Efficient evaluation of stochastic traffic flow models using Gaussian process approximation

Storm, Pieter Jacob; Mandjes, Michel; van Arem, Bart

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Efficient evaluation of stochastic traffic flow models using Gaussian process approximation☆

Pieter Jacob Storm [a,*], Michel Mandjes [b,c,d], Bart van Arem [e]

[a] *Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*
[b] *Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*
[c] *Eurandom, Eindhoven University of Technology, Eindhoven, The Netherlands*
[d] *Amsterdam Business School, Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands*
[e] *Delft University of Technology, P.O. Box 5048, 2600 GA, Delft, The Netherlands*

ARTICLE INFO

ABSTRACT

This paper studies a Gaussian process approximation for a class of stochastic traffic flow models. It can be used to efficiently and accurately evaluate the joint (in the spatial and temporal sense) distribution of vehicle-density distributions in road traffic networks of arbitrary topology. The Gaussian approximation follows, via a scaling-limit argument, from a Markovian model that is consistent with discrete-space kinematic wave models. We describe in detail how this formal result can be converted into a computational procedure. The performance of our approach is demonstrated through a series of experiments that feature various realistic scenarios. Moreover, we discuss the computational complexity of our approach by assessing how computation times depend on the network size. We also argue that the (debatable) assumption that the vehicles' headways are exponentially distributed does not negatively impact the accuracy of our approximation.

1. Introduction

This paper focuses on the efficient numerical evaluation of stochastic road traffic models. Traditionally, the development of deterministic models has been a dominant research line (Chanut and Buisson, 2003; Hoogendoorn and Bovy, 2001; Logghe and Immers, 2008; Maerivoet and De Moor, 2005; van Wageningen-Kessels et al., 2015; Wong and Wong, 2002). For such models, owing to their relatively simple dynamics, the vehicle-density propagation can typically be evaluated in an efficient manner, also for large-scale networks. Recently, a renewed interest has been seen in studying their stochastic counterparts; see, e.g., the research lines (Jabari and Liu, 2012, 2013; Lu and Osorio, 2018, 2021; Osorio and Wang, 2017). Such stochastic models aim to incorporate the deterministic kinematic properties of traffic flow as well as its inherent stochastic fluctuations. Consequently, they can be used to capture the uncertainty of various performance indicators (such as travel times), which can ultimately be applied when developing measures that improve network robustness and reliability. Due to their intrinsic complexity, stochastic models are computationally significantly more demanding than deterministic ones. This explains the lack of stochastic traffic flow models that are consistent with deterministic traffic flow models but that at the same time can be evaluated on medium to large-scale networks, cf. the discussion in Lu and Osorio (2018). Our paper aims to bridge this gap between deterministic, reasonably tractable models on one hand, and stochastic, laborious models on the other hand.

Central in this paper is a class of stochastic traffic flow models that generalizes the prominent class of discrete-space kinematic wave models (i.e., models based on LWR theory; cf. Lighthill and Whitham (1955) and Richards (1956)). The setup considered represents a road traffic network made up of d segments, and focuses on describing the distribution of the vehicle densities in these segments as a function of time. We work with Markov dynamics that are set up in such a way that the fundamental diagram of traffic flow is respected. In Mandjes and Storm (2021) it was shown that on road segments of (roughly) several hundreds of meters, the joint vehicle-density process in the d segments can be accurately described (as a function of time, that is) by a Gaussian process. Its means and (co-)variances, as a function of time, characterize the joint vehicle-density distribution, and can be found by evaluating a coupled system of ordinary differential equations, which makes the Gaussian framework highly efficient from a computational point of view. Formally, the results of Mandjes and Storm (2021) are in terms of a *scaling limit*: under an appropriate scaling of the traffic rates and the segment lengths, the joint per-segment vehicle-density process converges in distribution to a Gaussian process.

The translation of the results derived in Mandjes and Storm (2021) into practical computation recipes brings in its own challenges. It involves expressing a practical setting into the model's (hyper-)parameters and features (on-ramps, merges, etc.), and in addition various coupled systems of differential equations have to be numerically evaluated. The key objective of this paper is to provide a detailed description and discussion of the underlying computational aspects, as well as an in-depth account of the method's pros and cons. Concretely, this requires various non-trivial steps in which the formal convergence-in-distribution result of Mandjes and Storm (2021) is to be converted into an approximation of the joint per-segment vehicle-density process. The goal is to demonstrate how, by applying standard numerical software only, both the intra-segment correlations and the temporal correlations can be efficiently and accurately evaluated for networks of considerable size. A second objective concerns an assessment of the method's complexity. Concretely, we aim to systematically analyze the impact of the model parameters on the computation time.

We proceed by briefly reviewing related branches in the literature, after which we provide more background on our approach and contributions, and on how these relate to the existing literature. As mentioned, most traffic flow models in the literature are deterministic, but the stochastic nature of road traffic has been widely recognized, cf. the accounts in Jabari et al. (2014), Ngoduy (2021) and Qu et al. (2017).

There have been several attempts to model the observed fluctuations in traffic flows (or fundamental diagram scatter) by means of macroscopic stochastic models. A frequently followed approach works with the stochastic cellular transmission model (CTM), two variants of which have been explored in the literature. In the first variant one starts with a deterministic CTM and replaces the model parameters with random variables (Boel and Mihaylova, 2006; Li et al., 2012; Ngoduy, 2011; Panda et al., 2019; Sumalee et al., 2011; Zhong et al., 2013). In Lu and Osorio (2018) it is argued that the numerical evaluation of these models can become computationally intensive for large-scale networks, due to the sampling that is required. The second variant is based on a stochastic model in which the vehicle headway is seen as a random quantity. This approach, originally proposed in Jabari and Liu (2012, 2013), was generalized to incorporate multi-class fundamental diagrams and networks in Mandjes and Storm (2021). It is the latter approach that we follow in the present paper, its major asset being that it does not require any sampling; instead, as will be extensively demonstrated, it allows highly efficient numerical evaluation.

The stochastic link transmission model (LTM) is another approach of macroscopic stochastic traffic flow modeling, that builds on Newell's theory of kinematic waves in traffic (Newell, 1993). By recognizing that cumulative flows of vehicles across link boundaries fit the mathematical framework of (fictitious) queues inside a link, Osorio and Flötteröd (2015) succeeds in characterizing the per-link distribution of the number of vehicles using matrix exponential methods. In Lu and Osorio (2018) the model evaluation speed is significantly increased by using simplifying assumptions that allow for explicit formulas for the matrix exponential. The recent paper (Lu and Osorio, 2021) extends the model to networks.

Another class of stochastic models relies on the microscopic traffic flow modeling paradigm. The evaluation of these models is typically even more computationally intensive than the ones we discussed thus far, as it relies on simulation. With the key example being the class of stochastic car-following models (Laval and Leclercq, 2008; Laval et al., 2014; Lee et al., 2019; Ngoduy et al., 2019; Treiber and Kesting, 2017; Yuan et al., 2018), their focus lies on reproducing traffic phenomena such as capacity drop or traffic oscillations, and less on the evaluation of vehicle-density distributions in large-scale networks. Another class of microscopic models uses stochastic traffic cellular automata, as reviewed in Maerivoet and De Moor (2005), with Nagel and Schreckenberg (1992) presenting the seminal model.

Recently, with the increasing availability of highly granular data, data-driven approaches have been devised to predict vehicle densities, see, e.g., Ahmed et al. (2021) and Yuan et al. (2021). For instance, using UAV-images (Ahmed et al., 2021) one could incorporate a vast amount of information into predictions. Such a procedure has the potential to lead to highly accurate (online) predictions of vehicle densities. Relative to the use of mathematical models (such as the ones discussed above) it has the disadvantage of not providing fundamental insights and understanding.

The above overview indicates the need for stochastic traffic flow models that allow an efficient and accurate evaluation of the joint (in the spatial and temporal sense) distribution of vehicle-density distributions. These should work for road traffic networks of arbitrary topology, and should be able to incorporate a broad array of relevant features, such as speed limits, traffic heterogeneity, etc. As mentioned, in this paper we will adopt the framework developed in Mandjes and Storm (2021).

Now we provide a detailed account of the paper's contributions, also commenting on the relation with the context that has been provided above. Our work should be seen in the tradition of the pioneering works (Jabari and Liu, 2012, 2013), aiming at evaluating the probability distribution of the numbers of vehicles simultaneously present in the individual segments. In this approach it is implicit that the proposed stochastic model necessarily inherits the average behavior that is in line with the kinematic wave models. In Mandjes and Storm (2021) we succeeded in lifting the framework of Jabari and Liu (2012, 2013) to a higher level of

realism, by, e.g., incorporating multiple traffic classes (so as to be able to model the different behavior and impact of, say, normal cars and trucks, in line with the framework developed in Chanut and Buisson (2003)).

The main finding of Mandjes and Storm (2021) was that, under a natural scaling, the numbers of vehicles per vehicle class simultaneously present in d segments, can be accurately described (as a function of time, that is) by a Gaussian process. This Gaussian process can be taken as a starting point for a traffic flow model, the mean and covariances (spatial and temporal) can be found by evaluating matrix ordinary differential equations (ODEs). Since a multivariate Gaussian distribution is completely characterized by the associated mean vector and covariance matrix, the ODEs represent an extremely efficient way of evaluating the joint vehicle-density distribution. This paper discusses the various non-trivial decisions that are to be made when one wishes to use this Gaussian framework, focusing on the computational aspects and their complexity properties.

As mentioned, the results of Mandjes and Storm (2021) are in terms of a scaling limit, which effectively means that under an appropriate scaling of the traffic rates and the segment lengths, the joint per-segment vehicle-density process converges in distribution to a d-dimensional Gaussian process. This scaling is to ensure that there is enough aggregation, concretely meaning that any road segment should be sufficiently long to guarantee that the typical number of vehicles of any type is large enough for the central limit regime to kick in. Importantly, as has been extensively explored, typically central limit based approximations do not require an excessive scale in order to be reasonably accurate; as a rule of thumb, it suffices that any segment can accommodate roughly 15–20 vehicles of any type. The above in particular means that at any point in time, the joint distribution of all traffic densities is multivariate normal, as long as the level of aggregation is sufficient. In this paper we show that our Gaussian approximation can evaluate the joint per-cell vehicle-density distribution highly accurately and efficiently for networks of considerable size. The approximation nicely decomposes the deterministic component representing the system's mean behavior, which complies with the associated kinematic wave model, from the stochastic component, which reflects the fluctuations around the mean. In this sense our approximation inherits the structure of classical central-limit type approximations, where an aggregate of random quantities is approximated by the corresponding mean increased by a zero-mean normal random variable.

Finally, we provide a number of realistic experiments to demonstrate the potential of the Gaussian process approach. Their focus is on the effect of traffic surges (e.g. at on-ramps) on congestion, on the way in which the effect of accidents propagates, and combinations of these two elements. An extremely useful property of our approach is that it is straightforward to evaluate the *combined* impact of various traffic phenomena and control measures, as we demonstrate in one of our experiments. In particular, any cell can potentially be endowed with a different fundamental diagram to capture relevant properties specific to local traffic flows. In contrast, many existing traffic models focus on the effect of a single phenomenon or control measure, cf. Papageorgiou et al. (2003). We also include a complexity study, in that we systematically evaluate the impact of the model parameters on the computation time. Our method, being based on an underlying Markov model, uses the assumption that that cell sojourn-times are exponentially distributed (with congestion-dependent rates), but we show through a series of experiments that the performance of our method hardly degrades when these sojourn-times actually stem from other distributions.

The organization of this paper is as follows. In Section 2 we present the model from Mandjes and Storm (2021), and point out how the formal results are translated into our Gaussian approximation. We also discuss the advantages of our approach over competing computational techniques. In Section 3 we present a series of experiments in which the vehicle-density distribution is evaluated (jointly at multiple locations and multiple points in time, that is). Additional properties of the approach are presented in Section 4; this includes (i) a complexity study and (ii) an assessment of the robustness of the methodology under different choices for the cell lengths in the underlying road network. Section 5 reflects on our approach, in particular showing that the assumption of exponentially distributed headways can be relaxed while hardly compromising the model's accuracy. Finally, Section 6 presents concluding remarks.

2. Model, limit results, and methodological advantages

In the first part of this section we summarize the traffic flow model of Mandjes and Storm (2021), as well as the main results obtained there. We point out how this framework allows for an efficient assessment of stochastic traffic flows on networks of arbitrary topology using a Gaussian approximation. We then present a detailed discussion on the how to utilize the Gaussian approximation in a practical context, in particular explaining how to translate a specific choice for the model's (hyper-)parameters and features (on- and off-ramps, merges, diverges, etc.). We conclude this section by discussing the benefits of using our Gaussian approximation over alternative methods (which will be further backed in Section 3 through a series of representative experiments).

2.1. Model summary

We start by providing a compact model description; see Mandjes and Storm (2021, Section3) for more details. For ease of exposition we here focus on a segment consisting of a sequence of cells, modeling a road segment without any intermediate on-ramps and off-ramps, rather than a general network. As pointed out in Mandjes and Storm (2021, Section 7.1), however, this setup naturally extends to networks of arbitrary topology.

