dispersion through travels of cellphone users

2.2.1. Cellphone tracking data

In addition to the risk of acquiring dengue fever locally around their homes, individuals are also likely to be infected when traveling to other places. To depict their travel behavior, the cellphone tracking datasets were collected through a cellphone network composed of cell towers, and obtained from the Shenzhen Transportation Operation Command Center for research purposes. Each cell tower gives a cellular coverage to an area that is often approximated by a 2-dimensional non-overlapping Thiessen polygon (Song et al., 2010), hereinafter referred to as the coverage area. The radius of coverage areas varies from 200 m to 2 km, dependent on the density of towers (Fig. 1a). A cellphone is connected to a tower when entering the tower coverage area, and meanwhile obtains a current location as the latitude and longitude of the tower. Note that no information is known about the exact position of the cellphone in the coverage area. Therefore, the positional accuracy of cellphone records relies on the size of the tower coverage area.

With this locating technology, the cellphone tracking dataset, referred to as the Signaling System 7 (SS7) dataset, was generated from real-time monitoring of cellphone signals based on the SS7 protocols. Different from the widely used CDR data, this is an active tracking strategy in that the cellphone status, including its time and location, can be recorded at a regular basis, for example, once an hour, no matter whether a telecommunication transaction was made or not. Collected by the China Mobile Telecommunications Company, this dataset tracked 5.85 million anonymized cellphone users on an hourly basis during one weekday in 2012 without major events. With this dataset, a series of activity locations and time were derived for each individual user (Table 2).

2.2.2. Individual's total risk of dengue infection

In combination with the local risk map generated in Section 2.1, the total probability of an individual acquiring dengue virus during a typical day was formulated as a function of the local risk at locations being visited and the activity level of mosquitoes during the time of visit, expressed as Eq. (1):

Table 2
Hourly activities of one cellphone user between 7 a.m. and 1 p.m. recorded in the SS7 cellphone tracking dataset.

User ID	Time (1–24 h)	Cellphone tower ID
bd7a******9d30e	7	2901
bd7a******9d30e	8	2707
bd7a******9d30e	9	1615
bd7a******9d30e	10	1904
bd7a******9d30e	11	1904
bd7a******9d30e	12	1904
bd7a******9d30e	13	1904

$$Pr_i = 1 - \prod_{t=1}^{24} \left(1 - R_{L(t)}\right)^{A(t)} \quad (1)$$

Where Pr_i is the total probability of acquiring dengue fever virus for individual i during a day. t (=1,2,. . .,24) indicates the time of a day in hour. $L(t)$ represents the cellphone tower location being visited during time t. $R_{L(t)}$ is the local transmission risk at location $L(t)$, which can be derived from the case-based risk map (the map derived from Section 2.1). $A(t)$ is the activity level of mosquitoes in blood meal hunting during time t. The parameterization of Eq. (1) is further discussed in following two paragraphs. The detailed derivation of Eq. (1) can be referred to the Supplementary File.

The local acquisition risk $R_{L(t)}$ (in Eq. (1)) can vary dramatically within the coverage area of a cell tower, because an individual's exact location cannot be discerned within this coverage area. Cumulatively, this uncertainty would propagate into the final estimation of Pr_i. This study employed a stochastic approach to address such positional uncertainty. Specifically, if a cell tower coverage area was visited by an individual, this study assumed that every cell location inside the coverage area can be visited with equal chance. The value of $R_{L(t)}$ was then extracted from a randomly selected cell location inside the coverage area once it was needed for calculation.

Table 3
Daily observations on the blood meal hunting behavior of *Aedes albopictu* (Li et al., 2004) and the estimation of hourly activity levels.

t (h)	18	19	20	21	22	23	24	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Vectors	34	69	26	36	38	40	37	34	17	12	15	15	56	39	16	4	17	6	3	12	4	28	22	29
A(t)	0.49	1.0	0.38	0.52	0.55	0.58	0.54	0.49	0.25	0.17	0.22	0.22	0.81	0.57	0.23	0.06	0.25	0.09	0.04	0.17	0.06	0.41	0.32	0.42

For each individual i, the Pr_i was calculated for 1000 permutations to statistically quantify the uncertainty.

With regard to $A(t)$ in Eq. (1), this research adopted results from an observational study (Li et al., 2004) to parameterize the activity of mosquitoes at different hours. The observation was conducted in a city close to the study area and the researchers counted the number of female adult mosquitoes attracted by a blood meal every hour for 24 h (Table 3). In general, the vectors are hunting for blood meals all day long, but most active before and after the sunrise and sunset. All these hourly count numbers were standardized between 0 and 1 by dividing them to the maximum count at 19:00, and the resulting values were used to indicate hourly mosquito activity levels $A(t)$.

