Transportation Research Part B 46 (2012) 156–174

Contents lists available at SciVerse ScienceDirect

Transportation Research Part B

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

A stochastic model of traffic flow: Theoretical foundations

Saif Eddin Jabari, Henry X. Liu *

Department of Civil Engineering, University of Minnesota, 500 Pillsbury Drive S.E., Minneapolis, MN 55455, United States

ARTICLE INFO

Article history:
Received 5 June 2011
Received in revised form 11 September 2011
Accepted 11 September 2011

Keywords:
Stochastic traffic flow
Random time headways
Counting processes
Cell transmission
Fluid limit

ABSTRACT

In a variety of applications of traffic flow, including traffic simulation, real-time estimation and prediction, one requires a probabilistic model of traffic flow. The usual approach to constructing such models involves the addition of random noise terms to deterministic equations, which could lead to negative traffic densities and mean dynamics that are inconsistent with the original deterministic dynamics. This paper offers a new stochastic model of traffic flow that addresses these issues. The source of randomness in the proposed model is the uncertainty inherent in driver gap choice, which is represented by random state dependent vehicle time headways. A wide range of time headway distributions is allowed. From the random time headways, counting processes are defined, which represent cumulative flows across cell boundaries in a discrete space and continuous time conservation framework. We show that our construction implicitly ensures non-negativity of traffic densities and that the fluid limit of the stochastic model is consistent with cell transmission model (CTM) based deterministic dynamics.

1. Introduction

A variety of traffic management applications today require probabilistic models of traffic flow. These include traffic simulation, estimation of traffic conditions along freeways and signalized arterials when measurement data is limited (whether in real-time or not), and applications that involve short-term traffic prediction such as online (adaptive) traffic signal control. An issue that has received little attention in the literature pertaining to the stochastic modeling of traffic flow is that of the "physical relevance" of the sample paths of stochastic processes. Any stochastic model of traffic flow, in some *mean dynamic sense*, should describe queue build-up and dissipation in a manner that is consistent with well established (deterministic) traffic flow principles, such as predictions of the model of Lighthill, Whitham, and Richards (the LWR model) (Lighthill and Whitham, 1955; Richards, 1956). In fact, traffic flow variables in the LWR model are defined as *averages* (e.g., in (Lighthill and Whitham, 1955) traffic density is defined as "the average number of vehicles on the slice of road"). From the kinetic view of traffic flow Prigogine and Herman (1971), the LWR model arises as a temporal mean dynamic. In addition, such stochastic dynamics should implicitly ensure the non-negativity of traffic densities because ad hoc treatments of negative traffic densities may reinforce the issue of inconsistency with deterministic traffic flow models.

In the literature, the usual line of attack when constructing stochastic models of traffic flow is to select a deterministic traffic flow model, which usually consists of a conservation equation and a fundamental relationship, and then add "noise" terms to the model. These include the models of Gazis and Knapp (1971) and Szeto and Gazis (1972) for purposes of traffic state estimation and control (also later in Gazis and Liu (2003)), where noise is added to the conservation equation (and the measurement equation). This is also the case in Wang and Papageorgiou (2005) and Wang et al. (2007), where noise is added to the dynamic speed equation and to a speed-density relationship (their model of traffic state includes both a conservation

* Corresponding author.
E-mail addresses: jabar005@umn.edu (S.E. Jabari), henryliu@umn.edu (H.X. Liu).

0191-2615/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.trb.2011.09.006

of flow equation and a speed evolution equation). Other recent examples include Boel and Mihaylova (2006) and Sumalee et al. (2011), which constitute stochastic extensions of the cell transmission model (Daganzo, 1994; Daganzo, 1995b).

In such settings, and particularly when the deterministic dynamics are nonlinear, one cannot guarantee that the resulting mean dynamics are consistent with the original deterministic dynamics. To illustrate, let $\bar{z}(t)$ denote a dynamic variable of interest (i.e., one that varies with time, denoted by t) and suppose it evolves according to:

$$\bar{z}(t+\Delta t) = g(\bar{z}(t), \Delta t), \tag{1}$$

a deterministic dynamic. Let $\zeta(t)$ denote white noise. A stochastic dynamic, written as:

$$z(t+\Delta t) = g(z(t), \Delta t) + \zeta(t+\Delta t) \tag{2}$$

suffers two problems. First, if $g(\cdot,\cdot)$ is non-linear (a typical feature of traffic flow models), then it is easy to see that, in general, $\mathbb{E}z(t+\Delta t) \neq \bar{z}(t+\Delta t)$,[1] despite the fact that $\mathbb{E}\zeta(t) = 0$ for each $t \geqslant 0$. The second problem is that arbitrarily adding the white noise term, $\zeta(\cdot)$, results in negative sample paths. One may suggest, to overcome this problem, that the noise terms can simply be replaced with random variables that only assume positive values (e.g., truncated noise terms); unfortunately, such random variables cannot have zero mean (unless they are identically 0) and the first problem is reinforced. To the best of the authors' knowledge, this problem has not been identified in the literature of stochastic traffic flow models and has yet to be addressed in the theory of stochastic conservation laws (see for example Holden and Risebro (1997) for a brief discussion and numerical illustration). A notable exception, in the traffic flow literature, is the Markov compartment model (MARCOM) (Davis and Kang, 1994; Kang, 1995), which models traffic flows as state-dependent Poisson processes. While the issue of non-negativity of sample paths does not arise in MARCOM, the resulting mean dynamics generally do not conform to LWR theory as this was not an objective in their study.

To overcome these problems, this paper proposes a new stochastic model of traffic flow that operates in discrete space (i.e., we break road sections into small cells) and continuous time. The use of continuous time is a consequence of random time headways taking their values in $\mathbb{R}^+$; it is also motivated by our desire to apply the proposed model to actuated traffic signals, where signal phases may vary in length. The use of discrete space is analytical tractability. The proposed dynamics can be thought of as probabilistic versions of the cell transmission model (or, more generally, Godunov scheme based dynamics), but unlike previous literature that generalizes the cell transmission model (CTM), our concern is that the CTM arise as a mean dynamic. The source of randomness in the proposed model is the uncertainty inherent in driver gap choice, which is represented by random vehicle time headways. Generally speaking, adding noise may be appropriate for systems where uncertainty is due to exogenous stochastic forcing, but this is not the case when considering conservation of traffic on a road without sources or sinks; here, a dominant source of uncertainty is driver behavior. To achieve consistency with the CTM, random elements in the model depend on traffic state; namely, the traffic density. In the proposed model, when traffic densities becomes zero, the random variables describing cumulative traffic flow across cell boundaries will degenerate (i.e., their variance becomes zero). This assumption has physical relevance. Take flow rates across a certain boundary for instance: given the traffic density is equal to zero, one knows *with certainty* that the flow rate is zero as there are no vehicles to cross the boundary; also, when traffic is at jam density, one knows with certainty that the flow rate is zero as well. In general, uncertainty in traffic state increases with increasing vehicle interactions (e.g., lane-changes and deceleration behind slower vehicles); thus the need to incorporate state dependence in traffic variables.

In the proposed model, we do not suppose any particular headway distribution; we only assume finite variances; the proposed model can, hence, accommodate a variety of headway distributions, which is desirable from an application standpoint. We also allow various flux-density relationships, but shall restrict our analysis to concave relationships and our examples to triangular relationships. To show that our model is consistent with first-order traffic flow theory, we derive a (deterministic) *fluid limit* of the proposed stochastic model, which we show to be consistent with the CTM. Estimation of system states and parameters will not be covered in the present paper, but shall be treated in a sequel along with real world examples.

This paper is organized as follows: in Section 2, we give a brief background on deterministic dynamics and how they arise as long-run averages of general kinetic models of traffic flow. We then proceed to Section 3, the main section in this paper, in which we discuss our proposed approach for modeling vehicle time headways, how this is used to construct our proposed stochastic model, ensuring non-negativity of traffic densities, and a formal derivation of the fluid limit of our model. Throughout Section 3, we give small examples to both illustrate the generality of the model and attempt to help hone the intuition behind the mathematics. In Section 4, two numerical examples are presented to demonstrate the consistency with the CTM and the evolution of probability densities of time headways in space and time as a result of the propagation of a traffic disturbance. Section 5 concludes the paper and discusses, briefly, future research directions.

2. Background

2.1. The LWR model: properties and solution

To set the stage for the remainder of this paper, we give a brief review of the Lighthill–Whitham and Richards (LWR) model of macroscopic traffic flow (Lighthill and Whitham, 1955; Richards, 1956) and its most popular solution techniques,

[1] $\mathbb{E}$ denotes the *expectation* operator.

which due to their recursive, forward in time, nature have come to constitute traffic flow models in their own right. These latter models shall represent the mean dynamics of our proposed model.

Consider a long road without intermediate sources or sinks with a large number of vehicles N on the road that change their locations over time, such that the *average* number of vehicles at any time instance, t, and at any point in space, x, can be written as $Np(x,t) = \bar{\rho}(x,t)$, where $p(x,t)$ denotes concentration (or fraction) of vehicles at point x at time t and the interpretation of $\bar{\rho}(x,t)$ as the mean traffic density follows. Likewise, let $\bar{v}(x,t)$ denote the average vehicle speed at location x at time t; the average flux (flow rate) is then expressed as $\bar{\rho}(x,t)\bar{v}(x,t)$. The conservation of vehicles in any arbitrary stretch $[x_1,x_2]$ of the road, over an arbitrary time interval $[t_1,t_2]$, is written as:

$$\int_{x_1}^{x_2} \bar{\rho}(x,t_2)dx - \int_{x_1}^{x_2} \bar{\rho}(x,t_1)dx = \int_{t_1}^{t_2} \bar{\rho}(x_1,t)\bar{v}(x_1,t)dt - \int_{t_1}^{t_2} \bar{\rho}(x_2,t)\bar{v}(x_2,t)dt \tag{3}$$

The differential form of the conservation equation (the LWR model) is derived from (3) under the additional assumption that $\bar{\rho}(x,t)$ and $\bar{v}(x,t)$ are differentiable in both x and t, and is written as (see for example (LeVeque, 1992) for derivation):

$$\frac{\partial}{\partial t}\bar{\rho}(x,t) + \frac{\partial \bar{\rho}(x,t)\bar{v}(x,t)}{\partial x} = 0 \tag{4}$$

The average flux $\bar{\rho}(x,t)\bar{v}(x,t)$ is usually written as an equilibrium flux function of density $Q_e(\bar{\rho}(x,t))$ that is independent of both space and time, describing the stationary (long run) relationship between density and flux. In most cases, $Q_e(\cdot)$ is simply an empirical relationship; in fact the subscript "e" has been used to denote "empirical" rather than "equilibrium" (Helbing, 2001); the two interpretations are not mutually exclusive. While various shapes have been suggested for $Q_e(\cdot)$, in general, it can be described as a concave function of density; that is, the flux increases as the traffic density increases in under-congested traffic conditions, the flux then reaches a maximum value as congestion sets in, and beyond this maximum value, the flux begins to decrease as the traffic density increases signifying deterioration in traffic conditions due to over-saturation. The boundary condition is usually given as a smooth function that defines the density distribution at some initial time (usually $t = 0$) over space, $\bar{\rho}(x,0) = \bar{\rho}_0(x)$, and the spatio-temporal evolution of traffic density is the solution of:

$$\frac{\partial \bar{\rho}(x,t)}{\partial t} + \frac{dQ_e(\bar{\rho})}{d\bar{\rho}}\frac{\partial \bar{\rho}(x,t)}{\partial x} = 0 \tag{5a}$$

$$\bar{\rho}(x,0) = \bar{\rho}_0(x) \tag{5b}$$

It is well known that (5b) is an ill-posed problem; this is primarily due to the fact that the differentiability assumption used to derive (4) does not hold in general; namely when the initial conditions are non-uniform. The best one can then hope for is a weak solution, which in this case is a solution to the integral form of the problem, (3). Weak solutions, however, are not unique, and not all weak solutions capture the physics of traffic flow correctly. For this reason, one imposes additional conditions that pick out the "correct" weak solution. These conditions are usually either entropy conditions or simpler geometric conditions that ensure that (i) shock waves propagate at the correct speeds determined by the traffic conditions on either side of the shock, and (ii) when jams are dissipating, the characteristic lines spread (rather than propagate as a discontinuity); the latter is referred to as a rarefaction fan. These two conditions are illustrated in Fig. 1. See (Osher, 1984; LeVeque, 1992; Bui et al., 1992), and references therein for more details on weak solutions.