The segment is divided into $d \in \mathbb{N}$ cells, with cell i having length $\ell_i > 0$, for $i \in \{1, \ldots, d\}$. An important asset of our model is that we allow for multiple vehicle types; in the sequel we let $m \in \mathbb{N}$ denote the number of types. We focus on probabilistically describing the objects $X_{ij}(t)$, defined as the number of type-j vehicles in cell i at time $t \geqslant 0$, for $j \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, d\}$. The random variable $X_{ij}(t)$ attains values in $\{0, 1, \ldots, x_{ij}^{\text{jam}}\}$, where $x_{ij}^{\text{jam}} \in \mathbb{N}$ is an upper bound on the number of type-j vehicles that

can be simultaneously present in cell i. We denote by $X(t)$ the md-dimensional random vector with entries $X_{ij}(t)$. Define the type-j vehicle density in cell i at time $t \geqslant 0$ by

$$\rho_{ij}(t) := \frac{X_{ij}(t)}{\ell_i},$$

attaining values in $\{0, 1/\ell_i, \ldots, x_{ij}^{\mathrm{jam}}/\ell_i\}$, where $\rho_{ij}^{\mathrm{jam}} := X_{ij}^{\mathrm{jam}}/\ell_i$ can be interpreted as the jamming vehicle density. Analogously to $X(t)$, we let $\rho(t)$ be the md-dimensional random vector with entries $\rho_{ij}(t)$.

Vehicles traverse the consecutive cells, starting at cell 1 and ending at cell d. In a *cellular transition model* (CTM), vehicle-mass moves across cell boundaries at a rate that is given by a *discrete flux function*. This function, whose arguments are the vehicle densities in the sending and receiving cells (of all m types), is derived from a macroscopic fundamental diagram (MFD) by solving an associated Riemann problem (Mandjes and Storm, 2021, Section 3.2). More formally, this means that the discrete flux function from cell i to cell $i+1$ is given by a function of the type

$$\tilde{q}_i : \underset{j=1}{\overset{m}{\bigtimes}} \left([0, \rho_{ij}^{\mathrm{jam}}] \times [0, \rho_{i+1,j}^{\mathrm{jam}}] \right) \to \underset{j=1}{\overset{m}{\bigtimes}} [0, q_{ij}^{\mathrm{max}}].$$

In our setup, we consider stochastic inter-cell transition times, that we choose in such a way that the *mean* dynamics correspond with a CTM. More precisely, the time it takes a type-j vehicle to move from cell i to cell $i+1$ is an exponentially distributed random variable, with a mean that is in line with the discrete flux function (and thus depends on the value of the vehicle-density process in cells i and $i+1$). In addition to vehicles jumping between cells, vehicles enter the segment at cell 1 and depart from the segment at cell d. These transitions are to be handled slightly differently from the inter-cell transitions. Concretely, the type-j arrival rate at cell 1 depends on the vehicle densities in cell 1 but is in addition bounded by a given rate λ_j. Likewise, the type-j departure rate from cell d is a function of the vehicle densities in cell d, truncated at v_j. Since in our framework the transition times are exponentially distributed, the process under study is a continuous-time Markov chain.

We denote by $q_{0j}(\rho(t))$, $q_{ij}(\rho(t))$ and $q_{dj}(\rho(t))$, respectively, the arrival rate at cell 1, transition rate between cell i and cell $i+1$, and departure rate from cell d, given the state of the system $\rho(t)$, all of them pertaining to vehicles of type j; here $i \in \{1, \ldots, d-1\}$ and $j \in \{1, \ldots, m\}$. In the analysis presented in Mandjes and Storm (2021) we heavily rely on a description of the type-j density in cell i ($\rho_{ij}(t)$, that is) in terms of the densities in preceding cells in the past ($\rho(s)$ with $s \leqslant t$, that is). Concretely,

$$\rho_{ij}(t) = \rho_{ij}(0) + \frac{1}{\ell_i} \int_0^t Y_{i-1,j}\left(\int_0^t q_{i-1,j}(\rho(s))\,\mathrm{d}s \right) - \frac{1}{\ell_i} \int_0^t Y_{ij}\left(\int_0^t q_{ij}(\rho(s))\,\mathrm{d}s \right), \tag{1}$$

where $Y_{ij}(\cdot)$ are independent unit-rate Poisson processes, with $j \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, d\}$. Here $Y_{ij}(t)$ can be interpreted as the cumulative number of arrivals to cell i of type j over the interval $[0, t]$. We denote by $Y(\cdot)$ the process with entries $Y_{ij}(\cdot)$, for $i \in \{0, \ldots, d\}$, $j \in \{1, \ldots, m\}$.

2.2. Scaling limit

It is important to observe that, for realistic instances, the Markov chain defined in the previous section has an excessively large state space. As a consequence, direct numerical evaluation of performance metrics is typically not feasible; we get back to this issue in Section 2.3.2. The main idea presented in this section, is to approximate the random objects under study by a suitably chosen Gaussian counterpart that *does* allow efficient numerical evaluation. The formal backing of this procedure is given by the scaling limits presented in Mandjes and Storm (2021, Section 4), which we briefly summarize here.

In our approach we scale the cell lengths by a factor n, i.e., $\ell_i \mapsto n\ell_i$. Here, we let the scaling parameter n grow large to obtain central-limit theorem type of results. We emphasize, however, the well-known fact that typically the corresponding Gaussian approximations are already accurate for relatively low values of n (i.e., it suffices if there are on average a couple of tens of vehicles per segment). Simultaneously, we scale time by a factor n, i.e., $t \mapsto nt$, so that the expected flow of density between cells per unit of time remains invariant. We denote $\rho^n(t) := \rho(nt)/n$, with the cell lengths being given by $n\ell_i$, for $i \in \{1, \ldots, d\}$.

To keep notation light, let $Q(\rho(t))$ be the vector of length $(d+1)m$ with entries $q_{i-1,j}(\rho(t))$, $i \in \{1, \ldots, d+1\}$, $j \in \{1, \ldots, m\}$, ordered lexicographically, i.e., $Q_{(i-1)m+j} = q_{i-1,j}$. We define H to be the $dm \times (d+1)m$ matrix with $H_{kl} := \mathbb{1}_{\{k=l\}} - \mathbb{1}_{\{k+m=l\}}$. Finally, we set L as the $dm \times dm$-dimensional diagonal matrix, with the kth diagonal element being $1/\ell_i$ if $\lceil k/m \rceil = i$, for $k \in \{1, \ldots, dm\}$ and $i \in \{1, \ldots, d\}$. As pointed out in Mandjes and Storm (2021, Assumption 3.1), to formally prove the Gaussian convergence results below, a mild regularity condition has to be imposed on the functions $\tilde{q}_i$.

The first result is what is called a *fluid limit*, which can be seen as a law of large numbers at the process level. It states that if $\rho^n(0) \to \bar{\rho}(0)$, then, as $n \to \infty$,

$$\rho^n(t) \to \bar{\rho}(t) = \bar{\rho}(0) + \int_0^t F(\bar{\rho}(s))\,\mathrm{d}s, \tag{2}$$

with $F(\rho(t)) := LHQ(\rho(t))$; for the precise statement we refer to Mandjes and Storm (2021, Thm. 4.1). The proof extensively exploits the representation (1).

The second result is what is usually referred to as a *diffusion limit*, i.e., a central limit theorem at the process level. Suppose that $\lim_{n\to\infty} \sqrt{n}|\rho^n(0) - \rho_0| = 0$ for some ρ_0. Then it states that the process $\hat{\rho}^n(\cdot)$, defined through $\hat{\rho}^n(t) := \sqrt{n}(\rho^n(t) - \bar{\rho}(t))$, converges in distribution (as $n \to \infty$) to the process $\hat{\rho}(\cdot)$ solving the stochastic differential equation

$$\hat{\rho}(t) = \hat{\rho}(0) + \int_0^t \partial F(\bar{\rho}(s))\,\hat{\rho}(t)\mathrm{d}s + \int_0^t LH\,\Sigma(\bar{\rho}(s))\mathrm{d}B(s); \tag{3}$$

see for details (Mandjes and Storm, 2021, Thm. 4.3). Here $\partial F(\rho(t))$ is to be understood as the matrix of weak partial derivative of $F(\bar{\rho}(t))$, $\Sigma(\bar{\rho}(s))$ is the diagonal matrix with the square roots of $Q(\bar{\rho}(t))$ as entries, and $B(\cdot)$ is a vector of length $(d + 1)m$ of independent standard Brownian motions. This result was established again using the representation (1), in combination with results from martingale theory.

It is known that $\hat{\rho}(\cdot)$, as defined through (3), defines a Gaussian process. For our scaled system, it thus leads to the following approximation of the scaled joint vehicle-density process:

$$\rho^n(t) \approx \bar{\rho}(t) + \frac{1}{\sqrt{n}}\,\hat{\rho}(t).$$

The first term, which is deterministic, reflects the process' mean behavior, and has been constructed such that it complies with the associated kinematic wave model. The second term, which is stochastic, quantifies the process' inherent fluctuations around the mean. Observe the similarity with the commonly used central-limit type of approximations: with $X_1, \ldots, X_n$ a sequence of independent and identically distributed random variables with mean μ and variance σ^2, we approximate

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx \mu + \frac{1}{\sqrt{n}}\,\sigma\,N,$$

with N a standard normal random variable. One could thus say that our Gaussian approximation is a process level generalization of the conventional central-limit based normal approximation. A formal backing of the above reasoning is given in Mandjes and Storm (2021, Section 6.1.1).

The fact that $\hat{\rho}(\cdot)$ is a Gaussian process in particular entails that all linear combinations of the $\hat{\rho}_{ij}(t_k)$, for time points t_k, have a Gaussian distribution. The corresponding mean vector and covariance matrix are defined as

$$M(t) := \mathbb{E}[\hat{\rho}(t)], \quad \Gamma(s,t) := \mathrm{cov}\,(\hat{\rho}(s), \hat{\rho}(t)) = \mathbb{E}[(\hat{\rho}(s) - M(s))(\hat{\rho}(t) - M(t))^\top].$$

The covariance matrix of $\hat{\rho}(t)$ is denoted by $V(t) := \mathrm{cov}(\hat{\rho}(t), \hat{\rho}(t)) = \Gamma(t,t)$. As in Karatzas and Shreve (2012, Section 5.6, Problems 6.1, 6.2), $M(t)$ and $\Gamma(s,t)$ satisfy the explicit expressions

$$M(t) = \bar{\Phi}(t)\left[M(0) + \int_0^t \bar{\Phi}^{-1}(s)\mathrm{d}s\right],$$

$$\Gamma(s,t) = \bar{\Phi}(s)\left[V(0) + \int_0^{t\wedge s} \bar{\Phi}^{-1}(u)LH\,\Sigma(\bar{\rho}(u))\left(\bar{\Phi}^{-1}(u)\,LH\,\Sigma(\bar{\rho}(u))\right)^\top \mathrm{d}u\right]\bar{\Phi}^\top(t), \tag{4}$$

where $\bar{\Phi}(t) := \Phi(t, 0)$ and $\Phi(t, s)$ is the solution to the matrix-valued ordinary differential equation

$$\frac{\partial}{\partial t}\Phi(t,s) = \partial F(\bar{\rho}(t))\Phi(t,s), \quad \Phi(s,s) = I. \tag{5}$$

In addition, $M(t)$ and $V(t)$ solve the linear (matrix) differential equations

$$\frac{\mathrm{d}M(t)}{\mathrm{d}t} = \partial F(\bar{\rho}(t))\,M(t),$$

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t} = \partial F(\bar{\rho}(t))V(t) + V(t)(\partial F(\bar{\rho}(t)))^\top + LH\,\Sigma(\bar{\rho}(t))(LH\,\Sigma(\bar{\rho}(t)))^\top. \tag{6}$$

Solving this type of differential equations is a well-developed branch of research in the field of numerical mathematics, and can be done relying on standard computational software.

2.3. The Gaussian approximation and its advantages

We proceed by pointing out how the scaling limit results from the previous subsection can be used to approximate the joint vehicle-density distribution by a multivariate Gaussian distribution. Then we discuss the advantages of this method over competing approaches.

2.3.1. Obtaining the Gaussian process approximation

The above fluid and diffusion limit results imply that the (non-scaled) vehicle-density process $\rho(\cdot)$ can be approximated by $\bar{\rho}(\cdot)+\hat{\rho}(\cdot)$, cf. Mandjes and Storm (2021, Equation (22)), with $\bar{\rho}(\cdot)$ and $\hat{\rho}(\cdot)$ given by (2) and (3), respectively. As argued in Section 2.2, the first (deterministic) term takes care of working with the correct mean, while the second (zero-mean Gaussian) term takes care of the fluctuations around that mean. In particular, this means that $\rho(\cdot)$, for any sequence of time points $0 \le t_1 < \cdots < t_k < \infty$, with $k \in \mathbb{N}$, is approximately multivariate Gaussian, where

$$\mathbb{E}\left[(\rho(t_1), \ldots, \rho(t_k))\right] \approx (\bar{\rho}(t_1), \ldots, \bar{\rho}(t_k)), \quad \mathrm{cov}\left[(\rho(t_1), \ldots, \rho(t_k))\right] \approx \mathrm{cov}\left[(\hat{\rho}(t_1), \ldots, \hat{\rho}(t_k))\right].$$

It follows that we can approximate the distribution of the kmd-dimensional object $(\rho(t_1), \ldots, \rho(t_k))$ by a multivariate Gaussian distribution of dimension kmd, i.e.,

$$\big(\rho(t_1), \ldots, \rho(t_k)\big) \approx N_{kmd}(\mu, \Sigma), \tag{7}$$

where the associated mean vector μ and covariance matrix Σ are provided by numerically solving the integral equations in (2) and (4) at the time points $t_1, \ldots, t_k$.

