

MIT Open Access Articles

*Capturing Dependency Among Link Boundaries
in a Stochastic Dynamic Network Loading Model*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Osorio, Carolina, and Gunnar Flotterod. "Capturing Dependency Among Link Boundaries in a Stochastic Dynamic Network Loading Model." *Transportation Science* 49, no. 2 (May 2015): 420–431.

As Published: <http://dx.doi.org/10.1287/trsc.2013.0504>

Publisher: Institute for Operations Research and the Management Sciences (INFORMS)

Persistent URL: <http://hdl.handle.net/1721.1/101683>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Capturing dependency among link boundaries in a stochastic dynamic network loading model

Carolina Osorio * Gunnar Flötteröd †

April 13, 2013

Abstract

This work adds realistic dependency structure to a previously developed analytical stochastic network loading model. The model is a stochastic formulation of the link-transmission model, which is an operational instance of Newell's simplified theory of kinematic waves. Stochasticity is captured in the source terms, the flows, and, consequently, in the cumulative flows. The previous approach captured dependency between the upstream and downstream boundary conditions within a link (i.e. the respective cumulative flows) only in terms of time-dependent expectations without capturing higher-order dependency. The model proposed in this paper adds an approximation of full distributional stochastic dependency to the link model. The model is validated versus stochastic microsimulation in both stationary and transient regimes. The experiments reveal that the proposed model provides a very accurate approximation of the stochastic dependency between the link's upstream and downstream boundary conditions. The model also yields detailed and accurate link state probability distributions.

1 Introduction

A *network loading model* describes how a time-dependent travel demand advances through a network. The demand is typically given in terms of time-dependent origin/destination (OD) flows and a route choice model. Given these inputs, the network loading model then captures the progression of the demand through the network, accounting for congestion and the resulting delays.

A *stochastic network loading model* does essentially the same, but it additionally accounts for uncertainty in the modeling (e.g. of source terms, flows, network parameters) and captures distributional information of the network states. This paper focuses on analytical stochastic (i.e. probabilistic)

*Massachusetts Institute of Technology (MIT), Department of Civil & Environmental Engineering, Cambridge, MA 02139, USA, osorioc@mit.edu (corresponding author)

†KTH Royal Institute of Technology, Department of Transport Science, 11428 Stockholm, Sweden, gunnar.floetteroed@abe.kth.se

dynamic network loading models. The main technical limitation in developing such models is the dimensionality of the joint distribution, which is exponential in the number of spatial discretization units. Thus, the challenge is to approximate the dependency structure while deriving a computationally efficient approach.

In the following, we focus on the widely accepted kinematic wave model (KWM; Lighthill and Witham (1955); Richards (1956)). Both the KWM's original link model and its more recently developed node models (e.g. Daganzo (1995); Lebacque (1996); Lebacque and Khoshyaran (2005); Tampere et al. (2011); Flötteröd and Rohde (2011); Corthout et al. (2012)) are deterministic. They describe space/time average conditions but do not account for higher-order distributional information.

There has been a recent interest in the development of stochastic link models. Most studies have considered stochastic cell-transmission models (CTMs; Boel and Mihaylova; 2006; Sumalee et al.; 2011; Jabari and Liu; 2012). Boel and Mihaylova (2006) consider the sending and receiving functions of the CTM as random variables. The evaluation of the model involves computationally intensive sampling in order to estimate the main link performance measures. Jabari and Liu (2012) consider headways to be random variables. The fluid limit of their stochastic model is consistent with the CTM. This is also a simulation-based approach, where performance measure estimates are obtained via sampling. The stochastic CTM of Sumalee et al. (2011) allows for stochasticity in the sending and receiving functions and in the source terms. This model is analytical (i.e. not simulation-based). The stochasticity results from adding noise in the form of a second-order wide-sense stationary process to otherwise deterministic model variables. Jabari and Liu (2012) detail the limitations of using such types of noise terms.

Let us also briefly comment on the kinetic approach to stochastic traffic flow modeling. Here, one starts out from a probabilistic description of individual-vehicle interactions, which is typically solved by extracting dynamic equations for the first moments (in particular, mean values and variances) of aggregate traffic characteristics (e.g. Tampere et al.; 2003). To derive operational formulations, the assumption of "vehicular chaos" is typically made, meaning that the states of interacting vehicles are stochastically independent. Nelson and Kumar (2006) elaborate on the implications of omitting such dependencies. It appears that the complexity of kinetic models with realistic stochastic dependency structures has so far precluded their implementation in non-trivial network contexts (Helbing; 2001).

Osorio et al. (2011) recently proposed a stochastic formulation of the link-transmission model of Yperman et al. (2007), which is an operational instance of Newell's simplified theory of kinematic waves (Newell; 1993). Newell's model can in turn be derived from the variational theory of Daganzo (2005). The queueing-theoretical model of Osorio et al. (2011) accommodates stochastic source terms and flows across nodes. It is a (vehicle-)discretized, stochastic instance of the KWM, whereas the aforementioned stochastic CTMs constitute stochastic instances of (space-)discretized KWMs. That is, only the model of Osorio et al. (2011) is *directly* derived from the KWM.

The present article adds important dependency structure to this model, hereafter referred to as the *basic model*. The *basic model* exhibits the following additional features.

1. It is analytical. It captures the evolution of link state distributions through differentiable equations. Thus, the approach does not require computationally costly sampling to obtain distributional estimates. This approach allows to obtain valuable insights into stochastic network dynamics. It also provides a differentiable description of these dynamics, which can be exploited

when applying efficient optimization or calibration routines, e.g. for the design of signal control strategies (Osorio and Bierlaire; 2010) or the estimation of OD matrices (Flötteröd et al.; 2011).

2. The basic model represents the flow transmissions across a network node (connection of upstream and downstream links) in terms of a multivariate Poisson process. The node model yields a joint distribution of the downstream boundary conditions of the node's upstream link and the upstream boundary conditions of the node's downstream link.
3. The basic model represents a homogeneous link segment by two finite capacity queues that constitute stochastic counterparts of the cumulative curves used in Newell's simplified KWM, and it coincides with a discretized version of Newell's model when the randomness in all involved processes vanishes. That is, the basic link model is derived from the KWM (Newell; 1993; Yperman et al.; 2007).
4. The basic model captures dependency between a single link's upstream boundary conditions and the same link's downstream boundary conditions merely in terms of time-dependent expectations but ignores higher-order dependencies. This implies that dependency between upstream and downstream boundary conditions in a single link is not captured beyond what a deterministic model could describe.

Item 4 constitutes the main simplification in the basic model, which this article overcomes. The presentation therefore focuses on the joint modeling of boundary conditions within a link. Details on how across-node correlations (i.e. correlations with links further up- or downstream) are captured appear in Osorio et al. (2011). It should be noted that none of the aforementioned stochastic CTMs provide analytical expressions for the joint distribution of multiple cells: Boel and Mihaylova (2006) and Jabari and Liu (2012) resort to simulation; Sumalee et al. (2011) model the states of adjacent cell pairs as independent random variables.

The remainder of this article is organized as follows. Section 2 recalls the basic link model. Section 3 describes how realistic dependency structure is added, leading to the proposed new model. Comprehensive experiments are described in Section 4, where the distributional information extracted from the analytical model is compared to distributional estimates obtained via simulation. Finally, Section 5 concludes the article and gives an outlook on further research questions.

2 Basic link model

We briefly recall the original link model of Osorio et al. (2011). The presentation given here follows a different path than the original work in that it first recalls an operational formulation (Yperman et al.; 2007) of Newell's simplified KWM (Newell; 1993) and then formulates the stochastic model as a distributed version of Newell's model.

Yperman et al. (2007) phrase this model within the sending/receiving function framework of Daganzo (1994) and of Lebacque (1996). This framework postulates that, at any interface within the network, the instantaneously transmitted flow is the minimum of an upstream sending function and a downstream receiving function, reflecting the KWM's principle of local flow maximization (Ansorge;

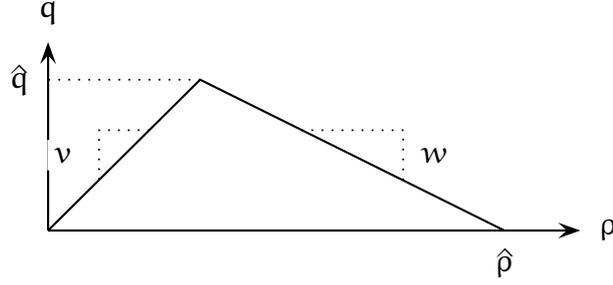


Figure 1: Deterministic fundamental diagram resulting from original two-queue system

1990). The embedding of a link in a network model hence requires a link model that defines, at every time instant t , a receiving function $R(t)$ (and a sending function $S(t)$, respectively) that reflects the boundary conditions the link provides to its upstream node (downstream node, resp.).

