



# Optimizing Traffic System Performance with Environmental Constraints: Tolls and/or Additional Delays

Xin Lin<sup>1</sup>  · Chris M. J. Tampère<sup>1</sup> · Stef Proost<sup>2</sup>

Published online: 3 August 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

Environmental concerns are increasingly important targets in urban traffic management. The way traffic spreads over routes in a network can affect substantially the environments well as the level of congestion. This paper explores the selection and location of control measures that influence route choice aimed at optimizing traffic system performance subject to environmental constraints (OSP-EC problems). We address two groups of link-based control instruments commonly considered in urban traffic management: tolls and additional delays. Tolls penalize drivers monetarily (e.g. congestion charging) while additional delays represent those measures directly increasing travel time (e.g. speed limits) on the controlled links. We first identify the parallels and differences of how tolls and delays reallocate traffic flows over routes and links in a network, how they affect total system cost and total emission, and finally how they produce different solutions for the OSP-EC problem. When only tolls can be used, one intuitive solution to the OSP-EC problem is a set of tolls that superimposes environmental shadow prices for actively constrained links onto the marginal system cost pricing that is to be charged on all links. When only delays are considered, the selection of the best locations to implement control variables is more complex. We show that the monotonicity of emission factors with speed plays a crucial role for the selection of the type of control variables. If both tolls and delays are available, tolls work better than additional delays in most cases. The theoretical results can contribute to the development of efficient algorithms for solving the OSP-EC problem, and help authorities to solve this policy problem.

**Keywords** Traffic emission · Bi-level · Toll · Delay · Speed reduction · OSP-EC problem · Environmental constraint

---

✉ Chris M. J. Tampère  
chris.tampere@kuleuven.be

## 1 Introduction

The environmental impact of traffic is one of the critical issues in urban areas. Conventionally, the primary goal of traffic management is traffic performance improvement, but in certain situations the environment could be the main reason for traffic management strategies. Apart from the primary target however, such strategies will not only influence the environmental aspects but also other goals, such as traffic performance. These impacts could be positive or negative (Nagurney 2000).

Various environmental goals exist in practice. For example, the reduction of total emission is a primary goal for most pollutants. There are the pollutants influencing large-scale areas and with a long-term effect, such as CO<sub>2</sub>, CH<sub>4</sub> and other greenhouse gasses. For these pollutants it does not matter where the reduction is organized so it is best to deal with them across all sectors and areas. Things are different for pollutants that are short lived, having a local but harmful influence on human health, such as NO<sub>x</sub>, CO and Particulate Matters (PM) (Du et al. 2012; Aziz et al. 2017) as well as noise (Kaddoura et al. 2017). It is natural to address such pollutants locally. As for these pollutants the health effects and damage is very uncertain, one prefers to limit the concentration of the harmful pollutants to a maximum level. In other words, the authorities need to impose local environmental constraints on certain links or in certain areas. This is the main topic of this paper.

This paper takes the local approach and focuses on the traffic flow management approach for satisfying hard environmental constraints at the link level. The environmental constraints satisfaction problem is one important topic in traffic management (Szeto et al. 2012; Zhang et al. 2013). In the literature there are approaches that focus on the environmental aspects in local traffic flow management, but only few of them formulated the traffic management problem under environmental constraints in a proper way. Some contributions take the environmental aspects as the primary target, for example, considering it as the only objective to be minimized (Benedek and Rilett 1998; Sugawara and Niemeier 2002; Szeto et al. 2008; Lin et al. 2010b; Zegeye et al. 2010; Haas and Bekhor 2017), or formulate it as the User Equilibrium with side constraints problem (UE-SC). The UE-SC problem seeks a traffic User Equilibrium assignment which satisfies the link constraints or zonal aggregate constraints. A corresponding set of measures penalizing travelers' travel cost is used as a trigger for travelers to reroute so that the resulting traffic assignment respects the UE-SC constraints. Different algorithms exist for the UE-SC problem (Ferrari 1995; Larsson and Patriksson 1995, 1999; Larsson et al. 2004; Zhong et al. 2011; Chen et al. 2011; Li et al. 2012). However, neither minimizing emission nor the UE-SC approach consider other traffic system performance criteria. To not underutilize the control measures, authorities should not just consider local constraints as targets, as they then fail to further improve the traffic system under these constraints. The two existing approaches thus underutilize the potential improvement of traffic system and in some cases, deteriorate the system performance, such as causing a large increase of the total system cost.

To involve both the environmental aspects and traffic aspects, multiple-criteria decision-making (MCDM) (Hendriks et al. 1992) is commonly used. The environmental aspects are then considered in a Pareto-efficiency framework as one of many

objectives besides other traffic performance criteria (Johansson 1997; Yin 2002; Johansson-Stenman 2006; Lin et al. 2010a; Aziz and Ukkusuri 2012; Wang et al. 2015; Mascia et al. 2017; Wang and Connors 2018). However, emissions in Pareto-efficient solutions are unbounded, and hence cannot guarantee that emissions are under the constraints. Besides, even in solutions of the set in which all the emissions are under the constraints, these solutions cannot guarantee that the control measures are not underutilized because there may be another Pareto-efficient set with better system traffic performance criteria and also with all emission constraints satisfied. Other researchers indeed do consider the environmental constraints together with other system performance (Yang et al. 2005, 2008; Jakkula and Asakura 2009; Feng et al. 2010), but the control measures are then limited to traffic demand management measures only. To study the differences of local traffic flow management measures for satisfying environmental constraints, we adopt in this work a framework for assessing the impact of environmental constraints on traffic system proposed by Lin et al. (2016). The main idea is to understand how environmental constraints reduce the feasible settings of available control measures for achieving optimal traffic performance. This allows looking for control measures satisfying the environmental constraints while inflicting minimal damage to the traffic performance. We denote such problem as the Optimal System Performance with Environmental Constraints (OSP-EC) Problem. The OSP-EC problem is a Stackelberg competition problem (Fisk 1984) because users themselves do not take the environmental goals into account. The controller needs to apply corresponding control measures to let users' behavior satisfy the environmental constraints.

Various control measures can be used for achieving traffic performance and/or environmental goals. Different control measures to reach the same environmental goal can have different impacts on the system performance. When considering control measures that influence route choice by changing travel costs of links, we here distinguish two families of control measures: tolls and additional delays.

Tolls are widely applied in literature on traffic management. The travelers have to pay a monetary penalty to access certain links (Yang and Bell 1997) or hotspots (Solé-Ribalta et al. 2018) so their travel decisions such as route choices are influenced. One practical example is the London congestion charging. If there is no environmental constraint, the first-best toll scheme, in which the marginal external congestion cost is tolled on each link, is guaranteed to guide the users to the so-called system optimum assignment (Yang and Huang 1998). Meanwhile, toll is also used in the traffic management for environmental reasons (Ferrari 1995; Yin and Lawphongpanich 2006; Chen and Yang 2012; Li et al. 2014; Kickhöfer and Nagel 2016). Moreover, tolls equal to the shadow prices for the side constraints are used for sustaining the traffic flow constraints in UE-SC problems (Larsson and Patriksson 1995). As a result, toll is one of the important control measures for traffic management related to environmental considerations.

Besides tolls, route choice can also be controlled by another family of practical measures that affect travel times. One typical example is the speed limit, which authorities may impose on links to make these links less attractive (Yang et al. 2012; Wang 2013). Other examples are speed control humps and other small infrastructural measures (Barbosa et al. 2000; De Borger and Proost 2013). We denote from here on this family of measures with the term “additional delays” referring to the

difference of time costs before/after the measure's implementation. Note that apart from their indirect impact on local emissions through rerouting of traffic, this category of measures is often considered for its direct impact, as speed and accelerations directly affect the emissions per vehicle as well (Keller et al. 2008; Baldasano et al. 2010; Madireddy et al. 2011). However, the impact of those measures on emission is less clear. For example, if the speed limit reduces from 50 km/h to 30 km/h, Madireddy et al. (2011) found that NO<sub>x</sub> and CO<sub>2</sub> can reduce by 25% in a urban street but Int Panis et al. (2011) indicated that CO<sub>2</sub> would increase and the change of NO<sub>x</sub> depended on the driving cycles.

In theoretical or even in practice, the time penalty (additional delay) is sometimes considered as an alternative of tolls because both measures increase the traveler's actual travel costs (Larsson and Patriksson 1999; Zhong et al. 2011). However, toll and delay may result in different traffic flows and different total travel costs (Yang et al. 2012; Wang 2013). Moreover, the emissions under the two control measures can be different. Yang et al. (2012) indicated that with arbitrary emission function, speed limits may perform better than tolls/subsidies regarding total emission and total travel time. Wang et al. (2015) explored the potential of a hybrid scheme containing tolls and speed limits, for a Pareto-efficient flow and speed patterns that minimize the total travel time and total emission. Those differences but the potentials for the hybrid schemes suggest that it is necessary to investigate and compare the two control measures, to better understand the mechanisms of the two measures influencing the traffic system, especially when considering environmental aspects.

Based on the scopes discussed above, this study adopts the OSP-EC framework to analyze the two different local traffic flow control measures, tolls and additional delays, in the local environmental constraints satisfaction problem. The study assumes static, deterministic User Equilibrium conditions. It is worth noting that in general the term of "system cost" could include any objectives, like a total travel cost (or efficiency) and a global emission cost component (e.g CO<sub>2</sub> or local pollutants). Without loss of generality, in this paper we use the "system cost" to indicate the travel costs (especially under the environmental constraints).

The contributions of this paper include

- Theoretically analyzing the differences between tolls and additional delays in traffic link flow patterns, total system cost, emission and the OSP-EC problem.
- Theoretically generalizing the proper toll or/and additional delay sets for optimizing the total system time cost with environmental constraints; identification of the role of monotonicity of emission factors herein.
- Mathematically limiting the candidate locations for implementing the two control measures, and suggesting the preferred type of control measures for the OSP-EC problem, to avoid unnecessary search directions when solving more complex optimization problems.
- Discussing application of the OSP-EC problem in practice, to help authorities select correct control targets and measures while considering environmental constraints.
- Illustrating practical applications of the theoretical concepts discussed in this paper, to help the authorities select better control measures and locations for the general OSP-EC problem.

The reminder of this paper is organized as follows. In Section 2, we show why tolls and additional delays should be considered separately in OSP-EC problems. To this end, we compare their impacts on traffic flow, total system cost, and emissions, based on mathematical formulations and a theoretical case study. We also argue that the monotonicity of emission factor as a function of average speed influences the control parameter selection. Section 3 and Section 4 discuss strategies for the OSP-EC problem in scenarios with only tolls, with only additional delays, and with both measures available. However the two sections use different emission factor functions: either monotonically decreasing (Section 3) or non-monotonically (Section 4). The practical applications are presented in Section 5, and we discuss conclusions and suggestions for further research in Section 6.

## 2 The Parallels and Differences Between Toll and Delays

In this section, we discuss the parallels and differences between tolls and additional delays measures. It will be shown that tolls and additional delays have the same effects in rerouting traffic flows. However, even though different types of control may thus yield similar link flow volumes, they do so while having different impacts on total system cost and on emissions. Moreover, we will show how the monotonicity of the emission factor function influences the emission and hence the solutions of OSP-EC problems. This indicates that the OSP-EC problem should be dealt with separately according to the different control measures and different emission factor relationships, as we do later in Sections 3 and 4.

### 2.1 On Re-Allocating Traffic Flow

In this sub-section, we argue that both toll and additional delay are effective control measures for rerouting traffic, and that their influence on the flow distribution is equivalent (however they do so at different system cost and different total emissions, as will be shown in subsequent sub-sections).

Considering a standard network  $\langle \mathcal{N}, \mathcal{A} \rangle$ , with  $\mathcal{N}$  and  $\mathcal{A}$  being the set of nodes and links, respectively;  $l_a$  is the length of link  $a$ ;  $f_a$  is the link flow on link  $a$ , combined for all links in the vector  $\mathbf{f} = (f_a, a \in \mathcal{A})^T$ .  $\mathcal{W}$  is the set of origin-destination pairs  $w \in (\mathcal{N} \times \mathcal{N})$  with demand  $D_w$ . Let  $\mathcal{R}_w$  denote the possible simple routes between  $w$ , and  $\mathbf{q}_w = (q_{wr}, r \in \mathcal{R}_w)^T$  denote the set of route flows of OD  $w$ . All route flows are defined as  $\mathbf{q} = (\mathbf{q}_w, w \in \mathcal{W})^T$ . Vectors  $\mathbf{q}$  and  $\mathbf{f}$  should satisfy the basic topological constraints, non-negative flow constraints, demand constraints. Let  $\Omega_Q$  and  $\Omega_F$  denote the feasible set of route flows and link flows, respectively, and can be defined as below (Yang et al. 2012)

$$\Omega_Q = \{ \mathbf{q} \mid q_{wr} \geq 0, \sum_{r \in \mathcal{R}_w} q_{wr} = D_w, r \in \mathcal{R}_w, w \in \mathcal{W} \} \tag{1a}$$

$$\Omega_F = \{ \mathbf{f} \mid f_a = \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}_w} q_{wr} \delta_{wra}, a \in \mathcal{A}, \mathbf{q} \in \Omega_Q \} \tag{1b}$$

Here  $\delta_{wra}$  is an indicator which equals 1 when the route  $wr$  uses link  $a$ , and 0 otherwise.

We introduce Assumptions 1–3, which are commonly considered in static traffic assignment.

- Assumption 1: Users are homogenous and follow deterministic user equilibrium route choice
- Assumption 2: Link cost function  $c_a$  is separable, continuous and strictly monotonically increasing with its link flow  $f_a$ .
- Assumption 3: The OD demands  $D_w$  are inelastic.

Then link cost function on link  $a$  in general depends on flow  $f_a$  on this link, and possibly also on crossing flows on adjacent links (e.g. through priority processes at the intersection). From here onwards however, like in the majority of theoretical traffic network analyses (Yang and Huang 1998; Larsson et al. 2004; Yin and Lawphongpanich 2006; Chen and Yang 2012; Yang et al. 2012; Wang 2013; Liu et al. 2017), we neglect so-called non-separable influences of adjacent link flows, and thus assume separable cost functions that only depend on the flow on link  $a$  itself:  $c_a = c_a(f_a)$  and the cost vector is  $\mathbf{c} = (c_a, a \in \mathcal{A})^T$ , with unit hour. The assumptions above also imply all routes are acyclic and the link flow patterns of User Equilibrium exist and are unique (Smith 1979). In purpose of better explanation, we use inelastic demands for this study to simplify the argumentation and the proofs. However, conclusions of Sections 2, 3 and 4 also hold for surplus maximization when standard elastic demand functions are used that are separable, continuous and monotonically decreasing with users' travel costs. Those proofs for elastic demand are available from the corresponding author.

- Assumption 4: The authorities can charge a toll and/or impose an additional delay on each link.

If there is no environmental constraint, the well-known “first best” toll pricing scheme is guaranteed to optimize system performance. Let  $(\mathbf{t}, \mathbf{d})$  denote the tolls and additional delays on the links, respectively. Both parameters are non-negative and normalized as the same unit of link cost, therefore, in hour. It is important to note that we use general additional delay measures instead of the speed reductions measures. For a detailed discussion about the additional delay measures, see Section 5.1.

- Assumption 5: Both control measures are perfectly implemented, and drivers fully comply with these control measures.
- Assumption 6: The charged toll will be fully returned to society, so they are a pure transfer and should not be considered as a system cost.

These indicate that both the compliance rate of the control measures and the revenue rate of tolls are 1. These are common assumptions in theoretical traffic system analysis (Sheffi 1984; Small and Verhoef 2007). It is worth noting that these theoretical assumptions do not hold exactly in practice: there, drivers may not fully comply with the controllers, and revenue rate could be smaller than 1 when the toll systems have non-negligible transaction cost (Proost and Van Dender 2001).

