

SCIENTIFIC REPORTS



OPEN

Understanding heat patterns produced by vehicular flows in urban areas

Rui Zhu¹, Man Sing Wong¹, Éric Guilbert² & Pak-Wai Chan³

Vehicular traffic has strong implication in the severity and degree of Urban Heat Island (UHI) effect in a city. It is crucial to map and monitor the spatio-temporal heat patterns from vehicular traffic in a city. Data observed from traffic counting stations are readily available for mapping the traffic-related heat across the stations. However, macroscopic models utilizing traffic counting data to estimate dynamic directional vehicular flows are rarely established. Our work proposes a simple and robust cell-transmission-model to simulate all the possible cell-based origin-destination trajectories of vehicular flows over time, based on the traffic counting stations. Result shows that the heat patterns have notable daily and weekly periodical circulation/pattern, and volumes of heat vary significantly in different grid cells. The findings suggest that vehicular flows in some places are the dominating influential factor that make the UHI phenomenon more remarkable.

Urban Heat Island (UHI) is an environmental phenomenon characterized by temperatures in urban areas being significantly higher than in surrounding rural areas. It is one of the major environmental issues caused by urbanization that generates more heat and adverse effects in local climate^{1,2}, increasingly receiving public concern during the recent decades. UHI causes various adverse impacts to society in terms of health risk^{3–5}, public security^{6–8}, and energy consumption^{9,10}. In addition to estimating the magnitude of the UHI intensity, literature has been explicitly studied on their formation mechanism. They gradually reached a consensus that UHI is caused by (i) loss of greenery area over urbanization^{11–14}; (ii) buildings blocking ventilation corridors and accumulating heat^{15,16}; (iii) construction materials with low specific heat capacities absorbing solar radiations or reflecting solar radiations in densely built up areas¹⁷; and (iv) increase of vehicles and growing electricity consumption producing more anthropogenic heat^{18–20}. With regard to anthropogenic heat, vehicles can generate large amount of heat, and heat dispersion can be slowed down because of dense road networks and clusters of high-rise buildings^{21,22}. For example, the highest UHI intensity can be observed in the Kowloon peninsula of Hong Kong along major roads and road intersections, with a significant number of vehicles passing through every day²³. Thus, vehicular traffic should be considered as one of the major causes that increase the severity of UHI especially in mega cities such as Hong Kong. The objective of this study is to develop a quantitative approach that can investigate the influence of vehicle movements on UHI.

Understanding the influence of vehicular flow on UHI requires an accurate estimation of the time-dependent traffic flows, i.e., the number of directional moving vehicles passing through a road network at a given time period. Traffic flow estimation in literature is mainly divided into two categories: microscopic traffic modeling which estimates the behavior of each individual vehicle^{24,25} and macroscopic traffic modeling which describes the characteristics of traffic flows using aggregated parameters such as density and average speed^{26,27}. Microscopic models normally collect sporadic data with spatial information (e.g. GPS locations) to construct the trajectory of each vehicle, which is depicted as a time-series of vehicle locations²⁸. Hence, the models can estimate heterogeneous traffic flows appropriately since origin–destination (OD) matrices can be derived explicitly. This approach is effective to reveal spatio-temporal traffic flow patterns but fails to provide reliable quantitative information of the vehicular traffic, since recording real-time location-based information of every vehicle is still a challenge. A possible solution is through supersampling to extrapolate a system²⁹, which requires a complex maximum entropy model. In contrast, macroscopic models usually utilize data collected from traffic sensors (e.g. traffic

¹Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, 181 Chatham Road South, Kowloon, Hong Kong. ²Department of Geomatics Sciences, Laval University, 1055 Avenue du séminaire, Local 1327, Québec (Québec), G1V 0A6, Canada. ³Hong Kong Observatory, 13A Nathan Road, Kowloon, Hong Kong. Correspondence and requests for materials should be addressed to M.S.W. (email: ls Wong@polyu.edu.hk)

counting stations), and can overcome this problem because data are available for aggregation with steady and frequent updating. Therefore, macroscopic models usually have fewer variables and need fewer properties^{26,27}, which can simplify the computation of heat flux accumulation with higher reliability. A recent study incorporated vehicle-driver behaviors into the macroscopic models³⁰, in which the behaviors were derived from microscopic traffic flows.

For macroscopic models, dynamic traffic assignment (DTA) can be used to estimate traffic flow patterns on the road network. It is because DTA is formed by a principle of travel option which can determine (i) departure times, (ii) origins and destinations, (iii) travel routes of the vehicles, and (iv) a traffic flow module, that it can trigger the propagation of traffic flows over time³¹. To obtain better inherent consistency of dynamic routing behavior, substantial studies have used a time-series of traffic counts to develop time-dependent origin-destination (TDOD) estimation^{32–36}. Specifically, several bi-level optimization models, which contain an upper-level problem to represent trip matrix of the OD demands and a lower-level problem to assign dynamic traffic flows, have been proposed by assuming the OD demand is either stochastic^{32–36} or deterministic^{37–41}. Even though the stochastic approach does not require prior knowledge to adjust functional relations between the parameters, the searching capability is relatively weak and is much time consuming. In contrast, the deterministic algorithms are much more effective to solve the upper-level problem since explicit OD demands can be determined by executing a DTA simulator only several times to aggregate traffic volumes measured from traffic stations⁴².

A cell-transmission-model (CTM), in which vehicles travel through a set of grid cells, has gained wide attentions because of its simplicity to apply the Lighthill-Whitham-Richards (LWR) model to describe traffic dynamics as density-speed related flows^{43,44}. Compared with LWR models, better results were achieved by employing higher-order models^{45,46}. However, these models rely on complex equations and algorithms. To simplify the model and to adapt this model to a big-data computational module at the best convenience, this study will still use a LWR model to investigate a series of parameters for refining the results.

A recent study incorporated a TDOD demand estimation with a CTM to simulate dynamic traffic flows on the road network⁴⁴. In the study, the OD demand estimation was transformed to solve the excess-demand DTA problem in order to optimize three equilibrium conditions (i.e., minimized routing costs, minimized traffic-count matching errors, and maximized OD demand entropies). In comparison, our work develops a similar but simpler and more robust bi-level TDOD demand estimation model that can (i) establish travel behaviors and TDOD trip matrix before iterative computations; (ii) focus on the vehicular flow assignment in each iterative computation; and (iii) achieve the model refinement through systematic post tests instead of instant optimization. In addition, this framework can model weekly dynamic traffic flows and can estimate accumulative heat flux generated from vehicles from weekday to weekend in Hong Kong. In this study, moving behavior of vehicles is firstly modeled by considering the turning probabilities of each vehicle at road intersections. Road networks are then discretized by a series of spatial continuity of homogeneous cells, and discretized networks are generalized as directional dummy paths in group of each cell such that the accessibility of vehicles from each traffic counting station to the cell boundaries and the accessibility from each boundary to its adjacent boundaries can be derived. Further, traffic counting stations are viewed as explicit origins and destinations, and the cells where parking lots and roadside parking areas are determined, so that deterministic TDOD demands can be constructed based on a time-series analysis. The above foundation can significantly reduce the computational load of estimating dynamic traffic flows in each iteration. Thus, trajectories of station-to-boundary and boundary-to-boundary traffic flows can be simulated by a new CTM in-cooperation with the elimination of duplicated counting of vehicles, and for estimation of the traveling speed. To validate the effectiveness of the proposed model, boundary-to-station vehicular flows with the quantitative information are simulated simultaneously for each time period so that it can be used to compare with the data observed from the stations. Lastly, heat flux produced by vehicles can be accumulated in each cell for each constant time period, which is hence used for the correlation analysis with air temperatures observed by automatic weather stations.

Road network used in this study was obtained from OpenStreetMap⁴⁷, and the traffic flow was simulated based on traffic counting data for the entire year of 2015, which were provided by the Transport Department of Hong Kong (TDHK). Source data for different types and numbers of vehicles and their fuel statistics were obtained from TDHK⁴⁸ and Electrical and Mechanical Services Department of Hong Kong⁴⁹ for accumulating heat flux. Hourly air temperatures for the same year acquired from the Hong Kong Observatory were used for the model validation.

