DTA2014 SYMPOSIUM:

Introducing the Core Probability Framework and Discrete-Element Core Probability Model for efficient stochastic macroscopic modelling

Simeon Calvert, s.c.calvert@tudelft.nl (corresponding author)
Henk Taale, h.taale@tudelft.nl
Maaike Snelder, maaike.snelder@tno.nl
Serge Hoogendoorn, s.p.hoogendoorn@tudelft.nl
All with the Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Transport & Planning, The Netherlands

Abstract

In this contribution the Core Probability Framework (CPF) is introduced with the application of the Discrete-Element Core Probability Model (DE-CPM) as a new DNL for dynamic macroscopic modelling of stochastic traffic flow. The model is demonstrated for validation in a test case and for computational efficiency on two simple networks. The CPF extends a base model, such as the Cell Transmission Model (CTM), by considering each traffic variable as a discrete stochastic variable denoted as a probability distribution of values for each traffic variable in time and space. Traffic is propagated along a link using the base model and through a larger network with the application of probability merging algorithms at the nodes. Due to the incorporation of probability in the core of traffic propagation, the necessity for multiple acts as an internalisation of the Monte Carlo routine in the CPF for fast and efficient calculation of uncertainty. Initial tests cases show that the DE-CPM has the potential to reduce computation time multi-tenfold compared to regular Monte Carlo simulation. Such developments allow the application of stochastic dynamics traffic models to be more readily applied in practice.

Keywords

Stochastic traffic flow modelling, Macroscopic traffic flow models, probabilistic modelling

1. INTRODUCTION

Stochastic variations and uncertainty typify human life and equally the world around us, and are no less present in the world of traffic flow. Application of traffic flow analysis carried out using traffic flow models aims to represent the world of traffic flow in simulation. Traffic models are simplifications of reality and make assumptions to allow for a fast and efficient modelling of situations. These assumptions should however have minimal effect on the deviation of results to what would be expected in reality. In the case of stochastic variations and the inclusion of uncertainty, many assumptions are made that have a greater effect on model outcomes than may be desirable.

In many traffic models stochastic variation is ignored or assumed to be of limited importance to the outcome of simulations. In many cases reducing the input of variables in a traffic model to average or representative values can have detrimental effect of the simulation results, and may lead to biased outcomes in relation to what may be found from empirical data (Calvert et al. 2012; Mahmassani et al. 2012; van Lint et al. 2012). It is argued that the stochasticity in traffic cannot be reduced prior to traffic flow simulation and cannot be expected to give the same outcome as if the reduction had not taken place. Instability in traffic, including network effects in congestion, lead to a non-linear propagation of stochastic variation, especially for the more extreme cases. In turn greater traffic flows and congestion will lead to higher values for travel times and delays than can be derived from averaged or representative input values (Calvert et al. 2012). It is therefore imperative to explicitly consider stochastic variation in traffic flow modelling, when this variation is present in the considered scenarios and networks.

In this contribution a new stochastic macroscopic framework is introduced which, combined with the relevant dynamic network loading (DNL) models, tackles many challenges in macroscopic modelling and is developed with a view for easy and efficient application in practice.

The Core Probability Framework (CPF) is a probabilistic framework for modelling multi-dimensional variations in capacity and traffic demand in dynamic macroscopic traffic flow. The CPF extends a base model, such as the Cell Transmission Model (CTM), by considering each traffic variable as a stochastic variable denoted as a probability distribution of the chance of values for each traffic variable. The CPF is accompanied by the Discrete-Element Core Probability Model (DE-CPM) as an example of a possible DNL model. The DE-CPM is introduced as an internalisation of the Monte Carlo routine in the core of the traffic model.

In section 2 the existing literature on this topic is reviewed, while in section 3 the main issues concerning stochastic macroscopic traffic flow modelling are described. This is followed in section 4 by a description of the conceptual framework of the CPF and the application of the DE-CPM DNL model. In section 5 an explanation is given how the model addresses the issues mentioned in section 3. Section 6 shows a demonstration case of the model in practice and the potential calculation time gains for two networks. Finally, section 7 describes the current developments of the model and conclusions of this paper.

2. STOCHASTIC MACROSCOPIC TRAFFIC MODELLING

Since the 1990's there has been a gradual increase in effort towards improving traffic flow modelling through the explicit inclusion of stochastic variation. Initially a focus was placed on Monte Carlo simulation and later the focus shifted more towards internalised stochastics. In Monte Carlo simulation various input values for the traffic variables are sampled and applied in simulation for a N number of simulation to approach a distribution of possible outcomes. Although Monte Carlo simulation has been widely applied, mainly due to its relative simplicity and effectiveness, the method has its drawbacks. Main concerns in traffic modelling in the past have been the computational load of the method (Chang et al. 1994; Chen et al. 2002; Sumalee et al. 2011) and the presence of correlation between input variables. The incorporation of variance reduction methods, such as Importance sampling or Latin Hypercube sampling, have helped to reduce the computational effort of such models as well as the use of more powerful computers (Jonnalagadda et al. 2001; Hess et al. 2006; van Lint et al. 2012; Calvert et al. 2014). Furthermore, recent developments in marginal simulation approaches offer an alternative solution to a heavy computational load in Monte Carlo approaches (Corthout et al. 2011). In marginal simulation a significant overlap between traffic flow in successive simulation iterations is presumed. By only simulating the marginal difference in traffic flow, repetitive network loading with a full dynamic macroscopic model is not required. Therefore the marginal simulation method only requires a single full initial model simulation and thereafter simulates the marginal differences using Monte Carlo simulation with a first-order based kinematic model, leading to a gain in computational efficiency. Correlation between input variables may be considered prior to simulation at the sampling stage (Chen et al. 2002). Variables with dependencies may also have probabilities which rely on the values sampled from other variables. In this way correlation between two or more variables is included and allows for a realistic simulation. However calculating non-bias outcomes in situations in which correlations are more complex and, furthermore, have dependencies on variables in the model, becomes much more difficult (Chang et al. 1994). In many approaches the extent of bias is presumed to be limited and therefore little attention is spent on this difficulty.

More recent developments in stochastic macroscopic modelling are found in stochastic extensions of existing mainstream traffic models. Boel and Mihaylova (2006) proposed an extension to the CTM with stochastic elements. Rather than reconstructing the CTM as piece-wise structure based on traffic states, they defined the sending and receiving functions from the CTM as random variables in which the dynamics of the average speed in each cell is stochastically varied. The purpose was to incorporate stochasticity in the heart of the model at link level, which may propagate through an entire network through cell interaction. However, as their approach only considers a single stochastic scenario at a time, repetitive simulations are required to compose a probability distribution of the outcomes. Similar approaches were proposed by Sun et al. (2003) focussing on the explicit defining of traffic states. A main reason for considering multiple traffic states is the avoidance of nonlinearity in the fundamental relation, which is difficult to quantify otherwise. Jabari and Liu (2012) argued that presuming non-linearity, while being mathematically beneficial, may lead to inconsistency with the original deterministic dynamics. Therefore Jabari and Liu (2012) proposed to include stochasticity as a function of the uncertainty in the driver gap choice, represented by the random vehicle headway. In doing so, they argue that non-linearity is avoided in continuous time as all traffic dynamics may be derived to the longitudinal car following behaviour. (Sumalee et al. 2011) proposed a further extension of the CTM in which traffic states are explicitly defined as a stochastic bilinear system. Their Stochastic CTM (S-CTM) avoids non-linearities in the original CTM and considers variation through propagation of the probability of traffic states and corresponding densities as the likely values and surrounding standard deviation. The S-CTM also demonstrated computational efficiency as a *one shot* model in which multiple iterations using a Monte Carlo routine are avoided. This greater efficiency is however also achieved through the simplification of the probability distributions to the aforementioned

Gaussian characteristics, such as the median and standard deviation. Although a legitimate choice, this reduces the accuracy of the probabilistic estimation by presuming a set distribution, which in many not be the case.

Analysis of stochastic variation due to randomness in driver behaviour has led to developments in stochastic modelling for both microscopic and macroscopic models.

Variations in traffic flow are easily viewed empirically from fundamental diagram plots. Therefore it is not surprising that stochasticity is also included in (macroscopic) traffic models by means of a stochastic fundamental diagram. Li et al. (2009) make a strong argument that a simple, but effective manner of probabilistic modelling is to make use of a stochastic fundamental diagram. Such a diagram is constructed through a flux function obtained from random elements observed from speed-density data. Kim and Zhang (2008) also previously described stochasticity in the fundamental diagram by defining the growth and delay of perturbations from random fluctuations in both the gap time and transitions between traffic states. In their work they closely examined fluctuations in car following to derive their defined *gap time*. Boel and Mihaylova (2006) also make use of similar fundamental diagrams in their stochastic switching traffic state model, previously mentioned. While these models address the incorporation of variation in the model, this is performed in a simplified fashion, such that traffic states are not all well defined (Sumalee et al. 2011), or fail to fully deal with other stochastic modelling challenges, such as spatiotemporal correlations.

Other models involving stochastic variation relate to a wide number of analytical approaches that have been suggested, especially in relation to travel time reliability (Du and Nicholson 1997; Clark and Watling 2005), however these are not purely considered as stochastic traffic flow models and are therefore not considered here. Furthermore, efforts in stochastic modelling are also present for analytical approaches and focus on queuing models combined with traffic flow theory. This area of research, while in stochastic modelling, applies inherently different solutions, and therefore is not elaborated on in detail here. Some literature for the interested reader can be found in (Hall 1999; Van Woensel and Vandaele 2007; Osorio et al. 2011).

Despite recent developments, challenges remain for the development of stochastic macroscopic traffic flow models and more so for their practical application. The important issues still facing stochastic macroscopic modelling, and yet not completely addressed in a single model, are discussed in the following section.

