



City size, network structure and traffic congestion

Theodore Tsekeris^{a,*}, Nikolas Geroliminis^b

^a Centre of Planning and Economic Research (KEPE), 11 Amerikis, 10672 Athens, Greece

^b Urban Transport Systems Laboratory (LUTS), École Polytechnique Fédérale de Lausanne (EPFL), Room GC C2 389, Station 18, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:

Received 9 June 2011

Revised 20 December 2012

Available online 1 February 2013

JEL classification:

L9

R1

R3

R4

Keywords:

City size

Land use

Transport network

Traffic congestion dynamics

Macroscopic fundamental diagram

ABSTRACT

This paper presents an alternative approach for analyzing the relationship between land use and traffic congestion by employing the Macroscopic Fundamental Diagram (MFD). The MFD is an empirically observed relationship between traffic flow and traffic density at the level of an urban region, including hypercongestion, where flow decreases as density increases. This approach is consistent with the physics of traffic and allows the parsimonious modeling of intra-day traffic dynamics and their connection with city size, land use and network characteristics. The MFD can accurately measure the inefficiency of land and network resource allocation due to hypercongestion, in contrast with existing models of congestion. The findings reinforce the 'compact city' hypothesis, by favoring a larger mixed-use core area with greater zone width, block density and number of lanes, compared to the peripheral area. They also suggest a new set of policies, including the optimization of perimeter controls and the fraction of land for transport, which constitute robust second-best optimal strategies that can further reduce congestion externalities.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The increasing economic and environmental concerns raised by the growth of private vehicle use in urban areas have resulted in the design and implementation of a number of planning and management strategies on the supply side (control of traffic signals, ramp metering, capacity enhancement, etc.) or the demand side (congestion pricing, parking restriction, etc.) to diminish efficiency losses. From the planning perspective, policies have favored more compact development patterns by revitalizing the city center and restricting urban sprawl, through density and boundary growth controls (Anas et al., 1998; McConnell et al., 2006). In this context, the appropriate selection of network design parameters is crucial for the efficient allocation of road investment in the early stages of planning, or when updating the urban master plan. Such design parameters may encompass the number of road links, average link length, block area and average number of lanes. Particularly, the question of the allocation of resources to large urban clusters or more spatially dispersed metropolitan areas is critical for the

development of countries with a rapidly growing urban population, such as China and India (Henderson, 2010).

The proper modeling, interpretation and treatment of the relationship between urban land use and congestion are necessary to address the above question. However, existing traffic models in urban economics pose severe theoretical and empirical limitations in realistic applications. This is because they employ link travel cost functions which cannot accurately specify the intra-day traffic dynamics and relate them to land use and urban-scale network characteristics in a way that is computationally tractable and consistent with the physics of traffic. This failure hinders the ability of economic models to support accurate and robust design proposals for the allocation of urban land and network resources to diminish congestion externalities.

Specifically, the traditional models of congestion simplistically assume that the travel time on each link is separable and monotonically increasing with link flow. This assumption is adopted by the Bureau of Public Roads (Branston, 1976) and the Vickrey congestion function. The form of such travel time functions implies the existence of a stationary traffic equilibrium regime and steady-state volume-delay relationships. Several studies have challenged the use of these functions because of the need to account for the non-monotonicity of travel time with traffic flow (McDonald et al., 1999) and showed the intrinsic inconsistency, infeasibility

* Corresponding author.

E-mail addresses: tsek@kepe.gr (T. Tsekeris), nikolas.geroliminis@epfl.ch (N. Geroliminis).

and instability of static models of congestion (Verhoef, 1999, 2001, 2005). As more recent studies (Geroliminis and Daganzo, 2008) have pointed out, plots between pertinent traffic variables (flow, speed, delay) on a spatially disaggregated (link) level are not in steady state, but are actually in highly-scattered conditions and do not follow a well-defined curve.

The general equilibrium models of urban land use and transport typically assume an uncongested regime or point-queuing. The latter assumption does not account for the temporal and spatial dimensions of congestion and is not consistent with the laws of physics, because travel speed is entirely determined by traffic density at a specific location and time (Ross and Yinger, 2000; Arnott, 2007). Even recent, conceptually sophisticated models of this type, such as the Regional Economy, Land Use and Transportation (RELU-TRAN) model (Anas and Liu, 2007; Anas, 2011), do not consider the intra-day traffic dynamics. Moreover, such models involve increased calibration costs for network modeling (including partitioning into a considerable number of zones) and require the availability of origin–destination trip matrices and traffic assignment procedures to produce a user equilibrium pattern of link travel times. The traffic assignment procedures are associated with strong assumptions about the route choice behavior of consumers, intense computational burden and mathematical properties that are difficult to analyze in realistically large networks (Peeta and Ziliaskopoulos, 2001; Chiu et al., 2011).

The bottleneck model (Vickrey, 1969; Arnott et al., 1993; Arnott, 1998), in which peak-period traffic congestion is represented as a queue behind a bottleneck with fixed flow capacity at the edge of CBD, provides an alternative congestion technology. However, an important problem pertaining to all the above models is that they ignore the downward-sloping part of the curve between flow and density, known as hypercongestion, on a single link that is typically homogenous with uniformly distributed capacity (Lo and Szeto, 2005). In hypercongestion, the flow decreases from the point where the density reaches a particular critical value that maximizes flow, until the flow falls to zero when density reaches its maximum value (jam density).

Furthermore, only a few studies have attempted to abstractly represent the network-wide relationships between traffic variables, by extending the constraints and dynamics involved in the bottleneck model. In some of them, like that of Small and Chu (2003), which allowed for hypercongestion, the dynamic analysis cannot ensure a stable macroscopic relationship that is consistent with the physics of traffic. In others, such a relationship cannot be generalized (Arnott and Inci, 2010), or it becomes computationally intractable for large-scale urban networks. There are also studies (e.g. Lago and Daganzo, 2007; Arnott and de Palma, 2011) which consider the suburbs and CBD as points in space, i.e., without having a physical dimension. The latter assumption hinders the analysis of intersection-based control strategies (e.g. metering of network access) and land use policies for reducing congestion costs.

The approach proposed here treats the above theoretical and computational shortcomings by abstracting the complexity of network traffic dynamics into a single graphical expression. Fig. 1 illustrates a stable and low-scattered graphical relationship, referred to as the Macroscopic Fundamental Diagram (MFD),¹ between space–mean flow (or produced amount of travel), the system-wide vehicle-kilometers traveled per hour (veh-km/h), and vehicle density, the system accumulation in vehicles per kilometer per lane (veh/km/lane), for two regions (neighborhoods) of a hypo-

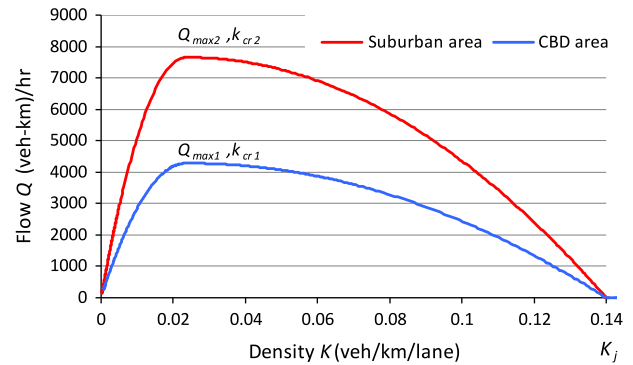


Fig. 1. Typical macroscopic fundamental diagram (network flow Q vs. network traffic density K) for the concentric city.

thetical concentric city (presented in Section 2). The MFD recognizes the inherent dynamics of macroscopic phenomena of congestion, the dependence of travel delays on the initial density conditions and the non-linearity of the average throughput–accumulation relationship. The system throughput is defined as the network outflow in vehicles per hour (veh/h), which is the rate that vehicles exit the network by driving to other neighborhoods or reaching their destinations.

In methodological terms, the MFD represents macroscopically (at each region) the flow of vehicles from their origin to destination at each instant (very short time interval) of the period of analysis. The modeling encompasses the rates at which vehicle trips enter and exit the network, and the queuing dynamics, according to the flow–density relationship given by the MFD curve of each region. This process ensures the conservation of flows between entry and exit rates and accounts for both states of congestion and hypercongestion at the rising and falling portion of the curve, respectively (see Fig. 1). The traffic capacity in some regions may change over time, when during hypercongestion the vehicle density (or accumulation) degrades throughput.

The MFD can also integrate the physical and functional characteristics of the urban land and transport network. This allows the expression of the spatial–temporal patterns of congestion as a function of city size, land use, network topology and traffic control. Hence, it is shown how different land use and network structures affect the formation and dynamics of congestion. Therefore, the MFD offers a tractable and parsimonious approach for modeling congestion that is consistent with the physics of traffic and relevant to economic analysis of policies to reduce it. The proposed alternative paradigm addresses the evaluation of land use policies and the design of transport systems, and suggests suitable combinations of strategies to alleviate congestion.

Existing studies in economics (see above) have mostly focused on the provision of new road capacity and imposition of tolls to reduce congestion externalities. Nonetheless, the implementation of these strategies is usually restricted by limited funding and enforcement capacities, particularly in developing countries. The MFD approach permits the analysis of a set of second-best policies, which could not be investigated in models that fail to include hypercongestion or are intractable for deployment in large-scale urban areas. Such policies include the use of advanced technology for metering of access (CBD perimeter control) and the reallocation of the existing network capacity and land use among urban zones. The resulting plans recognize that (i) commuting costs are not only a function of inflow, but also (non-linearly) depend on the previous level of congestion and (ii) an optimal solution may be reached even in hypercongestion.

The paper proceeds as follows. Section 2 describes the structure of a simple monocentric city model with concentric neighborhoods, which is used to apply the MFD approach. Section 3 pre-

¹ The existence of MFD has been empirically verified through aggregating highly scattered plots of flow vs. density from individual links as measured by fixed detectors and other readily available monitoring technologies in uniformly congested urban areas (Geroliminis and Daganzo, 2008).

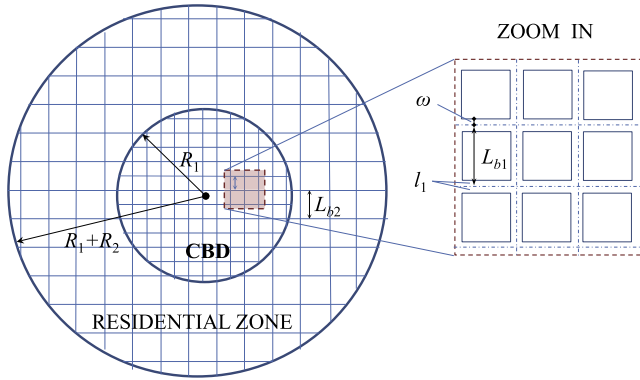


Fig. 2. Concentric city configuration.

sents the MFD approach and compares it with existing models of congestion. Section 4 describes simulation experiments for the optimal allocation of urban land and network resources in the presence of various constraints. Section 5 extends these experiments by allowing all of the available urban parameters to vary and it performs a sensitivity analysis of the results. It also analyzes how the MFD can be used for the robust design of alternative policies, such as metering of access through perimeter control and optimal allocation of land for transport, to reduce congestion externalities. In addition, it shows the results from implementing the MFD model into major cities with realistically representative settings. Section 6 summarizes and provides conclusions.