Assuming a triangular fundamental diagram as shown in Figure 1, having free flow velocity v , backward wave speed w (negative), flow capacity \hat{q} , and jam density $\hat{\rho}$, Yperman et al. (2007) present the following equations for a discrete-time simulation of the KWM (detailed derivations can be found in Yperman (2007)):

$$R(t) = \min \left\{ N \left(L, t + \delta - \frac{L}{|w|} \right) + \hat{\rho}L - N(0, t), \hat{q}\delta \right\} \quad (1)$$

$$S(t) = \min \left\{ N \left(0, t + \delta - \frac{L}{v} \right) - N(L, t), \hat{q}\delta \right\}, \quad (2)$$

where $R(t)$ is the amount of flow the link can receive at time instant t during the next time interval of length δ , $S(t)$ is the respective sending flow, and $N(x, t)$ is the cumulative flow having passed location x at time t , with $x \in [0, L]$ in the link of length L . The Courant-Friedrichs-Lewy condition requires $\delta \leq L/v$ to hold. See, for instance, Yperman (2007) for a detailed discussion of the possible wave configurations given a triangular fundamental diagram and Nagel and Nelson (2005) for a discussion of its empirical validity.

Defining the two quantities

$$UQ(t) = N(0, t) - N(L, t + \delta - L/|w|) \quad (3)$$

$$DQ(t) = N(0, t + \delta - L/v) - N(L, t) \quad (4)$$

allows to rewrite (1), (2) as follows:

$$R(t) = \min \{ \hat{\rho}L - UQ(t), \hat{q}\delta \} \quad (5)$$

$$S(t) = \min \{ DQ(t), \hat{q}\delta \}. \quad (6)$$

Formally, $UQ(t)$ and $DQ(t)$ are just summary representations of differences in cumulative flows. However, they also allow for a tangible interpretation of these otherwise rather abstract cumulative flow differences. For this, $UQ(t)$ is interpreted as the number of vehicles in a finite capacity *upstream*

queue (UQ) that keeps track of the upstream boundary conditions *within* the link, and DQ(t) is interpreted as the number of vehicles in a finite capacity *downstream queue* (DQ) that keeps track of the downstream boundary conditions *within* the link.

Allow both queues to hold at most $\hat{\rho}L$ vehicles. The receiving function in (5) is hence limited by the available space in UQ, which is $\hat{\rho}L - UQ(t)$. This means that the link behaves *as if* UQ was embedded within its upstream end, and *as if* vehicles trying to enter the link actually tried to enter UQ. Further, the sending function in (6) is limited by the number of vehicles in DQ, which is DQ(t). This means that the link behaves *as if* DQ was embedded within its downstream end, and *as if* vehicles leaving the link actually left DQ. (Charypar (2008) describes essentially the same approach in a microsimulation framework, but without any analytical considerations).

This interpretation carries further. Equations (3) and (4) can be recursively written as

$$UQ(t) = UQ(t - \delta) + \delta [q^{\text{in}}(t - \delta) - q^{\text{out}}(t - L/|w|)] \quad (7)$$

$$DQ(t) = DQ(t - \delta) + \delta [q^{\text{in}}(t - L/v) - q^{\text{out}}(t - \delta)] \quad (8)$$

where $q^{\text{in}}(t)$ is the link's instantaneous inflow rate at time t , $q^{\text{out}}(t)$ is the instantaneous outflow rate, and both quantities are held constant throughout a time step of duration δ , consistently with the underlying framework of Yperman et al. (2007). Equation (7) indicates that the change in UQ during $[t - \delta, t]$ is given by the difference between

- i) the flow that entered the link during that time
- ii) and the flow that left the link during $[t - L/|w|, t - L/|w| + \delta]$.

Similarly, Equation (8) indicates that the change in DQ during $[t - \delta, t]$ is given by the difference between

- i) the flow that entered the link during $[t - L/v, t - L/v + \delta]$
- ii) and the flow that left the link during that time.

That is, UQ and DQ evolve through time *as if* the link in- and outflows actually entered the respective queues. It needs to be re-iterated, though, that UQ and DQ are merely intuitive representations of the boundary conditions provided by the link to its up- and downstream nodes.

Figure 2 shows the fictitious embedding of UQ and DQ within the link. The lower path (solid arrows) represents the actual mass transfer: flow enters the link, is delayed by L/v time units (corresponding to the free-flow travel time) and then becomes available for departure in DQ. The upper path (dashed arrows) captures how vehicle departures eventually enable new vehicle entries, in that flows that have departed from DQ are delayed by $L/|w|$ time units (corresponding to the time it takes a kinematic backward wave to traverse the link) before they are removed from UQ.

The stochastic link model results from a stochastic modeling of UQ and DQ, relying on finite capacity queueing theory, where the evolution of the distribution of the number of vehicles in either queue is tracked through time. The dynamics of these queues are guided by time-dependent arrival and service rates as well as the probabilities of the queues being perfectly empty (i.e. being unable to send more

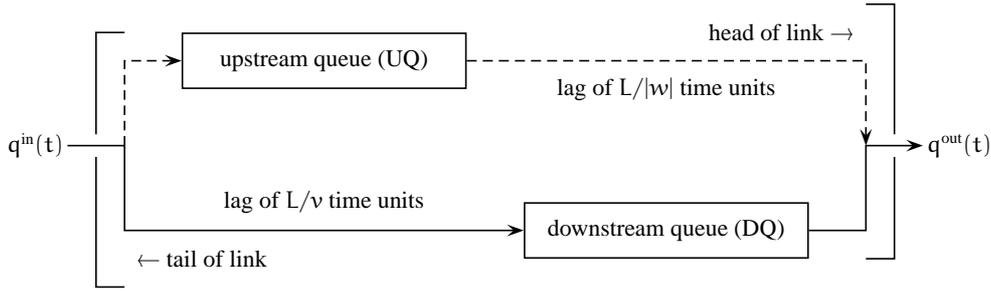


Figure 2: Original two-queue system

flow) or perfectly full (i.e. being unable to receive more flow). Since UQ and DQ represent differences in cumulative flows, this approach corresponds to Newell’s model with stochastic cumulative curves.

Assume a time-dependent Poisson arrival process with intensity $\lambda(t)$ (i.e. a non-homogenous Poisson process) at the upstream end of the considered link, where $\lambda(t) \leq \hat{q}$ in order to capture the second constraint in Equation (1). That is, inter-arrival times of vehicles to the link are exponentially distributed. This captures uncertainty in the source term (or, in a network context, in the demand from upstream). The probability that an arrival at time t encounters an available space in UQ is $P(\text{UQ}(t) < \ell)$ where ℓ is the link’s space capacity (corresponding to a rounded version of $\hat{p}L$). Allowing for losses (which would translate into spillback in the full node model not considered here), the effective inflow becomes $q^{\text{in}}(t) = \lambda(t)P(\text{UQ}(t) < \ell)$.

Furthermore, assume exponentially distributed service times at the downstream end of the link with rate $\mu(t) \leq \hat{q}$ in order to capture the second constraint in Equation (2), where $\mu(t)$ can be interpreted as a downstream capacity constraint (which would in a network embedding, amongst other things, capture spillback from downstream). Distributed service times may capture a distribution of headways across vehicles or stochastic flow interactions in the downstream node, the latter being possibly due to a gap acceptance distribution. The probability that there are vehicles ready to leave the link is given by the probability that there are vehicles in DQ. Thus, the outflow of the link is given by $q^{\text{out}}(t) = \mu(t)P(\text{DQ}(t) > 0)$.