3. IMPORTANT ISSUES

In the theoretical development, but also for the practical application of dynamic stochastic macroscopic traffic flow models, there are a number of issues that have not been solved in full or in combination with each other. In some cases one issue may be addressed at the expense of another. In this section four important issues are discussed:

1. Computational efficiency
2. Spatiotemporal dependency
3. Stochastic propagation of probability
4. Generality of stochastic variation

3.1 COMPUTATIONAL EFFICIENCY

Originally the issue of computational efficiency arose with the application of Monte Carlo simulation in traffic models. Often performing hundreds of simulations was time consuming and acted as a deterrent to apply stochastic variation. Even through the application of variance reduction techniques and faster and more powerful computers, this remains an issue. A trend that counteracts such advancements originates from a desire to apply more complex traffic models on larger and more detailed networks. Also an increasing number of (stochastic) variables demand a greater computational effort that somewhat undermines hardware and software advancement.

The development of *one shot* models, which largely do away with the necessity for repetitive simulations have a great potential to allow for stochastic simulation at a lesser computational cost. Such models as the S-CTM (Sumalee et al. 2011) and that of Jabari and Liu (2012) are at the forefront of these developments. A danger however is that a simplification of the stochastic input or propagation may be required to allow one shot models to be effective. The opposite effect may be an over-complicated model without simplification, but at a cost computational efficiency and even possible application. Therefore the challenge is not just in reducing computational load, but doing so in a way that a model is not reduced in stochastic and modelling accuracy. This is a balance that is still in the process of being optimised for stochastic modelling and is especially relevant for the practical application of models.

3.2 SPATIOTEMPORAL DEPENDENCY

Incorporation of spatial and temporal dependant variation from different sources brings a further issue of correlation on a number of levels. On a temporal plane it is clear that a stochastic element will affect traffic during a certain time frame, possibly with differing severity. A basic example is that of an accident that reduces road capacity. At the time an accident occurs, the capacity is affected differently than during the aftermath and the clean-up, but nevertheless the capacity reduction is correlated in time, as a natural consequence of a chain of events. In the same way there is also a spatial correlation. The capacity reduction affects the location of the accident, but due to congestion propagation, also affects both upstream capacity and traffic flow. A further complexity in dependence comes from not only considering a single stochastic influence variable, such as the capacity, but also the traffic demand. In the case of an accident, drivers may reroute, shift departure time, etc. This does not only affect traffic flow in time, but also in space. Furthermore, correlation effects also exist between the traffic demand and road capacity in some instances. When considering a greater number of variables, the dependency relations explode.

In many cases some of these dependencies are presumed non-existent for ease of modelling (Clark and Watling 2005; Sumalee et al. 2011). Especially for the interdependent correlations between variables this is readily the case, while spatiotemporal dependencies must be considered on some level to avoid disutility of a model. Even then, these correlations may be simplified by means of presumptions or transformations (Clark and Watling 2005; Jabari and Liu 2012). It should not immediately be presumed that a less than full consideration of dependency will have large detrimental effects on model outcomes, as there are cases in which this is clearly the case (Calvert et al. 2012), however the possibility thereof should always be considered.

3.3 STOCHASTIC PROPAGATION OF PROBABILITY

In traffic flow models it is commonplace for traffic to propagate through a link and network. However upon including stochastic probability in traffic flow modelling, the probabilities of traffic values also propagate in time and space with traffic (Lebacque et al. 2007; Hoogendoorn et al. 2008). For Monte Carlo simulation this is not an issue, as each simulation is a single probability value and therefore no probability value is required to be considered. For one shot models there is a challenge to propagate probability information without compromising model accuracy or one of the other important issues, such as computational efficiency.

In models which apply stochastic effects through the fundamental diagram, traffic flow is presumed to propagate in an identical fashion to that of a regular flow model. In a stochastic fundamental diagram, probabilities are stochastically applied in the shape of the diagram. In the S-CTM, median and standard deviations of traffic variables are propagated through time and space, dependent on the relevant traffic state. It is not uncommon to only consider a median and standard deviation, as this requires the least computational effort and still gives a good estimation of variational spread. However more in-depth analysis is harder as the underlying distribution is not preserved. Furthermore, such an approach often presumes probability distributions to be symmetrical according to a presumed shape, which is not always the case. In such a case biases are allowed, which may not accurately represent the underlying distribution. It should however be noted that these biases may be small compared to the overall error level.

3.4 GENERALITY OF STOCHASTIC VARIATION

Inclusion of stochastic variation does not only demand solid and accurate modelling, but also realistic and correct model input. The level of stochastic input depends on which variables are considered stochastic. These may be the time headway (or gap time) between vehicles, capacity values, traffic demand values, or even 'lower level' variables, such as vehicle population or probability of accidents. Depending on how a model processes the stochastic variables, these may be offered to the model as a complete distribution, either of a specific form or empirical, or as a description of variations, such as median, standard deviation and possibly a shape parameter. The difficulty with this issue is that of generality. A set parametric shape of probable values for a set variable may not be valid for every location on a network or under certain other conditions. Furthermore, such variables may not pertain to a set distribution type. Often presumptions are made to how general distributions or variations are. In many instances white noise may be applied to known representative values to imitate variation (Helbing et al. 2001; Jabari and Liu 2012). The validity of such approaches is not often considered and is taken as a model assumption. However here there is also room for improvement, when applying stochastic variation to traffic flow models. In the case of stochastic fundamental diagrams, the difficulty of generality may also arise. In some cases allowing specific local data to influence the extent of stochastic variation can help solve this.

3.5 SUMMARY OF ISSUES

It is of course the case that each issue influences the others in some way. This is a main reason why individual solutions for each issue do not necessarily yield an overall solution for all the issues. Figure 1 gives a rough estimation of the dependencies between the issues. We derive that especially the manner of stochastic propagation of probability in traffic is a key issue. There is a strong influence from this issue to both the manner in which the spatiotemporal dependency is influenced and the extent to which stochastic variables can be dealt with generically. It may be that certain presumptions for dealing with uncertainty propagation may limit how stochastic variables are defined. Furthermore each issue affects the computation time of a model and in most cases contributes to a lower computational efficiency. There are situations possible that may lead to shorter computational times, such when a process inherently or even implicitly allows for parallelisation. When setting out on tackling one of the issues, the effect on the others should not be ignored, moreover the effect should explicitly be considered for model usefulness.

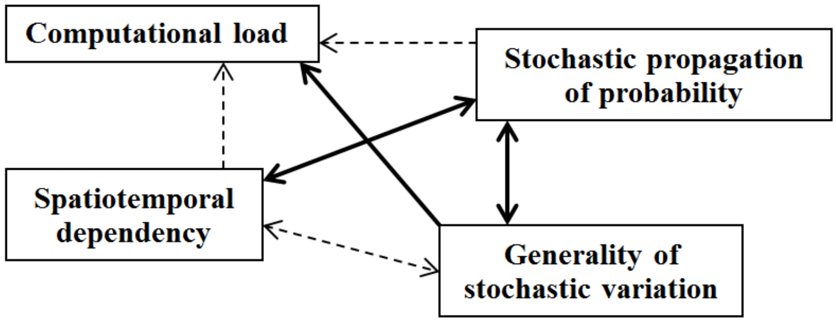

Figure 1 Interrelations between the main modelling issues (continuous and dashed lines indicates strong and weak relationships resp.)

4. CORE PROBABILITY MODELLING

In this section the framework for the Core Probability modelling and the underlying assumptions are explained. The Core Probability Framework (CPF) extends existing macroscopic traffic flow models to allow stochastic behaviour in traffic to be internalised in the traffic flow model which it extends. Internalisation here refers to the manner in which stochasticity is present in the model, where Monte Carlo simulation is a clear example of external stochastic influence. Initial application of the CPF makes use of the Cell Transmission Model (CTM) as base model. The basic premise entails replacing single traffic variables in time and space, such as the density, in a model with a distribution of that same traffic variable, also in space and time. The distribution, denoted as a vector, consists of predefined probabilities of various possible values of the considered traffic variable at a certain time and location, therefore transforming the traffic variables into stochastic variables. The general dynamics of the base model are kept the same as the deterministic version of the model. In such a way, traffic is propagated through a link (or network) considering possible valid values of each traffic variable with a set probability, using already validated traffic flow dynamics from the base model. The input distributions are empirically determined for specific locations and/or scenarios or from generic empirical analysis (Calvert et al. 2014; van Stralen et al. 2014).

The framework allows different probabilistic models for propagation of the stochastic traffic flows to be developed and applied. In this contribution we further present the Discrete-Element Core Probability Model that makes use of the framework. A more detailed description of the framework and this model are given in the subsequent subsections. This begins with a short explanation of the applied base model (4.1). The concept of the CPF is given in 4.2 and is followed in section 4.3 by the description of the manner in which probability is included in the DE-CPM, how it is propagated, and how congestion and traffic states are dealt with. A simple numerical example is shown to conclude the section (4.4).

4.1 BASE MODEL

The Core Probability Framework makes use of a base model, which dictates the manner in which traffic flow is propagated, and considers stochastic probabilities in the core of a macroscopic traffic model. The base model applied here is the first order Cell Transmission Model (CTM) (Daganzo 1994; Daganzo 1995). The CTM describes traffic using a discretised form of the Lighthill-Whitham-Richards (LWR) model (Lighthill and Whitham 1955). The LWR model is governed by the law of conservation of vehicles (eq.(1)), and the fundamental relation (eq.(2)):

$$\frac{\partial k(x,t)}{\partial t}+\frac{\partial q(x,t)}{\partial x}=0 \tag{1}$$

$$q(x,t)=Q_E(k(x,t)) \tag{2}$$

Here $\partial k(x,t)$ denotes the change in density in time, t, and space, x. $\partial q(x,t)$ denotes the same for the intensity, while Q_E is the fundamental relation between the density and flow, which is explained in more detail later on.

