Robust signal timing optimization with environmental concerns

Lihui Zhang\textsuperscript{a}, Yafeng Yin\textsuperscript{b,\ast}, Shigang Chen\textsuperscript{c}

\textsuperscript{a} School of Transportation and Logistics, Dalian University of Technology, Dalian, Liaoning 116024, China
\textsuperscript{b} Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, USA
\textsuperscript{c} Department of Computer & Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA

\textbf{A R T I C L E  I N F O}

Article history:
Received 31 August 2010
Received in revised form 6 January 2013
Accepted 9 January 2013

Keywords:
Traffic signal timing
Traffic delay
Air pollutant concentration
Bi-objective optimization

\textbf{A B S T R A C T}

This paper formulates a bi-objective optimization model to determine timing plans for coordinated traffic signals along arterials to minimize traffic delay and the risk associated with human exposure to traffic emissions. Based on a cell-transmission representation of traffic dynamics, a modal sensitive emission approach is used to estimate the tailpipe emission rate for each cell of a signalized arterial. A cell-based Gaussian plume air dispersion model is then employed to capture the dispersion of air pollutants and compute the roadside pollutant concentrations. A measure of mean excess exposure is further defined to represent the risk associated with human exposure to traffic pollutants under the wind uncertainty. A signal timing optimization model is formulated to optimize the cycle length, offsets, green splits and phase sequences to minimize the total system delay and the mean excess exposure simultaneously. The bi-objective optimization model is solved via a simulation-based genetic algorithm to find a set of Pareto optimal solutions. A numerical example is presented to demonstrate the model.

\Copyright 2013 Elsevier Ltd. All rights reserved.

1. Background

Because of its negative effects on health and living conditions, air pollution has long emerged as one of the most acute problems in many metropolitan areas. A major source of this pollution is the emissions from vehicular traffic (see e.g., Cai et al., 2009; Sharma et al., 2010; Environmental Protection Agency, 2011). They contribute considerably to the level of, e.g., carbon monoxide (CO), nitrogen oxides (NOx), and volatile organic compounds (VOCs) in the environment. In order to achieve a more sustainable mobility, reducing traffic emissions has been an ongoing endeavor of many governmental authorities over the past two decades. However, many emission reducing programs implemented thus far are passive in nature because they regulate only the rates (e.g., grams per mile) at which a vehicle generates these hazardous air pollutants. Depending on the amount of driving or the mode of driving, cars meeting an emission standard can generate as much air pollutants as cars not meeting the standard. More proactive strategies that aim to reduce the mileage vehicles traveled or improve the efficiency of the transportation network have recently gained more and more attention (e.g. Yin and Lawphongpanich, 2006; Cambridge Systematics, 2009).

Traffic signals at intersections impose significant impact on traffic emissions because they interrupt traffic flow (for good reasons) and create additional deceleration, idle and acceleration driving modes to the otherwise cruise driving mode. Traffic emissions are very sensitive to the driving modes. For example, accelerating vehicles produce 350% more CO than those in the cruise driving mode (Environmental Protection Agency, 2002). Previous studies have investigated the impacts of signal...
Timing on vehicular emissions. Rakha et al. (2000) showed that efficient signal coordination can reduce emissions up to 50% in a highly simplified scenario. Hallmark et al. (2000) evaluated signal coordination strategy on CO emission reduction with an activity-specific approach, where the vehicle activity profile was obtained with handheld laser range-finding devices. Unal et al. (2003) directly collected emission data using a portable, on-road vehicle date measurement device, to show that signal coordination was able to reduce pollutant emissions by approximately 10–20%. Coelho et al. (2005) integrated a mesoscopic traffic model and the modal emission approach to investigate the impact of various speed control signal settings on vehicle emissions. Li et al. (2009) integrated a comprehensive modal emission model (Barth et al., 2000) and PARAMICS to evaluate vehicular emissions at signaled intersections. Hirschmann and Fellendorf (2010) developed a simulation toolbox to estimate polluting emissions under different signal control strategies, and simulation results show 5–12% emission reduction depending on pollutant types and signal control strategies. Madireddy et al. (2011) recently examined a signal coordination strategy that reduces vehicle emissions by about 10% on a real arterial. Tao et al. (2011) also used field data to evaluate the effectiveness of signal coordination on reducing vehicle emissions during both peak and non-peak hours.

Previous studies have also attempted to develop signal timing optimization models to minimize both congestion and emissions. Li et al. (2004) optimized the cycle length and green splits for an isolated intersection to minimize a weighted sum of the delay, fuel consumption and emission. Liao and Machemehl (1996), Stevanovic et al. (2009) and Park et al. (2009) developed signal timing optimization models to minimize the fuel consumption and vehicle emission based on microscopic simulation. More recently, Ma and Nakamura (2010) developed an analytical procedure to obtain the optimized cycle length against vehicle emissions for isolated intersections. These studies consider tailpipe emissions instead of roadside air pollution concentrations, which decide the local air quality. In other words, dispersion of air pollutants is not captured in their models. The tradeoff between delays and emissions in street canyons should be very different from that in open rural environments, because the dispersion procedure of air pollutants relies heavily on the surface roughness (e.g., Wieringa, 1993). Since air pollutants are more difficult to disperse in the former, they likely lead to higher pollution concentrations and worse air quality in the area adjacent to the streets. Much research has been carried out to investigate the impact of terrain conditions on atmospheric dispersion (e.g., McElroy, 1969; Briggs, 1973; Bowne, 1974; Dennis, 1978). Nowadays, most air dispersion models available are capable of modeling different terrains, and many of them are specifically developed for estimating concentrations near roadways, e.g. the CAR model (Eerens et al., 1993), the CALINE model (California Department of Transportation, 1989), the HYROAD model (Carr et al., 2002) and the ADMS-Roads model (Cambridge Environmental Research Consultants, 2006).