2. Concentric city model

Let us assume a concentric city with two zones: the inner core, $z = 1$, and the periphery, $z = 2$. The total area is equal to $A = \sum_z A_z = A_1 + A_2 = \pi R_1^2 + \pi(R^2 - R_1^2)$, where the total radius $R = R_1 + R_2$ (see Fig. 2). The city is closed with a total population $P = \sum_z P_z = P_1 + P_2$, and two land uses, only residence at zone 2 and both residence and employment at zone 1. Both the housing and employment locations of each consumer are fixed. By defining the population density of zone z as $D_z = P_z/A_z$ and the density ratio between the two zones as $r_d = D_2/D_1$, then, the population at zone 2 can be expressed as a function of the density ratio,² i.e., $P_2 = A_2 r_d D_1$.

Assuming that each urban zone z is spatially organized in square blocks of average length L_{b_z} , which is equal to the average intersection spacing or link length (distance between two consecutive traffic lights per block, see the inset view of Fig. 2), the total number of blocks can be continuously approximated as $N_{b_z} = A_z/L_{b_z}^2$. By uniformly partitioning the total area of a region into a finite number of square block groups, the number of lane-km can be approximated by multiplying the total number of block sides with the average link length L_{b_z} and the average number of road lanes ℓ_z at that zone, as follows:

$$L_z \cong 2(N_{b_z} + \sqrt{N_{b_z}})L_{b_z}\ell_z, \quad (1)$$

The model sets certain assumptions about the average link length L_{b_z} and number of lanes ℓ_z (Section 4), which yield the total network length, in terms of total lane-km $TLK = L_1 + L_2$. By denoting as w the unit price per lane-km, in monetary units (mu/lane-km),

the total budget required for road construction and maintenance costs in zone z is wL_z . The present value of this budget can be obtained through Eq. (2), where ϕ is a depreciation factor:

$$B_z = \frac{wL_z\phi}{T_D}, \quad \text{where } \phi = \frac{r_0(1+r_0)^{T_D}}{(1+r_0)^{T_D} - 1} \quad (2)$$

This relationship expresses the present value of annuity payments, which refer to a stream of fixed payments over a specified period of time (annually), taking into account the time value of money, interest rate r_0 and duration T_D (in years) of the design period of analysis (Lasher, 2008). It is noted that the representation of capital depreciation is not constrained by Eq. (2) and other functions may well be used for this purpose.

The individual demand is assumed here as perfectly inelastic and there is absence of tolls; hence, there are not components of consumer surplus and toll revenues. Thus, the efficiency gains resulting from an optimal urban plan (in terms of the city size, land use and network configuration) can be defined by minimizing the deadweight loss,³ by saving per-capita average time cost or aggregate infrastructure capital and time costs, in monetary terms. The latter measure expresses the total social cost (for both the government and consumers). The present value of the total social cost, TSC_z , encompasses the infrastructure (budget) cost, B_z , and the total travel time τ_z . In order to take into account the population heterogeneity, the transformation of τ_z into monetary units relies on weighting with the corresponding value of travel time $VOTT_z$ of commuters whose origin is located at zone z .

Let us decompose the vehicle accumulation within the CBD as $n_1 = n_{11} + n_{12}$, where n_{11} and n_{12} are the numbers of vehicles in zone 1, which originated from zones $z = 1$ and $z = 2$, respectively (time index is omitted for brevity), and denote the vehicle accumulation in zone 2 as n_2 . Provided that interval t has duration of one minute, each vehicle traveling either in zone 1 or zone 2 adds to the trip cost of the corresponding zone one minute of travel during the period T of analysis. By aggregating the intervals in which each vehicle is traveling in the zones 1 and 2 over the period T , the total travel time in each of the zones (expressed in veh-h) is $\tau_1 = \int_T n_{11}(t) dt$ and $\tau_2 = \int_T (n_2(t) + n_{12}(t)) dt$, respectively. Then, the total social cost in the whole urban area can be expressed as:

$$TSC = \sum_z TSC_z = \sum_z [B_z + p_a \tau_z] \\ = [B_1 + p_a VOTT_1 \tau_1] + [B_2 + p_a VOTT_2 \tau_2], \quad (3)$$

where p_a refers to a factor projecting the estimated travel time τ_z for the period of analysis to the whole peak travel period of a typical day (weekday) on a yearly basis. By assuming that the trip costs are mostly accumulated in the morning peak period and the remaining period of a typical day accounts for the 50% of the trip costs of the peak period, and given that the typical days of a year amount to 260 (by excepting the weekends), it is obtained that $p_a = 1.5 * 260 = 390$. As it was explained before, the problem objective can be associated with the minimization of either the TSC or the total average per-capita travel time $ATT = (\tau_1 + \tau_2)/\rho P$, where ρ is the car ownership index (cars/person), assumed to be the same for both zones. Subsequently, it holds that the per-capita travel time spent by the consumers whose origin is located at zone $z = 1, 2$ is $ATT_z = \tau_z/(\rho P_z)$.

The minimization of the above objectives may be subject to social, fiscal, regulatory, operational and physical constraints. Such constraints may be related to budget and equity problems typically arising in the design of city characteristics. They involve the max-

² Although the population density gradually changes from the CBD to the residential area in several real cases, the conventional assumption of the concentric zoning system allows the plausible separation of urban space into two zones with distinct land uses (residential and mixed) and the investigation of various zone-specific urban policies, such as boundary growth control, region-wide changes of street network or block densities, perimeter controls and cordon toll pricing.

³ By assuming the lot sizes as fixed, consumers derive variable utility only from other goods, and with fixed overall resources, those available for other goods are maximized when aggregate transport costs are minimized.

imum available total budget U_B (in present value terms) and the spatial equity $SE = |1 - (ATT_2/ATT_1)| \leq U_S$ and fiscal equity $FE = |1 - (B_1/B_2)| \leq U_F$ of consumers between the two zones. The measurement of inequities, with respect to time or monetary variables, can offer a meaningful base for the fair reallocation of land and network resources. The constraint of spatial equity addresses the question of whether the inter-regional provision of transport services is equitable, in terms of the difference in the ability of commuters from various zones to access work locations (here, those located in the CBD). The fiscal equity constraint addresses the fair allocation of public expenditure among regions and the equitable coordination or regulation of fiscal competition between them, given a fixed budget for road infrastructure investment. The use of upper level constraints, U_S and U_F , gives favorable emphasis on the accessibility of commuters originating from the periphery, compared to those originating from the CBD, and the allocation of road expenditure on the CBD, compared to the periphery, respectively.

Based on Eq. (2), the budget constraint U_B does also impose a constraint on the maximum allowable (or feasible) consumption of land for transport in the whole urban area, in terms of the maximum total lane-km $TLK_{max} = U_B T_D / w\phi$. In empirical applications of the model into realistic metropolitan areas (see Section 5.4), this constraint can also be expressed in terms of keeping the TLK fixed. Then, it denotes that no additional portion of land should be consumed for transport in the city and allowable changes only concern the inter-zonal reallocation of land for this purpose. In addition, as described in Section 5.2, the above constraint can be modified or augmented to include a maximum fraction of land area $A_{maxz} \geq \omega L_z / A_z$ to be allocated for transport in region z , where L_z is estimated from Eq. (1) and ω is a typical average road width (e.g., 3 m, see Fig. 2).

Another constraint refers to the level-of-service (LoS) requirement of the road transport system, which can be defined as $U_o \geq n_z / n_{crz}$ for each zone $z = 1, 2$, where n_z is the vehicle accumulation at the last time interval f in which the period of analysis is partitioned and n_{crz} (or k_{crz}) is the critical accumulation (or density) at which the outflow is maximized (see Fig. 1). At accumulation $n_z > n_{crz}$ or corresponding density $K_z = (n_z / TLK_z / \ell_z) > k_{crz}$, the system enters into the hypercongestion regime (see Section 3). The latter constraint aims at ensuring the smooth dissipation of queues and elimination of hypercongestion at the end of the reference period of analysis.

3. The MFD approach for modeling congestion dynamics

3.1. Description of the MFD model

The macroscopic-network traffic models which have been recently proposed and tested (Daganzo, 2007; Daganzo and Geroliminis, 2008; Geroliminis and Daganzo, 2008), state that traffic in homogeneously loaded large urban regions can be modeled dynamically at an aggregate level, if such regions exhibit the following properties:

- (i) there exists a MFD⁴ linking the average car density with the average flow in each region, and

- (ii) there is a linear relationship between the flow and outflow in each region, i.e., the average trip length within each region is time invariant.

Consider a city partitioned into two concentric reservoirs (zones), 1 and 2 (Fig. 2). The homogeneity of traffic conditions assumed in each reservoir is consistent with the homogeneity that applies to flow and density when dealing with road congestion in urban economic models (Ohta, 2001; Small and Chu, 2003; Arnott and Inci, 2010). Specifically, the average travel time needed for vehicles to move within a region (or to pass over another region) is the same for a given interval t , which denotes the equalization of the (average) time cost among all users in that region. This is analogous to the (uniform) distributions adopted in modeling homogenous roads, including bottlenecks, where the time needed to pass across the link (or remain in that link) is considered to be the same.⁵ The above properties pertaining to the traffic dynamics imply, for each reservoir z , (i) the existence of a MFD, $Q_z(n_z)$, between accumulation n_z and network flow Q_z (in veh-km/min) and (ii) that equation $O_z(n_z) = Q_z(n_z) / \lambda_z$ holds, according to the well-known Little's formula for steady-state queuing systems (Little, 1961), where O_z is the outflow and λ_z the average trip length in that zone.⁶ It is also assumed that there exists an entrance function $C_{2 \rightarrow 1}(n_1)$, which describes the maximum inflow to zone 1 from zone 2. The causality of this function is that sufficiently large accumulations in the CBD restrict its inflow along the periphery.

By changing the parameters of traffic signals in the boundary of the two zones, principally through adjusting a control variable x for the demand-responsive metering of access to zone 1 from zone 2 (see subsection 5.3), one could develop perimeter control strategies to reduce hypercongestion in the total urban area and keep the accumulation in the city center at optimal levels. The dynamic equations between the state variables (n_1, n_2) and demand inputs (q_1, q_2) of the system are described as:

$$\frac{dn_1}{dt} = q_1 + \min(x \cdot C_{2 \rightarrow 1}(n_1), \frac{1}{\lambda_2} Q_2(n_2)) - \frac{1}{\lambda_1} Q_1(n_1), \quad (4a)$$

$$\frac{dn_2}{dt} = q_2 - \min(x \cdot C_{2 \rightarrow 1}(n_1), \frac{1}{\lambda_2} Q_2(n_2)). \quad (4b)$$

The second term of the right hand side, for both Eqs. 4(a) and 4(b), represent the transfer from zone 2 to zone 1; the third term of Eq. 4(a) represents the rate vehicles finish their trips inside reservoir 1. Based on the decomposition of vehicle accumulation within the CBD into n_{11} and n_{12} , an additional set of dynamic equations is obtained as follows:

$$\frac{dn_{11}}{dt} = q_1 - \frac{1}{\lambda_1} \frac{n_{11}}{n_1} Q_1(n_1), \quad (4c)$$

$$\frac{dn_{12}}{dt} = \min(x \cdot C_{2 \rightarrow 1}(n_1), \frac{1}{\lambda_2} Q_2(n_2)) - \frac{1}{\lambda_1} Q_1(n_1) \cdot \frac{n_{12}}{n_1}. \quad (4d)$$

Analytical relationships have been recently developed, as specified with the model of Daganzo and Geroliminis (2008), to explore the connection between network settings and a MFD for urban neighborhoods where cars are controlled by traffic signals. The

⁵ However, in the proposed modeling framework the assumption of homogeneity and continuity of each region can be relaxed, e.g., to represent physical borders or constraints in the spatial structure of the city, through suitably increasing the partitioning of the urban area into a larger number of zones.

