QUANTIFYING INPUT-UNCERTAINTY IN TRAFFIC ASSIGNMENT MODELS

Konstantinos Perrakis[a], Mario Cools[b,c], Dimitris Karlis[d], Davy Janssens[a] ,
Bruno Kochan[a], Tom Bellemans[a] and Geert Wets[a,]*

[a]Transportation Research Institute
Hasselt University
Wetenschapspark 5, bus 6
BE-3590 Diepenbeek
Belgium
Fax.:+32(0)11 26 91 99
Tel:+32(0)11 26 91{40, 28, 47, 27, 58}
Email: {konstantinos.perrakis, davy.janssens,
bruno.kochan, tom.bellemans, geert.wets}@uhasselt.be

[b]Centre for Information, Modeling and Simulation (CIMS)
Hogeschool-Universiteit Brussel
Warmoesberg 26
BE-1000 Brussels
Belgium
Tel: +32(0)479 52 62 99
Email: mario.cools@hubrussel.be

[c]Research Foundation Flanders (FWO)
Egmontstraat 5
BE-1000 Brussels
Belgium

[d]Department of Statistics
Athens University of Economics and Business
76 Patision Str, 10434, Athens
Greece
Tel:+30 210 8203920
Email: karlis@aueb.gr

Number of words: 6239
Number of Tables: 1
Number of Figures: 4
Total number of words: 6239 + 5*250 = 7489

Revised paper submitted: November 13, 2011

* Corresponding author

ABSTRACT

Traffic assignment methods distribute Origin-Destination (OD) flows throughout the links of a given network according to procedures related to specific deterministic or stochastic modeling assumptions. In this paper, we propose a methodology that enhances the information provided from traffic assignment models, in terms of delivering stochastic estimates for traffic flows on links. Stochastic variability is associated to the initial uncertainty related to the OD matrix used as input into a given assignment method, and therefore the proposed methodology is not constrained by the choice of the assignment model. The methodology is based on Bayesian estimation methods which provide a suitable working framework for generating multiple OD matrices from the corresponding predictive distribution of a given statistical model. Predictive inference for link flows is then straightforward to implement, either by assigning summarized OD information or by performing multiple assignments. Interesting applications arise in a natural way from the proposed methodology, as is the identification and evaluation of critical links by means of probability estimates. A real-world application is presented for the road network of the northern, Dutch-speaking region of Flanders in Belgium, under the assumption of a deterministic user equilibrium model.

1. INTRODUCTION

Traffic assignment is one of the most crucial parts of transportation analysis. Traffic assignment models take into account the dependencies among OD demand, link flows and path costs, and simulate the interactions between transportation supply and travel-demand in order to deliver an output which describes the mean state of a transportation system and its corresponding overall variability (*1*). Traffic assignment methods flourish in the relative literature; ranging from simple deterministic/stochastic uncongested network models and deterministic/stochastic user equilibrium and system optimum models, for the cases of congested networks, to more advanced methods such as equilibrium assignment with variable demand, multiclass assignment and dynamic process models. Descriptions of models and algorithms can be found in numerous books, as in Thomas (*2*), Patriksson (*3*) and Cascetta (*1*), to name a few. Extensive information for deterministic and stochastic equilibrium assignment is also found in the articles of Florian and Hearn (*4*) and Cantarella and Cascetta (*5*), respectively.

The output of traffic assignment models is generally vital to decisions related to infrastructure expansion and transport policy measures, and simple point estimates – even if they refer to the most likely values – are not sufficient for a proper and safe assessment of the risks associated with such decisions. Therefore, despite the wide range of available assignment models, the need of quantifying more precisely the uncertainty which is related to traffic assignment estimates is strongly present. De Jong et al. (*6*) address the matter of evaluating uncertainty in transport models and note that very few studies include some quantification of uncertainty, in the form of confidence intervals or other related measures. In addition, the authors distinguish between *input-uncertainty* and *model-uncertainty*, and further note that from the limited number of studies which take into account quantification of uncertainty, only one third deals with both types. Although the research of de Jong et al. (*6*) concerns transport models in general, and despite recent developments in quantification of uncertainty in other related fields as in modeling of travel times ((*7*), (*8*)), infrastructure management ((*9*), (*10*)) and travel-demand modeling (*11*), one can safely state that the conclusions of de Jong et al. (*6*) also, and perhaps particularly, hold for applications of traffic assignment models. Deterministic assignment models deliver by definition point estimates of link flows without a corresponding measure of statistical dispersion. In the case of stochastic assignment, statistical variability is associated with path choice behavior which is modeled according to random utility functions instead of deterministic functions. Therefore, applications of stochastic assignment models may incorporate model-uncertainty but not necessarily input-uncertainty. In addition, the behavioral uncertainty component of stochastic equilibrium assignment associates to short-term variations of user-perception (*12*), which makes stochastic assignment suitable for short-term travel-demand applications, when the OD flows refer, for instance, to hourly or daily intervals. For cases of long-term assignment of travel-demand, e.g. monthly or yearly intervals, deterministic user equilibrium assignment remains up to the present the common option.