Moreover, previous signal timing optimization studies either use a highly simplified approach to compute traffic emissions or rely on microscopic driving-cycle-based emission models. The former approach, which is adopted in Li et al. (2004) and Ma and Nakamura (2010), estimates emissions based on average traffic condition at the link or intersection level, and thus fails to capture spatially and temporally varying traffic state and may underestimate the emissions due to accelerations and decelerations (Lin and Ge, 2006). The latter approach requires instantaneous vehicle information to calculate individual vehicle emissions for each modeling time step. To obtain such detailed information, researchers have turned to microscopic traffic simulation (e.g., Park et al., 2009). When such an emission estimation component is incorporated into an optimization framework, the resulting microscopic-simulation-based optimization model will be computationally intensive for real network optimization.

In contrast, this paper presents a signal timing optimization model for coordinated signal control along arterials to minimize traffic delay and roadside air pollutant concentrations. The model is macroscopic and computationally tractable. At the same time, it can model traffic dynamics to capture the impact of time-dependent traffic characteristics on emissions. More specifically, the macroscopic cell transmission model (CTM) developed by Daganzo (1994, 1995) is adopted to describe traffic dynamics. CTM represents urban streets as a set of consecutive cells and propagates traffic flow through these cells based on the fundamental diagram and signal status. The computation is very efficient. Based on the CTM representation of traffic dynamics, traffic delay may be estimated with reasonable accuracy. More importantly, the driving mode of the vehicles in each cell at each time interval can be identified as deceleration, idle, acceleration or cruise. Subsequently, the time-dependent cell emission rates for air pollutants such as CO and NOx can be determined using the modal emissions approach (e.g., Frey et al., 2001). With the emission rate of each cell, a Gaussian plume dispersion model (e.g., Turner, 1994) is further utilized to capture the dispersion of air pollutants and compute the roadside pollutant concentrations. For a given wind direction and speed, a measure of total human emissions exposure is calculated to represent the potential adverse impacts by the roadside pollutant concentrations. Given a set of scenarios of wind direction and speed, we define a risk measure, i.e., the mean excess exposure, as the mean of the total human emissions exposures incurred by high-consequence wind scenarios.

In financial engineering, the risk measure is known as conditional value-at-risk or CVaR (Rockafellar and Uryasev, 2000). With the above consideration, a bi-objective signal timing optimization model is developed to determine cycle length, offsets, green splits and phase sequences to simultaneously minimize traffic delay and the mean excess exposure. A genetic-algorithm-based solution approach (Yin, 2002) is adopted to solve the timing optimization problem for a set of Pareto optimal solutions. The solutions form an efficient frontier that presents explicit tradeoffs between the total delay of the corridor and the mean excess exposure of the roadside area. Based on the decision maker’s consideration of social costs of delays and emissions, an optimal timing plan can be selected.

For the remainder, Section 2 briefly introduces the CTM description of traffic dynamics in a signalized arterial. Section 3 elaborates the cell-based modal emissions approach to compute the cell emission rates, the cell-based Gaussian plume dispersion model to capture air pollutant dispersion and the definition of mean excess exposure. Section 4 presents the bi-
objective signal timing optimization model and develops a simulation-based genetic algorithm to solve the bi-objective timing optimization model. Section 5 is a numerical example to demonstrate the model and the solution algorithm, followed by concluding remarks in the last section.

2. Modeling traffic dynamics via cell-transmission model

We attempt to determine timing plans for a series of coordinated traffic signals along a local arterial, which share a common cycle length, and prescribe offset, green splits and phase sequence for each signal in coordination. We assume that the traffic demand for the arterial is given, and do not consider the impact of signal timing on drivers’ route choices.

Modeling traffic dynamics is particularly important for signal timing optimization due to the need for accurately evaluating various feasible timing plans. The evaluation process should be efficient such that the optimization procedure can be computationally tractable. This paper adopts CTM to describe traffic dynamics, such as shockwave, and queue formation and dissipation.

CTM is a finite differencing solution scheme for the first-order hydrodynamic theory of traffic flow, i.e., the Lighthill–Whitham–Richards (LWR) model. Mathematically the model can be stated as the following equations:

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0$$  \hspace{0.5cm} (1)

$$q = f(k, x, t)$$  \hspace{0.5cm} (2)

where $k$ and $q$ denote traffic density and flow respectively, which may vary across location $x$ and time $t$. Eq. (1) is the flow conservation equation and Eq. (2) defines the traffic flow at location $x$ and time $t$ as a function (denoted as $f$) of the density, the so-called fundamental diagram. For a homogeneous roadway, Daganzo (1994) suggested using the following time-invariant flow–density relationship:

$$q = \min\{Vk, Q_i(k_{\text{jam}} - k)\}$$

where $k_{\text{jam}}$ is the jam density; $Q_i$ is the inflow capacity; $V$ is the free-flow speed and $W$ is the backward wave speed. By dividing the whole network into homogeneous cells with the cell length equal to the duration of a time step multiplied by the free-flow speed, the results of the LWR model can be approximated by a set of recursive equations in the simplest case:

$$n_i(t + 1) = n_i(t) + y_{i-1}(t) - y_i(t)$$  \hspace{0.5cm} (3)

$$y_i(t) = \min\{n_i(t), Q_i(t), \omega \cdot [N_{i+1, \text{max}} - n_{i+1}(t)]\}$$  \hspace{0.5cm} (4)

where $n_i(t)$ is the number of vehicles in cell $i$ during time step $t$; $y_i(t)$ is the number of vehicles that leave cell $i$ during time step $t$; $N_{i, \text{max}}$ is the maximum number of vehicles that can be accommodated by cell $i$; $\omega$ is equal to $W/V$; $Q_i(t)$ is the minimum of the capacity flows of cell $i$ and $i + 1$. Eq. (3) ensures the flow conservation and Eq. (4) determines the outflow for each cell at each time step.