⁶ The Little's formula verifies the assumption of average trip length in each region and it consistently allows describing the state-dependent discharge rate from a region as a concave function of the number of vehicles in that region. Although it holds when traffic is in a steady state, the effect of transitions between traffic states is small when the peak period is long compared to the duration of a trip (Daganzo, 2007; Daganzo et al., 2011).

⁴ The initial conjecture of a MFD was first proposed by Godfrey (1969). Collective behavior makes sense for some types of urban networks, since drivers naturally flock to underused parts of the network, entrances and exits might not disturb the distribution of speed and the city traffic could be treated macroscopically as a single-region dynamic (queuing) system with the number of vehicles n as the single state variable.

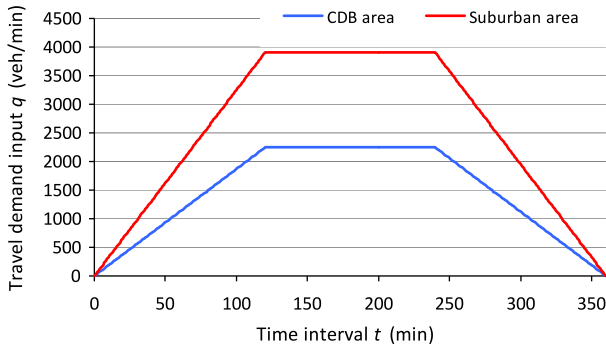


Fig. 3. Trapezoidal shape of travel demand profile (distribution of departure times) for the concentric city.

MFD for each region, $Q_z(n_z)$, can be expressed as a function of: (i) network variables: average link length, L_{bz} , and average number of lanes, ℓ_z ; (ii) link variables for one lane (common for both regions): free-flow speed, u_{ff} , jam density of a fully congested road, k_j , saturation flow, s (maximum flow of cars during green time), and congested wave speed (speed of queue length increase when vehicles at saturation flow approach a red light and stop), $u_w = u_{ff}/(k_j u_{ff}/s - 1)$; and (iii) intersection variables: signal cycle offset, δ (the time difference at the beginning of green between adjacent traffic signals), cycle time, C , and green time, G . The scalability of flows from a series of links to large traffic networks is not a straightforward transformation. Route or network capacity can be significantly smaller than the sum of capacities of constituent single links, because of the correlations developed through the different values of offsets. The offsets affect network capacity considerably when blocks are short.

In the MFD, any set of flow Q_z (or outflow O_z) and accumulation n_z (or network traffic density K_z) values relates to some level of (hyper)congestion, which depends on the network traffic capacity of region z . The capacity refers to the maximum flow Q_{maxz} in the region, at the critical accumulation n_{crz} (or critical density k_{crz}), and it depends on the zone width R_z and the average intersection spacing (link length) L_{bz} and number of lanes ℓ_z . Based on the dynamic traffic loading of the two regions according to the demand profile shown in Fig. 3 and following Eqs. 4(a)–4(d), the non-monotonic pattern of the MFD for region z is derived by fitting two distinct second-degree polynomials of the form $Q_z = \alpha_1 n_z^2 + \alpha_2 n_z + \alpha_3$. The set of parameters $(\alpha_1, \alpha_2, \alpha_3)$ corresponds to a specific link length L_{bz} and is calibrated separately for the upward (congested) and downward (hypercongested) parts of the curve (Fig. 1). The optimal plan resulting from this pattern relates the flow $Q_z(n_z)$, for a given vehicle accumulation n_z and traffic density K_z in the region, to a given land use density (or block density, in terms of the number of blocks with a specific land use per unit area). The MFD model can ensure the maximum flow and throughput in the case where critical density is reached, beyond which the network becomes hypercongested and the flow decreases.

Through the MFD pattern, the diseconomies of scale and density in each region are taken into account. Hence, in contrast with the traditional economic models, the present one can plausibly explain how congestion costs (diseconomies of agglomeration) in the city will force changes in the optimal amount of land and network resources, or their reallocation from one region to another. In this way, the MFD allows directly the use of flow and car density as two policy instruments which affect the city size, land use and block density, and network structure.

Fig. 1 shows an example of a MFD specified for the concentric city model (Fig. 2), with total area $A = 120 \text{ km}^2$ and equal zone widths $R_1 = R_2 = 3.09 \text{ km}$. The network variables are defined as

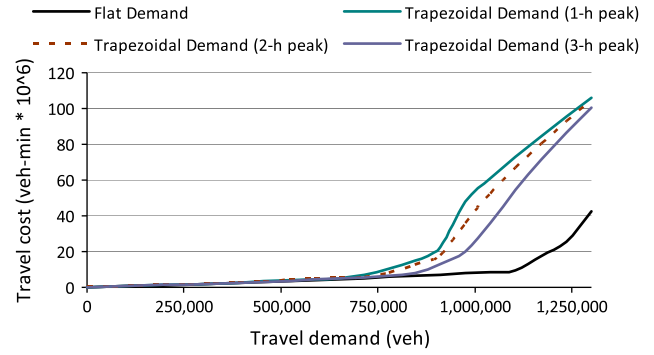


Fig. 4. Relationship between travel demand and cost under different demand profiles for the concentric city.

$L_{b1} = L_{b2} = 100 \text{ m}$ (meters), $\ell_1 = 2.0$ and $\ell_2 = 1.5$, while the values of link and intersection variables are the same to those used in Section 4. In this section, we assume that the demand profile for both zones has a symmetrical trapezoidal shape, with a total morning period of analysis equal to 360 min (minutes) and a peak period length of 120 min, and zonal populations $P_1 = 1.076$ and $P_2 = 1.872$ million residents (with car ownership index $\rho = 0.5$). These values make the system exhibit hypercongestion. The magnitude of demand input in each zone is determined in analogy with the regional allocation of population (Fig. 3).

3.2. Comparison with existing models

Consider the dynamical system of a two-region city (Section 2) governed by Eqs. 4(a)–4(d). Simulations run with different demand sizes (total trips per day) and lengths of peak period, i.e., 1 h (highest peak), 2 h, 3 h and 6 h (flat demand). The system returns to uncongested conditions by the end of simulation period, i.e., the same number of trips is completed for runs having the same demand size. The results are summarized in Fig. 4. It is clear that when traffic conditions are undersaturated and demand is low, the effect of variation in demand (as expressed with the peak period length) is negligible. The reason is that the MFD has almost linear behavior for small values of accumulation (see Fig. 1) and the total delay is not sensitive to small variations of flow within the given period. But, once the dynamical system of the two-region city enters the congested regime of the MFD for one of the regions, then assuming an average cost-demand curve creates significant errors. As expected, the more concentrated peak periods (and, hence, higher peak-period accumulations) are associated with higher increase of travel cost in relation to demand size, compared to the more dispersed and flat peak periods.

Fig. 5 illustrates the differences in the results obtained from the use of the MFD and a simple, static model (SM) of network traffic congestion. The SM adopts a simple supply function, where travel time at every instance is an increasing function of the inflow.⁷ In the SM, the travel cost conditions are stationary and the network traffic inflow solely depends on the given (trapezoidal with a 2 h peak period) pattern of demand and traffic capacity constraints, based on the specific city configuration. Namely, it ignores the network traffic dynamics, as described in the Eqs. 4(a)–4(d), which take into account time-varying changes of trip flows and travel cost resulting from previous time intervals due to queuing (in consistency with the physics of hypercongestion), and uncompleted trips, which is a function of the network outflow.

⁷ This traditional network supply curve was first introduced by Pigou (1920) and applied in most marginal-cost pricing models.

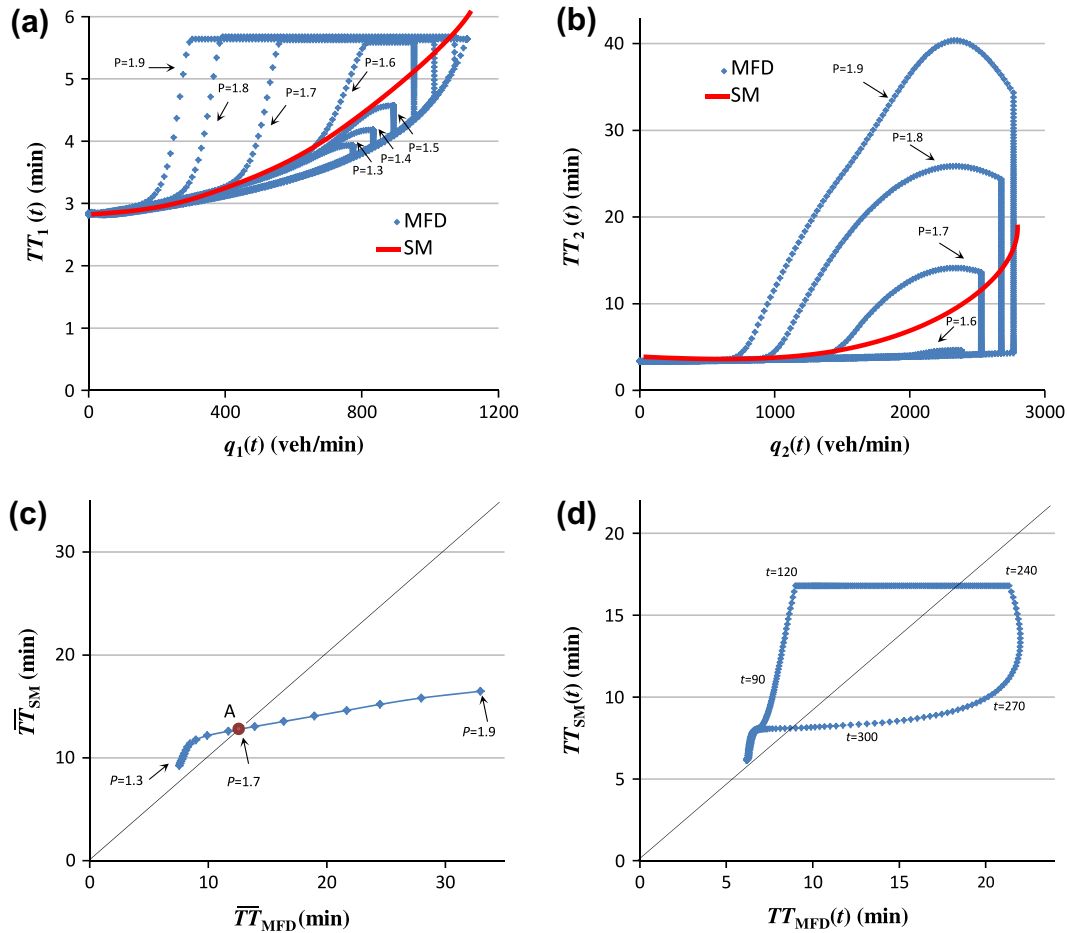


Fig. 5. Comparison of results obtained from the MFD and a static model (SM) of network traffic congestion based on supply functions for different population sizes (in million) for (a) region 1 and (b) region 2, and estimating (c) the day period-average travel time per vehicle trip for different population sizes, and (d) the per-vehicle average travel time for each time interval (when $P = 1.7$ million).