Results

Moving probability in the shortest path. This work assumes that all the vehicles follow the shortest paths (SPs) from their origins o to the destinations d . For example, urban population gives higher promises to drive on major roads or highways from their homes to workplaces since these paths are easily accessible and time effective. By assigning higher weights to the major roads, turning probability p_i that a vehicle turns at the road intersection can be calculated as the weight of the road it turns to divided by the sum of the weights it theoretically can turn. However, it is rare for vehicles to turn round directly at the intersection. Thereby, the probability that a vehicle moves from o to d by the SP can be calculated by accumulating all the turning probabilities at all the road intersections as shown in Equation 1.

$$P_{o \rightarrow d} = \prod_{i=1}^n p_i, \left(p_i = \frac{wgt_{turns}}{\sum(wgt_{can_turn}) - wgt_{has_traveled}} \right) \quad (1)$$

Road networks discretization and generalization. Since UHI is a field phenomenon, summarizing heat flux produced by vehicles with enclosed regions is convincing. To maintain the universality, road networks are thus discretized by a set of homogeneous grid cells that *touch*⁵⁰ with each other in the same cell resolution

cr , and each cell $c(h, v)$ is referenced by horizontal and vertical indices starting from the cell $c(1,1)$. Certainly, each $c(h, v)$ has four boundaries in the east, west, south and north directions and each boundary b is noted as e, w, s and n , respectively. In each cell, average length of all the directional SPs that vehicles can move from one boundary to each of the four boundaries is summarized as $D_{b \rightarrow b'} = \frac{\sum \text{length}_{sp}}{\text{num}_{sp}}$, ($b, b' \in \{e, w, s, n\}$). Similarly, average length of the directional SPs that vehicles can travel from one traffic counting station s_p ($p = 1, \dots, q$) to any one of the four boundaries b' can also be generalized and noted as $D_{s_p \rightarrow b'}$. Notably, SP calculation for both $D_{s_p \rightarrow b'}$ and $D_{b \rightarrow b'}$ is only based on the discretized road networks that fall in the single cell. Further, based on the maximum limited speed S for each directional road line segment, the averaged maximum speed for each SP $sp_{s_p \rightarrow b'}$ and $sp_{b \rightarrow b'}$ are calculated as $S_o = \frac{\sum \text{maxSpd}_{line}}{\text{num}_{line}}$, respectively.

Station-to-boundary and boundary-to-boundary accessibilities. A road may intersect with a boundary as a *node*, and there can be several *nodes* $\{nd_i^{(h,v,b)}, (i = 0, \dots, j)\}$ in the same boundary that vehicles can move to from a counting station. Thus, the station-to-boundary (S-B) accessibility can be represented by the *joint probability* that a vehicle theoretically can move from a traffic counting station s_p to a neighbouring boundary b' by the SPs, and the accessibility can be calculated as the sum of all the SP-based moving probabilities to the same boundary as represented in Equation 2. Simultaneously, a vehicle can also move from b to b' in the same cell, and the probability that it moves from b to b' similarly can be represented by the boundary-to-boundary (B-B) accessibility and computed in Equation 3. Since the two accessibilities represent the direct accessing abilities of vehicles between S-B and B-B without crossing any boundaries, both calculations are based on the discretized roads in a single cell. Thus, S-B and B-B dummy paths in each cell are established as $DP_{s_p \rightarrow b'} = \langle D_{s_p \rightarrow b'}, A_{s_p \rightarrow b'}, S_o \rangle$ and $DP_{b \rightarrow b'} = \langle D_{b \rightarrow b'}, A_{b \rightarrow b'}, S_o \rangle$ for traffic flow estimation in the next stage.

$$A_{s_p \rightarrow b'} = \sum_{j=0}^n P_{s_p \rightarrow nd_j} \tag{2}$$

$$A_{b \rightarrow b'} = \sum_{i=0}^m \sum_{j=0}^n P_{nd_i \rightarrow nd_j} \tag{3}$$

Elimination of duplicatedly-counted vehicles. For an easy representation, let a tuple $SW = (t_{wsize}, t_{wstep}, t_w)$ be a temporal *sliding window* that t_{wsize} slides forward at each time instant t_w for each t_{wstep} time period named as the *window step*, such that non-overlapping streaming data can be used for incremental statistics when t_{wsize} equals to t_{wstep} . Let $T = w \cdot t_{wstep}$ denote an incremental time period for the past w window steps. Utilizing traffic counting stations to simulate dynamic vehicular flows is useful for accurately estimating the number of the moving vehicles, which, however, causes a non-ignorable problem that some vehicles may pass through two or several counting stations located in the same statistical cell. This implies that these vehicles will be duplicatedly counted, since all the vehicles that pass through counting stations located in the same cell will be accumulated for each time period t_{wsize} . To eliminate this scenario, a concept of the station-to-station (S-S) accessibility is proposed. For each cell $c(h, v)$ and for each pair of counting stations s_p and s_q located in the same cell, it computes the probability that a vehicle moves from upstream s_p to downstream s_q along the SP. Let $A_{s_p \rightarrow s_q}$ denote the S-S accessibility and equal to the continued multiplication of the turning probabilities when the vehicle turns at the intersections in the SP. Different from the S-B and B-B accessibility calculation, original roads without the cell-clipping are used to calculate the S-S accessibility. Therefore, the generated SP allows a vehicle to cross boundaries of the same cell several times to continuously stay in the original cell during the t_{wsize} . If a number of n vehicles pass through s_p at t_w , it is reasonable to infer that dispersion of these vehicles follow the multinomial distribution and certain number of the vehicles will travel to s_q . The number of the dispersed vehicles generally follows the multinomial distribution as shown in Equation 4 with the *maximum expected value*. When $A_{s_p \rightarrow s_q}$ is obtained, a number of $n \cdot A_{s_p \rightarrow s_q}$ vehicles will arrive at the s_q during the current t_{wsize} in a high possibility. Thus, the duplicated counting is eliminated when the number of the vehicles from the s_p is adjusted to $n \cdot (1 - A_{s_p \rightarrow s_q})$.

$$P(X_1 = n_1, \dots, X_j = n_j) = \begin{cases} n! \prod_{i=1}^j \frac{A_{s_p \rightarrow s_q}^{n_i}}{n_i!}, & \sum_{i=1}^j n_i = n \wedge s_q \in \{s_1, \dots, s_j\} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Figure 1 describes an instance that road networks in Hong Kong are discretized by grid cells in 800 meters, and there are 9 traffic counting stations locating in the cell $c(41,13)$. Correspondingly, detailed description of the B-B, S-B, and S-S accessibilities for the cell are represented in a table directly plotted from the table in PostgreSQL 9.5 (Table 1). It is suggested that the highest probability for vehicles traveling to the downstream east boundary is from the north boundary. Simultaneously, vehicles traveling forward to the east boundary can also start from five counting stations in theory, in which two stations (their IDs are 2209 and 2402) are more likely to lead the vehicles to the east boundary. Investigating the S-S accessibility, vehicles can move from the 2216 counting station to six other counting stations, and half of the vehicles most probably will move to the 2207 station. Therefore, it is intuitively believed that Table 1 can effectively reveal dummy paths characteristics for each statistical cell.

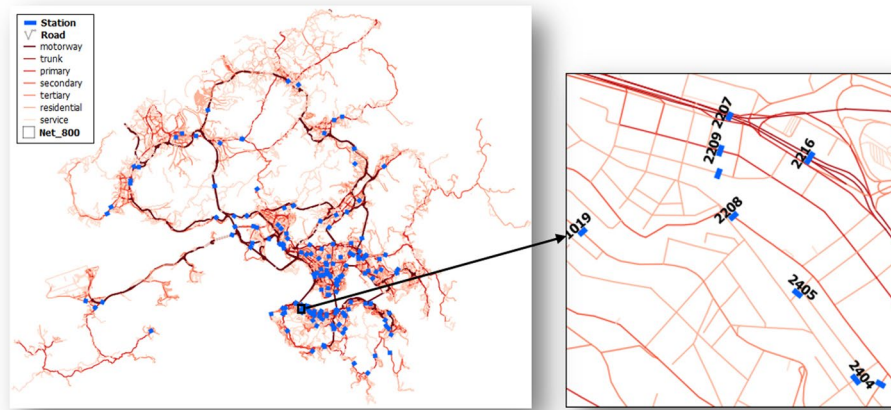


Figure 1. Illustration of the S-S, S-B, and B-B accessibilities. Road networks in Hong Kong are classified into seven types, which are intersected by 169 traffic count stations. Resolution of the discretized grid cell in this example is 800 meters, and the enlarged cell is $c(41,13)$, in which there are nine stations and each one accumulates the passing vehicles in one or more lanes. The figure is originally created by using ArcMap 10.0 (URL link: <http://support.esri.com/Products/Desktop/arcgis-desktop/arcmap/10>), in which the data are in the ESRI shapefile format.