In the CTM the traffic flow at the interfaces between two cells, q, is determined by a sending and receiving function, denoted here as the demand, D, and supply, S, which closely represent the available capacity in a cell and the desired traffic flow into a cell:

$$q^{x_m \to x_{m+1}}(k(x,t))=\min(D_m(k(x,t)),S_{m+1}(k(x,t))) \tag{3}$$

The demand function D is calculated by the largest flow or capacity of cell m in relation to eq.(2), and the supply function S by the desired outflow from the previous cell according to the fundamental traffic characteristics of the preceding cell. The base model is applied in its discrete form for use in the Core Probability Framework and governs the main dynamics of traffic flow.

4.2 CORE PROBABILITY FRAMEWORK

The main premise of the Core Probability Framework (CPF) is the incorporation of uncertainty in the core of the model as probability distributions. While regular Monte Carlo simulation applies uncertainty through multiple simulation iterations, for the CPF these are internalised. This approach allows for a *one shot* simulation run and an increased efficiency in simulation. The uncertainty is applied in the form of (discrete) empirical probability distributions, which describe the variations in traffic variables in the model and are primarily applied as cumulative probability functions of the traffic demand at the origins and the capacity of each cell. A graphical description of the Core Probability Framework is shown in Figure 2b, alongside the general framework of a Monte Carlo routine as a comparison over a similar macroscopic traffic model for a simple three cell road stretch (Figure 2a).

The figure clearly shows the evasion of multiple simulations in the case of the CPF in comparison to a Monte Carlo routine over the same base model. The CPF is in its self not a DNL, but rather the framework which states that distributions are explicitly propagated through time and space in combination with the dynamics of the base model. The example of the Discrete-Element Core Probability Model (DE-CPM) is given in this contribution as a possible DNL model, that may be applied in the framework, which describes how the distributions of the stochastic traffic variables are propagated through the network. In the rest of this section, the CPF is explained for the application of the DE-CPM. Other core probability models may also be applied to the CPF and are in development, but are not discussed any further in this paper.

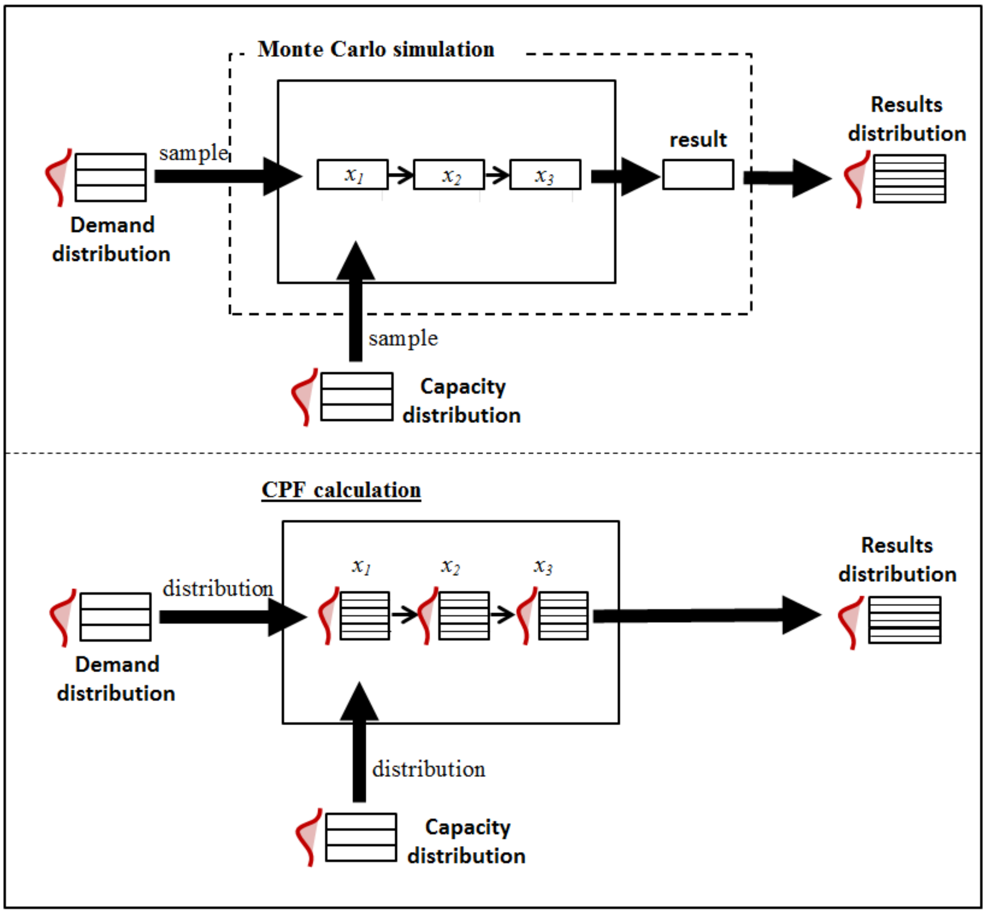

Figure 2a-b Conceptual overview of the (a - above) Monte Carlo traffic simulation framework and (b - below) the Core Probability Framework

4.3 Discrete-Element Core Probability Model

- ***Concept***

The Discrete-Element Core Probability Model (DE-CPM) is DNL model that makes use of the Core Probability Framework to propagate traffic through a link and network. The DE-CPM describes the traffic variables as a distribution, denoted as a vector, which consists of static probabilities of various possible values of the considered traffic variable at a certain time and location. For each variable at each time step identical static probability elements are used in the distribution. Each discrete element in the distribution is explicitly kept from interaction with other elements as the flow distributions are propagated through the network. This approach basically creates an internalisation of the Monte Carlo routine, in which each discrete element or 'scenario' is kept separate. In such a way, traffic is propagated through a link (or network) considering possible valid values of each traffic variable with a set probability, using already validated traffic flow dynamics from the base model.

In the following paragraphs the Core Probability Framework is defined for application of the Discrete-Element CPM as network loading model.

- ***Inclusion of probability***

In classical first order models, each variable is represented by a single value for each point in time, t, and space, x. In the core-probability approach a further variable is added, which represents the probability of the density occurring, and sequentially the traffic flow, $\boldsymbol{q}$, and the speed, $\boldsymbol{v}$. This further transforms the variables from a single value in time and space into a probability distribution in the same time and space, represented by their corresponding vector.

Presuming static values for the probability elements avoids the necessity to explicitly define the probabilities of the values corresponding to the probabilities for each cell (m) in each time step (n).

Initially in the continuation of the description, for the reason of clarity, a further presumption is made that each value in the probability vector has identical probability. This assumptions also entails that the discrete probability values for each probability element are set for the entire simulation for all time steps (n), cells (m), and for each variable (k, q, v):

$$\boldsymbol{p}:\ p_1 = p_2 = \ldots = p_i \quad \forall(k, q, v) \tag{4}$$

We are currently in the process of developing a discretisation technique that presumes the individual probability elements not to be equal, but rather generically set to values that capture the extremities of the distribution to a better extent and in doing so, reduce the required number of elements in the discrete variable.

Now let the random variable $K(x,t)$ denote the density on a cell [x, x+dx] and at time t. Let $p_i(x,t)$ denote the accompanying probabilities. Such a relation is given as:

$$P(K(x,t) = k) = p_i(x,t) \tag{5}$$

Note that the values of k are discrete and hence a discrete probability function can be used. However such a notation indicates a variable probability as a function of given densities. The CPM presumes set probability elements, and therefore the random density variable $K(x,t)$ is defined as a function of set probabilities instead.

So for example $K(x,t)$, now written as vector $\boldsymbol{k}(x,t\,;\boldsymbol{p})$, denotes all possible values of the density for a moment in time and a location, given the probabilities of these densities. The density vector can also be written as:

$$\boldsymbol{k}(x,t\,;\boldsymbol{p}) = \begin{Bmatrix} k_1(x,t)\ with\ probability\ p_{d.1} \\ k_2(x,t)\ with\ probability\ p_{d.2} \\ \ldots \\ k_i(x,t)\ with\ probability\ p_{d.i} \end{Bmatrix} \tag{6}$$

This notation is much closer to that applied in Fuzzy Logic, in which a crisp number is denoted as having multiple possible values, each with their own probability (Buckley 2005). Here, the notation is borrowed from Fuzzy Logic Theory, while applying General Probability Theory, which states in this case that k is a stochastic variable, which has various values with predefined probabilities.

From now on we will only use the short form for the density vector, rather than the description on the right hand side of eq. (5). The addition of the vector $\boldsymbol{p}$ includes all possible values of the appropriate variable with identical probabilities of each value in time and space, so that:

$$\boldsymbol{p} = p_1 + p_2 + \ldots + p_i = 1 \tag{7}$$

Here, i is further limited to a finite value, which is applied as an input parameter of the model. The equations for the conservation of vehicles (eq.1) and the fundamental relation (eq.(2)) now incorporate a further dimension for the probability in time and space, and become dependent on the probability of their value:

$$\frac{\partial \boldsymbol{k}(x,t\,;\boldsymbol{p})}{\partial t} + \frac{\partial \boldsymbol{q}(x,t\,;\boldsymbol{p})}{\partial x} = 0 \tag{8}$$

$$\boldsymbol{q}(x,t\,;\boldsymbol{p}) = Q_E(\boldsymbol{k}(x,t\,;\boldsymbol{p})) \tag{9}$$

The conservation of vehicles therefore remains intact by definition, as each considered element in the probability distribution vector acts as an individual case of the CTM for which conservation has been proven (Daganzo 1994).

- ***Application of stochastic demand and traffic propagation***

External stochastic traffic demand is applied in the model at the peripherals of a network on the inflowing cells. From there on traffic may propagate applying eq (7) and (8) according to the dynamics of the base model. The initial traffic demand contains j_d times j_c number of elements in the probability vector $\boldsymbol{p}$, where j_d is the number of probability elements in the vector for the demand and j_c is the number of probability elements for the capacity, such that each probability vector $\boldsymbol{p}$ is constructed of all possible combinations of p_{jd} and p_{jd}. The initial flow at the network origins is therefore:

$$\boldsymbol{q}(x_0,t_1\,;\boldsymbol{p}) = \{q_{p1}, q_{p2}, \ldots q_{p(jd\cdot jc)}\} \tag{10}$$

Where the probability vector $\boldsymbol{p}$ exists of j_d times j_c elements. This multiplication is performed to accommodate a position in the probability distribution for the outcomes of each combination of traffic demand and capacity.