The MFD model supply function is dynamically specified by plotting the per-vehicle average travel time for each interval t for the inner region 1 (Fig. 5a) and outer region 2 (Fig. 5b), $TT_1(t)$ and $TT_2(t)$, as a function of the demand input of each region, $q_1(t)$ and $q_2(t)$, respectively. The different scenarios are analyzed for population sizes ranging from $P = 1.3$ million to $P = 1.9$ million. The supply function of the SM is correspondingly determined for each region and population size by calibrating a best-fit curve to represent the static relationship between travel time and inflow demand for each interval t (Fig. 5a and b). In contrast with the SM, the supply curves of the MFD model form loops. These loops signify that transport networks are not memory-less, since the same inflow will create higher travel times in a more congested state, compared to an initially less congested (or uncongested) state. They also show that vehicles at the offset of hypercongestion experience higher travel times for the same inflow than at the onset of hypercongestion. In region 2 (Fig. 5b), for population $P < 1.6$ million, there is no significant congestion and all curves produced by the MFD model are similar. For both regions and smaller population sizes, the initial segments of the curves formed by the SM and the MFD tend to coincide at the uncongested part. The error resulting from the SM is magnified with the increase of congestion and growth of population (for $P \geq 1.6$ million).

Fig. 5c summarizes the comparison of the results of the SM and MFD for the day period-average travel time per vehicle trip, \overline{TT}_{SM} and \overline{TT}_{MFD} , respectively, for increasing population (every $P = 0.03$ million) between $P = 1.3$ million and $P = 1.9$ million. It is found that the results of the two models are the same ($\overline{TT}_{SM} = \overline{TT}_{MFD}$) when

$P = 1.7$ million (at point A), the SM overestimates ($\overline{TT}_{SM} > \overline{TT}_{MFD}$) the day period-average travel cost when $P < 1.7$ million and underestimates it ($\overline{TT}_{SM} < \overline{TT}_{MFD}$) when $P > 1.7$ million. However, as it is shown in Fig. 5d, even when $TT_{SM} = TT_{MFD}$, the SM is not able to accurately estimate the trip cost in the urban area. This is because the per-vehicle average travel time $TT_{SM}(t)$ resulting from the SM may present significant intra-day deviations (depending on the time-of-day traffic conditions) from the per-vehicle average travel time $TT_{MFD}(t)$ resulting from the MFD. Specifically, when $P = 1.7$ million, the overestimation and underestimation errors of the SM, in relation to the MFD, are equalized and canceled out over the day period (Fig. 5d). The inaccurate modeling of the flow-density relationship and congestion costs through the SM may lead to misallocation of land and network resources in the urban area, and it cannot be utilized for deploying intra-day dynamic traffic management schemes, such as perimeter control and congestion pricing.

4. Model application with both fixed and variable parameters

In the current simulations, the two-region city has a total base population $P = 1.3$ million, fixed density ratio $r_d = 0.5$ (i.e., the population density in the periphery is half that of the CBD) and car ownership index $\rho = 0.5$. The urban topology characteristics are described by six parameters: the average block (or link) lengths, L_{b1} and L_{b2} , and average number of lanes, ℓ_1 and ℓ_2 , in zones 1 and 2, respectively, zone width ratio R_2/R_1 , which produces R_1 and R_2 (for the case of fixed R), and zone width R_2 , which produces R_2/R_1

ratio (for the case of variable R_2). In all the experiments, the permissible values of average link length are ranging from 70 to 500 m and of average number of lanes from 1 to 3, for both zones.

Each combination of optimized parameter values results in a different plan, with respect to the city size, land use and network configuration, whose efficiency is evaluated with either the total social cost, TSC , or the average per-capita travel time, ATT . In order to create some intuition about how topological parameters affect the different objectives, some preliminary analysis is performed first, wherein only two parameters (out of the six) are optimized considering the others as fixed. This situation could be due to imposition of physical, administrative or other constraints to the planning process. More specifically, the first set of experiments (plans 1–4) assumes that the total radius is fixed $R = 6.18$ km, hence, the total urban area is $A = 120$ km². The second set of experiments (plans 5–6) relaxes the above restriction and assumes a fixed CBD radius $R_1 = 3$ km and variable size of the outer zone with a minimum width $R_2 = 3$ km. The optimized parameters are as follows: L_{b1} and L_{b2} for plan 1, R_2/R_1 and L_{b2} for plan 2, R_2/R_1 and ℓ_2 for plan 3, L_{b2} and ℓ_2 for plan 4, R_2 and L_{b2} for plan 5, and R_2 and ℓ_2 for plan 6. Later, in Section 5, a full optimization framework is demonstrated that allows identifying optimal solutions for all available parameters.

The critical values of the model parameters have been experimentally evaluated in previous simulated and real case studies (Geroliminis and Daganzo, 2008). They are assumed to be proportional to the size of reservoirs: free-flow speed $u_{ff} = 43.25$ km/h, jam density $k_j = 0.14$ veh/m, saturation flow $s = 1700$ cars per hour per lane during green time, traffic signal cycle length $C = 90$ s, green time $G = 40.5$ s and a satisfactory signal offset to provide efficient coordination between traffic signals. It is assumed that the average trip lengths in zones 1 and 2 are $\lambda_1 = 1.2R_1$ and $\lambda_2 = 0.8R_2$. The average value of travel time for commuters in zones 1 and 2 are $VOTT_1 = 10$ mu/hour and $VOTT_2 = 6$ mu/hour. The temporal profile of demand is assumed to follow a trapezoidal shape (Fig. 3), with total period of analysis equal to 360 time intervals and (morning) peak period equal to 120 intervals. In terms of the infrastructure cost parameters (see Eq. (2)), the unit price $w = 15$ million mu/lane-km, the interest rate $r_0 = 0.1$ and the design period of analysis $T_D = 15$ years.

Let us define n_{max} as the maximum vehicle accumulation in the analysis period. Based on suitably scaling the total base population, three scenarios about the levels of demand and resulting congestion conditions for both zones are defined as follows:

- (a) reduced demand, which refers to low or mild congestion with $0.5 \leq n_{max}/n_{cr} < 0.9$ during the peak period,
- (b) moderate demand, which refers to system operation in close-to-capacity conditions, such that $0.9 \leq n_{max}/n_{cr} < 1.0$ during the peak period, and
- (c) increased demand, which refers to system operation in over-saturated conditions (hypercongestion), such that $n_{max}/n_{cr} \geq 1.0$ during the peak period.

In each scenario, these changes yield a different optimal set of parameters and selected urban plans. In the plans 1–4, wherein the population density is kept fixed for the total urban area, only inter-zonal changes in density are allowed. In the plans 5–6, the population density is only influenced by changes in the optimal land provision in the periphery. Tables 1–6 summarize the results for different levels of congestion (low, close-to-capacity, hypercongestion) conditions, by showing the values of network parameters (average block length and lane number), zone widths, TSC and ATT , spatial and fiscal equity, and total required budget (in billion mu or bmu).

In all the cases considered, the increase of demand level and congestion conditions, from reduced demand (Tables 1 and 2) to moderate demand (Tables 3 and 4) and, then, increased demand (Tables 5 and 6), leads to increased TSC and ATT , as well as larger investment needs, since the number of commuters who must be served also increases. The higher budget is invested to create more dense networks to keep the service quality for the increased demand at an acceptable level. For instance, in plan 1, when minimizing the TSC , block lengths decrease (and budget increases) as the demand moves from low to moderate conditions, i.e., from $L_{b1} = 140$ m and $L_{b2} = 480$ m ($B = 0.193$ bmu) to $L_{b1} = 130$ m and $L_{b2} = 310$ m ($B = 0.279$ bmu) (plan 1a), respectively, and then to hypercongestion conditions, to $L_{b1} = 120$ m and $L_{b2} = 220$ m ($B = 0.360$ bmu).

In Tables 3 and 4, plans 1a and 1b are comparatively examined to evaluate the impact of different CBD boundary constraints, in terms of setting the value of R_2/R_1 ratio equal to 1.00 and 0.50, respectively, on various performance measures. The increase of land occupied by the CBD, relative to the periphery (plan 1b), leads to the reduction of TSC and, particularly, of ATT , and the increase of total budget. Similarly, the results of plans 2 and 3, where the total radius R is held fixed, demonstrate that reducing the R_2/R_1 ratio, which implies increasing the share of commuting within the CBD, as traffic congestion increases, leads to diminishing the TSC and ATT .

The current findings are important because they take into account the exact fraction of land that is assigned to each use (residential and mixed), in contrast with the traditional urban economic analysis, where land use patterns typically refer to exclusive zones or rings for each use. Hence, they can provide useful insight into the optimal share of land use mixing in the urban area, even in hypercongestion conditions. The results verify previous findings in the relevant economic literature that fully interspersed (residential and employment) land use is never optimal (considering congestion effects) and that mixing in the CBD becomes optimal when traffic demand equilibrates at some level of congestion (Wheaton, 2004). The results are also consistent with those of McDonald (2009), in which the selective movement of households with commuters from an outer ring to an inner ring where they work was found to reduce congestion costs and improve efficiency in a monocentric city.

Table 1
Optimal values of parameters for minimization of TSC and ATT (in parenthesis) objective functions under the reduced demand scenario.

Variable	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6
R_2/R_1	1.00	1.40 (1.40)	1.10 (1.10)	1.00	1.00(1.00)	1.00(1.00)
R_1 (km)	3.09	2.58 (2.58)	2.94 (2.94)	3.09	3.00	3.00
R_2 (km)	3.09	3.60 (3.60)	3.24 (3.24)	3.09	3.00 (3.00)	3.00 (3.00)
L_{b1} (m)	140 (110)	110	110	110	110	110
L_{b2} (m)	480 (260)	480 (280)	280	500 (450)	500 (340)	340
ℓ_1	2.00	2.00	2.00	2.00	2.00	2.00
ℓ_2	1.50	1.50	1.14 (2.00)	3.00 (1.55)	1.50	1.15 (3.00)

Note: Bold values in the white cells indicate the optimized parameters, non-bold values in the white cells indicate dependent variables obtained from the optimized parameters and the shadowed cells include input values in each scenario. The same notation is used in Tables 3 and 5.

Table 2
Results of simulation analysis for different planning objectives and policy plans under the reduced demand scenario.