Given these facts there exists a necessity for more robust traffic assignment estimates, which will incorporate long-term, travel-demand uncertainty. This necessity did not pass unnoticed, as in recent years a significant amount of studies is orientated towards that research direction. Waller et al. (*13*) showed that the expected performance of a network is, not only, not equivalent to the performance of the system under the expected value of travel-demand but that the latter case is also suboptimal. Ukkusuri and Waller (*14*) extended previous work and investigated the performance of seven point estimates of OD demand under the user equilibrium assumption. In the studies of Gardner et al. ((*15*), (*16*)), the impact of demand uncertainty on the pricing of transportation networks is explored by evaluating network performance resulting from single point demand-approximations, multiple points of inflated/deflated demand and meta-heuristic approaches. Sampling approaches are demonstrated in Duthie et al. (*17*), where correlated OD demand realizations are sampled from multivariate truncated-at-zero normal and multivariate lognormal distributions, and iteratively used as input into the user equilibrium model. Other approaches aim to incorporate long-term demand stochasticity directly in the assignment problem, resulting to modifications of the deterministic user equilibrium formulation as a bi-level, non-linear, non-convex mathematical optimization problem. Ukkusuri et al. (*18*) proposed a robust network design problem and utilized a genetic algorithm for the solution, while Ukkusuri and Patil (*19*) formulated a flexible network design problem which can be solved under complementarity constraints.

In this paper we present a methodology for incorporating and quantifying input-uncertainty in traffic assignment which adds to the growing body of recent literature on this subject. The approach is sampling-based and results as an extension of previous work presented in Perrakis et al. (*20*), where a Bayesian covariate-modeling approach for OD matrices derived from census studies is presented. As illustrated in that study, one desired property of Bayesian methods, is the existence of a *predictive distribution* from which multiple predictive OD datasets can be randomly generated, resulting in distributional estimates for all OD pairs. With this research, we demonstrate how the Bayesian OD predictions can be used as input into an assignment model by introducing two distinct inputting methods. In method 1, a specific summary of the predictive OD datasets is calculated first and then assigned to the network, while with method 2 repeated assignments are executed individually for all predictive datasets. The first technique is faster and suitable for obtaining point and interval estimates, whereas the second technique is computationally more intense but also delivers *full information* over the variability of link flows, in the form of Bayesian predictive distributions conditional on the assignment model. These distributions allow for flexibility and innovation in various applications related to infrastructure management and transport policy-making. For instance, as presented in this study, uncertainty concerning identification and evaluation of critical links can be reduced by investigating this matter under purely probabilistic terms.

The proposed methodology, in general, can be viewed as a 2-stage modeling procedure. The first-stage model lies on firm statistical ground and involves the common sequence of model formulation, estimation, comparison – in case of more than one competing models – and finally prediction and validation with respect to the initial OD data. The second-stage model is a traffic assignment model which takes as input the predictions of the first-stage statistical model by utilizing one of the proposed methods presented in this study. The approach is generally applicable, in the sense that first-stage statistical modeling can be applied on short-term as well as long-term OD matrices, while the selection of the traffic assignment model in the second-stage is independent from the statistical model, allowing for deterministic as well as stochastic assignment procedures.

The proposed methods are applied on a realistic context, namely the road network of Flanders, under the assumption of a Deterministic User Equilibrium (DUE) assignment model. Flanders is the northern, Dutch speaking region of Belgium accounting for approximately 60% of the country's population. The corresponding road network contains 308 zones, 97,450 links and runs a total length of 65,296.72 kilometers.

The paper is organized as follows. A brief description of statistical covariate-based OD modeling is presented in section 2. In section 3, we demonstrate how Bayesian OD predictions can be used as input into an assignment model. Results for the road network of Flanders are presented in section 4. Finally, conclusions are summarized in section 5.

2. STATISTICAL COVARIATE-BASED OD MODELING

Statistical OD modeling with covariates is investigated in Perrakis et al. (*20*), where it is demonstrated that the OD estimation problem may be expressed as a purely statistical problem when reliable historical OD information exists, e.g. from census studies. The goal then is to find an appropriate set of covariates, adopt sound distributional assumptions, estimate the parameters of scientific interest and finally evaluate the predictions of the model.