The most popular numerical schemes used to solve (5b) are based on the Godunov scheme (Godunov, 1959; Lebacque, 1996), and are sometimes referred to as cell transmission models (CTM), due to Daganzo (1994) who first introduced the nomenclature. In Daganzo's CTM, triangular and trapezoidal flux-density relationships are used and a fixed ratio of cell

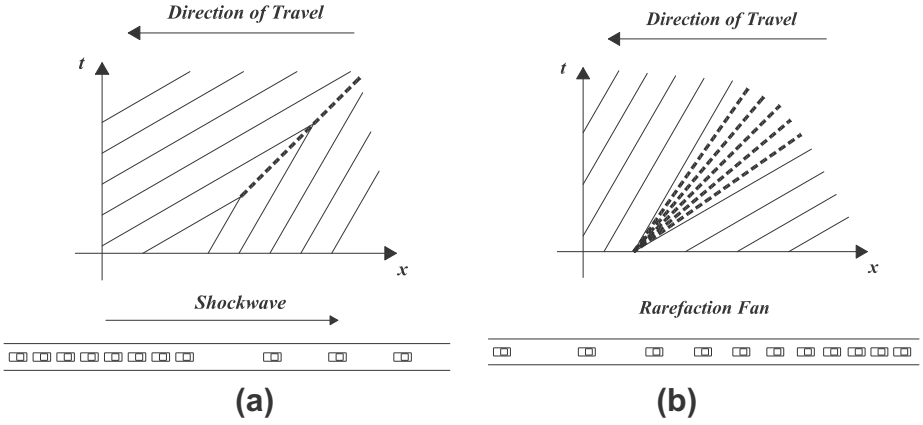

Fig. 1. (a) Shockwave profile, (b) rarefaction fan.

length to discrete time interval length is used based on free-flow speed ($l_x/\Delta t = v_f$), where l_x denotes the length of cell x and Δt denotes the discrete time interval length. In this paper, a CTM shall refer to a broader Godunov scheme based model that allows for a variety of concave relationships and any ratio that honors the condition: $l_x/\Delta t \leqslant v_f$. The CTM offers an approximate solution to the LWR model, where densities remain constant over discrete time intervals of length Δt and constitute averages over discrete spatial intervals (cells). Let $\bar{\rho}(x,t)$ denote the mean traffic density in cell x at time t and denote by $\lambda([\bar{\rho}(x,t)\ \bar{\rho}(x+1,t)]^T)$ the mean flux at the downstream boundary of cell x; then the numerical solution is obtained by recursively computing:

$$\bar{\rho}(x,t+\Delta t) = \bar{\rho}(x,t) + \frac{\Delta t}{l_x}\left(\lambda([\bar{\rho}(x-1,t)\ \bar{\rho}(x,t)]^T) - \lambda([\bar{\rho}(x,t)\ \bar{\rho}(x+1,t)]^T)\right) \tag{6}$$

$$\lambda([\bar{\rho}(x,t)\bar{\rho}(x+1,t)]^T) = \begin{cases} \min\limits_{\bar{\rho}(x,t)\leqslant\kappa\leqslant\bar{\rho}(x+1,t)} Q_e(\kappa) & \text{if } \bar{\rho}(x,t) \leqslant \bar{\rho}(x+1,t) \\ \max\limits_{\bar{\rho}(x+1,t)\leqslant\kappa\leqslant\bar{\rho}(x,t)} Q_e(\kappa) & \text{if } \bar{\rho}(x,t) \geqslant \bar{\rho}(x+1,t) \end{cases} \tag{7}$$

In the case of a concave flux function, (7) can be written more simply as:

$$\lambda([\bar{\rho}(x,t)\ \bar{\rho}(x+1,t)]^T) = \min\{S_e(\bar{\rho}(x,t)), R_e(\bar{\rho}(x+1,t))\} \tag{8}$$

where $S_e(\cdot)$ and $R_e(\cdot)$ are, respectively, sending and receiving functions obtained from $Q_e(\cdot)$ as shown in Fig. 2 (Daganzo, 1995a).

Eq. (6) can also be written in terms of *cumulative* flows as follows:

$$\bar{\rho}(x,j\Delta t) = \bar{\rho}(x,0) + \frac{1}{l_x}\left(\sum_{k=0}^{j-1}\lambda([\bar{\rho}(x-1,k)\ \bar{\rho}(x,k)]^T)\Delta t - \sum_{k=0}^{j-1}\lambda([\bar{\rho}(x,k)\ \bar{\rho}(x+1,k)]^T)\Delta t\right) \tag{9}$$

and upon letting $\Delta t \to 0$, one writes the continuous time version of the conservation Eq. (6), as:

$$\bar{\rho}(x,t) = \bar{\rho}(x,0) + \frac{1}{l_x}\left(\int_0^t \lambda([\bar{\rho}(x-1,u)\ \bar{\rho}(x,u)]^T)du - \int_0^t \lambda([\bar{\rho}(x,u)\ \bar{\rho}(x+1,u)]^T)du\right) \tag{10}$$

This, combined with Eq. (8), shall constitute the desired mean dynamics of our stochastic model.

2.2. The LWR model as a long-run temporal mean dynamic

Our proposed model is based on the premise that first order macroscopic traffic equations arise as a long run temporal mean dynamic of a more general stochastic process. In this section, we illustrate this from a simplified kinetic point of view, as was previously derived in (Prigogine and Herman, 1971).

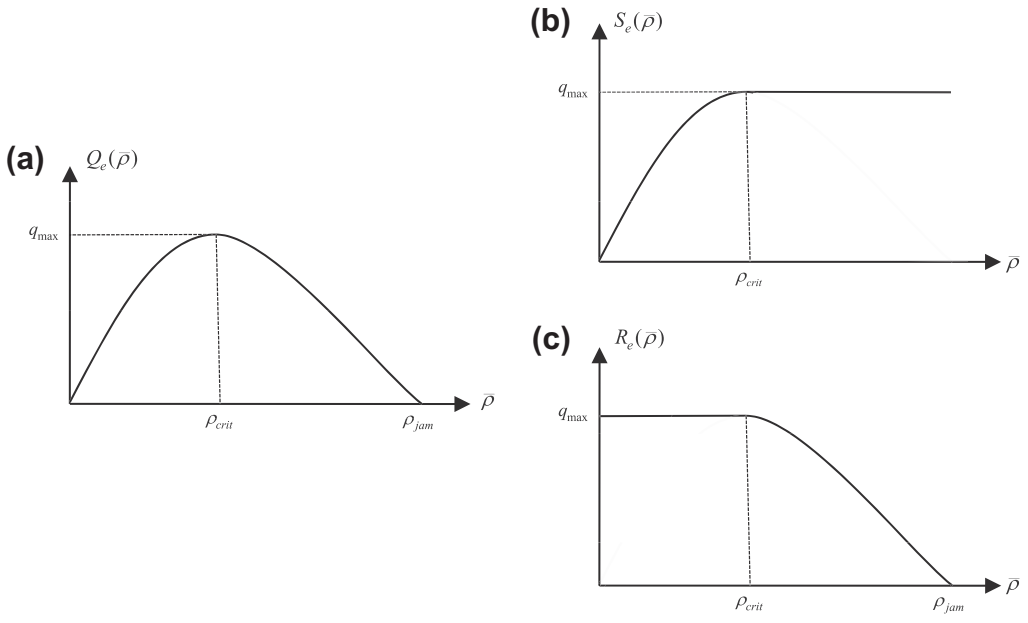

Fig. 2. (a) A typical concave flux function, (b) the sending function, (c) the receiving function.

The same setting in Section 2.1 of an infinite road with a large number of vehicles N is assumed. Let $f(x,t,v)$ denote the joint probability density of vehicles at location x at time t and traveling with speed v. The concentration $p(x,t)$ is now interpreted as a marginal probability density and the traffic density is also written as $N\,p(x,t)$. Let $\mu(\{x,v\},\{y,w\})$ denote the instantaneous transition rate from traffic state $\{x,v\}$ to traffic state $\{y,w\}$, where the first coordinate denotes location and the second coordinate denotes speed. A master equation (a.k.a. Kolmogorov backward equation) defining the evolution of the probability density is written as:

$$\frac{d}{dt}f(x,t,v)=\int_0^\infty\int_{-\infty}^\infty \mu(\{y,w\},\{x,v\})f(y,t,w)dy\,dw-\int_0^\infty\int_{-\infty}^\infty \mu(\{x,v\},\{y,w\})f(x,t,v)dy\,dw \tag{11}$$

The first such probabilistic formulation appears in (Prigogine and Andrews, 1960), in which heuristic arguments are made in order to establish vanishing relationships that define "relaxtion" (or transience) towards a desired speed distribution and vehicle interaction terms as fast vehicles overtake slow ones, which also vanish in the limit (hereafter, this model shall be referred to as Prigogine's model). The sum of the relaxation and interaction terms would define a special case of the right hand-side of (11). Paveri-Fontana (1975) proposed improvements to the relaxation and interaction terms in Prigogine's model that overcome undesirable behavior; nonetheless, his adjusted model can still be considered as a version of (11) with desired speed as an additional implicit parameter.

Returning to the master equation; assume that in the long run (as $t\to\infty$), the probability density becomes stationary; that is, in the limit

$$\frac{d}{dt}f(x,t,v)=0 \tag{12}$$

Expanding the total derivative in (12), we get:

$$\frac{\partial}{\partial t}f(x,t,v)+\frac{dx}{dt}\frac{\partial}{\partial x}f(x,t,v)+\frac{dv}{dt}\frac{\partial}{\partial v}f(x,t,v)=0 \tag{13}$$

and upon integrating over v and assuming that acceleration $dv/dt\equiv a$ is independent of speed, we get:

$$\frac{\partial}{\partial t}p(x,t)+\frac{\partial}{\partial x}p(x,t)\int_0^\infty vf(v|x,t)dv=0 \tag{14}$$

The integral in (14) defines the conditional mean speed $\mathbb{E}(v|x,t)\equiv\bar{v}(x,t)$. Upon multiplying both sides of (14) by N, the LWR model is recovered:

$$\frac{\partial}{\partial t}\bar{\rho}(x,t)+\frac{\partial}{\partial x}(\bar{\rho}(x,t)\bar{v}(x,t))=0 \tag{15}$$

Unfortunately, the precise form of the transition function $\mu(\cdot,\cdot)$ is unknown and assumed forms are usually established on heuristic grounds and, in some cases, result in unreasonable traffic flow dynamics. However, as long as $\mu(\cdot,\cdot)$ satisfies the conditions required for convergence to hold (e.g., that $\mu(\cdot,\cdot)$ defines a contraction semi-group), (15) will arise as a mean dynamic.[2]

In summary, the LWR model can be described as a relationship between average quantities that arises as a temporal mean dynamic and the stochastic model we propose below honors this view.

3. The proposed stochastic traffic flow model

Take the simple setting of a homogeneous roadway without intermediate sources or sinks, divide it into cells, and let $\mathcal{C}$ denote the set of cell indices. The conservation equation is written as:

$$\rho(x,t)=\rho(x,0)+\frac{1}{l_x}(Q(x-1,t)-Q(x,t)) \tag{16}$$

where l_x is the length of cell $x\in\mathcal{C}$, $\rho(x,t)$ is the traffic density in cell x at time t, a random variable, $Q(x,t)$ is the cumulative number of vehicles that have departed cell x by time t. We shall treat $Q(x,\cdot)$ as a (random) counting process with $Q(x,0)\equiv 0$ for all x and assume, for each $x\in\mathcal{C}$, that $\rho(x,0)$ is a discrete random variable that is independent of $Q(x,t)$ (thus, $\rho(x,t)$) for all $t>0$.