Variable	Objective	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6
TSC (bmu)	TSC min	0.403	0.377	0.447	0.425	0.419	0.428
	ATT min	0.478	0.425	0.505	0.497	0.459	0.538
ATT (min)	TSC min	7.144	6.721	7.818	6.864	6.335	6.761
	ATT min	6.501	6.357	7.070	6.232	6.236	6.137
Budget (bmu)	TSC min	0.193	0.187	0.235	0.224	0.232	0.230
	ATT min	0.287	0.246	0.312	0.312	0.274	0.356
Spatial equity	TSC min	1.108	1.263	1.287	1.188	1.021	1.191
	ATT min	1.036	1.095	1.041	0.924	0.979	0.927
Fiscal equity	TSC min	0.487	0.194	0.296	0.894	0.467	0.937
	ATT min	0.043	0.290	0.259	0.115	0.014	0.258

Table 3
Optimal values of parameters for minimization of TSC and ATT (in parenthesis) objective functions under the moderate demand scenario.

Variable	Plan 1a	Plan 1b	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6
R_2/R_1	1.00	0.50	0.85 (0.45)	0.75 (0.40)	1.00	1.38 (1.27)	1.00 (1.00)
R_1 (km)	3.09	4.12	3.34 (4.26)	3.53 (4.41)	3.09	3.00	3.00
R_2 (km)	3.09	2.06	2.84 (1.92)	2.65 (1.77)	3.09	4.15 (3.80)	3.00 (3.00)
L_{b1} (m)	130 (100)	120 (100)	100	100	100	100	100
L_{b2} (m)	310 (100)	220 (130)	280 (130)	130	500 (340)	500 (260)	260
ℓ_1	2.00	2.00	2.00	2.00	2.00	2.00	2.00
ℓ_2	1.50	1.50	1.50	1.00 (1.40)	2.50 (3.00)	1.50	1.30 (3.00)

Table 4
Results of simulation analysis for different planning objectives and policy plans under the moderate demand scenario.

Variable	Objective	Plan 1a	Plan 1b	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6
TSC (bmu)	TSC min	0.822	0.802	0.731	0.832	1.043	0.599	0.641
	ATT min	0.963	0.906	0.920	0.960	1.118	0.671	0.766
ATT (min)	TSC min	11.514	8.181	8.430	8.454	15.640	7.533	8.154
	ATT min	11.096	7.718	7.710	7.868	15.216	7.153	7.520
Budget (bmu)	TSC min	0.279	0.360	0.311	0.375	0.280	0.260	0.266
	ATT min	0.437	0.490	0.498	0.495	0.377	0.333	0.416
Spatial equity	TSC min	1.845	0.851	0.983	0.898	2.826	1.078	0.962
	ATT min	1.711	0.857	0.851	0.842	2.699	0.883	0.741
Fiscal equity	TSC min	0.359	0.936	0.521	0.262	0.316	0.391	0.321
	ATT min	0.420	0.384	0.567	0.927	0.257	0.166	0.428

Table 5
Optimal values of parameters for minimization of TSC and ATT (in parenthesis) objective functions under the increased demand scenario.

Variable	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6
R_2/R_1	0.50	0.75 (0.45)	0.65 (0.40)	0.50	1.10 (1.08)	1.10 (1.08)
R_1 (km)	4.12	3.53 (4.26)	3.75 (4.41)	4.12	3.00	3.00
R_2 (km)	2.06	2.65 (1.92)	2.43 (1.77)	2.06	3.30 (3.25)	3.30 (3.25)
L_{b1} (m)	120 (100)	100	100	100	100	100
L_{b2} (m)	220 (130)	270 (130)	130	340 (230)	350 (130)	130
ℓ_1	2.00	2.00	2.00	2.00	2.00	2.00
ℓ_2	1.50	1.50	1.00 (1.40)	3.00 (3.00)	1.50	1.30 (3.00)

Table 6
Results of simulation analysis for different planning objectives and policy plans under the increased demand scenario.

Variable	Objective	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6
TSC (bmu)	TSC min	0.861	0.797	0.894	1.481	0.883	0.886
	ATT min	0.942	0.956	0.998	1.538	1.045	1.001
ATT (min)	TSC min	8.712	8.665	8.619	18.093	12.972	13.025
	ATT min	7.891	7.868	8.035	17.765	12.473	12.235
Budget (bmu)	TSC min	0.360	0.331	0.391	0.420	0.264	0.265
	ATT min	0.490	0.498	0.495	0.494	0.442	0.407
Spatial equity	TSC min	0.805	0.974	0.887	2.488	1.851	1.865
	ATT min	0.840	0.837	0.828	2.409	1.545	1.481
Fiscal equity	TSC min	0.936	0.722	0.510	0.606	0.346	0.338
	ATT min	0.384	0.567	0.927	0.101	0.478	0.408

This outcome concerning the land use mixing (or share of commuting) in the CBD can be attributed to the decrease of trip delays due to the reduced total distance traveled. This is because of the increasing share of intra-zonal trips, compared to the inter-zonal trips, as the width ratio R_2/R_1 decreases. Besides, the reduction of the *ATT* (plan 1a vs. plan 1b) and R_2/R_1 ratio, as the congestion level increases, result in improvement (toward zero) of the spatial equity metric, while the impact on fiscal equity is reversed. Therefore, in the presence of urban boundary constraints, the growing levels of congestion favor more compact city development, through mixing land uses in the inner area, to reduce delay externalities and improve spatial equity. The requirements (set as $U_{0.12} = 0.50$, for the level-of-service, $U_S = 2.0$, for spatial equity, and $U_F = 1.0$, for fiscal equity) are found to be ensured (not binding) in all cases, except for plan 4 in the moderate and increased demand scenarios (Tables 4 and 6, respectively), where $SE > 2.0$. In the latter case, the spatial equity constraint must be relaxed to yield a feasible optimal plan.

At all levels of traffic congestion, plan 2 (in the first set of experiments) and plan 5 (in the second set of experiments) are found to produce the lowest *TSC* and *ATT*. Namely, the average block length (or link length) in zone 2 can provide the best possible design instrument to mitigate congestion externalities, in conjunction with changing either the share of urban land in favor of the CBD or the total urban boundary for a fixed CBD area. Specifically, increase of the outer area block density, i.e., a shorter average link length L_{b2} (improvement of the street network connectivity in zone 2), as well as increase of the CBD area, entails improvement of the system capacity. This is because of the increase of the density of access roads and throughput, i.e., the rate vehicles enter into the CBD. On the other hand, increase of the average block density entails a larger total number of links and, hence, a higher amount of road investment (construction and maintenance cost).

5. An optimization framework for parameter estimation

5.1. Unconstrained optimization results

In this subsection, the *TSC* and *ATT* are minimized with respect to all available parameters without imposing constraints. The least possible assumptions are made to ensure a plausibly allowable size of road infrastructure (see Section 4) and distinct separation be-

tween the two urban zones, to facilitate understanding of the model and interpretation of the results. The plans are investigated for different levels of demand, by gradually (using a factor of 1.5) increasing the total population size, i.e., $P = 1.30, 1.95, 2.60$ and 3.25 million. The unit prices and parameter values of infrastructure costs and the lengths of the total and peak period are the same to those adopted in Section 4. The optimization process is based on multiple runs of a steepest descent (SD) search routine to avoid local optima. It is known that the traditional SD method can be trapped to local minima. To address this weakness, the SD approach is combined with an initial random search in each consecutive run to identify a set of good initial solutions and improve the 'best' solution of the previous run. Table 7 presents the optimal values for the *TSC* and *ATT*, and the estimated parameter values and measures of equity and budget (which do not express constraints but they are outcomes of the optimization process), for different population size and urban growth scenarios.

The optimization results, which involve a higher degree of freedom, are found to be consistent with those described in the previous section. Specifically, larger population sizes are associated with higher expenditure and delay externalities, as reflect the increased values of the *TSC* and *ATT*. In addition, they suggest a more compact pattern of urban development, as implied by the shorter average block length, particularly for zone 2 (L_{b2}), and the smaller ratio R_2/R_1 , for both urban boundary growth scenarios (i.e., with fixed R and fixed R_1). The reduction of commuting share from zone 2, compared to the CBD, decreases the average travel time ATT_2 of vehicles originating from that zone, in relation to the ATT_1 , thus improving (moving closer to zero) the measure of spatial equity (based on *ATT* minimization). On the other hand, the measure of fiscal equity deteriorates (moves away from zero). This is because the CBD attracts more investments as it increasingly consumes more of the available land, relative to the periphery. The next subsection presents a sensitivity analysis of various parameter values and constraints.

5.2. Sensitivity analysis

The rapid growth of large metropolitan areas worldwide sets forth the need for a sensitivity analysis of initial plans on which the urban development process has been based. Given the optimal plan settings obtained for different population sizes (Table 7), the

Table 7
Optimization results with respect to *TSC* and *ATT* for different population size and urban growth scenarios.

Objective	Variable	Population in millions (with fixed R)				Population in millions (with variable R_2)			
		1.3	1.95	2.6	3.25	1.3	1.95	2.6	3.25
TSC minimization	TSC (bmu)	0.382	0.656	0.852	1.172	0.360	0.531	0.705	0.912
	ATT (min)	5.749	7.736	7.811	8.020	6.412	6.356	6.419	6.570
	L_{b1}	0.070	0.103	0.070	0.070	0.175	0.127	0.098	0.070
	L_{b2}	0.496	0.274	0.358	0.263	0.442	0.368	0.341	0.141
	ℓ_1	2.290	2.007	2.075	2.724	2.332	2.719	2.849	2.435
	ℓ_2	2.071	1.414	2.366	2.580	1.677	2.026	2.675	1.678
	R_2/R_1	2.203	1.832	1.101	0.880	2.111	1.736	1.060	1.001
	Budget	0.227	0.304	0.382	0.580	0.159	0.228	0.272	0.333
	Spatial equity	1.239	0.995	0.930	0.907	0.929	1.162	1.434	0.805
	Fiscal equity	0.176	0.582	0.283	0.540	0.797	1.531	3.233	2.695
	ATT minimization	TSC (bmu)	0.524	0.762	1.035	1.266	0.481	0.594	0.869
ATT (min)		5.827	6.475	7.067	7.738	5.596	6.061	6.247	6.447
L_{b1}		0.097	0.070	0.070	0.070	0.117	0.095	0.076	0.070
L_{b2}		0.292	0.258	0.206	0.118	0.368	0.313	0.186	0.182
ℓ_1		2.949	2.731	2.909	2.971	2.551	2.686	2.831	3.000
ℓ_2		2.130	2.584	2.643	1.539	2.829	2.032	2.676	3.000
R_2/R_1		1.449	1.384	1.096	0.882	2.561	2.131	1.960	1.826
Budget		0.361	0.587	0.621	0.682	0.311	0.341	0.472	0.523
Spatial equity		1.156	1.023	0.971	0.966	0.956	0.982	0.866	0.826
Fiscal equity		0.170	0.177	0.052	0.279	0.140	1.207	0.480	0.614