In summary, the problem of finding the joint vehicle-density distribution is, in practical terms, reduced to solving the integral equations (2) and (4) that provide us with the μ and Σ in (7). We now present a detailed description of the complete procedure.

We suppose that the network topology is given including the lengths of the road segments in the network (i.e., the lengths of the edges in the underlying graph). Then, a choice must me be made for the number of cells per segment and the cell lengths. Concretely, as before focusing on a segment consisting of a sequence of cells, the number of cells $d \in \mathbb{N}$ needs to be chosen, together with the length $\ell_i > 0$ for each cell i; this determines the matrices L and H appearing in (4). There is an incentive to choose the cell lengths ℓ_i relatively small, so as to provide a description of the vehicle-density distribution that is as detailed as possible. This should, however, be done under the constraint that the cells are sufficiently large to make sure that the use of our Gaussian approximation is justified: there should be a sufficient level of aggregation so that the central limit theorem kicks in. Much is known about the minimum aggregation level required; see e.g. the rules of thumb discussed in Schader and Schmid (1989). In the setting of this paper, one could work with the practical requirement that each cell should be able to accommodate, say, 15–20 vehicles of any type. This reasoning also explains why, in a practical context, the number of types should not become excessively large. In Section 4.3, we demonstrate that given sufficient aggregation, the model's outcomes are robust for different choices of the cell lengths ℓ_i.

- First, a choice for a discrete flux function $\tilde{q}$ (and, hence, fundamental diagram) is made, together with a choice for the involved parameters (e.g., maximum vehicle velocity, backwards wave velocity, and jamming capacity). This choice also involves the parameter m, where $m = 1$ corresponds to single class and $m = 2, 3, \ldots$ to multi-class. The flux function essentially encodes how the vehicle velocity depends on the vehicle densities in the sending and receiving cell. In Mandjes and Storm (2021, Section 3.3), two possible choices for a discrete flux function have been presented for a network without merging and diverging; we will use the merging and diverging rates of Mandjes and Storm (2021, Example 1), which are originally from Daganzo (1995), and which will be further discussed in Section 3.1.1.

- The choice for a specific discrete flux function determines the functions Σ and F (both via the function Q) appearing in the integral equations (2) and (4). Determining ∂F entails (a) computing the Jacobian of F whenever it exists, and (b) choosing the value that ∂F takes whenever the Jacobian of F does not exist. The validity of this reasoning is provided in Mandjes and Storm (2021, Section 4.2) and a detailed example is worked out in Jabari and Liu (2013, Section 4.1) for the discrete flux function that corresponds to the Daganzo fundamental diagram.

- Having determined the functions F, Σ, and ∂F, as well as the matrices H and L, one can numerically solve the integral equations (2) and (4). The solution to the former can be found by feeding its corresponding differential form

$$\dot{\bar{\rho}}_t = F(\bar{\rho}_t), \quad \bar{\rho}_0 = \bar{\rho}$$

to an ODE-solver. Computing $\Gamma(s, t)$ in (4) requires numerical integration of the functions F, Φ, and Φ^{-1}. The functions Φ and Φ^{-1} can be evaluated by solving (5) and its adjoint equation (cf. Hale (1969, Lemma III.1.4)),

$$\frac{\partial}{\partial t} \Phi^{-1}(t, s) = -\Phi^{-1}(t, s) \partial F(\bar{\rho}(t)), \quad \Phi^{-1}(s, s) = I$$

respectively, at a sequence of time points $0 = t_1 < \cdots < t_k = t$ for some $m \in \mathbb{N}$ using an ODE-solver. Then, $\Gamma(s, t)$ is computed for every pair (t_i, t_j), for $i, j \in \{1, \ldots, k\}$. Note that the discretization $t_1, \ldots, t_k$ of $[0, t]$ determines the time points at which the multivariate distribution of $\rho(\cdot)$ is computed. Moreover, the mesh of the discretization prescribes the accuracy at which the integrals are to be evaluated. While, by and large, one can rely on off-the-shelf numerical software, in specific cases one has to slightly adapt the input to obtain meaningful results. Most notably, as extensively discussed in Mandjes and Storm (2021), in case the discrete flux function does not obey a certain regularity condition (i.e., being 'Lipschitz'), numerical issues can arise, which can be resolved by applying a standard mollifier.

Note that if one is only interested in *spatial* covariances (not the temporal ones), then solving (6) rather than (4) is sufficient. This is even faster since numerical integration is not required.

The resulting Gaussian approximation can be used to approximate the probability of any event that can be expressed in terms of the random vector (7). The evaluation of these probabilities will be efficient, as it require basic numerical operations; specifically, we are to solve the integral equations (2) and (4) for the mean vector and covariance matrix, respectively, which have a standard functional form. In the remainder of this section we compare our approximation, in terms of computational efficiency, with two evident competing approaches.

2.3.2. Advantages over alternative methods

A first alternative for evaluation of the finite-dimensional distributions concerns the use of computational tools for continuous-time Markov chains. The objective is to numerically evaluate the matrix $P(t)$ with entries $\mathbb{P}(\rho(t) = k \mid \rho(0) = \ell)$, for k, ℓ in the state space S of $\rho(\cdot)$. Denoting by Q the transition rate matrix associated to the Markov chain $\rho(\cdot)$, then, by the Kolmogorov forward equation, we need to evaluate the following matrix exponential: for $t \geqslant 0$,

$$P(t) := \exp(Qt) = \sum_{k=0}^{\infty} \frac{(Qt)^k}{k!}.$$

Both $P(t)$ and Q are matrices of the dimension $|S| \times |S|$, where $|S|$ is the dimension of the state space S of $\rho(\cdot)$. Note that $|S|$ depends on, e.g., the cell lengths, vehicle types and fundamental diagram. For the case of a single vehicle type and homogeneous cell-lengths, we have $|S| = |(X^{\mathrm{jam}} + 1)|^d$, with X^{jam} the maximum number of vehicles that can be simultaneously present in a cell. It is directly seen that already for a model of modest size (e.g., $d = 10$ and $X^{\mathrm{jam}} = 20$), this number is too large to make this approach computationally viable as a method for computing $P(t)$. There are other ways to numerically evaluate $P(t)$, for instance by solving a system of linear differential equations, but these approaches will also break down as the dimension of the state space grows.

An other alternative is to rely on simulation. This approach, however, has intrinsic conceptual complications as well. The most prominent issue is that one has to deal with estimation errors. For instance, suppose we wish to *reliably* estimate (i.e., aiming at estimates with errors that are smaller than a given, small, fraction of the estimates themselves) the distribution of $\rho(\cdot)$ at times $t_1, \ldots, t_k$. Then, many runs are needed, as a consequence of the fact that all states (in the large state space) have to be visited sufficiently often. Evidently, if one is interested in specific events that effectively do not involve the full state space, or just one point in time, the computational load will be more manageable. It is noted, however, that there are various relevant performance metrics whose reliable estimation is prohibitively time-consuming; for instance, as was pointed out in the introduction of Jabari and Liu (2013), estimating the time-dependent covariance matrices is particularly challenging.

3. Vehicle-density distribution evaluation in networks

With a precise understanding of our model and Gaussian process approximation, we now turn to the main contribution of this paper: we demonstrate that the model can efficiently evaluate the joint (both in the spatial and temporal sense) vehicle-density distribution for a road-traffic network. The resulting numerics can be used to assess the probability of 'risky events', such as excessive congestion. As pointed out earlier, the distributions found are consistent with Godunov solutions for conservation laws, i.e., our setup generalizes kinematic wave models.

The main objective of this section is to demonstrate that the model is capable of evaluating the joint vehicle-density distribution in a wide range of real-life traffic scenarios: when congestion arises due to on-ramp traffic, jam formation and dissipation due to accidents, and traffic propagation when the road capacity drops due to a lane closure. Moreover, we illustrate how relevant probabilities can be computed, e.g., the probability of exceeding a critical density threshold (i.e., corresponding to heavy congestion). An attractive feature of our approach is that it allows the evaluation of the *combined* impact of various traffic phenomena and control measures, as we demonstrate in one of our experiments. In addition, every cell can have its own specific fundamental diagram to model effects specific to local traffic flows. We illustrate this by taking different parameters for the flux functions in various cells, but in principle the flux functions themselves could also be different. We in addition compare the computation times with those corresponding to estimating the relevant means and variances by simulation. In Section 4, we present an in-depth study of the computation times of our approach.

3.1. Experimental setup

We first provide a detailed setup of all real-life scenarios that we evaluate using our approach; in the sequel we refer to these by *experiments*. We throughout use the same topology, as well as the same initial condition, so as to facilitate comparing the outcomes of the experiments. We first describe these commonalities, after which we in-depth discuss the experiments. As our objective is to show that our approach is capable of quickly evaluating the joint vehicle-density distribution in a network, we decided to keep the setup of our experiments as focused as possible. We work with single-class traffic-density models with all 'topological features' (i.e., on-ramps, merges, etc.) that are contained in networks. For several multi-class examples on a road segment, we refer to Mandjes and Storm (2021, Section 6).

3.1.1. Commonalities

The topology of our network is graphically represented in Fig. 1. It consists of 4 road segments with both an on-ramp and off-ramp between every consecutive pair of road segments. The four road segments, denoted r_i in the figure, consist of $d_i \in \mathbb{N}$ cells, $i = 1, 2, 3, 4$. In front of the first cell of r_1, a cell is placed at which vehicles seek to arrive with rate λ_0. Similarly, at the end of r_4, a cell is placed from which vehicles depart with a rate that is bounded by v_0.

The topology of the part of the network that contains an on-ramp and off-ramp (called a 'block') is also schematically represented in Fig. 1. Each of the three blocks consists of 5 cells: a diverge cell (D), a merge cell (M), an on-ramp cell (on) to which vehicles arrive, an off-ramp cell (off) from which vehicles depart, and an intermediate cell (I). We note that the latter is necessary to comply with the framework of merge and diverge models as proposed in Daganzo (1995) and reviewed in Tampère et al. (2011). From the diverge cell in block 1, vehicles prepare to leave the network with probability p_1^{div}, or they stay on the network with probability $1 - p_1^{\mathrm{div}}$.
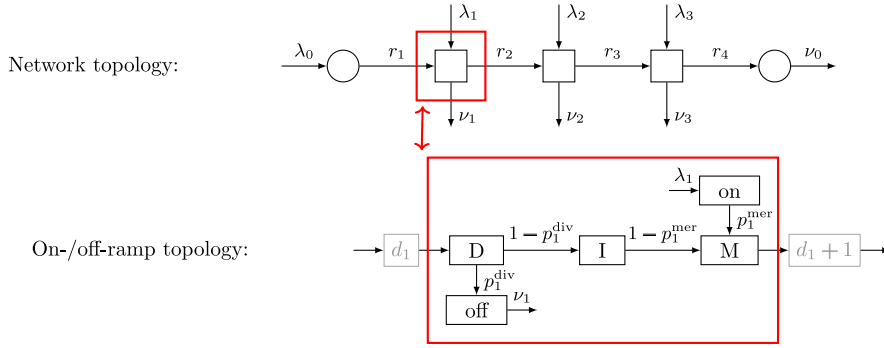
Fig. 1. Schematic representation of the network, consisting of 4 roads r_i between circular nodes, with $d_i \in \mathbb{N}$ cells per road, $i \in \{1, \ldots, 4\}$. Between consecutive pairs of roads there is both an on-ramp and off-ramp for traffic to arrive and depart, respectively, as represented by the part in the red box. Arrows with label λ_0 or ν_0 correspond to, respectively, cells with arriving and departing vehicles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Similarly, vehicles in the intermediate cell and at the on-ramp, merge into the merge cell, where priority is given to the vehicles from the on-ramp with probability p_1^{mer}. Similar parameters are defined for merging and diverging behavior in blocks 2 and 3. The arrival and departure rate of on-ramp j and off-ramp j are positive numbers and are denoted, respectively, by λ_j and ν_j, for $j = 1, 2, 3$.

From Fig. 1 it can be seen that in this network, diverge cells are in one direction followed by an 'intermediate' cell (I) which continues the road segment. In the other direction after the diverge cell, we have placed an off-ramp (off) where vehicles depart the network. Instead, we could have continued the network in this direction, similar to how we continued the network after each intermediate cell, resulting in a network of highway segments that vehicles can switch between. Hence, conceptually there is no difference between a single stretch of highway with one or more merge and diverge cells, and a concatenation of stretches of highways connected by merge and diverge cells. This concretely means that a network consisting of the elements that are covered in Fig. 1 contains all features necessary to build large-scale traffic networks.