Our primary concern now is to characterize the $Q(\cdot,\cdot)$ processes. To do so, first let $A_k(x)$ denote the time at which the kth vehicle departs cell x; to minimize clutter in our notation, we shall also say that vehicle i has crossed location x (the downstream boundary of cell x). Let $\{X_i(x)\}_{i\geqslant 1}$ denote the sequence of time headways measured at location x defined as $X_i(x)=A_i(x)-A_{i-1}(x)$, where $A_0(x)\equiv 0$. It then also follows that:

[2] A more general integral version of (15) can also be derived as a limiting dynamic.

$$A_k(x) = \sum_{i=1}^{k} X_i(x) \tag{17}$$

We shall assume that the time vehicle i crosses x is *conditionally* independent of all past crossing times $\{A_k: 0 \leqslant k \leqslant i-2\}$ given the crossing time of vehicle $i-1$. That it is independent of all future arrivals is assumed in order to maintain causality in our model. In essence, these assumptions mean that the sequence $\{A_k(x)\}$ is Markovian. $Q(x,\cdot)$ is simply the reciprocal of the crossing process $A_k(x)$, defined for each $t \geqslant 0$ by:

$$Q(x,t) = \max\{k : A_k(x) \leqslant t\} \tag{18}$$

Eqs. (16)–(18) constitute the proposed model. We should emphasize that our model can accommodate a variety of time headway distributions, which are allowed to depend on traffic state. We next give a more detailed discussion of vehicle time headways as they constitute a main building block for our proposed model and the subsequent derivation of the fluid limit.

3.1. Time headways

Time headways measured at location x depend on traffic conditions in the immediate upstream and immediate downstream of x; let $y(x,t) = [\rho(x,t)\ \rho(x+1,t)]^T$ denote such a vector of "relevant" traffic conditions at location x.[3] We shall think of the random variables $\{X_i(x)\}$ as having a general continuously differentiable conditional distribution function $G(\cdot|y(x,t))$ with mean $\bar{h}(y(x,t),\beta) \equiv 1/\lambda(y(x,t),\beta)$, i.e., the reciprocal of a concave flux function, or fundamental diagram (see Eq. (8) and Fig. 2), and where β is a vector of coefficients of the flux function (e.g., β could include the critical density, the capacity, and the jam density). The specific form of the distribution function and the flux function (i.e., the parameters β) are to be determined empirically and may vary from one site to another; below, we give examples highlighting some of the popular forms that appear in the literature.

Temporally, $X_i(x)$ depends on the traffic state at the beginning of the ith interval; that is, its conditional expectation is written more precisely as: $\bar{h}(y(x,A_{i-1}(x)),\beta)$. Furthermore, define the variance of time headways by[4] $\sigma^2(y,\beta) = (1_{\{y_1>0\}}\bar{c}\cdot\bar{h}(y,\beta))^2$, where $y_1 \equiv \rho(x,t)$, $\bar{c}$ is a coefficient of variation of the time headways. We shall assume, without loss of generality that $\bar{c}$ is a constant that is to be determined empirically. The role of the indicator, $1_{\{y_1>0\}}$, is to ensure that the random variable degenerates when cell x is empty; this is the main idea in ensuring non-negativity of the sample path, to be discussed below.

To ensure that the mean and the variance are bounded, we shall first fix a small number $\varepsilon > 0$ and then define the range of $\bar{h}(\cdot,\beta)$ by $[\varepsilon,\rho_{jam}] \times [0,\rho_{jam}-\varepsilon] \subset \mathbb{R}^2$, where ρ_{jam} denotes the jam density; that is, we have bounded the flux away from zero, but we can get arbitrarily close to zero flux with appropriate choice of ε; this, in turn, results in headways that are bounded from above. In practice, the boundedness of the individual time headways follows from the boundedness of the analysis time horizon; i.e., $t \in [0,U]$, where $U < \infty$. This is a mild restriction and will be relaxed later (namely, for the counting processes).

The construction we propose can accommodate a variety of headway distributions as special cases. This is illustrated in the following two examples.

3.1.1. Example 1: exponential headways

Let $G(\cdot|y) \equiv G(\cdot)$ be independent of traffic state (or simply take the same form for all possible values of y) and define the mean headway as: $\bar{h}(y,\beta) \equiv 1/\lambda$, where $\lambda > 0$ is a known scalar. Then, writing the distribution function as:

$$G(a) = 1 - e^{-\lambda a}, \quad a \geqslant 0, \tag{19}$$

one obtains a model with independent and identically distributed (IID) exponential headways with $\bar{c} = 1$, which is appropriate for modeling free-flow traffic conditions. (Note that $1_{\{\cdot\}}$ was omitted here because of state-independence).

3.1.2. Example 2: a mixed headway distribution

In this example, we show that the M4 headway model of (Cowan, 1975) can also be accommodated as a special case. Let $B(\cdot)$ be an arbitrary distribution function defined for a positive random variable with finite second moment that represents a "tracking" component of time headway. Define $\theta \in [0,1]$, let $\lambda > 0$ be as defined in example 1, and again assume independence of traffic state. Then write:

$$\bar{h}(y,\beta) = \int_0^\infty u dB(u) + \frac{1-\theta}{\lambda} \tag{20}$$

which is a constant given $B(\cdot)$. Writing the distribution function as:

$$G(a) = \theta B(a) + (1-\theta)\int_0^a B(a-u)\lambda e^{-\lambda u} du, \quad a \geqslant 0 \tag{21}$$

[3] The vector $y(\cdot,\cdot)$ may be generalized to include further information related to traffic conditions, but for our purposes we only consider densities upstream and downstream location x.

[4] It is to be understood that $y \equiv y(x,t)$; we shall often suppress (x,t) for notation clarity.

one obtains the M4 mixed headway distribution of Cowan (1975). Here, it is easy to check that $\bar{c}$ is also a constant, which is derived from $B(\cdot)$ and Eq. (20). Cowan's M2 and M3 models are special cases of M4 and the M1 model (also a special case of M4) is the case of IID exponentials illustrated in the previous example.

It is also possible to derive similar special cases to obtain the mixture distribution of Branston (1976) and the Pearson type III based model of Hoogendoorn and Bovy (1998) (for a single vehicle type). We will now illustrate the effect of state dependence in the example below.

3.1.3. Example 3: state dependence

Suppose we have a triangular flux function, $Q_e(\rho)$, shown in Fig. 3. The sending and receiving functions are written as $S_e(\rho) = \min\{v_f\rho, q_{max}\}$ and $R_e(\rho) = \min\{q_{max}, w(\rho_{jam} - \rho)\}$, respectively. For this flux function, $\beta = [v_f\ q_{max}\ w\ \rho_{jam}]^T$; let

$$\lambda(y,\beta) = \min\{S_e(y_1), R_e(y_2)\}, \tag{22}$$

where $y = [y_1\ y_2]^T$ is the vector of densities in cells x and $x+1$, respectively (i.e., $y_1 \equiv \rho(x,t)$ and $y_2 \equiv \rho(x+1,t)$). We now have that

$$\bar{h}(y,\beta) = \frac{1}{\lambda(y,\beta)},\quad y \in (0,\rho_{jam}] \times [0,\rho_{jam}) \tag{23}$$

where for the example shown in Fig. 3, $\beta = [v_f\ q_{max}\ w\ \rho_{jam}]^T = [45mph\ 1800vph\ 15mph\ 160vpm]^T$ and, consequently, $\lambda(y,\beta) = \min\{45y_1, 1800, 15(160 - y_2)\}$ (in vehicles per hour).

For $G(\cdot|y)$, suppose the random variables $\{X_i(x)\}$ are Gamma distributed with shape parameter 2 and mean $\bar{h}(y,\beta)$ (i.e., $\bar{c} = \sqrt{1/2}$). Then

$$G(a|y) = \int_0^{a\lambda(y,\beta)} ue^{-u}du,\quad a \geqslant 0 \tag{24}$$

and we have the following probability density function:

$$f(a|y) = \frac{d}{da}G(a|y) = a\lambda^2(y,\beta)e^{-a\lambda(y,\beta)},\quad a \geqslant 0 \tag{25}$$

To illustrate the impact of state dependence, Fig. 4 is a plot of the probability density functions of the time headways measured at the downstream boundary of cell x, $f(\cdot|y)$, for three cases: (i) $y = [10\ 10]^T$ representing free-flow conditions in both upstream and downstream cells, (ii) $y = [100\ 10]^T$ representing queue discharge conditions, and (iii) $y = [100\ 100]^T$ representing congested conditions in both upstream and downstream cells.

We see in Fig. 4 that under free-flow conditions ($y = [10\ 10]^T$), larger headways are "more likely" to arise with smaller headways occurring less frequently. This is in contrast to a queue discharge situation ($y = [100\ 10]^T$), where small headways tend to occur with high frequency and larger headways are "less likely" to occur. Under congested conditions ($y = [100\ 100]^T$), we see something that falls in between the first two cases; one can think of this as a dispersed version of case (ii) as vehicles tend to encounter slower moving traffic in the downstream.

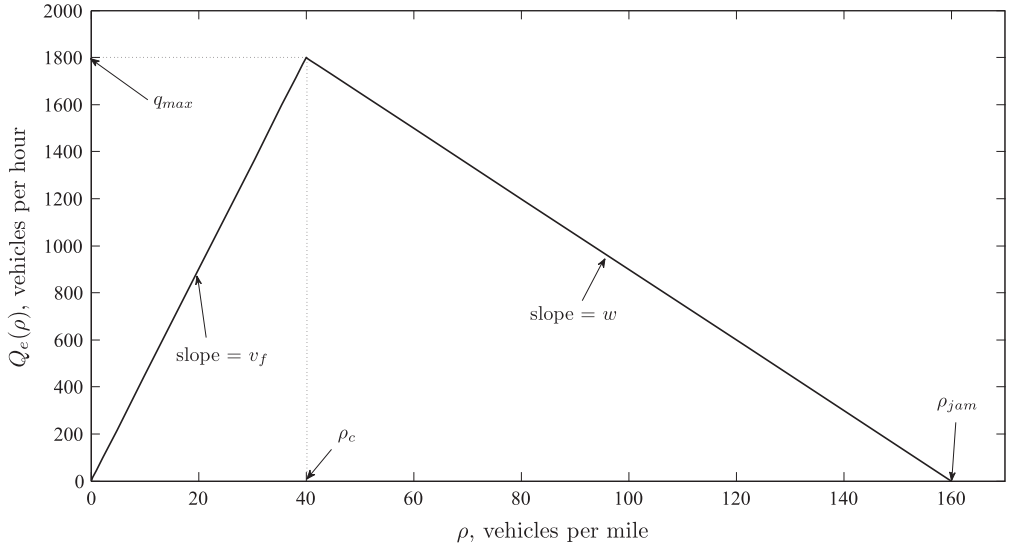

Fig. 3. Example flux function.

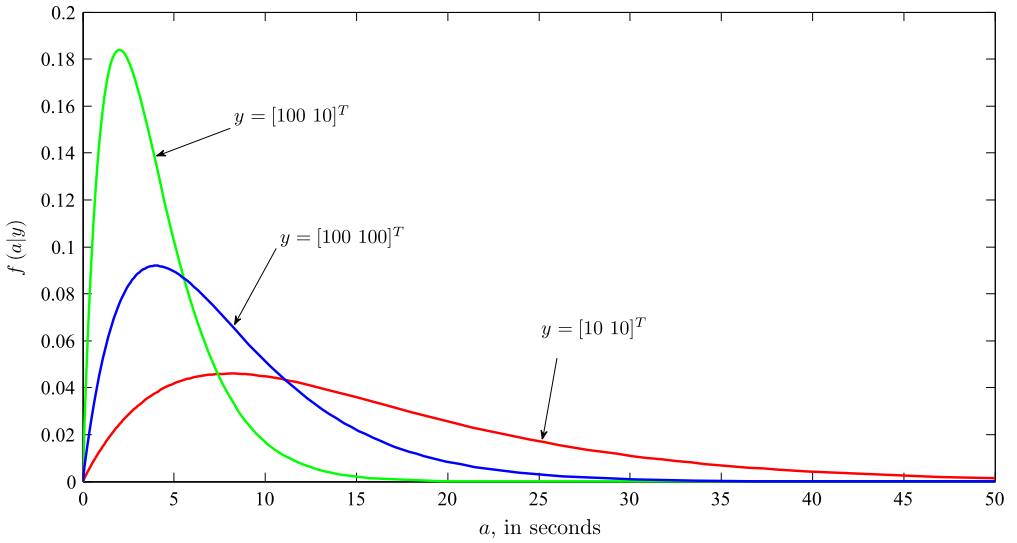

Fig. 4. Example probability density plots for state dependent headways.