CTM has been extensively employed in the literature for modeling traffic dynamics in various applications, e.g. signal timing optimization (e.g., Lo, 1999; Lo et al., 2001; Lin and Wang, 2004; Karoonsoontawong and Waller, 2009, 2010; Zhang et al., 2010; Li, 2011), dynamic traffic assignment (e.g., Ziliaskopoulos, 2000; Lo, 2001), dynamic network design (e.g., Karoonsoontawong and Waller, 2006), and evacuation planning (e.g. Xie et al., 2010; Ben-Tal et al., 2011). This paper adopts the CTM implementation that Zhang et al. (2010) proposed for a general signal-controlled arterial network. In their implementation, all the cells comprising the network are categorized into seven groups: ordinary, origin, destination, non-signalized diverge, non-signalized merge, signalized diverge and signalized merge cells, as shown in Fig. 1 from (a) to (g) correspondingly.

Fig. 1a denotes an ordinary cell (cell $i$) whose flow propagation is described by Eqs. (3) and (4). Using a technique recently proposed by Pavlis and Recker (2009), the equations can be equivalently converted into a linear system of equalities and inequalities with integer variables. Fig. 1b is an origin cell (cell $i$) with the inflow fixed as the exogenous demand input, denoted as $d_i$ in the figure. The flow propagation for this type of cells is similar to Eqs. (3) and (4). The destination cell, as shown in Fig. 1c, is also similar to the ordinary cell, but the outflow is unlimited, implying that all the vehicles currently residing in the cell are able to flow out of the system at the next time step. Fig. 1d illustrates a non-signalized diverge cell where the geometry or capacity changes and traffic diverge into different lanes for their respective destinations. In the figure, traffic in cell $i$ diverges into cells $j_1$ and $j_2$ according to some pre-determined proportion parameters. Fig. 1e shows non-signalized merge, where the traffic in upstream cells $i_1$ and $i_2$ merge into the downstream cell $j$. The merging flow rates are determined based on the traffic condition and priority setting at the merge point. Fig. 1f sketches a signalized diverge cell, where the sign $S$ indicates a traffic signal. A signalized diverge occurs within a signalized intersection. Traffic from one direction enters the intersection during a green phase and then leaves the intersection while diverging into two or more directions of traffic. Lastly, Fig. 1g presents an example of traffic merge under signal control. According to the signal settings, three streams of

---

1 To simplify the notation, we here use the same $t$ to indicate the discretized time interval.
traffic entering the intersection, $y_i$, $y_j$, and $y_k$ in the figure, are associated with three individual signal phases that conflict with each other. Practically there is only one stream of traffic entering the intersection at one time step based on the signal status. For a detailed mathematical description of flow propagation through these cells, please refer to Zhang et al. (2010). In summary, the CTM description of traffic dynamics along a general signal-controlled arterial network leads to a linear system of equalities and inequalities with integer variables. The linear system captures the impact of signal timing on the flow propagation and cell traffic states.

In this paper, signals are assumed to be under fixed-time control and signal timings are optimized with the NEMA phasing structure. Fig. 2a illustrates the standard NEMA, dual ring phase structure for a four-way intersection. With the barrier in the middle, the structure is divided into four portions (phases 1 and 2, phases 3 and 4, phases 5 and 6, phases 7 and 8). Introducing a binary variable to indicate which phase appears first in each portion, a phase sequence can be represented by a quadruplet ($k_1, k_2, k_3, k_4$) where, if $k=0$, the odd phase appears first, otherwise, the even phase leads. For example, the phase sequence shown in Fig. 2b can be represented as $(1,0,0,0)$.

One of the objectives for signal timing optimization is to minimize the total system delay of an urban arterial. With the above CTM implementation, the objective function is to minimize the total area (as in Fig. 3) between the cumulative arrival curves of the origin cells and the cumulative departure curves of the destination cells, expressed as the following linear function:

$$TD = \min \left( \sum_{i=0}^{T} \sum_{t=1}^{t} d_i(t) - \left( \sum_{i=0}^{T} \sum_{t=1}^{t} y_i(t) \right) \right) \tag{5}$$

where $TD$ is the total system delay; $O$ is the set of origin cells and $D$ is the set of destination cells; $d_i(t)$ is the demand at origin cell $i$ during time step $t$; $T$ is the modeling horizon. It is straightforward to observe that if the demands at origin cells are given, the objective function is equivalent to maximizing the second component, i.e., the area under the cumulative departure curves.

3. Emission and roadside pollutant concentrations

3.1. Emissions model

Existing emission modeling systems can estimate and predict traffic emission at regional, facility and individual-vehicle levels. See e.g., Yu et al. (2009), for a recent review on the emission models. For the analysis of signalized intersection emis-
sions, the modal and instantaneous emission models are more applicable because they predict second-by-second tailpipe emissions as a function of the vehicle’s operating mode (e.g., Cernuschi et al., 1995; Barth et al., 2000; Rakha et al., 2004).

In this paper, emissions factors are estimated using the modal emissions approach proposed by Frey et al. (2001, 2002) where a vehicle’s operating mode is divided into four modes of idle, acceleration and deceleration and cruise, and the average emissions factors for each mode are determined. As demonstrated below, the resolution and data requirement of this approach are very compatible with the CTM representation of traffic dynamics. Moreover, it is very efficient to be incorporated into the signal timing optimization procedure.