2.2.3. Importation risk to coverage areas

With the individuals' total risk Pr_i being computed from Eq. (1), the total importation risk into a cell tower's coverage area was further calculated by aggregating total risk of every visitor, formulated as Eq. (2):

$$IR_C = \sum_{i=1}^{N_C} Pr_i \quad (2)$$

Where IR_C denotes the importation risk to a cell tower C's coverage area, i.e., the number of imported infections per day from other areas of the city. i represent an individual visitor and N_C is the total number of individuals who had visited a coverage area C during a day. Note that the 'importation' here does not refer to the travel of dengue cases from the outside into the city, but the travel into a coverage area from another coverage area within the same city. Since the cellphone users only represent approximately half of the city population, Eq. (2) does not estimate the actual number of imported infections, but the outcomes would be sufficient for statistical comparison.

To account for the positional uncertainty, the 1000 permutations of Pr_i were used to produce 1000 possible values of IR_C for each cell tower. Descriptive statistics of IR_C were then derived for each cellphone tower to quantify such uncertainty, including the mean, median, standard deviation, coefficient of variation (a ratio of standard deviation to mean), and 95% confidence interval. The top 50 areas with the highest importation risk and the top 100 areas with the greatest uncertainty were identified for intervention design

3. Results and discussions

3.1. Random forest model for local risk mapping

Model validation with the testing dataset indicated a strong model performance with a predictive error rate around 5%. The Receiver-operator characteristic (ROC) plot suggested an area under the curve (AUC) of 0.8, which implies a robust model to predict presence and absence of dengue transmission for grid cells (Fig. S1 in Supplementary File). A temporal test also indicated the validity of model to predict over time (Fig. S2 in Supplementary File). The random forest algorithm also estimated the importance of each variable by looking at how much prediction error increases when data for that variable is permuted while all others are left