The variation in the capacity of the network is applied for each cell corresponding to the probability of the capacity of that cell in a similar way to the traffic flow $\boldsymbol{q}$. In a simplified case only bottleneck cells will have varied capacity values, with the other cells yielding identical capacity values for each element in $\boldsymbol{p}$. The capacity contains j_c probability elements for the capacity in both time and space, although for most cells, variation in the capacity has little to no influence where flow is sub-critical:

$$\boldsymbol{q}_{cap}(x_m,t_n\,;\boldsymbol{p}) = \begin{Bmatrix} q_{cap.1}(x,t)\ with\ probability\ p_{c.1} \\ q_{cap.2}(x,t)\ with\ probability\ p_{c.2} \\ \ldots \\ q_{cap.i}(x,t)\ with\ probability\ p_{c.i} \end{Bmatrix} \tag{11}$$

Once the stochastic traffic is on the network, the traffic propagates through the network dependent on the corresponding demand and following the dynamics as previously shown in eq.(8) and eq.(9).

Spatiotemporal dependence is applied as a conditional probability at the entrance of a network, between the initial demand (applied to connector links to get initial densities) and capacity variables. Propagation of this dependence entails that each element in the probability vector of the density corresponds to the same place in the probability vector of the density of the following time step. This is described as the *chain-rule*, as graphically shown in Figure 4 and is further described later in this paragraph and is given in Eq. 14. The *chain-rule* ensures an identical number of elements in the resulting probability vector for propagation through the network, and therefore avoids an explosion of marginal probability elements. Basically, this creates a set of values which can be seen as scenarios of unique traffic demand and capacity combinations.

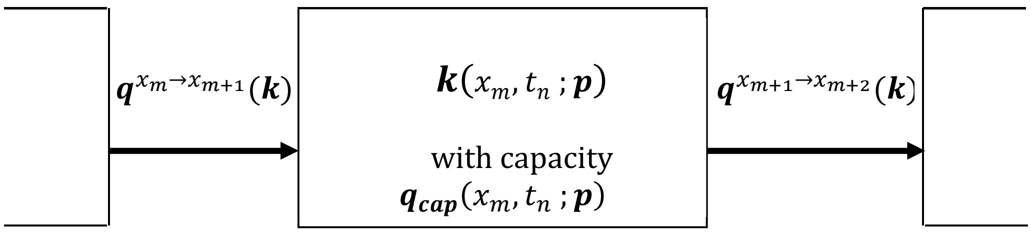

Figure 3 Traffic propagation in the DE-CPM

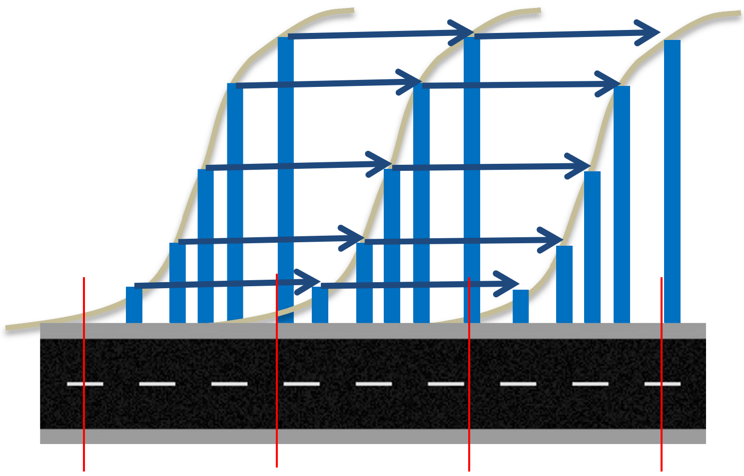

Figure 4 Chain-rule for propagation of traffic variables as discrete elements of a distribution in the DE-CPM

The process is explained as such: there is a traffic demand $\mathrm{q}(\mathrm{x}_1, \mathrm{t}_1)$ with a set of possible values, q_p, corresponding to certain probabilities:

$$q(x_1, t_1 \,; p) = \{q_{p1}, q_{p2}, \ldots q_{pi}\} \tag{12}$$

Calculations in the model are performed using the density, therefore q is transformed using eq. (2) to:

$$k(x_1, t_1 \,; p) = \{k_{p1}, k_{p2}, \ldots k_{pi}\} \tag{13}$$

In the following time step there is a new q and k at location x_1, in line with traffic flow in and out of the cell and in keeping with the conservation of vehicles (eq(1)) :

$$k(x_1, t_2\,; p) = \{k_{p1}, k_{p2}, \dots k_{pi}\} \tag{14}$$

However the position of each element in the $(x_1, t_2\,; p)$corresponds only to that of the element in the same position in the following time step in $\boldsymbol{k}(x_1, t_2\,; \boldsymbol{p})$, so that for each element, i, applies:

$$q(x_1, t_2\,; p_i) \rightarrow q(x_1, t_1\,; p_i) \tag{15}$$

This strict *'chain-rule'*, that for each location in consecutive time steps the same probability must apply, protects the validity of the initial conditional dependence between the capacity and traffic demand in both time and space.

Although the CTM base model, and therefore also CPF / DE-CPM, calculates traffic using the density, it is often required to translate this to the traffic flow $\boldsymbol{q}(x, t\,; \boldsymbol{p})$, for determination of the flux for example. This is performed using the fundamental relation shown in eq.(2), in which each value of q is transformed using a deterministic fundamental diagram. The resulting values of $\boldsymbol{q}(x, t\,; \boldsymbol{p})$ from $\boldsymbol{k}(x, t\,; \boldsymbol{p})$ maintain the same probabilities for each time step and cell in space.

In the same way, the traffic flow on the subsequent cells is also calculated. The only difference is that the supply and demand refer to those of the following cells, x_j. In such a way, one can speak of multiple scenarios in a single procedure, as each element of the marginal probabilities are considered individually for a single variable.

- ***Determination of Congestion***

The sending and receiving functions, or rather demand and supply, $\boldsymbol{d}$ and $\boldsymbol{s}$, are in part determined by the traffic state. Traffic states are in turn determined by the density of traffic in a cell at a specific time. Under congestion, the demand function is equal to the capacity, and the supply function of the outgoing traffic flow:

$$\boldsymbol{d}(x_m, t_n\,; \boldsymbol{p}) = \boldsymbol{q_{cap}}(x_{m-1}, t_n\,; \boldsymbol{p}) \tag{16}$$
$$\boldsymbol{s}(x_m, t_n\,; \boldsymbol{p}) = \boldsymbol{q}(x_m, t_n\,; \boldsymbol{p}) \tag{17}$$

For uncongested states, the demand function is the incoming traffic flow, and the supply function is the available capacity:

$$\boldsymbol{d}(x_m, t_n\,; \boldsymbol{p}) = \boldsymbol{q}(x_{m-1}, t_n\,; \boldsymbol{p}) \tag{18}$$
$$\boldsymbol{s}(x_m, t_n\,; \boldsymbol{p}) = \boldsymbol{q_{cap}}(x, t\,; \boldsymbol{p}) \tag{19}$$

For the Core Probability Framework without capacity variation, congestion is determined by comparison between the probable density and the critical density of a cell:

$$\boldsymbol{Cong}(x, t\,; \boldsymbol{p}) = \boldsymbol{k}(x, t\,; \boldsymbol{p}) \geq k_{crit}(x, t) \tag{20}$$

However, when capacity is also varied, the congestion equation states a distribution vector on either side of the operator:

$$\boldsymbol{Cong}(x,t\,;\boldsymbol{p}) = \boldsymbol{k}(x,t\,;\boldsymbol{p}) \geq \boldsymbol{k_{crit}}(x,t\,;\boldsymbol{p}) \quad (21)$$

- ***Network flow over nodes***

For modelling traffic in networks, it is imperative to consider traffic flow over the nodes. This is usually performed using a node model which deals with the manner in which traffic propagates at convergence and divergence point in a network, but also how other traffic waves, such as congestion may propagate in an upstream direction. This contribution does not aim at developing a stochastic node model, and therefore they will not be reviewed here. For an overview of the state-of-art of node models we refer to (Tampère et al. 2011). The inherent characteristics of the chain-rule, as used in the DE-CPM for the propagation of distributions as an internalisation on the Monte Carlo routine, determine that just about any arbitrary node model that is applicable for the base model may be applied in the CPF.

This is demonstrated for the merge model as described by Daganzo (Daganzo 1995) for an uncongested flow. The merge model describes the maximised flow, Q, from two incoming links, *i = 1,2*, into a single outgoing link 3. As seen already from the CTM, sending flows perpetuate from the upstream links (see eq. (3)). These flows are constrained by the maximum flows that may leave each link: S_1, S_2. Likewise, the receiving downstream link also has a maximum flow that it is capable of receiving: R_2. Therefore we can easily see that traffic flow is constrained by either the traffic demand from the inflowing links or the supply of capacity from the receiving link according to:

$$q_i \leq S_i \qquad \forall i \ \in \{1\,,2\} \quad (22)$$

$$\sum_{i=1,2} q_i \leq R_3 \quad (23)$$

Considering the constraints and convergence of the flow from eq. (22) and (23), it becomes apparent that the flow into the receiving downstream link for uncongested circumstances is:

$$\mathrm{Q} = \min\{S_1 + S_2\,; R_3\} \quad (24)$$

Extension of the node model for use in the DE-CPM extends eq. (22) and (23) by considering each variable as a discrete stochastic variable in which the chain-rule is valid between the corresponding elements of the variables. Hence, equations (22-24) become:

$$\boldsymbol{q}_i(t_1\,;\boldsymbol{p}) \leq \boldsymbol{S}_i(t_1\,;\boldsymbol{p}) \quad (25)$$

$$\sum_{i=1,2} \boldsymbol{q}_i(t_1\,;\boldsymbol{p}) \leq \boldsymbol{R}(t_1\,;\boldsymbol{p}) \quad (26)$$

$$\mathbf{Q}(t_1\,;\boldsymbol{p}) = \min\{\boldsymbol{S}_1(t_1\,;\boldsymbol{p}) + \boldsymbol{S}_{(t_1\,;\boldsymbol{p})1}\,; \boldsymbol{R}(t_1\,;\boldsymbol{p})\} \quad (27)$$

In eq. (25-27) $\boldsymbol{p}$ indicates the entire distribution vector for which is valid p *for* $\forall\ p \in \boldsymbol{p}$ according to the previously defined chain-rule for an arbitrary variable, *f*: $f\,(p_i) \ \rightarrow f(p_i)$. Graphically, it is very easy to observe how the propagation of traffic in the DE-CPM does not require special attention for nodes beyond the introduced theory that is also applicable for stretches. Convergence and divergence of traffic flow at a node are again dealt with according to the dynamics of the base node modal, where each element from the stochastic variables is processed independently.