Under the Assumptions 1–6, the additional link cost on link  $a$  is  $z_a \equiv t_a + d_a$ , where  $t_a, d_a$  are the toll and the additional delay on link  $a$ , respectively:  $\mathbf{z} = \mathbf{t} + \mathbf{d}$ . Based on the assumptions and the concept of deterministic user equilibrium, we easily reach Proposition 1, and Corollaries 1.1–1.4 show the influences on the link traffic flow.

- **Proposition 1:** toll and additional delays have equivalent effects on re-allocating traffic flow, so  $\mathbf{f}(\mathbf{z} - \Delta\mathbf{d}, \Delta\mathbf{d}) = \mathbf{f}(\mathbf{z}, \mathbf{0}) = \mathbf{f}(\mathbf{0}, \mathbf{z})$
- **Corollary 1.1:**  $\frac{\partial \mathbf{f}}{\partial \mathbf{t}} = \frac{\partial \mathbf{f}}{\partial \mathbf{d}} = \frac{\partial \mathbf{f}}{\partial \mathbf{z}}$
- **Corollary 1.2:**  $\frac{\partial f_a}{\partial z_a} = \frac{\partial f_a}{\partial t_a} = \frac{\partial f_a}{\partial d_a} \leq 0 \forall a \in \mathcal{A}$
- **Corollary 1.3:** if  $\frac{\partial f_a}{\partial z_a} = 0 \forall a \in \mathcal{A}$ , then  $\frac{\partial f_b}{\partial z_a} = 0, \forall b \in \mathcal{A}, b \neq a$
- **Corollary 1.4:** if  $\frac{\partial f_a}{\partial z_a} = 0 \forall a \in \mathcal{A}$ , either  $f_a = 0$  or  $f_a = \sum_w D_w$  for  $w$  where  $\sum_r q_{wr} \delta_{wra} > 0$

**Proof:** see Appendix 1

Proposition 1 is obvious since users are homogeneous and only consider their total costs of route choice, so tolls and additional delays have identical effects on user link costs and route costs. Proposition 1 and Corollary 1.1 therefore states the same effects on flows and on the change of flows. Corollary 1.2 states that an increase of additional link cost implies that the flow on that link will always decrease or remain the same. It proves that the basic concept for reducing flow on the target link is increasing the additional link cost on that link. Corollary 1.3 states that it is impossible that a measure on one link would cause flow changes on other links but not on the controlled link itself. This implies that a change of additional link cost of a target link can only directly influence the flow on the target link where the additional cost is changed, and the flows on the other links are indirectly influenced by the flow change of the target link. Corollary 1.4 indicates that if the control measure has no effect on that link’s flow, either the link flow has been already reduced to zero, or the OD pairs using the link do not have alternative routes bypassing the link and that are active under current route costs.

The effect of control measures on link traffic flows discussed above resemble those in the studies on User Equilibrium with side constraints (Ferrari 1995; Larsson and Patriksson 1999), which proposed charging additional cost on and only on the links exceeding the link side constraints. This is also convenient for the practical implementation. For instance, if a town would reduce the emission or traffic flow of the link through the center, the authorities simply add the additional cost on that link, for example, by imposing speed limits.

Furthermore, it can be shown that with elastic demand, there must be no flow on the link  $a$  if  $\frac{\partial f_a}{\partial z_a} = 0$  according to Corollary 1.3, Corollary 1.4 and the monotonicity of the demand functions. In other words, the control measures always work for reducing the target link flow in the elastic demand case, but does not work for the inelastic demand case if the link carries an OD, of which all route flows must use that link (under current costs).



### 2.2 On the Total System Time Cost

In this sub-section, we argue that although both control measures have the same effects on re-allocating traffic flows, they realize the objective in different ways and the system cost in the case of tolls is always lower than the same traffic flow pattern sustained through additional delays.

Toll sustains the target traffic flow by adding monetary costs on each link, so those costs do not directly change the driving patterns. Additional delays increase the actual traffic cost of users, and directly influence their driving patterns, for example, by slowing down the speed. Moreover, the toll is transferred between stakeholders in the system and is hence not lost, whereas delay is a pure (time) cost for the drivers. Hence their effect on total cost differs. Considering that total cost is  $C = (c + d) \cdot f = (c + d)^T f$ , we give Proposition 2 and Corollary 2.1 without proofs.

- **Proposition 2:** Additional delays lead to a higher total travel system cost compared to tolls for sustaining the same target traffic flow pattern.
- Corollary 2.1:

$$\left\{ \begin{aligned} \frac{\partial TC}{\partial t} &= f^T \frac{\partial c}{\partial t} + (c + d)^T \frac{\partial f}{\partial t} = \left( \frac{\partial c}{\partial f}^T f + c + d \right)^T \frac{\partial f}{\partial t} \\ \frac{\partial TC}{\partial d} &= f^T \left( \frac{\partial c}{\partial d} + \frac{\partial d}{\partial d} \right) + (c + d)^T \frac{\partial f}{\partial d} = \left( \frac{\partial c}{\partial f}^T f + c + d \right)^T \frac{\partial f}{\partial d} + f^T = \frac{\partial TC}{\partial t} + f^T \end{aligned} \right. \quad (2)$$

Because  $\frac{\partial f}{\partial d} = \frac{\partial f}{\partial t}$  according to Corollary 1.1 and  $f \geq 0$ , we get  $\frac{\partial TC}{\partial d} \geq \frac{\partial TC}{\partial t}$ . Therefore, a target traffic flow pattern sustained by tolls could reduce the total system cost while the same target flows sustained by additional delays may increase the total system cost. A simple example would be the system optimum assignment. A system optimum assignment gives the minimal system cost if it is sustained by tolls  $(t_{SO}, \mathbf{0})$ . But, the system could have higher cost than no control User Equilibrium if sustained by additional delays  $(0, t_{SO})$ . It is necessary to understand their different effects on the total system time cost to avoid incorrectly using additional delays, because many sustaining target traffic flows strategies exist for the reduction of total system time cost. However, this does not mean additional delays would always make the system worse off, since  $\frac{\partial TC}{\partial t_a} + f_a$  can be smaller than 0 if  $\frac{\partial TC}{\partial t_a} < -f_a$ . A typical example is Braess' Paradox (Braess 1968): Imposing a delay at the middle link of Braess's Paradox network would reduce the total system cost (Wang 2013), but less efficiently than toll would.

### 2.3 On the emission

This sub-section presents a key finding of this paper, showing that although toll and additional delays have the same effects on re-allocating traffic flows, toll may do so while producing less or more emissions compared to additional delays, depending on the sign of the sensitivity of the emission factor for changes in the average speed being



negative (case discussed further in Section 3) or positive (case discussed further in Section 4) respectively.

Proposition 1 and Proposition 2 show that tolls should always work better (in terms of total system cost) than additional delays in sustaining the same target traffic flow pattern. We now compare the influences of the two control measures on emission. Let  $\varepsilon_{ap}$  denote the emission factor with unit gram/veh.km of pollutant  $p$  on link  $a$ , and assume:

- Assumption 7: The average speed  $v_a$  of link flow  $f_a$  is the only variable influencing emission factor  $\varepsilon_{ap}$  through the driving patterns.
- Assumption 8: All vehicles have homogenous emission factors.

In general, an emission factor is related to driving patterns which include speed, speed variances, road condition such as slopes and other parameters (Smit et al. 2010; Szeto et al. 2012). In practice, it is not easy to acquire all the parameters and investigate the relationship among all of them. Therefore, only parameters are selected that are appropriate for the modelling scale used. We are investigating the insights of tolls and additional delays in a static (or: stationary) traffic network (corresponding to a macroscopic aggregate scale, with time-averaged flow representation), so only average speed is considered. The average speed is accurate enough for estimating the emission factor at the urban level and widely used in macroscopic and mesoscopic traffic emission models (Smit et al. 2010). Furthermore, although different vehicle types such as Euro 3 Light Duty Car and Euro 5 Heavy Duty Truck have different emission relationships with driving patterns (Carslaw et al. 2011), vehicles can be simplified as having homogenous emission factors by adopting weighted emission factors based on vehicle composition, provided this composition is stationary (static modeling assumptions).

We recall the preliminary discussion by Lin et al. (2016), showing the derivative of link emission  $E_{ap}$  with respect to control measures  $M_i$  is:

$$\frac{\partial E_{ap}}{\partial M_i} = \frac{\partial \varepsilon_{ap} f_a l_a}{\partial M_i} = \left[ \frac{\partial \varepsilon_{ap}}{\partial M_i} f_a + \frac{\partial \varepsilon_{ap}}{\partial V_a} \frac{\partial V_a}{\partial M_i} f_a + \left( \frac{\partial \varepsilon_{ap}}{\partial V_a} \frac{\partial V_a}{\partial f_a} + \varepsilon_{ap} \right) \frac{\partial f_a}{\partial M_i} \right] l_a \quad (3)$$

Here  $M$  is the set of available control measures variables.  $E_{ap}$ ,  $\varepsilon_{ap}$ ,  $V_a$  and  $f_a$  are arbitrarily assumed as differentiable functions of  $M$ .  $E_{ap}$  is the link emission of pollutant  $p$  at link  $a$ .  $V_a$  is the driving patterns of traffic, and can be replaced by average speed  $v_a$  according to Assumption 7. Further,  $\varepsilon_{ap} > 0$  because of the nature of an emission factor. Equation (3) shows that the control measures could influence the total emission in three different ways: directly influencing the emission factor, directly influencing the driving patterns, and indirectly influencing the driving patterns by re-allocating the traffic flow. One example of the first way is clean fuel. Because neither toll nor additional delay directly influence the emission factor,  $\frac{\partial \varepsilon_{ap}}{\partial t}$  and  $\frac{\partial \varepsilon_{ap}}{\partial d}$  are both 0. Moreover,  $\frac{\partial v_a}{\partial t} = 0$  because toll does not directly influence the driving patterns. Because  $M_i \in \{(t, d)\}$ , the partial derivatives of link emission to tolls and additional delays become

$$\frac{\partial E_{ap}}{\partial \mathbf{t}} = \left( \frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial f_a} + \varepsilon_{ap} \right) \frac{\partial f_a}{\partial \mathbf{t}} l_a \tag{4}$$

$$\frac{\partial E_{ap}}{\partial \mathbf{d}} = \left[ \frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial \mathbf{d}} f_a + \left( \frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial f_a} + \varepsilon_{ap} \right) \frac{\partial f_a}{\partial \mathbf{d}} \right] l_a = \frac{\partial E_{ap}}{\partial \mathbf{t}} + \frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial \mathbf{d}} f_a l_a \tag{5}$$

Since  $\frac{\partial f_a}{\partial \mathbf{t}} = \frac{\partial f_a}{\partial \mathbf{d}}$  from Corollary 1.1, besides flow reallocating effects, additional delays also influence the emission factor by changing the current driving patterns. Because  $v_a = \frac{l_a}{c_a+d_a}$ ,  $\frac{\partial v_a}{\partial d_a} = -\frac{l_a}{(c_a+d_a)^2} < 0$ . As a result, we have Proposition 3.

- **Proposition 3:** If  $\frac{\partial \varepsilon_{ap}}{\partial v_a} > 0$ , additional delay results in less emission than toll for sustaining the same link flow pattern, and vice-versa.

### 2.4 On the OSP-EC Problem

Finally, we recall the concept of selecting correct control measures for environmental constraints in traffic management, which was shortly discussed in Section 1. Mathematically, with all given assumptions, the OSP-EC problem can be described as the bi-level programming problem below:

$$\min_{\mathbf{M} \in \mathcal{M}} SP(\mathbf{M}) \tag{6}$$

Subject to :  $E(\mathbf{M}) \leq EC$

$$c(\mathbf{M}) \cdot (\mathbf{f} - \mathbf{f}(\mathbf{M})) \geq 0 \forall \mathbf{f} \in \Omega_F(\mathbf{M}) \tag{7}$$

Here  $SP(\mathbf{M})$  is the system performance function and it could be varied according to the definition of system performance.  $\mathcal{M}$  is the set of all available control measures.  $c(\mathbf{M})$ ,  $\mathbf{f}(\mathbf{M})$  and  $E(\mathbf{M})$  are respectively the link user cost vector, link flow vector and emissions under implemented control measure  $\mathbf{M}$ .  $\Omega_F(\mathbf{M})$  implies that the control measures may influence the feasible set of traffic flows, for example, when the demand is elastic. It is a bi-level programming problem where the lower level of variable inequality problem (7) reflects Wardrop’s first principle (Wardrop 1952).

If only tolls and additional delays are available as control measures, the environmental constraints are link based, the traffic demands are inelastic, and if the system performance function of interest is the total system time cost, the bi-level programming becomes:

$$\min_{(\mathbf{t}, \mathbf{d}) \in (\mathbb{R}^A, \mathbb{R}^{A^+})} TC(\mathbf{t}, \mathbf{d}) = [c(\mathbf{f}(\mathbf{t}, \mathbf{d})) + \mathbf{d}] \cdot \mathbf{f}(\mathbf{t}, \mathbf{d}) \tag{8}$$

Subject to :  $E_{ap}(\mathbf{t}, \mathbf{d}) \leq EC_{ap} \quad \forall a \in A_{EC}$  (9)

$$[c(f(t, d)) + t + d] \cdot (f - f(t, d)) \geq 0 \forall f \in \Omega_F \tag{10}$$

Here  $A$  is the number of links, and  $\mathcal{A}_{EC}$  is the set of links on which an environmental constraint is imposed. According to the previous analysis, toll and additional delay have the same effects on re-allocating traffic flow, but different effects on total system cost and emission. Tolls always works better than additional delays to the purpose of reduce system time cost. Although the two control measures influence the traffic emission in different ways, we do not know which control measures cause less traffic emission unless the sign of  $\partial \varepsilon_{ap} / \partial v_a$  is predicted. Therefore, the two control measures have the same effects on (10), but different effects on (8) and (9). They cannot be considered as equivalent control measures.

### 2.5 Case study

This sub-section illustrates the theoretical results obtained in the previous sub-sections. We select one of the simplest, most intuitive case studies: a 2-route network. We show that – even though the analytical results in the previous sub-sections may seem relatively straightforward – their implications even on this simple network may not always be that trivial, which is why we feel the topic deserves further analysis.

As shown in Fig. 1, Link 1 is a shorter route across the city and Link 2 is a longer highway around the city. The cost functions of Link 1 and Link 2 are respectively  $c_1 = 12 + 0.01f_1$ (minutes) and  $c_2 = 18 + 0.002f_2$ (minutes). The cost functions imply that the free flow speeds on both links are respectively 50 km/h and 100 km/h, but the urban route is more sensitive to congestion. The fixed demand from origin to destination is 3000 vehicle/h. The emission factor  $\varepsilon_{ap}$  (gram /veh/km) of the target pollutant is shown below. Similar as for NOx, the emission factor decreases with average speed when the average speed is below 60 km/h and increases when the average speed is above 60 km/h.

$$\varepsilon_{ap} = \begin{cases} 1.05 - 0.015v_a & v_a < 60 \text{ km/hour} \\ 0.001(v_a - 60) + 0.15 & v_a \geq 60 \text{ km/hour} \end{cases} \tag{11}$$

Without any control measures, the user equilibrium link flows are  $f = (1000, 2000)$ . The authorities want to sustain a target traffic flow<sup>1</sup> and consider two types of controls: tolls-only and additional-delays-only. Figure 2 shows how control, total system cost, emission factors and total emission respectively evolve as a function of the target flow to be observed on link 1 under the two different control measures. The points indicated by letters are used to illustrate the OSP-EC problem.

Figure 2a presents the control parameters for sustaining the targeted traffic flows. It is worth mentioning that the control sets are not unique (flows respond to the difference in control values, not to absolute values on each link), and we select two simple sets:

<sup>1</sup> Different from the environmental constraints, the target traffic flow is not an upper limit but rather a flow that needs to be met. We use this target flow to better illustrate the parallels and differences of the two measures learned from Section 2. Following discussion about the figure will also show why the two measures can produce different solutions for the same environmental constraint (the upper limit).