3.2. Non-negativity of sample paths

The counting processes, $Q(\cdot,\cdot)$ are by definition non-negative. Negative traffic densities arise when more vehicles than available in the cell artificially depart the cell. To see that the proposed model does not result in negative traffic densities, fix $x \in \mathcal{C}$ and $t \in [0,U]$; we may suppose that $\rho(x,0)$ and $Q(x-1,t)$ are zero, since they only have positive contribution to the traffic density, $\rho(x,t)$ (see Eq. (16) above). Now, note that $l_x\rho(x,t)$ may only assume integer values since it is, essentially, the difference between counting processes. Suppose time t marks the time instance of the most recent departure from cell x and suppose this corresponds to the ith vehicle, then the vector $y(x,t) = [\rho(x,t)\ \rho(x+1,t)]^T$ constitutes the relevant information for determining $X_{i+1}(x)$. If $l_x\rho(x,t) \geqslant 1$, then regardless of which value $X_{i+1}(x)$ may assume, $\rho(x,\cdot)$ will still be non-negative at time $t + X_{i+1}(x)$. We are, thus, interested in the limiting case (when ε approaches zero) where $l_x\rho(x,t) = 0$ (i.e., $\rho(x,t) = 0$). In this case, $\min\{S_e(\rho(x,t)), R_e(\rho(x+1,t))\} = S_e(\rho(x,t)) = 0$ and, hence, $\bar{h}(y(x,t),\beta) \to \infty$. Recall, that the variance of $X_{i+1}(x)$ is $(1_{\{\rho(x,t)>0\}}\bar{c}\cdot\bar{h}(y(x,t),\beta))^2$. Then, when $\rho(x,t) = 0$, the variance is zero and at this point $X_{i+1}(x)$ is deterministic and equal to its expectation, which is infinite. This then implies that as long as $\rho(x,t) = 0$, no departures will take place and the density remains non-negative.

3.3. The (deterministic) fluid limit

The fluid limit of a stochastic process can be thought of as a long-run temporal mean of the process. Thus, to determine the fluid limit, one usually "accelerates time" and "aggregates space"[5] by monotonically increasing factors. This practice is referred to as "scaling" in the queueing theory literature. See for example (Chen and Yao, 2001 and Whitt, 2002). In this section we first state our result formally and in mathematical detail. We then present simple examples to help hone the intuition behind scaling.

Consider a sequence of stochastic processes of the form (16) indexed by the variable $n \geqslant 1$ (the *scaling factor*), such that the nth process is *accelerated* by the factor n and *aggregated* by the factor $1/n$. That is, $\rho^n(x,t) = \rho(x,nt)$, $Q^n(x,t) = Q(x,nt)$, $\rho^n(x,0)$ is a random variable pertaining to the nth process that is independent of $\rho^n(x,t)$ and $Q^n(x,t)$ for all $t > 0$, and the conservation equation for the nth process is

$$\frac{1}{n}\rho^n(x,t) = \frac{1}{n}\rho^n(x,0) + \frac{1}{nl_x}(Q^n(x-1,t) - Q(x,t)) \tag{26}$$

Let $\lambda(y^n,\beta) \equiv 1/\bar{h}(y^n,\beta)$ denote the flux function pertaining to the nth process. Specifically, $\lambda(y^n,\beta) = \lambda((1/n)[\rho(x,nt)\ \rho(x+1,nt)]^T,\beta)$. The following theorem summarizes our main result in this section, where the fluid limit in Eq. (28) coincides with the CTM, Eq. (10). Before stating the theorem, we note the following notational convention: since the main objective in this section is the determination of limiting *processes*, we use the "$\cdot$" notation in brackets to emphasize we are concerned with convergence of sample paths; when this appears in the upper limit of a sum or an integral, we use "$\bullet$" for visual clarity.

[5] Here, by "space" what is meant is the value assumed by the stochastic process and not the location x of the process.

Theorem 3.1 (*The Fluid Limit*). *Suppose, for each $x \in \mathcal{C}$, that the flux functions possess the Lipschitz property and that*

$$\frac{1}{n}\rho^n(x,0) \underset{n\to\infty}{\to} \bar{\rho}(x,0) \text{ a.s.} \tag{27}$$

where $\bar{\rho}(x,0)$ is a constant. Then, for each $x \in \mathcal{C}$, as $n \to \infty$, the process $(1/n)\rho^n(x,\cdot)$ on $[0,U]$ converges weakly (i.e., in distribution) to the deterministic process:

$$\bar{\rho}(x,\cdot) = \bar{\rho}(x,0) + \frac{1}{l_x}\int_0^{\bullet} (\lambda(\bar{y}(x-1,u),\beta) - \lambda(\bar{y}(x,u),\beta))du \tag{28}$$

where $\bar{y}(x,t) = [\bar{\rho}(x,t)\ \ \bar{\rho}(x+1,t)]^T$ is a deterministic vector.

An essential ingredient in the proof of Theorem 3.1 is convergence of the $Q^n(\cdot,\cdot)$ processes, which we shall establish first, but to do so, we will need the following two Lemmas. To simplify notation, y_i^n shall denote the vector of traffic conditions (represented by the nth process) at the time instance vehicle $i-1$ departs location x; i.e., pertaining to the ith time headway, $X_i(x)$.

Lemma 3.1. *Fix $x \in \mathcal{C}$ and let $\{X_i(x)\}_{i\geqslant 1}$, $Q^n(x,t)$, and $\bar{h}(y_i^n,\beta)$ be as defined above. Then, for each $n \geqslant 1$,*

$$\left\| \frac{1}{n}\left(Q^n(x,t) - \sum_{i=1}^{Q^n(x,t)} \frac{X_i(x)}{\bar{h}(y_i^n,\beta)} \right) \right\|_1 \underset{n\to\infty}{\to} 0 \tag{29}$$

where $\|\cdot\|_1 \equiv \mathbb{E}|\cdot|$ denotes the L_1-norm.

Proof. First note that the left hand side of (29) may be written as:

$$\begin{aligned}\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^{Q^n(x,t)}\left(1-\frac{X_i(x)}{\bar{h}(y_i^n,\beta)}\right)\right|\right) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{Q^n(x,t)}\left(1-\frac{X_i(x)}{\bar{h}(y_i^n,\beta)}\right)\middle| Q^n(x,t) \geq \sum_{i=1}^{Q^n(x,t)}\frac{X_i(x)}{\bar{h}(y_i^n,\beta)}\right)\\ &+\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{Q^n(x,t)}\left(\frac{X_i(x)}{\bar{h}(y_i^n,\beta}-1\right)\middle|\sum_{i=1}^{Q^n(x,t)}\frac{X_i(x)}{\bar{h}(y_i^n,\beta)} \geq Q^n(x,t)\right)\end{aligned} \tag{30}$$

But $Q^n(x,t) \to \infty$ as $n \to \infty$ so that the first term on the right hand side in (30) is equal to its unconditional expectation in the limit and the second term is zero. Then, in the limit, (30) becomes:

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{Q^n(x,t)}\left(1-\frac{X_i(x)}{\bar{h}(y_i^n,\beta)}\right)\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{Q^n(x,t)}\lambda(y_i^n,\beta)(\bar{h}(y_i^n,\beta)-X_i(x))\right) \tag{31}$$

Let $\mathcal{G}_i$ represent the σ-field with respect to which $\bar{h}(y_i^n,\beta)$ and $\lambda(y_i^n,\beta)$ are measurable (i.e., the history of the process up to the time of departure of vehicle $i-1$), that is, $\mathbb{E}(X_i(x)|\mathcal{G}_i) = \bar{h}(y_i^n,\beta)$. Then, expanding the expectation of the random sum, (31) can be written as:

$$\begin{aligned}\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{Q^n(x,t)}\lambda(y_i^n,\beta)(\bar{h}(y_i^n,\beta)-X_i(x))\right) &= \frac{1}{n}\sum_{k=1}^{\infty}\sum_{i=1}^{k}\mathbb{E}\left(\lambda(y_i^n,\beta)(\bar{h}(y_i^n,\beta)-X_i(x))|Q^n(x,t)=k\right)\cdot P(Q^n(x,t)=k)\\ &= \frac{1}{n}\sum_{k=1}^{\infty}\sum_{i=1}^{k}\mathbb{E}\left(\mathbb{E}\left(\lambda(y_i^n,\beta)(\bar{h}(y_i^n,\beta)-X_i(x))|\mathcal{G}_i\right)|Q^n(x,t)=k\right)\cdot P(Q^n(x,t)=k)\\ &= \frac{1}{n}\sum_{k=1}^{\infty}\sum_{i=1}^{k}\mathbb{E}\left(\lambda(y_i^n,\beta)\mathbb{E}(\bar{h}(y_i^n,\beta)-X_i(x)|\mathcal{G}_i)|Q^n(x,t)=k\right)\cdot P(Q^n(x,t)=k)\\ &= \frac{1}{n}\sum_{k=1}^{\infty}\sum_{i=1}^{k}\mathbb{E}\left(\lambda(y_i^n,\beta)(\bar{h}(y_i^n,\beta)-\bar{h}(y_i^n,\beta))|Q^n(x,t)=k\right)\cdot P(Q^n(x,t)=k)\\ &= \frac{1}{n}\sum_{k=1}^{\infty}\sum_{i=1}^{k}0\cdot P(Q^n(x,t)=k)=0\end{aligned} \tag{32}$$

Lemma 3.2. *Fix $x \in \mathcal{C}$ and let $\{X_i(x)\}_{i\geqslant 1}$, $Q^n(x,t)$, $\bar{h}(y_i^n,\beta)$, and $\lambda(y_i^n,\beta)$ be as defined above. Then, for each $t \in [0,U]$, we have*

$$\frac{1}{n}\sum_{i=1}^{Q^n(x,t)}\frac{X_i(x)}{\bar{h}(y_i^n,\beta)} \underset{n\to\infty}{\to} \int_0^t \lambda(y^n(x,u),\beta)du \text{ weakly} \tag{33}$$

Proof. Since $\lambda(y_i^n,\beta) = 1/\bar{h}(y_i^n,\beta)$, we have that:

$$\frac{1}{n}\sum_{i=1}^{Q^n(x,t)}\frac{X_i(x)}{\bar{h}(y_i^n,\beta)} = \sum_{i=1}^{Q^n(x,t)}\lambda(y_i^n,\beta)\frac{X_i(x)}{n} \tag{34}$$

Define

$$M(x) = \max_{1\leqslant i\leqslant Q^n(x,t)} X_i(x) \tag{35}$$

and note that boundedness of the random variables $\{X_i(x)\}_{i\geqslant 1}$ implies that

$$\frac{M(x)}{n} \xrightarrow[n\to\infty]{} 0 \text{ a.s.} \tag{36}$$

From Eq. (18), we have that $A_i(x)\in[0,nt]$ whenever $i\in\{1,\dots,Q(x,nt)\}$. We can then think of the set $\{1,\dots,Q(x,nt)\}$ as a set of indices on a partition of $[0,nt]$ with the set $\{X_i(x)\}_{1\leqslant i\leqslant Q(x,nt)}$ representing the elements of the partition. Now, note that the two events $A_i(x)\in[0,nt]$ and $A_i(x)/n\in[0,t]$ are equal in probability; then, letting $n\to\infty$ and noting that the flux functions are bounded and continuous, the weak limit can be interpreted as the Riemann integral:

$$\int_0^t \lambda(y^n(x,u),\beta)du \tag{37}$$

□

Corollary 3.1. *Fix $x\in\mathcal{C}$ and let $Q^n(x,t)$ and $\lambda(y^n,\beta)$ be as defined above. Then, for each $t\in[0,U]$,*

$$\frac{1}{n}Q^n(x,t) \xrightarrow[n\to\infty]{} \int_0^t \lambda(y^n(x,u),\beta)du \text{ weakly} \tag{38}$$

Proof. The result follows immediately from Lemmas 3.1 and 3.2. □

We are now ready to prove Theorem 3.1, but we first note that the flux functions possess the Lipschitz property; that is, for $a,b\in[0,\rho_{jam}]\times[0,\rho_{jam}]$, and some constant $0<K<\infty$, we have that

$$\|\lambda(a,\beta)-\lambda(b,\beta)\|_U \leqslant K\|a-b\|_U \tag{39}$$

where $\|\cdot\|_U$ is the uniform norm on the space $C([0,U],\mathbb{R}^m)$ of continuous functions from $[0,U]$ into $\mathbb{R}^m$, defined as follows: let $z(\cdot)$ denote a continuous function on $[0,U]$ that takes its values in $\mathbb{R}^m$, then

$$||z(\cdot)||_U \equiv \sup_{0\leqslant t\leqslant U}\max_{1\leqslant j\leqslant m}|z_j(t)| \tag{40}$$

That the flux function $\lambda(y(x,\cdot),\beta)$ is Lipschitz continuous follows from the Lipschitz continuity of the sending and receiving functions from a concave fundamental relationship.