X	Y	b	b'	$A_{b \rightarrow b'}$	$D_{b \rightarrow b'} (m)$	$S_{b \rightarrow b'} (km/h)$
41	13	w	e	0.053712	1147.18	70.00
41	13	n	e	0.927163	907.43	70.39
41	13	e	e	0.584512	787.62	70.00
41	13	s	e	0.045546	1043.20	70.00
X	Y	sp	b'	$A_{s_p \rightarrow b'}$	$D_{s_p \rightarrow b'} (m)$	$S_{s_p \rightarrow b'} (km/h)$
41	13	2207	e	0.500000	526.25	76.00
41	13	2209	e	0.776263	686.77	70.00
41	13	2210	e	0.286170	977.96	70.00
41	13	2216	e	0.500000	308.53	73.33
41	13	2402	e	0.760590	911.37	70.00
X	Y	s _p	s _q	$A_{s_p \rightarrow s_q}$		
41	13	2216	2207	0.500000		
41	13	2216	2208	0.000133		
41	13	2216	2209	0.001678		
41	13	2216	2402	0.000114		
41	13	2216	2404	0.000092		
41	13	2216	2405	0.000046		

Table 1. Accessibility of vehicles for the dummy paths of boundary-to-boundary, station-to-boundary, and station-to-station.

Station-to-boundary traffic flow simulation. This work assumes that all the vehicles start from traffic counting stations and move in the SPs to approach their destinations, which means origins of all the OD pairs are easily determined as a set of $\{s_p\}$. To simulate S-B vehicular flows (i.e. $f_{s_p \rightarrow b'}$), it imports streaming data of $\{n_{s_p}^w\}$ into the road networks at each t_w , which simulates the process that n number of vehicles simultaneously pass through each specific counting station s_p in the bidirectional or unidirectional streets at t_w . These vehicles are supposed to move to the four boundaries of the cell where they are located in, and dispersion of these vehicles to the boundaries also follow the multinomial distribution. In this scenario, the number of the vehicles dispersing to each b' equals to $A_{s_p \rightarrow b'} \cdot n_{s_p}^w$ given the assumption that they always follow the *maximum expected value*. To accumulate the total number of the vehicles $N_{s_p \rightarrow b'}^w$ that are traveling from s_p to b' during the current t_{wsize} , the number of vehicles have already arrived at or passed through b' during the past t_{wsize} have to be determined. This can be achieved by recording the traveled distance information for each set of vehicles during the past $(w - 1)$ window steps as $\{(n_{s_p \rightarrow b'}^i, d_{s_p \rightarrow b'}^i)\} (i = 1, \dots, w - 1)$, which is updated based on an alternative option setting as shown in Equation 5. Thus, the number of vehicles that have passed through b' during the past t_{wsize} can be accumulated and

noted as $N_{s_p \rightarrow b'}^{w-1}$. Hence, $N_{s_p \rightarrow b'}^w$ can be computed in Equation 6. Based on the $N_{s_p \rightarrow b'}^w$ that has already obtained, the distance that each set of the vehicles can travel from s_p to b' during the current t_{wsize} is computed as $l_{s_p \rightarrow b'} = spd_{s_p \rightarrow b'} \cdot t_{wsize}$ and the traveled distance is updated in each iteration in Equation 7.

$$n_{s_p \rightarrow b'}^i = \begin{cases} n_{s_p \rightarrow b'}^i, & \text{if } d_{s_p \rightarrow b'}^i < D_{s_p \rightarrow b'} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$N_{s_p \rightarrow b'}^w = N_{s_p \rightarrow b'}^{w-1} + A_{s_p \rightarrow b'} \cdot n_{s_p}^w - N_{s_p \rightarrow b'}^{w-1} \quad (6)$$

$$d_{s_p \rightarrow b'}^i = d_{s_p \rightarrow b'}^i + l_{s_p \rightarrow b'}, \quad \forall i \in \{1, \dots, w\} \quad (7)$$

Boundary-to-boundary traffic flow simulation. Vehicles which have passed through b' from s_p are in the downstream cell $c(h+k, v+l)$ ($(k, l) \in \{(\pm 1, 0), (0, \pm 1)\}$), sharing the same boundary b' with the $c(h, v)$. These vehicles are assumed to move continuously from b' to the boundary b'' ($b'' \in \{e, w, s, n\}$) of the downstream cell. To simultaneously simulate B-B vehicular flows (i.e. $f_{b' \rightarrow b''}$) for the current t_{wsize} , the number of vehicles that have respectively moved into b' and moved out of b'' during the past t_{wsize} has been determined firstly. Similar to the approach above, the number of vehicles and travel distances of the vehicles that moved from b to b' in each $c(h, v)$ are recorded as $\{(n_{b \rightarrow b'}^i, d_{b \rightarrow b'}^i)\}$ ($i = 1, \dots, w-1$). Vehicles that move into $c(h+k, v+l)$ through b' at t_w can come from s_p and b in the $c(h, v)$ that satisfy $d_{s_p \rightarrow b'}^i \geq D_{s_p \rightarrow b'}$ and $d_{b \rightarrow b'}^i \geq D_{b \rightarrow b'}$, respectively. Thereby, the total number of these vehicles can be noted as $n_{\Rightarrow b'}^w$ and be aggregated in Equation 8. Assuming all the vehicles from b' to b'' also follows the *maximum expected value* in the multinomial distribution for the B-B accessibility, the newly-entered vehicles disperse to each specific b'' can be determined as $A_{b' \rightarrow b''} \cdot n_{\Rightarrow b'}^w$. Simultaneously, the total number of the vehicles that have left $c(h+k, v+l)$ from b' via b'' during the past t_{wsize} can be summarized and denoted as $N_{b' \rightarrow b''}^{w-1}$, meeting the condition that $d_{b' \rightarrow b''}^i \geq D_{b' \rightarrow b''}$. Hence, total number of vehicles $N_{b' \rightarrow b''}^w$ that are moving from b' to b'' in $c(h+k, v+l)$ during the current t_{wsize} is updated in Equation 9. Thus, the distance that vehicles can travel from b' to b'' for each $c(h+k, v+l)$ can be estimated (i.e. $l_{b' \rightarrow b''} = spd_{b' \rightarrow b''} \cdot t_{wsize}$) and total travel distance of all the vehicles that are still in the current cell can be incrementally summarized in Equation 10.

$$n_{\Rightarrow b'}^w = \sum_{p=1}^q n_{s_p \Rightarrow b'}^{w-1} + \sum n_{b \Rightarrow b'}^{w-1}, \quad \forall b \in \{e, w, s, n\} \quad (8)$$

$$N_{b' \rightarrow b''}^w = N_{b' \rightarrow b''}^{w-1} + A_{b' \rightarrow b''} \cdot n_{\Rightarrow b'}^w - N_{b' \rightarrow b''}^{w-1} \quad (9)$$

$$d_{b' \rightarrow b''}^i = d_{b' \rightarrow b''}^i + l_{b' \rightarrow b''}, \quad \forall i = \{1, \dots, w\} \quad (10)$$

Destination estimation. Assuming volume of the vehicles is constant, decreases of number of moving vehicles on the road networks indicate more vehicles in the parking lots. Therefore, temporal distribution of the total number of vehicles parking at the lots is opposite to the number of moving vehicles on the streets. Let p_w denote hourly vehicular-flow percent of counting station s_p at time instant t_w , which is equal to the number of vehicles passing through s_p of an hour divided by the total number of vehicles passing through this counting station of a day. To have unique temporal distribution of the parking-vehicles for each specific grid cell containing counting stations, averaged vehicular-flow percent in each $c(h, v)$ is summarized as p_w^{ave} firstly based on the provided daily statistics so that the daily maximum and minimum values can be determined as p_{max} and p_{min} . Then, the revised temporal distribution of p_w is computed as $p_w^{pk} = (p_{max} + p_{min} - p_w^{ave})$ such that $\{p_w^{pk}\}$ can represent temporal distribution of vehicles parking at the lots in the corresponding $c(h, v)$. However, an overall temporal distribution of the percent of parking-vehicles is calculated for the grid cells that not containing any counting stations, utilizing the records of vehicular-flow percent of all the counting stations. The number of the parking lots in each cell $c(h, v)$ can be summarized and denoted as $\{n_{(h,v)}^{pk}\}$. Assuming vehicles moving in a specific $f_{b' \rightarrow b''}$ in the $c(h, v)$ and some of them parking at a set of lots (i.e. $n_{(h,v)}^{pk}$) located at the $c(h, v)$ at t_w , the number of vehicles parking at each lot follows a homogeneous temporal distribution as what has been established. For each time period t_{wsize} and for each $f_{b' \rightarrow b''}$ of $c(h, v)$ that the vehicles are moving, the number of stopped vehicles can be calculated as $n_{b' \rightarrow b''}^{pk} = n_{(h,v)}^{pk} \cdot p_w^{pk} \cdot n_{b' \rightarrow b''}^w$. Thus, the number of moving vehicles in the $f_{b' \rightarrow b''}$ can be updated accordingly.