Frey et al. (2001) defined the idle mode as zero speed and zero acceleration. For the acceleration mode, the vehicle speed must be greater than zero with an acceleration rate averaging at least 1 mph/s for 3 s or more. Deceleration is defined in a similar manner as acceleration but with a negative acceleration rate. The cruising mode is approximately steady-speed driving, but some drifting of speed is allowed. In the CTM implementation, the traffic state at each cell is determined at each time interval and the driving mode of the vehicles in the cell can then be identified through comparing the densities of the current and the immediately downstream cells.

\[ 2 \text{ In CTM, vehicles in one cell are treated homogenous.} \]
Fig. 4 illustrates how the driving mode of an ordinary cell is identified. The current cell is \( i \) and the downstream cell is \( i + 1 \) and their densities are denoted as \( k_i \) and \( k_{i+1} \). The fundamental diagram of the signalized arterial is also presented in the figure to facilitate the presentation, where \( k_{jam} \) is the jam density and \( k_c \) is the critical density. In Fig. 4, the relationship between \( k_i \) and \( k_{i+1} \) falls into one of the four areas, i.e., C, A, D and I. Area "C" stands for the cruise mode and consists of two portions: the first portion is a rectangle for the situations when \( k_i \leq k_c \) and \( k_{i+1} \leq k_c \), i.e., the vehicles in both cells are traveling at the free-flow speed; the other is the line separating the areas of "A" and "D", where \( k_i = k_{i+1} < k_{jam} \), implying that vehicles in both cells are traveling at the same positive speed. The end of that line is the area (point) of "I", representing the idle mode. At this point, \( k_i = k_{i+1} = k_{jam} \) and all vehicles have a zero speed. Albeit a point, it occurs frequently in the cells near signalized intersections during the red intervals. The lower area of "D" represents the deceleration mode where \( k_i < k_{i+1} \), and \( k_{i+1} > k_c \). This implies that vehicles in the current cell are traveling faster than the downstream vehicles. Because these vehicles will move into the downstream cell, they have to decelerate to adapt to the prevailing downstream speed. Similarly, Area "A" represents the acceleration mode where \( k_i > k_{i+1} \) and \( k_i > k_c \), implying that vehicles in the current cell are traveling slower than the downstream vehicles, and are accelerating. In summary, by simply comparing the cell density with that of the immediately downstream cell, one can easily identify the driving mode of the vehicles in each cell. As an example, Fig. 4 shows a case that the vehicles in cell \( i \) are in the deceleration mode.

3.2. Pollution dispersion model

Frey et al. (2001, 2002) developed the average emissions factors for each driving mode. Applying these factors, the cell emission rate can be calculated as follows:

\[
ER_i(t) = n_i(t) \times EF_i(t)
\]

where \( ER_i(t) \) denotes the emission rate of cell \( i \) at time interval \( t \); \( n_i(t) \) is the number of vehicles in the cell; \( EF_i(t) \) is the emission factor for a driving mode in units of mg/s/vehicle or g/s/vehicle.

Fig. 4. Driving mode of cell \( i \).
With the cell emission rates, an atmospheric dispersion model can be used to estimate the roadside air pollutant concentrations. To facilitate the presentation of the modeling framework, this paper adopts the simplest steady-state Gaussian plume dispersion model (e.g., Turner, 1994). Other more advanced dispersion models can be applied in a similar way, particularly those non-steady-state models to capture the time-varying emission rates computed above. However, we note that the dispersion model should be computationally tractable to maintain the efficiency of signal timing optimization procedure.

To apply the Gaussian plume dispersion model, it is assumed that the pollution source of each cell is located in the cell center. Within the modeling horizon, e.g., 1 h, the average cell emission rate $ER_i$ is estimated as $ER_i = \sum_{t=1}^{T} ER_i(t)/T$. Subsequently, at any point in space $(x, y, z)$, the incremental pollutant concentration $C_i(x, y, z)$ from the source cell $i$ can be approximately estimated as follows:

$$C_i(x, y, z) = \frac{ER_i}{2\pi U \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left\{ \exp\left[\frac{-(z-h)^2}{2\sigma_z^2}\right] + \exp\left[\frac{-(z+h)^2}{2\sigma_z^2}\right] \right\}$$

where $U$ is the wind speed at the pollutant release height $h$; $x$ is the distance along the wind direction; $y$ is the distance along the cross-wind direction; $z$ is the vertical distance; $\sigma_y$ is the cross-wind dispersion coefficient and $\sigma_z$ is the vertical dispersion coefficient.

For numerical computation, the roadside area is cut into grids and the pollutant concentrations of each grid are computed with the coordinates of the center of the grid. Fig. 5 presents two concentration contour maps for one particular signalized arterial. The $Y$ axis is along the center line of the main street, while the $X$ axis represents the direction of cross streets, with $(0,0)$ at the south end of the whole network. The maps are generated using the same wind speed but different directions. The wind direction in Fig. 5a is $\pi/4$ clockwise to the north while in Fig. 5b it is the north. These two examples demonstrate the need for considering pollutant dispersion as the impacts of traffic emissions will be substantially different in these two cases.

### 3.3. Mean excess exposure

We define a surrogate measure to represent the potential negative impacts of roadside pollutant concentrations. The so-called total human emissions exposure is calculated as follows:

$$TE = \sum_{i=1}^{I} \int_y \int_x p(x, y, z) \cdot C_i(x, y, z) \, dx \, dy \, dz$$

where $TE$ is the total human emissions exposure; $I$ is the total number of cells of the signalized corridor and $p(x, y, z)$ is the population density function beside the corridor.