To be able to perform any numerical calculations, we need to specify the model's discrete flux function, i.e., the entries of the vector $Q(\rho(t))$, being a function of the current state of the joint vehicle-density process $\rho(t)$. For this, we rely on the expressions derived in Daganzo (1994, 1995) corresponding to Daganzo's fundamental diagram. As pointed out in Chanut and Buisson (2003), it follows from Lebacque (1996) that for transitions between two cells, the discrete flux-function can be written as the minimum of a sending and receiving function, which are then given in Daganzo (1994). These functions were generalized to include merging and diverging in Daganzo (1995). We summarize the rate functions $q_i(\rho(t))$ below.

- The rate at which a vehicles transitions between two cells i and $i + 1$ (unless they are merge or diverge cells) is given by

$$q_i(\rho(t)) = \min\{S_i(\rho(t)), R_i(\rho(t))\},$$

where

$$S_i(\rho(t)) := \min\{v_i^f \rho_i(t), q_i^{\max}\} \quad R_i(\rho(t)) := \min\{-w_i(\rho_i(t) - \rho_i^{\max}), q_i^{\max}\};$$

the minimum in the definition of $q_i(\rho(t))$ reflects the fact that this rate is determined by the most binding of the 'sending rate' $S_i(\rho(t))$ and the 'receiving rate' $R_i(\rho(t))$. The arrival rate to cell 1 and departure rate from cell d are given by, respectively,

$$q_0(\rho(t)) = \min\{\lambda_0, R_1(\rho(t))\}, \quad q_d(\rho(t)) = \min\{S_d(\rho(t)), \nu_0\}.$$

- When two cells i and j merge into a cell k, with priority probability p for cell i, then the rates $q_i(\rho(t))$ and $q_j(\rho(t))$ at which vehicles depart cells i and j, respectively, are given by

$$\begin{aligned} q_i(\rho(t)) &= S_i(\rho(t)), \\ q_j(\rho(t)) &= S_j(\rho(t)), \end{aligned} \quad \text{if } S_i(\rho(t)) + S_j(\rho(t)) \le R_k(\rho(t)),$$

in case $S_i(\rho(t)) + S_j(\rho(t)) \le R_k(\rho(t))$, and

$$\begin{aligned} q_i(\rho(t)) &= \mathrm{mid}\{S_i(\rho(t)), R_k(\rho(t)) - S_j(\rho(t)), pR_k(\rho(t))\}, \\ q_j(\rho(t)) &= \mathrm{mid}\{S_j(\rho(t)), R_k(\rho(t)) - S_i(\rho(t)), (1-p)R_k(\rho(t))\}, \end{aligned}$$

in case $S_i(\rho(t)) + S_j(\rho(t)) > R_k(\rho(t))$, where $\mathrm{mid}\{a, b, c\}$ is the median of the three numbers a, b, and c.

- Similarly, when a cell i diverges into cells j and k, with a fraction p going into cell j, then the departure rate from cell i is given by

$$q_i(\rho(t)) = \min\{S_i(\rho(t)), R_j(\rho(t))/p, R_k(\rho(t))/(1-p)\},$$

and the rate into cells j and k is given, respectively, by $p\,q_i(\rho(t))$ and $(1-p)\,q_i(\rho(t))$ (where it is noted that $q_i(\rho(t))$ depends on j and k).

We refer to Jabari and Liu (2013) for an account of how to properly take derivatives of the functions S_i and R_i. Our choice regarding the weak derivatives matches the one applied in Jabari and Liu (2013).

Each experiment is initialized from a state complying with the stationary distribution of the model with a set of baseline parameters. The underlying idea is that we want to make sure that the effect of the real-life scenarios that we modeled, can all be compared to the same reference scenario. The set of baseline parameters corresponds to a relatively lightly-loaded network configuration, so that including an additional on-ramp or an accident has a clear effect. In addition, in the instance we consider there is no traffic arriving or departing at any of the intermediate on- or off-ramps (but, evidently, we could have chosen different scenarios as well). Specifically, we let $d_i = 5$ for each road i, with each cell in the network having length 500 m. We choose the following parameters of the fundamental diagram. In each cell, the maximum velocity is $v^f = 80$ km/h, the backward wave speed is $w = 20$ km/h, and the maximum flow is $q^{\max} = 1800$ veh/h. We let the critical density equal $\rho^{\max} = 108$ veh/km for each cell. Finally, the arrival rate at the first cell is $\lambda_0 = 1200$ veh/h, $\nu_0 = 1800$ veh/h, and $p_j^{\text{div}} = \lambda_j = 0$ (so that we do not need to specify p_j^{mer} and ν_j) for $j = 1, 2, 3$.

3.1.2. Experiments

We now provide a detailed description of the three experiments.

On-ramp influence In this experiment, we add an extra stream of arriving vehicles at the second on-ramp. More specifically, we set $\lambda_2 = 1200$ veh/h and $p_2^{\text{mer}} = 0.5$, on top of the baseline setting. This means that when the second merging cell is congested, the receiving flow for the merging cell is divided over the second on-ramp cell and second intermediate cell. We evaluate the flow of traffic for half an hour.

By imposing a higher load on the second merging cell ($\lambda_0 + \lambda_2 = 2 \times 1200 = 2400$) than it can handle ($q^{\max} = 1800$), we anticipate that congestion should arise. Specifically, congestion arises both on the on-ramp and in the intermediate cell, due to $p_2^{\text{mer}} = 0.5$, which divides the load evenly. Consequently, congestion should propagate upstream from the intermediate cell.

Accident modeling For this second experiment, we start with the baseline setting, and divide a window of half an hour into three intervals $[0, \tau_1], [\tau_1, \tau_2], [\tau_2, \tau_3]$, with $\tau_1 = 300$ s, $\tau_2 - \tau_1 = 600$ s, and $\tau_3 - \tau_2 = 900$ s. At τ_1, an accident occurs at cell $d_1 + 3$, that is, at the third cell of the second road. We model this by restricting the flow into cell $d_1 + 3$ by a factor 0.75, i.e., we scale the rate $q(\rho_{d_1+2}, \rho_{d_1+3})$ by a factor 0.25. At $\tau_2 = 900$ s, we let the accident be resolved, and traffic can again flow as before. We then let the network evolve until $\tau_3 = 1800$ s.

Combined phenomena An appealing feature of our approach is that it allows easy evaluation of the *combined* effect of multiple different phenomena and control measures, in contrast to many existing methodologies that primarily focus on quantifying the impact of a single measure phenomenon or measure, cf. Papageorgiou et al. (2003). It means that we can analyze, for example, the combination of the impact of on-ramps and accidents, but also scenarios in which in addition specific control measures (e.g., ramp metering and maximum-speed measures) are incorporated, or design choices (e.g., extra lanes and new roads), or natural phenomena (e.g. capacity drops and weather changes). By our final experiment we demonstrate this, by considering a scenario that combines various features in a single experiment. More specifically, we 'activate' our on-ramps and off-ramps, we change the maximum velocity on r_1 to 100 km/h instead of 80 km/h, and we impose time-varying arrival rates at the arrival cell and on-ramp 3. The ODE-solver evaluates the joint effect of these changes (relative to the baseline model).

The effects are evaluated over the time span of one hour. We set $\lambda_1 = \lambda_2 = 600$ veh/h and $\lambda_3 = \lambda_0(t)$ (explained below), with $p_1^{\text{mer}} = p_2^{\text{mer}} = p_3^{\text{mer}} = 0.5$. All ν are set to 1800 veh/h, with $p_1^{\text{div}} = 0.7$, $p_2^{\text{div}} = 0.5$, and $p_3^{\text{div}} = 0$. Finally, for the time-varying arrivals at the arrival cell and on-ramp 3, we take a piece-wise constant arrival rate function $\lambda_0(t)$, defined as

$$\lambda_0(t) = \begin{cases} 400, & t \in [0, 600), \\ 1100, & t \in [600, 1200) \cup [3000, 3600), \\ 1800, & t \in [1200, 1800) \cup [2400, 3000), \\ 2400, & t \in [1800, 2400), \end{cases}$$

that is, we model an arrival pattern that step-by-step increases from 400 to 2400, and then decreases down to 1100 again. More complicated arrival patterns can be evaluated, such as daily patterns of arrivals. To keep our setup relatively simple, we decided to work with a piece-wise constant arrival rate function, but it is stressed that one could work with considerably more general non-homogeneous Poisson processes (notably, with arrival rates that change continuously with time). In a similar way, the diverge and merge probabilities could be made time-varying as well. However, to keep the setup as transparent as possible, we decided to refrain from this in this experiment, and work with ramp-specific diverge probabilities instead.

3.2. Experiment outcomes

We now present the outcomes of each of our experiments. For every experiment, we graphically illustrate how the model has evaluated the vehicle-density distributions. In addition, we demonstrate how probabilities of certain relevant events can be computed in a straightforward manner. In addition, we provide the corresponding computation times, both using our approach and using simulation.
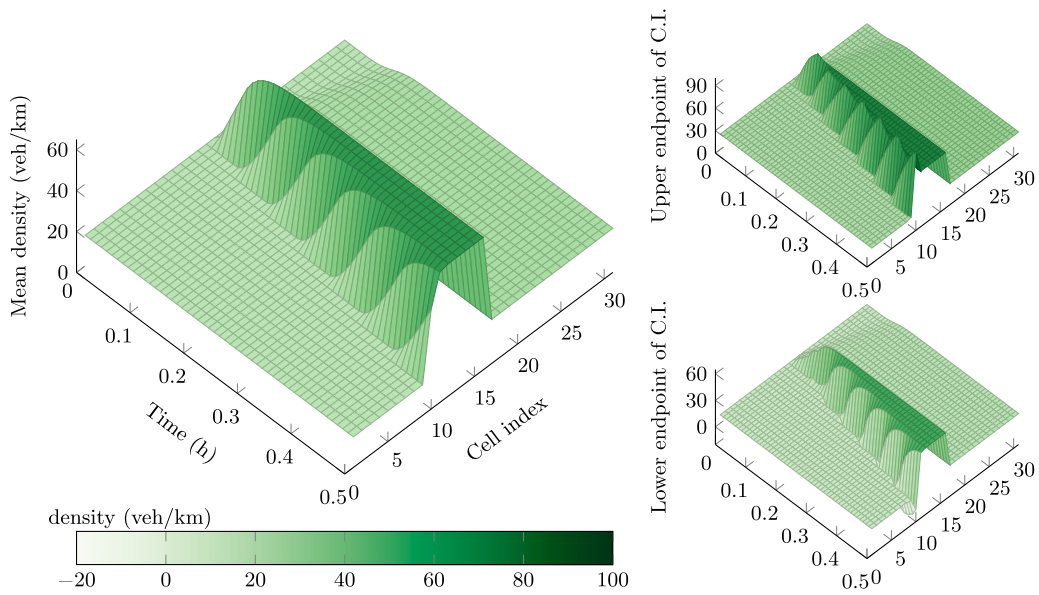
Fig. 2. Vehicle-density propagation under congestion arising from an on-ramp. Both the mean (left) and the upper and lower bound for a confidence interval (right) are plotted on top of the (time, cell index)-plane.

3.2.1. On-ramp influence

To present how the vehicle-density distribution for all cells jointly evolves over time, we plot the mean vehicle density $\bar{\rho}_i(t)$, as well as the upper and lower endpoints of the associated confidence interval $[\bar{\rho}_i(t) - \sigma_i(t), \bar{\rho}_i(t) + \sigma_i(t)]$ in various different ways, where $\sigma_i(t) = \sqrt{V_{ii}(t)}$ with $V(\cdot)$ as in (6). In the first place, we have plotted these quantities in Fig. 2, for $t \in [0, 0.5]$ and i corresponding to the cells in the main segment (excluding on- and off-ramps) in the order they are visited by vehicles. The figure illustrates how the vehicle density builds up in the cells upstream of the on-ramp in a backpressure-like manner as a consequence of the congestion caused by the on-ramp at cell 17.

The evolution of traffic densities and the corresponding confidence intervals is more clearly represented in Fig. 3 for a selection of representative cells. Specifically, the mean vehicle density and its associated confidence intervals are shown for the on-ramp, an upstream cell, the merge cell, and a downstream cell. From this figure, it becomes apparent that cells that are upstream from the on-ramp become more congested. In contrast, cells downstream from the on-ramp fill up to the point that the vehicle velocity would start to suffer from congestion.

Both Figs. 2 and 3 illustrate that the variance obtained from the Gaussian approximation reflects the backwards propagation of congestion. The spike in the variance of the upstream cell between $t = 0.1$ and $t = 0.2$ can be explained from the traffic density passing through the point where the derivative of the fundamental diagram has a discontinuity, cf. Mandjes and Storm (2021, Remark 2). A possible remedy to obtain a better variance approximation is to increase the cell lengths, thereby obtaining better approximations cf. Mandjes and Storm (2021, Section 6.1.2), and rescaling the variance afterwards. We discuss this scaling property in greater detail in Section 4.2.