Heat emission accumulation. The study simulates the B-B vehicular flows for each t_{wsize} based on the deterministic OD pairs. Origins of these flows are cells of the locations of traffic counting stations, and destinations are cells with the parking lots. For each cell $c(h, v)$ and for each T_j ($T_j = n \cdot t_{wsize}$, where $j \in \{1, \dots, w/n\}$) time period, it summarizes the total length that all vehicles have traveled in Equation 11. To estimate the accumulation of heat emission of these vehicles, fuel combustion of different types of the vehicles have to be determined since it can cause significant variation of heat emissions. However, traffic counting stations are not able to identify the type of vehicles. Previous work has proposed an effective model to compute hourly heat emission released from the fuel combustion of vehicles⁵¹. In this regard, our study refines this model as Equations 12 and 13 to calculate heat emission from vehicles by assuming an evenly distribution of different types of vehicles. Equation 12

Variables	Meaning
$NHC_f (J kg^{-1})$	Net combustion of fuel type f
$\rho_f (kg l^{-1})$	Fuel density
$FE_{ef} (m l^{-1})$	Mean fuel economy for vehicles in type e using fuel type f
fr_y	Ratio of the vehicles consuming fuel type f
er_x	Ratio of the vehicles in type e
$\frac{n \cdot t_{wsize}}{3600} (s)$	Time period for each accumulation
$TL_j^{(h,v)} (m)$	Total length that all the vehicles have traveled in $c(h, v)$ for T_j

Table 2. Specific meaning of the variables used in the equations.

Fuel type	Net heat combustion ($\times 10^6 J kg^{-1}$)	Fuel density ($kg l^{-1}$)
Petrol	46.4	0.75
Diesel	42.8	0.85
LP Gas	50.2	0.58

Table 3. Net heat combustion and fuel density of fuels.

calculates the energy ($J m^{-1}$) that each vehicle in type e can generate when it travels each meter consuming fuel type f , and Equation 13 accumulates the energy (J) that all the vehicles have generated in $c(h, v)$ for T_j . Specific meanings of other variables used in the two equations are represented in Table 2.

$$TL_j^{(h,v)} = \sum_{i=1+(j-1)n}^{jn} (N_{b' \rightarrow b''}^i \cdot l_{b' \rightarrow b''}^i) [m] \tag{11}$$

$$EV_{ef} = \frac{NHC_f \cdot \rho_f}{FE_{ef}} [J m^{-1}] \tag{12}$$

$$E_j^{(h,v)} = \sum_{x=1}^e \sum_{y=1}^f TL_j^{(h,v)} \cdot fr_y \cdot er_x \cdot EV_{ef} \cdot \frac{n \cdot t_{wsize}}{3600} [J] \tag{13}$$

Vehicles in Hong Kong mainly consume three types of fuels, including petrol, diesel, and liquefied petroleum (LP) gas. Their net heat combustion (NHC_p) and the corresponding fuel density (ρ_p) are summarized in Table 3. To compute $E_j^{(h,v)}$, ratio of the fuel type (fr) and ratio of the vehicle type (er) that uses fr are also required, which are computed as combined ratio (r) in Table 4 together with the corresponding mean fuel economy (FE_{ef}). Based on the assumption that different types of the vehicles have uniform distribution in Hong Kong respectively, $E_j^{(h,v)}$ is thus computed when $TL_j^{(h,v)}$ has been obtained during each T_j .

Accuracy test. To test the accuracy of simulated vehicular flows, the number of vehicles collected from counting stations are utilized to compare with those from estimation. Since the model has recorded the number of vehicles $n_{\Rightarrow b'}$ that enter into a new grid cell from the boundary b' during each t_{wsize} benefiting from the B-B vehicular flow simulation, these vehicles are supposed to continuously move forward and arrive at the counting stations $\{s_p\}$ in the current grid cell. This can be achieved by calculating the boundary-to-station (B-S) accessibility that, for each grid cell where counting stations are located in, vehicles move from each b' to each s_p . Using the same method as proposed, the probability that each vehicle can move from a node nd_j intersected by a road and a boundary b' to a counting station s_p can be calculated and noted as $P_{nd_j \rightarrow s_p}$. Theoretically, vehicles tending to travel on the roads have higher weights, and dispersion of the vehicles on the boundary b' can be determined based on the given weights of the roads. In this consideration, the accessibility that vehicles at node nd_j moving from b' to s_p is calculated as in Equation 14. Ultimately, the number of vehicles moving from b' to s_p is obtained as $A_{b' \rightarrow s_p} \cdot n_{\Rightarrow b'}$ for accuracy assessment.

$$A_{b' \rightarrow s_p} = \sum_{j=0}^n \left(\frac{wgt_{nd_j} P_{nd_j \rightarrow s_p}}{\sum wgt_{nd_j}} \right) \tag{14}$$

Weights (wgt) for different types of roads were determined based on the averaged number of total vehicles passing through the corresponding types of roads, and wgt for *motorway*, *trunk*, *major*, *secondary*, *tertiary*, *residential*, and *service* is equal to 31, 39, 14, 6, 7, 3, and 1, respectively. To investigate the influence of different spatial granularities on accuracy of estimation, cell resolutions (cr) in 800 and 400 meters were used for two independent testings. For each cr , two vital parameters were then investigated, i.e., *capacity* (cap) to estimate travel speed and *window size* (t_{wsize}) to determine each simulation time period. Figures 2 and 3 depicts the curves of the correlation

Vehicle type	Fuel type	Total num	Ratio (%)	Mean fuel economy ($m l^{-1}$)
Motor cycle	Petrol	47710	0.066470	24390.24
Private car	Petrol	512808	0.714450	7800.31
Private car	Diesel	5280	0.007356	9259.26
Taxi	LP Gas	18100	0.025217	6341.15
Single deck bus	Diesel	7687	0.010710	4140.79
Double deck bus	Diesel	5786	0.008061	1592.36
Light bus	Diesel	3612	0.005032	5479.45
Light bus	LP Gas	3795	0.005287	3412.97
Light goods vehicle	Petrol	801	0.001116	8547.01
Light goods vehicle	Diesel	69567	0.096922	7633.59
Medium goods vehicle	Diesel	35999	0.050154	3636.36
Heavy goods vehicle	Diesel	6621	0.009224	2617.80

Table 4. Ratio and heat flux for different types of the vehicles. Total numbers of different types of vehicles are summarized based on the source data from Transport Department of Hong Kong⁴⁸, and the mean fuel economy values of the vehicles are calculated based on the source data from Electrical and Mechanical Services Department of Hong Kong⁴⁹. Principle of the categorization is based on the amount of fuels that vehicles consume.

coefficients (R^2) between the observed and the estimated number of vehicles listed in an ascending order in two resolutions, using the data in All-day statistics to significantly reduce the time consumption of the estimation. In particular, when cap , t_{wsize} , and cr are respectively equal to 4500 vehicles, 40 seconds, and 800 meters, refined algorithms and improved execution program in PostgreSQL can reduce the time cost from the original of more than 5 minutes for 480,000 records to 16~20 seconds for 800,000 records to complete each iterative computation in the database. Time cost does not significantly extend with finer spatio-temporal scales ($cr = 400$ meters and $t_{wsize} = 20$ seconds). Assuming that each road at the maximum capacity allows vehicles to pass through the road from 0.5 second per vehicle (i.e. 7200 vehicles per hour) to 1.0 second per vehicle (i.e. 3600 vehicles per hour) with setting t_{wsize} constantly equaling 40 seconds, Fig. 2(a) shows all curves getting similar accuracy trend, and approximate 85% of the $\{R^2\}$ are higher than 0.8. While, the curve with cap equaling to 4500 vehicles has slightly better performance. By fixing this $capacity$, Fig. 2(b) presents the curves when t_{wsize} is ranged from 20 (i.e. 180 iterations per hour) to 40 seconds (i.e. 90 iterations per hour). However, the figure shows that accuracy decreases with an increase of the number of iterative computation. Using same setting of these two parameters, Fig. 3(a) shows better and more stable results with different capacities when cr is 400 meters. For better comparison, cap was also determined as 4500 vehicles to test the accuracy by verifying t_{wsize} since this capacity can still achieve a satisfactory result. Coincidentally, Fig. 3(b) suggests that t_{wsize} equaling to 40 seconds can also achieve a sound result, which even has more high-correlations (25 more stations are higher than 0.9 for R^2 compared with that of cr equaling to 800 meters).