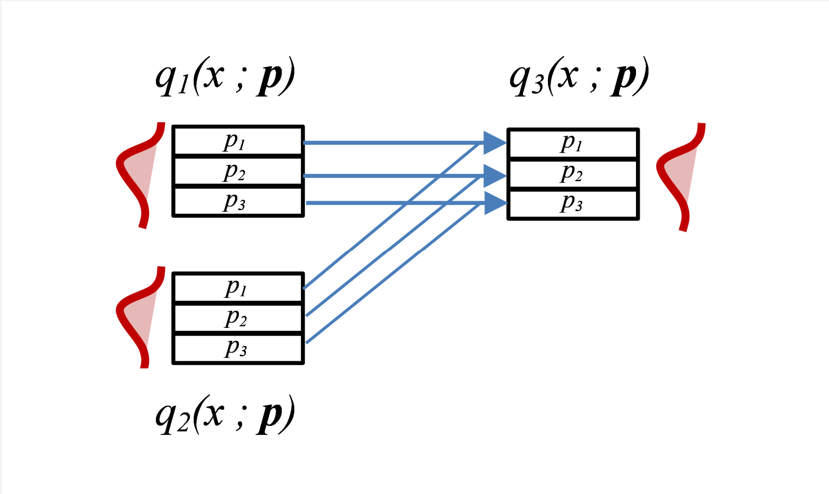

Figure 5 Graphical representation of the DE-CPM for a node merge

The same simple extension applies to other node models and the additional equations that describe the congested states in the node models for application in the DE-CPM. As the chain-rule explicitly keeps the individual elements of the discrete distribution separated for calculation, these act in the same fashion as the deterministic case for which the models are already developed.

4.4 Simple numerical example (both capacity and demand varied)

To demonstrate the manner in which the DE-CPM works, a simple numerical example is given as demonstration. A more elaborate demonstration is given in section 6. The traffic demand at the network peripherals is given as an intensity with a set probability. In this example there is a 50% chance of two different inflow values, and there is 50% of two different capacity values. Therefore there are 4 elements in the demand vector, because the size of $\boldsymbol{q}(x, t\,; \boldsymbol{p})$ is equal to j_d times j_c (see eq. (10)):

$$\boldsymbol{q}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \begin{Bmatrix} 1900 \\ 1900 \\ 2200 \\ 2200 \end{Bmatrix} \tag{28}$$

The capacity values of the cell are also given in the, j_d times j_c number of elements, capacity flow vector:

$$\boldsymbol{q_{cap}}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \begin{Bmatrix} 2100 \\ 2300 \\ 2100 \\ 2300 \end{Bmatrix} \tag{29}$$

Note that the sequences for the values of the flow in the demand vector (Eq 28) are differently arranged over the j_d times j_c elements in comparison to the capacity flow vector (Eq 29).

This flow vector, $\boldsymbol{q}(x, t\,; \boldsymbol{p})$, in eq. (28) is transformed to a density vector, $\boldsymbol{k}(x, t\,; \boldsymbol{p})$, using the fundamental relation $q = Q_E(k)$ in which the critical density is $k_{crit} = 25$. This gives:

$$\boldsymbol{k}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \begin{Bmatrix} 22 \\ 20 \\ 26 \\ 24 \end{Bmatrix} \qquad (30)$$

The probability of congestion is calculated using Eq. (20):

$$\boldsymbol{Cong}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \boldsymbol{k}(x, t\,; \boldsymbol{p}) \geq k_{crit}(x, t) = \left[\begin{Bmatrix} 22 \\ 20 \\ 26 \\ 24 \end{Bmatrix} \geq 25\right] = \begin{Bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{Bmatrix} \qquad (31)$$

Therefore, based on eq.(16) through eq.(19), the demand D and supply S, can be calculated as:

$$\boldsymbol{D}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \begin{Bmatrix} 1900 \\ 1900 \\ 2200 \\ 2200 \end{Bmatrix} \quad \text{and} \quad \boldsymbol{S}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \begin{Bmatrix} 2100 \\ 2300 \\ 2100 \\ 2300 \end{Bmatrix} \qquad (32)$$

The flux between two cells is defined and given as:

$$\boldsymbol{q}^{x_i \to x_{i+1}}\left(x_1, t_1, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \min\big(\boldsymbol{D}(\boldsymbol{k}), \boldsymbol{S}_{i+1}(\boldsymbol{k})\big) = \begin{Bmatrix} 1900 \\ 1900 \\ 2100 \\ 2200 \end{Bmatrix} \qquad (33)$$

The density therefore in the current and following cells in the following time step, t_2, is given by the previous density adjusted by the flux into and out of that cell, during the size of the time step, h. Here we presume an identical inflow into cell x_1 for t_2 as in t_1:

$$\boldsymbol{k}\left(x_1, t_2, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \boldsymbol{k}(x_1, t_1\,; \boldsymbol{p}) + (\boldsymbol{q}^{x_0 \to x_1} - \boldsymbol{q}^{x_1 \to x_2}) \cdot h = \begin{Bmatrix} 22 \\ 20 \\ 26 \\ 24 \end{Bmatrix} + \left(\begin{Bmatrix} 1900 \\ 1900 \\ 2200 \\ 2200 \end{Bmatrix} - \begin{Bmatrix} 1900 \\ 1900 \\ 2100 \\ 2200 \end{Bmatrix}\right) \cdot h \qquad (34)$$

Similarly, the flow into the yet unoccupied cell x_1 is calculated:

$$\boldsymbol{k}\left(x_1, t_2, \boldsymbol{p} = \begin{Bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{Bmatrix}\right) = \boldsymbol{k}(x_1, t_1\,; \boldsymbol{p}) + (\boldsymbol{q}^{x_0 \to x_1} - \boldsymbol{q}^{x_1 \to x_2}) \cdot h = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} + \left(\begin{Bmatrix} 1900 \\ 1900 \\ 2100 \\ 2200 \end{Bmatrix} - \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}\right) \cdot h \qquad (35)$$

This same process repeats itself for each cell in each time step and so on.

5. ADDRESSING THE MAIN ISSUES

This section describes the manner in which the important issues from section 3 are addressed in the CPF and DE-CPM and improve on the current state-of-art.

For **computational efficiency**, the main challenge is to reduce computational load and in doing so, do it in a way that the model is not reduced in stochastic and modelling accuracy. Compared to a Monte Carlo simulation, the CPF does not require multiple repetitive simulations before arriving at a distribution, as the distribution of the traffic variables is explicit to the methodology. Therefore the computational load will be lighter if a single (DE-)CPM simulation run is quicker than the sum of the required number of Monte Carlo simulations on the same base model. It is hypothesised that this is the case, as the DE-CPM has a single computational overhead for the entire distributions, while a Monte Carlo simulation has a computational overhead for each simulation iteration. Furthermore, a lower detail of discretisation is hypothesised to be required for the DE-CPM as the model calculates using distributions throughout. In section 6.2 a demonstration is given of the potential computational gains. Monte Carlo simulation makes use of less efficient random process of sampling, which reduces the completeness of a distribution and therefore requires a greater number of simulations to reach the same level of accuracy, therefore increasing the computational load. On simple network or corridors, the efficiency effect will be limited, however for larger networks and for a greater spread of variation the gains should be greater. It should be noted that Monte Carlo simulation allows for parallelisation, which can significantly improve computation time.

Other CPM-models in the CPF have the potential to be much more efficient, as the consideration of 'scenarios' as internal Monte Carlo draws can be relaxed and allow natural probabilities to be calculated. This cannot be verified however until these models are further developed and tested.

Spatiotemporal dependency is catered for in the DE-CPM through the explicit consideration of correlations at the peripheral of the model and maintenance thereof in propagation through the chain-rule. For other DNL models in the CPF, the manner in which the dependency is dealt with may vary. Reduced to two dependant variables, the traffic demand and road capacity, correlations between possible values of both are explicitly considered in the distributions entering a network at the peripherals. Values in the initial distribution vector of the traffic demand entering the network correspond on an element-to-element bases to that of values of the capacity distribution vector at the same element location. This was explained in section 4. By explicitly maintaining this chain-rule throughout the traffic propagation, independency between traffic demand and capacity is maintained. Dependency in time for both the demand and capacity is also explicitly dealt with outside the model. Input values for certain elements in the distribution vectors follow those of the preceding time step and therefore already consider a logical and dependant propagation from the input vectors in time. Spatial dependency is dealt with in the same way as in the base model and therefore requires no further attention. Simplified, each element in a distribution vector may be seen as a single input value for a single Monte Carlo simulation, therefore it may also be considered as independent from other elements just as a single Monte Carlo iteration is from another Monte Carlo iteration.