The long tradition of transportation models provides sufficient indications regarding the choice of covariates or explanatory variables. Variables which are used within travel-demand and supply modeling, such as population densities, employment ratios, income levels, car-ownership ratios, measurements related to road networks, traffic and distances, are all suitable candidates for statistical OD modeling. Regarding the distributional assumptions, the OD flows are modeled by appropriate discrete distributions, namely either by the Poisson or by the negative-binomial distributions. The choice between a Poisson and a negative-binomial likelihood depends on the degree of *overdispersion* present in the data, i.e. the ratio of the variance over the mean. For OD flows which are balanced and exhibit no overdispersion, the Poisson likelihood will be most probably sufficient. On the other hand, when the variance of the OD flows exceeds the mean significantly, then a negative-binomial likelihood is more suitable. In this section, we provide a brief description of the Poisson and negative-binomial models within the Bayesian framework and demonstrate how to generate predictions of OD flows from the corresponding predictive distribution of each model.

Let n denote the number of OD pairs and p the number of explanatory variables included in a model. In addition, let $\mathbf{y} = \left(y_1, y_2, ..., y_n\right)^T$ denote the vector of OD flows, $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \beta_2, ..., \beta_p\right)^T$ the vector of regression parameters and $\mathbf{X}$ the design matrix of dimensionality $n \times (p+1)$, containing the intercept and the p explanatory variables, with $\mathbf{x}_i = \left(x_{i0}, x_{i1}, x_{i2}, ..., x_{ip}\right)^T$ being the i-th row of $\mathbf{X}$ related to OD flow y_i and $i = 1, 2, ...n$. The formulation for the Poisson model is

$$y_i \mid \boldsymbol{\beta} \sim Pois(\mu_i), \text{ with } \mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) \qquad \text{(likelihood of OD flows)}$$
$$\boldsymbol{\beta} \sim \mathbf{N}_{\mathbf{p+1}}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \qquad \text{(prior of regression parameter vector } \boldsymbol{\beta}\text{)}$$

and the resulting posterior distribution is

$$p(\boldsymbol{\beta} \mid \mathbf{y}) \propto \prod_{i=1}^{n} \left[\exp\left[-\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right] \left[\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right]^{y_i} \right] \times \exp\left(-\frac{1}{2} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right) \right). \quad (1)$$

For the case of the negative-binomial model we have that

$$
\begin{aligned}
& y_i \mid \boldsymbol{\beta}, \theta \sim NB(\mu_i, \theta), \text{ with } \mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) && \text{(likelihood of OD flows)} \\
& \boldsymbol{\beta} \sim \mathbf{N}_{\mathbf{p+1}}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) && \text{(prior of regression parameter vector } \boldsymbol{\beta}\text{)} \\
& \theta \sim Gamma(a, b) && \text{(prior of dispersion parameter } \theta\text{)}
\end{aligned}
$$

with corresponding posterior distribution

$$
\begin{aligned}
p(\boldsymbol{\beta}, \theta \mid \mathbf{y}) \propto \prod_{i=1}^{n} \left[\frac{\Gamma\left(y_i + \theta\right)}{\Gamma\left(\theta\right)} \frac{\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)^{y_i} \theta^{\theta}}{\left[\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) + \theta\right]^{y_i + \theta}} \right] \times \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \left(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right) \times \quad (2) \\
\times \theta^{a-1} \times \exp\left(-b\theta\right).
\end{aligned}
$$

Hyperparameters $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, a and b may be set accordingly, if prior knowledge is available. Otherwise, $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = n \times \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \times 10^3$ would result to a non-informative prior for the regression parameter vector, namely to one of the "benchmark" priors discussed in Fernández et al. (*21*), and $a = b = 10^{-3}$ would result to a non-informative prior for the dispersion parameter with mean equal to 1 and variance equal to 1,000, as in Ntzoufras (*22*).

The two un-normalized posterior distributions in expressions (1) and (2) do not result in known distributional forms and therefore direct inference on the parameters is not feasible. Nevertheless, efficient Markov Chain Monte Carlo (MCMC) methods (*23*), such as the Metropolis-Hastings algorithm (*24*), can be applied in order to simulate from these non-standard posterior distributions. Statistical inference is then straightforward, since the posterior sample generated by MCMC may be utilized to calculate directly any summary of interest.