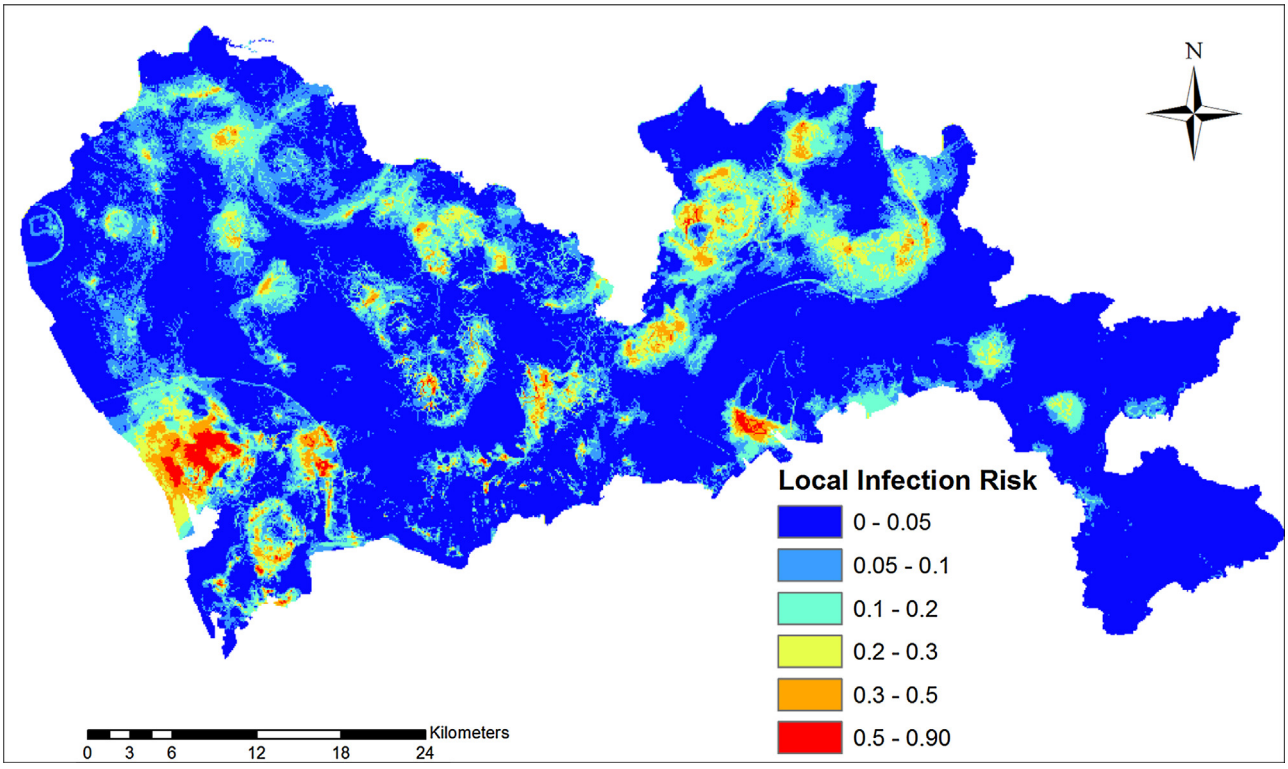

Fig. 2. Predicted local acquisition risk of dengue fever in Shenzhen city at a 100 m resolution, produced from the random forest classification model.

unchanged. The average rainfall, maximum temperature and workplace density were the top 3 most important variables, while the distance to water bodies, distance to green space, and wetness index were the least important (See Table S1 in Supplementary File).

The predicted local risk varies dramatically across the city and exhibit a number of scattered clusters (Fig. 2). There are three notable risk clusters: 1) one situated at the southwestern corner of the city, characterized by a medium size area and very high infection risk; 2) one at the northeast corner covering an extensive geographic area with moderate infection risk; and 3) the last one located in the central south with a small area but high risk. The southwestern cluster may be explained by its highly dense road network and population, while the northeastern cluster is attributable to its characteristically large number of water bodies, such as rivers and reservoirs. The central south area of the city had the highest average rainfall that may contribute to the formation of the third cluster.

3.2. Human mobility analysis

Understanding the local movement of population is critical for preventing disease dispersion in the city. Fig. 3 shows that movement flows with large volume are mostly within short distances. The radius of gyration (ROG), a measure of spatial extent of a cellphone user's daily activities (Song et al., 2010), indicates that 88% of cellphone users circulate within 3 km during a day. However, there is a small proportion of cellphone users (4%) traveling long distances with a ROG greater than 7 km (See Table S2 Supplementary File). Although small in number, these travels can provide sufficient shortcuts for diseases to rapidly spread throughout the entire city, widely known as the small-world effect (Watts and Strogatz, 1998). The hourly tracking records also show the temporal pattern of human movements during a day (See Fig. S2 Supplementary File). The total movements for each hour suggest that the dawn peak of mosquito activities (6:00–7:00 in Table 3) corresponds to a quite low-level human movements, while the dusk peak of mosquito activities (18:00–19:00 in Table 3) coincides with a high level of human movements. The movements decline dramatically after 20:00, even though mosquitoes remain quite active between 20:00 and 24:00. The dusk time should, thus, be paid more attention when aiming to reduce the importation risk.

3.3. Risk of importation and uncertainty

Fig. 4 displays the median importation risk to cellphone coverage areas estimated from 1000 computer permutations. A vast majority of coverage areas were not subject to heavy risk importation (IR < 5000). There is only one cluster of extremely high importation risk, standing out at the southwestern corner. This is very likely to be an outcome from exceptionally high volume of travel influx into this area. Due to less travels into the northwestern and central southern region, the risk of imported dengue fever only reached a moderate level even though the risk of local acquisition remains high there.

The coefficient of variation (CV) reflects the variability of estimated importation risk around its mean value (Fig. 5). For 64% of coverage areas, their estimates only oscillate between ±10% of the mean value (CV < 0.1) and thus are reliable. There are a few coverage areas (137 out of 5941) with extremely high variability (CV > 0.25). The spatial statistic, Moran's I, shows that the CVs also exhibit spatial clustering patterns, in which the high–high clusters are mainly distributed in the central part of the city. Caution is thus needed when using risk estimates in these high–high clusters. Such variability is originated from the positional uncertainty of cellphone tracking data, and then magnified by multiple factors, such as the number of visitors to a coverage area, the number of stops for each individual visitor, and the heterogeneity of local risk in coverage areas.