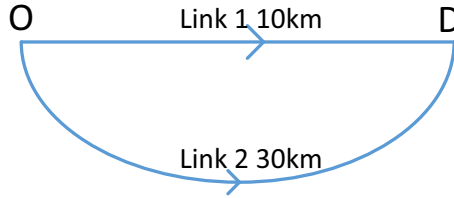
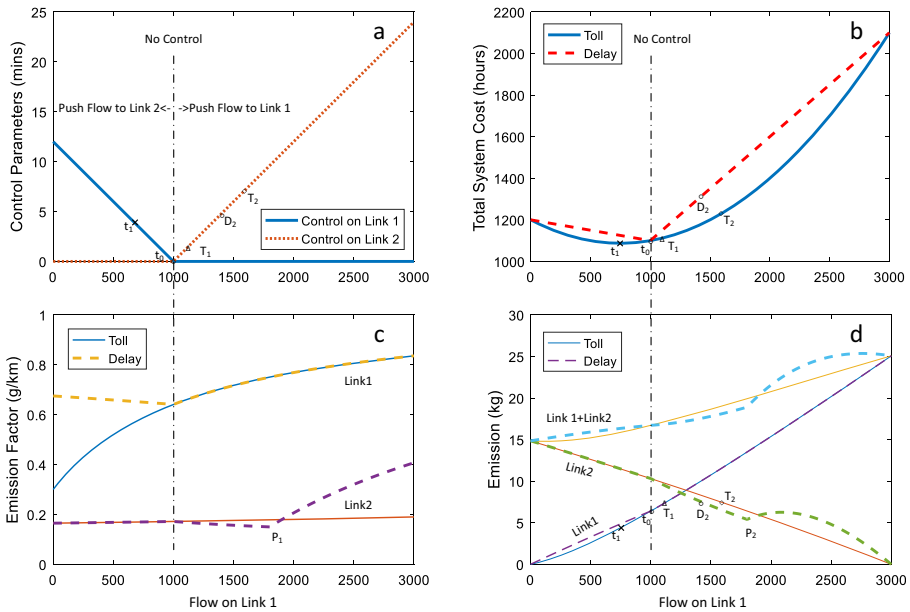


Fig. 1 Two links network

when the target requires pushing flow to link 2, a control on link 1 will be imposed; and vice-versa. The figure shows that tolls and additional delays have the same influence on sustaining the traffic flow patterns, which is coincident with Proposition 1. Figure 2b represents Proposition 2: although the same value of tolls and additional delay could reallocate the same traffic flow, doing this through additional delays always causes higher (or at best: equal) system cost than through tolls. Figure 2c indicates that also the emission factors differ: the upper two lines are the emission factors on link 1 and the lower two lines are the emission factors on link 2, respectively with tolls and additional delays. Because the maximal speed on link 1 is 50 km/h, the emission factor is always monotonically decreasing with the speed  $\frac{\partial \varepsilon_{ap}}{\partial v_a} < 0$ . When imposing a toll on link 1 (upper left), the emission factor will always decrease because  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial f_a} \frac{\partial f_a}{\partial t_a} < 0$ . By imposing the delay, the emission factor increases. The latter happens when  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \left( \frac{\partial v_a}{\partial d_a} + \frac{\partial v_a}{\partial f_a} \frac{\partial f_a}{\partial d_a} \right) > 0$ , or  $\frac{\partial v_a}{\partial d_a} < -\frac{\partial v_a}{\partial f_a} \frac{\partial f_a}{\partial d_a}$ . When imposing a toll on link 2 (lower right),  $v_a \geq 60$  km/h, so the emission factor is increasing with the speed  $\frac{\partial \varepsilon_{ap}}{\partial v_a} > 0$ . But additional delays will reduce the emission factor on link 2 until the speed drops below 60 km/h at point  $P_1$ . This happens because speed reduction due to the delay is stronger than the speed gain due to congestion reduction, so net speed decreases; above 60 km/h this leads to emission factor reduction, whereas below 60 km/h the emission factor increases strongly with reduced speed. Looking at total emissions in Fig. 2d, the delays on link 2 lead to the reduction of emission on link 2 (lower right), but delays on link 1 increase the emission on link 1 (lower left). The emission curves for both links combined illustrate Proposition 3: total emission under the same link flow pattern can be lower when these flows are triggered by delay as compared to toll; this happens indeed when imposing a delay on link 2 in a regime where the emission factor increases with speed, until reaching point  $P_2$ .

Now, consider some OSP-EC problems: In a first case, suppose that the environmental constraint requires that the emission on Link 1 may not exceed 7.5 kg/h. The authorities seek the optimal controls for satisfying the constraints, and suppose only toll is available. From Fig. 2d, to satisfy the constraint, the flow on link 1 should be smaller than  $f_1^{T1}$ . The toll for the equivalent flow can be found on point  $T_1$  in Figure 2a, and the total cost is indicated at Point  $T_1$  in Figure 2b. However, Figure 2b also shows that point  $t_1$  can further reduce the total system cost by satisfying the constraint on Link 1. Point  $t_1$  is the optimal solution of this OSP-EC problem. Note that at  $t_1$ , the environmental constraint is not activated.

For a second case, suppose the same constraint is imposed on Link 2 instead of Link 1, and both control measures are available. In Figure 2d, point  $T_2$  and point  $D_2$  are



**Fig. 2** Control Parameters (a), Total System Cost (b), Emission factors (c) and Emissions (d) as a function of the target flow on link 1

respectively the maximal flows on link 2 (look the x-axis from right to left) with tolls and delays due to different emission factors. It shows that the toll measure needs to push more flow from Link 2 to Link 1. The corresponding control measures  $T_2$  and  $D_2$  can be found on Figure 2a, and likewise the minimal total system cost of the two different measures can be read from Figure 2b. The total system cost with toll is smaller than the total system cost with delays. The toll set at  $T_2$  is the best control set for optimizing the system performance with the given environmental constraint. Note that in this case, the environmental constraint is actively constraining the optimal solution because the unconstrained optimum  $t_1$  is not within the feasible range of emission values on link 2.

The two examples represent the simple process for solving OSP-EC problem under tolls and additional delays shown in eq.(8) – eq.(10). However, the process is more complicated in a general network with multiple link constraints, and there remain some unsolved issues:

- it is difficult to seek points T and D in general networks because of the lack of efficient algorithms for the bi-level programming.
- it is unclear whether the maximal flow constraints will be activated in the OSP-EC problem solution.
- delays may cause less emission but higher system cost, so a method is needed to select proper controls in a general situation.
- the examples above do not have an optimum using the combination of the two control measures, but in general cases can exist where a combination works better than a single control measure.

To answer the questions above, we explore further the role of tolls and additional delays strategies in the OSP-EC problem. We recall from Proposition 3 that the monotonicity of emission factor function is crucial for additional delays to have an advantage in reducing emissions. Therefore, the remainder of the discussion discriminates between two cases. Section 3 discusses the strategies when  $\partial\varepsilon_{ap}/\partial v_a \leq 0$  (emission factor monotonically decreasing with speed, i.e. strictly decreasing or constant). In Section 4, we discuss the case where the emission factor is not monotonic in the speed. In each section, three types of control measures are discussed separately: only tolls are available (the tolls-only strategy), only additional delays are available (the delays-only strategy), and both tolls and additional delays are available (the combination strategy).

### 3 Strategies with Monotonically Decreasing Emission Factor

With average speed increasing, combustion engines become more fuel-efficient, and drive patterns contain less stop-and-go activities, so the emission factors (expressed in unit gram per kilometer) could become smaller. In traffic management, the emission factor is therefore generally considered as constant or monotonically decreasing with the speed over a related range below the optimal speed  $v_p^{\text{OPT}}$  (see discussion in Section 4).

In this section we consider the quite common case where  $\partial\varepsilon_{ap}/\partial v_a \leq 0$ . In the respective sub-sections we discuss and prove that in this case:

- under tolls-only strategy, one of the toll sets achieving system optimum time cost subject to emission constraints is the combination of marginal system time cost tolls on all the links, supplemented by shadow prices on the active constrained links.
- the same delays-only strategy might not work, for the delays optimizing the OSP-EC problem may be on other links than on the constrained links.
- if both control measures are available, tolls always work better than additional delays; hence an optimal control is again the combination of marginal system time cost tolls on all links plus shadow prices on active constrained links.

#### 3.1 Tolls-Only Strategy

In this sub-section, we show that when the emission factor is monotonically decreasing with speed in a tolls-only strategy, a set of flow constraints can be used to replace the original set of environmental constraints. Moreover, we show that these constraints can be met by imposing a marginal system cost externality toll on all links, supplemented by a shadow toll on the actively constrained links. This analytical result simplifies the problem of finding OSP-EC tolls by reducing the dimension of the search space from unknown tolls on all links to unknown shadow price supplements only on the actively constrained subset of links.

Equation (4) sets the relationship between total emission and tolls. Because  $\frac{\partial v_a}{\partial f_a} = \frac{\partial c_a}{\partial c_a}$   $\frac{\partial c_a}{\partial f_a} \leq 0$  according to Assumption 2,  $\varepsilon_{ap} > 0$  and  $l_a > 0$ , it follows that  $\left(\frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial f_a} + \varepsilon_{ap}\right) l_a > 0$ . Therefore, in the tolls-only strategy, link emission  $E_{ap}$  is strictly

monotonically increasing with link flow  $f_a$  and  $E_{ap}$  is a bi-injective function with link flow. We can conclude that:

- **Proposition 4:** If the emission factor is monotonically decreasing with the speed and only tolls are available as the control measure, then a link-based primary pollutant constraint is equivalent to a flow constraint and the emission is monotonically increasing with the flow.

If  $E_{ap}(t^*, \mathbf{0}) \leq EC_{ap}$ , we can always find an equivalent flow constraint  $f_{ap}^{ECeq}$  which ensures  $f_a(t^*, \mathbf{0}) \leq f_{ap}^{ECeq}$  is equivalent to  $E_{ap}(t^*, \mathbf{0}) \leq EC_{ap}$ . A set of flow constraints  $f^{ECeq} = \left( \min_{p \in P} f_{ap}^{ECeq}, a \in \mathcal{A}_{EC} \right)^T$  can be used to replace the original set of environmental constraints (9).

$$f_a(t, \mathbf{0}) \leq f_a^{ECeq} \quad \forall a \in \mathcal{A}_{EC} \tag{12}$$

According to Corollary 1.2 and Corollary 1.4, the toll on each constrained link always reduces the flow on that link unless there is no flow on that link or the flows do not have (active) alternative routes. An obvious strategy for satisfying the equivalent flow constraints would be to impose tolls on the constrained links only. Whereas this strategy can guarantee the satisfaction of constraints, it cannot in general ensure the minimum of system time cost; charging tolls on unconstrained links may be required (note that this is essentially the difference between the UE-SC and OSP-EC problems).

Proposition 5 proposes one possible tolls-only solution for the OSP-EC problem with monotonically decreasing emission factors. The strategy consists of charging, besides link marginal external congestion cost toll equal to  $c_a^{mecc} \triangleq \frac{\partial c_a}{\partial f_a} f_a$  on all the links, additional positive shadow prices  $s_a \geq 0$  for satisfying the environmental constraints only on the links whose environmental constraints are activated.

- **Proposition 5:** one of the OSP-EC toll sets for satisfying the link environmental constraints is:

$$t_a = \begin{cases} c_a^{mecc} + s_a & f_a = f_a^{EC} \forall a \in \mathcal{A}_{EC} \\ c_a^{mecc} & f_a < f_a^{EC} \forall a \in \mathcal{A}_{EC} \\ c_a^{mecc} & \forall a \notin \mathcal{A}_{EC} \end{cases} \tag{13}$$

**Proof:** See Appendix 2

It is worth noting that the toll set from Proposition 5 is one of the proper toll sets but not the only toll set optimizing the OSP-EC problem, because of the non-uniqueness property of toll sets for system optimum problems (Yang and Bell 1997). However, this particular toll set is intuitive and makes practical sense. On the one hand, it is intuitive to charge toll on the constrained links whenever flows exceed the constraint. On the other hand, controllers need to internalize the marginal system cost on all links to minimize the system time cost besides the shadow prices for constraints. Note that the



analytical form of these marginal external congestion cost tolls is identical to the well-known formula for unconstrained system optimum toll:  $\frac{\partial c_a}{\partial f_a} f_a$ , even though the numerical values of the toll will differ as the flow value in the formula will be different.

Proposition 5 can be recognized as a combination of two types of toll pricing: marginal system cost tolls for system optimum and the shadow price for the side constraints. It reduces the dimension of the search space for finding optimal controls from  $A$  (number of independent link tolls in the network) to  $A_{EC}$  (number of independent shadow prices on constrained links in the network), which is usually a substantially lower number. It would also help to select proper algorithms from references, and to develop new algorithms for given problems. Existing algorithms for solving the UE with side-constraints (Larsson and Patriksson 1995; Larsson et al. 2004) can be used to solve the toll-only OSP-EC problem, by replacing the link traffic cost  $c_a(f_a)$  by link marginal system cost  $\hat{c}_a(f_a) = c_a^{mccc} + c_a(f_a)$ .

### 3.2 Additional-Delay-Only Strategy

In this section, we argue based on hypotheses and a counter example that a “shadow-delay” strategy on the constrained links similar to the shadow-price strategy of Proposition 5, is in general not optimal and might at best yield approximate solutions to the OSP-EC problem.

Fully in parallel to Proposition 4 for tolls, Proposition 6 is obvious because  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial d}$  and  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial f_a}$  are both positive in Equation (5), and so is Corollary 6.1 by comparing Equations (4) and (5), and Corollary 6.2 because  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial d} = 0$  when the delay is not implemented on the constrained link.

- **Proposition 6:** If the emission factor is monotonically decreasing with the speed and only additional delays are available as a control measure, then a link-based primary pollutant constraint is equivalent to a flow constraint and the emission is monotonically increasing with the flow.
- Corollary 6.1: If for a given set of emission constraints, the equivalent flow constraints are  $f_{a,delay}^{ECeq}$  and  $f_{a,toll}^{ECeq}$  for delay-only and tolls-only strategy, respectively, then it must hold that:  $f_{a,delay}^{EC} \leq f_{a,toll}^{EC}$ .
- Corollary 6.2: If for a given set of emission constraints, the equivalent flow constraints are  $f_{a,DOC}^{ECeq}$  and  $f_{a,DOUC}^{ECeq}$  for delay-on-the-constrained-link strategy and delay-on-the-unconstrained-links strategy, respectively, then it must hold that:  $f_{a,DOC}^{ECeq} \leq f_{a,DOUC}^{ECeq}$ .

Because according to Corollary 2.1, additional delays cause an additional cost  $\frac{\partial TC}{\partial d} - \frac{\partial TC}{\partial t} = f^T$  compared to tolls and hence the proof for Proposition 5 cannot work here, also Hypothesis 1 below cannot be proven. Because additional delays will increase the traffic cost of traffic flow still on the constrained link, we suggest Hypothesis 2 (which we will also prove to be wrong later on): control the route flows using the constrained link, but instead of doing this by delay on the constrained link, do it elsewhere along the

route on a link that has the least influenced traffic flow. For instance, the upstream and downstream links, which also carry route flows using the constrained link, are potential locations. Here, Braess links denote links on which a proper delay could reduce the total system cost (Braess 1968). Literature indicates Braess links exist in urban networks (Youn et al. 2008) but they are not very common (Nagurney 2010).

- **Hypothesis 1 (proven wrong):** Shadow delays for the OSP-EC problem should always be implemented on the constrained links (except for Braess links).
- **Hypothesis 2 (proven wrong):** Shadow delays for the OSP-EC problem should always be implemented on links which carry the route flows using the constrained links (except for Braess links).