Based on the results above, 40 seconds of t_{wsize} and 4500 vehicles of cap were used to simulate three continuous days of the vehicular flows from weekday, Saturday, to Sunday. This simulation can also reduce computational cost with less iteration. Figure 4 is a comparison of two sets of correlations for all counting stations in three corresponding days derived from 800 and 400 meters of the statistical grid cells. It suggests that both have optimized to high correlations, and all curves grow dramatically to 0.6 at 20 from the very beginning, reach 0.8 at 50, resulting to higher correlations for the rest stations. Particularly for cr equaling to 400 meters, weekday ($dow = 1$) and Saturday ($dow = 2$) have slightly higher correlations than that of cr equaling to 800 meters when $\{R^2\}$ are respectively higher than 0.7 and 0.8. While Sunday ($dow = 3$) for cr equaling to 400 meters generally has lower correlations for the first 45 stations, and it reaches almost the same correlations for the rest compared with that of cr equaling to 800 meters. Thus, heat emission from the estimated vehicles can be accumulated based on two statistical cell resolutions since both can achieve reliable results.

As shown in Fig. 5, the study accumulated hourly heat generated by vehicles for all grid cells that contain road networks for 72 continuous hours from weekday, to Saturday, and to Sunday. The accumulation was respectively based on cr equaling to 800 and 400 meters such that the two sets of the grid cells are spatially coincided with each other. Figure 6 depicts the heat variations for nine enlarged grid cells as represented in Fig. 5. Several distinctive phenomena are found from Fig. 6. First, the total amount of heat generated by vehicles continuously decreases from weekday, Saturday, to Sunday. Second, heat accumulates dramatically since 7 am, reaches the peak at 9 am, and continuously decreases to a valley point at 2 pm, followed by the second peak at 7 pm during the weekday. However, this phenomenon is not significant on Saturday and Sunday. Third, beginning time of rapid heat accumulation in the morning is postponed from 7 to 9 am and the slopes turn to be more gentle from weekday to Sunday. Fourth, the minimum amount of heat occurs during the Sunday night and the early morning of Monday. All of these indicate that periodical human activities in work-and-off days in mega cities such as Hong Kong can influence the micro-environment in the urban areas considerably. i.e., volume of the heat varies through time following the spatio-temporal distribution of moving vehicles. Comparing Fig. 6(a) and (b), the total amount of heat for nine grid cells in each day are almost identical. However, when investigating at each individual grid cell, heat in the $c(42, 13)$ and $c(41, 13)$ shall be overestimated while heat in the $c(40, 11)$ and $c(40, 13)$ are supposed to be underestimated in the 800 m grid cells compared with that when cr equals 400 m. The reason is that area of

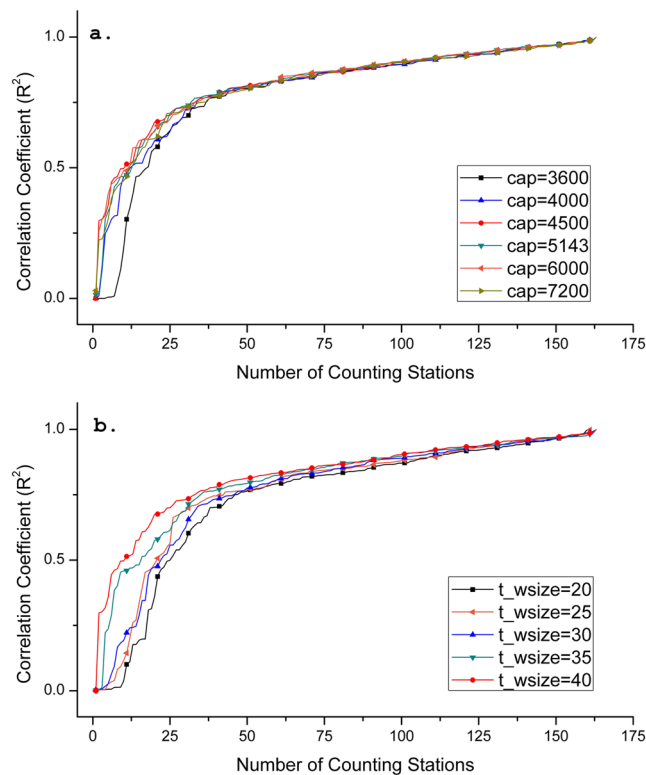


Figure 2. Accuracy test for the vehicular-flow estimation investigating two fundamental parameters when cell resolution is 800 meters. **(a)** Correlation coefficients (R^2) between measured and estimated vehicles for all the counting stations when the *capacity* is set from 3600 and 7200. **(b)** Correlation coefficients (R^2) between measured and estimated vehicles for all the counting stations when t_{wsize} is set from 20 to 40 seconds.

the grid cell in 800 m is four times larger than that in 400 m, such that more numbers of vehicles are alternatively determined as staying in the current cell or arriving at a new cell for each iterative vehicular flow estimation. Thus, it is possible to cause over estimation of moving vehicles allocating into current grid cell or its surrounding grid cells. This analysis also suggests that heat summarized by 400 m grid cells has better quantitative coherence in spatial context.

Air temperature measured by a weather station located in rural area of the New Territories of Hong Kong was used as the reference of rural temperature, so that relative temperatures (i.e. air temperatures measured by other automatic weather stations minus the rural temperature) can be represented as the UHI intensity⁵². Pearson correlation coefficient (PCC denoted as r) is used to present the linear correlation between the UHI intensity and the generated heat in the same statistical grid cell. T-test ($p < 0.01$) is also used to explore the significance of correlation. Results show strong positive correlations (when $r \geq 0.7$) with high confidence (when $t \geq 2.819$) for the two cell resolutions, i.e., 14 out of 17 ($cr = 400$ m) and 23 out of 27 ($cr = 800$ m) statistical grid cells satisfy this condition. More specifically, Fig. 7 presents hourly UHI intensity of each grid cell, representing locations of weather stations and road networks, averaged heat in the corresponding grid cells, and their r and t values with the grid-based locations in the horizontal axes.

As shown in Fig. 8, heavy traffic with slow movements or even congestion frequently occurring in the grid cells $c(42, 12)$ and $c(44, 17)$ should be the main or dominant cause of UHI, because these heavy traffic not only produced the maximum/huge volume of heat but also had strong correlations with UHI intensities (Fig. 7(a)). High-rise buildings surrounded by the roads in the two grid cells are supposed to increase the severity of UHI. Similarly, traffic hubs which a large number of vehicles passing through $c(48, 20)$ (a green-land and impervious mixed area) and $c(78, 39)$ (an impervious open area) can also be one of the main causative factors since the same factors can be derived in Fig. 7(a) and (b). In contrast, moving vehicles in $c(71, 42)$ (Fig. 7(b)), as the only major heat resource in the seaside, had very strong and reliable correlation with the UHI intensities. Even though they produced very small amount of heat, UHI intensities reached as high as 7 degree Celsius. The same phenomenon is also found in the green area ($c(36, 41)$ and $c(45, 11)$ in Fig. 7(a), and $c(90, 21)$ in Fig. 7(b)) where there are less buildings and vehicles are considered to be the only major heat resource. In these two cells, strong and reliable correlations were determined and the heat was insignificant, while intensities still reached up to 7.5 degree Celsius. A similar but different situation occurred in $c(44, 8)$ in Fig. 7(a) and $c(87, 15)$ in Fig. 7(b) that vehicles accumulated notable amount of heat in the seaside and influenced UHI phenomenon obviously, which is the same as the grid cell $c(96, 40)$ (Fig. 7(b)) across green area.