3.4. Policy implications

The resultant maps of local transmission risk, importation risk, and uncertainty could better inform the design of spatially resolved control polices in the urban area. First, the local acquisition risk map (Fig. 2) suggests three clusters where the application of insecticide

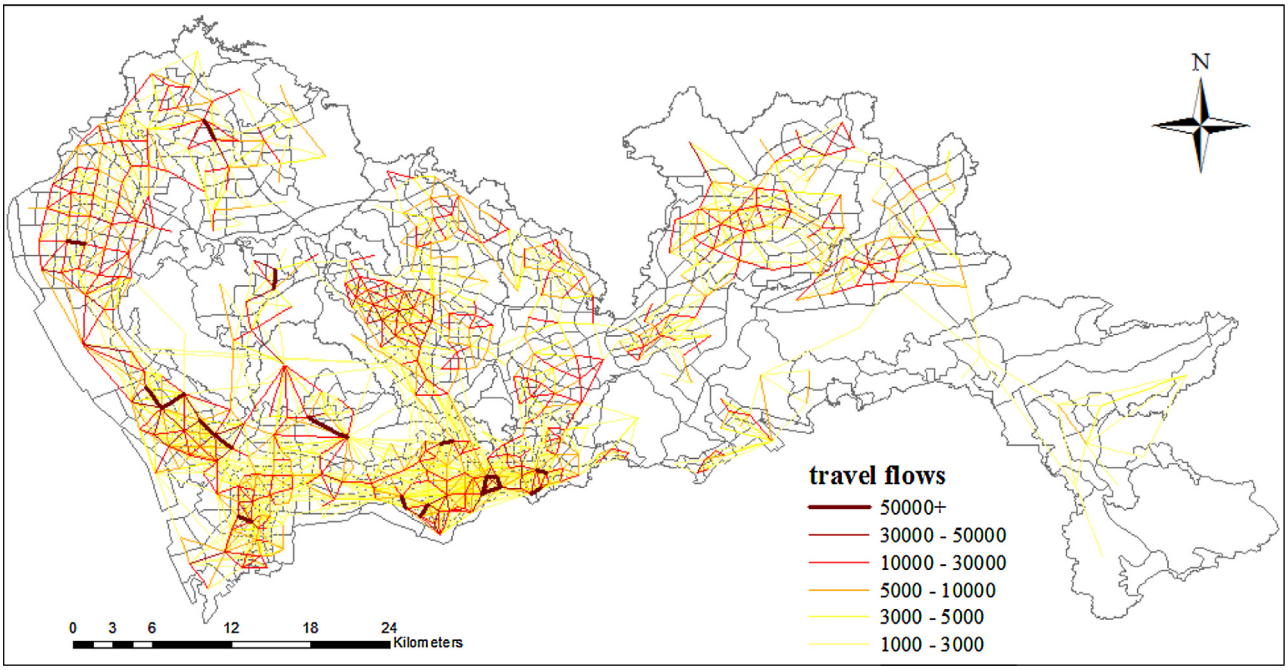

Fig. 3. Total movement flow volume of cellphone users between traffic analysis zones during a day, derived from the hourly tracking records. Travel flows under 1000 people were omitted for content clarity.