Proof of Theorem 3.1. Since we are investigating the convergence of a process with jumps to a continuous process, we will need a more general notion of distance. Let $D([0,U],\mathbb{R})\supset C([0,U],\mathbb{R})$ denote the space of right continuous functions with left limits and $\|\cdot\|_{M,U}$ the Skorokhod M_1 metric (see Skorokhod (1956) and Whitt, 2002 Chapters 3 and 11), which induces a weaker topology of convergence than the uniform metric; that is $\|a\|_{M,U}\leqslant\|a\|_U$, for any $a\in C([0,U],\mathbb{R})$. Then

$$\begin{aligned}\left\|\frac{1}{n}\rho^n(x,\cdot)-\bar{\rho}(x,\cdot)\right\|_{M,t} \leqslant{}& \left\|\frac{1}{n}\rho^n(x,0)-\bar{\rho}(x,0)\right\|_{M,t} + \left\|\frac{1}{l_x}\left(\frac{1}{n}Q^n(x-1,\cdot)-\int_0^{\bullet}\lambda(y^n(x-1,u),\beta)du\right)\right\|_{M,t}\\ &+\left\|\frac{1}{l_x}\left(\frac{1}{n}Q^n(x,\cdot)-\int_0^{\bullet}\lambda(y^n(x,u),\beta)du\right)\right\|_{M,t}\\ &+\left\|\frac{1}{l_x}\left(\int_0^{\bullet}\lambda(y^n(x-1,u),\beta)du-\int_0^{\bullet}\lambda(\bar{y}(x-1,u),\beta)du\right)\right\|_{M,t}\\ &+\left\|\frac{1}{l_x}\left(\int_0^{\bullet}\lambda(y^n(x,u),\beta)du-\int_0^{\bullet}\lambda(\bar{y}(x,u),\beta)du\right)\right\|_{M,t}\end{aligned} \tag{41}$$

where the inequality follows from the triangle inequality.

As $n\to\infty$ the first term on the right hand side of (41) converges to zero by assumption. The second and third terms converge to zero in accordance with Corollary 3.1.

For the last two terms, define the Lipschitz constant $K_x = K/l_x$ and note, by Lipschitz continuity of the flux functions and the triangle inequality, that

$$\left\|\frac{1}{l_x}\left(\int_0^{\bullet}\lambda(y^n(x,u),\beta)du-\int_0^{\bullet}\lambda(\bar{y}(x,u),\beta)du\right)\right\|_{M,t} \leqslant K_x\int_0^t\|y^n(x,\cdot)-\bar{y}(x,\cdot)\|_{M,u}du \tag{42}$$

By the boundedness of traffic densities, we can write

$$\|y^n(x,\cdot) - \bar{y}(x,\cdot)\|_{M,u} \leqslant H\left\|\frac{1}{n}\rho^n(x,\cdot) - \bar{\rho}(x,\cdot)\right\|_{M,u} \tag{43}$$

for some constant $1 \leqslant H < \infty$.

Let Ξ^n denote the sum of the first three terms on the right-hand side of (41) and recall that, as $n \to \infty$, $\Xi^n \to 0$ weakly. Let $\widetilde{K} = 2HK_x$, then combining (41) with (42) and (43), we get the following bound:

$$\left\|\frac{1}{n}\rho^n(x,\cdot) - \bar{\rho}(x,\cdot)\right\|_{M,t} \leqslant \Xi^n + \widetilde{K}\int_0^t \left\|\frac{1}{n}\rho^n(x,\cdot) - \bar{\rho}(x,\cdot)\right\|_{M,u} du \tag{44}$$

and applying the Bellman–Gronwall inequality (see Appendix A), we get

$$\left\|\frac{1}{n}\rho^n(x,\cdot) - \bar{\rho}(x,\cdot)\right\|_{M,t} \leqslant \Xi^n e^{\widetilde{K}t} \tag{45}$$

The right-hand side of (45) converges to zero weakly as $n \to \infty$, providing the desired result. □

The example below illustrates the intuition behind scaling.

3.3.1. Example 4: scaling

Let $\{X_i\}$ be IID Gamma distributed time headways, with shape parameter 2 and mean headway $\bar{h} = 10$ s/veh. Again, A_k denotes the arrival time of the kth vehicle and $Q(t)$ is the reciprocal counting process. Fig. 5 below plots a single sample path of the counting process for the first 10 arrivals.

The scaled version of the process, $(1/n)Q^n(t)$, in essence involves a *larger* number of *smaller* jumps. To see this, Fig. 6 plots a sample path of the scaled counting process for $n = 3$.

As n gets larger, the variance of the process shrinks, and in the limit the process degenerates to a deterministic process. To illustrate this, the following figures compare plots of the jump points of 500 sample paths of $Q(t)$ (i.e., $n = 1$) with 500 sample paths of $(1/n)Q^n(t)$ for $n = 10$, $n = 100$, and $n = 1000$, in Figs. 7–9, respectively, for a time horizon $U = 300$.

Clearly, as n gets larger, process variation gets smaller. In Fig. 9, it is difficult to distinguish between the scaled process and the (continuous) deterministic process $\overline{Q}(t) = \lambda t = 0.1t$. This is the fluid limit of $Q(t)$, in accordance with Corollary 3.1. It is worth noting that the fluid limit for counting IID time headways is much easier to derive than the state-dependent case that we derived in Theorem 3.1 (interested readers can refer to Appendix B for the IID case).

3.4. Model implementation and discussion

In essence, the proposed stochastic model starts with (i) a conditional headway distribution, $G(\cdot|y(x,t))$, where $y(x,t) = [\rho(x,t)\ \rho(x+1,t)]^T$ is a vector of relevant traffic state information at location x and at time t; (ii) a flux function $Q_e(\cdot)$ from which the conditional mean time headways and conditional variances in time headways are derived: $\bar{h}(y,\beta)$

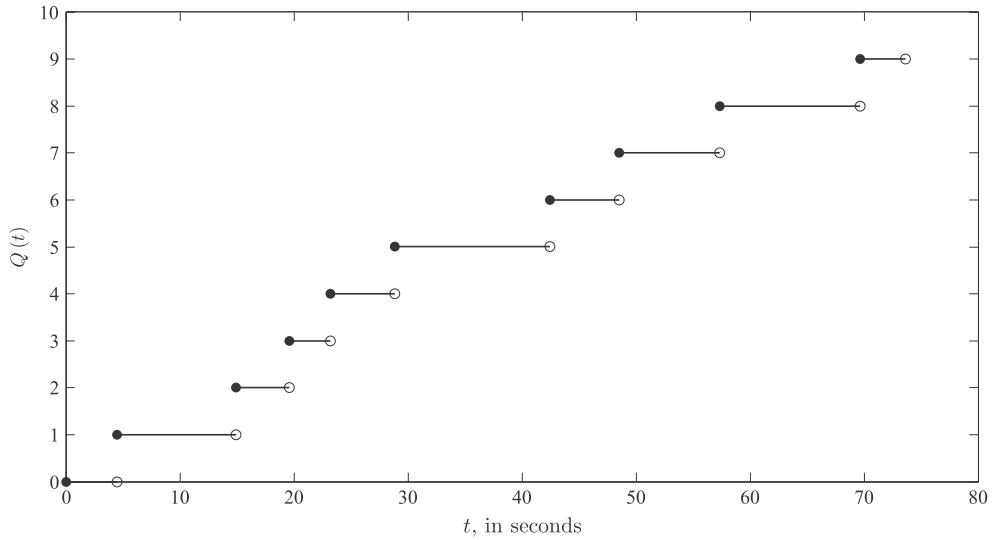

Fig. 5. Example 4 sample path for the original process ($n = 1$).

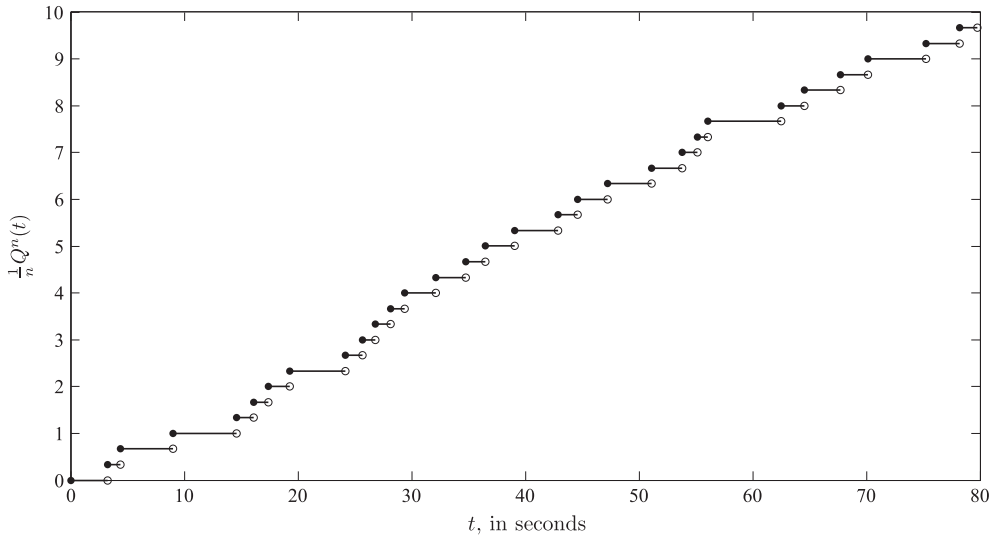

Fig. 6. Example 4 sample path for a scaled process with $n = 3$.

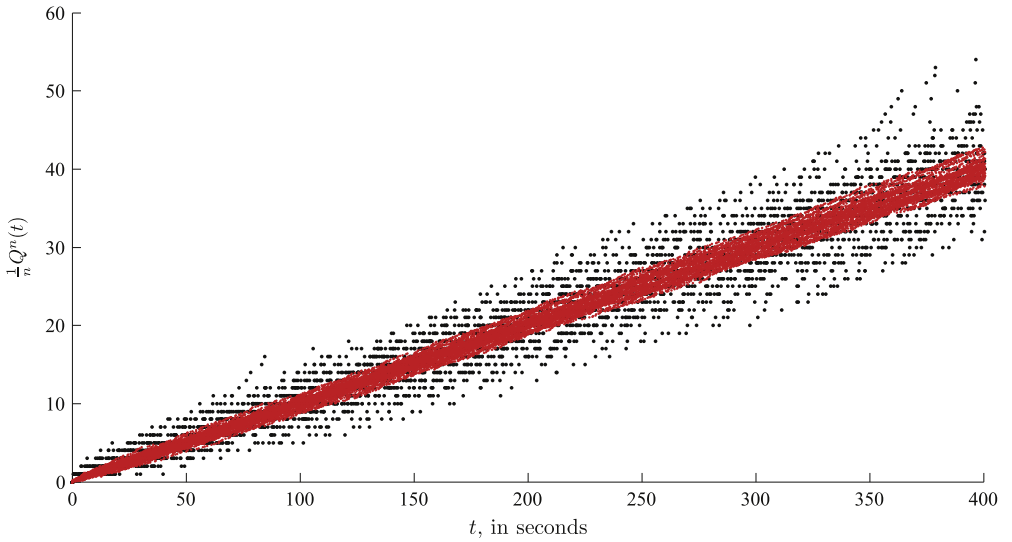

Fig. 7. Example 4 sample path comparisons; black: $n = 1$, red: $n = 10$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and $(1_{\{y_1>0\}}\bar{c}\cdot\bar{h}(y,\beta))^2$; and (iii) an initial cell density $\rho(x,0)$, which could be a random variable or a deterministic quantity, but in either case $l_x\rho(x,0)$ should be a bounded non-negative integer. To obtain a sample path of the process (i.e., to simulate a single realization of the process), for each cell $x \in \mathcal{C}$ one uses the initial density to compute $\bar{h}(y,\beta)$, which is then used to simulate a random time headway (as a $G(\cdot|y(x,t))$ distributed pseudo-random number), $\{X_1(x)\}_{x\in\mathcal{C}}$. These random numbers constitute the times at which the cumulative flows $Q(\cdot,\cdot)$ are increased by a single unit, which are, in turn, used in the conservation equation to compute the new densities. The procedure is then repeatedly carried out where in each step i the set of headways $\{X_i(x)\}_{x\in\mathcal{C}}$ are simulated and the densities are updated, and until the horizon time, U, is reached.