As mentioned before, the Gaussian approximation can be used to assess the probability of a wide range of relevant events. As an example, we consider here the probability $\mathbb{P}_i(t)$ that the density in cell i exceeds the critical threshold of 30 veh/km at time $t \geq 0$, which is the value at which vehicles drive slower than 60 km/h. Using the notation of (7), write $\Sigma(t) = (\sigma_{ij}(t))$ for the covariance matrix at time t, with i, j ranging over the indices of the cells. For every cell i, the marginal distribution is then $N(\bar{\rho}_i(t), \sigma_{ii}^2(t))$. This makes computation of the probability of the above event straightforward once the matrix $\Sigma(t)$ has been determined by the ODE-solver.

The Gaussian approximation in (7) also provides us with covariances, which play a role in computing the probability of events that involve the state of *multiple* cells (possibly at different points in time). To illustrate this, consider also the probability $\mathbb{P}_i^{(3)}(t)$ that cells i, $i + 1$, and $i + 2$ *all* have a vehicle density larger than 30 veh/km at time t. To compute the probability of such events, we require spatial covariances (at a single point in time); temporal covariances are not required here.

In Fig. 4 we have plotted both $\mathbb{P}_i(t)$ and $\mathbb{P}_i^{(3)}(t)$ for all cells i in the main segment (i.e., omitting on- and off-ramps) and $t \in [0, 0.5]$ as heatmaps. The computed probabilities indicate that the risk of congestion is very high (almost probability one) in front of the on-ramp. We also observe how the risk of congestion propagates backwards over time. The left heatmap shows that the presence of the on-ramp arrivals also increases the risk of congestion downstream from the on-ramp. Finally, we present the times it takes to establish the Gaussian process approximation, i.e., the time required to solve (2) and (6), with the time one would need when relying on simulation. To evaluate the differential equations we used an ODE-solver from the SciPy package in Python. Both the code used by the simulation and the ODE-solver have not been (fully) optimized; the presented numbers serve as an indication of the
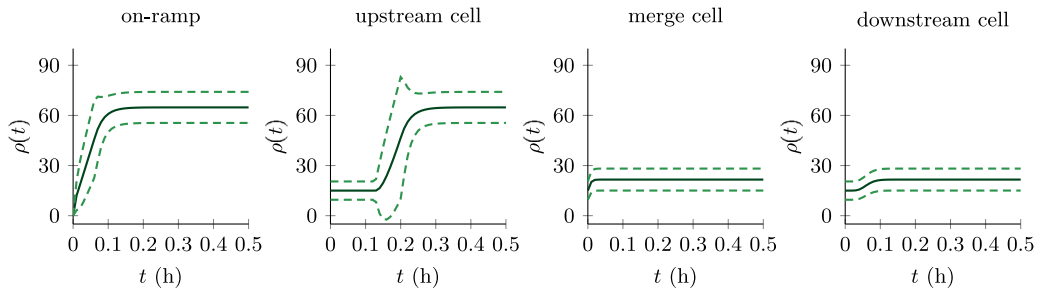
Fig. 3. Detailed plots of the qualitative change in cell density due to congestion arising from an on-ramp. Starting from the left, both the mean vehicle density (solid line) and a confidence interval (dashed lines) of width two standard deviations, are plotted for: the on-ramp, a upstream, the merge, and a downstream cell, as seen from the second on-/off-ramp block.

difference in computation times. All numerical experiments were performed on a high-performance laptop, where it is anticipated that computation times can be improved by using specialized computation clusters. The computation time for the ODE-solver was about 12 s. The simulation required about 20 min to return $N = 100$ sample paths of $\rho(\cdot)$, and over 3 h to return $N = 1000$ sample paths.

Note that simulated results provide an estimate of the 'true' distributions of $\rho(\cdot)$ (on a grid of time points), which evidently comes with an estimation error. This error can be made smaller by increasing the sample size N. As an example, we present α-confidence intervals for the estimated mean and standard deviation of the vehicle density in cell 10 at $t = 0.5$ h, the latter obtained using a bootstrap of size 10^4. For $\alpha = 0.95$, we obtain $(67.41, 70.15)$ and $(70.02, 70.88)$ as α-confidence intervals for the mean, using $N = 100$ and $N = 1000$ respectively. The α-confidence intervals for the standard deviation are $(7.97, 8.73)$ and $(7.30, 9.42)$. Recall the rule that in order to reduce the standard deviation of an estimator (and hence the width of the corresponding confidence interval) by a factor 10, the number of samples should be increased by a factor 100. Therefore, even for this relatively small size of the network and with an optimized simulation implementation, simulation times are anticipated to be of the order a few days at the minimum. Parallel simulation may substantially reduce simulation time, but an excessively large number of cores would be needed to be on par with our Gaussian approach.

The ODE solver provides an approximation to the desired vehicle-density distributions, which comes with an approximation error. This error will not fully vanish, since the vehicle densities $\rho(\cdot)$ tend to the Gaussian process only in the limit. By making the cells larger, though, the approximation becomes better over aggregated time intervals. Moreover, the ODE solver also introduces an error, but this error is negligible compared to the other ones discussed; the ODE-solver is typically set to accept absolute errors of size 10^{-6} at any time t in our experiments.

Finally, we note that in the present case, we focused on spatial covariances only, and were therefore not required to solve (4). The evaluation of spatial covariances requires longer computation times, but this increase is not significant. In particular, our experiments that we performed indicate that the computation time is still orders of magnitude smaller than the simulation time, so that the numbers presented above can be considered representative. In this context we refer to the experiments performed in Mandjes and Storm (2021, Example 6.2), in which the travel-time distribution is evaluated in a multi-class setting, requiring the evaluation of temporal covariances.

3.2.2. Accident modeling

For the second experiment, we again present how the vehicle-density distribution evolves over time by plotting the mean vehicle density $\bar{\rho}_i(t)$ and the corresponding confidence interval with width $2\sigma_i(t)$. The space–time behavior of the distribution is depicted in Fig. 5. The figure quantifies how, due to reduced flow in the third cell of r_2 during $t \in [1/12, 1/4]$, congestion builds up in cells upstream from that cell. In contrast, cells downstream start to empty, as one would expect. After the accident has been resolved and the flow is no longer reduced, the formed traffic jam starts to dissipate during $t \in [1/4, 1/2]$. However, the jam does propagate backwards, i.e., congestion temporarily increases in upstream cells.

The behavior of vehicle densities in cells of r_2 is plotted in detail in Fig. 6. Here, we show the mean vehicle density and its confidence interval as a function of time for the first cell of r_2, the cell where the accident happens, and the last cell of r_2. The different stages of time are colored differently to emphasize what happens before, during, and after the accident. Again, it is apparent that the path of the variance found by the Gaussian process approximation roughly follows the same shape as the path of the mean, which illustrates that the same traffic flow phenomena are captured by the mean and the variance. As an aside, note that, due to the specific we in which we in this experiment modeled the accident's impact, *in the accident cell itself* the density is relatively low; to make sure that the accident cell itself fills, one could model the accident's impact by reducing the flow into the cell *beyond* the accident cell.

To quantify the computational advantage of our approach, we ran a simulation with $N = 100$ samples paths, and another one with $N = 1000$ sample paths, which took 13 min and over 2 h, respectively. The computation time required by the ODE solver, on the other hand, was as low as 25 s.
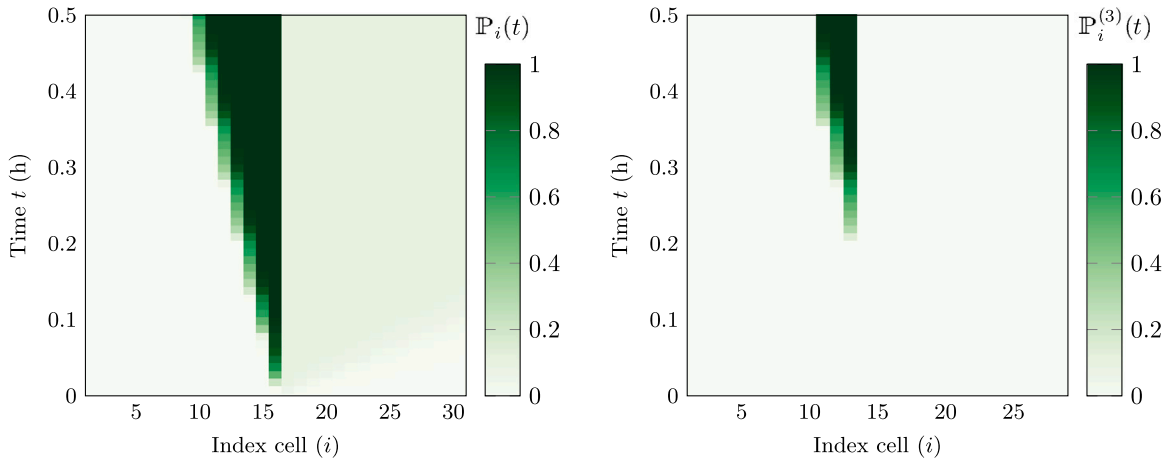
Fig. 4. Heatmaps, depicting the (Gaussian) probability $\mathbb{P}_i(t)$ of cell i exceeding a density of 30 veh/km at time t (left) and the probability $\mathbb{P}_i^{(3)}(t)$ of cells i, $i+1$, and $i+2$ all exceeding a density of 30 veh/km at time t (right).
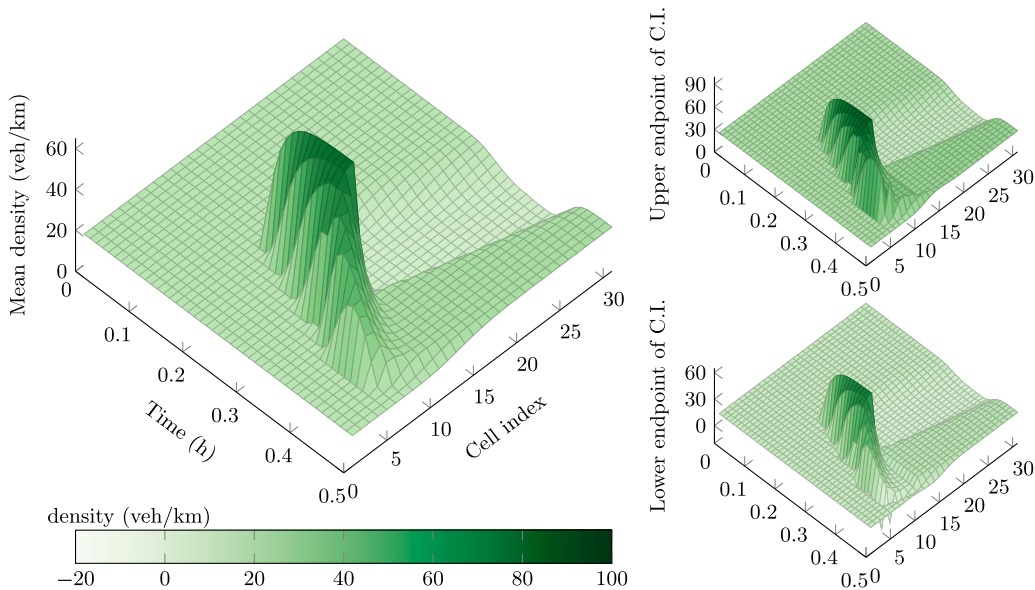

Fig. 5. Jam formation and dissipation due to an accident. Both the mean (left) and the upper and lower bound for a confidence interval (right) are plotted on top of the (time, cell index)-plane.

3.2.3. Combined phenomena

In our final experiment, we again show the evolution of vehicle-density distributions by plotting the mean vehicle density $\bar{\rho}_i(t)$ and associated confidence interval with width $2\sigma_i(t)$; see Fig. 7. It shows that the time-dependent arrivals at the arrival cell influence vehicle densities at r_1 and r_2 (the latter mildly), but hardly at r_3. This effect is partly due to most vehicles departing from r_1 at the first off-ramp. In addition, it shows that on-ramp 3 has enough arrivals to create congestion at r_3. This even happens at the low arrival rates, due to the arrivals at on-ramp 1 and on-ramp 2. Importantly, from the upper and lower endpoints of the confidence interval, it can be seen that the variance obtained is consistent with the fluid limit, also in this case of combined phenomena.

In Fig. 8, we present how the vehicle densities behave in r_1 and the first on-/off-ramp block. The figure demonstrates that the influence of time-varying arrival rate and diverging traffic is handled appropriately, also in terms of the (co-)variance of the vehicle-density process. The effect of the time-varying arrival rate on the vehicle densities in these cells is clearly visible. Also, we see how the off-ramp has a higher vehicle density than the intermediate cell, which was to be expected due to the large diverge probability of 0.7 in the first diverge cell.