The basic model captures uncertainties in upstream demand patterns and downstream supply conditions. Specifically, this model assumes arrivals to arise from a Poisson process. Two common criticisms of this assumption are mitigated by features of the present model, which are absent in other Markovian-type road traffic models. As discussed in Osorio et al. (2011), the *dynamics* allow to capture temporal dependency effects (e.g. platooning) deterministically through the joint dynamics of the time-dependent rates of all involved Poisson processes. The *finite capacity* assumption of this model ensures that unrealistically high flows do not arise. Such flows could arise in an infinite capacity queue setting, due to the Poisson distribution’s fat right tail. Ultimately, these assumptions will require empirical justification.

This link model is a simplification in the sense that UQ and DQ are modeled independently. To clarify this, let the *lagged inflow* $\text{LI}(t)$ at time t be the stochastic amount of flow that has entered the link between $t - L/v$ and t . In Figure 2, this corresponds to the flow on the lower left path, which has already entered the link but has not yet entered DQ. Further, let the *lagged outflow* $\text{LO}(t)$ at time t be the stochastic amount of flow that has left the link between $t - L/|w|$ and t . In Figure 2, this corresponds to the flow on the top right path, which has already left the link but has not yet left UQ.

Mass conservation then requires

$$LI(t) + DQ(t) = UQ(t) - LO(t) \quad (9)$$

to hold. Both sides denote the number of vehicles on the link at time t , such that a substantial dependency between the distribution of UQ and DQ can be expected.

The remainder of this article develops and analyzes an improved link model formulation that almost perfectly captures this dependency while maintaining consistency with the network modeling framework of Osorio et al. (2011).

3 New link model

The main difficulty when trying to capture a joint distribution of UQ and DQ is the fact that these evolve in reaction to the same inflows and outflows but evaluate these flows with different time lags.

Assuming as from now a discrete model where k is the time index, k^{fwd} is a rounded version of L/v , k^{bwd} is a rounded version of $L/|w|$, and ℓ is a rounded version of $\hat{\rho}L$, a joint distribution of all time-lagged model variables would result in an exponential state space increase as the involved time lags get larger.

The proposed solution to this problem is to add only two additional dimensions to the (UQ, DQ) state space, which are called the *lagged inflow queue* (LI) and the *lagged outflow queue* (LO). The LI queue captures, at an aggregate level, the distribution of all link entries that have not yet reached DQ , cf. Equation (8). Symmetrically, the LO queue captures, at an aggregate level, the distribution of all link exits that have not yet been removed from UQ , cf. Equation (7). Modeling a joint evolution of a four-dimensional state space (UQ, LI, DQ, LO) is expected to capture relevant aspects of the dependency between UQ and DQ .

Figure 3 gives an overview. In discrete time, the lags of inflows and outflows correspond to moving them through a sequence of k^{fwd} (respectively, k^{bwd}) buffers. LI (resp. LO) contains the sum of the lagged inflows (resp. outflows) in the corresponding buffers. That is, each dotted box contains three alternative ways of representing a lag: the original lag in continuous time, the discrete-time representation as a series of buffers, and the aggregation into one single queue.

Let $UQ(t; k)$ denote the number of vehicles in UQ at continuous time t within time interval k of duration δ . Similarly, we define $LI(t; k)$, $DQ(t; k)$ and $LO(t; k)$. At a given point in time, we have:

$$UQ(t; k) = DQ(t; k) + LI(t; k) + LO(t; k). \quad (10)$$

Let us emphasize the purpose of each of these queues. DQ represents the number of vehicles that could possibly leave the link. LI represents the number of vehicles that have entered the link but are not yet available for departure due to the finite link traversal time. LO represents the number of “spaces” that correspond to departures from the link that, due to the finite backward wave speed, have not yet propagated to the link’s upstream end. Thus, UQ in (10) represents all vehicles on the link *plus* those vehicles that have recently left the link but whose available space has yet to become available for use upstream.

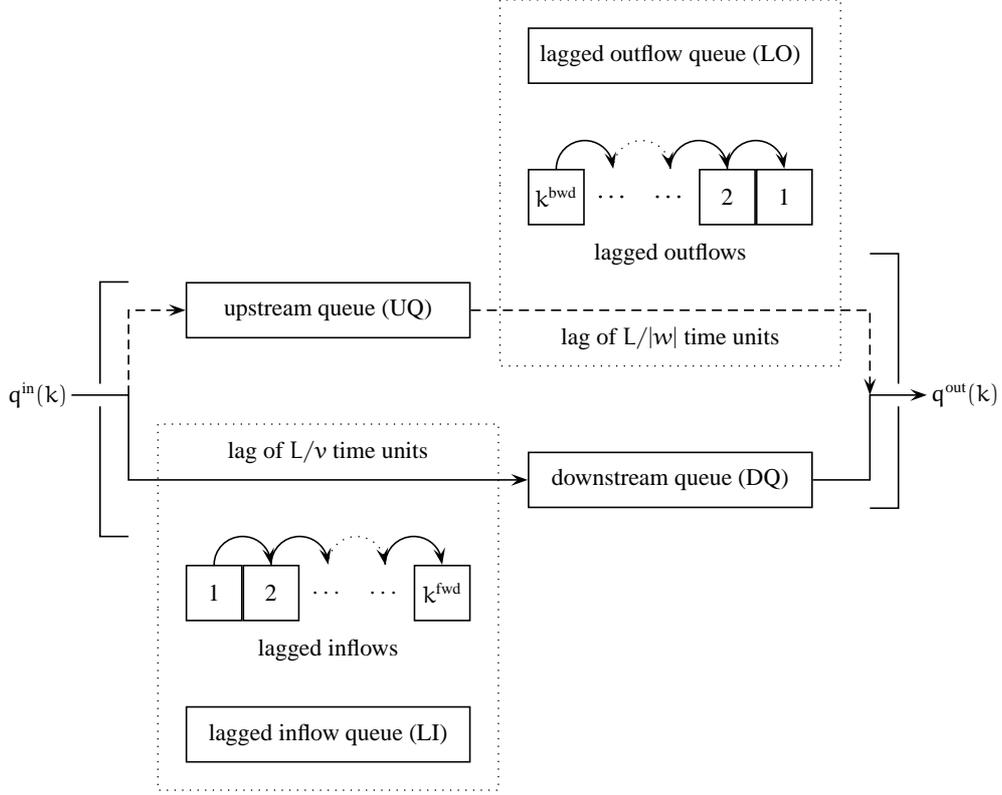


Figure 3: Lagged in/outflow buffers, aggregated into four-queue system

We only model the joint evolution of the three independent queues (LI,DQ,LO), noting that the state of the fourth queue can be deduced from (10). The state space of a link consists of all feasible values of this triplet of random variables. It is defined as:

$$\{(li, dq, lo) \in \mathbb{N}^3, li + dq + lo \leq \ell\}. \quad (11)$$

Let $p(t;k)$ denote the joint transient probability distribution of (LI,DQ,LO) at continuous time t within time interval k of duration δ . The evolution of this distribution is given in continuous time t from 0 to δ by the following linear system of differential equations (see, for instance, Reibman (1991)):

$$\frac{dp(t;k)}{dt} = p(t;k)Q(k) \quad \forall t \in [0, \delta] \quad (12)$$

where $p(t;k)$ is a probability vector and $Q(k)$ is a square matrix, known as the transition rate matrix, which is described below. Initial conditions ensure continuity at the beginning of the time interval:

$$p(0;k) = p(\delta;k-1). \quad (13)$$

The general solution to Equations (12) and (13) is given by:

$$p(t;k) = p(0;k)e^{Q(k)t} \quad \forall t \in [0, \delta]. \quad (14)$$

Equation (14) is a discrete-time differentiable expression, which guides the transition of the queue distributions from one time step to the next. It holds under the assumption of constant link boundary

initial state s	new state j	rate $Q(s, j; k)$	condition
(li, dq, lo)	$(li + 1, dq, lo)$	$\lambda(k)$	$li + dq + lo < \ell$
(li, dq, lo)	$(li - 1, dq + 1, lo)$	$\mu^{LI}(li; k)$	$li > 0$
(li, dq, lo)	$(li, dq - 1, lo + 1)$	$\mu^{DQ}(k)$	$dq > 0$
(li, dq, lo)	$(li, dq, lo - 1)$	$\mu^{LO}(lo; k)$	$lo > 0$

Table 1: Transition rates between queues LI, DQ and LO.

conditions during a time step, i.e. $Q(k)$ is constant during time interval k . This assumption was already introduced in the framework of Section 2.