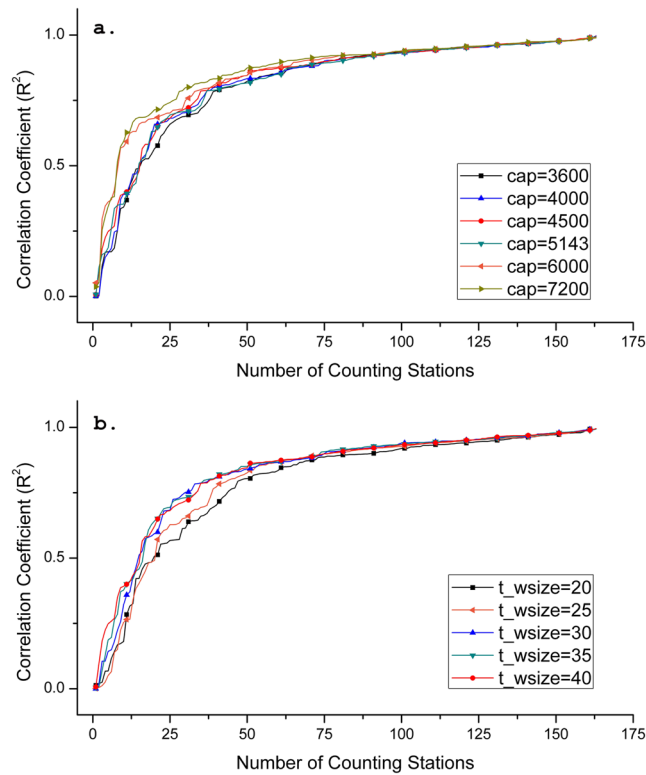


Figure 3. Accuracy test for the vehicular-flow estimation investigating two fundamental parameters when cell resolution is 400 meters. **(a)** Correlation coefficients (R^2) between measured and estimated vehicles for all the counting stations when the *capacity* is set from 3600 to 7200. **(b)** Correlation coefficients (R^2) between measured and estimated vehicles for all the counting stations when t_{wsize} is set from 20 to 40 seconds.

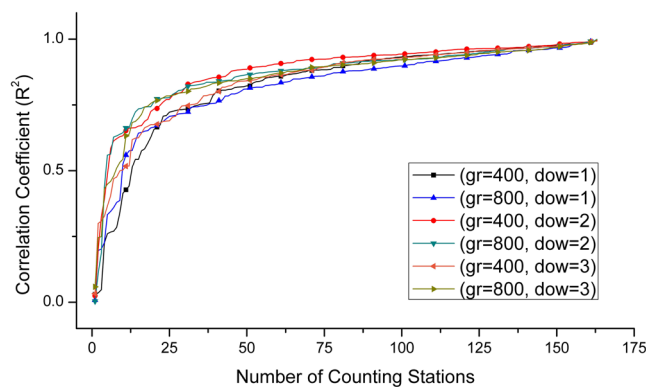


Figure 4. Correlation coefficients (R^2) between measured and estimated vehicles for all the counting stations and for three continuous days from weekday, Saturday, to Sunday when t_{wsize} and *cap* respectively equal to 40 seconds and 4500 vehicles.

Discussion

In order to investigate the influence of moving vehicles on UHI phenomenon across urban areas, this study designed a cell-transmission-model to simulate cell-based vehicular flows and to estimate the accumulation of the heat generated from vehicles over continuous time. The study has three important findings. First, heat patterns produced by vehicular flows strictly follow the spatio-temporal patterns of moving vehicles, e.g., heat has peaks and valleys periodically on a daily basis and it also has periodical circulations on a weekly basis. Second, UHI intensities in some specific land covers, such as core area of downtown with a large number of vehicles passing by, and even the green area in the suburbs and seaside with lack of man-made structures, are mainly or dominantly caused by vehicular flows. In other words, vehicular flow is also a considerable influential factor to anthropogenic heat or UHI phenomenon that cannot be ignored. Third, slow movement of vehicles would produce large amount of the heat, which suggests that construction of urban infrastructures (e.g. flyovers and highways) would promote

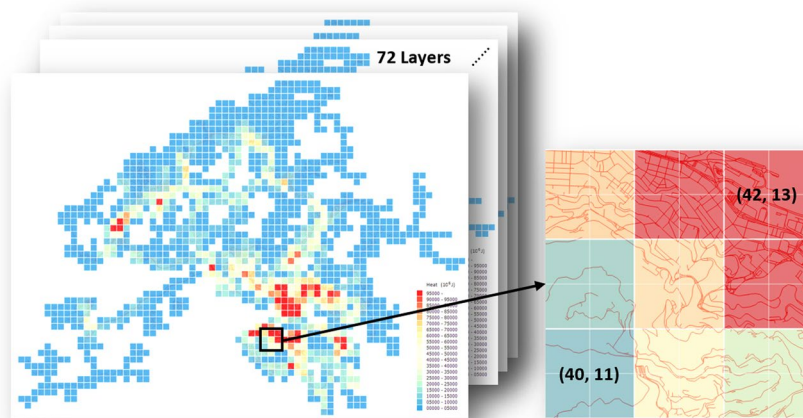


Figure 5. Heat emission from vehicles is accumulated for each grid cell and for each every hour for the three continues days in two different cell resolutions. Spatial continuity of nine grid cells ($cr = 800$ m) between $c(40, 11)$ and $c(42, 13)$ contain 36 smaller grid cells ($cr = 400$ m) so that heat of each every four smaller grid cells located in the larger one are summarized for the comparison of that of the larger grid cell. QGIS Desktop 2.8.2 (<http://www.qgis.org/en/site/>) is used to connect the DBMS PostgreSQL 9.5 (<https://www.postgresql.org/>) so that the 72 layers stored in the database can be read and visualized in sequence to have a screenshot to create the figure in Adobe Photoshop 7.0 (<http://www.adobe.com/products/photoshopfamily.html>).

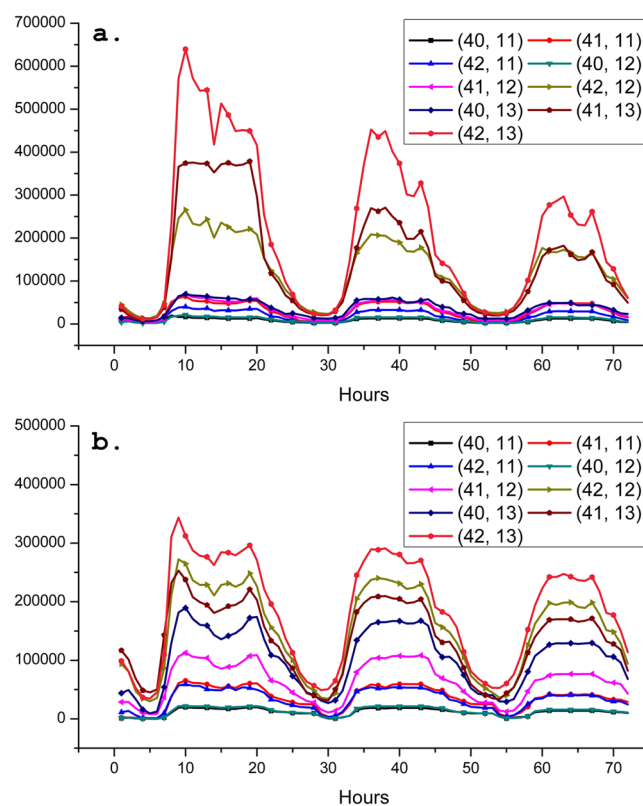


Figure 6. Comparison of the estimated heat summarized by nine spatial continuity of grid cells ($cr = 800$ m) from $c(40, 11)$ to $c(42, 13)$ for continues of weekday, Saturday, and Sunday. (a) Hourly accumulated heat for nine spatial continuity of grid cells when cr is 800 meters. (b) Hourly accumulated heat for each very four spatial continuity of grid cells ($cr = 400$ m) located in the same region of nine larger grid cells.

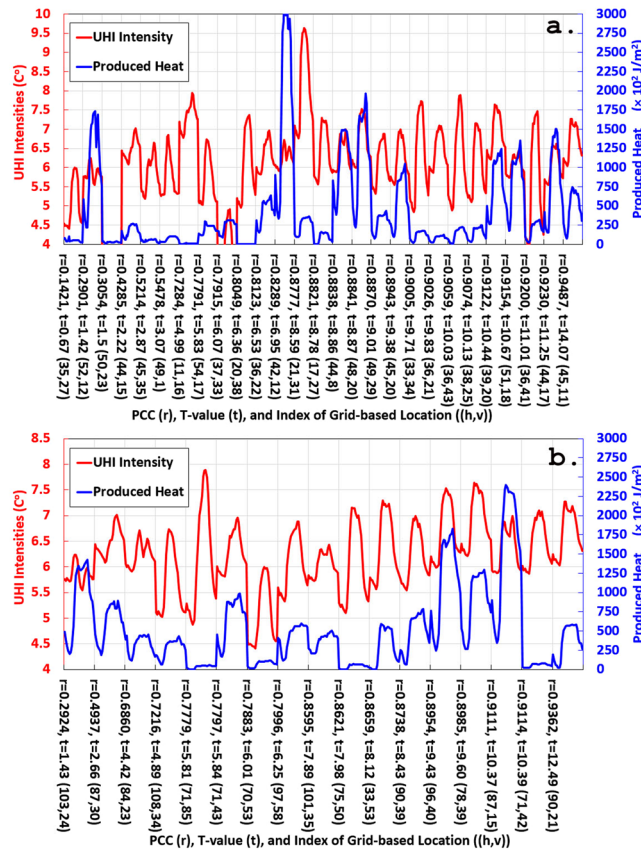


Figure 7. Time series based correlation analysis between the heat produced by vehicles and the measured UHI intensities that are located in the same grid cell. (a) Correlation between the estimated heat and the UHI intensities when cr is 800 meters. (b) Correlation between the estimated heat and the UHI intensities when cr is 400 meters.