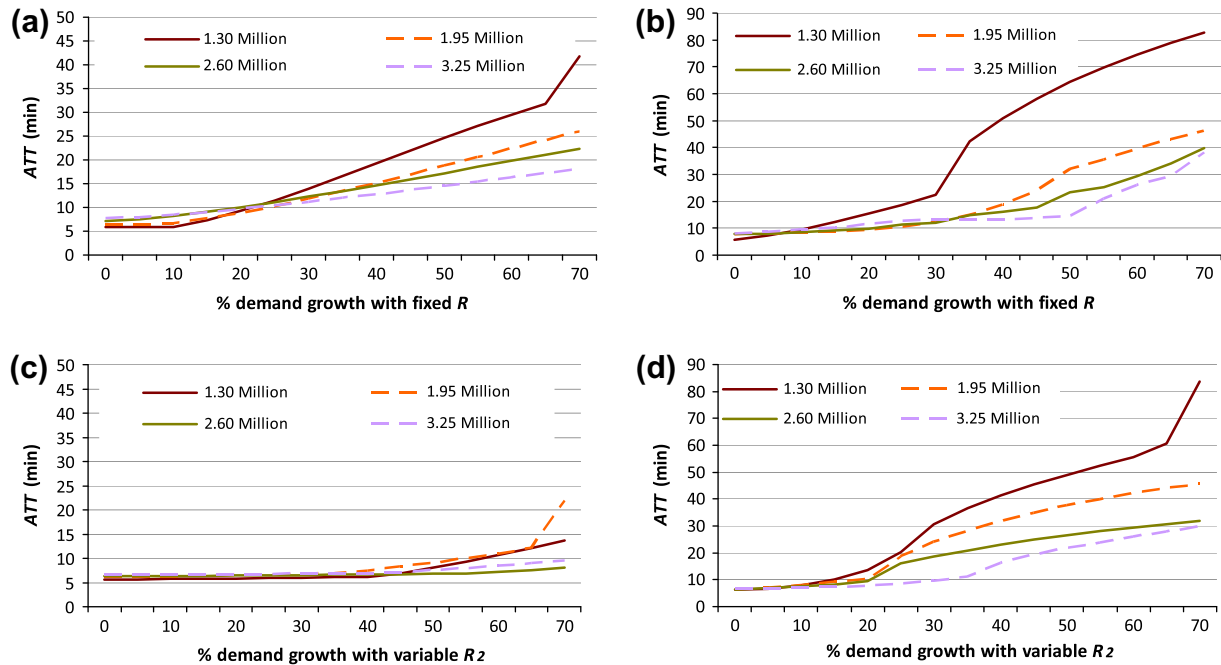


Fig. 6. Sensitivity analysis of the *ATT* with respect to percentage increase in demand for the cases of fixed total radius under (a) minimizing the *ATT* and (b) minimizing the *TSC*, and fixed CBD radius under (c) minimizing the *ATT* and (d) minimizing the *TSC*.

TSC and *ATT* are re-estimated for increasing (%) rates of travel demand growth. As it is shown in Fig. 6a–d, the optimal plans corresponding to smaller population sizes (1.3 and 1.95 million) yield higher *ATT* as the demand growth rate increases, compared to those corresponding to larger population sizes (2.6 and 3.25 million), for both minimization problems and boundary growth scenarios.

Specifically on the basis of the *ATT* minimization, when R is fixed (see Fig. 6a), the *ATT* remains small up to a 25% increase in demand, but then it remarkably increases from 5.8 min to 42 min, for initial population size of 1.3 million, and from 6.5 min to 26 min, for initial population size of 1.95 million, when the level of demand increases by 70%. Values of *ATT* more than 30 min indicate that the system is unable to return to uncongested conditions by the end of the period. However, the *ATT* increases only from 7.1 min to 22 min, for initial population size of 2.6 million, and from 7.7 min to 18 min, for initial population size of 3.25 million, for the maximum level of demand increase. These differences can be attributed to the higher total budget and, hence, increased road capacity allocated for cases of larger initial population size (see Table 7).

In the case where R is fixed (Fig. 6a), the *ATT* increases much sharper with the growth of population, compared to the case where only R_1 is fixed (Fig. 6c). For instance, for initial population 1.30 million, the city becomes highly congested ($ATT > 13$ min) when population grows by about 30% in the former case, while this occurs when population grows by about 70% in the latter case. Therefore, the increase of *ATT* is getting smaller when planning the city for higher (possibly expected in the future) population sizes and when only R_1 is kept fixed (R_2 is variable). These non-linear changes of *ATT* with respect to population growth can be interpreted by the non-monotonic flow-density relationship that reflects the resulting MFD pattern for each region. The MFD accounts for the sensitivity to initial conditions and plausibly explains the diseconomies of a sprawled periphery in relation to the CBD, compared to traditional economic models of urban congestion (see Section 1).

Specifically, this outcome can be attributed to the fact that considering (or predicting) a more crowded urban area, keeping the boundary of the CBD fixed and controlling (compressing) the width of the residential area, leads to a larger increase of block (or street network) density. This pattern entails a more compact city development and higher share of within-CBD commuting, compared to considering a less crowded urban area and keeping the total urban boundary fixed. Although it induces higher expenditure, the increased network density results in a MFD shape that is associated with less hypercongestion and an enhanced system capacity and throughput of the residential area. Also, it allows drivers to more easily divert to alternative paths (reroute) towards their destination at the CBD and, hence, avoid or experience less congestion delays. Such a plan can offer a more robust and sustainable urban development, particularly in the presence of rapid population growth.

Next, the optimization of *ATT* is investigated for different population sizes and available levels of budget, in terms of the fraction of the CBD land area devoted to roads (to enhance the physical intuition of the results), given the zonal widths. This fraction is assumed to be fixed for the periphery and equal to 1/4 of the corresponding value for the CBD area. The average road width in both zones is set equal to $\omega = 3$ m. Population size reflects the density of car trips generated per km^2 per day. The results are summarized as a contour plot in Fig. 7. The plot is divided into six regions: each region represents a distinct level-of-service of the auto mobility in the CBD and the suburban area, as denoted by the ratio $n_z^* = n_z/n_{crz}$ of the current accumulation to the critical accumulation (when $n_z^* > 1$, then hypercongestion occurs).

The data points of the plot in region A denote free-flow conditions ($ATT < 6$ min) for the whole duration of the period, while data points in region F denote cases where regions reach gridlock ($ATT > 30$ min), at the value of jam density where vehicles stop moving. The evolution of n_z^* over time in the CBD and suburban area is graphically illustrated in Fig. 7 for several cases. Planners should arguably decide to design the urban network to operate in regions B, C or D. Region A mostly entails a huge infrastructure

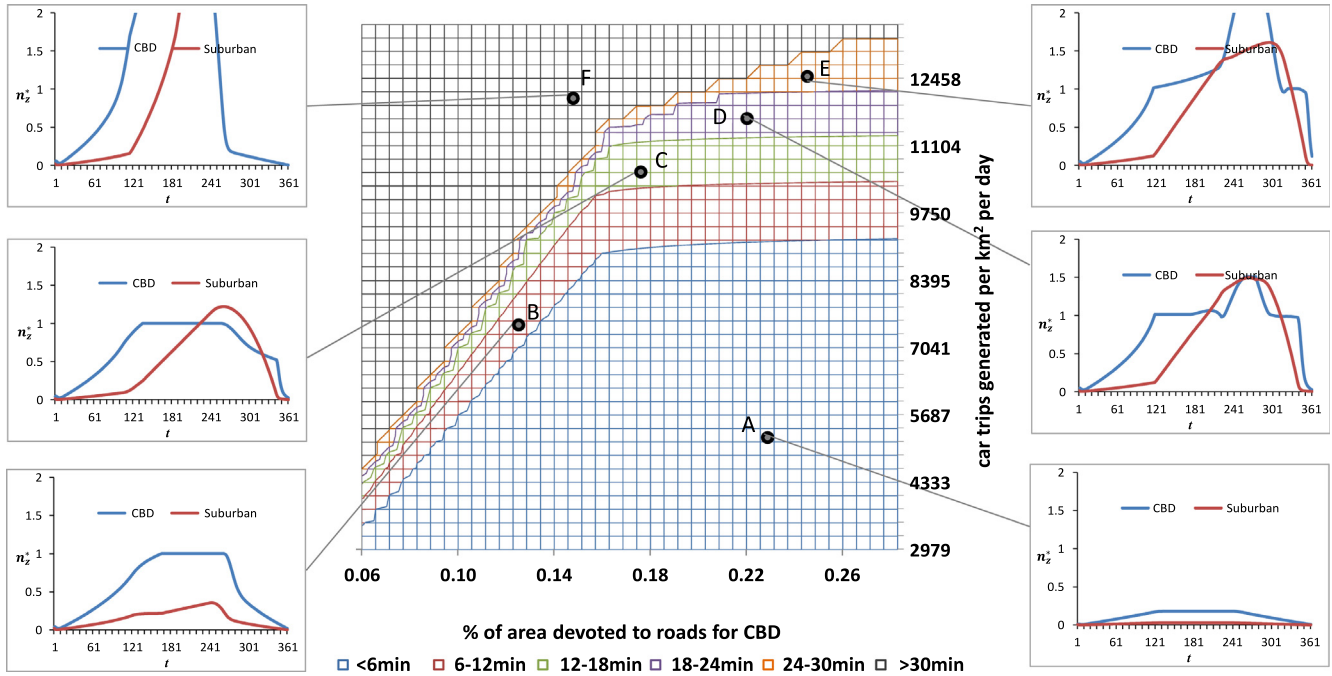


Fig. 7. Contour plot of ATT for different values of density of car trip generation (in trips per km² per day) and fractions of CBD area devoted to roads (this fraction is 4 times smaller for the suburban area in all cases).

investment cost, as reflects the high portion of CBD land devoted to roads, while regions E and F experience large reductions of the level-of-service, namely, hypercongestion conditions ($n_2^* > 1$) for a considerable length of the peak period.

Furthermore, the contour plot shows that there are some areas in the upward-moving part of the regions B, C, D and E where a small increase in the car trip density can significantly increase the level of congestion and reduce level-of-service. These non-linear changes are observed for the smaller fractions, up to about 15%. Hypercongestion arises in region C (for the suburban area) and regions D, E and F (for both the CBD and suburban area), when time approaches the mid of the peak period (after $t = 200$ time units). Thus, the allocation of land (for residence, employment and transport) per unit area and the intra-day evolution of congestion dynamics play an important role on how the vehicle accumulation degrades throughput, given the city size and network operating characteristics. These findings suggest that cities would likely select to trade investment cost for increased levels of robustness in network operating conditions, away from those areas where the cost impact of small fluctuations in demand or capacity cannot be predicted (e.g. due to special events, accidents, etc.). In this way, planners can make decisions on what percentage of car trips should be shifted to public transport to avoid excessive system delays. In all the results presented in Fig. 7, an advanced control of traffic signals has been applied, as it will be described in the next subsection.

5.3. The effect of perimeter control

In the case of highly congested networks (for instance, due to high population growth relative to network capacity, as shown in Fig. 6), the solution of increasing the urban block (or network) density (Section 5.2) can be impractical due to severe fiscal, land availability and other constraints. The deployment of improved traffic control strategies may have an effect similar to that of building more capacity, since it allows more vehicles to use the system within a given period (Button, 2010). An advanced, real-time

perimeter control strategy is required to prevent overcrowding in the CBD, through the demand-responsive metering of access to maintain the auto mobility at a stabilized level. For instance, longer red times can be applied into the traffic signal settings across the perimeter of the CBD area during peak hours for phases which direct vehicles to the center.