An interesting feature of Bayesian methods is the possibility of simulating *predictions* or *replications* of data by implicitly averaging over the parameters of each model. Let us assume that the final MCMC samples for both models are of size K, that is, draws of parameters $\boldsymbol{\beta}^{(k)}$ – common in both models – and $\theta^{(k)}$, for $k = 1, 2, ..., K$, are available, and let $\mathbf{y}^{\mathbf{pred}} = \left(y_1^{pred}, y_2^{pred}, ..., y_n^{pred}\right)^T$ denote one predictive vector containing n OD flows. Then, for the case of the Poisson model the elements of $\mathbf{y}^{\mathbf{pred}}$ are simulated from $y_i^{pred} \sim Pois\left(\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right)$, for $i = 1, 2, ...n$. Repetition over $\boldsymbol{\beta}^{(k)}$, for $k = 1, 2, ..., K$, results to K such predictive vectors; in vector notation we generate

$$
\mathbf{y}^{\mathbf{pred}(k)} \sim Pois\left(\exp\left(\mathbf{X}\boldsymbol{\beta}^{(k)}\right)\right). \quad (3)
$$

Similarly, the corresponding predictions from the negative-binomial model will be simulated from

$$
\mathbf{y}^{\mathbf{pred}(k)} \sim NB\left(\exp\left(\mathbf{X}\boldsymbol{\beta}^{(k)}\right), \theta^{(k)}\right), \quad (4)
$$

for $k = 1, 2, ..., K$. Note that the negative-binomial model may be viewed as a marginal model derived from an hierarchical Poisson-gamma mixture model (*25*). This property is particularly useful within the Bayesian framework, since the gamma distribution assigned to the *mixing parameters* or *random effects* is conjugate to the Poisson, and therefore the posterior distribution of the random effects is also a known gamma distribution. That implies that prediction from the hierarchical Poisson-gamma structure is straightforward to implement with some additional simulation. Let $\mathbf{u} = (u_1, u_2, ..., u_n)^T$ denote the vector of random effects, where each element u_i corresponds to a random effect for OD flow y_i, for $i = 1, 2, ..., n$. The posterior distribution for each u_i is $Gamma\left(y_i + \theta, \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) + \theta\right)$, and therefore – in a vector notation – generating over k results to K vectors $\mathbf{u}$, that is $\mathbf{u}^{(k)} \mid \mathbf{y}, \boldsymbol{\beta}^{(k)}, \theta^{(k)} \sim Gamma\left(\mathbf{y} + \theta^{(k)}, \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}^{(k)}\right) + \theta^{(k)}\right)$ for $k = 1, 2, ..., K$. Prediction from the hierarchical Poisson-gamma model is then straightforward; for $k = 1, 2, ..., K$ we generate

$$\mathbf{y}^{\mathbf{pred}(k)} \sim Pois\left(\exp\left(\mathbf{X}\boldsymbol{\beta}^{(k)}\right)\mathbf{u}^{(k)}\right). \quad (5)$$

For cases of large scale and extremely overdispersed OD datasets, predictions from the hierarchical Poisson predictive distribution in expression (5) are preferred to predictions from the negative-binomial distribution in expression (4), since the random effect for each OD pair is explicitly taken into account.

In general, the predictive OD vectors/matrices may be used for model validation purposes through various posterior predictive checks and also for evaluation of predictive scenarios. However, this aspect of research is not pursued further in this study. More detailed information on model formulation and also predictive validation can be found in Perrakis et al. (*20*). In the following section, we demonstrate how the OD predictions can be used as input in an assignment model, as a mean of quantifying input-uncertainty.

3. OD PREDICTIONS AS INPUT IN TRAFFIC ASSIGNMENT

Some additional notation must be introduced for this section; let us denote by m the total number of links in the network and let $\mathbf{z} = (z_1, z_2, ..., z_m)^T$ be the corresponding vector of link flows, in addition let $A = (D, S)$ represent the type of assignment procedure, deterministic or stochastic, respectively. Finally, in order to simplify notation, the OD predictions are now denoted by $\mathbf{y}^{(k)}$, for $k = 1, 2, ..., K$.

3.1. Method 1 – Assigning OD Summaries

In the first method the predictive OD vectors $\mathbf{y}^{(k)}$, for $k = 1, 2, ..., K$, are used to calculate a summary statistic of the OD matrix, denoted by $S(\mathbf{y})$. In general, the OD summary vector is a function of the K predictions, that is $S(\mathbf{y}) = f(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(K)})$. By assigning $S(\mathbf{y})$ to a network, a corresponding estimate for link flows, denoted by $S(\mathbf{z})$, is obtained.