To assess the efficiency our approach, we again compare with simulation. Generating $N = 100$ sample paths took as long as 40 min, while $N = 1000$ took even almost 7 h. The ODE-solver returned output in just 16 s.
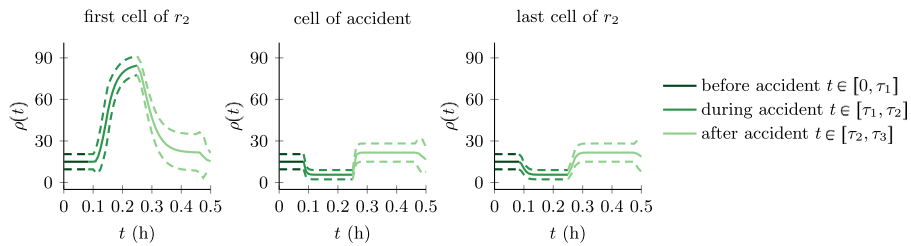
Fig. 6. Detailed plots of the qualitative behavior of vehicle densities in cells in and around the accident cell the third cell of r_2. The vehicle density (solid line) and associated confidence interval (dashed lines) have been plotted, with the different stages in time around the accident colored differently. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
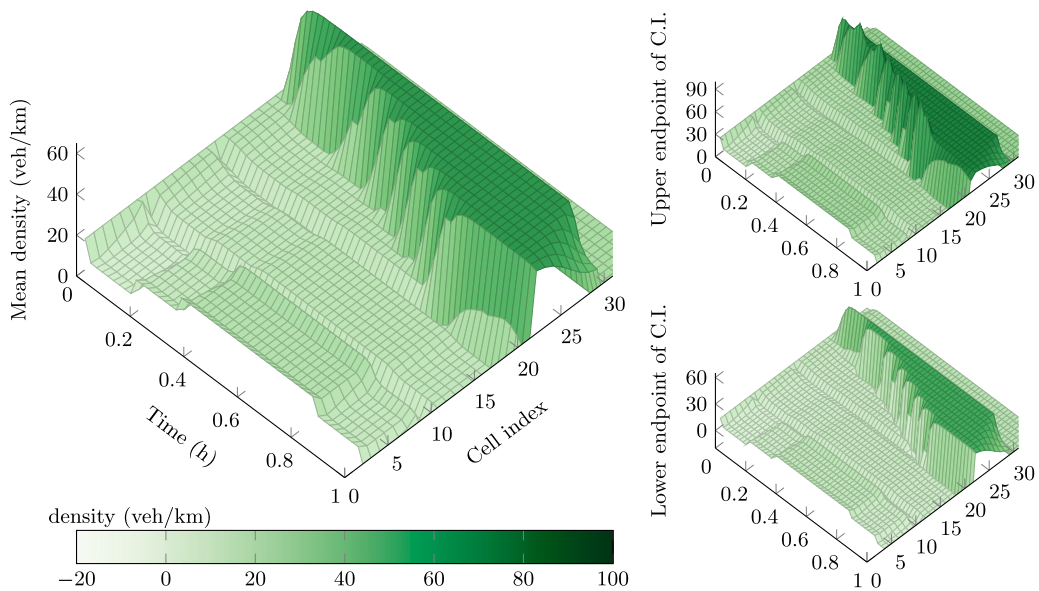

Fig. 7. Vehicle-density propagation under the combined influence of different phenomena (on-ramps, off-ramps, and time-varying arrivals). Both the mean (left) and the upper and lower bound for a confidence interval (right) are plotted on top of the (time, cell index)-plane.
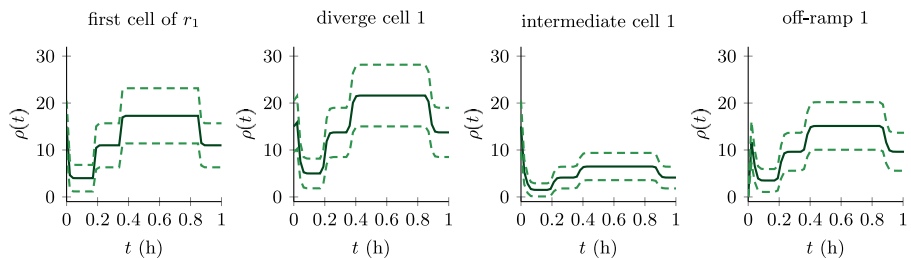

Fig. 8. Detailed plots of the qualitative change in cells of r_1 and the first on-/off-ramp block under the combined influence of different phenomena (on-ramps, off-ramps, and time-varying arrivals). Starting from the left, both the mean vehicle density (solid line) and a confidence interval (dashed lines) of width two standard deviations, are plotted for: the first cell of r_1 and the first diverge, intermediate and off-ramp cells in the segment.

4. Complexity and scaling properties

This section provides an in-depth analysis of the computation time of our approach. Specifically, we quantify the time it takes to evaluate the differential equations for the mean and (co-)variance of the Gaussian process. We assess the dependence of the computation times on the number of cells d (i.e., the computational complexity), establish that computation times are invariant
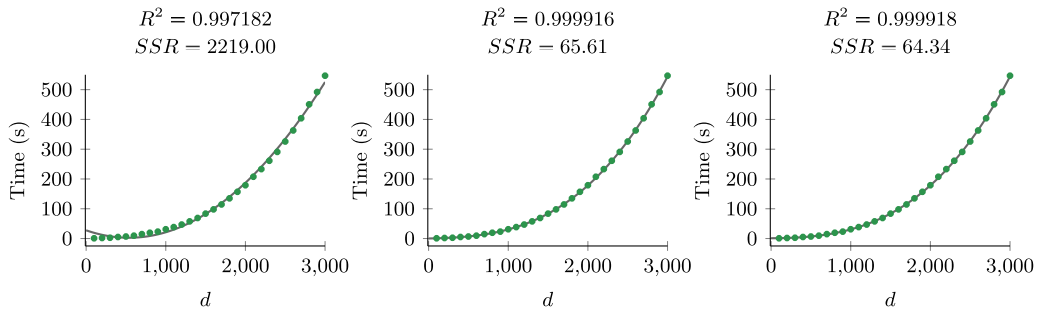
Fig. 9. Three plots of the best-fit polynomial (gray line) for the computation times (green dots), with the associated R^2 and sum of squared residuals (SSR). The maximal degrees of the polynomials are: two (left), three (middle), and four (right).

under scaling of time and space, and compare computation times for different discrete flux functions (i.e., different fundamental diagrams). In addition, we focus on robustness with respect to the precise choice of the cell lengths: we demonstrate that, under the constraints for choosing the cell lengths that were discussed in Section 2.3.1, the way in which the road segment is divided into cells does not significantly affect our methodology's output, as desired.

4.1. Number of cells

We start by studying the computation time as a function of the number of cells d in the network. Recall that the computation time of the model is essentially the time that it takes an ODE solver to compute the solution to the differential equation for (2) and (6) jointly. Since d determines the size of the matrices and vectors occurring in these differential equations, and since solving ODEs is effectively dominated by matrix multiplications, we would expect that computation times are roughly of the order $C \times d^\alpha$, for some $C > 0$ and $\alpha \approx 3$. Strictly speaking, the value of α depends on the efficiency of the implementation of matrix multiplication: standard software typically achieves $\alpha \approx 2.8$, but for large matrices more efficient implementations have been developed. In general, a network can be composed of merge/diverge cells, arrival/departure cells, and cells that are none of the former. The number of each of these types typically influences just the constant C involved in the order of complexity, i.e., not the order α. Therefore, for simplicity, we focus on a network that is a single road segment, without any on-/off-ramps, i.e., without merge and diverge cells and with one arrival and one departure cell.

The number of cells d is taken between 100 and 3000, with increments of 100. Each cell has unit length. We use the fundamental diagram of Daganzo, with $v^f = 100$ km/h, $w = 20$ km/h, $\rho^{\max} = 108$ veh/km, $q^{\max} = 1800$ veh/h, and an arrival and departure rate equal to, respectively, $\lambda = 1800$ veh/h and $\nu = 1200$ veh/h. The network is initialized as empty, and is evaluated on the time interval $[0, 1/3]$, with time expressed in hours.

To confirm our idea of the computation times being roughly d^3, we apply functional regression to the computation times. To present the outcomes, we have plotted in Fig. 9 the computation times along with the best-fit order-k polynomial, for $k = 2, 3, 4$. At the top of each of the plots we give the condition number R^2 and the sum-of-squared residuals SSR. The graphs and numbers demonstrate that the fit improves significantly from $k = 2$ to $k = 3$, but hardly from $k = 3$ to $k = 4$. We conclude that the computation time of the model has a complexity of roughly order d^3, as anticipated.

Some improvement can be achieved by exploiting the special (sparse) structure of the matrices involved, the complexity can be reduced. Exploiting the sparsity, specializing to the case $m = 1$, the computation fluid limit theoretically takes $O(d)$, while computing all objects featuring in the diffusion approximation of Eq. (7) is $O(d^2)$. In our numerical experiments we have relied on Python, in particular NumPy in conjunction with ODE-solvers from SciPy. We verified that switching to sparse matrices in SciPy did not significantly improve computation times. To achieve the theoretical complexity of $O(d^2)$, a tailor-made low-level implementation of the numerical procedures may be required. It is noted that, given the fact that the Σ matrix, featuring in the Gaussian approximation (7), contains $\frac{1}{2}(kmd)^2 + \frac{1}{2}kmd$ distinct entries, the theoretical complexity cannot be below $O(d^2)$.

A similar analysis applies when studying the impact of the number of vehicle types m, since the dimension of F (and thus of the matrices in (6)) is also determined by m. However, since m is typically relatively small, we omit this analysis.

4.2. Scaling properties

An interesting property of the model is that if the cell lengths and time are scaled by the same constant $c > 0$, then the means stays invariant while the (co-)variances scale by $1/c$ (i.e., standard errors by $1/\sqrt{c}$). This property shows that when cell lengths become larger, the approximation becomes more accurate. However, when c becomes too large, the process dynamics degenerate on small time intervals, in that the process becomes constant on these intervals. Additionally, increasing c removes the randomness from the system, which is shown by vanishing (co-)variances. We briefly elaborate on these properties.

To formally establish the above scaling property, first consider (2). Now, for some $c > 0$, we scale time (i.e., $t \mapsto ct$) and space (i.e., $\ell_i \mapsto c\ell_i$ for all cells i). Then,

$$\bar{\rho}(ct) = \bar{\rho}(0) + \int_0^{ct} \frac{1}{c} LH\, Q(\bar{\rho}(s))\mathrm{d}s = \bar{\rho}(0) + \int_0^t LH\, Q(\bar{\rho}(cs'))\mathrm{d}s',$$

where we recall that $F(\rho) = LH\, Q(\rho)$ for the matrices L and H, and function $Q(\rho)$, and where we apply the transformation $s = cs'$. This shows that $\bar{\rho}(ct)$, with cell lengths scaled by c, is equal to $\bar{\rho}(t)$ with the unscaled cell lengths.

Performing similar manipulations on $\Phi(\cdot)$ and $\Gamma(\cdot, \cdot)$, and using (5) (as well as its adjoint equation) and (4), one can show

$$\Gamma(cs, ct) = \bar{\Phi}(cs)\left[V(0) + \int_0^{c(t\wedge s)} \bar{\Phi}^{-1}(u) LH\, \Sigma(\bar{\rho}(u))\left(\bar{\Phi}^{-1}(u) LH\, \Sigma(\bar{\rho}(u))\right)^\top \mathrm{d}u\right]\bar{\Phi}^\top(ct)$$

$$= \bar{\Phi}(cs)\left[V(0) + \frac{1}{c}\int_0^{(t\wedge s)} \bar{\Phi}^{-1}(cu)LH\, \Sigma(\bar{\rho}(cu))\left(\bar{\Phi}^{-1}(cu) LH\, \Sigma(\bar{\rho}(cu))\right)^\top \mathrm{d}u\right]\bar{\Phi}^\top(ct).$$

It therefore follows that if $V(0)$ is scaled by $1/c$, we obtain that $\Gamma(cs, ct)$, with cell lengths scaled by c, is equal to $\Gamma(s,t)/c$ with unscaled cell lengths.

To empirically validate the scaling property that we established above, we consider the network segment featuring in Section 4.1 for $d = 20$ and each cell having length 0.5 km. Specifically, for $c \in C := \{1, 10, 20, 30, \ldots, 1000\}$, we scale the cell length $\ell_i \mapsto c\ell_i$ for each cell i, and evaluate the model on $[0, c\tau]$ with $\tau = 600/3600$ h. We consider for each $c \in C$ the mean $\bar{\rho}(cs)$ and variance $cV(cs)$ at equidistant points in time (i.e., times $s \in \{t_0, \ldots, t_{50}\} \subset [0, c\tau]$, with $t_i = c\tau i/50$ and $i = 0, 1, \ldots, 50$). These are then compared against the solution obtained for $c = 1$:

$$\max_{c \in C} \max_{t_k} \|\bar{\rho}(t_k) - \bar{\rho}(c\, t_k)\| = 7.18 \cdot 10^{-11},$$

$$\max_{c \in C} \max_{t_k} \|V(t_k) - V(c\, t_k)\| = 2.60 \cdot 10^{-9},$$

where the norms over vectors/matrices are taken as the maximum absolute difference between corresponding entries. We conclude that this numerical experiment confirms that the solutions indeed match.