Even though they sound plausible, the counter example in Appendix 3 proofs that both Hypothesis 1 and Hypothesis 2 are incorrect. The intuition behind Hypothesis 1 being wrong is double. Firstly, additional delay on a link causes additional system time cost. If the constraint on a link with high flow is not too strong, imposing delay there would inflict additional cost on a rather high flow remaining on the link. Thus we should preferably impose delay on a link with a low flow. Then it may be better to target a smaller flow elsewhere in the network if this also makes flow in the constrained link drop. Secondly, because of the decreasing nature of the emission factor, imposing delay on a link increases the emission factor. So if we target the constrained link, then because of this ‘negative bonus’ we should suppress traffic flow even more to remain below the constraint; this causes additional system time cost and hence is not optimal. If a delay elsewhere is also effective in rerouting traffic away from the constrained link, this may be preferable. The ‘negative bonus’ is illustrate in Appendix 4.

We conclude that the OSP-EC problem controlled by additional delays has more complex solutions than control by tolls only. A global searching algorithm is necessary, in principle over the full  $A$ -dimensional solution space. It can be expected however that the locations for optimal control by additional delays are dependent on the level of constraints: the stricter the environmental constraint is, the more likely it is that optimal delay-control needs to be imposed only on the constrained link because  $\frac{\partial \epsilon_{av}}{\partial v_a} \frac{\partial v_a}{\partial d} f_a$  becomes very small in Equation (5). If local optima are acceptable (e.g. in a more practical, heuristic setting), Hypotheses 1 and 2 could be considered for reducing the dimension of control variable space.

### 3.3 Combined Tolls-Delays Strategy

If both control measures can be applied in the traffic network for the OSP-EC problem, instead of locations and values, the first question for the authorities is the selection of the most appropriate control measures. Proposition 7 states that if the emission factor is monotonically decreasing with average speed, tolls always work better than additional delays.

- Proposition 7: If both tolls and additional delays are available and the emission factor is monotonically decreasing with respect to average speed, then only tolls should be considered for the OSP-EC problem.

**Proof:** See Appendix 5.

Even though the authorities could use both control measures, they only need to consider tolls and ignore the additional delays measure. The authorities therefore do not need to consider the complex influences of additional delays, and only need to employ the strategy based on Proposition 5 for solving the OSP-EC problems with tolls as discussed in Section 3.1.

## 4 Strategies with Non-Monotonic Emission Factor

In Section 3.3, the emission factor was assumed to be monotonically decreasing. In practice however, if the average speed exceeds a certain optimal speed  $v_p^{\text{eOPT}}$  (usually around 60–70 km/h), the emission factor of combustion engines could increase with respect to the average speed because of higher energy consumption and less post-treatment efficiency (Zhang et al. 2014). Such kind of increasing relationship is also considered in some average-speed-based emission models (Smit et al. 2010; Gkatzoflias et al. 2012). In this case, the condition  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \leq 0$  does not hold, and  $\frac{\partial \varepsilon_{ap}}{\partial v_a} > 0$  if the average speed  $v_a$  is larger than the environmental optimal speed  $v_p^{\text{eOPT}}$ .

With increasing emission factor, Proposition 4 cannot hold and additional delays may reduce the emission factors. But, in this section, we show that in most real cases, the shadow price strategy still works. Delays should only be considered as a complementary measure on the links with low flow and the speed above  $v_p^{\text{eopt}}$  if both control measures are available. We do this again in subsequent sub-sections for the case of tolls-only, additional-delay-only, and the combined strategy.

### 4.1 Tolls-Only Strategy

The conclusion of Proposition 4 that a link emission constraint is equivalent to a link flow constraint was based on the condition  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \leq 0$ . Now however, because (over a certain speed range)  $\frac{\partial \varepsilon_{ap}}{\partial v_a} > 0$ , the total emission may not necessarily increase with the traffic flow and thus the equivalent environmental link flow is theoretically not unique. To examine whether we could nevertheless work with an equivalent link flow constraints, we investigate the relationship between total emission and traffic flow on a link in an alternative way. We will now show that under reasonable assumptions on travel time functions and emission factor functions, total emissions are still increasing with traffic flow on a link, and hence our earlier conclusions on equivalence between emission constraint and link flow constraint still hold.

The equation below, based on Assumption 7, calculates the total emission:

$$E_{ap}(\mathbf{t}, 0) = \varepsilon_{ap}(v_a(\mathbf{t}, 0))f_a(\mathbf{t}, 0)l_a \quad (14)$$

where  $v_a(t, 0) = \frac{l_a}{c_a(t, 0)} = \frac{l_a}{c_a(f_a(t, 0))}$ . We will now rewrite the total emission as a function of average speed, enabling us to investigate its derivative with speed. Since  $\frac{\partial E_{ap}}{\partial f_a} = \frac{\partial E_{ap}}{\partial v_a} \cdot \frac{\partial v_a}{\partial f_a}$  with  $\frac{\partial v_a}{\partial f_a} < 0$ , we find that indeed total emissions increase with flow ( $\frac{\partial E_{ap}}{\partial f_a} > 0$ ) whenever  $\frac{\partial E_{ap}}{\partial v_a} < 0$ . We now show that this is the case under reasonable assumptions.

Suppose the link cost function is specified as a BPR function (Sheffi 1984):  $c_a(t, 0) = \frac{l_a}{v_a^{free}} \left( 1 + \alpha \left( \frac{f_a}{f_a^{cap}} \right)^\beta \right)$ . With  $v_a(t, 0) = \frac{l_a}{c_a(t, 0)}$  and  $f_a^{cap}$  is the flow capacity, we can eliminate the flow dependency from the total emissions (14):

$$E_{ap}(t, 0) = \varepsilon_{ap}(v_a) \left( \frac{v_a^{free} - v_a}{\alpha v_a} \right)^{1/\beta} f_a^{cap} l_a \tag{15}$$

When we further specify the emission factor function and the parameters, then we can finally analyze the derivative of (15). Let  $(\alpha; \beta) = (0.15; 4)$ , the relationship between normalized link emission rate  $\frac{E_{ap}}{l_a f_a^{cap}}$  and average speed  $v_a$  under different free flow speeds can be obtained in Figure 3, with  $E_{ap}$  from Equation (15) and  $\varepsilon_{ap}(v_a)$  from COPERT4 (Gkatzoflias et al. 2012).

We find for NOx that despite the emission factor increasing with the average speed when the average speed become larger than 70 km/h, still total link emissions decrease monotonically with the average speed because the link flow also decreases, which – as shown above – implies that total emissions indeed increase with flow.

We have verified for other pollutants (CO and PM) and realistic emission functions (except for NOx with a quartic link cost function), and a reasonable range of  $\alpha$  and  $\beta$  (unless parameters  $\beta$  is very large such as  $\beta \geq 8$ ), the link flow and total link emission all have the same strictly increasing relationship. We also notice that the CO emission

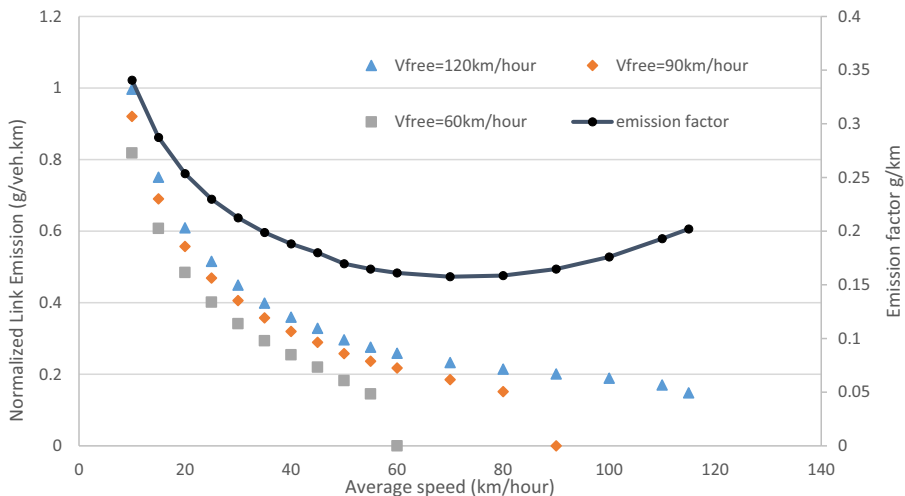


Fig. 3 Relationship between link NOx emission and average speed on the link

model from TRANSYS-7F (Wallace et al. 1984), which is widely used in traffic environmental-related research (Yin and Lawphongpanich 2006; Nagurney et al. 2010; Chen et al. 2011; Chen and Yang 2012), also leads to a strictly monotonically increasing relationship with the traffic flow, although the model CO emission factor is non-monotonic in speed. Moreover, according to (European Environment Agency 2016) the emission factors of new vehicles (post Euro 5) are monotonically decreasing over the entire speed range between 0 km/h and 130 km/h. This is because the engine and post-treatment are more efficient in high speed ranges in new vehicles.

As a result we conclude that, although theoretically and for some components (Jakkula and Asakura 2009) the link emission curve can increase when the average speed increases, for major pollutants, the link emission with actual emission factor is strictly increasing with the link flow.

An important consequence is that for tolls-only scenarios we can in practice, for most pollutants and regardless of the sign of the emission factor derivative, always consider equivalent link flow constraints for a given set of link emission constraints. As a consequence, the strategy of tolls developed for monotonically decreasing emission factors can be applied as well.

Moreover, for those few cases where the relationship between link emission and link flow would nevertheless turn out to be non-monotonic, the unique equivalent link flow constraint cannot work. The link flow which satisfies the emission constraint could then be located in two ranges: one range is  $0 \leq f_{ap}^* \leq f_{ap}^{ECeq}$  and the average speed is close to the free flow speed; and the other range is where the average speed of the flow should be around the optimal speed  $v_p^{OPT}$ . A branch-and-bound algorithm combined with the shadow price strategy could help to solve such either-or constraints problem.

## 4.2 Delays-Only Strategy

In the range of speeds with an increasing emission factor, additional delays on the constrained link will reduce the emission factor. Contrary to Corollary 6.2 in Section 2,  $f_{a,DOC}^{ECeq} \geq f_{a,DOUC}^{ECeq}$  if unique equivalent link flow constraints exist. When, compared to unconstrained optimal delay traffic assignment, a controller uses additional delays on a constrained link to reduce emissions there, the measure may now have a positive bonus (in contrast to the negative bonus discussed in Section 3.2, and see Appendix 4 for illustration): (i) because of the delay, traffic reroutes, hence flow decreases and so also total emission, and (ii) in addition the remaining flow having lower speed emits less per veh/km. Thanks to this bonus, it may suffice to use a milder delay (compared to the case with decreasing emission factor) to push emissions below the constraint.

Moreover, if a single suitable additional delay on the constrained link can satisfy the environmental constraint and also reduces the average speed from the above- $v_p^{eOPT}$  range to below- $v_p^{eOPT}$  range, a better control measure can exist: a milder delay on the constrained link to let its average speed be  $v_p^{eOPT}$  and another milder delay on some unconstrained links which can potentially further reduce the flow on the constrained link (see Appendix 4 for illustration). This combination can inflict less additional delay because the equivalent flow is less strict, and thus can possibly damage the system less.

### 4.3 Combined Tolls-Delays Strategy

In this sub-section, we show that if both control strategies are available and emission factors do not always decrease with speed, the tolls-only strategy of Proposition 5 (and 7) can in general not guarantee optimality: in theory there may exist cases where, in addition to the marginal time cost and shadow price tolls, adding delay to a constrained link may be beneficial. However, we argue that such cases are rather uncommon, so that in practice mostly the toll strategy of Proposition 5 suffices. We point however at an exception where a constraint imposed by toll only would yield an infeasible flow solution whereas the positive bonus (discussed in the previous section) allows the constrained link to carry more flow.

If both control measures are available, and the emission factors are non-monotonic, the additional delays could reduce the emission factor. As a consequence, Proposition 7 does not hold anymore, because we may no longer count on the fact used in its proof that  $E_{ap}(\mathbf{t}^* + \mathbf{d}^*, 0) \leq E_{ap}(\mathbf{t}^*, \mathbf{d}^*)$ , i.e. that emission control by imposing delay can at most be as effective as an equivalent toll but is usually less effective. In the range of increasing emission factors, additional delays now reduce (instead of increase) the system cost compared to tolls. We give Proposition 8 without proofs as the proofs are similar to that of Proposition 7 in Appendix 5.

- **Proposition 8:** to minimize the system cost with the environmental constraint, the following sets of tolls and additional delays are needed:
- for unconstrained links and links with inactive constraints: only a toll at marginal external congestion cost on each link;
- for links with average speed below environmental optimal speed  $V_{opt}$ : only a toll which combines the marginal congestion externality price with the shadow price of the constraint (see Proposition 5).
- for links with average speed above environmental optimal speed  $v_p^{eOPT}$  : a combination of toll and delay is possibly able to improve the system.

In this part, we further discuss when the delay can be used as a complementary measure to toll for the links with  $v_a > v_p^{eOPT}$ . Assume the network with multiple link constraints is optimized by tolls-only strategy  $(\mathbf{t}^*, 0)$ , the total cost difference when adding additional delay  $\Delta d_a$  and extracting link toll  $\Delta t_a$  on the constrained link  $a$  is:

$$\begin{aligned}
 TC(\mathbf{t}^* - \Delta t_a \delta_a, \Delta d_a \delta_a) - TC(\mathbf{t}^*, \mathbf{0}) &= [c_a(f_a^1) f_a^1 - c_a(f_a^2) f_a^2] + \Delta d_a f_a^1 \\
 &+ [\Phi_a(f_a^1) - \Phi_a(f_a^2)] \tag{16}
 \end{aligned}$$

Here  $\delta_a$  is the vector of  $\{0,1\}$  indicator to identify the constrained link.  $f_a^1 = f_a(\mathbf{t}^* - \Delta t_a \delta_a, \Delta d_a \delta_a)$  and  $f_a^2 = f_a(\mathbf{t}^*, \mathbf{0})$ . The difference includes three parts: the change of total travel cost on the constrained link, additional delay cost on the constrained link (note that in this discussion, we treat additional delay and travel cost  $c_a(f_a)$  separately, even though they both contribute to the total time costs for the users), and the change of  $\Phi_a$ , the optimal value of total cost of other links excluding link  $a$ , under given flow on link  $a$ .

Proposition 7 implies that the reason of adding additional delay on link  $a$  is to allow more flows to use the constrained link by reducing the emission factors, so  $\Delta t_a > \Delta d_a$  and  $f_a^2 < f_a^1$ . It increases the total travel cost of that link (1st term). Besides, the additional delay adds a delay cost that is not experienced when using only tolls (2nd term). The first two parts always increase. Because it is an OSP-EC problem, other links are always optimized.  $\Phi$  is monotonically increasing with the flows that other links should carry  $\sum_{b \in A} f_b - f_a$ , or in other words,  $\Phi$  is monotonically decreasing with  $f_a$ .

Because  $f_a^1 > f_a^2$ ,  $\Phi(f_a^1) \leq \Phi(f_a^2)$ . Hence the third part in expression (16) is always negative. According to the monotonicity of the three terms, the benefit of considering additional delays on the links can only come from the congestion relaxation effect of other links (3rd term), and it is only beneficiary to implement additional delays if that benefit is larger than the increasing total time cost from the first two terms.

Based on Proposition 8 and the discussion above, we can know the characteristics of links where it makes sense to implement additional delays together with tolls:

- First, the average speed on that link should be high ( $> v_p^{\text{OPT}}$ ) so that we are in the regime of emission factors increasing with speed (e.g. not too congested motorways);
- Second, the environmental constraint on that link is very strict, which means that in the 2nd term of eq. (16), delay is multiplied by only a small flow remaining on the constrained link;
- Third, other alternative links of the network are very congested (so that the 3rd term of eq. (16), which represents the benefit of reducing flow there, is large).

These characteristics mean that additional delays are only useful for motorway links or high-speed arterial roads with a low traffic flow and strict environmental constraint.