Here, the main issue (related to simulating the process) is that each of the cells maintains a different (random) clock; this could introduce difficulties from a computational stand-point. We note that in the special case where $G(\cdot|y(x,t))$ is the conditional distribution of an exponential random variable, this problem does not arise since exponential random variables are memoryless. In this case, in each step i, we take the minimum value of the simulated set $\{X_i(x)\}_{x\in\mathcal{C}}$ and only update the densities of the cells in the immediate upstream and immediate downstream of the location (cell boundary) with the minimum time headway using the conservation equation and update a universal time clock accordingly; an example is given in Section

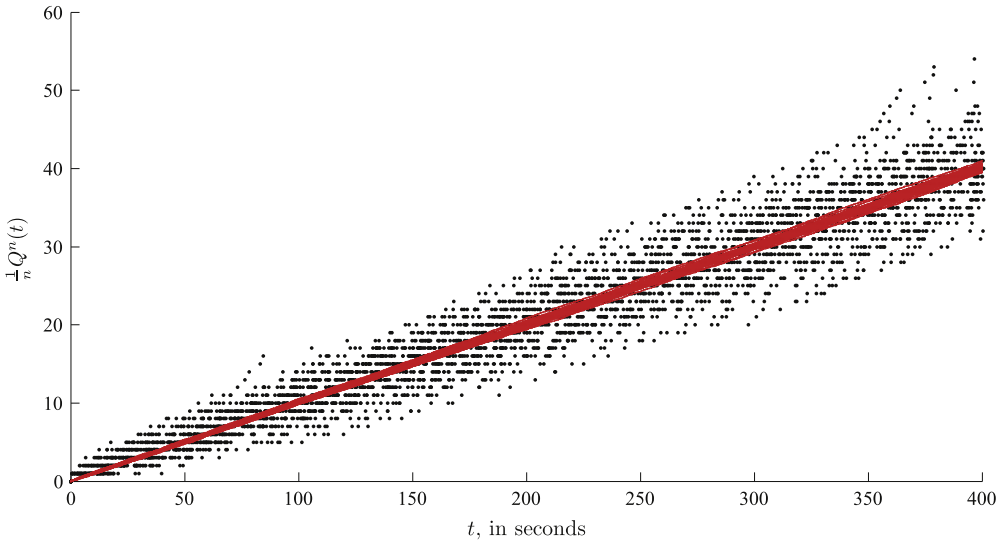

Fig. 8. Example 4 sample path comparisons; black: $n = 1$, red: $n = 100$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

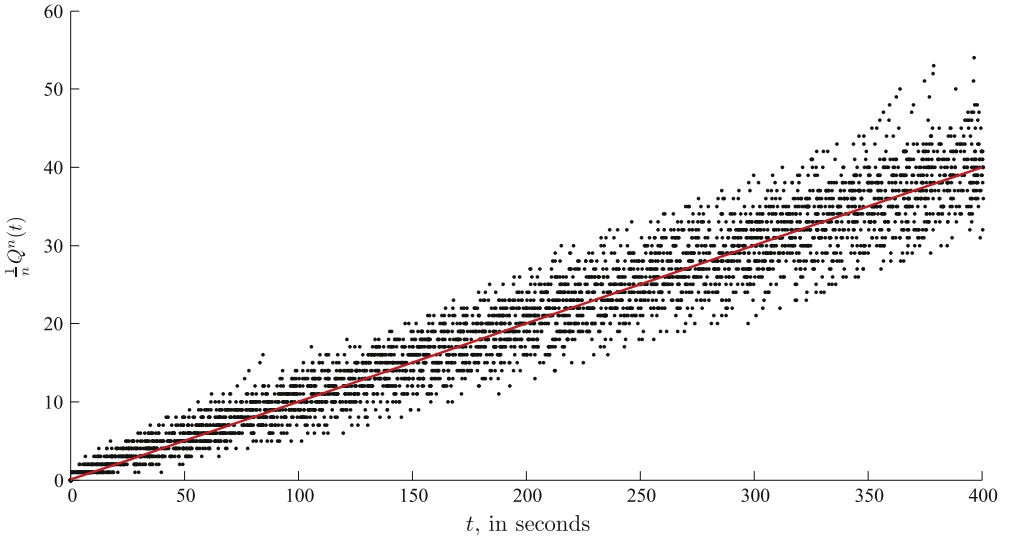

Fig. 9. Example 4 sample path comparisons; black: $n = 1$, red: $n = 1000$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1 below. This issue does not impact our analytical result in the general case; namely that the fluid limit in (28) will still agree with the CTM, Eq. (10). Implementation issues shall be addressed in a sequel.

4. Numerical examples

In the following two examples, we give a simulation algorithm based on an assumption of conditionally exponentially distributed vehicle time headways. We shall address the more general situation (simulating processes with non-exponential time headways) in a sequel. In Section 4.1, we give an example that is easy to reproduce to illustrate the consistency of the scaled processes with the CTM as n gets large. The second example (Section 4.2) illustrates the evolution of probability densities in the presence of a shockwave.

4.1. Two cell example

In this section, we give an algorithm to simulate traffic densities based on our proposed model. We shall use this example to illustrate consistency with the cell transmission model. We use the two cell setting that appears in (Sumalee et al., 2011,

Appendix A) with the upstream cell being labeled cell 1 and the downstream cell labeled cell 2. The choice of such a simple setting is to allow the reader to reproduce our results with ease. The simulation time horizon is $U = 200$ s; the cells are 100 m long each. Below are the flux functions (based on a triangular fundamental diagram):

$$\lambda(y(0,t),\beta) = \min\{5000, 20(400 - \rho(1,t))\} \tag{46a}$$

$$\lambda(y(1,t),\beta) = \min\{60\rho(1,t), 6000, 20(400 - \rho(2,t))\} \tag{46b}$$

$$\lambda(y(2,t),\beta) = \min\{60\rho(2,t), 6000g(t)\} \tag{46c}$$

where $g(t) = 0$ if $t \in [50,70)$ and $g(t) = 1$ otherwise, which is used to artificially create congestion. Eq. (46a) is the flow rate into cell 1 at time t, based on a constant potential inflow rate of 5000 vehicles per hour (i.e., a constant sending function); Eq. (46b) is the flow rate between cells 1 and 2; and Eq. (46c) is the flow rate out of cell 2.

For simplicity, we shall assume that vehicle time headways are conditionally exponentially distributed; that is, $G(a|y(x,t)) = 1 - exp(-\lambda(y(x,t),\beta)a)$, $\bar{c} = 1$, and $\bar{h}(y(x,t),\beta) = 1/\lambda(y(x,t),\beta)$. To ensure boundedness of vehicle time headways, in this example, we use $\varepsilon = 10^{-6}$. Under these assumptions, a sample path of the proposed stochastic model can be simulated using the following algorithm:

Algorithm 1.

Initialization:
t:=0
ε:=10^{-6}
$\rho^n(x,0)$:=0, for all $x \in \mathcal{C}$
Iteration:
while $t < U$ **do**
  for $x = 0 \rightarrow |\mathcal{C}|$ **do**
    state λ_x:=$\lambda(y^n(x,t),\beta)$ /3600 (flow in vehicles per second)
    if λ_x == 0 **then**
      λ_x:=ε
    end if
    v_x:=*uniform*(0,1) (uniform pseudo-random number in [0,1])
    t_x:= $-\log(1 - v_x)/\lambda_x$ (exponential pseudo-random number)
    t_x:=t_x/n (scaled time headways)
  end for
  $i := y : t_y == \min_{x\in\mathcal{C}}(t_x)$($i$:= index of the smallest headway)
  t:=$t + t_i$ (update system clock)
  Compute (scaled) numbers of vehicles leaving cells:
  for $x = 0 \rightarrow |\mathcal{C}|$ **do**
    if x == i **then**
      $q_x^n := 1/n$
    else
      $q_x^n := 0$
    end if
  end for
  Compute cell densities:
  for $x = 1 \rightarrow |\mathcal{C}|$ **do**
    $\rho^n(x,t) = \rho^n(x, t - t_i) + (1/l_x)(q_{x-1}^n - q_x^n)$
  end for
end while

In essence, the algorithm proceeds as follows: (i) compute the fluxes at the cell boundaries, (ii) use the fluxes to generate $|\mathcal{C}|$ exponential pseudo-random numbers (headways), (iii) find the minimum headway and scale it, t_x, (iv) set the flow across the cell boundary with the smallest headway to $1/n$ and all other boundary flows to 0, (v) compute the new densities, and (vi) repeat the process until $t == U$.

The figures below, show the results (evolution of traffic densities) from four simulations with $n = 1$, $n = 10$, $n = 100$, and $n = 1000$ and compare each to the CTM (computed using $\Delta t = 0.6$ s). Each simulation gives a single sample path of the process. Fig. 10 shows the evolution of traffic densities in cell 1 for the four scaling cases while Fig. 11 illustrates this for cell 2. In both figures it can be seen that as n gets larger sample paths tend to get closer and closer to the CTM. We note that for the same example in (Sumalee et al., 2011), the mean traffic densities at time 31 differ from the densities that one would compute using the CTM.

4.2. Probability profile in the presence of a shockwave

In this section, we construct a simple hypothetical example of a homogeneous 1 mile pipeline, that is divided into cells of length $l = 0.02$ miles, with a traffic signal at the downstream end. The purpose of this exercise is to illustrate the evolution of probability densities of time headways along the road in the presence of a shockwave that is created by the traffic signal.

As in Section 4.1, we shall assume that the time headways are conditionally exponentially distributed, given the system state. We will also assume that the underlying fundamental relationship is a triangular relationship with $\beta = [v_f \; q_{max} \; w \; \rho_{jam}]^T = [60 \text{ min/h } 1800 \text{ veh/h } 10 \text{ min/h } 210 \text{ veh/m}]^T$. We then have the following flux functions at the cell boundaries:

$$\lambda(y(x,t),\beta) = \min\{\lambda_0(t), R_e(\rho(x+1,t))\}, \quad x = 0 \tag{47a}$$

$$\lambda(y(x,t),\beta) = \min\{S_e(\rho(x,t)), R_e(\rho(x+1,t))\}, \quad 1 \leqslant x \leqslant |\mathcal{C}| - 1 \tag{47b}$$

$$\lambda(y(x,t),\beta) = \min\{S_e(\rho(x,t)), g(t)q_{max}\}, \quad x = |\mathcal{C}| \tag{47c}$$

where $\lambda_0(t)$ is a deterministic exogenous inflow rate and $g(t)$ is a signal status indicator at time t; i.e., $g(t) = 1$ if the signal status is green at time t and $g(t) = 0$ if the signal status is red at time t. The initial densities in all cells are assumed to be identically zero; that is, $\rho(x,0) \equiv 0$ for all $x \in \mathcal{C}$.

The conditions at the boundaries for this example, with $U = 700$ s, are: $\lambda_0(t) = 1600$ veh/h for $t \in [0,600]$ and $\lambda_0(t) = 800$ veh/h for $t \in [600,700]$, while $g(t) = 0$ for $t \in [100,200]$ and $g(t) = 1$ for $t \in [0,100) \cup (200,700]$.

We then follow the algorithm presented in Section 4.1 to recursively compute the traffic densities as a single realization of the proposed stochastic model. Fig. 12 below illustrates the evolution of traffic densities along the road.

The horizontal axis in Fig. 12 represents space (increasing from left to right) and the vertical axis represents time (increasing from bottom to top). Darker colors mean lower densities and lighter colors higher density. The lower part of the lighter diagonal band in the figure is the trajectory of a shockwave that propagates into the upstream at a speed of ≈10 min/h. The top part of the light diagonal band appears blurrier than the top part, since it constitutes queue dissipation dynamics, described by rarefaction fans (see Fig. 1 above).