The most common point estimate is the mean, $S(\mathbf{y}) = \bar{\mathbf{y}}$, which can be calculated as follows

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{pmatrix} = \frac{1}{K}\left[\begin{pmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_n^{(1)} \end{pmatrix} + \begin{pmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_n^{(2)} \end{pmatrix} + \cdots + \begin{pmatrix} y_1^{(K)} \\ y_2^{(K)} \\ \vdots \\ y_n^{(K)} \end{pmatrix}\right] = \frac{1}{K}\begin{pmatrix} y_1^{(1)} + y_1^{(2)} + ... + y_1^{(K)} \\ y_2^{(1)} + y_2^{(2)} + ... + y_2^{(K)} \\ \vdots \\ y_n^{(1)} + y_n^{(2)} + ... + y_n^{(K)} \end{pmatrix}.$$

The $n \times 1$ mean vector $\bar{\mathbf{y}}$ can then be used as the OD-input in an assignment model and yield a $m \times 1$ vector $\bar{\mathbf{z}}$, which will correspond to the mean estimate of link flow vector $\mathbf{z}$

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{pmatrix} \xrightarrow{A} \bar{\mathbf{z}} = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \vdots \\ \bar{z}_m \end{pmatrix}.$$

Specifically, $\bar{\mathbf{z}}$ is an estimate of $E(\mathbf{z} \mid \bar{\mathbf{y}}, A)$, the expected vector of link flows conditional on the predictive expectations of OD flows but also conditional on the assignment model, and therefore $E(\mathbf{z} \mid \bar{\mathbf{y}}, D)$ under deterministic assignment will not be necessarily the same with $E(\mathbf{z} \mid \bar{\mathbf{y}}, S)$ under stochastic assignment.

For interval estimates the appropriate summary statistics are percentile vectors, i.e. $S(\mathbf{y}) = \mathbf{y}^p$ for the *p*-th percentile. Estimation of a percentile vector is not as straightforward as the calculation of the mean vector. Calculating individually the corresponding percentile of each OD pair would result in a percentile vector which will be highly unlikely to occur, especially for percentiles near 0 or 100 and for a large number of OD pairs, i.e. for a large *n*. Therefore, it is preferable to derive percentile vectors based on a function of vectors $\mathbf{y}^{(k)}$ which will operate as a criterion and constrain the estimates within realistic limits.

A natural selection for the criterion is the sum of each vector $\mathbf{y}^{(k)}$, i.e. $s^{(k)} = \sum_{i=1}^{n} y_i^{(k)}$, for $k = 1, 2, ..., K$. Under this approach, percentile vector $\mathbf{y}^p$ is derived by calculating first the corresponding percentile s^p from the values $\{s^{(1)}, s^{(2)}, ..., .s^{(K)}\}$. Then, the vector which has the closest sum to s^p is chosen as $\mathbf{y}^p$. In case there are two or more sums which satisfy that condition, then $\mathbf{y}^p$ is set as the average of the corresponding vectors. In general,

$$\mathbf{y}^{p} = \left(\overline{\overline{L_p}}\right)^{-1} \sum_{l=1}^{\overline{\overline{L_p}}} \mathbf{y}^{(l)},$$

where $L_p = \left\{ \mathbf{y}^{(l)} : \left| s^{(l)} - s^{p} \right| = \min_k \left\{ \left| s^{(k)} - s^{p} \right| \right\},\ k = 1,\dots,K \right\}$ and $\overline{\overline{L_p}}$ is the cardinality of set L_p. A common percentile pair is $\left(\mathbf{y}^{2.5}, \mathbf{y}^{97.5}\right)$ which corresponds to a 95% interval estimate. Note that $\mathbf{y}^{50}$, the vector corresponding to the median, can also be calculated through this procedure and be used as an additional point estimate to the mean.

For the special cases of the minimum and maximum vectors, we simply have to find $s^{\min}$ and $s^{\max}$, respectively, and then calculate vectors $\mathbf{y}^{\min}$ and $\mathbf{y}^{\max}$ as follows

$$\mathbf{y}^{\min} = \left(\overline{\overline{L_{\min}}}\right)^{-1} \sum_{l=1}^{\overline{\overline{L_{\min}}}} \mathbf{y}^{(l)},$$

$$\mathbf{y}^{\max} = \left(\overline{\overline{L_{\max}}}\right)^{-1} \sum_{l=1}^{\overline{\overline{L_{\max}}}} \mathbf{y}^{(l)},$$

where

$$L_{\min} = \left\{ \mathbf{y}^{(l)} : s^{(l)} = \min_k \left\{ s^{(k)} \right\},\ k = 1,\dots,K \right\},$$

$$L_{\max} = \left\{ \mathbf{y}^{(l)} : s^{(l)} = \max_k \left\{ s^{(k)} \right\},\ k = 1,\dots,K \right\},$$

with $\overline{\overline{L_{\min}}}$ and $\overline{\overline{L_{\max}}}$ being the corresponding cardinalities of sets $L_{\min}$ and $L_{\max}$.

After calculating a specific percentile vector $\mathbf{y}^{p}$ that is of interest, the corresponding estimate for link flows $\mathbf{z}^{p}$ is obtained by assigning $\mathbf{y}^{p}$ to the network. The results will once again depend on the choice of the assignment model.