Stochastic propagation of probability in traffic flow is performed as described in section 4.3 for the DE-CPM and is also touched upon in the previous paragraph on spatiotemporal dependency. Dealing of this issue is also DNL model dependent and not generic for the CPF. A complete distribution of possible values per traffic variable is present as a distribution in the form of a vector. This vector exists of more elements than is necessary, so to allow each possible value of that vector to correspond to the elements of other vectors and therefore to avoid correlation difficulties. As these distribution vectors are propagated in space and time, there is no need to reduce variables to a representation of the distribution using a set distribution type, median, standard deviation, shape parameter or such like. Although this may lead to a higher computational effort, it maintains a guaranteed accuracy of the propagation of the traffic variables and their probabilities, as the distributions remain intact in the process of propagation. Therefore a greater accuracy can be achieved in comparison to methods that do transform distributions to characteristics of the distribution, mostly to some parametric form.

For the CPF, the question of **generality** is one that is less relevant to the model itself, but rather to the quality of the data and distributions that it is fed with. As the CPF performs calculations using discrete distributions, a reduction of the input data may only happen in the case of rediscretisation for the sake of computational efficiency. Therefore the necessity to apply accurate input distributions for the traffic demand and road capacity is applicable for the local circumstances or from a general distribution if the local situation is not known. Construction of generic input distributions for this purpose, taken from wide spread empirical analysis, makes it easy to apply the CPF without requiring extensive data analysis for each application of the model (Calvert et al. 2014; van Stralen et al. 2014). Nevertheless, this issue is one that is less explicit to the model, as the quality of input data is relevant and independent to all models. However the manner in which a model deals with accurate input is important. The CPF does not simplify input by moulding it to a parametric function, therefore maintaining high level of accuracy and avoiding additional unnecessary biases, contrary to many other models. The CPF makes use of empirical distributions which maintain the characteristics of each distribution as it propagates through a network.

6. TEST CASES DE-CPM

6.1 TRAFFIC PROPAGATION ON A SINGLE ROAD SECTION

Demonstration of the application and validity of the Discrete-Element Core Probability Model (DE-CPM) is given in a test case. The test case aims to show that traffic propagation along a road section in the DE-CPM can accurately resemble traffic flow found from empirical observations. As the case is carried out on a single stretch, there is not much that can be said about the computational efficiency. This is considered in the following sub-section.

The test case is carried out for the A12 motorway in The Netherlands between Utrecht and The Hague (see Figure 6). On this motorway in 2009[1], a lane drop was

[1] Since 2009, this location has been upgraded to four lanes along the entire stretch to eradicate the bottleneck.

present from four to three lanes, which acted as a structural bottleneck at location A. Daily congestion starting at this location near the town of Woerden would be present, especially during the evening peak period. A section of 11 kilometres is considered, of which 10 km upstream and 1 km downstream of the bottleneck. The DE-CPM is fed with data from 63 afternoon peak period observations of the traffic flow between 2 PM and 9 PM from 2009 as a representation of the probability of certain traffic flows appearing. The input for the model is taken exclusively from the most upstream location. Therefore the validation is that of the stochastic traffic propagation. Each observation is considered as an equal probability of a real traffic demand for this location and is therefore given a 100/63 = 1.6% probability for the input at the inflow of the corridor. These traffic flows are fed into the network at the most upstream location.

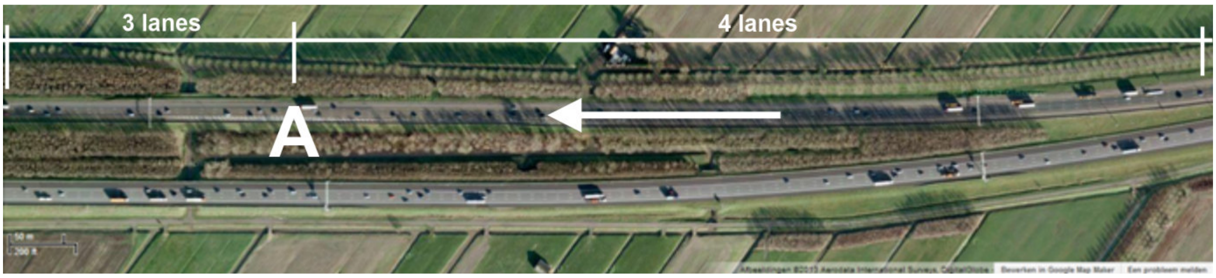

Figure 6 Bottleneck location near Woerden at the considered road section on the A12 used in the case study

A comparison is made based on the ability of the model to accurately predict the propagation of the probabilities of traffic flow and corresponding traffic states between the outcome of the DE-CPM simulation and the empirical data. For this, the unfiltered traffic states in time and space are gathered on the entire corridor. The comparison focusses on the time of traffic breakdown, congestion duration, spill-back distance, and the specific speed values in time and space. This is shown for the median probability (most likely traffic situation) and a further demonstration of the results are given in the form of a 3D congestion probability plot. The results of the median probability are shown in the time-space Figure 6.

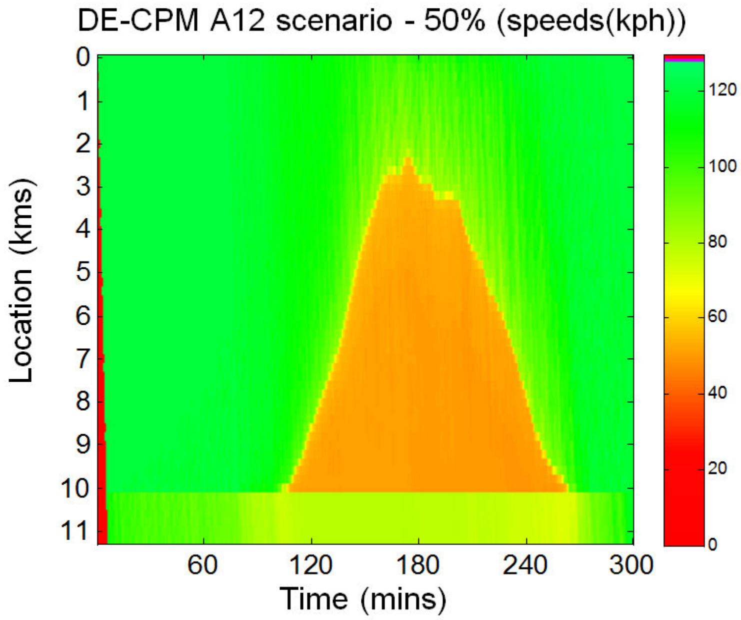

Figure 7 Modelled speed diagram for the median probability in the A12 test case

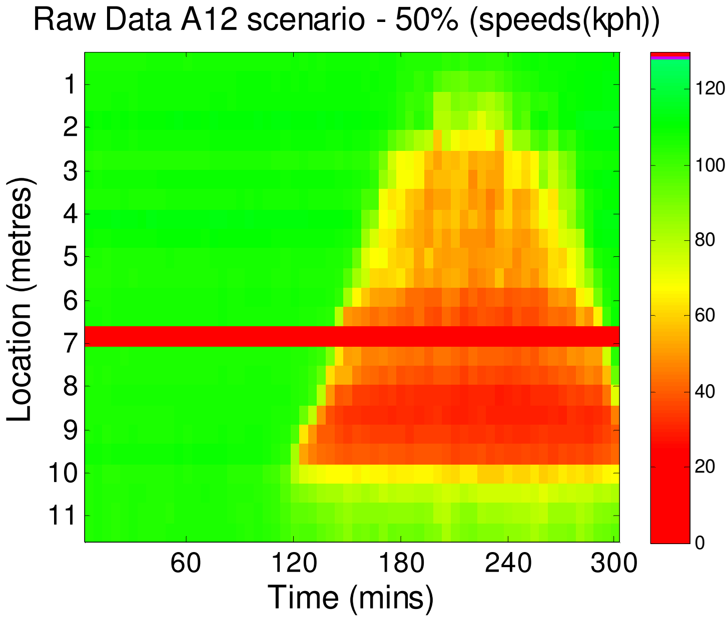

Figure 8 Empirical speed data for the median observation in the A12 test case [2]

The initial results shown in Figure 7 show the simulated median (50%) results from the model, compared to the median from the empirical data shown in Figure 8. The speed values are shown as these give a good indication of where congestion is present, how extreme congestion is and how traffic flow changes the time. Initially the extent of congestion appears to be relatively well modelled. Nevertheless there are certain deviations in comparison to the empirical data. The onset of congestion occurs approximately 10 minutes earlier in the simulation, while congestion lasts for 158 minutes compared to 190 minutes in the data. However, the spillback of congestion in both is of a similar magnitude and deviates no more than 200 meters over a distance of some 9 kilometres. The speed in the heavily congested area of traffic is lower in the empirical data compared to the model (ca. 30 kph versus 40 kph). This may also be a main reason why the duration of congestion differs, as traffic in the simulation may proceed at a slightly higher speed and therefore let congestion disperse earlier. Despite these minor deviations, this initial test case gives cause for optimism. A further fine-tuning of the model parameters when applied in practice may easily compensate for the observed differences.

The CPF allows a vast amount of data to be produced and presented as a probability distribution or in another forms as a direct consequence of the way the CPF works. As each traffic variable is considered as a distribution of possible values, each can therefore be calculated or shown as such at each time step and location. This is demonstrated in Figure 9 in which the congestion probability at each location and for every time step is given. Congestion is defined as such when the critical density is exceeded, while the probability thereof indicates the frequency that congestion is expected to occur for an arbitrary location and time along the corridor. It is possible to show more complex results in a greater number of dimensions, i.e. including the probability as a variable in a diagram, however this leads to difficulties in the interpretation of diagrams. Nevertheless, broad analyses are made much easier and more extensive with the results

[2] The red horizontal line indicates a location at which a faulty detector is present. The speed at this location is returned as null.

from the CPF. Significant computational gains are not found on a single corridor, but rather are expected for networks and for greater variations in stochastic variables. This is further looked at in the following section.