However, local level environmental constraints are usually set to protect residents, schools or other vulnerable activities alongside the road and hence typically target urban roads. It is then weird to impose such strict environmental constraint on a motorway or arterial and not on local congested links. In other words, the case where one would benefit from adding delay to the network is rather uncommon, which means that in practice a tolls-only strategy is all you need for urban network OSP-EC problems. Nevertheless, if such uncommon scenario would be required, a decentralized greedy searching algorithm, which iteratively optimizes the toll and additional delays on those highway links, can be employed (Rinaldi and Tampère 2015).

Besides the discussion above, another advantage of the additional delays is to relax the constraints. If imposing the additional delays on the constrained links where the average speeds are above  $v_p^{\text{OPT}}$ , it follows that the equivalent link flow constraints by delay  $f_{a,\text{delay}}^{\text{ECeq}} \geq f_{a,\text{toll}}^{\text{ECeq}}$ . This shows the additional delays could relax the constraints. It helps for finding a solution when the constraints are conflicting or too strict so that tolls-only strategy only cannot satisfy all constraints in the inelastic demand cases. In Appendix 6 we give a simple illustration of this case.



## 5 Illustrations

In this section, we discuss the relevance of our theoretical results on the OSP-EC problem for traffic measures that are used in practice. Section 5.1 shows how the tolls and additional delays can be realized by existing speed reduction measures, such as speed limits. In Section 5.2, the concept of Low-emission zone is considered as a measure related to the OSP-EC problem, and the selection of possible available measures is discussed. In Section 5.3, we discuss the optimal location of control measures for an emission problem created by trucks.

### 5.1 Implementing Tolls / Additional Delays Measures in Practice

In this sub-section, we identify the empirical control measures which are similar to the theoretical first-best-tolls and the theoretical additional delays. Moreover, we also discuss the selection of parameters for an approximation of the traffic performance.

For tolls measures, Propositions 1 and 2 demonstrate the first-best-tolls guide the traffic flows without imposing additional costs to the system. The tolls-only strategy can obtain the OSP-EC flow patterns. Next the authorities could seek alternative control measures for sustaining (or in a more practical, heuristic setting: approximating) the same flow patterns. The concepts learned from tolls can be applied in the implementations of those measures below

- Tolls: The toll set from Proposition 5, Alternative toll sets, Trade permits
- Flow restrictions: Autonomous driving system with obligatory route guidance, License Plate rationing, Blocking

It is worth noting that we assumed the users are homogenous and the system cost is total time cost. If the users are heterogenous in value of time, tolls and obligatory flow restrictions can have different impact in system monetary time cost (Nie and Liu 2010; Nie 2017).

In this paper, we introduced the term “additional delays” to represent the family of control measures which increase the travelers’ time cost. The way we formulated additional delays ensures the uniqueness property of link flows under delays and allows a clear comparison to tolls. When implementing practical measures, imposing an additional delay for a link can be realized in two ways: reducing the speed or lengthen the travel distance. Although some control measures such as Chicanes can both reduce the speed and lengthen actual drive distance, most control measures mainly take effect in one way only. Furthermore, the additional delays can be imposed uniformly along the link or at specific points of the link. Figure 4 shows examples of practical control measures intending to add additional delays.

The relationship between additional delays and actual control variables can be identified according to the actual control variables’ perspectives on adding delays.

Most control measures are of the speed reduction family. In general, the control measure should increase travel time by the same amount as the desired optimal additional delay. Assuming the speed profile function under the new control measures is  $v_a(m, \kappa)$  where  $m$  is the control variable and  $\kappa$  is the travel time, the total travel time

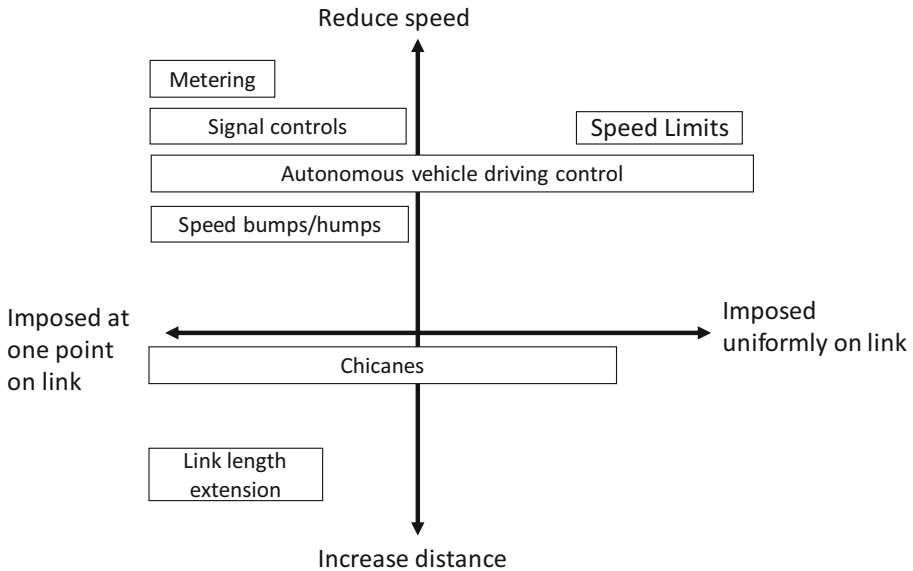


Fig. 4 practical control measures using the two ways to add delays

along link should become equal to  $c_a(f_a) + d_a$  or mathematically  $\int_0^{c_a(f_a) + d_a} v_a(m, \kappa) d\kappa = l_a$ . Take the uniform link speed limit as an example, because the speed limit does not change along the link,  $v_a^{\max}$  is independent of  $\kappa$ . The equivalent variable speed limit  $v_a^{\max} = l_a / (c_a(f_a) + d_a)$  should be considered for implementing optimal delay controls  $d_a$ , if the traffic flow assignment under the speed limits set is unique (for non-uniqueness property of speed limits, we refer to Appendix 6, Liu et al. (2017) and Yang et al. (2012)). For metering at the link exit, the control variable is the queueing delay and therefore the queue length should be controlled such that queueing delay is equal to the desired optimal additional delay.

Measures extending the travel distance work in the other way: take the link length extension measures as an example. A practical example of this could be a residential area, where authorities consider reorganizing a straight connection by a longer winding path to avoid rat-running. When we assume the extended length  $\Delta l_a$  is continuous,  $\Delta l_a = d_a / d' / (c_a(f_a) + d_a)$ . However, the link length extension also influences the total emission from Equation (3), and thus the actual  $\Delta l_a$  is larger than the equivalent extended length, or the solution can even be infeasible.

It is worth noting that we mainly focus on the macroscopic level for a long-term scenario, so the average speed is the only parameter influencing the emission factor. The control measures of different delay locations resulting in the same average speed can have different impacts on the spatial and temporal variabilities of the speed, flow and emission. For example, speed profiles for speed bumps (Barbosa et al. 2000) show that the acceleration and deceleration happen more frequently and such driving patterns will further increase the average emission factors. If studying the control measures at a microscopic level, these variabilities need to be investigated.

## 5.2 Low-Emission Zone (LEZ)

In this sub-section, we show, using the concept of OSP-EC, how to find the proper control measures for a LEZ. Based on the “Ecopass” case of Milan (Rotaris et al. 2010; Percoco 2013), we will show the importance of considering the combination of the shadow price for satisfying the constraints and the toll for optimizing the system.

Conventionally, a LEZ indicates an area with a dirty-vehicles-access restriction (DVAR), i.e. only sufficiently clean vehicles are allowed to enter (Holman et al. 2015). The main purpose of a low-emission zone is to ensure an acceptable level of local pollutants at that hot spot, in other words: an area where the local pollutants should not exceed certain environmental constraints. This corresponds to the concept of the OSP-EC problem, with however a clear difference: OSP-EC also respects the environmental constraints, but does so while optimizing the system cost. Let us consider correspondences and differences for several sets of control measures. Figure 5 schematically represents an LEZ problem. If the authorities want to restrict the emissions in the LEZ, several control measures are available; we consider here DVAR, tolls, and additional delays.

According to Propositions 1–3, 7 and 8, if non-delay control measures are available, these control measures have a lower total system cost. Generally, there are two types of DVAR measures: soft restrictions and hard restrictions. Soft restrictions indicate that the dirty vehicles should pay additional charges for entering the LEZ (e.g. London), and hard restriction implies that dirty vehicles are prohibited from entering the LEZ (e.g. Amsterdam and most German LEZ). Both types of restrictions do not impose additional delays at the LEZ, so the system has a lower total cost than the system with the same traffic flow sustained by additional delays. From Equation (3), the DVAR measure also directly reduces the emission factor, so it emits less traffic emission than the system sustained by the tolls. Conventionally, dirty vehicles have already been sorted into a certain number of categories, so the traffic flow patterns sustained by DVAR are discrete for a hard restriction. The tolls/subsidies for re-allocating the traffic flow could further reduce the system cost if the system optimum is not achieved under DVAR. For soft restrictions, besides the charges for the environmental constraints for dirty vehicles, the external congestion costs of all vehicles should also be considered for achieving the system optimum. The external system cost can either be charged to all the vehicles, or can be integrated in tolls for the dirty vehicles. The next part about the Milan “Ecopass” case further explains the selection between the dirty-vehicle-toll-only instrument and the homogenous toll to achieve the system optimum. For delays,

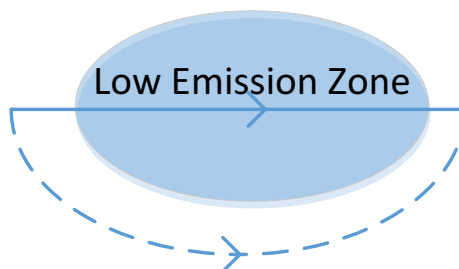


Fig. 5 Low-emission zone

because the low-emission zone is usually considered in an area where the average speed is lower than the optimal speed ( $v_p^{eOPT}$ ), additional delay measures, such as speed limits should not be considered for LEZ purposes unless other non-delay measures are not available.

In this part, we briefly analyze the Milan “Ecopass” case based on the concepts learned from the OSP-EC problem. From 2008 onwards, in order to mitigate the environmental problems and congestion in the sensitive area, Milan introduced the “Ecopass” system: the dirty vehicles must pay a toll for accessing the constrained area. At the early stage, the “Ecopass” worked efficiently for reducing both pollutants and congestion. However, because the drivers gradually replaced their dirty vehicles by clean vehicles, the toll became less efficient to address congestion. From 2012 onwards, the conventional homogenous congestion toll “area C” was implemented to replace the “Ecopass” in order to solve the congestion problems.

If we can assume that the initial targeted total emission level is the environmental constraint, the authorities want to use toll measures for achieving the target and also want to improve the system performance. It can be considered as a OSP-EC problem. In the Milan case, two sub-toll sets are available: dirty-vehicle toll and homogeneous toll. Based on Proposition 5, a toll set combining the marginal system cost and the shadow price should be considered. Compared to the homogeneous toll, the dirty-vehicle toll reduces the composition of the dirty vehicles in LEZ, so reduces the average emission factor in the LEZ. As a result, the equivalent flow constraint (referring to Proposition 4) under the dirty-vehicle toll can be larger than the equivalent flow constraint with the homogenous toll. At the introduction of the LEZ in Milan when the fraction of dirty vehicles was still high, the environmental constraint in the LEZ was active, hence in order to respect it, the OSP-EC flow was smaller than the unconstrained system-optimum flow in LEZ. The higher equivalent flow constraint of the dirty-vehicle toll measure compared to homogeneous tolls thus led to a traffic flow that was closer to the system optimum while satisfying the same environmental constraints. In contrast, at the later stage when there were less dirty vehicles, the environmental constraint was no longer active. The marginal system cost for optimizing the system, which all vehicles need to pay, had become more important. Therefore, the homogenous toll became more efficient than the initial dirty-vehicle toll.

### 5.3 Restricting Improper By-Passing

In this sub-section, we show how the difference of tolls and additional delays from Propositions 4–8 help authorities to select proper control locations for a given set of available measures.

Figure 6 is a map of an area in Belgium, where trucks from the directions of Antwerp (Belgium) and Aachen (Germany), travel towards Brussels through Aarschot. In Belgium, the trucks are already subjected to a kilometer charge that equals the average external cost per km of travelling through Belgium. Around Aarschot, although most trucks coming from Antwerp use road N223 to enter the motorway E314, some trucks still use the bypass roads N19 and N229 for entering the motorway.

Trucks cause extra environmental damage when they pass the towns of Putkapel (point A) and Wezemaal (point C), so the authorities want to reduce the emission in these towns, by making the bypasses less attractive.

If an additional toll can be charged on trucks for this purpose, based on Proposition 5, besides the first best toll, a shadow price should be imposed only in the area where the environmental constraints are violated. The location for additional charge should hence be in the towns (A and C), and the level of additional charge depends on the level of the emission constraints. Moreover, the additional toll for the satisfaction of the emission constraint also ensures the optimality of the system by reference to total cost. Mind that because of the additional shadow-price tolls at A and C, flows over routes DCB and DCA may decrease and those on the parallel alternative over the E314 motorway may increase, and as a consequence also the first-best toll on those routes should be correspondingly modified.

If for whatever reason it would be infeasible to raise additional toll and only additional delays can be used, we argue that traffic managers can better not impose these delays at their own towns (A and C). To understand this, it is important to notice that in the towns A and C, the speed limit is 50 km/h or less, which is lower than the optimal speed and hence we are in the case of non-monotonic emission factors, where emissions may increase by adding delays (Proposition 3). According to Proposition 6 (and related discussion), locations A and C may then be suboptimal and other non-constrained links along the truck route (e.g. points B and D) are candidate locations for imposing these delays. As location B would unnecessarily increase costs for traffic from C towards Brussels, point D is the better choice here, as it only affects the targeted flows coming from Aarschot or upstream. Compared to the toll-only scenario however, additional delays – even when imposed at the optimal location D – will inevitably increase the system cost.



Fig. 6 Control Location selection problem for trucks

## 6 Conclusion and Future Work

If authorities impose environmental constraints to a traffic network, traffic managers need additional controls to achieve these constraints. However, they need to be aware that they would underutilize the control measures if they would only aim at satisfaction of the environmental constraints while not simultaneously trying to optimize the overall traffic system performance. In this paper we defined such problems as optimal system performance with environmental constraints or OSP-EC problems. The solution of the OSP-EC problem depends on the available control measures. We revealed the critical role of the emission factor as a function of speed: more specifically whether emissions per vehkm increase or decrease with increasing speed. In our paper, we investigate tolls and additional delays, two conventional control measures in traffic management, and explore the selection of proper control parameters for given environmental constraints. The result is summarized in Table 1 at the end of this section.

For the tolls-only scenario, whether the emission factor is monotonically decreasing or not, marginal external cost should be charged on the links and the shadow price should be added to the links with active environmental constraints. For the additional-delay-only scenario, the same strategy does not work for monotonically decreasing emission factors, but in the monotonic increasing emission factor range, delays on the constrained links have more potential to reduce emission efficiently. For the scenario where both control measures are available, tolls measures are preferable; that is: tolls are the way to push the environmental impact below the accepted level and minimize the impact on system cost.

This result does not only contribute to the theoretical analysis and to algorithms for solving the OSP-EC problem but it also helps traffic managers to select the proper control parameters for satisfying newly imposed or revised environmental constraints. For example, in order to meet the environmental constraint in a link passing through the city, the traffic manager should increase the toll on the constrained link: in addition to the marginal external congestion cost he should charge the shadow price needed to enforce the environmental constraints. Moreover, wherever rerouting affects the flows in the network, he needs to change the tolls correspondingly to keep them equal to the marginal external costs of the new traffic volumes. However, if only additional delays can be used, it could be more beneficial to add the additional delay on other links instead of on the constrained links; these can be other links along routes leading to or from the constrained link, but in certain cases one needs to consider other links in the network that relate less obviously to the constrained link.

Alternatively, the insights of this paper may be used to understand how imposing local environmental constraints may harm system optimality. Indeed, in general, the activation of an environmental constraint will reduce the potential of a traffic system to reach optimality, as the strength of the available control measures can now no longer be exclusively used to optimize system performance but has to meet environmental goals at the same time. Authorities and traffic managers can exploit these insights in an ex-ante analysis, seeking for a way to formulate environmental constraints with minimal loss of traffic system performance.