In general, method 1 provides a suitable framework for obtaining point and interval estimates of link flows, in a fast and computationally easy way. For instance, the mean and the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles, which are often adequate for basic inferential purposes, can be obtained with three individual assignments. Subsequent inference can either focus on specific links of the network that might be of particular interest or expand on the entire network by visualizing the results of the assignment.

3.2. Method 2 – Assigning Multiple OD's

Method 2 involves assigning all K OD predictions individually in order to obtain K corresponding vectors of link flows. This method is computationally more intense than method 1, but also delivers full information for link flows in the form of distributional estimates.

In this case, K individual assignments must be implemented for each $\mathbf{y}^{(k)}$, with $k = 1, 2, \dots, K$. In vector notation, we have that

$$\mathbf{y}^{(1)} = \begin{pmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_n^{(1)} \end{pmatrix} \xrightarrow{A} \mathbf{z}^{(1)} = \begin{pmatrix} z_1^{(1)} \\ z_2^{(1)} \\ \vdots \\ z_m^{(1)} \end{pmatrix},$$

$$\mathbf{y}^{(2)} = \begin{pmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_n^{(2)} \end{pmatrix} \xrightarrow{A} \mathbf{z}^{(2)} = \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \\ \vdots \\ z_m^{(2)} \end{pmatrix},$$

$$\vdots \qquad\qquad \vdots$$

$$\mathbf{y}^{(K)} = \begin{pmatrix} y_1^{(K)} \\ y_2^{(K)} \\ \vdots \\ y_n^{(K)} \end{pmatrix} \xrightarrow{A} \mathbf{z}^{(K)} = \begin{pmatrix} z_1^{(K)} \\ z_2^{(K)} \\ \vdots \\ z_m^{(K)} \end{pmatrix}.$$

Overall summary statistics can be calculated, only this time directly from the vectors $\mathbf{z}^{(k)}$. For instance, the mean is calculated as follows

$$\overline{\mathbf{z}} = \begin{pmatrix} \overline{z}_1 \\ \overline{z}_2 \\ \vdots \\ \overline{z}_m \end{pmatrix} = \frac{1}{K}\left[\begin{pmatrix} z_1^{(1)} \\ z_2^{(1)} \\ \vdots \\ z_m^{(1)} \end{pmatrix} + \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \\ \vdots \\ z_m^{(2)} \end{pmatrix} + \cdots + \begin{pmatrix} z_1^{(K)} \\ z_2^{(K)} \\ \vdots \\ z_m^{(K)} \end{pmatrix}\right] = \frac{1}{K}\begin{pmatrix} z_1^{(1)} + z_1^{(2)} + \ldots + z_1^{(K)} \\ z_2^{(1)} + z_2^{(2)} + \ldots + z_2^{(K)} \\ \vdots \\ z_m^{(1)} + z_m^{(2)} + \ldots + z_m^{(K)} \end{pmatrix}.$$

Note that this an estimate of the expectation of vector $\mathbf{z}$ conditional on *all* OD predictions and on the assignment model, that is $E\left(\mathbf{z} \mid \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(K)}, A\right)$. Once again, the estimate depends on the assignment model and therefore the estimate $E\left(\mathbf{z} \mid \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(K)}, D\right)$ under deterministic assignment will not be the same with the estimate $E\left(\mathbf{z} \mid \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(K)}, S\right)$ under stochastic assignment.

With this method, estimates of percentile vectors are straightforward to calculate. Any percentile vector $\mathbf{z}^{p} = \left(z_1^{p}, z_1^{p}, \ldots, z_m^{p}\right)^T$ is calculated directly from the K vectors $\mathbf{z}^{(k)}$, i.e. z_j^{p}, for $j = 1, 2, \ldots, m$, is estimated individually as the *p*-th percentile obtained from the corresponding sample values $\left\{z_j^{(1)}, z_j^{(2)}, \ldots, z_j^{(K)}\right\}$. In general, the vectors $\mathbf{z}^{(k)}$, for $k = 1, 2, \ldots, K$, contain all necessary information for the links of the network; in case we wish to infer on a specific link z_j, then point, interval and dispersion estimates, such as the variance and the standard deviation, or even the distribution of z_j can be estimated directly from the sample values $\left\{z_j^{(1)}, z_j^{(2)}, ..., z_j^{(K)}\right\}$.

Despite the fact that this method is computationally demanding, it delivers full information over the links of a specific network. This information does not only involve measures related to location and dispersion but also distributional estimates which can be used to calculate specific probabilities related to link flows and the capacity of a given network, e.g. probabilities of exceeding threshold values of volume (link flow) over capacity ratio.

4. AN APPLICATION FOR FLANDERS UNDER DUE ASSIGNMENT

Flanders is the northern, Dutch-speaking region of Belgium with a population of 6,058,368 residents, accounting for approximately 60% of the country's population. The initial OD matrix is derived from the 2001 Belgian census study and contains information about work and school trips for a normal weekday and for all travel modes. The OD matrix consists of 308 zones, which correspond to the municipalities of Flanders, and has 94,864 OD pairs.