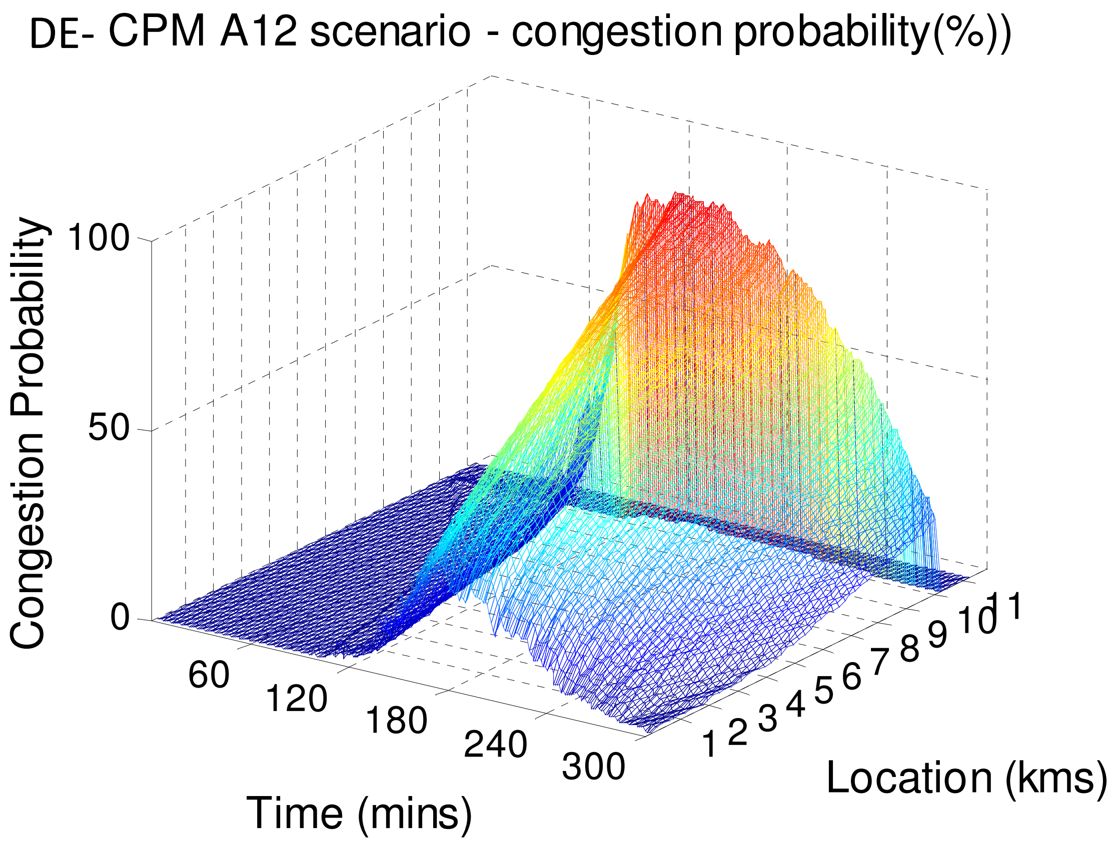

Figure 9 Modelled congestion probability in time and space for the A12 test case

6.2 NETWORK COMPUTATIONAL PERFORMANCE

Performance of the DE-CPM for computational efficiency is tested on two simple networks. In comparison to the previously considered road stretch, variation in traffic flow can interact much more as it propagates through a network and will also include network effects. The considered networks are shown in Figure 10 and 11. Network 1 is a 7 link network with two origins and one destination, while network 2 is constructed from 17 bi-directional links with four origins and destinations. A comparison is made between the application of identical input distributions and capacity distributions in the DE-CPM against a CTM Monte Carlo simulation on the same networks in a MATLAB implementation. In both models the main CTM code is identical, naturally with the addition of the core probability components for the DE-CPM model. Furthermore, both models make use of exactly the same route model, which presumes static turning fractions and all other variables and parameters are kept identical in both cases.

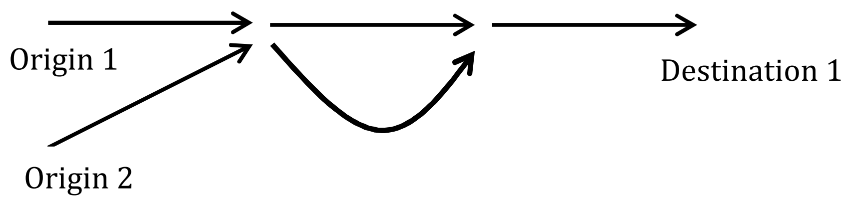

Figure 10 Test network 1

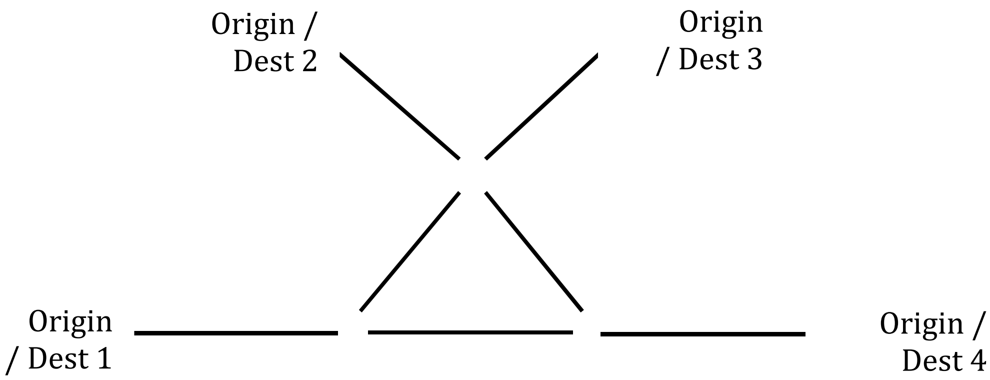

Figure 11 Test network 2

Input for the models are the *network definition*, which includes network characteristics and geometry, *stochastic dynamic demand matrices*, and *stochastic capacity values*. The demand and capacity distributions are kept to a limited number of discrete elements, which also act as the input for the DE-CPM and as each combination for the Monte Carlo routine. The input distributions therefore do not require further discretisation. Besides tests on two different networks, various 'total number of time steps' and various 'number of discrete elements in the input distributions' are applied, as shown in Table 1.

For each scenario, at least five simulations[3] are performed of which the average computation times are given in the last two columns of Table 1. Five

Table 1 Computational speed tests for the DE-CPM

Network	Links (incl inflow)	Origin	Dest	Nodes (excl. connectors)	Time step size (sec)	Time steps	Cap stoch elements	Demand stoch elements	Unique combin-ation of elements	MC run time (sec)	DE-CPM run time (sec)
Netw 1	7	2	1	2	6	200	10	4	40	25	1.1
Netw 1	7	2	1	2	6	400	10	4	40	51	2.3
Netw 1	7	2	1	2	6	200	5	4	20	13	1.0
Netw 1	7	2	1	2	6	800	10	4	40	101	4.3
Netw 1	7	2	1	2	6	800	5	4	20	51	3.9
Netw 1	7	2	1	2	6	200	5	2	10	7.0	1.0
Netw 1	7	2	1	2	6	400	5	2	10	13	2.0
Netw 1	7	2	1	2	6	400	5	4	20	25	2.1
Netw 1	7	2	1	2	6	800	5	2	10	26	3.8
Netw 2	17	4	4	3	6	200	10	4	40	66	2.7
Netw 2	17	4	4	3	6	400	10	4	40	131	5.5
Netw 2	17	4	4	3	6	800	10	4	40	261	11
Netw 2	17	4	4	3	6	200	5	4	20	33	2.6
Netw 2	17	4	4	3	6	200	5	2	10	16	2.5

[3] The reason for multiple simulations is to be sure that there are no or limited variations cased by the computer. Although five simulations are performed for each scenario and model, the differences in calculation time for the five simulations in all cases on the same machine consequentially varied minimally, generally below 2%.

The results of the computation time tests as a function of the number of time steps and discrete elements from the distributions, show some interesting trends. A graphical representation of the results are shown in Figures 12 and 13 for the CTM Monte Carlo and DE-CPM respectively for network 1. The relationship between the number of time steps and the calculation time is approximately linear for both models and has its origin near to a time of zero. The relationship between the number of discrete elements and the calculation time is also approximately linear in both cases. However there is a significant difference between the CTM Monte Carlo and DE-CPM for the incremental increase in relation to the number of discrete elements from the distributions. As may be expected, the CTM Monte Carlo model increases linearly with an origin near to time zero. This is expected as each Monte Carlo simulation makes exactly the same calculations for each combination of inputs, with each calculation taking approximately the same amount of time. The DE-CPM, however, requires a relatively shorter additional time to calculate additional number of discrete elements from the input (and here in the propagation). This is found for both networks and can be clearly observed in Figures 12 and 13. Also it appears that the linear increase with the number of elements does not originate at zero seconds, which indicates some sort of small start-up time. Comparison between the two networks would indicate that the start-up time is dependent on the size of the network.