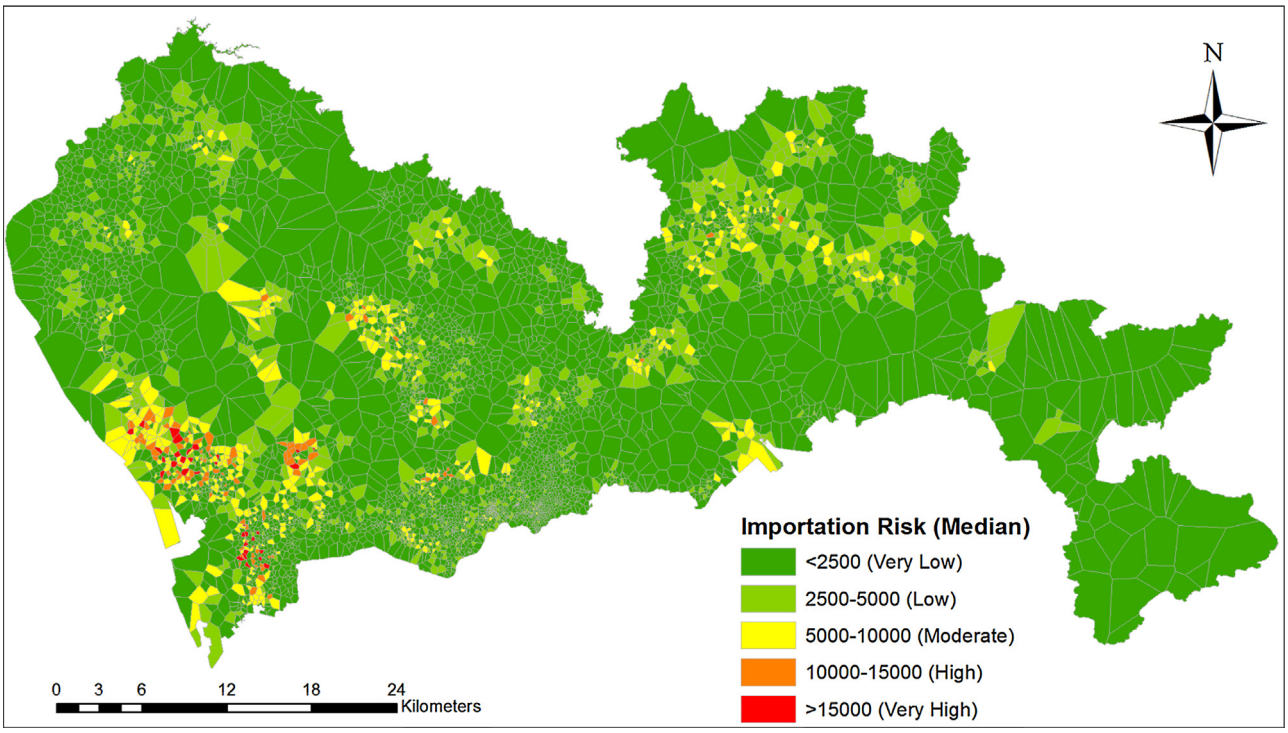

Fig. 4. The median value for the importation risk (No. of imported infections per day) to cellphone coverage areas, estimated from 1000 permutation results.

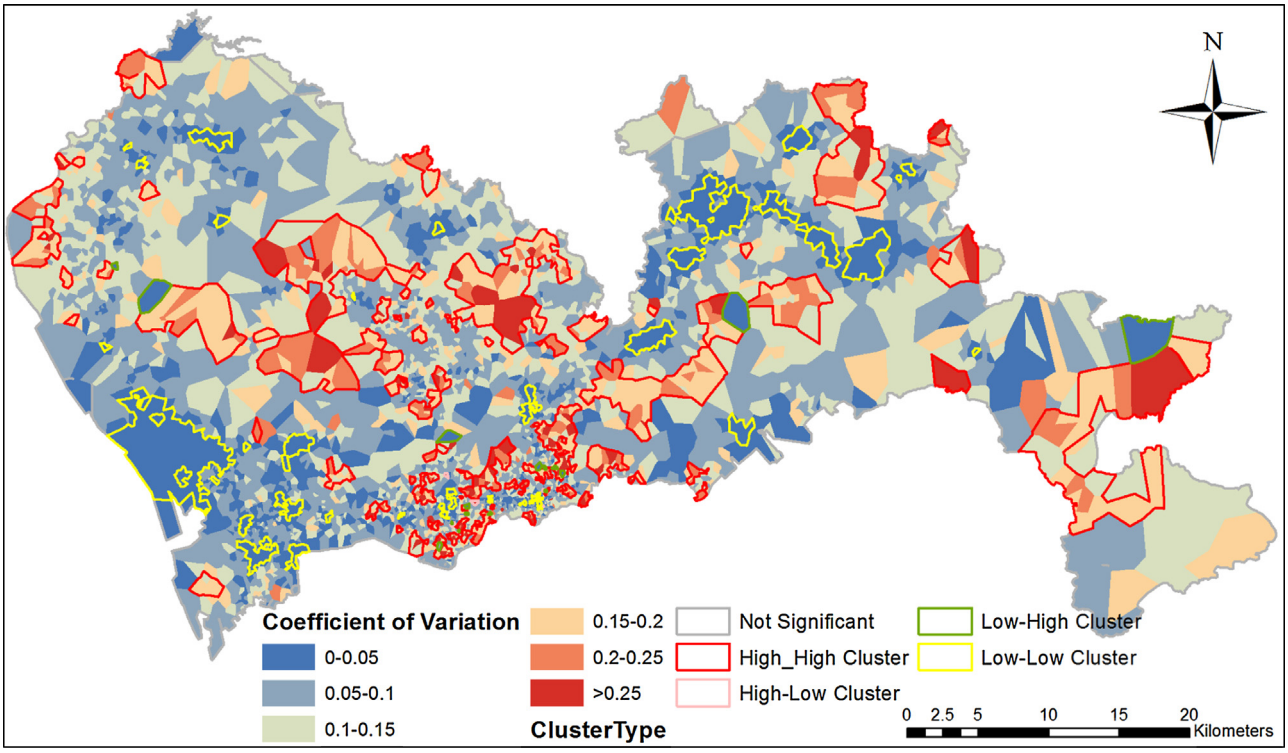

Fig. 5. The coefficient of variation for the importation risk to cellphone towers' coverage areas estimated from 1000 permutation results, and its spatial clusters identified by the local Moran's I index.

should be targeted, in order to achieve effective vector control and minimize local vector transmission. For a further goal of elimination, these efforts would not be sufficient since human movements from other medium/low risk areas can reintroduce dengue fever into areas that had just been cleaned. Here, the coverage areas with top 50 importation risk and with top 100 uncertainty were identified and overlaid to determine where necessary further measures should be implemented (Fig. 6).