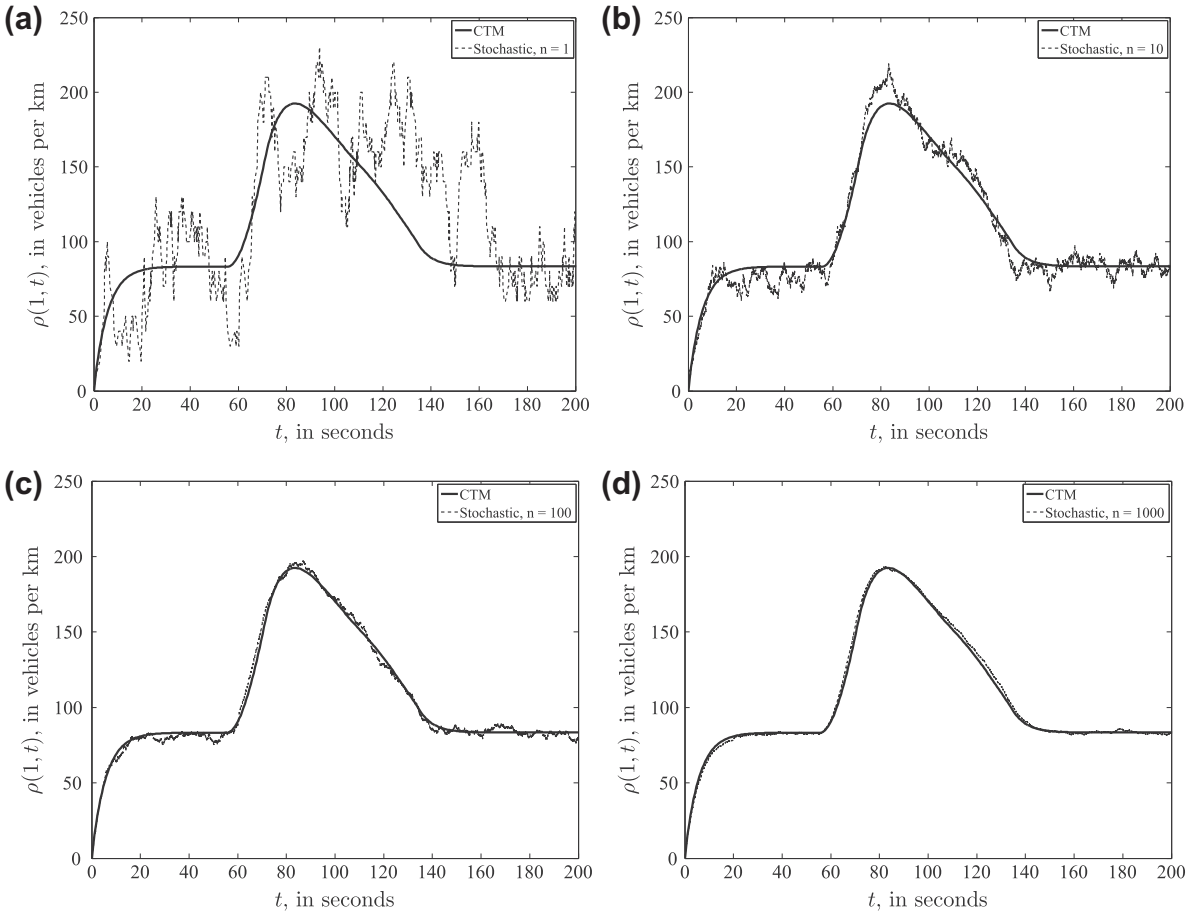

Fig. 10. Comparisons with the CTM for density evolution in cell 1, (a) $n = 1$, (b) $n = 10$, (c) $n = 100$, (d) $n = 1000$.

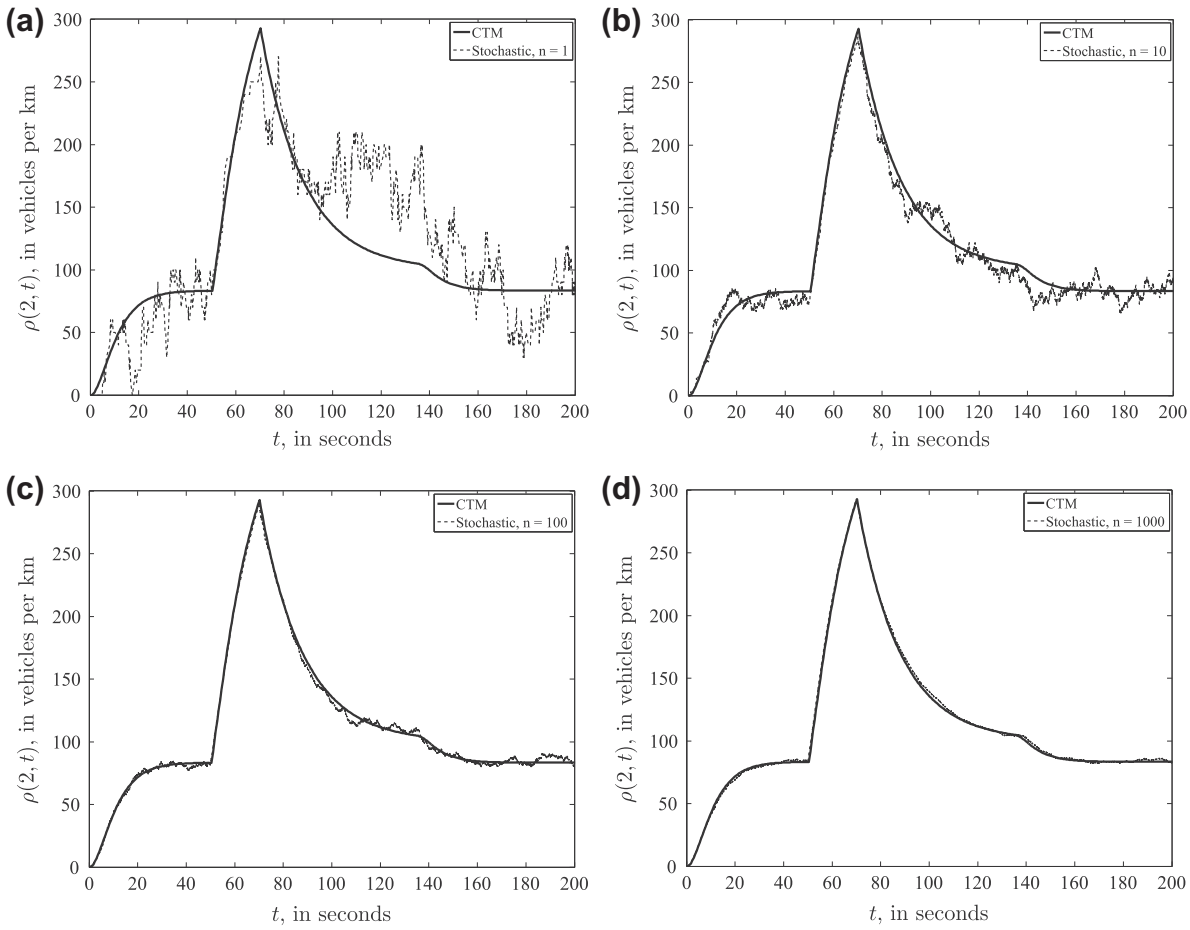

Fig. 11. Comparisons with the CTM for density evolution in cell 2, (a) $n = 1$, (b) $n = 10$, (c) $n = 100$, (d) $n = 1000$.

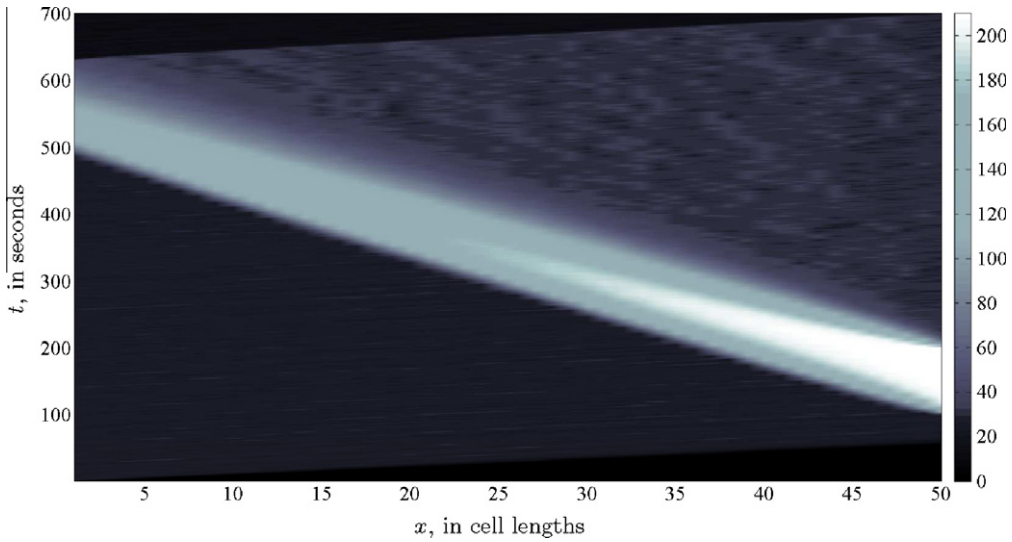

Fig. 12. Simulated traffic densities.

Figs. 13–16 show four snapshots of the probability densities of time headways along the road. Cross-sections along the x-axis represent exponential probability densities of time headways at location x at a particular time instance.

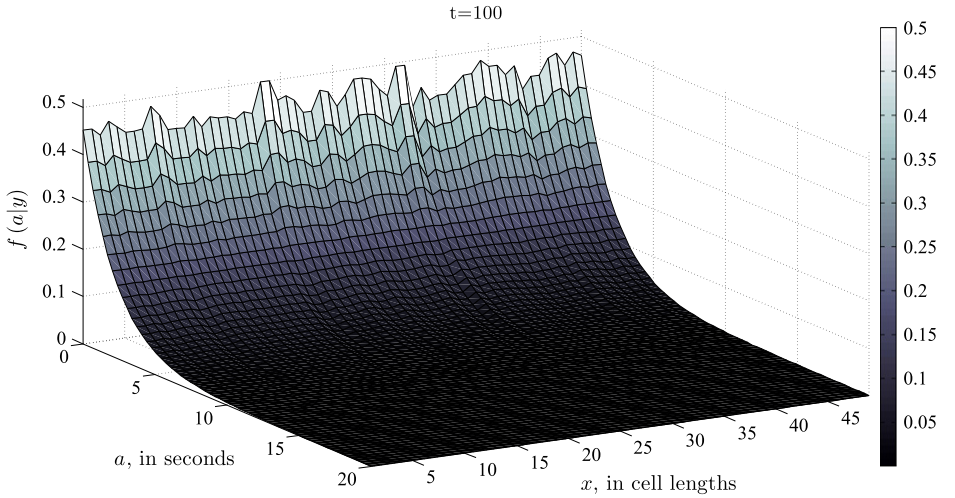

Fig. 13. Snapshot of probability densities at time $t = 100$.

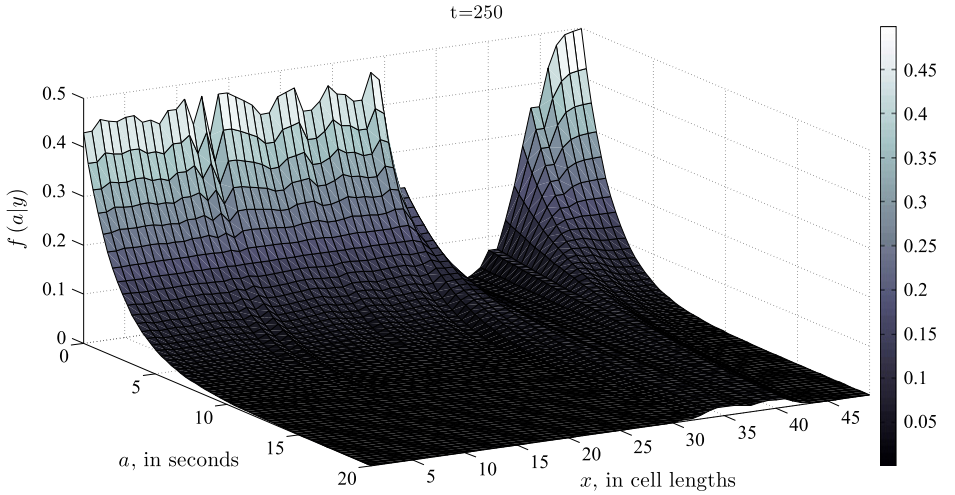

Fig. 14. Snapshot of probability densities at time $t = 250$.