In this research, we studied tolls and additional delays under the ideal situation that both control measures can be implemented on all links. However, in practical settings, the two control measures may not be available on all links. For tolls, such problems

**Table 1** Summarized Results

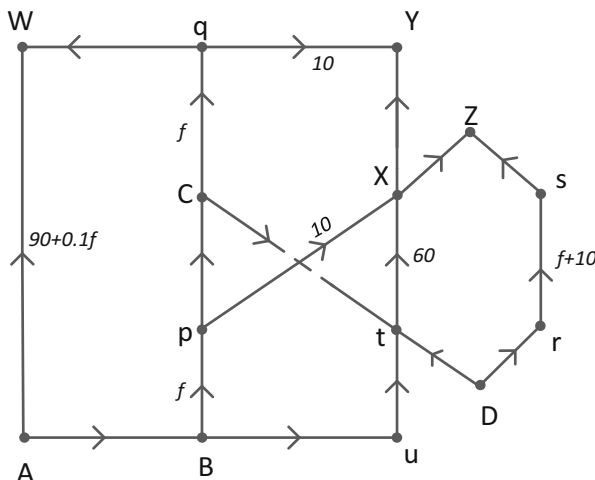
	Toll-only	Delay-only	Both measures available
Reallocating traffic flow (Section 2.1)	Same effect as delay	Same effect as toll	Same effect as toll or delay
Total system (time) cost (Section 2.2)	No additional system cost	Additional system cost due to delays	Additional system cost due to delays
Emission (Section 2.3)	Less emission when emission factor is monotonically increasing with speed	Less emission when emission factor is monotonically increasing with speed	The combination of tolls and delays is determined by the relationship between emission factor and speed.
OSP-EC problem (Monotonically Decreasing Emission Factor)	Shadow price on constrained link; marginal external congestion cost on all links (Section 3.1)	Complex. Normally the delay should be imposed on an unconstrained link up- or downstream of the constrained link, and which causes the least delay to other traffic (Section 3.2)	Only toll should be considered, and the variable is the same as the toll-only scenario. (Section 3.3)
	System optimality	Higher minimal system cost compared to toll (Section 3.2)	Same as toll-only scenario. (Section 3.2)
OSP-EC (Non-Monotonic Emission Factor)	Most practical cases: shadow price on constrained link; marginal external congestion cost on all links (Section 4.1)	Location depends on the level of constraint. (Section 4.2)	In general toll is the preferred measure, with possibly some complementary delay on constrained links with an average speed higher than environmental optimal speed (e.g. 70 km/h) (Section 4.3)
	System optimality	Higher minimal system cost but the difference becomes smaller compared to Monotonically Decreasing Emission Factor (Section 4.2)	The delay can further reduce the minimal system cost only on the links with very high speeds and low traffic flow. (Section 4.3)



**Table 2** Results of different control measures for link flow constraint  $f_{ix} \leq 30$

Control parameters		No control	Delay on link tX	Delays on links {Dt, ut, Ct}	Delay on link pC
		0	1.395	{0,1.43,1.43}	3
AW	$q_{AW}$	100	127.9	128.6	130
	$q_{ABqW}$	30	2.1	1.4	0
	$C_{AW}$	100	102.8	102.86	103
BX	$q_{ButX}$	30	0.7	0	0
	$q_{BpX}$	20	49.3	50	50
	$C_{BW}$	60	61.4	61.43	60
CY	$q_{CtXY}$	30	0.7	0	0
	$q_{CqY}$	20	49.3	50	50
	$C_{CY}$	60	61.4	61.43	60
DZ	$q_{DtXZ}$	30	28.6	30	30
	$q_{DrsZ}$	50	51.4	50	50
	$C_{DZ}$	60	61.4	60	60
$f_{iX} = q_{ButX} + q_{CtXY} + q_{DtXZ}$		90	30	30	30
Total cost	$TC(0, d)$	23,800	24,416	24,315	24,190

become a second-best toll scheme problem, and (the proof of) Proposition 4 cannot hold. The second-best toll scheme may be quite different from the first best toll scheme and may depend strongly on the available locations for controls, and the location and value of constraints. Given the lack of elegant analytical results that help in the search for optimal controls, the only remaining option may be to optimize controls in an expensive bi-level programming calculation, or to settle for heuristics (rules-of-thumb). Finding such heuristics with acceptable system costs is an interesting topic for future research.



**Fig. 7** Network rejecting hypotheses 1 and 2

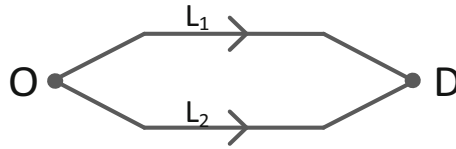


Fig. 8 the two routes network used in Appendix 6

Secondly, this research was framed in the static modeling approach; similar strategies in a dynamic traffic system and for non-separable link-cost functions may be different and need to be investigated. Besides tolls and additional delays, more control measures can be considered in a dynamic OSP-EC framework. For some control measures, algorithms for solving corresponding Network Design Problem (Farahani et al. 2013) can be applied, but for others more efficient problem-dependent algorithms are required.

Finally, although we conclude that for major pollutants, the link emission is strict-monotonically increasing with the link flow, for minor components, this relationship may be non-monotonic. In that case, the unique equivalent link-flow constraints cannot be guaranteed to work. Finally, we only consider total system time cost as system performance, but it is also possible to use, in the framework of the OSP-EC problems, other system performance criteria such as total welfare, climate, safety, control implementing cost, and other costs.

**Abbreviations** *DVAR*, Dirty vehicles access-restriction; *EC*, Environmental constraints; *LEZ*, Low-emission zone; *LO*, Loss of optimality; *NO<sub>x</sub>*, Nitrogen oxides; *OD*, Origin-destination; *OSP-EC*, Optimizing system performance with environmental constraints; *SO*, System optimum; *TAP-SC*, Traffic assignment problem with side constraints; *UE*, User equilibrium; *UE-SC*, User equilibrium with side constraints

**Symbols**  $T, c^T$  Transpose of the matrix/vector;  $\cdot$ ,  $c \cdot f$  Dot product of two vertical vectors,  $c \cdot f = c^T f$

**Notations**  $a$ , Link;  $A$ , Number of links;  $\mathcal{A}$ , Set of links;  $b$ , Link dummy index;  $c$ , Link cost;  $c$ , Link cost vector;  $C$ , Route cost;  $d$ , Link additional delay;  $d$ , Link additional delays vector;  $D$ , Demand;  $E$ , Emission (g/h);  $E$ , Emissions (g/h);  $EC$ , Environmental constraint;  $EC$ , Environmental constraints;  $f$ , Link traffic flow;  $f$ , Link traffic flow vector;  $l$ , Link length;  $m$ , Control variable of each control measure;  $M$ , Vectorized implemented control measures;  $\mathcal{M}$ , Set of available control measures (control pool);  $\mathcal{N}$ , Set of nodes;  $q$ , Route flow;  $q$ , Route flows vector;  $r$ , Route;  $\mathcal{R}$ , Set of routes between one OD-pairs;  $s$ , Link shadow price;  $SP$ , System performance function;  $t$ , Toll;  $t$ , Link tolls vector;  $TC$ , Total System Cost;  $v$ , Traffic speed;  $V$ , Driving patterns;  $w$ , Origin destination pairs;  $\mathcal{W}$ , Set of origin-destination pairs;  $z$ , Link additional cost ( $t + d$ );  $z$ , Link additional cost vector;  $\alpha$ , BPR parameter;  $\beta$ , BPR parameter;  $\delta$ ,  $\{0,1\}$  indicator identifying whether the link is included;  $\delta$ ,  $\{0,1\}$  vector identifying the target link(s);  $\varepsilon$ , Emission factor;  $\kappa$ , Time;  $\mu$ , Unit shadow price;  $\rho$ , Link traffic flow density;  $\Phi$ , Optimal system partial cost;  $\Omega$ , Feasible set of flow

**Additional Abbreviations/Notations Only Used as Sub/Super-Scripts** *cap*, Physical capacity; *DOUC*, Delay-on-the-unconstrained-links; *ECEq*, The constraint equivalent to the environmental constraint; *F*, Link flows; *mecc*, Marginal external congestion cost; *DOC*, Delay-on-the-constrained-link; *free*, Parameters under free flow; *eOPT*, Parameter according to environmental optimum; *Q*, Route flows

### Appendix 1: Proof of Proposition 1 and Corollaries 1.1–1.4

- **Proposition 1:** toll and additional delays have equivalent effects on re-allocating traffic flow, so  $f(z - \Delta d, \Delta d) = f(z, 0) = f(0, z)$

**Proof:** The traffic link flow patterns  $f(z - \Delta d, \Delta d)$  and  $f(z, 0)$  can be solved by the respective VI problems which are similar to Equation (10).

$$(c(z - \Delta d, \Delta d) + z - \Delta d + \Delta d) \cdot (f - f(z - \Delta d, \Delta d)) \geq 0 \quad \forall f \in \Omega_F \tag{17}$$

$$(c(z, 0) + z + 0) \cdot (f - f(z, 0)) \geq 0 \quad \forall f \in \Omega_F \tag{18}$$

Obviously, the two problems are equivalent and the solution  $f(z - \Delta d, \Delta d)$  from (17) is also the solution for Problem (18). Under Assumptions 2, the UE link flow patterns are unique (Smith 1979), so  $f(z - \Delta d, \Delta d) = f(z, 0)$ . If taking  $\Delta d = z$ ,  $f(z - \Delta d, \Delta d) = f(z, 0) = f(0, z)$ .

- **Corollary 1.1:**  $\frac{\partial f}{\partial t} = \frac{\partial f}{\partial d} = \frac{\partial f}{\partial z}$

**Proof:** From Proposition 1, for any change of control variable  $\Delta m_a \neq 0$  on any link  $a \in \mathcal{A}$ ,  $f(t + \delta_a \Delta m_a, d) = f(t, d + \delta_a \Delta m_a)$ . Here  $\delta_a$  is the  $\{0, 1\}$  vector identifying the link, where  $\delta_{b \in \mathcal{A}} = 1$  when  $b = a$  and  $\delta_{b \in \mathcal{A}} = 0$  when  $b \neq a$ .

Moreover,  $[f(t + \delta_a \Delta m_a, d) - f(t, d)] / \Delta z_a = [f(t, d + \delta_a \Delta m_a) - f(t, d)] / \Delta m_a$  also holds. Let  $\Delta m_a \rightarrow 0$ ,  $\frac{[f(t + \delta_a \Delta m_a, d) - f(t, d)]}{\Delta m_a} = \frac{\partial f}{\partial t_a}$  and  $[f(t, d + \delta_a \Delta m_a) - f(t, d)] / \Delta m_a = \frac{\partial f}{\partial d_a}$ . So  $\frac{\partial f}{\partial t_a} = \frac{\partial f}{\partial d_a}$ .

According to the chain rule and  $\frac{\partial f}{\partial t_a} = \frac{\partial f}{\partial d_a}$ , we find that  $\frac{\partial f}{\partial z_a} = \frac{\partial f}{\partial t_a} \frac{\partial t_a}{\partial z_a} + \frac{\partial f}{\partial d_a} \frac{\partial d_a}{\partial z_a} = \frac{\partial f}{\partial t_a} \left( \frac{\partial t_a}{\partial z_a} + \frac{\partial d_a}{\partial z_a} \right) = \frac{\partial f}{\partial t_a} \left( \frac{\partial(t_a + d_a)}{\partial z_a} \right) = \frac{\partial f}{\partial t_a}$ .

As a result, for any link  $a$ ,  $\frac{\partial f}{\partial t_a} = \frac{\partial f}{\partial d_a} = \frac{\partial f}{\partial z_a}$ , and so  $\frac{\partial f}{\partial t} = \frac{\partial f}{\partial d} = \frac{\partial f}{\partial z}$ .

- **Corollary 1.2:**  $\frac{\partial f_a}{\partial z_a} = \frac{\partial f_a}{\partial t_a} = \frac{\partial f_a}{\partial d_a} \leq 0 \forall a \in \mathcal{A}$

**Proof:** According to Corollary 1.1, we only need to prove  $\frac{\partial f_a}{\partial t_a} \leq 0$ .

Assuming  $\Delta t_a \rightarrow 0$  is the change of the toll on link  $a$ , the traffic link flow patterns  $f^0 \triangleq f(t, d)$  and  $f^1 \triangleq f(t + \delta_a \Delta t_a, d)$  can be solved by VI problems.

$$(c(f^0) + z) \cdot (f - f^0) \geq 0 \quad \forall f \in \Omega_F \tag{19}$$

$$(c(f^1) + z + \delta_a \Delta t_a) \cdot (f - f^1) \geq 0 \quad \forall f \in \Omega_F \tag{20}$$

Where  $z = t + d$ .

Because both  $f^0 \in \Omega_F$  and  $f^1 \in \Omega_F$ , so the two inequalities below should hold.

$$(c(f^0) + z) \cdot (f^1 - f^0) \geq 0 \tag{21}$$

$$(c(f^1) + z + \delta_a \Delta t_a) \cdot (f^0 - f^1) \geq 0 \tag{22}$$

From (21) (22), we can acquire

$$(c(f^1) + \delta_a \Delta t_a - c(f^0)) \cdot (f^0 - f^1) \geq 0 \tag{23}$$

Therefore

$$(c(f^1) - c(f^0)) \cdot (f^0 - f^1) + \Delta t_a \delta_a \cdot (f^0 - f^1) \geq 0 \tag{24}$$

Because of the monotonicity of link cost function (Assumptions 2),  $(c(f^1) - c(f^0)) \cdot (f^0 - f^1) \leq 0$ . In order to let (24) hold, the inequality below must hold

$$\Delta t_a \delta_a \cdot (f^0 - f^1) \geq 0 \tag{25}$$

(25) is equivalent to  $\Delta t_a (f_a^0 - f_a^1) \geq 0$  and further  $(f_a^0 - f_a^1) / \Delta t_a \geq 0$  for  $\Delta t_a \neq 0$ . As a result, if letting  $\Delta t_a \rightarrow 0$ , then  $\frac{\partial f_a}{\partial t_a} \leq 0$ .

– **Corollary 1.3:** if  $\frac{\partial f_a}{\partial z_a} = 0 \quad \forall a \in \mathcal{A}$ , then  $\frac{\partial f_b}{\partial z_a} = 0, \forall b \in \mathcal{A}, b \neq a$

**Proof:** According to Corollary 1.1, we also only prove the condition of  $\frac{\partial f_a}{\partial t_a} = 0$ .

Continue the proof of Corollary 1.2. If  $\frac{\partial f_a}{\partial t_a} = 0, f_a^0 - f_a^1 = 0$  and  $\Delta t_a \delta_a \cdot (f^0 - f^1) = 0$ . If Equation (24) holds,  $(c(f^1) - c(f^0)) \cdot (f^0 - f^1) \geq 0$ . However,  $(c(f^1) - c(f^0)) \cdot (f^0 - f^1) \leq 0$  because of Assumption 2. As a result,  $(c(f^1) - c(f^0)) \cdot (f^0 - f^1) = 0$ , and  $f^0 = f^1$ .

So  $\frac{\partial f_b}{\partial t_a} = 0, \forall b \in \mathcal{A}, b \neq a$ .