For a given system of queues, $Q(k)$ is a function of the arrival rates and service rates of each of the queues. This matrix contains the transition rates between all pairs of states. The non-diagonal elements, $Q(k)_{sj}$ for $s \neq j$, represent the rate at which the transition from state s to state j takes place. The diagonal elements are defined as $Q(k)_{ss} = -\sum_{j \neq s} Q(k)_{sj}$. Thus, $-Q(k)_{ss}$ represents the rate of departure from state s .

The non-diagonal and non-null elements of the transition rate matrix are given in Table 1. Assume an initial state of (LI, DQ, LO) equal to (li, dq, lo) . The first line of the table describes arrivals to the link. They occur with rate $\lambda(k)$ and may enter the link as long as $li + dq + lo < \ell$, i.e. they may enter as long as UQ is less than the space capacity, ℓ . Flow from LI to DQ (line two of the table) is transmitted with rate $\mu^{LI}(li; k)$, and this can occur as long as LI is nonempty ($li > 0$). Line three describes departures from the link. They occur at rate $\mu^{DQ}(k)$ as long as DQ is nonempty. The last line describes departures from LO, which occur at rate $\mu^{LO}(lo; k)$.

The link dynamics are described via the queueing parameters $\lambda(k)$, $\{\mu^{LI}(li; k)\}_{li=1, \dots, \ell}$, $\mu^{DQ}(k)$, and $\{\mu^{LO}(lo; k)\}_{lo=1, \dots, \ell}$, which are derived in the following.

- $\lambda(k)$ is exogenous to the single-link model considered here.
- $\mu^{DQ}(k)$ defines the rate at which flow may leave the link. It also is considered exogenous in this paper.
- $\mu^{LI}(li; k)$ is the rate at which the LI queue discharges into DQ, given that LI contains li vehicles.

At the beginning of time interval k , queue LI contains all arrivals to the link at time intervals $k - k^{fwd}$, $k - k^{fwd} + 1$, ..., $k - 1$. It represents a sequence of k^{fwd} buffer cells, where cell j contains the entries to the link during time interval $k - j$. The number of vehicles in LI is the sum of the vehicles in these k^{fwd} cells. The vehicles that can leave LI during time interval k are those that are in LI's last (i.e. most downstream) cell, which is denoted LLI. That is, LLI represents the k^{fwd} th buffer cell of LI. Hereafter, we use $LI(k)$ to denote $LI(0; k)$, i.e. the number of vehicles in LI at the beginning of time interval k ; an according notation is used for all other time-dependent quantities as well.

The flow from LI to DQ during time interval k is given by the number of vehicles in LLI at the beginning of time interval k , i.e. $LLI(k)$. We proceed by deriving $E\{LLI(k) \mid LI(k) = li\}$, which is the expected flow transferred from LLI into DQ during time interval k , given that LI contains li vehicles.

The arrival process to the link during time interval k is a Poisson process with rate $\lambda(k)$. Arrivals may enter the link as long as it has not spilled back. This occurs with probability $P(UQ(k) < \ell)$. Thus, the vehicles enter the link according to a Poisson process with rate

$$q^{\text{in}}(k) = \lambda(k)P(UQ(k) < \ell), \quad (15)$$

where $q^{\text{in}}(k)$ represents the expected inflow to the link during time interval k . Updating the arrival rates across time intervals allows the model to account for temporal dependency between arrivals. This occurs, for instance, in a network context without losses, where vehicles that were blocked (i.e. could not enter) in previous time steps enter in later time steps. Additionally, temporal dependence is introduced through UQ , which modulates the inflow through its probability of not being full in Equation (15) and evolves relatively slowly along the time axis.

The only simplifying assumption made here and in the following is to neglect the stochastic temporal dependence between inflows (and, as explained later, outflows) at different time steps. The experiments of Section 4 demonstrate the very minor effect of this approximation. Given this, the arrivals to each inflow buffer cell constitute independent Poisson processes with corresponding rates $q^{\text{in}}(k-1), \dots, q^{\text{in}}(k-k^{\text{fwd}})$, and $LI(k)$ is the sum of these independent processes. Thus, the conditional distribution of $LLI(k)$ given that $LI(k) = li$ is Binomial $B(li, r(k))$ with

$$r(k) = \frac{q^{\text{in}}(k - k^{\text{fwd}})}{\sum_{j=1}^{k^{\text{fwd}}} q^{\text{in}}(k - j)} \quad (16)$$

being the probability of encountering a randomly selected vehicle from $LI(k)$ in $LLI(k)$. For a derivation of this result (i.e., the Binomial distribution and its corresponding parameters) see Section 2.12.4 of Larson and Odoni (1981).

In consequence,

$$E\{LLI(k) \mid LI(k) = li\} = li \cdot r(k). \quad (17)$$

The service rate $\mu^{\text{LI}}(li; k)$ can now be derived. By definition, the service rate is the inverse of the expected time between successive departures from LI . In order to derive this expectation, we observe that departures from LI form a Poisson process, and the expected number of departures given that LI has li vehicles is given by $E\{LLI(k) \mid LI(k) = li\}$.

Given that m events of a Poisson process have occurred during a fixed time interval δ , then inter-event times are independently, uniformly distributed over the fixed time interval of interest with expected inter-arrival time δ/m (see, for instance, Section 2.12.3 of Larson and Odoni (1981)). In our case, $m = E\{LLI(k) \mid LI(k) = li\}$, such that the service rate becomes

$$\mu^{\text{LI}}(li; k) = \frac{li}{\delta} \cdot \frac{q^{\text{in}}(k - k^{\text{fwd}})}{\sum_{j=1}^{k^{\text{fwd}}} q^{\text{in}}(k - j)}. \quad (18)$$

- $\mu^{\text{LO}}(lo; k)$ is the rate at which “spaces” resulting from downstream vehicle departures become available upstream. Queue LO contains all departures from DQ at time intervals $k - k^{\text{bwd}}, k -$

parameter	value
v	36 km/h
w	-18 km/h
$\hat{\rho}$	200 veh/km
\hat{q}	2400 veh/h = 0.67 veh/s
$\mu(k)$	1080 veh/h = 0.3 veh/s
$\lambda(k)$	varies by experiment
$\ell, L, k^{\text{fwd}}, k^{\text{bwd}}$	varies by experiment

Table 2: Link parameters

Time interval k:	[0,999]	[1000,1999]	[2000,2999]
Profile 131	0.1	0.3	0.1
Profile 151	0.1	0.5	0.1
Profile 353	0.3	0.5	0.3

Table 3: Arrival rate profiles, $\lambda(k)$, in veh/s

$k^{\text{bwd}} + 1, \dots, k - 1$. Symmetrically to the derivation that led to Equation (18), the departure rate from LO given that LO contains l_0 vehicles is

$$\mu^{\text{LO}}(l_0; k) = \frac{l_0}{\delta} \cdot \frac{q^{\text{out}}(k - k^{\text{bwd}})}{\sum_{j=1}^{k^{\text{bwd}}} q^{\text{out}}(k - j)}. \quad (19)$$

In summary, the overall link model is solved by repeated evaluations of Equation (14), using the exogenous parameters $\lambda(k)$ and $\mu^{\text{DQ}}(k)$ and the endogenous transmission rates defined in Equations (18) and (19). The linkage between Equation (14) and these rates is given through Table 1.

4 Experiments

A single-lane link with parameters shown in Table 2 is considered. Nine experiments are conducted, combining three different arrival rate profiles and three different link lengths (and, hence, different space capacities and forward/backward lags).

Each experiment starts with an initially empty link at time zero and runs for 3000 one-second time steps. The arrival profiles are displayed in Table 3. Profile 131 (resp. 151) corresponds to a step-change from undercritical to marginally critical (resp. overcritical) conditions and back. Profile 353 corresponds to a step-change from marginally critical to overcritical conditions and back.

The considered space capacities are $\ell = 10, 20, 30$, resulting in link lengths $L = 50, 100, 150$ m, forward time lags $k^{\text{fwd}} = 5, 10, 15$ and backward time lags $k^{\text{bwd}} = 10, 20, 30$. Table 4 labels the experiments for the resulting nine parameter combinations as concatenations of the respective arrival profile and space capacity.