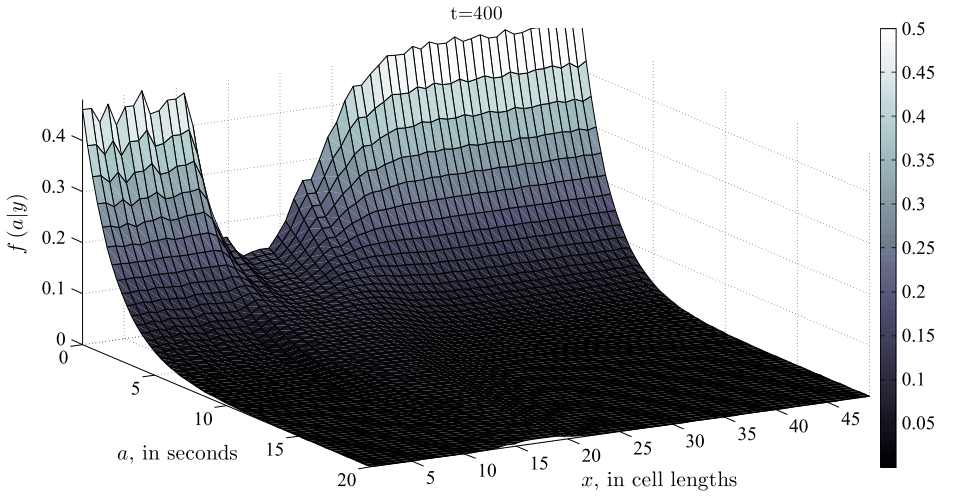

Fig. 15. Snapshot of probability densities at time $t = 400$.

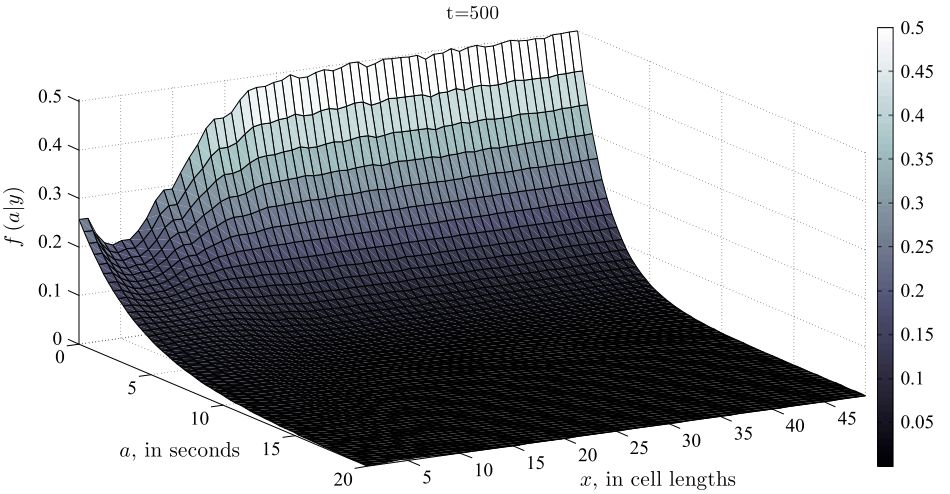

Fig. 16. Snapshot of probability densities at time $t = 500$.

What we can see in the Figs. 13–16 is that probability densities tend to be flatter with higher traffic densities signifying that shorter time headways are "less likely" to occur. We can also see the propagation of the disturbance into the upstream as time increases. At time $t = 100$ the shockwave has yet not formed, thus the uniformity in the probability densities along the road in Fig. 13. At time $t = 250$, there is heavy congestion near the downstream, seen as a valley in Fig. 14, where probability densities are almost flat at locations where traffic is not moving, but there is also traffic being discharged further downstream and in between the immobile traffic and the traffic being discharged we see a gradual steepening of the probability densities, similar to a rarefaction fan. At time $t = 400$, the valley in the probability densities has propagated further upstream (Fig. 15), but is not as deep as it was earlier and further in the upstream, since the disturbance is felt further upstream but is not as severe. At time $t = 500$ (Fig. 16), the valley has almost reached the upstream end of the road, but traffic conditions are not as bad.

5. Conclusion and future research

This paper presents a new stochastic model, for which a long run temporal mean dynamic is shown to be consistent with a well established first order traffic flow models, the CTM. At the same time the proposed model implicitly ensures non-negativity of traffic densities. In contrast to previous stochastic models of traffic flow, where noise is added to deterministic equations, the source of uncertainty in our model is driver gap choice. Our treatment of vehicle time headways is general in the sense that we can accommodate a variety of time headway distributions that are dependent on traffic states. Numerical examples are given to illustrate scaling and convergence to the CTM and to demonstrates the evolution of probability densities of time headways in space and time as a result of the propagation of a traffic disturbance.

This work could be extended in a number of directions for future research. A natural follow-up study would be the estimation and validation of the proposed model using real-world traffic data, in various settings including freeways and signalized arterials. One could also extend the proposed model by considering merging and diverging behavior in a stochastic context, which could lead to a stochastic model of network traffic. Also, due to its probabilistic nature, the proposed model can be applied to the reliable/robust control of ramp meters and traffic signals.

Appendix A. Bellman–Gronwall inequality

Theorem 5.1 (*Bellman–Gronwall inequality*). *Let $z(t)$ be a non-negative function, such that, for $t \in [0, U]$,*

$$z(t) \leqslant \Xi + \Upsilon \int_0^t z(u)du \tag{A.1}$$

for some constants Ξ and Υ. Then, for $t \in [0, U]$,

$$z(t) \leqslant \Xi e^{\Upsilon t} \tag{A.2}$$

Proof. See Bellman (1943) for the proof of a slightly more general version, and see (Oksendal, 2007 Exercise 5.17) for the version that appears here. □

Appendix B. A functional strong law of large numbers for counting IID time headways

The derivation below is a well-known derivation; see, for instance (Resnick, 1992, Theorem 3.3.2) and (Chen and Yao, 2001, Lemma 5.8). We include it here for the sake of completeness.

Let $\{X_i\}_{i\geqslant 1}$ denote a sequence of IID random variables with common distribution function $G(\cdot)$ and common mean $\bar{h}$. Then note that $(1/n)A_n = (1/n)\sum_{i=1}^{n} X_i$ converges, as $n \to \infty$, to $\bar{h}$ by the classical strong law of large numbers; also note that the counting process $Q(nt)$ goes to ∞ as $n \to \infty$. From Eq. (18), we have:

$$A_{Q(nt)} \leqslant nt < A_{Q(nt)+1}, \tag{B.1}$$

so that

$$\frac{1}{A_{Q(nt)}} \geqslant \frac{1}{nt} > \frac{1}{A_{Q(nt)+1}}, \tag{B.2}$$

and upon multiplying by $Q(nt)$, we get

$$\frac{Q(nt)}{A_{Q(nt)}} \geqslant \frac{Q(nt)}{nt} > \frac{Q(nt)+1}{A_{Q(nt)+1}} \frac{Q(nt)}{Q(nt)+1} \tag{B.3}$$

As $n \to \infty$, $(1/Q(nt))A_{Q(nt)} \to \bar{h}$ and $Q(nt)/(Q(nt)+1) \to 1$, which implies that

$$\frac{Q(nt)}{n} \underset{n\to\infty}{\to} \frac{1}{\bar{h}} t = \lambda t \tag{B.4}$$

as expected. Note that the convergence in this case is *almost sure*.

References

Bellman, R., 1943. The stability of solutions of linear differential equations. Duke Mathematical Journal 10 (4), 643–647.
Boel, R., Mihaylova, L., 2006. A compositional stochastic model for real-time freeway traffic simulation. Transportation Research Part B 40 (4), 319–334.
Branston, D., 1976. Models of single lane time headway distributions. Transportation Science 10 (2), 125–148.
Bui, D., Nelson, P., Narasimhan, S., 1992. Computational realizations of the entropy condition in modeling congested traffic flow. Federal Highway Administration Report FHWA/TX-92/1232-7.
Chen, H., Yao, D., 2001. Fundamentals of Queueing Networks. Springer, New York.
Cowan, R., 1975. Useful headway models. Transportation Research 9 (6), 371–375.
Daganzo, C., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B 28 (4), 269–287.
Daganzo, C., 1995a. A finite difference approximation of the kinematic wave model of traffic flow. Transportation Research Part B 29 (4), 261–276.
Daganzo, C., 1995b. The cell transmission model, part II: network traffic. Transportation Research Part B 29 (2), 79–93.
Davis, G., Kang, J., 1994. Estimating destination-specific traffic densities on urban freeways for advanced traffic management. Transportation Research Record: Journal of the Transportation Research Board 1457, 143–148.
Gazis, D., Knapp, C., 1971. On-line estimation of traffic densities from time-series of flow and speed data. Transportation Science 5 (3), 283–301.
Gazis, D., Liu, C., 2003. Kalman filtering estimation of traffic counts for two network links in tandem. Transportation Research Part B 37 (8), 737–745.
Godunov, S., 1959. A difference scheme for numerical computation of discontinuous solutions of hydrodynamic equations. Mathematics of the Sbornik 47 (3), 271–306.
Helbing, D., 2001. Traffic and related self-driven many-particle systems. Reviews of Modern Physics 73 (4), 1067–1141.
Holden, H., Risebro, N., 1997. Conservation laws with a random source. Applied Mathematics and Optimization 36 (2), 229–241.
Hoogendoorn, S., Bovy, P., 1998. New estimation technique for vehicle-type-specific headway distributions. Transportation Research Record: Journal of the Transportation Research Board 1646, 18–28.
Kang, J., 1995. Estimation of destination-specific traffic densities and identification of parameters on urban freeways using Markov models of traffic flow. Ph.D. thesis, University of Minnesota.
Lebacque, J., 1996. The Godunov scheme and what it means for first order traffic flow models. In: Lesort, J.B. (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Elsevier, Lyon, France, pp. 647–677.
LeVeque, R., 1992. Numerical Methods for Conservation Laws, second ed. Birkhäuser, Berlin.
Lighthill, M., Whitham, G., 1955. On kinematic waves. I: flood movement in long rivers, II.A theory of traffic flow on long crowded roads. In: Proceedings of the Royal Society (London) A229. pp. 281–345.
Oksendal, B., 2007. Stochastic Differential Equations, sixth ed. Springer, Berlin.
Osher, S., 1984. Riemann solvers, the entropy condition, and difference approximations. SIAM Journal on Numerical Analysis 21 (2), 217–235.
Paveri-Fontana, S., 1975. On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis. Transportation Research 9 (4), 225–235.
Prigogine, I., Andrews, F., 1960. A Boltzmann-like approach for traffic flow. Operations Research 8 (6), 789–797.
Prigogine, I., Herman, R., 1971. Kinetic Theory of Traffic Flow. Elsevier, New York.
Resnick, S., 1992. Adventures in Stochastic Processes. Birkhäuser, Boston.
Richards, P., 1956. Shock waves on the highway. Operations Research 4 (1), 42–51.
Skorokhod, A., 1956. Limit theorems for stochastic processes. Theory of Probability and Its Applications 1 (3), 261–290.
Sumalee, A., Zhong, R., Pan, T., Szeto, W., 2011. Stochastic cell transmission model (SCTM): a stochastic dynamic traffic model for traffic state surveillance and assignment. Transportation Research Part B 45 (3), 507–533.
Szeto, M., Gazis, D., 1972. Application of Kalman filtering to the surveillance and control of traffic systems. Transportation Science 6 (4), 419–439.
Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. Transportation Research Part B 39 (2), 141–167.
Wang, Y., Papageorgiou, M., Messmer, A., 2007. Real-time freeway traffic state estimation based on extended Kalman filter: a case study. Transportation Science 41 (2), 167–181.
Whitt, W., 2002. Stochastic Process Limits. Springer, New York.