By employing the MFD approach, a simple way to deploy this strategy is to monitor the system so that, when traffic density passes its critical value, to restrict entry flow from the suburban area to the CBD. In this way, the oversaturated traffic flow is converted into queues and hypercongestion is eliminated by the end of the peak period. In the case where both regions of the city are highly congested, a more complex strategy should be developed. This control strategy is formulated as an optimization problem, where optimal control values $x(t)$ are identified to minimize the ATT, as follows:

$$\min_{x(t)} ATT = \left(\int_T (n_1(t) + n_2(t)) dt \right) / \rho P$$

subject to
 Dynamic equations 4(a)–4(d)

$$0 \leq x(t) \leq 1.5, \quad \forall t$$

$$0 \leq n_1(t), n_2(t), \quad \forall t \tag{5}$$

The above problem is solved by using the generalized reduced gradient algorithm, which is a well-known, efficient and robust methodology for solving non-linear programs of general structure. For demonstration purposes, the outcome of Fig. 6b is compared with the ATT yielded after implementing the perimeter control, for different population sizes. The results are summarized in Fig. 8 for initial population 1.3 million and the same network settings. The plots show the evolution of the number of vehicles in each of the two regions (CBD and suburban area) over time and the value of parameter $x(t)$, for three population sizes (1.65, 1.85 and 1.95 million), with and without perimeter control.

While the perimeter control only slightly improves the ATT for smaller population sizes, the benefits are found to become signifi-

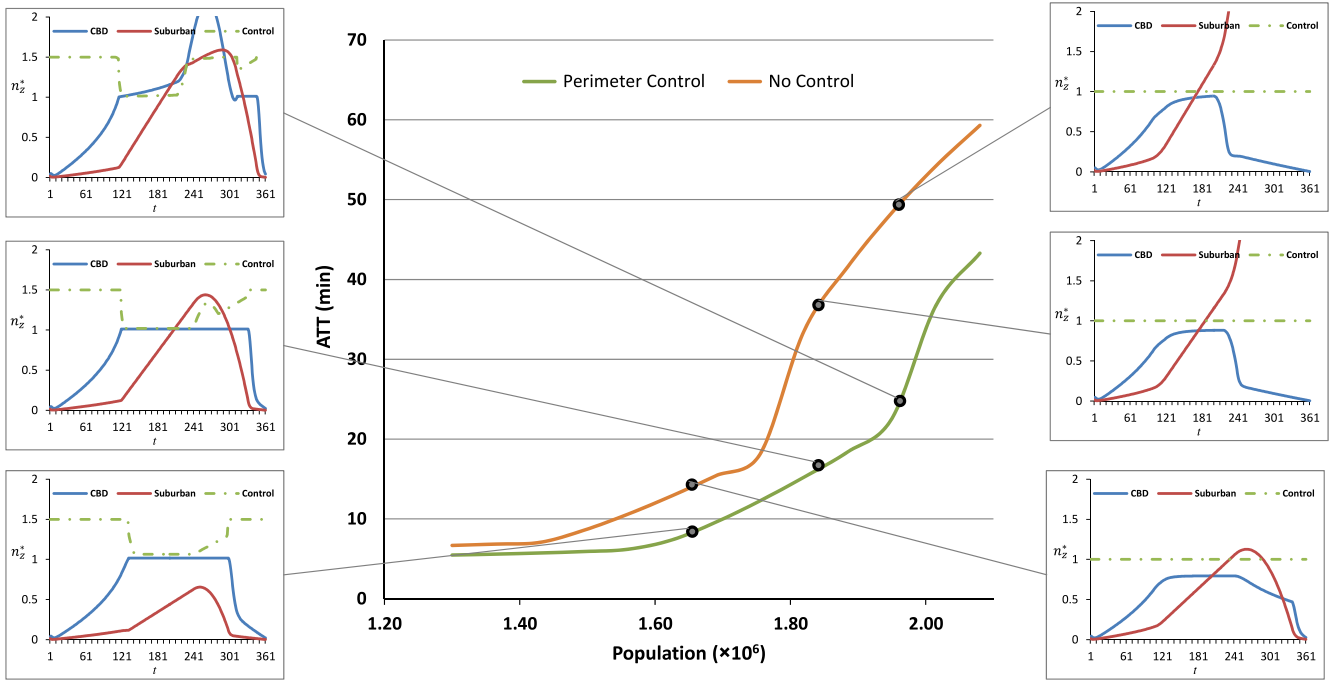


Fig. 8. Results of ATT as a function of population growth for a given network structure with and without perimeter control between the CBD and the suburban area.

cantly larger for higher population sizes. Note that the system does not reach gridlock even for 50% population growth (1.95 million) once the perimeter control is applied, while increased congestion is present for a much smaller population growth (1.65 million) without control (i.e., $x(t) = 1 \forall t$). In the case of perimeter control, hypercongestion occurs in the suburban area only for population of 1.85 and 1.95 million, and n_2^* always returns to the uncongested state at the end of the period. On the contrary, hypercongestion always occurs in the suburban area without control, while n_2^* takes very large values for $t > 200$ time units and it never returns to the uncongested state for populations of 1.85 and 1.95 million.

These results indicate that a better utilization of existing road capacity through installing an advanced traffic signal control software can provide an efficient and relatively low-cost policy option to serve the growing amount of car trips and retain the auto mobility within the city center at a desired level. This option is preferable than the costly and usually impractical solution of increasing network density. Such a policy may be regarded as a desirable alternative (no-toll) second-best optimal strategy compared to the first-best optimum network capacity provision with marginal-cost pricing. This particularly holds in cases where dynamic congestion charging is infeasible due to very expensive (high-technology) investment, increased administrative costs for operating the system and limited enforcement capacities. Other dynamic traffic management schemes such as time-varying parking fees could also be adopted to enhance time cost reductions. Finally, in the case of high demand increase, more sustainable, multi-modal provision strategies may be necessary, in order to shift consumers to other modes, such as bus and metro (out of the road network). The latter issue is outside the scope of this paper, but it definitely constitutes an important research priority.

5.4. Results for typical settings of major cities

The previous findings showed that relatively small changes in the network capacity (either due to provision or technology enhancement) can result in significant reduction in trip costs. In the present subsection, this is demonstrated by employing typical

settings for large metropolitan areas worldwide, including London, Johannesburg, Mexico City, New Delhi and Mumbai. For the given population, car ownership index, value of time (based on income data from the World Bank), spatial characteristics (R_1, A_1 and R_2, A_2) and total available budget (as calculated in terms of the TLK) of these urban areas, optimal plans are estimated with respect to changing network design parameters. Other parameters, such as those referring to link and intersection variables, and infrastructure costs, are assumed to be the same (as defined in previous subsections) for comparison purposes. In order to represent each city according to the simplified concentric urban model (Section 2), a uniform block structure is assumed, without regarding the complexity associated with the hierarchy of road network.

Each city is simulated as having two concentric zones, i.e., the inner (CBD) area (zone 1) and the outer (residential) area (zone 2). For all the cities, the total period of analysis is set equal to 360 time intervals (6 h), with a peak period of 240 intervals (4 h). Based on equation (1), the total lane-km of each city can be expressed as:

$$TLK = L_1 + L_2 = 2(N_{b1} + \sqrt{N_{b1}})L_{b1}\ell_1 + 2(N_{b2} + \sqrt{N_{b2}})L_{b2}\ell_2, \quad (6)$$

By assuming (e.g., based on map observations) the average block length of zone 2, L_{b2} , the average lane number of zone 2, ℓ_2 , and the relationship between ℓ_1 and ℓ_2 ($\lambda = \ell_1/\ell_2$) as given, then, for fixed TLK (budget constraint), a number of optimal plans can be obtained for different values of ℓ_1 . The average block length of zone 1, L_{b1} , can be easily estimated, based on Eq. (6), as follows:

$$L_{b1} = \frac{2\lambda A_1 \ell_2 L_{b2}}{TLK \ell_2 - 2\lambda \sqrt{A_1} \ell_2 L_{b2} - 2A_2 \ell_2 - 2\sqrt{A_2} \ell_2 L_{b2}} \quad (7)$$

Table 8 presents the results of optimal plans for the settings corresponding to each city. In addition to the performance measures used in the previous subsections, the measure of congestion index is also calculated here. This index is obtained from the ratio τ/τ_{ff} of the estimated total travel time ($\tau = \tau_1 + \tau_2$) for the whole period to the corresponding total travel time τ_{ff} at free-flow conditions (where $n_{max}/n_{cr} \leq 0.1$). Moreover, the level of congestion index depicts the magnitude at which the trip cost is underestimated

Table 8Results of optimization plans for typical settings corresponding to major cities for different $\lambda = \ell_1/\ell_2$ values. Source for population and geographical data: www.urban-age.net.

		London	Johannesburg	Mexico City	New Delhi	Mumbai
Inner city population		2,771,700	713,140	3,839,772	6,171,798	7,361,549
Outer city population		4,416,400	2,512,468	4,881,144	7,678,709	4,616,901
Inner city area (sq.km)		319	314	306	314	215
Outer city area (sq.km)		1253	1330	1178	1483	223
Total lane-km		14,676	7519	10,350	24,885	7917
L_{b1}	Current ($\lambda = 1.25$)	0.141	0.455	0.150	0.087	0.078
	Optimum $\lambda = 1.25$	0.090	0.377	0.139	0.101	0.077
	Optimum $\lambda = 1.50$	0.108	0.455	0.166	0.121	0.093
	Optimum $\lambda = 2.00$	0.134	0.500	0.223	0.123	0.124
L_{b2}	Current ($\lambda = 1.25$)	0.500	0.500	0.460	0.275	0.450
	Optimum $\lambda = 1.25$	0.500	0.500	0.500	0.500	0.500
	Optimum $\lambda = 1.50$	0.500	0.500	0.500	0.500	0.500
	Optimum $\lambda = 2.00$	0.500	0.500	0.500	0.480	0.500
ℓ_1	Current ($\lambda = 1.25$)	1.700	1.313	1.250	1.563	1.250
	Optimum $\lambda = 1.25$	1.313	1.250	1.250	2.250	1.250
	Optimum $\lambda = 1.50$	1.575	1.500	1.500	2.700	1.500
	Optimum $\lambda = 2.00$	2.000	2.000	2.000	3.000	2.000
ℓ_2	Current ($\lambda = 1.25$)	1.360	1.050	1.000	1.250	1.000
	Optimum $\lambda = 1.25$	1.050	1.000	1.000	1.800	1.000
	Optimum $\lambda = 1.50$	1.050	1.000	1.000	1.800	1.000
	Optimum $\lambda = 2.00$	1.000	1.000	1.000	1.500	1.000
TSC (bmu)	Current ($\lambda = 1.25$)	6.735	1.713	3.447	4.217	1.909
	Optimum $\lambda = 1.25$	4.994	1.500	2.936	3.942	1.835
	Optimum $\lambda = 1.50$	4.853	1.162	2.822	3.916	1.815
	Optimum $\lambda = 2.00$	4.709	1.122	2.742	3.893	1.789
ATT (min)	Current ($\lambda = 1.25$)	35.169	36.478	30.381	27.583	28.034
	Optimum $\lambda = 1.25$	22.897	22.819	23.258	19.965	25.694
	Optimum $\lambda = 1.50$	21.900	15.734	21.671	19.237	25.064
	Optimum $\lambda = 2.00$	20.911	15.279	20.558	18.679	24.240
Spatial equity	Current ($\lambda = 1.25$)	0.287	0.254	0.303	0.436	0.105
	Optimum $\lambda = 1.25$	0.536	0.463	0.413	0.614	0.121
	Optimum $\lambda = 1.50$	0.570	0.747	0.449	0.646	0.124
	Optimum $\lambda = 2.00$	0.632	0.783	0.480	0.712	0.129
Fiscal equity	Current ($\lambda = 1.25$)	0.123	0.672	0.006	0.167	5.748
	Optimum $\lambda = 1.25$	0.750	0.606	0.165	0.301	6.581
	Optimum $\lambda = 1.50$	0.750	0.606	0.165	0.301	6.581
	Optimum $\lambda = 2.00$	0.887	0.521	0.165	0.652	6.581
Congestion index	Current ($\lambda = 1.25$)	4.079	4.838	3.901	3.103	4.572
	Optimum $\lambda = 1.25$	2.601	2.898	2.946	2.202	4.181
	Optimum $\lambda = 1.50$	2.481	0.980	2.733	2.115	4.076
	Optimum $\lambda = 2.00$	2.359	0.949	2.583	2.041	3.940

when ignoring the effects of congestion in the network. This magnitude, whose values range from 0.95 to 4.84 (Table 8), provides supporting evidence of the need to appropriately measure the congestion cost by the proposed MFD model.