λ -profile ℓ	131	151	353
10	“Exp 131 Cap 10”	“Exp 151 Cap 10”	“Exp 353 Cap 10”
20	“Exp 131 Cap 20”	“Exp 151 Cap 20”	“Exp 353 Cap 20”
30	“Exp 131 Cap 30”	“Exp 151 Cap 30”	“Exp 353 Cap 30”

Table 4: Experiments

λ	0.1 veh/s	0.3 veh/s	0.5 veh/s
ℓ			
10	0.57	0.68	0.52
20	0.45	0.76	0.50
30	0.38	0.81	0.46

Table 5: Stationary correlations between UQ and DQ

Particular attention is paid in the following to the stochastic dependency between up- and downstream conditions within the link, corresponding to dependency between UQ and DQ. For this, the results of the proposed analytical model are compared to empirical distributions obtained from 10^6 replications of an event-based microsimulation.

This microsimulation implements only an UQ and a DQ. It does not resort to the LI and LO approximation but instead explicitly implements the time lags experienced by vehicles entering the link until they reach DQ and by spaces becoming available due to exiting vehicles until they reach UQ. The stochasticity of the arrival and service process is explicitly simulated by drawing corresponding random numbers. That is, the microsimulation implements an instance of the proposed model that comes with no approximations of the time lags – at the cost of being able to only draw from the underlying distributions (as opposed to an analytical approach). Since the microsimulation perfectly captures all dependencies, it serves as a benchmark for the analytical model.

Figure 4 shows for all nine experiments the evolution of the correlation between UQ and DQ over time. The red crosses represent results from the analytical model, and the blue circles represent results from the event-based simulation. Figure 5 (resp. 6) shows in greater detail the transient dynamics of the correlation around second 1000 (resp. 2000). As a first impression, the deviations between simulation and analytical model are visually negligible, indicating an excellent overall fit.

Table 5 contains the stationary correlations between UQ and DQ for the three space capacities $\ell = 10, 20, 30$ and the three stationary arrival rates $\lambda = 0.1, 0.3, 0.5$ veh/s. Figures 7(a),(g),(d) show the corresponding (UQ,DQ) distributions for $\ell = 30$, obtained with the analytical model. The correlation values can be interpreted based on the joint distributions in the following way.

- All correlations are positive and quite large. This is plausible given that both UQ and DQ represent aspects of the link’s occupancy and have substantial overlap, cf. also Equation (10).
- For each given space capacity ℓ , the correlation is in marginally critical conditions higher than in under- or overcritical conditions. Under- and overcritical conditions differ from marginally critical conditions in this regard because Equation (10) implies that UQ is always fuller than