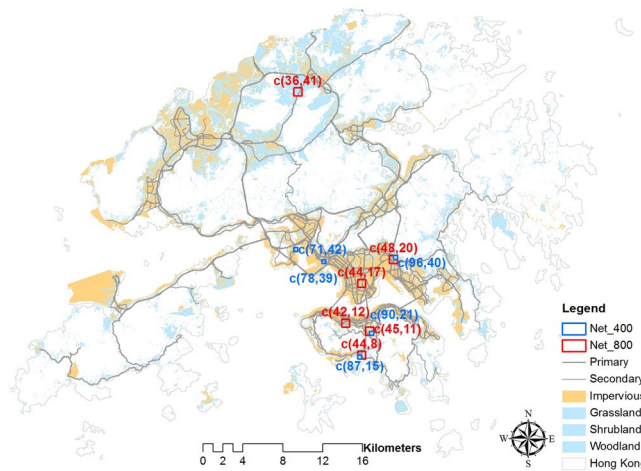


Figure 8. UHI intensities and the heat produced by vehicles are highly correlated in the statistical grid cells, where resolutions of the red and blue cells are 800 m and 400 m respectively. The figure is created in ArcMap 10.0 (<http://support.esri.com/Products/Desktop/arcgis-desktop/arcmap/10>) with the ESRI shapefile data format.

high-speed traffics and hence mitigate the UHI phenomenon effectively for heavily-congested cities. At the same time, if the estimated heat can be precisely transferred as the changed temperature by considering several interactive factors (i.e., atmospheric pressure, heat capacity of the air, and density of air mass under different air temperature in aerodynamics), then the changed temperature can have straight-forward comparison with UHI intensities.

This study designed a unique map generalization method to create dummy paths for each grid cell, so accurate locations of vehicles in the cell is not required, resulting in a significant improvement of effective simulation to fulfill the needs of heat accumulations, i.e., handling with more moving objects over longer time period, but with less time cost, less computational cost and hard disk storage for the intermediate results. This study also solves a fundamental problem of spatio-temporal modeling in transportation by estimating cell-based trajectories of vehicles. In this model, the OD matrix is deterministic because the origins are traffic counting stations and the destinations are grid cells with parking lots and/or roadside parking areas. The OD demand is dynamic since each vehicular flow with the quantitative information varies depending on time. Thus, the model can also be used for other studies. For example, it can be used to simulate patterns of human mobility with a given population census in a detailed spatial scale, and hence to estimate the accumulation of metabolic heat emission of human bodies quantitatively. This metabolic heat emission is also an influential factor of anthropogenic heat that has been widely acknowledged in literature.

Methods

Traffic counting data. The Transport Department of Hong Kong provides annual traffic census (ATC) data, which depict up-to-date vehicular flow information of Hong Kong in the year of 2015. Two items of ATC data were used in this study, i.e., geographic information with locations of 169 counting stations that may cross one or more lanes of roads, and textual information including (i) annual average daily traffic (AATD) describing the total number of vehicles passing through the counting stations in the whole year summarized by All-Day, week-day (Monday to Friday), Saturday, and Sunday; and (ii) hourly percentage of the number of vehicles in 24 hours that are also summarized in aforementioned four categories. Thus, the number of vehicles that pass through each station in each hour can be calculated by multiplying two items of the textual information.

Hourly air temperature. Hourly air temperature collected by 49 automatic weather stations for the whole year of 2015 were acquired from the Hong Kong Observatory. All weather stations were set up around 1–2 meters above the ground in both urban and rural areas. Therefore, air temperature can be an effective indicator to examine the effect of heat emission of vehicular flows since the weather stations and the vehicles are almost at the same heights.

Road network. It is necessary to determine different types of road network in Hong Kong to assign weight values of the roads for moving probability calculation. For this purpose, this work obtained directional roads from OpenStreetMap, which is an open-source map including volunteered information of road properties. To improve computing efficiency, irrelevant roads such as walking paths were removed and roads having the following seven properties are retained, i.e., *motorway*, *trunk*, *major*, *secondary*, *tertiary*, *residential*, and *services*.

Shortest path calculation. Since estimating vehicular flows covering the whole of Hong Kong in a high spatio-temporal resolution is a typical Big Data analytic task, road network and ATC data were thus imported into the PostgreSQL 9.5, an open-source object-relational database management system, to calculate the shortest path and to optimized execution efficiency at the maximum capability. The Dijkstra's algorithm was used to calculate the shortest path by pgRouting (an extension of PostgreSQL) to provide geospatial routing functionality. To release the computer RAM occupation that can extremely extend the time cost during each iterative computation for the four accessibilities simulation cell after cell and the vehicular flow estimation hour by hour, the central core is built in Eclipse to call the SQL functions and manages all the processes.

Travel speed estimation. Average speeds of the traffic flows in S-S, S-B, and B-B for each $c(h, v)$ are computed based on an established BPR function⁵³ as shown in Equation 15, where spd , S_o , v , and c respectively indicate the average speed in the current traffic condition (i.e. *average link speed*), the maximum speed of the free traffic-flow (i.e. *free-flow link speed*), the number of vehicles that are moving in the streets at the moment (i.e. *volume*), and the number of vehicles that can move at capacity (i.e. *capacity*). The coefficient a determines the ratio of the travel time in the free-flow to the travel time at capacity, and b determines how rapidly travel time increases starting from the free-flow travel time. The higher the b is the less sensitive the estimated travel time will be until the ratio (v/c) approaches 1. a and b are widely set as 0.15 and 4 respectively, and the maximum capacity ratio v/c is normally between 0.8 and 1.

$$spd = \frac{S_o}{\left[1 + a\left(\frac{v}{c}\right)^b\right]} \quad (15)$$

References

1. Zhou, D. *et al.* Spatiotemporal trends of urban heat island effect along the urban development intensity gradient in China. *Science of the Total Environment* **544**, 617–626 (2016).
2. Li, X. *et al.* The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States. *Science of the Total Environment* **605–606**, 426–435 (2017).
3. Ding, Z. *et al.* Impact of diurnal temperature range on mortality in a high plateau area in southwest China: A time series analysis. *Science of the Total Environment* **526**, 358–365 (2015).
4. Kenney, W. L., Craighead, D. H. & Alexander, L. M. Heat Waves, Aging, and Human Cardiovascular Health. *Med. Sci. Sports Exerc.* **46**, 1891–1899 (2014).
5. Morabito, M. *et al.* Air temperature-related human health outcomes: Current impact and estimations of future risks in Central Italy. *Science of the Total Environment* **441**, 28–40 (2012).