In Perrakis et al. (*20*) we fitted a Poisson and a negative-binomial model to the census OD matrix. Both models included the same set of 24 regression parameters and a Metropolis-Hastings algorithm was utilized in order to sample from the posterior distributions presented in expressions (1) and (2) of section 2. Due to the extreme overdispersion of the OD flows the negative-binomial model was found to provide a substantially better fit to the data. Predictions of OD flows were generated from the hierarchical Poisson predictive distribution presented in expression (5) of section 2.

In this section we present results from the 2 inputting methods by utilizing 200 predictive OD datasets under a DUE model. As link performance function, the commonly-used Bureau of Public Roads (*26*) formulation is adopted, with calibration parameters alpha and beta set equal to their historical values 0.15 and 4, respectively. The corresponding road network of Flanders has a total length of 65,296.72 kilometers and contains 97,450 links, 8.58% of which correspond to highways, including entrances and exits to highways, 15.49% to main regional roads, 21.1% to small regional roads, 52.91% to local municipal roads, while the remaining 1.98% of the links correspond to walk and bicycle paths. The road network of Flanders with the corresponding boundaries of the 5 Flemish provinces and the capitals – and major municipal centers – of each province, is presented in Figure 1.

The analysis that follows is focused on the morning peak hour between 7 am and 8 am for traffic flows concerning going-to-work/school trips. Links within the metropolitan area of Brussels are not included in the analysis; nevertheless the highway ring around Brussels area is included. In addition, inference is restricted necessarily on trips made by Flemish residents and therefore potential trips, within the Flemish region, made by residents of Brussels area or by residents of the southern Walloon region of Belgium are not taken under consideration.

Global visualization is a first illustrative step which provides a general idea for the overall state of the network. The mean link flow vectors resulting from both methods were used in order to visualize the expected state of the network. These graphs are not presented here due to the extremely large scale of the network and also due to the difficulty in marking small differences between the two methods on a global scale. Nevertheless, the key findings from global visualization are the following. The link flows and the corresponding volume over capacity (V/C) ratios are larger near the important municipal centers, especially near Antwerp, the capital and largest city of Flanders, as well as Ghent and Leuven. Flows are also dense on the three major highways connecting Antwerp with Ghent to the west (highway E17), with Brussels to the south (highway E19) and with Hasselt to the east (highway E313). In addition, dense traffic flows are observed on the northern part of the highway ring of Brussels metropolitan area (highway ring R0) and on the highways which connect Brussels with Ghent and Bruges (highway E40 west), and Brussels with Leuven (highway E40 east).

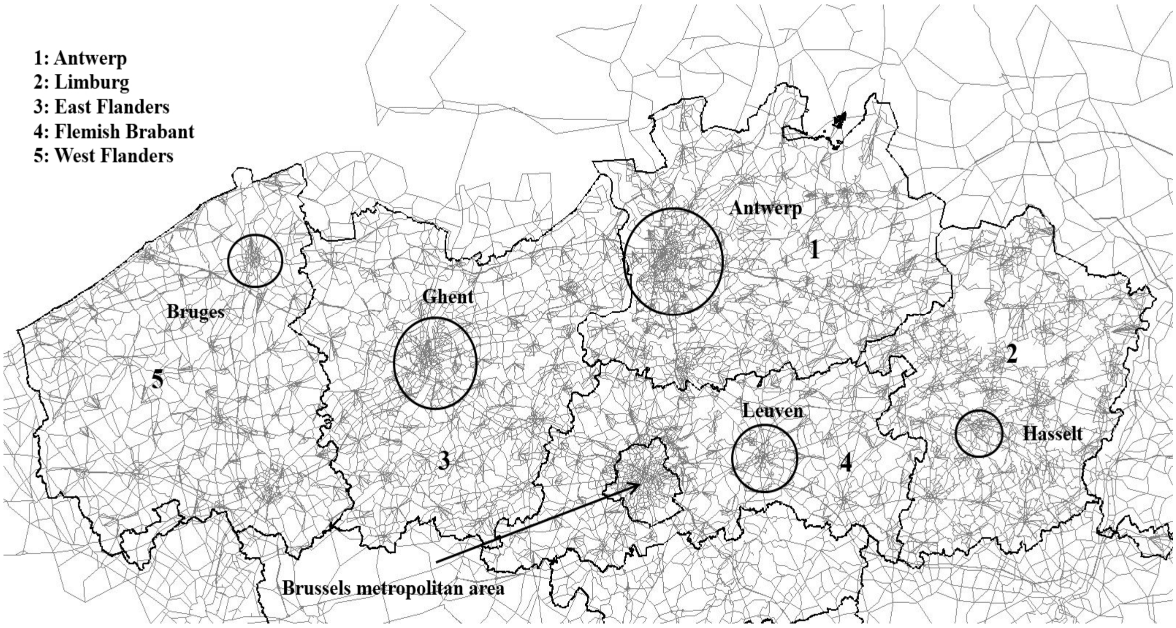

FIGURE 1 The road network of Flanders and the 5 Flemish provinces of Antwerp, Limburg, East Flanders, Flemish Brabant and West Flanders with corresponding capitals; Antwerp, Hasselt, Ghent, Leuven and Bruges.