The above definition is written for one particular pollutant for illustration purposes. If multiple pollutants are considered, Eq. (6) can be applied to each pollutant and the total human emissions exposure is the weighted sum of the exposures to all pollutants. The weighting factors are determined based on, e.g., the social cost of each pollutant (Delucchi et al., 2002).
The above human emissions exposure is computed for a certain wind speed $U$ and direction $WD$. These two parameters change constantly while the signal plan implemented in the field remains the same for a certain period of time, e.g., years. Therefore, the changing wind speed and direction should be proactively considered. To represent such wind uncertainty, we introduce a set of wind scenarios $\Omega = \{1, 2, 3, \ldots, S\}$. A scenario consists of a certain wind speed and direction, i.e., $\omega^s = (U^s, WD^s)$, and its probability of occurrence is $p^s$. The wind scenarios can be specified using the wind rose of the study area. Fig. 6 is the 30-year wind rose map from United States Department of Agriculture (USDA) for the month of July obtained from a wind station in San Francisco bay area.

Given the set of wind scenarios, it is straightforward to estimate the mean of the human emissions exposures across all scenarios, and then optimize signal timings to minimize it. However, decision makers may be more concerned with worst-case scenarios where substantial emissions exposure may occur. To address such a risk-averse attitude and avoid being too conservative at the same time, we optimize the signal timings against a set of worst-case scenarios (Yin, 2008). More specifically, we minimize the expected emissions exposure incurred by those high-consequence scenarios whose collective probability of occurrence is $1 - \alpha$, where $\alpha$ is a specified confidence level, e.g., 80%. In financial engineering, the performance measure is known as conditional value-at-risk or CVaR (Rockafellar and Uryasev, 2000) and we name it as mean excess (emissions) exposure.

Mathematically, for a wind scenario $\omega^s$, the total emissions exposure can be computed, denoted as $TE_s$. Consider all wind scenarios and order the emissions exposure as $TE_1 < TE_2 < \ldots < TE_S$. Let $s_\alpha$ be the unique index such that:

$$\sum_{s=1}^{s_\alpha} p_s > \sum_{s=1}^{s_\alpha-1} p_s$$

In words, $TE_{s_\alpha}$ is the maximum emissions exposure that is exceeded only with probability $1 - \alpha$. Consequently, the expected emission exposure exceeding $TE_{s_\alpha}$, i.e., the mean excess exposure, can be computed as follows:

$$MEE_\alpha = \frac{1}{1-\alpha} \left[ \left( \sum_{s=1}^{s_\alpha} p_s - \alpha \right) TE_{s_\alpha} + \sum_{s=s_\alpha+1}^{S} p_s TE_s \right]$$

(7)

Fig. 6. Example of wind rose map. Source: USDA.
The second component in the bracket is simply to compute the mean value, and the first is to split the probability ‘atom’ at the emissions exposure point $TE_a$, to make the collective probability of scenarios considered in the bracket exactly equal to $1 - \alpha$. Rockafellar and Uryasev (2002) showed that minimizing Eq. (7) is equivalent to minimizing the following equation:

$$\min_{i} Z_{2} = \zeta + \frac{1}{1-\alpha} \sum_{s=1}^{5} p_{s} \cdot \max(TE_s - \zeta, 0)$$

where $\zeta$ is a free decision variable. The max operator in the above equation can be replaced by another decision variable and two additional linear constraints, thereby simplifying Eq. (8) to be a linear function. For a further illustration of the concept, please refer to Zhang and Yin (2008).

4. Model formulation and solution algorithm

4.1. Bi-objective optimization model

Given a particular network, the cell representation should be first constructed based on the geometry and signal setting. The cells are then classified into seven categories and the corresponding constraints for flow propagation can be written for each cell as previously described in Section 2. The constraints comprise a linear system with integer variables. Other constraints include those for signal timings such as minimum and maximum green times, and relationships of phase sequences. For numerical computation of roadside pollutant concentrations, the roadside area is cut into grids and the constraints for computing pollutant concentrations can be written. All these constraints are essentially linear (with respect to decision variables) but involve integer variables. With the linear objective functions to minimize total system delay (Eq. (5)) and the mean excess exposure (Eq. (8)), the optimization problem is a bi-objective mixed-integer linear program. One portion of the optimal solution to the program specifies the signal timing, denoted as a vector $(l', o', g', \lambda')$, where $l'$, $o'$, $g'$ and $\lambda'$ are vectors of optimal cycle length, offsets, green splits and phase sequences. The first three are integer variables and phase sequences are in binary format.

Conceptually, the bi-objective signal optimization model is summarized as follows:

$$\min_{(l, o, g)} \{ TD, Z_{2} \}$$

s.t. Constraints for traffic dynamics and relating signal timing to flow propagation
Constraints for feasible signal settings
Constraints for computing pollutant concentrations and human emissions exposures

There may not exist an unambiguous optimal solution that minimizes both the total delay and mean excess exposure simultaneously. Hence, a set of Pareto optimal solutions or non-dominated solutions are sought instead. Those solutions are optimal in the sense that no improvement can be achieved in any objective without degradation in the other. All these solutions form a Pareto frontier. Based on the decision maker’s consideration of social costs of delays and emissions, an optimal timing plan can be selected.

4.2. Simulation-based bi-objective genetic algorithm

When the optimization horizon is long and the network size is large, the bi-objective model formulated will be a very large mixed-integer linear program. Therefore, conventional algorithms, such as branch and bound, are not able to solve it efficiently. Previous studies have attempted to solve such problems via meta-heuristics, mostly genetic algorithms (GAs) (see e.g. Lo et al., 2001; Sun et al., 2006; Jabari et al., 2009; Meng and Khoo, 2010; Li, 2011). Similarly we here apply a simulation-based binary GA, where the “simulation-based” means that the fitness function value is calculated through a CTM-based macroscopic simulation.