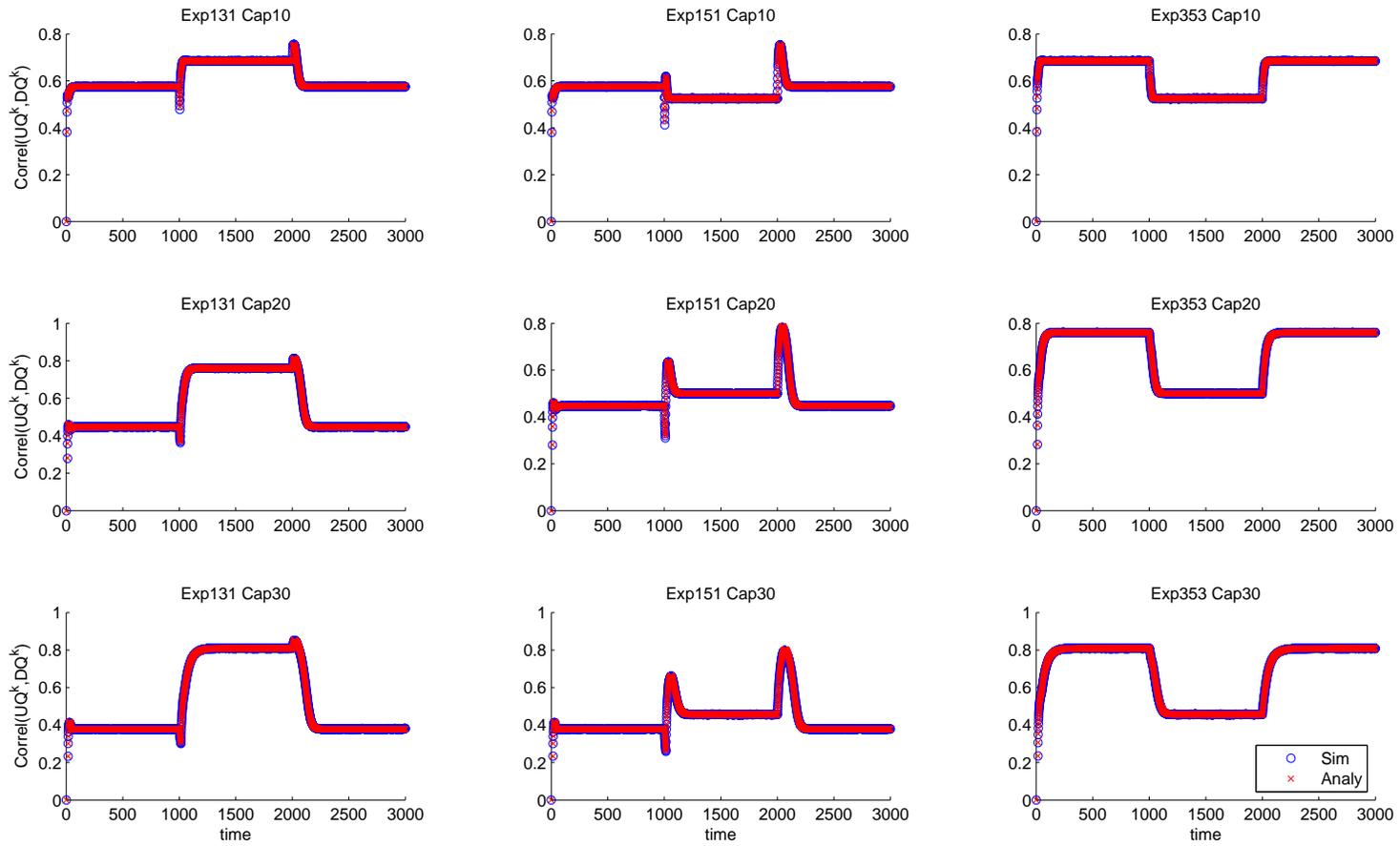


Figure 4: Correlation between UQ and DQ over time

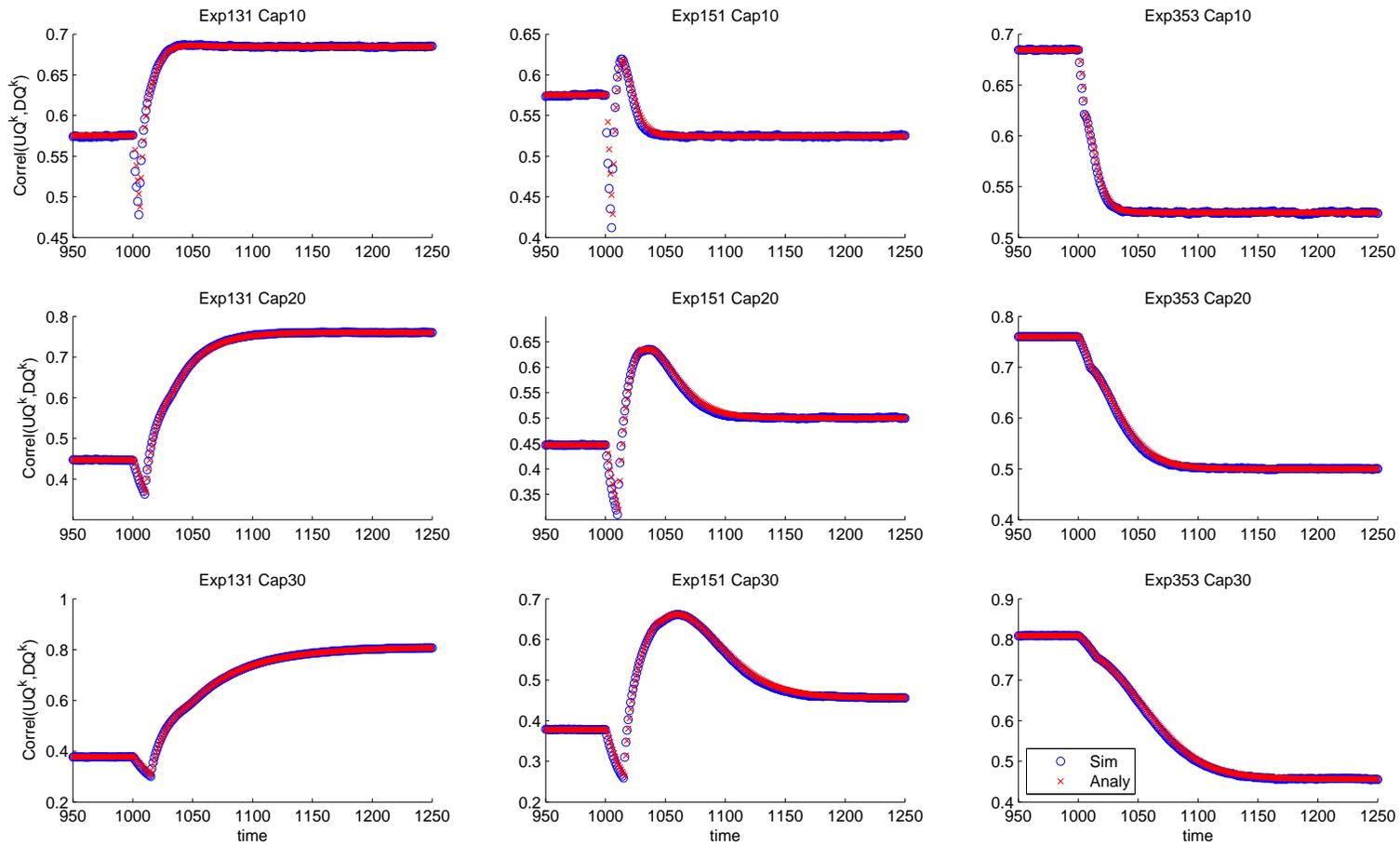


Figure 5: Correlation between UQ and DQ during the transition around second 1000

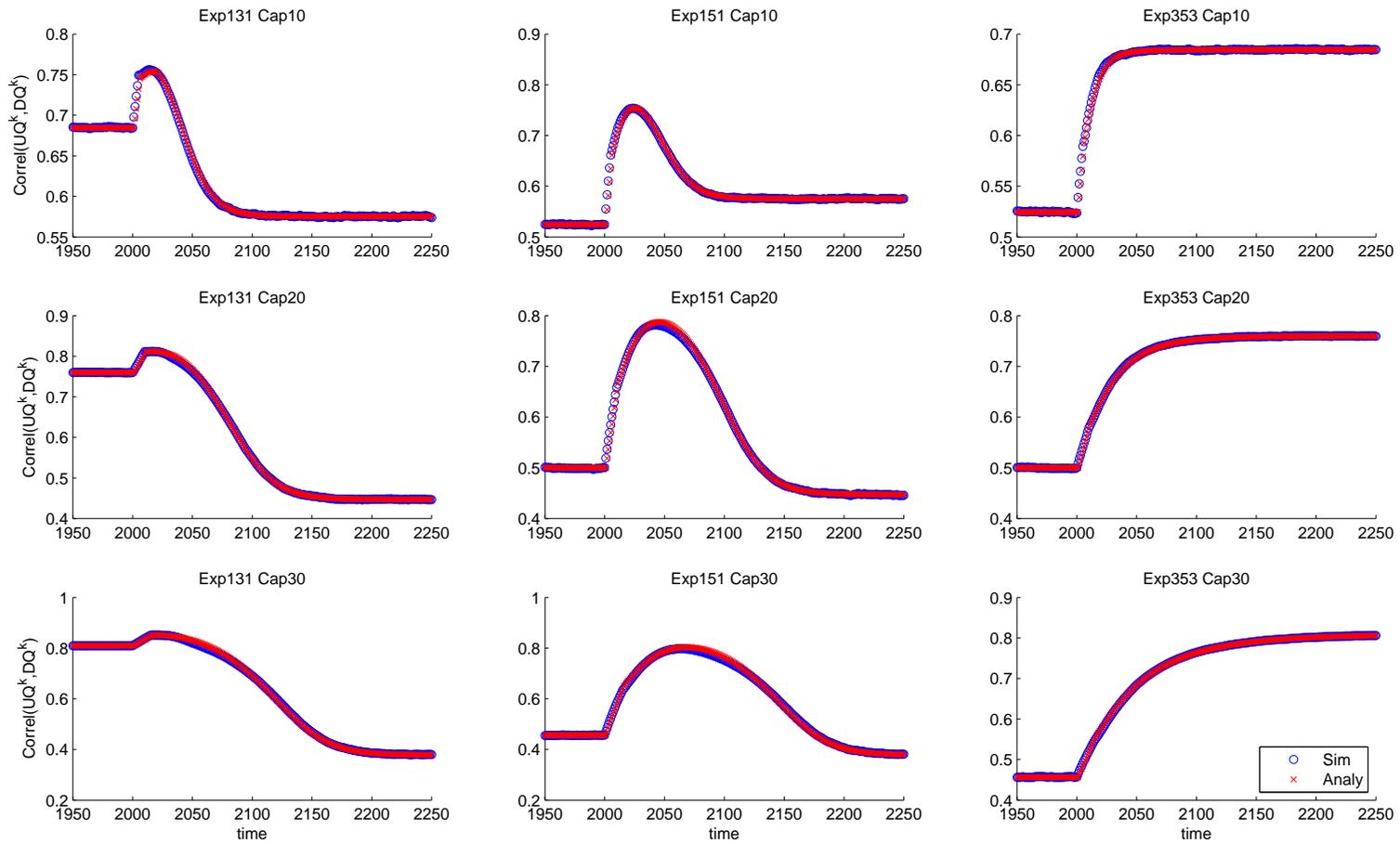


Figure 6: Correlation between UQ and DQ during the transition around second 2000

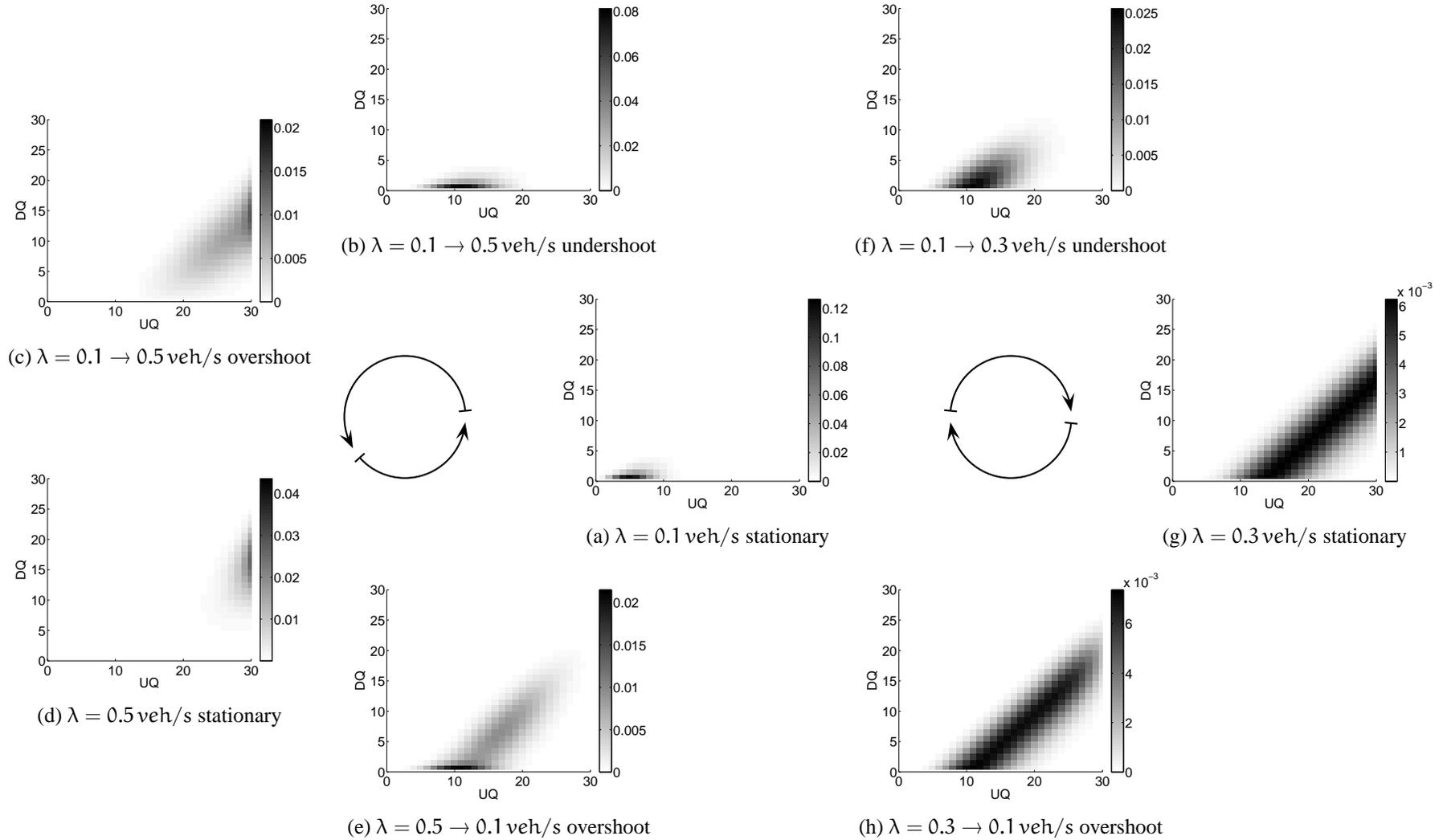


Figure 7: (UQ,DQ) joint distributions. Figures (a), (g), and (d) show stationary distributions for $\lambda = 0.1, 0.3, 0.5$ veh/s, respectively. Arranged between these figures in circles along the indicated directions are (UQ,DQ) distributions at the moments of largest under- and overshoot during the respective transients, cf. Figures 4, 5, and 6. All results are obtained with the analytical model and $\ell = 30$.

DQ. In undercritical conditions, this limits joint downwards fluctuations of UQ and DQ because DQ is already close to zero; the (UQ,DQ) distribution is truncated at DQ=0 in Figure 7(a). In overcritical conditions, this limits joint upwards fluctuations because UQ is already close to ℓ ; the (UQ,DQ) distribution is truncated at UQ= ℓ in Figure 7(d). In marginally critical conditions, the probability of these bounds taking effect is relatively low, allowing joint fluctuations of UQ and DQ to occur most freely in Figure 7(g).

- For $\lambda = 0.1$ and 0.5 veh/s, the correlation decreases with increasing ℓ . This is so because in undercritical (resp. overcritical) conditions, DQ (resp. UQ) dictates the link dynamics, and the respective other queue follows. The longer the link is, the more room exists for independent fluctuations of the queues, resulting in reduced correlation. In terms of Equation (10), the LI + LO addend of $UQ = DQ + (LI + LO)$ increases relative to and somewhat independently of DQ, moving away from what would be a perfect $UQ = DQ$ dependency.
- For $\lambda = 0.3$ veh/s, the correlation increases with increasing ℓ . As explained before, the distributions of UQ and DQ evolve most freely in marginally critical conditions. Indeed, Figure 7(g) reveals that the (UQ,DQ) distribution stretches in marginally critical conditions all the way from undercritical to overcritical conditions. As the link gets longer, this distribution stretches even further, resulting in the observed increase in correlation.

Figure 5 (resp. Figure 6) provide a more detailed evaluation of the correlation dynamics during the transient periods around second 1000 (resp. 2000). These transitions are captured very accurately by the analytical model. Figure 7 shows snapshots of the (UQ,DQ) distribution for $\ell = 30$ during stationarity and the times of largest correlation under- and overshoot. Here, one observes the following.

- For second 1000 and arrival profile 131 (first column in Figure 5), the correlation undershoots before attaining its new stationary value. These dynamics are reflected by the sequence of (UQ,DQ) distributions shown in Figure 7(a),(f),(g). The undershoot can be explained by UQ starting to increase k^{fwd} time steps before DQ, such that UQ initially changes independently of DQ. The (UQ,DQ) distribution first stretches out only horizontally before expanding also vertically. Eventually, this effect ceases as the two queues synchronize again. Consistently with this, the time of maximum undershoot coincides with the respective forward time lag k^{fwd} .
- For second 1000 and arrival profile 151 (second column in Figure 5), the correlation first undershoots and then overshoots before attaining its new stationary value. These dynamics are reflected by the sequence of (UQ,DQ) distributions shown in Figure 7(a)(b)(c)(d). The undershoot is explained in the previous item (initial horizontal expansion of (UQ,DQ) distribution from Figure 7(a) to 7(b)). The overshoot results because, whatever the initial (most likely, upwards) fluctuation of UQ is, the same fluctuation reaches DQ after k^{fwd} time steps. The relatively empty DQ is able to follow quite freely UQ's earlier fluctuation. This is reflected by the diagonal stretch of the (UQ,DQ) distribution in Figure 7(c). Eventually, the upper bound on UQ then takes effect, reducing the probability of joint fluctuations, which results in Figure 7(d).