It is interesting that the top 50 coverage areas with the highest importation risk are all clustered in the southwestern corner of the city and most of them are small in size (average area = 0.12 km^2 in Table 2), which make it easier for implementing travel restriction or

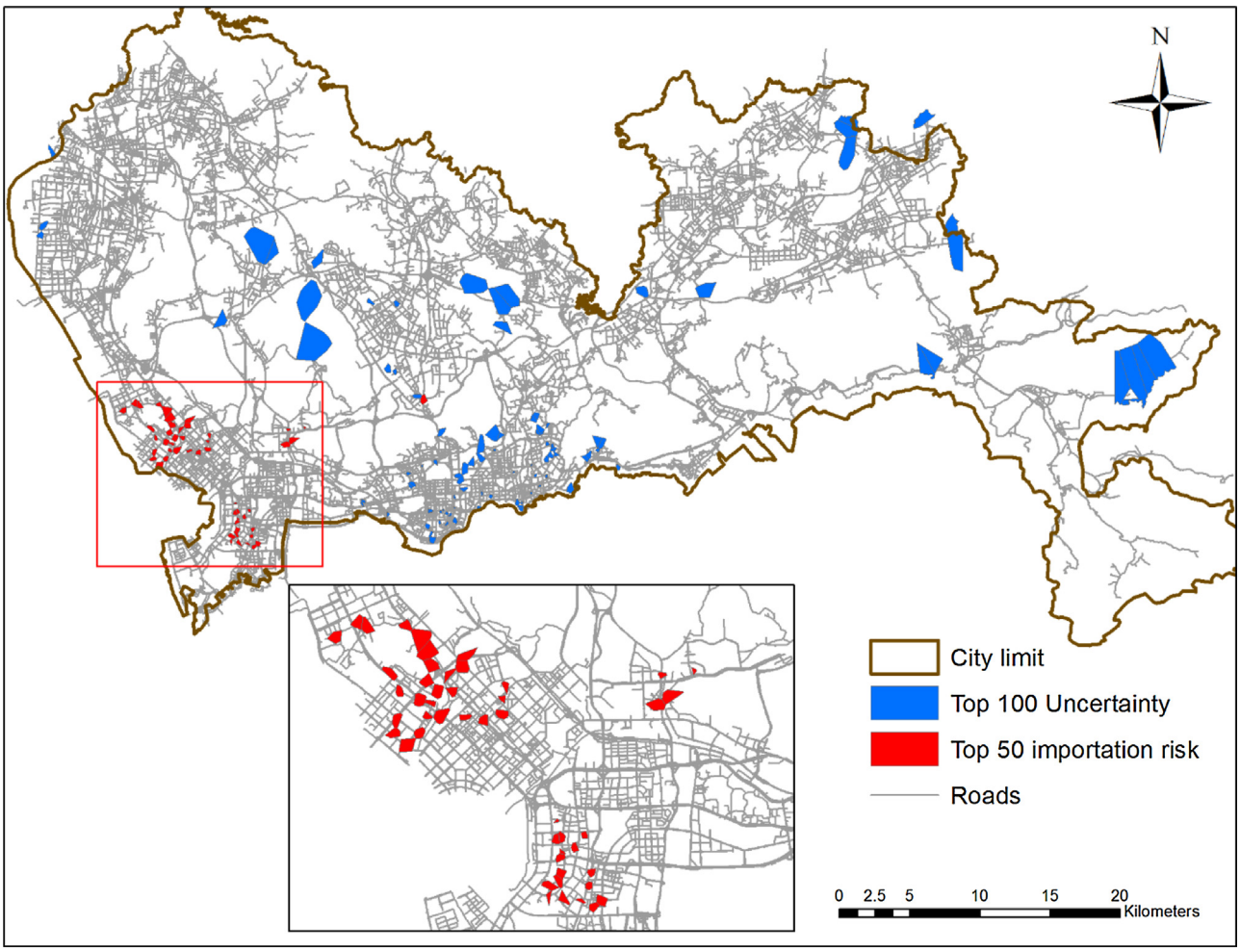

Fig. 6. Coverage areas with top 50 importation risk and top 100 uncertainty to inform policy making.