6. Cohn, E. G. & Rotton, J. Weather, seasonal trends and property crimes in Minneapolis, 1987–1988. A moderator-variable time-series analysis of routine activities. *Journal of Environmental Psychology* **20**, 257–272 (2000).
7. Field, S. The effect of temperature on crime. *British Journal of Criminology* **32**, 340–351 (1992).
8. Rotton, J. & Cohn, E. G. Outdoor temperature, climate control, and criminal assault. *Environment and Behavior* **36**, 276–306 (2004).
9. Fung, W. Y., Lam, K. S., Hung, W. T., Pang, S. W. & Lee, Y. L. Impact of urban temperature on energy consumption of Hong Kong. *Energy* **31**, 2623–2637 (2006).
10. Papakostas, K., Mavromatis, T. & Kyriakis, N. Impact of the ambient temperature rise on the energy consumption for heating and cooling in residential buildings of Greece. *Renewable Energy* **35**, 1376–1379 (2010).
11. Susca, T., Gaffin, S. R. & Dell’Osso, G. R. Positive effects of vegetation: Urban heat island and green roofs. *Environmental Pollution* **159**, 2119–2126 (2011).
12. Chen, D. *et al.* Urban vegetation for reducing heat related mortality. *Environmental Pollution* **192**, 275–284 (2014).
13. Grover, A. & Singh, R. B. Analysis of Urban Heat Island (UHI) in Relation to Normalized Difference Vegetation Index (NDVI): A Comparative Study of Delhi and Mumbai. *Environments* **2**, 125–138 (2015).
14. Mariani, L. *et al.* Climatological analysis of the mitigating effect of vegetation on the urban heat island of Milan, Italy. *Science of the Total Environment* **569**, 762–773 (2016).
15. Allegrini, J., Dorer, V. & Carmeliet, J. Influence of the urban microclimate in street canyons on the energy demand for space cooling and heating of buildings. *Energy and Buildings* **55**, 823–832 (2012).
16. Bonamente, E. *et al.* An energy-balanced analytic model for urban heat canyons: comparison with experimental data. *Advances in Building Energy Research* **7**, 222–234 (2013).
17. Phelan, P. E. *et al.* Urban Heat Island: Mechanisms, Implications, and Possible Remedies. *Annu. Rev. Environ. Resour.* **40**, 285–307 (2015).
18. Quah, A. K. L. & Roth, M. Diurnal and weekly variation of anthropogenic heat emissions in a tropical city, Singapore. *Atmospheric Environment* **46**, 92–103 (2012).
19. Chow, W. T. L. *et al.* A multi-method and multi-scale approach for estimating city-wide anthropogenic heat fluxes. *Atmospheric Environment* **99**, 64–76 (2014).
20. Lee, S. H., McKeen, S. A. & Sailor, D. J. A regression approach for estimation of anthropogenic heat flux based on a bottom-up air pollutant emission database. *Atmospheric Environment* **95**, 629–633 (2014).
21. Smoliak, B. V. *et al.* Dense Network Observations of the Twin Cities Canopy-Layer Urban Heat Island. *Journal of Applied Meteorology and Climatology* **54**, 1899–1917 (2015).
22. Yuan, C. & Chen, L. Mitigating urban heat island effects in high density cities based on sky view factor and urban morphological understanding: a study of Hong Kong. *Architectural Science Review* **54**, 305–315 (2011).
23. Nichol, J. E., Fung, W. E., Lam, K. & Wong, M. S. Urban heat island diagnosis using ASTER satellite images and ‘in situ’ air temperature. *Atmospheric Research* **94**, 276–284 (2009).
24. Chen, S. R. & Wu, J. Modeling stochastic live load for long-span bridge based on microscopic traffic flow simulation. *Computers and Structures* **89**, 813–824 (2011).
25. Zamith, M. *et al.* A new stochastic cellular automata model for traffic flow simulation with drivers’ behavior prediction. *Journal of Computational Science* **9**, 51–56 (2015).
26. Spiliopoulou, A., Kontorinaki, M., Papageorgiou, M. & Kopelias, P. Macroscopic traffic flow model validation at congested freeway off-ramp areas. *Transportation Research* **41**, 18–29 (2014).
27. Delis, A. I., Nikolos, I. K. & Papageorgiou, M. Macroscopic traffic flow modeling with adaptive cruise control: Development and numerical solution. *Computers and Mathematics with Applications* **70**, 1921–1947 (2015).
28. Nyhan, M. *et al.* Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model. *Atmospheric Environment* **140**, 352–363 (2016).
29. Sagarra, O., Szell, M., Santi, P., Diaz-Guilera, A. & Ratti, C. Supersampling and Network Reconstruction of Urban Mobility. *PLoS ONE* **10**, e0134508 (2015).
30. Dolfin, M. From vehicle-driver behaviors to first order traffic flow macroscopic models. *Applied Mathematics Letters* **25**, 2162–2167 (2012).
31. Li, A. C. Y. *et al.* Approximate Solution Procedure for Dynamic Traffic Assignment. *Journal of Transportation Engineering* **139**, 822–832 (2013).
32. Li, B. B. Bayesian inference for origin-destination matrices of transport networks using the EM algorithm. *Technometrics* **47**, 399–408 (2005).
33. Chen, A., Ji, Z. & Recker, W. Travel time reliability with risk sensitive travelers. *Transportation Research Record* **1783**, 27–33 (2002).
34. Hazelton, M. L. Estimation of origin-destination matrices from link flows on uncongested networks. *Transportation Research Part B* **34**, 549–566 (2000).
35. Hazelton, M. L. Estimation of origin-destination trip rates in Leicester. *Applied Statistics* **50**, 423–433 (2001).
36. Parry, K. & Hazelton, M. L. Estimation of origin-destination matrices from link counts and sporadic routing data. *Transportation Research Part B* **46**, 175–188 (2012).
37. Cascetta, E. & Nguyen, S. A unified framework for estimating or updating origin-destination matrices from traffic counts. *Transportation Research Part B* **22**, 437–455 (1988).
38. Davis, G. A. Estimating freeway demand patterns under impact of uncertainty on ramp controls. *Journal of Transportation Engineering* **119**, 489–503 (1993).
39. Siripipote, T., Sumalee, A., Ho, H. W. & Lam, W. H. K. Statistical approach for activity-based model calibration based on plate scanning and traffic counts data. *Transportation Research Part B* **78**, 280–330 (2015).
40. Ashok, K. & Ben-Akiva, M. E. Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. *Transportation Science* **36**, 184–198 (2002).
41. Antoniou, C., Ben-Akiva, M. & Koutsopoulos, H. N. Nonlinear Kalman Filtering Algorithms for On-Line Calibration of Dynamic Traffic Assignment Models. *IEEE Transactions on Intelligent Transportation Systems* **8**, 661–670 (2007).
42. Shafiei, M., Nazemi, M. & Seyedabrishami, S. Estimating time-dependent origin-destination demand from traffic counts: extended gradient method. *The International Journal of Transportation Research* **7**, 210–218 (2015).
43. Qiu, S., Abdelaziz, M., Abdellatif, F. & Claudel, C. G. Exact and grid-free solutions to the Lighthill-Whitham-Richards traffic flow model with bounded acceleration for a class of fundamental diagrams. *Transportation Research Part B* **55**, 282–306 (2013).
44. Xie, C. & Duthie, J. An Excess-Demand Dynamic Traffic Assignment Approach for Inferring Origin-Destination Trip Matrices. *Networks and Spatial Economics* **15**, 947–979 (2015).
45. Costeseque, G. & Lebacque, J. A variational formulation for higher order macroscopic traffic flow models: Numerical investigation. *Transportation Research Part B* **70**, 112–133 (2014).
46. Mohan, R. & Ramadurai, G. Submission to the DTA2012 Special Issue: A Case for Higher-Order Traffic Flow Models in DTA. *Networks and Spatial Economics* **15**, 765–790 (2015).
47. OpenStreetMap. <https://www.openstreetmap.org/#map=11/22.3567/114.1363> (Last Accessed: October 06, 2017).
48. Licensed Vehicles by Fuel Type in Hong Kong in Year 2015. http://www.td.gov.hk/filemanager/en/content_4708/table_44.pdf (Last Accessed: October 06, 2017).
49. Fuel Economy of Vehicles in Hong Kong. http://ecib.emsd.gov.hk/en/indicator_trp.htm (Last Accessed: October 06, 2017).

50. Long, Z. & Li, S. A complete classification of spatial relations using the Voronoi-based nine-intersection model. *International Journal of Geographical Information Science* **27**, 2006–2025 (2013).
51. Ferreira, M. J. & Oliveira, A. P. Anthropogenic heat in the city of São Paulo, Brazil. *Theoretical and Applied Climatology* **104**, 43–46 (2011).
52. Zhu, R., Guilbert, E. & Wong, M. S. Object-oriented tracking of the dynamic behavior of urban heat islands. *International Journal of Geographical Information Science* **31**, 405–424 (2016).
53. Mtoi, E. T. & Moses, R. Calibration and Evaluation of Link Congestion Functions: Applying Intrinsic Sensitivity of Link Speed as a Practical Consideration to Heterogeneous Facility Types within Urban Network. *Journal of Transportation Technologies* **4**, 141–149 (2014).

Acknowledgements

This work was supported in part by the grant of Early Career Scheme (project id: 25201614) and General Research Fund (project id: 515513) from the Research Grants Council of Hong Kong; the grants G-YM85 and 1-ZVAJ from the Hong Kong Polytechnic University. The authors thank the Hong Kong Transport Department, Lands Department, Hong Kong Observatory for the annual traffic census data, GIS road network data, and weather and climate data.

Author Contributions

Rui Zhu designed the model, performed experiments, data analysis and prepared the figures and the manuscript. Man Sing Wong finalized the model and prepared the manuscript. Éric Guilbert and Pak-Wai Chan also contributed to the manuscript preparation.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017