The proposed simulation-based GA follows the general framework of GA and its core components are discussed below:

4.2.1. Chromosome configuration

The chromosome is defined according to the decision variables, which include the cycle length, offsets, green splits and phase sequences. Fig. 7 presents an example of the chromosome that represents a three-signal arterial. There are in total 144 bits, of which the first six bits represent the cycle length and the rest is equally divided into three portions, 46 bits for each signal. Consider the first signal. Bits 7–10 define the phase sequences for the signal. The next seven bits, i.e., 11–17, represent the offset for the signal. A seven-bit binary number can represent a decimal number from 0 to 127. Because the offset is expressed as a percentage of the cycle length in this paper, one additional constraint on the binary number is in place to ensure the feasibility of the offset. Bits 18–24 represent the barrier point, which is also expressed as a percentage of the cycle length. The next four clusters of bits represent four green times $G_1, G_2, G_3$ and $G_4$, which are the green durations of the phases that lead in the respective portions within the NEMA phasing structure. $G_1$ and $G_3$ are in percentage of the barrier time while $G_2$ and $G_4$ are in percentage of the difference between cycle length and barrier time.
4.2.2. Fitness evaluation

For each generation, every individual needs to be evaluated so as to decide its priority to breed the next generation. Here each individual represents a specific signal timing plan and the fitness evaluation is to determine its corresponding system delay and mean excess emission exposure. More specifically, for each individual signal plan, we run a macroscopic simulation based on CTM with all wind scenarios and calculate the corresponding system control delays and mean excess exposure. The fitness value of each signal plan is then determined using the modified distance method proposed by Osyczka and Kundu (1996). Denote the existing Pareto solution set as \( H = \{1, 2, 3, \ldots, L\} \). A “latent potential value” \( PV \) is associated with each Pareto solution, and the maximum one is denoted as \( PV_{\text{max}} \). A newly generated solution will yield different distances measured in the delay-emission space from all the existing Pareto solutions, denoted as \( (DS_1, DS_2, DS_3, \ldots, DS_L) \). The latent potential
value of the nearest Pareto solution is identified as $PV_{\text{near}}$, and the corresponding minimum distance is denoted as $DS_{\text{min}}$. Subsequently, the fitness value of a new solution ($FV_{\text{new}}$), and its latent potential value ($PV_{\text{new}}$) are determined based on its Pareto dominance relation with the existing Pareto solutions, as shown in Table 1.

### Table 1
Calculation of the fitness value of each new solution.

<table>
<thead>
<tr>
<th>Dominance relations</th>
<th>It is a new Pareto solution, and dominates some Pareto solutions</th>
<th>It is a new Pareto solution, but dominates no other Pareto solutions</th>
<th>It is not a Pareto solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update 1: fitness value</td>
<td>$FV_{\text{new}} = PV_{\text{max}} + DS_{\text{min}}$</td>
<td>$FV_{\text{new}} = PV_{\text{near}} + DS_{\text{min}}$</td>
<td>$FV_{\text{new}} = PV_{\text{near}} - DS_{\text{min}}$</td>
</tr>
<tr>
<td>Update 2: latent potential value</td>
<td>$PV_{\text{new}} = FV_{\text{new}}$</td>
<td>$PV_{\text{new}} = FV_{\text{near}}$</td>
<td>$PV_{\text{new}} = \frac{PV_{\text{near}}}{C_0}$</td>
</tr>
<tr>
<td>Update 3: maximum latent potential value</td>
<td>$PV_{\text{max}} = PV_{\text{new}}$</td>
<td>$PV_{\text{max}} = \max (PV_{\text{max}}, PV_{\text{new}})$</td>
<td>Unchanged</td>
</tr>
</tbody>
</table>

Fig. 8. Cell representation of El Camino Real.
4.2.3. Crossover
To increase diversity, we use multi-point crossover other than one-point crossover. After the selection of two parents, several crossover points will be randomly generated but ensure that one is among the first six bits, which influences the cycle length, and another one for each signal, which may change the setting for each signal. Therefore, if there are \( n \) signals, there will be \( n + 1 \) crossover points in total.

4.2.4. Mutation
Each crossover operation will generate two offspring and the mutation operation is subsequently conducted. The mutation rate used in the numerical test is 5\( \% \).

In the first generation of the proposed GA, a set of Pareto optimal signal timing plans is obtained based on the randomly generated timing plans. The Pareto set is updated every generation with the reproduction, crossover and mutation operations. For a more detailed description of the algorithm, please refer to Yin (2002).

**Table 2**
Emission factors for different driving modes (mg/s/vehicle).

<table>
<thead>
<tr>
<th>Driving mode</th>
<th>Idle</th>
<th>Accelerate</th>
<th>Decelerate</th>
<th>Cruise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emission factor</td>
<td>20.04</td>
<td>2013.96</td>
<td>77.15</td>
<td>447.11</td>
</tr>
</tbody>
</table>