4.3. Impact of cell length on approximation accuracy

In light of the above scaling properties of the model, one may wonder how the accuracy of the Gaussian approximation is influenced by the choice of the cell lengths. The considerations in the previous subsection entail that when cells become too large, the dynamics of traffic flow are no longer captured accurately. Therefore, as we argued in Section 2.3.1, cells should be taken as small as possible, while still being large enough to accommodate sufficiently many vehicles to justify the Gaussian approximation. In this subsection we demonstrate the robustness of this approach. Concretely, for two different choices of cell lengths satisfying the above constraints, we show that the outcomes are hardly distinguishable, as desired.

In the instance considered, we take a road segment of 8 km long. In Model 1 we divide the segment into $d = 16$ equally-sized cells of length 500 m, whereas in Model 2 we take $d = 10$ equally-sized cells of length 800 m. The fundamental diagram in both models is the one used in Section 4.1. The arrival rate to the segment is $\lambda = 800$ veh/h and the maximal departure rate is $\nu = 1800$ veh/h. Both models are initialized with a density of 5 veh/km with a standard deviation of 1. Note that with these settings, in both models the per-cell aggregation level is large enough to validate a Gaussian approximation.

The two models are evaluated using both the Gaussian approximation and simulation, on a time interval of 500 s. For the simulation, we generate 1000 sample paths per model, from which we estimate the d-dimensional vector with the mean numbers of vehicles per cell, as well as the corresponding $d \times d$ covariance matrix, both at 51 equidistant time points $\{t_0, t_1, \ldots, t_{50}\}$ in $[0, 500]$. The same quantities are computed using the Gaussian approximation.

To compare the output for $d = 16$ and $d = 10$, we evaluate (at every timepoint t_k) the mean and variance of the number of vehicles corresponding to a subsegment of equal length. We take for this the final 4 km of the road segment, corresponding in both models with an integer number of cells (i.e., 8 cells for Model 1 and 5 cells for Model 2). We denote by $\mu_n(t_k)$ and $\sigma_n(t_k)$, respectively, the mean and standard deviation of the number of vehicles in these cells at time t_k for Model n, computed via the Gaussian approximation method. We denote by $\hat{\mu}_n(t_k)$ and $\hat{\sigma}_n(t_k)$ the same quantities, but then estimated from the simulated sample paths. The quantities $\mu_n(t_k)$ and $\sigma_n(t_k)$ can be obtained in the evident manner, i.e., $\mu_n(t_k)$ by summing over the final $d/2$ entries in the mean vector, and $\sigma_n(t_k)$ by summing the entries of the right-lower $(d/2) \times (d/2)$ submatrix of the covariance matrix.

In Fig. 10, we have plotted $\mu_n(t)$ and $\mu_n(t) \pm \sigma_n(t)$, as well as $\hat{\mu}_n$ and $\hat{\mu}_n(t) \pm \hat{\sigma}_n(t)$, as a function of time, for $n = 1, 2$. The figure shows that both models produce highly similar means and variances, in both the Gaussian method and the simulation. In addition, the maximum relative differences between the compared quantities using the Gaussian estimates are

$$\max_{k\in\{0,\ldots,50\}} \frac{|\mu_1(t_k) - \mu_2(t_k)|}{\frac{1}{2}\left(\mu_1(t_k) + \mu_2(t_k)\right)} = 0.0189, \qquad \max_{k\in\{0,\ldots,50\}} \frac{|\sigma_1(t_k) - \sigma_2(t_k)|}{\frac{1}{2}\left(\sigma_1(t_k) + \sigma_2(t_k)\right)} = 0.0725, \qquad (8)$$

and for the simulated estimates we have

$$\max_{k\in\{0,\ldots,50\}} \frac{|\hat{\mu}_1(t_k) - \hat{\mu}_2(t_k)|}{\frac{1}{2}\left(\hat{\mu}_1(t_k) + \hat{\mu}_2(t_k)\right)} = 0.0204, \qquad \max_{k\in\{0,\ldots,50\}} \frac{|\hat{\sigma}_1(t_k) - \hat{\sigma}_2(t_k)|}{\frac{1}{2}\left(\hat{\sigma}_1(t_k) + \hat{\sigma}_2(t_k)\right)} = 0.0683. \qquad (9)$$
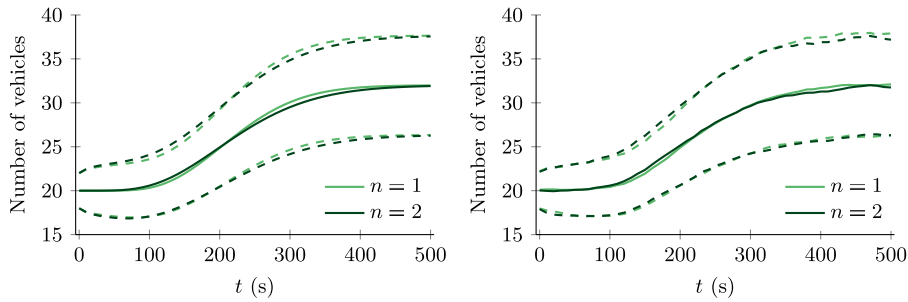
Fig. 10. The left figure is a graphical comparison of $\mu_1(t)$ and $\mu_2(t)$ (solid lines) and of $\mu_2(t) \pm \sigma_2(t)$ and $\mu_2(t) \pm \sigma_2(t)$ (dashed lines), as a function of $t \in [0, 500]$. On the right, the same comparison is made for the quantities $\hat{\mu}_n(t)$ (solid lines) and $\hat{\mu}_n(t) \pm \hat{\sigma}_n(t)$ (dashed lines), with $n = 1, 2$.
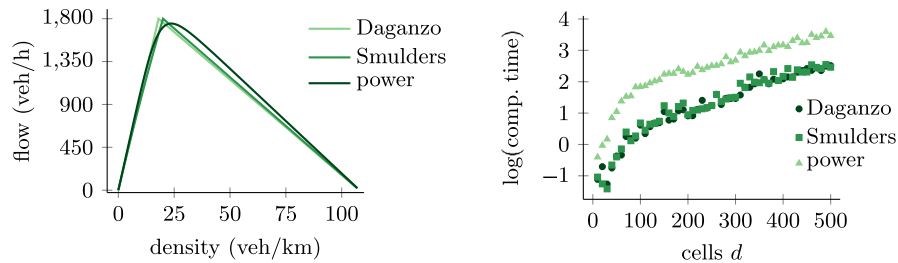

Fig. 11. Three different fundamental diagrams (left) and corresponding computation times as a function of the number of cells d (right).

Fig. 10 and the numbers in (8) and (9) show that Models 1 and 2 give highly similar predictions for the distribution of the number of vehicles simultaneously present on the subsegment. This indicates that the choice of the cell lengths has a modest impact on the accuracy, as long as the constraints stated in Section 2.3.1 are met.

4.4. Complexity for different fundamental diagrams

We conclude this section by studying the influence of fundamental diagrams on the computation times. Recall that the discrete flux function in our model is derived from an (for now unspecified) fundamental diagram by solving the associated Riemann problem. We now aim at quantifying, for various types of fundamental diagrams, the computation time as a function of the number of cells d. For typical fundamental diagrams, the number of calculations to evaluate the flow between two cells does not depend on d. As such, one expects that changing the discrete flux function (or, equivalently, the fundamental diagram) does not have any significant impact on the computational complexity, in that the discrete flux functions depend in roughly the same way on d, potentially with different proportionality constants.

To corroborate the above principle, we consider three discrete flux functions, derived from frequently used fundamental diagrams. In addition to the one by Daganzo, which we have been using thus far, we consider Smulders' fundamental diagram (Smulders, 1990) and the power-function variant of the generic fundamental diagram given in Kessels (2019, Section 2.2.1) (see the references for the precise functional form). We take the parameters for the Daganzo discrete flux function as in Section 4.1. The parameters for the Smulders and power-function discrete flux function are chosen such that the free-flow velocity, backwards wave speed, and maximum possible flow are matched as close as possible. See the left window in Fig. 11 for a graphical representation of the three fundamental diagrams. The right window of Fig. 11 shows the computation time for the three flux functions, as a function of the number of cells d, confirming the expected behavior described above. When working with a mixture of the fundamental diagrams (e.g. some cells having the Daganzo discrete flux function and other the Smulders one), an interpolation provides an impression of the computation time needed.

5. Discussion and generalizations

At this point, we have demonstrated the potential of the Gaussian approximation for efficient numerical evaluation of stochastic traffic flows. In our setup the vehicle-density process is Markovian, meaning that the underlying transition times are exponentially distributed (with a parameter that depends on the vehicle densities in the sending and receiving cells). In this section we experimentally show that this exponentiality assumption is of minor impact, in that can be substantially relaxed without compromising the accuracy of the approximation.

In addition, we discuss a series of generalizations that further enhance our model's scope. So far, to keep our exposition as clear as possible, we have been working with a relatively basic variant, but, as it turns out, our approach can be extended in various ways.

5.1. Exponentiality assumption

In our Markovian model, the gaps between vehicles (measured in units of time), also referred to as headways, are exponentially distributed, with a mean that depends on the current network population (and follows from the discrete flux function associated to the chosen fundamental diagram). The exponentiality assumption is particularly convenient, as it brings us into the realm of Markovian models, thus allowing the formal derivation of the Gaussian approximation. As in practice headways are typically non-exponential, one wonders what the impact of our exponentiality assumption is.

The exponential distribution is known to have a standard deviation that is equal to its mean, or, in other words, it has a *coefficient of variation* (defined as the ratio of the standard deviation and the mean) equal to one. For headway distributions, this coefficient of variation is typically strictly smaller than one, in particular in congested or high traffic-flow regimes (Ye and Zhang, 2009, Table 1) when the headways are relatively deterministic. In this subsection we show that for headway distributions with smaller coefficients of variation, the Gaussian methodology still provides accurate approximations.

To test the impact of the headway distribution we use the class of Coxian distributions. This class is a subset of the class of phase-type distributions, that contains distributions with an arbitrarily small coefficient of variation. In our setup, we consider a specific type of Coxian distribution, namely where the associated non-negative random variable X is given by

$$X = \begin{cases} X_1 & \text{with probability } 1 - p_1, \\ X_1 + X_2 & \text{with probability } p_1(1 - p_2), \\ X_1 + X_2 + X_3 & \text{with probability } p_1 p_2, \end{cases}$$

where X_1, X_2 and X_3 are independent exponential random variables, say with parameter $\mu > 0$. Observe that X always has one exponential phase, with probability p_1 a second phase on top of this, and then with probability $p_1 p_2$ even a third. It is readily verified that

$$\mathbb{E}X = \frac{1 + p_1 + p_1 p_2}{\mu}, \quad \text{Var } X = \frac{1 + 2p_1(1 + p_2) - p_1^2(1 + p_2^2)}{\mu^2}.$$

To systematically assess the robustness of the Gaussian approximations, we consider the following adapted version of our model in which the headways have a given coefficient of variation (that differs from 1). We let the time until the next type-j vehicle jumps from cell i to cell $i + 1$ now be given by a Coxian random variable (rather than an exponentially distributed one). The parameter μ, that was underlying the exponential distribution that we used in our original model, is (time-dependently) chosen such that $\mathbb{E}X = 1/q_{ij}(\rho(t))$, to make sure that the mean of the headway distribution is in line with the fundamental diagram By choosing the parameters $p_1, p_2 \in [0, 1]$ appropriately, we can make sure that the headway distribution has the desired coefficient of variation. Observe that, technically speaking, with these non-exponential inter-cell transition-time distributions, the model is no longer Markovian. One can make it Markovian again, however, by also keeping track of the residual number of exponential phases that the leading vehicle in each cell has to complete.

In order to verify the performance of approach for different headway distributions, we perform the following experiment. We focus on $d = 3$ cells of length 500 m, in the time interval $[0, 1000/3600]$ (time in hours). We take $m = 1$, with arrival rate $\lambda = 1800$ veh/h and $\nu = \lambda/8$. Furthermore, we take Daganzo's variant of the fundamental diagram (see, e.g., Mandjes and Storm (2021, Example 3.2)), with parameters $v^f = 100$ km/h, $w = 20$ km/h, $\rho^{\max} = 108$ veh/km, and $q^{\max} = 1800$ veh/h. Based on 1000 simulations we estimate the mean and standard deviation of $\rho(\cdot)$ at a 1000 equidistant time points in $[0, 1000/3600]$, i.e., as a function of time. This we do for two Coxian headway distributions, corresponding to coefficients of variation equal to 0.7 and 0.5; from Ye and Zhang (2009), we anticipate these numbers to be representative for congested and high traffic-flow regimes. We compare our output to the mean and standard deviations for $\rho(\cdot)$, at the same time points, as provided by our Gaussian approximations (being based on the model with exponentially distributed headways, having a coefficient of variation equal to 1).

The output of this experiment can be found in Figs. 12 and 13. We have plotted dashed lines for the confidence interval centered around the mean, with a width that equals twice the standard deviation. Importantly, observe that the means (corresponding to the three values of the coefficient of variation) almost match. As expected, when reducing the coefficient of variation, the standard deviation goes down as well, but this effect is relatively modest. We thus conclude that our approach is robust, in that for headway distributions with a coefficient of variation of 0.7 or 0.5, the Gaussian approximation remains accurate. The slightly non-smooth behavior for small t is due to the chosen fundamental diagram having a point at which it is non-differentiable; see Mandjes and Storm (2021, Remark 4.4). We also note that working with exponentially distributed headways can be considered as conservative: headways with a coefficient of variation smaller than 1 lead to (slightly) smaller standard deviations.