- For second 2000 and arrival profiles 131 and 151 (first and second column in Figure 6), the correlation overshoots before attaining its new stationary value. These dynamics are reflected by the sequence of (UQ,DQ) distributions shown in Figures 7(g)(h)(a) and 7(d)(e)(a). The

overshoot in arrival profile 151 can be explained by the link going all the way from over- to undercritical conditions, passing through a transient state of marginal criticality shown in Figure 7(e). The transient overshoot in arrival profile 131, cf. Figure 7(h), exceeds the already extreme correlation in marginally critical stationary conditions, cf. Figure 7(g). A careful comparison of these two figures reveals that the transient distribution is slightly *narrower* than the stationary distribution.

- For arrival profile 353 (last column in in Figure 5 and 6), there are no over- or undershoots. In marginally critical and overcritical conditions, UQ takes at least partial effect on the link dynamics. Since the change in boundary conditions also occurs upstream, the immediate reaction of UQ absorbs the link dynamics that in the previous cases resulted from the time-lagged interactions of UQ and DQ. This leads to a smooth transition between the band-like (UQ,DQ) distribution in marginally critical conditions (Figure 7(g)) and the more constrained overcritical distribution (Figure 7(d)).

So far, only correlation as a measure of linear dependency was considered. Figure 8 shows the joint distribution of LI, DQ and LO for different arrival profiles and at particularly interesting points in time (shortly after the jump-changes in the arrival profile). Only results for $\ell = 10$ are shown; the figures for $\ell = 20, 30$ do not reveal additional information. The horizontal axis represents the indices of the different states, and the vertical axis represents their probabilities. All feasible states of (LI, DQ, LO) are represented. One observes an almost perfect match between simulated and analytical results, across all experiments.

These experiments demonstrate an extremely high precision of the analytical model when approximating an event-based microsimulation of the exact stochastic KWM model for a homogeneous link. It hence is possible to analytically capture full link state distributions in consistency with a stochastic Newell model.

5 Summary and outlook

This article presents a new model for traffic flow along a linear link, which analytically captures queue length distributions. When compared to a previous approach, the new model adds realistic dependency structure between the link's upstream and downstream boundary conditions. In order to maintain tractability of the new model, some simplifications are adopted, which result in a negligible loss of precision when compared to a stochastic microsimulator.

The relationship of the proposed model to the kinetic theory of traffic flow may be worth investigating further. In kinetic models, stochasticity enters at the level of individual vehicle interactions, rendering the stochastic performance of a whole link an accumulation of such interactions. The present work limits itself to (demand) stochasticity at a link's upstream end and (supply) stochasticity at its downstream end. This in combination with the Poissonian assumption leads to an operational but arguably simplified representation of real traffic. An effort to derive the present model from (or to link it to) kinetic theory may enable richer distributional assumptions and may also facilitate the derivation of more operational kinetic models.