Table 4
Spatial and demographic characteristics of notable coverage areas.

	Local infection Probability		Imported risk (infections/day)		Population Density (/km^2)		Area (km^2)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. dev.
Top 50 importation risk	0.39	0.20	16190	2618	15150	5603	0.12	0.07
Top 100 uncertainty	0.03	0.05	809	1222	11988	7931	0.53	0.98

regional quarantine. These top risk areas are also located in a highly developed region with millions of visits per day. The restriction of travels into and out of these areas would impose remarkable effects on the dispersion of dengue fever across the entire city. Regarding to the top 100 areas with the uncertain estimation, they are widely distributed across the city and the population density and size vary dramatically (Table 2). Fortunately, there is no overlap between top risk areas and top uncertain areas, thus increasing the reliability of targeting the southwestern corner. As shown in Table 4, most of the highly uncertain areas are associated with low local acquisition risk and low importation risk, even after the variability is considered, and thus are not necessarily to be prioritized in control programs.

3.5. Limitations

There are a few issues regarding to the cellphone data worthy of discussion. First, the SS7 cellphone data is superior to the conventional CDR data because it allows recording human movements in a regular (hourly) fashion. The SS7 data used here, however, only depict one-day period and this study simply assumes that users will repeat the same activity pattern during the course of epidemic. Although questionable, this assumption can be justified by several recent reports that demonstrated a high regularity and predictability of human daily activities (Gonzalez et al., 2008; Lu et al., 2013; Song et al., 2010). Second, the cellphone data only represent approximately half of the city population, and the age, gender, income distributions remain unknown. Selection bias in terms of demographics and geography could be introduced into the estimation, but recent studies have suggested that this is not likely to present a substantial bias (Wesolowski et al., 2013). Third, the mosquito biting activities were derived from only one empirical observation, and the uncertainty had not been reported, thus not being considered in this study. Anyhow, if there were such uncertainty data, e.g., statistical distributions, our stochastic simulation approach can easily accommodate this by randomly sampling $A(t)$ in Eq. (1) and include such uncertainty in the risk assessment. Last, this study had not considered the travels of dengue fever carriers from the outside into the city, due to the lack of such data. No doubt

that these travelers would increase the transmission risk across the city, and thus our research only presents a conservative estimates. All these limitations warrant a future research with more datasets and more sophisticated methods, but they do not override the significance of proposed framework here.

4. Conclusions

This research presents a new framework to integrate big cellphone data into the risk assessment for intra-urban disease dispersion. This framework has first downscaled the current literature to a fine urban context, and explicitly considered the local human movements that are critical to disease spread but have been simplified before. Second, the framework also refines the temporal scale of risk assessment by using hourly based cellphone tracking records. Compared to the irregularly logged cellphone data used before, this new dataset depicts more continuity of local human movements and allows a further consideration of hourly vector activity levels. Third, this framework is the earliest attempt to include the positional uncertainty of cellphone records in the risk assessment. The uncertainty issue may not significantly impact previous assessment at the national or provincial scale, but become important when downscaling to the urban or community scale. The stochastic permutation approach proposed here allows a quantitative measurement of such uncertainty, and thus permits more reliable policy decision making. Given the availability of cellphone records in many countries, particularly developing countries, the framework proposed here can be adopted by other cities to design spatio-temporally resolved programs for disease control.

Authors' contributions

LM conceived of the study, designed the study, developed the random forest model, coordinated the study and drafted the manuscript; LY collected the cellphone tracking data and other major geographic datasets, designed the analysis of these datasets, participated in the design of the study and the completion of the manuscript; XS carried out the cellphone data statistical analysis and importation risk analysis; SM collected the field data of dengue infection cases, and helped draft the manuscript.

Funding

This projected is co-funded by the International Science and Technology Collaboration Project of Guangdong Province (#2014A050503053), the Natural Science Foundation of Guangdong Province (#2014A030313684), and the Basic Research Project of Shenzhen City (JCYJ20140610151856728).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.actatropica.2016.06.029.

References

Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Beyer, H.L., 2012. Geospatial Modelling Environment (Version 07.2.1), URL http://www.spatialecology.com/gme.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., 2013. The global distribution and burden of dengue. Nature 496, 504–507.