5.2. Models generalizations

Our experiments predominantly focused on systems in which the parameters are constant over time. We included one example where the arrival rates at on-ramps were varying with time, but used piece-wise constant rates for these processes. In general, our framework allows the external arrival processes (parametrized by λ_j), the departure processes (parametrized by ν_j), and the turning rates in merge and diverge cells to vary with time. The technical condition for this to hold, is that the resulting vehicle-density processes (i.e., the counterparts of (1)) can be described by non-homogeneous Poisson processes that have their rate function as fluid limit.
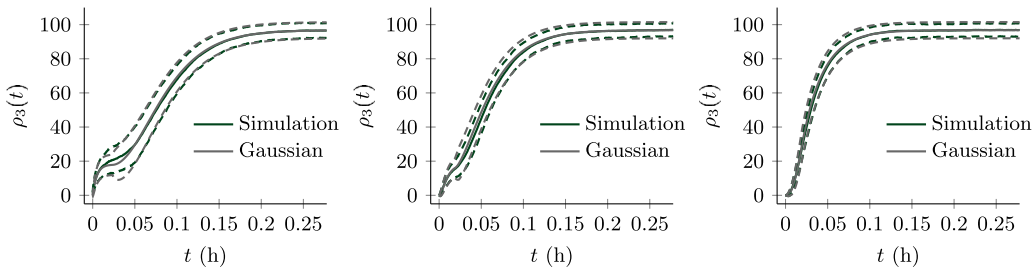
Fig. 12. Comparison of mean and standard deviation of $\rho(\cdot)$, for the Coxian case with coefficient of variation equal to 0.7 (simulation) and the exponential case (Gaussian).

Fig. 13. Comparison of mean and standard deviation of $\rho(\cdot)$, for the Coxian case with coefficient of variation equal to 0.5 (simulation) and the exponential case (Gaussian).
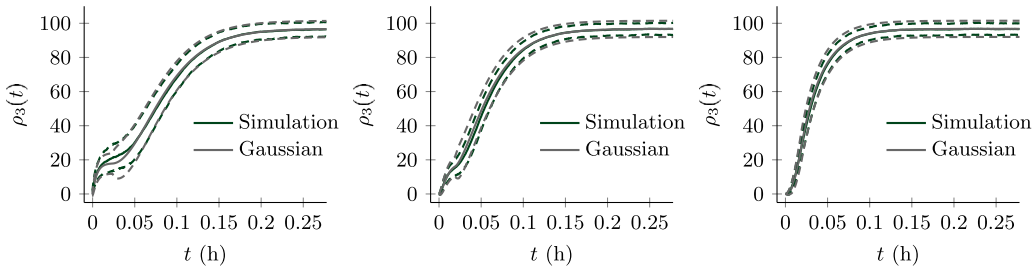
A relevant aspect that we did not emphasize, is that the flux function can be chosen cell-specific. In technical terms, the condition is that the flux between pairs of cells is Lipschitz continuous in the cell-densities involved. Therefore, it is allowed to take different parameters for different cells, but also to vary the functional form of the flux function per cell. For instance, if there are compelling reasons to do so, one could take Daganzo's flux function in one pair of cells and Smulders' in another. Therefore, it is possible to accurately calibrate the model to a practical situation, by estimating the fundamental diagram for every segment separately.

In this paper we exclusively analyzed single-class examples, but, as mentioned previously, the approach remains valid for multi-class fundamental diagrams. In Mandjes and Storm (2021), several experiments are conducted using a two-class model, which enables one to assess the impact of, e.g., trucks on a road-traffic network. With a fundamental diagram describing the interaction between automated vehicles and ordinary vehicles, one could use our approach to numerically evaluate relevant future scenarios.

We conclude this section with a short discussion on the class of models we have been working with. The most detailed class of stochastic models distinguish individual vehicles, keeping track of where they reside at every point in time. Evidently, this approach would lead to a model with an excessively large dimension that would not allow any computations. In models in which one 'zooms out', one predominantly observes the fluid limit, in line with the conservation laws, but not providing any insight into fluctuations around the deterministic limit. In that sense, our approach can be seen as a 'middle ground': there is some aggregation, so as to facilitate computability (in particular yielding huge reduction of complexity in comparison to models that work with individual vehicles), but not too much, to be able to still capture the system's inherent randomness in a sound manner. A consequence is that the precise position of individual vehicles in cells are abstracted away from. This for instance means that, in a literal sense, we do not model a situation in which a platoon of cars drives behind a truck. The corresponding fundamental diagram, however, *does* incorporate that such situations every now and then occur. In this sense, microscopic effects are included in the model in a macroscopic manner through the fundamental diagram (or discrete flux function) that is used.

6. Concluding remarks

In this paper, it was shown how a Gaussian approximation, resulting from the scaling limit results of Mandjes and Storm (2021), can be used in road traffic networks to efficiently compute the vehicle-density distribution, jointly at both different points in the network and at different times. The setup is easy to use since it can be implemented through off-the-shelf ODE-solvers, is orders of magnitude faster than simulation, and is capable of evaluating the impact of combinations of effects and/or measures (accidents, increased traffic load, velocity regulation, ramp metering, etc.) simultaneously. Moreover, the presented approach is highly flexible as the user can choose to implement any fundamental diagram, and works on networks of arbitrary topology. Our framework effectively generalizes the class of discrete-space kinematic wave models, in that it reproduces their deterministic part while at the same time it is capable of capturing their stochastic fluctuations.

Future research could concern exploring how the model under consideration can be extended to include more complex road traffic dynamics. For instance, the basic kinematic wave models have been criticized for not being able to capture phenomena

such as capacity drop, vehicle acceleration and hysteresis. Since the model in this paper is a generalization of the discrete-space kinematic wave models, it will not capture these phenomena either. The challenge lies in identifying model extensions under which the Gaussian process approximation still applies. A different approach could be to consider the differential equations for the mean and (co-)variances of the Gaussian processes, and modify them such that they cover more complex road traffic dynamics. Another potential research direction could relate to further optimizing the efficiency of the underlying computational procedures.

CRediT authorship contribution statement

Pieter Jacob Storm: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Michel Mandjes:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Bart van Arem:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing.

References

Ahmed, A., Ngoduy, D., Adnan, M., Baig, M., 2021. On the fundamental diagram and driving behavior modeling of heterogeneous traffic flow using UAV-based data. Transp. Res. A 148, 100–115.

Boel, R., Mihaylova, L., 2006. A compositional stochastic model for real time freeway traffic simulation. Transp. Res. B 40 (4), 319–334.

Chanut, S., Buisson, C., 2003. Macroscopic model and its numerical solution for two-flow mixed traffic with different speeds and lengths. Transp. Res. Rec. 1852, 209–219.

Daganzo, C., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp. Res. B 28, 269–287.

Daganzo, C., 1995. The cell transmission model part II: network traffic. Transp. Res. B 29, 79–93.

Hale, J., 1969. Ordinary Differential Equations. In: Pure and Applied Mathematics: a Series of Texts and Monographs, Wiley.

Hoogendoorn, S., Bovy, P., 2001. State-of-the-art of vehicular traffic modelling. Proc. Inst. Mech. Eng. I 215 (4), 283–303.

Jabari, S., Liu, H., 2012. A stochastic model of traffic flow: Theoretical foundations. Transp. Res. B 46, 156–174.

Jabari, S., Liu, H., 2013. A stochastic model of traffic flow: Gaussian approximation and estimation. Transp. Res. B 47, 15–41.

Jabari, S., Zheng, J., Liu, H., 2014. A probabilistic stationary speed density relation based on Newell's simplified car-following model. Transp. Res. B 68, 205–223.

Karatzas, I., Shreve, S., 2012. Brownian Motion and Stochastic Calculus. Springer.

Kessels, F., 2019. Traffic Flow Modelling. Springer.

Laval, J.A., Leclercq, L., 2008. Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. Transp. Res. B 42 (6), 511–522.

Laval, J.A., Toth, C.S., Zhou, Y., 2014. A parsimonious model for the formation of oscillations in car-following models. Transp. Res. B 70, 228–238.

Lebacque, J., 1996. The godunov scheme and what it means for first order traffic flow models. In: International Symposium on Transportation and Traffic Theory. pp. 647–677.

Lee, S., Ngoduy, D., Keyvan-Ekbatani, M., 2019. Integrated deep learning and stochastic car-following model for traffic dynamics on multi-lane freeways. Transp. Res. C 106, 360–377.

Li, J., Chen, Q.-Y., Wang, H., Ni, D., 2012. Analysis of LWR model with fundamental diagram subject to uncertainties. Transportmetrica 8 (6), 387–405.

Lighthill, M., Whitham, G., 1955. On kinematic waves. I: Flood movement in long rivers. II: A theory of traffic flow on long crowded roads. Proc. R. Soc. 229A, 281–345.

Logghe, S., Immers, L., 2008. Multi-class kinematic wave theory of traffic flow. Transp. Res. B 42, 523–541.

Lu, J., Osorio, C., 2018. A probabilistic traffic-theoretic network loading model suitable for large-scale network analysis. Transp. Sci. 52, 1509–1530.

Lu, J., Osorio, C., 2021. On the approximation of queue-length distributions in large-scale urban networks. arXiv e-prints, arXiv–2104.

Maerivoet, S., De Moor, B., 2005. Traffic Flow Theory. Technical Report 05–154, Katholieke Universiteit Leuven.

Mandjes, M., Storm, J., 2021. A diffusion-based analysis of a multi-class road traffic network. Stoch. Syst. 11, 60–81.

Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. J. Physique 2, 2221–2229.

Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic part I: General theory. Transp. Res. B 27 (4), 281–287.

Ngoduy, D., 2011. Multiclass first-order traffic model using stochastic fundamental diagrams. Transportmetrica 7 (2), 111–125.

Ngoduy, D., 2021. Noise-induced instability of a class of stochastic higher order continuum traffic models. Transp. Res. B.

Ngoduy, D., Lee, S., Treiber, M., Keyvan-Ekbatani, M., Vu, H., 2019. Langevin method for a continuous stochastic car-following model and its stability conditions. Transp. Res. C 105, 599–610.

Osorio, C., Flötteröd, G., 2015. Capturing dependency among link boundaries in a stochastic dynamic network loading model. Transp. Sci. 49 (2), 420–431.

Osorio, C., Wang, C., 2017. On the analytical approximation of joint aggregate queue-length distributions for traffic networks: A stationary finite capacity Markovian network approach. Transp. Res. B 95, 305–339.

Panda, M., Ngoduy, D., Vu, H.L., 2019. Multiple model stochastic filtering for traffic density estimation on urban arterials. Transp. Res. B 126, 280–306.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., Wang, Y., 2003. Review of road traffic control strategies. Proc. IEEE 91, 2043–2067.

Qu, X., Zhang, J., Wang, S., 2017. On the stochastic fundamental diagram for freeway traffic: Model development, analytical properties, validation, and extensive applications. Transp. Res. B 104, 256–271.

Richards, P., 1956. Shock waves on the highway. Oper. Res. 4, 42–51.

Schader, M., Schmid, F., 1989. Two rules of thumb for the approximation of the binomial distribution by the normal distribution. Amer. Statist. 43, 23–24.

Smulders, S., 1990. Control of freeway traffic flow by variable speed signs. Transp. Res. B 24, 111–132.

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. Transp. Res. B 45, 507–533.

Tampère, C.M., Corthout, R., Cattrysse, D., Immers, L.H., 2011. A generic class of first order node models for dynamic macroscopic simulation of traffic flows. Transp. Res. B 45 (1), 289–309.

Treiber, M., Kesting, A., 2017. The intelligent driver model with stochasticity-new insights into traffic flow oscillations. Transp. Res. Procedia 23, 174–187.

van Wageningen-Kessels, F., van Lint, H., Vuik, K., Hoogendoorn, S., 2015. Genealogy of traffic flow models. EURO J. Transp. Logist. 4, 445–473.

Wong, G., Wong, S., 2002. A multi-class traffic flow model–an extension of LWR model with heterogeneous drivers. Transp. Res. A 36, 827–841.

Ye, F., Zhang, Y., 2009. Vehicle type–specific headway analysis using freeway traffic data. Transp. Res. Rec. 2124 (1), 222–230.

Yuan, K., Laval, J., Knoop, V.L., Jiang, R., Hoogendoorn, S.P., 2018. A geometric Brownian motion car-following model: towards a better understanding of capacity drop. Transp. B: Transp. Dyn..

Yuan, Y., Zhang, Z., Yang, X.T., Zhe, S., 2021. Macroscopic traffic flow modeling with physics regularized Gaussian process: A new insight into machine learning applications in transportation. Transp. Res. B 146, 88–110.

Zhong, R., Sumalee, A., Pan, T., Lam, W., 2013. Stochastic cell transmission model for traffic network with demand and supply uncertainties. Transp. A: Transp. Sci. 9 (7), 567–602.