– **Corollary 1.4:** if  $\frac{\partial f_a}{\partial z_a} = 0 \quad \forall a \in \mathcal{A}$ , then either  $f_a = 0$  or  $f_a = \sum_w D_w$  for  $w$  where

$$\sum_r q_{wr} \delta_{wra} > 0$$

**Proof by Contradiction** We assume  $\frac{\partial f_a}{\partial z_a} = 0$  happens when  $f_a \neq 0$  either  $f_a \neq \sum_w D_w$  for  $w$  where  $\sum_r q_{wr} \delta_{wra} > 0$ . This must be at least one OD pair, which has at least one route (with positive route flow) using link  $a$  and also have at least one alternative route (with positive route flow) without using link  $a$ . We define these two route are  $r_1$  and  $r_2$  respectively, with route flow and costs  $q_1, q_2, C_1$  and  $C_2$ . According to Wardrop’s first principle (Wardrop 1952), the route costs  $C_1 = C_2$ . If adding a small positive cost  $\Delta z_a$  on link  $a$  satisfying  $\frac{\partial f_a}{\partial z_a} = 0$ , then according to Corollary 1.3, the flows on all links do not change. As a result,  $C_2$  remains constant because neither the link cost nor the control measures on other links change.  $C_1$  will increase by  $\Delta z_a$ . In the new assignment,  $C_1^{new} = C_1 + \Delta z_a > C_2 = C_2^{new}$ , and  $q_1^{new} = q_1 > 0$  because the flows on all links do not change. This is contradicted to Wardrop’s first principle where  $q_1^{new} = 0$  when  $C_1^{new} > C_2^{new}$ . Therefore, either  $f_a = 0$  or  $f_a = \sum_w D_w$  and Corollary 1.4 holds.

### Appendix 2: Proof of the Shadow Price Strategy for OSP-EC with Tolls

We first solve the optimization problem of total system cost with design variables route flows  $q$ .

$$\min_{q \in \Omega_Q} TC(f) = c(f) \cdot f \tag{26}$$

$$\text{Subject to : } f_a \leq f_a^{ECeq} \quad \forall a \in \mathcal{A}_{EC} \tag{27}$$

$$f_a = \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}_w} q_{wr} \delta_{wra} \quad \forall a \in \mathcal{A} \tag{28}$$

If expanding the feasible set of  $\Omega_Q$  by (1a) and replacing link flow  $f_a$  with route flow by (28), the Lagrangian function of this problem is

$$\begin{aligned} L = TC(f) &+ \underbrace{\sum_{w \in \mathcal{W}} \lambda_w \left( D_w - \sum_w \sum_r q_{wr} \right)}_H + \underbrace{\sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}_w} \mu_{wr} (-q_{wr})}_{G1} \\ &+ \underbrace{\sum_{a \in \mathcal{A}_{EC}} \mu_a^{EC} \left[ \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}_w} \delta_{wra} q_{wr} - f_a^{ECeq} \right]}_{G2} \end{aligned}$$

According to Karush–Kuhn–Tucker conditions (Boyd and Vandenberghe 2004):

$$\begin{aligned} \frac{\partial L}{\partial q_{wr}} &= \frac{\partial TC(f)}{\partial q_{wr}} + \frac{\partial H}{\partial q_{wr}} + \frac{\partial G1}{\partial q_{wr}} + \frac{\partial G2}{\partial q_{wr}} \\ &= \sum_{a \in \mathcal{A}} \delta_{wra} \hat{c}_a(f_a) - \lambda_w - \mu_{wr} + \sum_{a \in \mathcal{A}_{EC}} \delta_{wra} \mu_a^{EC} = 0 \quad \forall w \in \mathcal{W}, \forall r \in \mathcal{R}_w \end{aligned} \tag{30}$$

Reconstruct the Equations (30) by link:

$$\sum_{a \in A \setminus A_{EC}} \delta_{wra} \hat{c}_a(f_a) + \sum_{a \in A_{EC}} \delta_{wra} [\hat{c}_a(f_a) + \mu_a^{EC}] - \mu_{wr} = \lambda_w \forall w \in \mathcal{W}, \forall r \in \mathcal{R}_w \quad (31)$$

Here  $\hat{c}_a(f_a) = c_a(f_a) + \frac{\partial c_a}{\partial f_a} f_a$  is the marginal system cost. By combining other necessary KKT conditions we can obtain the route flow conditions below for optimizing the problem (26)–(28).

$$\begin{cases} \sum_{a \in A \setminus A_{EC}} \delta_{wra} \hat{c}_a(f_a) + \sum_{a \in A_{EC}} \delta_{wra} [\hat{c}_a(f_a) + \mu_a^{EC}] = \lambda_w & \text{if } q_{wr} \geq 0 \\ \sum_{a \in A \setminus A_{EC}} \delta_{wra} \hat{c}_a(f_a) + \sum_{a \in A_{EC}} \delta_{wra} [\hat{c}_a(f_a) + \mu_a^{EC}] \geq \lambda_w & \text{if } q_{wr} = 0 \\ \mu_a^{EC} (f_a^{EC} - f_a) = 0; \mu_a^{EC} \geq 0; f_a^{EC} - f_a \geq 0 & \forall a \in A_{EC} \end{cases} \quad (32)$$

Comparing to the UE with tolls in inelastic demand of the VIP problem (Smith 1979), where

$$\begin{cases} \sum_{a \in A} \delta_{wra} [c_a(f_a) + t_a] = \lambda_w^{UE} & \text{if } q_{wr} \geq 0 \\ \sum_{a \in A} \delta_{wra} [c_a(f_a) + t_a] \geq \lambda_w^{UE} & \text{if } q_{wr} = 0 \end{cases} \quad (33)$$

we can easily prove that the flow patterns which satisfy (in)equations (33) with the toll set below must let all (in)equations (32) be satisfied because  $f$  is unique.

$$t_a = \begin{cases} \frac{\partial c_a}{\partial f_a} f_a + \mu_a^{EC} + \lambda_w^{UE} - \lambda_w & \text{if } f_a = f_a^{EC} \forall a \in A_{EC}, \mu_a^{EC} \geq 0 \\ \frac{\partial c_a}{\partial f_a} f_a + \lambda_w^{UE} - \lambda_w & \text{if } f_a < f_a^{EC} \forall a \in A_{EC} \\ \frac{\partial c_a}{\partial f_a} f_a + \lambda_w^{UE} - \lambda_w & \forall a \in A \setminus A_{EC} \end{cases} \quad (34)$$

The result means that the user equilibrium with toll set (25) will sustain the same traffic flow patterns which are system optimum with the additional link equivalent flow constraints. Let  $\lambda_w^{UE} = \lambda_w$  and by known  $s_a \triangleq \mu_a^{EC} \geq 0$ , we conclude that one of the OSP-EC toll sets is Equation (3.13). The proof process is similar to the proof process of first-best toll pricing strategy but with an extra term  $G2$  related to link flow constraints.

### Appendix 3: Counter Example for Rejecting the Hypotheses 1 and 2

Figure 7 shows the network. Here AW, BX, CY and DZ are four OD pairs, with respectively demand 130, 50, 50 and 80. The link traffic costs are:

$$\begin{aligned}
 c_{AW} &= 90 + 0.1f_{AW} & c_{tX} &= 60 \\
 c_{Bp} &= f_{Bp} & c_{px} &= 10 \\
 c_{Cq} &= f_{Cq} & c_{qY} &= 10 \\
 c_{sr} &= f_{sr} + 10
 \end{aligned}$$

and costs of other links are equal to zero. The cost functions and demands satisfy Assumptions 2–3. If requiring strict monotonic cost functions, all constant cost functions could add a term related to the link flow with very small coefficient to make them strictly monotonically increasing, such as:  $c_{tX} = 60 + 10^{-5}f_{tX}$ . In order to simplify the calculation, we make constant. The network was chosen such that each OD can only have two routes, which are indicated in Table 2.

Suppose available control measures are additional delays on all links. It is easy to figure out that if there is no constraint, the optimal delays for the network are all 0 (i.e. there is no Braess Paradox link). Any additional delay on the network will increase the total system cost. In this no-delay and no-constraint scenario, the traffic flow on link tX is 90 according to User Equilibrium. If the emission factor is constant and the equivalent link flow constraint on link tX is  $f_{tX} \leq 30$ , additional delays should be imposed on some links in order to satisfy the constraint. According to Hypothesis 1, additional delay should be implemented on link tX, or according to Hypothesis 2, additional delays should be implemented on (all or some of the) links of routes leading to or from tX: {Dt, ut, Ct, Bu, XY, XZ}. According to the network’s symmetric topology, only three links {Dt, ut, Ct} need to be considered. Delay on link pC is a special location suggested neither by Hypothesis 1 nor by Hypothesis 2. Table 2 shows the optimal delays, the route flows  $q$ , link flow  $f$ , route cost  $C$  and the total system costs  $TC$ , according to location sets for imposing additional delays.

Here  $C_{AW}, C_{BX}, C_{CY}$  and  $C_{DZ}$  are route costs of respective OD Pairs. The result shows that neither the location indicated by Hypothesis 1 nor the locations indicated by Hypothesis 2 lead to the least system cost with environmental constraints satisfaction, but the delay on link pC does. Neither Hypothesis 1 nor Hypothesis 2 is correct.

Comparing the traffic assignments between the no-control scenario and  $d_{pC}$ -scenario, link flows  $f_{Bp}$  and  $f_{Cq}$  do not change and both remain 50. The route flows on link Bp and Cq are different though. The network has an amplifier phenomenon in which a control measure on link pC doubles the effect for link flow  $f_{tX}$  compared to the original link pC. Therefore, the flow reduction on the constrained link is more efficient due to less influenced flow.

### Appendix 4: ‘Negative Bonus’ and ‘Positive Bonus’ of Delays by the Monotonicity of Emission Factor

Here, we use the same network in Appendix 3 to illustrate the influence of ‘negative bonus’ and ‘positive bonus’, which are discussed in Section 3.2, Section 4.2 and Section 4.3. Considering the subnetwork with OD:AW and two routes  $R_{AW}$  and  $R_{ABqW}$  from Figure 7, Link AB and Link qW carry the same compositions of traffic flow  $f_{AB} = f_{qW} = q_{ABqW}$ , and  $\frac{\partial f_{AB}}{\partial d_{AB}} = \frac{\partial f_{qW}}{\partial d_{qW}}$ . Instead of the constrained link tX, The environmental constraint is imposed on link AB.



If the emission factor is strict monotonically decreasing with average speed, additional delays on the unconstrained links cause an even less negative effect on the system cost than that on the constrained link. The delay on the constrained link raises the emission factor, and thus link tX requires stricter equivalent flow constraint  $f_{a,d_{AB}}^{ECeq} \leq f_{a,d_{qW}}^{ECeq}$ . Imposing a delay on the constrained link AB causes more congestion damage to the system than a delay on Link qW. This is the ‘negative bonus’ discussed in Section 3.2.

If the emission factor is in the range of increasing with average speed,  $f_{a,d_{AB}}^{ECeq} \geq f_{a,d_{qW}}^{ECeq}$ . A milder delay on Link AB can be imposed and thus the ‘negative bonus’ becomes a ‘positive bonus’. Moreover, if the environmental constraint is strict, a single delay on Link AB reduces its average speed from above  $v_p^{OPT}$  to below  $v_p^{OPT}$ . Delays should be implemented both on Link AB and Link qW: a delay on Link AB for reducing the average speed to  $v_p^{OPT}$ , and another delay on Link qW for further reducing the flow to the equivalent value under  $v_p^{OPT}$ . These two concepts are discussed in Section 4.2.

### Appendix 5: Proof of Proposition 7

Suppose a combination of control measures  $(t^*, d^*)$  can achieve the minimal system cost while satisfying the environmental constraints where  $d^* \neq 0$ . The different set  $(t^* + d^*, 0)$  sustains the same traffic flow patterns according Proposition 1, and  $TC(t^* + d^*, 0) \leq TC(t^*, d^*)$  according to Proposition 2. Because  $TC(t^*, d^*) \leq TC(t^* + d^*, 0)$  due to  $TC(t^*, d^*)$  is minimal,  $TC(t^* + d^*, 0) = TC(t^*, d^*)$ . For link emission,  $\frac{\partial E_{ap}}{\partial t} \leq \frac{\partial E_{ap}}{\partial d} = \frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial d} f_a l_a + \frac{\partial E_{ap}}{\partial t}$  because  $\frac{\partial \varepsilon_{ap}}{\partial v_a} \frac{\partial v_a}{\partial d} \geq 0$  according to Equation (3.4) and (3.5). Therefore,  $E_{ap}(t^* + d^*, 0) \leq E_{ap}(t^*, d^*) \leq EC_{ap} \forall a \in \mathcal{A}_{EC}$ . It indicates  $(t^* + d^*, 0)$  is also in the feasible space for the environmental constraints. As a result,  $(t^* + d^*, 0)$  is one of the solutions for the OSP-EC problem. Therefore, only considering tolls is enough for minimizing the system cost with environmental constraints while both control measures are available.

### Appendix 6: An Example of Additional Delays Increasing the Feasible Space

In this part, we use an example where the solution of the OSP-EC problem under tolls-only strategy or under tolls-and-speed-limits strategy does not exist, but additional delays could help to satisfy the given environmental constraints. (Figure 8)

The OD demand is 2000 veh/h, the two links have the same length 6 km, and the same link cost function  $c_a = 0.05(1 + \frac{1}{2000} f_a)$  hour. Both links are constrained and with the same constraint value 160 g/h/km. The emission factor (g/km) function is the same one as in Section 2.5.

$$\varepsilon_{ap} = \begin{cases} 1.05 - 0.015 v_a & v_a < 60 \text{ km/hour} \\ 0.001(v_a - 60) + 0.15 & v_a \geq 60 \text{ km/hour} \end{cases} \tag{35}$$

Obviously, the emission factor is not monotonic and the optimal speed for emission is  $v = 60$  km/hour.

In tolls-only strategy, the equivalent flow on both links can be solved by the equation  $\varepsilon_{ap}(l_a/c_a(f_a))f_a = 160$  and the equivalent flow for both link is 930 veh/h. Because the constraints are conflicting, the total maximal allowed flows are 1860 veh/h under tolls-only strategy, smaller than the total demand. There is no solution for the OSP-EC problem.

By the help of additional delays on links, the emission factor will decrease by reducing the average speed. According to the discussion in Section 4.3, the optimal control measures  $(t, d)$  with the unit hour are (0.025; 0.0107). Here the toll is the marginal external congestion cost, which be omitted from both routes with inelastic demand (where only a toll *difference* affects the assignment). The delays reduce the average speeds and emission factors, so the equivalent flow constraints are relaxed. The solution can now be acquired. Furthermore, the delay of 0.0107 h is the solution for OSP-EC problem. If the minimal total emission is required, the delay should be 0.025 h to let  $v = 60$  km/hour. The total emission reduces from 1920 g/h to 1800 g/h, but the total cost will increase from 171.42 h to 200 h.

Moreover, the solution of the problem with toll and delays combined suggests the equivalent speed limit should be 70 km/h on each link. However, if both links are with the speed limit 70 km/h, the link flow patterns of deterministic static user equilibrium is not unique (Yang et al. 2012; Liu et al. 2017). Some of link flow patterns satisfy the constraints, but some of them do not. The speed limits cannot ensure that the emission constraints must be satisfied.