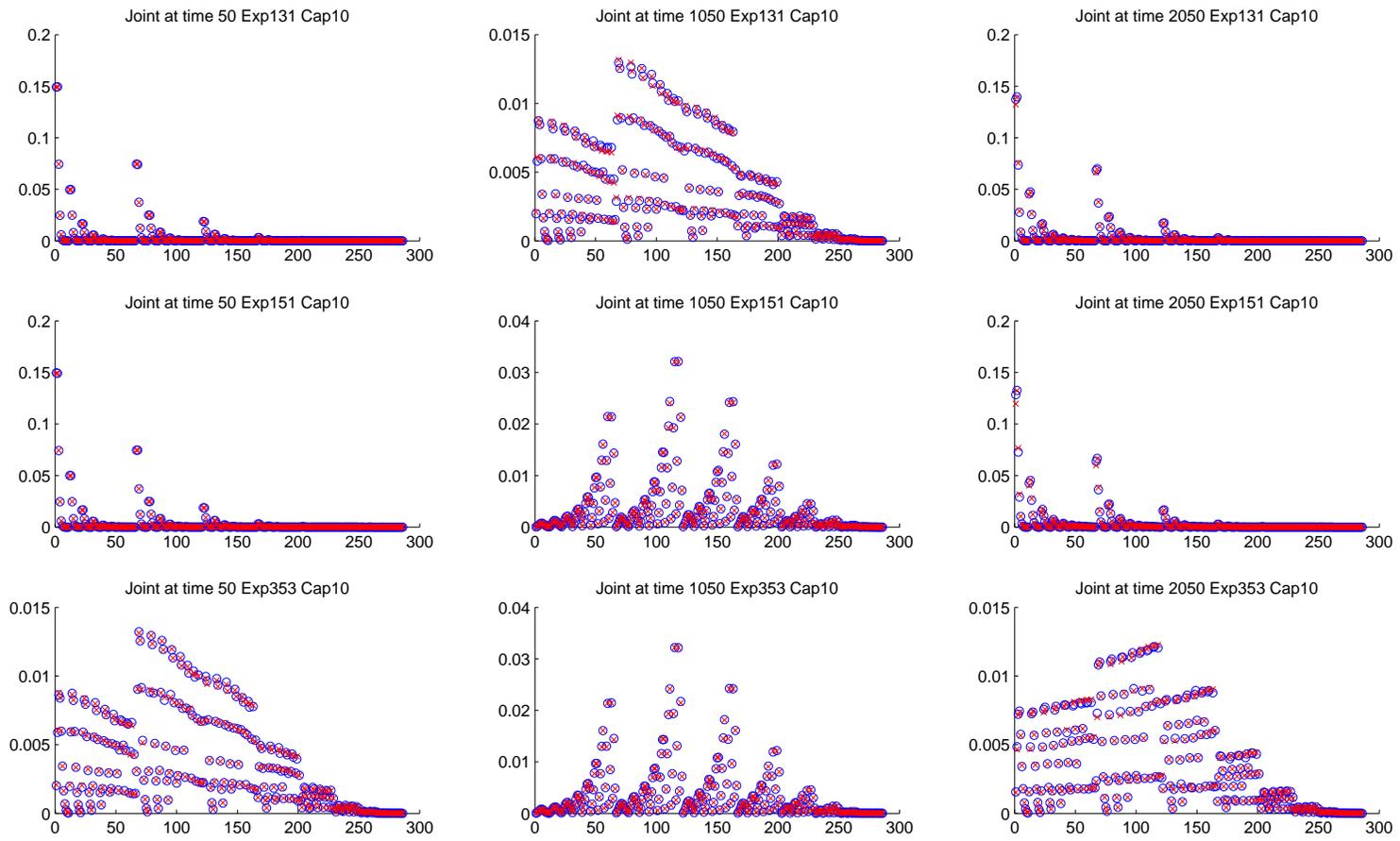


Figure 8: Joint distribution of LI, DQ and LO

Further efforts will focus on network modeling. This article describes how dependency across a homogeneous link can be captured; previous work demonstrated how dependency across a node can be captured (Osorio et al.; 2011). A logical next step is to combine these two models into an approximation of the joint queue length distributions in a complete network.

References

- Ansorge, R. (1990). What does the entropy condition mean in traffic flow theory, *Transportation Research Part B* **24**(2): 133–143.
- Boel, R. and Mihaylova, L. (2006). A compositional stochastic model for real time freeway traffic simulation, *Transportation Research Part B: Methodological* **40**: 319–334.
- Charypar, D. (2008). *Efficient algorithms for the microsimulation of travel behavior in very large scenarios*, PhD thesis, Swiss Federal Institute of Technology Zurich.
- Corthout, R., Flötteröd, G., Viti, F. and Tampere, C. (2012). Non-unique flows in macroscopic first-order intersection models, *Transportation Research Part B* **46**(3): 343–359.
- 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. (1995). A finite difference approximation of the kinematic wave model of traffic flow, *Transportation Research Part B* **29**(4): 261–276.
- Daganzo, C. (2005). A variational formulation of kinematic waves: basic theory and complex boundary conditions, *Transportation Research Part B* **39**(2): 187–196.
- Flötteröd, G., Bierlaire, M. and Nagel, K. (2011). Bayesian demand calibration for dynamic traffic simulations, *Transportation Science* **45**(4): 541–561.
- Flötteröd, G. and Rohde, J. (2011). Operational macroscopic modeling of complex urban intersections, *Transportation Research Part B* **45**(6): 903–922.
- Helbing, D. (2001). Traffic and related self-driven many-particle systems, *Rev. Mod. Phys.* **73**: 1067–1141.
- Jabari, S. and Liu, H. (2012). A stochastic model of traffic flow: theoretical foundations, *Transportation Research Part B* **46**(1): 156–174.
- Larson, R. C. and Odoni, A. R. (1981). *Urban operations research*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA.
- Lebacque, J. (1996). The Godunov scheme and what it means for first order traffic flow models, in J.-B. Lesort (ed.), *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Pergamon, Lyon, France.

- Lebacque, J. and Khoshyaran, M. (2005). First-order macroscopic traffic flow models: intersection modeling, network modeling, in H. Mahmassani (ed.), *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, Elsevier, Maryland, USA, pp. 365–386.
- Lighthill, M. and Witham, J. (1955). On kinematic waves II. a theory of traffic flow on long crowded roads, *Proceedings of the Royal Society A* **229**: 317–345.
- Nagel, K. and Nelson, P. (2005). A critical comparison of the kinematic-wave model with observational data, in H. Mahmassani (ed.), *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, Elsevier, Maryland, USA, pp. 145–163.
- Nelson, P. and Kumar, N. (2006). Point constriction, interface, and boundary conditions for kinematic-wave model, *Transportation Research Record* **1965**: 60–69.
- Newell, G. (1993). A simplified theory of kinematic waves in highway traffic, part I: general theory, *Transportation Research Part B* **27**(4): 281–287.
- Osorio, C. and Bierlaire, M. (2010). A simulation-based optimization approach to perform urban traffic control, *Proceedings of the Triennial Symposium on Transportation Analysis (TRISTAN)*, Tromsø, Norway.
- Osorio, C., Flötteröd, G. and Bierlaire, M. (2011). Dynamic network loading: a stochastic differentiable model that derives link state distributions, *Transportation Research Part B* **45**(9): 1410–1423.
- Reibman, A. (1991). A splitting technique for Markov chain transient solution, in W. J. Stewart (ed.), *Numerical solution of Markov chains*, Marcel Dekker, Inc, New York, USA, chapter 19, pp. 373–400.
- Richards, P. (1956). Shock waves on highways, *Operations Research* **4**: 42–51.
- Sumalee, A., Zhong, R. X., Pan, T. L. and Szeto, W. Y. (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.
- Tampere, C., Corthout, R., Cattrysse, D. and Immers, L. (2011). A generic class of first order node models for dynamic macroscopic simulations of traffic flows, *Transportation Research Part B* **45**(1): 289–309.
- Tampere, C., van Arem, B. and Hoogendoorn, S. (2003). Gas-kinetic flow modeling including continuous driver behavior models, *Transportation Research Record* **1852**: 231–238.
- Yperman, I. (2007). *The link transmission model for dynamic network loading*, PhD thesis, Katolieke Universiteit Leuven. Chapter 4.
- Yperman, I., Tampere, C. and Immers, B. (2007). A kinematic wave dynamic network loading model including intersection delays, *Proceedings of the 86. Annual Meeting of the Transportation Research Board*, Washington, DC, USA.