**Table 3**
30-Year wind probability data.

<table>
<thead>
<tr>
<th>Wind speed (m/s)</th>
<th>1.80–3.34</th>
<th>3.34–5.40</th>
<th>5.40–8.49</th>
<th>8.49–11.06</th>
<th>&gt;11.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-clock</td>
<td>(vehicle-hour)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (north)</td>
<td>0</td>
<td>0.001087</td>
<td>0.000842</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \pi/8 )</td>
<td>0.018862</td>
<td>0.007461</td>
<td>0.013302</td>
<td>0.002029</td>
<td>0.000114</td>
</tr>
<tr>
<td>( 2\pi/8 )</td>
<td>0.031752</td>
<td>0.094264</td>
<td>0.120749</td>
<td>0.071519</td>
<td>0.01879</td>
</tr>
<tr>
<td>( 3\pi/8 )</td>
<td>0.007061</td>
<td>0.050743</td>
<td>0.085233</td>
<td>0.038261</td>
<td>0.012153</td>
</tr>
<tr>
<td>( 4\pi/8 )</td>
<td>0.002639</td>
<td>0.023677</td>
<td>0.020374</td>
<td>0.004624</td>
<td>0.000883</td>
</tr>
<tr>
<td>( 5\pi/8 )</td>
<td>0</td>
<td>0.004702</td>
<td>0.004019</td>
<td>0.001046</td>
<td>0</td>
</tr>
<tr>
<td>( 6\pi/8 )</td>
<td>0</td>
<td>0 0.000291</td>
<td>0.000428</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 7\pi/8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 8\pi/8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 9\pi/8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 10\pi/8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 11\pi/8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 12\pi/8 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 13\pi/8 )</td>
<td>0</td>
<td>0.001401</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 14\pi/8 )</td>
<td>0.001069</td>
<td>0.009813</td>
<td>0.000186</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( 15\pi/8 )</td>
<td>0</td>
<td>0.005581</td>
<td>0.000294</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Fig. 9.** Pareto frontier from the GA-based algorithm.
5. Numerical example

The numerical experiment was carried out on a stretch of El Camino Real in the San Francisco Bay Area of California between Crystal Springs Rd. and 5th Ave. Fig. 8 is the cell representation of the arterial. The speed limit on the major street is 35 mph or 50 feet per second while 25 mph or 36 feet per second on the side streets. Because the traffic dynamics is modeled second by second, the cell length for the major and side streets is 50 and 36 feet respectively.

Traffic demand data were collected from loop detectors for peak hours in duration of 10 working days in July 2008. The data are reported in Zhang et al. (2010). The population distribution beside the arterial was assumed to be the following:

\[ p(x, y, z) = 3 \cdot 10^{-8} \cdot (100 - z) \cdot (30000 - |x|) \]

where \( z \leq 100 \), and \(-30000 \leq x \leq 30000\). In our coordinate system, the \( y \) axis is located along the arterial centerline. The above implies that the population density is less for a location farther away from the arterial and the ground level, i.e., with a larger \( x \) and \( z \).

To facilitate the presentation, we considered one pollutant only in this example. Table 2 presents the emission factors of carbon monoxide (CO) under different driving modes (Environmental Protection Agency, 2002).

Using the 30-year historical wind data from USDA, we specified the wind scenarios. Table 3 reports the probabilities of occurrence for all the scenarios we considered. The wind speed is categorized into five bins while the wind direction is captured every \( \pi/8 \). In total, there are 80 scenarios, but only 37 of them have positive probability of occurrence.

The simulation-based GA was applied to solve the bi-objective signal timing optimization model. The resulting Pareto frontier is shown in Fig. 9 where one point represents a particular signal timing plan. For example, Plan A minimizes the mean excess exposure and Plan B minimizes the total system delay. The plans in between are all other Pareto optimal solutions, not dominated by any other timing plan. It can be observed from Fig. 9 that the mean excess exposure resulting from these timing plans varies from 30,350 to 15,867 person-grams, a 48% reduction. Correspondingly, the total system delay changes from 657 to 1269 vehicle-hours, a 93% increase. The frontier presents the tradeoff between congestion and emissions, allowing decision makers to learn about the problem before committing to a final decision of an optimal timing plan.

As an example, Table 4 presents the optimized timing Plans A and B. Plan A minimizes the mean excess exposure, while Plan B minimizes the system delay during the modeling horizon. To demonstrate the impacts of these two substantially different timing plans, Figs. 10 and 11 present their corresponding emissions exposure contour maps. The former is for the average emissions exposure across all the possible scenarios at different height levels while the latter focuses on the 20% high-consequence scenarios. Although the contour maps for Plans A and B share similar shapes, the emissions exposure under Plan B is larger than that of Plan A.

The difference is apparent around both 5th Avenue and 4th Avenue (the bottom two intersections in the contour maps) in Fig. 10c and d, as well as Fig. 11c and d. From Table 4, it can be seen that Plan B allocates majority of the green time to the main streets, thereby resulting in a smaller system delay. In contrast, Plan A prescribes less green time to the main streets, which inevitably yields a much larger system delay, but interestingly much less emission impact. Due to that the CO emissions exposure across all the possible scenarios at different height levels while the latter focuses on the 20% high-consequence scenarios. Although the contour maps for Plans A and B share similar shapes, the emissions exposure under Plan B is larger than that of Plan A.

The difference is apparent around both 5th Avenue and 4th Avenue (the bottom two intersections in the contour maps) in Fig. 10c and d, as well as Fig. 11c and d. From Table 4, it can be seen that Plan B allocates majority of the green time to the sync phases, thereby resulting in a smaller system delay. In contrast, Plan A prescribes less green time to the main streets, which inevitably yields a much larger system delay, but interestingly much less emission impact. Due to that the CO emissions from idle vehicles are very small compared to those in the other travel modes, especially acceleration and cruise, an exposure-reducing plan will tend to maintain vehicles at the idle mode as often as possible. Since the modeling horizon is fixed, and vehicles on the streets do not need to be cleared, it is not surprising to see that Plan A allocates small amount of green time to the main street, making the arterial congested and more vehicles idle. Certainly, practitioner is unlikely to implement such an extreme signal plan in the field. Table 4 also presents an intermediate Pareto optimal plan, which appears much similar to Plan B, but reduces the emission exposure by 8% at the price of increasing system delay by 16%.