Boström, H., 2007. Estimating class probabilities in random forests, machine learning and applications, 2007. ICMLA 2007. Sixth International Conference on. IEEE, 211–216.

Chen, C., Bian, L., Ma, J., 2014. From traces to trajectories: how well can we guess activity locations from mobile phone traces? Transp. Res. Part C: Emerg. Technol. 46, 326–337.

Cohen, J.M., Ernst, K.C., Lindblade, K.A., Vulule, J.M., John, C.C., Wilson, M.L., 2008. Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. Malar. J. 7, 40.

Cohen, J.M., Dlamini, S., Novotny, J.M., Kandula, D., Kunene, S., Tatem, A.J., 2013. Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. Malar. J. 12, 10.1186.

Frank, L.D., Engelke, P.O., 2001. The built environment and human activity patterns: exploring the impacts of urban form on public health. J. Plan. Lit. 16, 202–218.

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns. Nature 453, 779–782.

Hii, Y.L., Zhu, H., Ng, N., Ng, L.C., Rocklöv, J., 2012. Forecast of dengue incidence using temperature and rainfall. PLoS Negl. Trop. Dis.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., Varshavsky, A., 2011. Identifying important places in people's lives from cellular network data. In: Pervasive Computing. Springer, pp. 133–151.

Jaiantilal, A., 2009. Randomforest-matlab. Google.

Jin, X., Lee, M., Shu, J., 2015. Dengue fever in China: an emerging problem demands attention. Emerg. Microbes Infect. 4, e3.

Jing, X., Wang, X., Jiang, Y., 1998. Investigation on the hover ability and spread range of Aedes albopictus. Chin. J. Vector Biol. Control 9, 165–167.

Knudsen, A.B., Slooff, R., 1992. Vector-borne disease problems in rapid urbanization: new approaches to vector control. Bull. World Health Organ. 70, 1.

Le Menach, A., Tatem, A.J., Cohen, J.M., Hay, S.I., Randell, H., Patil, A.P., Smith, D.L., 2011. Travel risk, malaria importation and malaria transmission in Zanzibar. Sci. Rep. 1.

Li, R., Li, J., He, Y., 2004. Daily observations on blood meal hunting behavior of Aedes albopictu. J. Med. Pest Control 20, 27.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

Lu, X., Wetter, E., Bharti, N., Tatem, A.J., Bengtsson, L., 2013. Approaching the limit of predictability in human mobility. Sci. Rep. 3.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197.

Qi, F., Du, F., 2013. Tracking and visualization of space-time activities for a micro-scale flu transmission study. Int. J. Health Geogr. 12.

Rahman, A., Kogan, F., Roytman, L., 2006. Analysis of malaria cases in Bangladesh with remote sensing data. Am. J. Trop. Med. Hyg. 74, 17–19.

Song, C., Qu, Z., Blumm, N., Barabási, A.-L., 2010. Limits of predictability in human mobility. Science 327, 1018–1021.

Stoddard, S.T., Morrison, A.C., Vazquez-Prokopec, G.M., Paz Soldan, V., Kochel, T.J., Kitron, U., Elder, J.P., Scott, T.W., 2009. The role of human movement in the transmission of vector-borne pathogens. PLoS Negl. Trop. Dis. 3, e481.

Tatem, A.J., Qiu, Y., Smith, D.L., Sabot, O., Ali, A.S., Moonen, B., 2009. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. Malar. J. 8, http://dx.doi.org/10.1186/1475-2875-8-287.

Tatem, A.J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D.K., Smith, D.L., Cohen, J.M., Graupe, B., Uusiku, P., 2014. Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. Malar. J. 13, 52.

WHO, 2016. Increased risk of urban yellow fever outbreaks in Africa.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

Wesolowski, A., Eagle, N., Tatem, A.J., Smith, D.L., Noor, A.M., Snow, R.W., Buckee, C.O., 2012. Quantifying the impact of human mobility on malaria. Science 338, 267–270.

Wesolowski, A., Eagle, N., Noor, A.M., Snow, R.W., Buckee, C.O., 2013. The impact of biases in mobile phone ownership on estimates of human mobility. J. R. Soc. Interface 10, 20120986.

Wu, P.-C., Lay, J.-G., Guo, H.-R., Lin, C.-Y., Lung, S.-C., Su, H.-J., 2009. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. Sci. Total Environ. 407, 2224–2233.

Zhao, Z., Shaw, S.-L., Xu, Y., Lu, F., Chen, J., Yin, L., 2016. Understanding the bias of call detail records in human mobility research. Int. J. Geogr. Inform. Sci., 1–25.