The findings verify that, in certain cases, there can be remarkable reduction of the congestion externalities through relatively small changes in specific network parameters. The reduction of the TSC and ATT is magnified with the increase of the λ value (from 1.25 to 2.0). This outcome signifies the importance of increasing the capacity, in terms of the average lane number, of the (higher density and more congested) inner area, relative to that of the outer area, in order to decrease congestion delays and total travel times. Nonetheless, the improvements considerably vary with the characteristics of each city. More specifically, for the case where $\lambda = 2.0$, Mumbai and New Delhi, which have increased population densities, experience the lowest reduction in the TSC (−6.3% and −7.7%), ATT (−13.5% and −32.3%) and congestion index (−13.8% and −34.2%). On the contrary, Johannesburg and London, which have relatively lower population densities, experience the highest reduction in the TSC (−34.5% and −30.1%), ATT (−58.1% and −40.5%) and congestion index (−80.4% and −42.2%).

However, the resulting optimal plans lead to larger spatial inequity in London and Johannesburg (increasing by 2–3 times the relevant metric), compared to the other cities. The optimal plans also result in larger fiscal inequity in all the cities. The equity is generally decreasing by allocating more capacity in the inner zone, in

terms of increasing λ value (from 1.25 to 2.0). This reduction of equity can be attributed to the spatial constraints imposed to the boundaries of both zones as well as the budget constraints, in terms of keeping the TLK fixed. Other policies, such as investment on a public transit mode and imposing time-of-day tolls to shift the rush hour to other times, could further reduce congestion costs (McDonald, 2009). Subsequently, they will decrease the amount of infrastructure capital redistribution needed to improve efficiency and the adverse effects on equity.

6. Conclusions

This paper addresses the problem of optimal city size and network structure by considering the traffic congestion dynamics in large urban areas. In contrast with traditional urban economic models, the MFD approach allows the parsimonious time-varying modeling of the non-monotonic travel cost vs. flow relationship, including hypercongestion, at the level of homogenous urban regions, which is consistent with the physics of traffic and economic theory. In particular, it treats the complexity of congestion dynamics using only a handful of physically sound assumptions and it is tractable, involving relatively few degrees of freedom. It has reduced needs for data collection and processing, employing basic flow and cost measures as inputs, which can be easily calibrated, compared to other models that necessitate detailed network modeling, origin–destination trip estimation and traffic assignment

procedures. Hence, it makes the given problem easy to solve, while keeping its realistic size and insightfulness regarding the analysis of land use and the urban economy.

It is shown that the identification of a stable MFD during the peak period can yield optimal urban plans, in terms of the total and inter-zonal allocation of land and network resources, without or subject to various constraints, and in response to alternative land use and policy scenarios. The findings suggest the increased mixing of commuting within the CBD and compactness of the urban area, through boundary growth control and increasing block density (network connectivity), as travel demand grows, to maximize efficiency and improve spatial equity.

Besides, the model can facilitate the effective implementation of policies which (i) for a given city size, make the system more robust, based on critical fractions of land allocated to transport, and (ii) for a specific network structure, fine-tune resulting plans through deploying simple control strategies, only relying on some critical value of traffic density. In the latter case, relatively low-cost investments for installing and operating advanced traffic control technology across the perimeter of the CBD, compared to dynamic congestion pricing and massive investment on a public transport mode, can lead to significant improvement in the desired level of auto mobility.

Increasing the accuracy of predictions of population growth and travel demand, particularly in rapidly urbanized areas, can help to suitably designate the boundaries and allocate network resources in each zone so that enhance the system capacity and sustainable metropolitan development. Moreover, redistribution of the existing road capacity among the urban zones may significantly reduce congestion costs, without need for increasing network density in the whole city. The amount of these improvements relies on the present level of available budget (road infrastructure stock), initial level of congestion, and spatial and socio-demographic characteristics of each metropolitan area.

The proposed model can be extended to include a range of urban forms, such as multiple centers and periphery with mixed rather than exclusive use. Further extensions may consider the hierarchical structure of the road network, departure time and activity location choices, time-varying congestion charging of commuters entering the CBD and integration of a public transport mode. Such travel demand management policies as staggered/flexible work hours may increase the length of peak period through temporal spreading of congestion (Henderson, 1981). In turn, this would influence the design and resulting performance of optimal land and network resource allocation strategies among urban zones, in terms of further reducing total required budget and delay externalities.

Acknowledgments

The authors are grateful to the editor William Strange and two anonymous reviewers for their insightful and constructive comments, and Prof. Eric Gonzales for proof-reading the paper. T. Tsekeris acknowledges the Urban Transport Systems Laboratory of École Polytechnique Fédérale de Lausanne in Switzerland for funding and hosting him as an academic visitor to conduct a part of this research.

References

Anas, A., Arnott, R., Small, K.A., 1998. Urban spatial structure. *Journal of Economic Literature* 36 (3), 1426–1464.

- Anas, A., Liu, Y., 2007. Regional economy, land use, and transportation model (RELU-TRAN): formulation, algorithm design, and testing. *Journal of Regional Science* 47 (3), 415–455.
- Anas, A., 2011. Metropolitan Decentralization and the Stability of Travel Time. Working Paper WP5, Department of Economics, State University of New York at Buffalo, Amherst, NY.
- Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak period congestion: a traffic bottleneck with elastic demand. *American Economic Review* 83 (1), 161–179.
- Arnott, R., 1998. Congestion tolling and urban spatial structure. *Journal of Regional Science* 38 (3), 495–504.
- Arnott, R., 2007. Congestion tolling with agglomeration externalities. *Journal of Urban Economics* 62 (2), 187–203.
- Arnott, R., Inci, E., 2010. The stability of downtown parking and traffic congestion. *Journal of Urban Economics* 68 (3), 260–276.
- Arnott, R., de Palma, E., 2011. The corridor problem: preliminary results on the no-toll equilibrium. *Transportation Research Part B: Methodological* 45 (5), 743–768.
- Branston, D., 1976. Link capacity functions: a review. *Transportation Research* 10 (4), 223–236.
- Button, K., 2010. *Transport Economics*, third ed. Edward Elgar, Cheltenham, UK.
- Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., Hicks, J., 2011. *Dynamic Traffic Assignment: A Primer*, Transportation Research Circular E-C153. Transportation Research Board, Washington, DC.
- Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41 (1), 49–62.
- Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 42 (9), 771–781.
- Daganzo, C.F., Gonzales, E.J., Gayah, V.V., 2011. Traffic Congestion in Networks, and Alleviating it with Public Transportation and Pricing, Working Paper UCB-ITS-VWP-2011-7. Institute of Transportation Studies, University of California at Berkeley, CA, UC Berkeley Center for Future Urban Transport.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42 (9), 759–770.
- Godfrey, J.W., 1969. The mechanism of a road network. *Traffic Engineering and Control* 11 (7), 323–327.
- Henderson, J.V., 1981. The economics of staggered work hours. *Journal of Urban Economics* 9 (3), 349–364.
- Henderson, J.V., 2010. Cities and development. *Journal of Regional Science* 50 (1), 515–540.
- Lago, A., Daganzo, C.F., 2007. Spillovers, merging traffic and the morning commute. *Transportation Research Part B: Methodological* 41 (6), 670–683.
- Lasher, W.R., 2008. *Practical Financial Management*, sixth ed. Cengage South-Western Learning, Mason, OH.
- Little, J.D., 1961. A proof of the queueing formula $L = \lambda W$. *Operations Research* 9 (3), 383–387.
- Lo, H.K., Szeto, W.Y., 2005. Road pricing modeling for hyper-congestion. *Transportation Research Part A: Policy and Practice* 39 (7–9), 705–722.
- McConnell, V., Walls, M., Kopits, E., 2006. Zoning, TDRs and the density of development. *Journal of Urban Economics* 59 (3), 440–457.
- McDonald, J.F., d'Ouille, E.L., Liu, L.N., 1999. *Economics of Urban Highway Congestion and Pricing*. Kluwer, Dordrecht, The Netherlands.
- McDonald, J.F., 2009. Calibration of a monocentric city model with mixed land use and congestion. *Regional Science and Urban Economics* 39 (1), 90–96.
- Ohta, H., 2001. Probing a traffic congestion controversy: density and flow scrutinized. *Journal of Regional Science* 41 (4), 659–680.
- Peeta, S., Ziliaskopoulos, A.K., 2001. Foundations of dynamic traffic assignment: the past, the present and the future. *Networks and Spatial Economics* 1 (3–4), 233–265.
- Pigou, A.C., 1920. *The Economics of Welfare*. Macmillan, London, UK.
- Ross, S.L., Yinger, J., 2000. Timing equilibria in an urban model with congestion. *Journal of Urban Economics* 47 (3), 390–413.
- Small, K., Chu, X., 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37 (3), 319–352.
- Verhoef, E.T., 1999. Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. *Regional Science and Urban Economics* 29 (1), 341–369.
- Verhoef, E.T., 2001. An integrated dynamic model of road traffic congestion based on simple car-following theory: exploring hypercongestion. *Journal of Urban Economics* 49 (3), 505–542.
- Verhoef, E.T., 2005. Speed-flow relations and cost functions for congested traffic: theory and empirical analysis. *Transportation Research Part A: Policy and Practice* 39 (7–9), 792–812.
- Vickrey, W., 1969. Congestion theory and transport investment. *American Economic Review (Papers and Proceedings)* 59 (2), 251–260.
- Wheaton, W.C., 2004. Commuting, congestion, and employment dispersal in cities with mixed land use. *Journal of Urban Economics* 55 (3), 417–438.