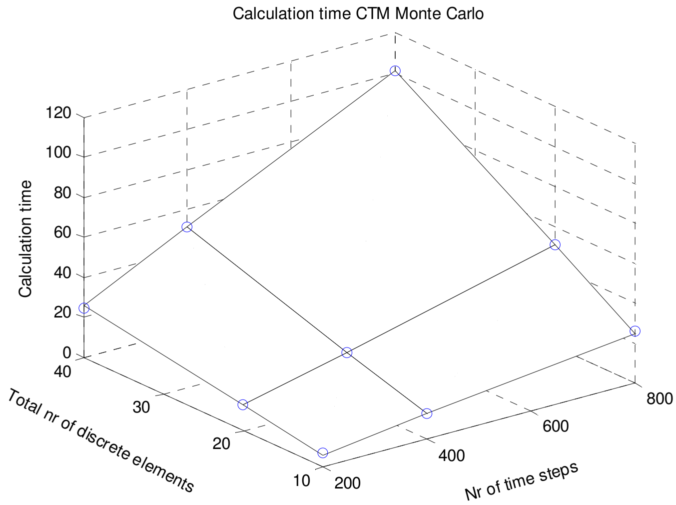

Figure 12 Calculation time CTM Monte Carlo for network 1

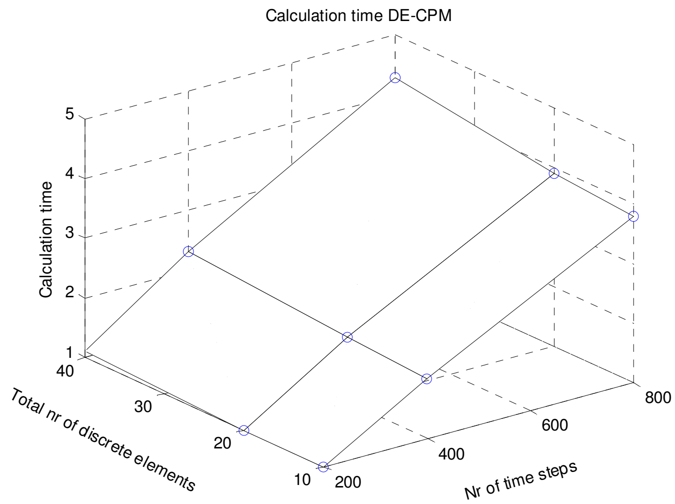

Figure 13 Calculation time DE-CPM for network 1

The consequence of this low coefficient for increasing number of discrete elements is that the DE-CPM is far more capable of efficiently dealing with traffic flows with large amounts of stochasticity in comparison to the compared Monte Carlo routine. This allows simulations to be carried out in which a greater detail of uncertainty may be incorporated at a marginal cost to the computational time.

In absolute terms, the calculation times for the DE-CPM outperform those for the CTM Monte Carlo by a factor 5-20 depending on the size of the network and number of stochastic elements. The results from Table 1 show that for larger networks and for a greater number of discrete elements that the DE-CPM outperforms the Monte Carlo routine to a greater extent. This is in line with the expectations that this model shows its effectiveness best for larger networks and under greater levels of uncertainty. The extent of the calculation time decreases came as a surprise, and was not expected to be so large. However further checks have reconfirmed the results. The possibility of parallelisation for Monte Carlo routines can reduce the computation time, however even compared to parallelisation such gains of 20 times or more for larger networks with the DE-CPM may even be competitive in comparison.

7. CONCLUSIONS AND CURRENT DEVELOPMENTS

In this contribution the Core Probability Framework (CPF) has been introduced with the application of the Discrete-Element Core Probability Model (DE-CPM) as a new DNL for dynamic macroscopic modelling of stochastic traffic flow. An initial validation case has been also been shown as well as an indication of the computational performance on networks. The CPF extends current deterministic traffic flow models by redefining traffic variables in the core of the model as distribution vectors of probable values for each traffic variable. In such a way stochastic variation in traffic is internalised in the model and does away with the necessity of repetitive Monte Carlo simulation. Furthermore a greater degree of flexibility in analysis is obtained, as each individual traffic variable in time and space may be given as a function of their probability. Moreover, the underlying distribution of each traffic variable in space and time is preserved such that the introduction of distribution fitting errors is limited to a minimum. Important issues facing stochastic traffic flow modelling are given in the contribution, and are identified as *computational efficiency*, *spatiotemporal dependency*, *stochastic propagation of probability*, and *stochastic generality*. The DE-CPM addresses each of these issues through element based calculation using the chain-rule and in doing so demonstrates the ability to advance developments in the area of stochastic traffic modelling. In particular the CPF aims to further the possibilities for reliable, accurate, efficient, and most of all, practically applicable stochastic macroscopic traffic flow modelling. The outcome of the calculation time tests on simple networks compared to a CTM Monte Carlo model showed that the DE-CPM has great significant potential to reduce computation times, especially for larger networks and for greater stochasticity. This is mainly due to the small marginal computational costs incurred when increasing the level of uncertainty in the discrete model. With the DE-CPM DNL model, a first step within the framework is taken. Further expansions in the form of more advanced model developments within the framework are ongoing and focus on the propagation of the stochastic variables as distributions without the application of the chain-rule. These developments have the potential to deal with stochastics to a more efficient extent.

ACKNOWLEDGEMENTS

This research is jointly funded by TNO, Netherlands Organisation for Applied Scientific Research, and TrafficQuest, a collaboration between TNO, Delft University of Technology, and Rijkswaterstaat, highway agency of the Dutch Ministry of Infrastructure and Environment.

REFERENCES

Boel, R. and L. Mihaylova (2006). "A compositional stochastic model for real time freeway traffic simulation." Transportation Research Part B: Methodological **40**(4): 319-334.

Buckley, J. J. (2005). Fuzzy probabilities: new approach and applications, Springer.

Calvert, S. C., H. Taale and S. P. Hoogendoorn (2014). "Quantification of motorway capacity variation: influence of day type specific variation and capacity drop." ((Submitted for publication, expected 2014)).

Calvert, S. C., H. Taale, M. Snelder and S. P. Hoogendoorn (2012). Probability in traffic: a challenge for modelling. Proceedings of the Fourth International Symposium on Dynamic Traffic Assignment (DTA 2012), MA, USA.

Calvert, S. C., H. Taale, M. Snelder and S. P. Hoogendoorn (2014). "Improving Probabilistic Traffic Modelling through Advanced Sampling (submitted for publication)." IEEE Transactions on Intelligent Transportation Systems.

Chang, C., Y. Tung and J. Yang (1994). "Monte Carlo Simulation for Correlated Variables with Marginal Distributions." Journal of Hydraulic Engineering **120**(3): 313-331.

Chen, A., H. Yang, H. K. Lo and W. H. Tang (2002). "Capacity reliability of a road network: an assessment methodology and numerical results." Transportation Research Part B: Methodological **36**(3): 225-252.

Clark, S. and D. Watling (2005). "Modelling network travel time reliability under stochastic demand." Transportation Research Part B: Methodological **39**(2): 119-140.

Corthout, R., C. M. Tampère, R. Frederix and L. H. Immers (2011). Marginal dynamic network loading for large-scale simulation-based applications. 90th annual meeting of the Transportation Research Board.

Daganzo, C. F. (1994). "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory." Transportation Research Part B: Methodological **28**(4): 269-287.

Daganzo, C. F. (1995). "The cell transmission model, part II: network traffic." Transportation Research Part B: Methodological **29**(2): 79-93.

Du, Z. P. and A. Nicholson (1997). "Degradable transportation systems: sensitivity and reliability analysis." Transportation Research Part B: Methodological **31**(3): 225-237.

Hall, R. W. (1999). Transportation queueing. Handbook of Transportation Science, Springer**:** 109-150.

Helbing, D., A. Hennecke, V. Shvetsov and M. Treiber (2001). "MASTER: macroscopic traffic simulation based on a gas-kinetic, non-local traffic model." Transportation Research Part B: Methodological **35**(2): 183-211.

Hess, S., K. E. Train and J. W. Polak (2006). "On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice." Transportation Research Part B: Methodological **40**(2): 147-163.

Hoogendoorn, S. P., H. van Lint and V. L. Knoop (2008). "Macroscopic modeling framework unifying kinematic wave modeling and three-phase traffic theory." Transportation Research Record: Journal of the Transportation Research Board **2088**(1): 102-108.

Jabari, S. E. and H. X. Liu (2012). "A stochastic model of traffic flow: Theoretical foundations." Transportation Research Part B: Methodological **46**(1): 156-174.

Jonnalagadda, N., J. Freedman, W. A. Davidson and J. D. Hunt (2001). "Development of microsimulation activity-based model for San Francisco: destination and mode choice models." Transportation Research Record: Journal of the Transportation Research Board **1777**(-1): 25-35.

Kim, T. and H. Zhang (2008). "A stochastic wave propagation model." Transportation Research Part B: Methodological **42**(7): 619-634.

Lebacque, J.-P., S. Mammar and H. H. Salem (2007). Generic second order traffic flow modelling. Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17.

Li, J., Q. Y. Chen, H. Wang and D. Ni (2009). Analysis of LWR model with fundamental diagram subject to uncertainties. The 88th Transportation Research Board Annual Meeting. Washington DC.

Lighthill, M. J. and G. B. Whitham (1955). "On Kinematic Waves .2. A Theory of Traffic Flow on Long Crowded Roads." Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences **229**(1178): 317-345.

Mahmassani, H. S., T. Hou and J. Dong (2012). "Characterizing Travel Time Variability in Vehicular Traffic Networks." Transportation Research Record: Journal of the Transportation Research Board **2315**(1): 141-152.

Osorio, C., G. Flötteröd and M. Bierlaire (2011). "Dynamic network loading: A stochastic differentiable model that derives link state distributions." Procedia-Social and Behavioral Sciences **17**: 364-381.

Sumalee, A., R. Zhong, T. Pan and W. Szeto (2011). "Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment." Transportation Research Part B: Methodological **45**(3): 507-533.

Sun, X., L. Muñoz and R. Horowitz (2003). Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control, IEEE.

Tampère, C. M. J., R. Corthout, D. Cattrysse and L. H. Immers (2011). "A generic class of first order node models for dynamic macroscopic simulation of traffic flows." Transportation Research Part B: Methodological **45**(1): 289-309.

van Lint, J. W. C., O. Miete, H. Taale and S. P. Hoogendoorn (2012). "A systematic framework for the assessment of traffic measures and policies on the reliability of traffic operations and travel time." 91th meeting of the Transportation Research Board.

van Stralen, W., S. C. Calvert and E. Molin (2014). The influence of adverse weather conditions on the probability of congestion on Dutch motorways. the 91st Annual Meeting of the TRB, Washington DC.

Van Woensel, T. and N. Vandaele (2007). "Modeling traffic flows with queueing models: a review." Asia-Pacific Journal of Operational Research **24**(04): 435-461.