## References

- Aziz HMA, Ukkusuri SV (2012) Integration of Environmental Objectives in a System Optimal Dynamic Traffic Assignment Model. *Comput Civ Infrastruct Eng* 27:494–511. <https://doi.org/10.1111/j.1467-8667.2012.00756.x>
- Aziz HMA, Ukkusuri SV, Zhan X (2017) Determining the Impact of Personal Mobility Carbon Allowance Schemes in Transportation Networks. *Netw Spat Econ* 17:505–545. <https://doi.org/10.1007/s11067-016-9334-x>
- Baldasano JM, Goncalves M, Soret A, Jimenez-Guerrero P (2010) Air pollution impacts of speed limitation measures in large cities: The need for improving traffic data in a metropolitan area. *Atmos Environ* 44: 2997–3006. <https://doi.org/10.1016/j.atmosenv.2010.05.013>
- Barbosa HM, Tight MR, May AD (2000) A model of speed profiles for traffic calmed roads. *Transp Res Part A Policy Pract* 34:103–123. [https://doi.org/10.1016/S0965-8564\(98\)00067-6](https://doi.org/10.1016/S0965-8564(98)00067-6)
- Benedek CM, Rilett LR (1998) Equitable traffic assignment with environmental cost functions. *J Transp Eng* 124:16–22
- Boyd S, Vandenberghe L (2004) *Convex Optimization*. Cambridge University Press, New York
- Braess D (1968) Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforsch Oper Res - Rech Opérationnelle* 12:258–268. <https://doi.org/10.1007/BF01918335>
- Carlsaw DC, Beevers SD, Tate JE et al (2011) Recent evidence concerning higher NOx emissions from passenger cars and light duty vehicles. *Atmos Environ* 45:7053–7063. <https://doi.org/10.1016/j.atmosenv.2011.09.063>
- Chen A, Zhou Z, Ryu S (2011) Modeling Physical and Environmental Side Constraints in Traffic Equilibrium Problem. *Int J Sustain Transp* 5:172–197. <https://doi.org/10.1080/15568318.2010.488277>
- Chen L, Yang H (2012) Managing congestion and emissions in road networks with tolls and rebates. *Transp Res Part B Methodol* 46:933–948. <https://doi.org/10.1016/j.trb.2012.03.001>

- De Borger B, Proost S (2013) Traffic externalities in cities: The economics of speed bumps, low emission zones and city bypasses. *J Urban Econ* 76:53–70. <https://doi.org/10.1016/j.JUE.2013.02.004>
- Du X, Wu Y, Fu L et al (2012) Intake fraction of PM<sub>2.5</sub> and NO<sub>x</sub> from vehicle emissions in Beijing based on personal exposure data. *Atmos Environ* 57:233–243. <https://doi.org/10.1016/j.atmosenv.2012.04.046>
- European Environment Agency (2016) EMEP/EEA air pollutant emission inventory guidebook 2016. Publications Office of the European Union, Luxembourg
- Farahani RZ, Miandoabchi E, Szeto WY, Rashidi H (2013) A review of urban transportation network design problems. *Eur J Oper Res* 229:281–302. <https://doi.org/10.1016/j.ejor.2013.01.001>
- Feng T, Zhang J, Fujiwara A, Timmermans HJP (2010) An integrated model system and policy evaluation tool for maximizing mobility under environmental capacity constraints: A case study in Dalian City, China. *Transp Res Part D Transp Environ* 15:263–274. <https://doi.org/10.1016/j.trd.2010.03.001>
- Ferrari P (1995) Road pricing and network equilibrium. *Transp Res Part B Methodol* 29:357–372. [https://doi.org/10.1016/0191-2615\(95\)00013-4](https://doi.org/10.1016/0191-2615(95)00013-4)
- Fisk CS (1984) Game theory and transportation systems modelling. *Transp Res Part B Methodol* 18:301–313. [https://doi.org/10.1016/0191-2615\(84\)90013-4](https://doi.org/10.1016/0191-2615(84)90013-4)
- Gkatzoflias D, Kouridis C, Ntziachristos L, et al (2012) COPERT 4 Computer programme to calculate emissions from road transport User manual (version 9.0)
- Haas I, Bekhor S (2017) An Alternative Approach for Solving the Environmentally-Oriented Discrete Network Design Problem. *Netw Spat Econ* 17:963–988. <https://doi.org/10.1007/s11067-017-9355-0>
- Hendriks MMWB, De Boer JH, Smilde AK, Doornbos DA (1992) Multicriteria decision making. *Chemom Intell Lab Syst* 16:175–191. [https://doi.org/10.1016/0169-7439\(92\)80036-4](https://doi.org/10.1016/0169-7439(92)80036-4)
- Holman C, Harrison R, Querol X (2015) Review of the efficacy of low emission zones to improve urban air quality in European cities. *Atmos Environ* 111:161–169. <https://doi.org/10.1016/j.atmosenv.2015.04.009>
- Int Panis L, Beckx C, Broekx S et al (2011) PM, NO<sub>x</sub> and CO<sub>2</sub> emission reductions from speed management policies in Europe. *Transp Policy* 18:32–37. <https://doi.org/10.1016/j.TRANPOL.2010.05.005>
- Jakkula N, Asakura Y (2009) Accuracy of Optimum Road Pricing considering Local Emissions of Road Traffic Network. *IFAC Proc Vol* 42:161–167. <https://doi.org/10.3182/20090902-3-US-2007.0055>
- Johansson-Stenman O (2006) Optimal environmental road pricing. *Econ Lett* 90:225–229. <https://doi.org/10.1016/j.econlet.2005.08.008>
- Johansson O (1997) Optimal road-pricing: Simultaneous treatment of time losses, increased fuel consumption, and emissions. *Transp Res Part D-Transport Environ* 2:77–87
- Kaddoura I, Kröger L, Nagel K (2017) User-specific and Dynamic Internalization of Road Traffic Noise Exposures. *Netw Spat Econ* 17:153–172. <https://doi.org/10.1007/s11067-016-9321-2>
- Keller J, Andreani-Aksoyoglu S, Tinguely M et al (2008) The impact of reducing the maximum speed limit on motorways in Switzerland to 80 km h<sup>-1</sup> on emissions and peak ozone. *Environ Model Softw* 23:322–332. <https://doi.org/10.1016/j.emvsoft.2007.04.008>
- Kickhöfer B, Nagel K (2016) Towards High-Resolution First-Best Air Pollution Tolls. *Netw Spat Econ* 16: 175–198. <https://doi.org/10.1007/s11067-013-9204-8>
- Larsson T, Patriksson M (1999) Side constrained traffic equilibrium models - analysis, computation and applications. *Transp Res PART B* 33:233–264
- Larsson T, Patriksson M (1995) An augmented lagrangean dual algorithm for link capacity side constrained traffic assignment problems. *Transp Res Part B Methodol* 29:433–455. [https://doi.org/10.1016/0191-2615\(95\)00016-7](https://doi.org/10.1016/0191-2615(95)00016-7)
- Larsson T, Patriksson M, Rydbergren C (2004) A column generation procedure for the side constrained traffic equilibrium problem. *Transp Res Part B Methodol* 38:17–38. [https://doi.org/10.1016/s0191-2615\(02\)00092-9](https://doi.org/10.1016/s0191-2615(02)00092-9)
- Li Z-C, Wang Y-D, Lam WHK et al (2014) Design of Sustainable Cordon Toll Pricing Schemes in a Monocentric City. *Netw Spat Econ* 14:133–158. <https://doi.org/10.1007/s11067-013-9209-3>
- Li ZC, Lam WHK, Wong SC, Sumalee A (2012) Environmentally sustainable toll design for congested road networks with uncertain demand. *Int J Sustain Transp* 6(3):127–155
- Lin S, De Schutter B, Zegeye SK, et al. (2010a) Integrated urban traffic control for the reduction of travel delays and emissions. In: the 13th International IEEE Conference on Intelligent Transportation Systems. Madeira Island, pp 677–682
- Lin X, Immers BH, Viti F, Tampere C (2010b) From environmental constraints to network-wide environment-friendly traffic patterns: Inverting the planning approach for sustainability. In: 12th World Conference on Transport Research. Lisbon
- Lin X, Tampère CMJ, Viti F, Immers B (2016) The Cost of Environmental Constraints in Traffic Networks: Assessing the Loss of Optimality. *Netw Spat Econ* 16:349–369. <https://doi.org/10.1007/s11067-014-9228-8>

- Liu Z, Yi W, Wang S, Chen J (2017) On the Uniqueness of User Equilibrium Flow with Speed Limit. *Netw Spat Econ* 17:763–775. <https://doi.org/10.1007/s11067-017-9343-4>
- Madireddy M, De Coensel B, Can A et al (2011) Assessment of the impact of speed limit reduction and traffic signal coordination on vehicle emissions using an integrated approach. *Transp Res Part D Transp Environ* 16:504–508. <https://doi.org/10.1016/j.trd.2011.06.001>
- Mascia M, Hu S, Han K et al (2017) Impact of Traffic Management on Black Carbon Emissions: a Microsimulation Study. *Netw Spat Econ* 17:269–291. <https://doi.org/10.1007/s11067-016-9326-x>
- Nagurney A (2010) The negation of the Braess paradox as demand increases: The wisdom of crowds in transportation networks. *EPL (Europhysics Lett)* 91:48002. <https://doi.org/10.1209/0295-5075/91/48002>
- Nagurney A (2000) Congested urban transportation networks and emission paradoxes. *Transp Res Part D Transp Environ* 5:145–151. [https://doi.org/10.1016/S1361-9209\(99\)00031-0](https://doi.org/10.1016/S1361-9209(99)00031-0)
- Nagurney A, Qiang Q, Nagurney LS (2010) Environmental Impact Assessment of Transportation Networks with Degradable Links in an Era of Climate Change. *Int J Sustain Transp* 4:154–171. <https://doi.org/10.1080/15568310802627328>
- Nie MY (2017) Why is license plate rationing not a good transport policy? *Transp A Transp Sci* 13:1–23. <https://doi.org/10.1080/23249935.2016.1202354>
- Nie MY, Liu Y (2010) Existence of self-financing and Pareto-improving congestion pricing: Impact of value of time distribution. *Transp Res Part A Policy Pract* 44:39–51. <https://doi.org/10.1016/J.TRA.2009.09.004>
- Perco M (2013) Is road pricing effective in abating pollution? Evidence from Milan. *Transp Res Part D Transp Environ* 25:112–118. <https://doi.org/10.1016/j.trd.2013.09.004>
- Proost S, Van Dender K (2001) The welfare impacts of alternative policies to address atmospheric pollution in urban road transport. *Reg Sci Urban Econ* 31:383–411. [https://doi.org/10.1016/S0166-0462\(00\)00079-x](https://doi.org/10.1016/S0166-0462(00)00079-x)
- Rinaldi M, Tampère CMJ (2015) An extended coordinate descent method for distributed anticipatory network traffic control. *Transp Res Part B Methodol* 80:107–131. <https://doi.org/10.1016/J.TRB.2015.06.017>
- Rotaris L, Danielis R, Marcucci E, Massiani J (2010) The urban road pricing scheme to curb pollution in Milan, Italy: Description, impacts and preliminary cost–benefit analysis assessment. *Transp Res Part A Policy Pract* 44:359–375. <https://doi.org/10.1016/J.TRA.2010.03.008>
- Sheffi Y (1984) *Urban transportation networks: equilibrium analysis with mathematical programming methods*. Prentice-Hall, Englewood Cliffs
- Small KA, Verhoef ET (2007) *The economics of urban transportation*. Routledge, New York
- Smit R, Ntziachristos L, Boulter P (2010) Validation of road vehicle and traffic emission models – A review and meta-analysis. *Atmos Environ* 44:2943–2953
- Smith MJ (1979) The existence, uniqueness and stability of traffic equilibria. *Transp Res Part B Methodol* 13: 295–304. [https://doi.org/10.1016/0191-2615\(79\)90022-5](https://doi.org/10.1016/0191-2615(79)90022-5)
- Solé-Ribalta A, Gómez S, Arenas A (2018) Decongestion of Urban Areas with Hotspot Pricing. *Netw Spat Econ* 18:33–50. <https://doi.org/10.1007/s11067-017-9349-y>
- Sugawara S, Niemeier D (2002) How Much Can Vehicle Emissions Be Reduced?: Exploratory Analysis of an Upper Boundary Using an Emissions-Optimized Trip Assignment. *Transp Res Rec J Transp Res Board* 1815:29–37. <https://doi.org/10.3141/1815-04>
- Szeto WY, Li X, O'Mahony M. 2008. Simultaneous occurrence of Braess and emission paradoxes. *Proceedings of the 6th International Conference on Traffic and Transportation Studies*, pp. 625–634
- Szeto WY, Jaber X, Wong SC (2012) Road Network Equilibrium Approaches to Environmental Sustainability. *Transp Rev* 32:491–518. <https://doi.org/10.1080/01441647.2012.690000>
- Wallace CE, Courage KG, Reaves DP, et al (1984) *TRANSYT-7F USER'S MANUAL*
- Wang J, Connors R (2018) Urban Growth, Urban Growth, Transport Planning, Air Quality and Health: A Multi-Objective Spatial Analysis Framework for a Linear Monocentric City Networks *Spat Econ* 18. <https://doi.org/10.1007/s11067-018-9398-x>
- Wang S (2013) Efficiency and equity of speed limits in transportation networks. *Transp Res Part C Emerg Technol* 32:61–75. <https://doi.org/10.1016/J.TRC.2013.04.003>
- Wang XL, Ye H, Yang H (2015) Decentralizing Pareto-efficient network flow / speed patterns with hybrid schemes of speed limit and road pricing. *Transp Res E* 83:51–64
- Wardrop JG (1952) Some Theoretical Aspects of Road Traffic Research. *Proc Inst Civ Eng* 1:325–362. <https://doi.org/10.1680/ipeds.1952.11259>
- Yang H, Bell MGH (1997) Traffic restraint, road pricing and network equilibrium. *Transp Res Part B-Methodological* 31:303–314
- Yang H, Huang H-J (1998) Principle of marginal-cost pricing: how does it work in a general road network? *Transp Res Part A Policy Pract* 32:45–54. [https://doi.org/10.1016/S0965-8564\(97\)00018-9](https://doi.org/10.1016/S0965-8564(97)00018-9)

- Yang H, Wang X, Yin Y (2012) The impact of speed limits on traffic equilibrium and system performance in networks. *Transp Res Part B Methodol* 46:1295–1307. <https://doi.org/10.1016/j.trb.2012.08.002>
- Yang Z, Chen G, Yu B (2008) Car ownership level for a sustainability urban environment. *Transp Res Part D Transp Environ* 13:10–18. <https://doi.org/10.1016/j.trd.2007.10.003>
- Yang Z, Feng T, Jia P (2005) Maximizing Car Ownership under Constraints of Environment Sustainability in a. *City J East Asia Soc Transp Stud* 6:3077–3089. <https://doi.org/10.11175/easts.6.3077>
- Yin Y, Lawphongpanich S (2006) Internalizing emission externality on road networks. *Transp Res Part D Transp Environ* 11:292–301
- Yin YF (2002) Multiobjective bilevel optimization for transportation planning and management problems. *J Adv Transp* 36:93–105
- Youn H, Gastner MT, Jeong H (2008) Price of Anarchy in Transportation Networks: Efficiency and Optimality Control. *Phys Rev Lett* 101:128701. <https://doi.org/10.1103/PhysRevLett.101.128701>
- Zegeye SK, De Schutter B, Hellendoorn J (2010) Model Predictive Traffic Control to Reduce Vehicular Emissions - An LPV-Based Approach. *Am Control Conf* 2010:2284–2289
- Zhang S, Wu Y, Hu J et al (2014) Can Euro V heavy-duty diesel engines, diesel hybrid and alternative fuel technologies mitigate NOX emissions? New evidence from on-road tests of buses in China. *Appl Energy* 132:118–126. <https://doi.org/10.1016/J.APENERGY.2014.07.008>
- Zhang X, Liu P, Li Z, Yu H (2013) Modeling the Effects of Low-carbon Emission Constraints on Mode and Route Choices in Transportation Networks. *Procedia Soc Behav Sci* 96:329–338. <https://doi.org/10.1016/J.SBSPRO.2013.08.040>
- Zhong RX, Sumalee A, Friesz TL, Lam WHK (2011) Dynamic user equilibrium with side constraints for a traffic network: Theoretical development and numerical solution algorithm. *Transp Res Part B Methodol* 45:1035–1061. <https://doi.org/10.1016/j.trb.2011.05.004>

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

## Affiliations

Xin Lin<sup>1</sup> · Chris M. J. Tampère<sup>1</sup> · Stef Proost<sup>2</sup>

Xin Lin  
xin.lin@live.com

Stef Proost  
stef.proost@kuleuven.be

<sup>1</sup> Centre for Industrial Management / Traffic and Infrastructure, KU Leuven, Celestijnenlaan 300A, Bus 2422, 3001 Heverlee, Belgium

<sup>2</sup> Faculty of Economics and Business, KU Leuven, Heverlee, Belgium