<table>
<thead>
<tr>
<th>Timing plans</th>
<th>Cycle length</th>
<th>Phase sequence</th>
<th>Offset</th>
<th>P1</th>
<th>P2*</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6*</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Avenue</td>
<td>118</td>
<td>(0, 0, 0)</td>
<td>0</td>
<td>9</td>
<td>90</td>
<td>8</td>
<td>11</td>
<td>12</td>
<td>87</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>4th Avenue</td>
<td>118</td>
<td>(0, 1, 0)</td>
<td>41</td>
<td>28</td>
<td>33</td>
<td>32</td>
<td>25</td>
<td>9</td>
<td>52</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>3rd Avenue</td>
<td>118</td>
<td>(0, 0, 1)</td>
<td>15</td>
<td>25</td>
<td>46</td>
<td>9</td>
<td>38</td>
<td>18</td>
<td>53</td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td>2nd Avenue</td>
<td>118</td>
<td>(0, -, -, -)</td>
<td>63</td>
<td>8</td>
<td>8</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crystal Springs Rd.</td>
<td>118</td>
<td>(-, 1, -)</td>
<td>63</td>
<td>-</td>
<td>30</td>
<td>-</td>
<td>88</td>
<td>22</td>
<td>8</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Plan B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Avenue</td>
<td>104</td>
<td>(0, 0, 1, 1)</td>
<td>0</td>
<td>10</td>
<td>73</td>
<td>8</td>
<td>13</td>
<td>10</td>
<td>73</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>4th Avenue</td>
<td>104</td>
<td>(0, 1, 0, 0)</td>
<td>104</td>
<td>10</td>
<td>65</td>
<td>21</td>
<td>8</td>
<td>9</td>
<td>66</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>3rd Avenue</td>
<td>104</td>
<td>(1, 0, 1, 0)</td>
<td>12</td>
<td>9</td>
<td>75</td>
<td>8</td>
<td>12</td>
<td>21</td>
<td>63</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>2nd Avenue</td>
<td>104</td>
<td>(0, -, -, -)</td>
<td>52</td>
<td>8</td>
<td>88</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crystal Springs Rd.</td>
<td>104</td>
<td>(-, -, 0, -)</td>
<td>7</td>
<td>-</td>
<td>78</td>
<td>-</td>
<td>26</td>
<td>10</td>
<td>68</td>
<td>26</td>
<td>-</td>
</tr>
<tr>
<td>Plan C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Avenue</td>
<td>96</td>
<td>(1, 0, 0, 1)</td>
<td>0</td>
<td>8</td>
<td>49</td>
<td>30</td>
<td>9</td>
<td>10</td>
<td>47</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>4th Avenue</td>
<td>96</td>
<td>(0, 0, 0, 0)</td>
<td>75</td>
<td>16</td>
<td>45</td>
<td>27</td>
<td>8</td>
<td>12</td>
<td>49</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>3rd Avenue</td>
<td>96</td>
<td>(0, 0, 1, 1)</td>
<td>66</td>
<td>10</td>
<td>56</td>
<td>18</td>
<td>12</td>
<td>8</td>
<td>58</td>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td>2nd Avenue</td>
<td>96</td>
<td>(0, -, -, -)</td>
<td>65</td>
<td>10</td>
<td>66</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crystal Springs Rd.</td>
<td>96</td>
<td>(-, 0, -)</td>
<td>78</td>
<td>-</td>
<td>71</td>
<td>-</td>
<td>25</td>
<td>9</td>
<td>62</td>
<td>25</td>
<td>-</td>
</tr>
</tbody>
</table>

* Coordinated phases: phase 2 for the northbound; phase 6 for the southbound.
Fig. 10. Average emissions exposure across all scenarios.
Fig. 11. Average emission exposure across high-consequence scenarios.
6. Conclusion

This paper has formulated a bi-objective traffic signal optimization model to make an explicit tradeoff between delays and roadside air pollution concentrations. The modal emission approach is integrated with the cell transmission model to compute the pollutant emission rates at each time step. By capturing the impact of spatially and temporally varying traffic state on emission, the approach provides more accurate emission estimation. A cell-based Gaussian plume dispersion model is encapsulated to capture the pollutant dispersion process to estimate the human emissions exposure. A risk measure, namely, the mean excess exposure, is then defined to represent the mean of the human emissions exposures against high-consequence wind scenarios. The signal timing optimization optimizes the cycle length, offsets, green splits and phase sequences to minimize the total system delay and the mean excess exposure simultaneously.

A simulation-based GA is employed to solve the bi-level mixed-integer problem. Both the model and the solution algorithm are demonstrated in a numerical example using field data from a selected real-world arterial network. A set of Pareto optimal signal timing plans are generated that form an efficient frontier. The frontier exhibits an obvious tradeoff between these two objective functions, providing a foundation for a sound decision making.

Lastly, we note that this paper assumed that all the vehicles are passenger cars. If heavy vehicles, like trucks, are major concern, a multiclass cell transmission model (e.g., Wong and Wong, 2002; Tuierprasert and Aswakul, 2010) should be used to represent the traffic dynamics, and correspondingly the emission factors for heavy vehicles can be used to compute their emissions. The proposed modeling framework can be extended to accommodate such vehicle heterogeneity without difficulty. Future research may also be carried out to validate the proposed model with real-world congestion and emission data.

Acknowledgements

The authors thank two anonymous reviewers for their helpful comments. This research was partly funded by National Science Foundation (CNS-0931969) and National Natural Science Foundation of China (71228101). The first author was also supported by the Humanities and Social Science Research Project, Ministry of Education (11YJCZH242). The authors also wish to thank Meng Li at California PATH, University of California at Berkeley for providing the data used in the numerical example.

References


Lin, J., Ge, Y.E., 2006. Impacts of traffic heterogeneity on roadside air pollution concentration. Transportation Research Part D 11, 166–170.