In order to summarize the results, the link flows were aggregated according to the type of road they correspond, the link flow estimates are presented in Table 1. According to the rounded estimates of Table 1, centrality measures, i.e. the means and the medians, derived from the two methods are relatively close, since the differences can be measured in terms of hundreds relative to the large magnitudes of the estimates. This is not the case, for the 95% interval and range estimates, where the estimates of method 2 are much wider than the corresponding estimates of method 1. This is to be expected; as described in section 3.1, percentile estimates from method 1 are approximate since they are derived conditional on the sum of each OD vector, and therefore result to smaller intervals which underestimate the overall input-uncertainty. Method 2, additionally delivers standard deviation estimates which are also included in Table 1.

The link flow distributions can be estimated directly from the 200 link flow vectors obtained from method 2; kernel estimates for flows in highways, main regional roads, small regional roads and local roads are presented in Figure 2. Gaussian kernels were used as smoothing kernels, with bandwidths set equal to the corresponding standard deviations of the smoothing kernels.

TABLE 1 Mean, Median, 95% Interval (2.5%-97.5%), Range and Standard Deviation Estimates from the 2 Methods for the 4 Types of Link Traffic Flows

Method 1					
Links	Mean	Median	95% interval	Range	St.dev.
Highways	1,565,723	1,565,536	(1,564,267-1,566,497)	(1,561,068-1,570,341)	-
Main regional roads	1,292,689	1,293,363	(1,291,685-1,294,257)	(1,288,723-1,295,259)	-
Small regional roads	796,803	797,499	(796,357-800,656)	(793,479-797,115)	-
Local roads	450,905	450,263	(449,754-450,293)	(449,653-451,522)	-
Method 2					
Links	Mean	Median	95% interval	Range	St.dev.
Highways	1,565,694	1,565,480	(1,559,776-1,571,443)	(1,556,818-1,573,290)	3,177
Main regional roads	1,292,411	1,292,261	(1,288,687-1,297,411)	(1,287,121-1,298,399)	2,113
Small regional roads	797,400	797,102	(794,400-803,345)	(793,451-805,032)	1,987
Local roads	450,656	450,599	(449,397-452,041)	(448,485-452,546)	728

A useful application from method 2 is the identification of *critical links*. Critical link identification is commonly a subject of *vulnerability analysis*, where state-of-the-art approaches are based on full network scan algorithms ((*27*), (*28*)). Nevertheless, the associated computational burden constitutes the implementation of such approaches prohibitive, as yet, for applications on large-scale, congested networks where execution of traffic assignment within the full scan algorithm is additionally required. From this point of view, method 2 provides a workable alternative for identifying critical links explicitly in terms of probability.

As critical links we define those links in which the V/C ratio exceeds a specific threshold value c with a certain probability, i.e. $P(\mathrm{V/C} > c)$. Evaluation of critical links in terms of probability estimates is safer in comparison to point estimates and also reduces the margin of uncertainty. For instance, the expected value of V/C may be smaller than c, nevertheless the probability of exceeding c may be significantly greater than zero. This point is illustrated further in the proceeding analysis. In general, the value $c=1$ is the common option, since values greater than 1 imply that the traffic flow exceeds the theoretical capacity of a link, thus resulting in congestion. In our case though, inference is limited in traffic flows for work and school trips made only by Flemish residents, which means that we would expect the V/C ratios to be higher if the proportion of traffic related to other trip-purposes and to non-Flemish residents was included in the analysis. From this perspective threshold values smaller than 1 may also be regarded as critical, but since the exact proportion of trips which is unaccounted for is not known exactly, c can only be selected on a heuristic basis. As a conservative choice and in order not to overestimate the number of critical links the value of 0.95 is adopted. Eleven links are identified as critical for $c = 0.95$, the V/C distributions of these links with the corresponding expected values of V/C ratio and the probabilities of exceeding 0.95 are presented in Figure 3. Note that if the analysis was based on the expected

V/C ratio instead of the probability of exceeding a V/C ratio of 0.95, then four out of the eleven links would not have been identified as critical.

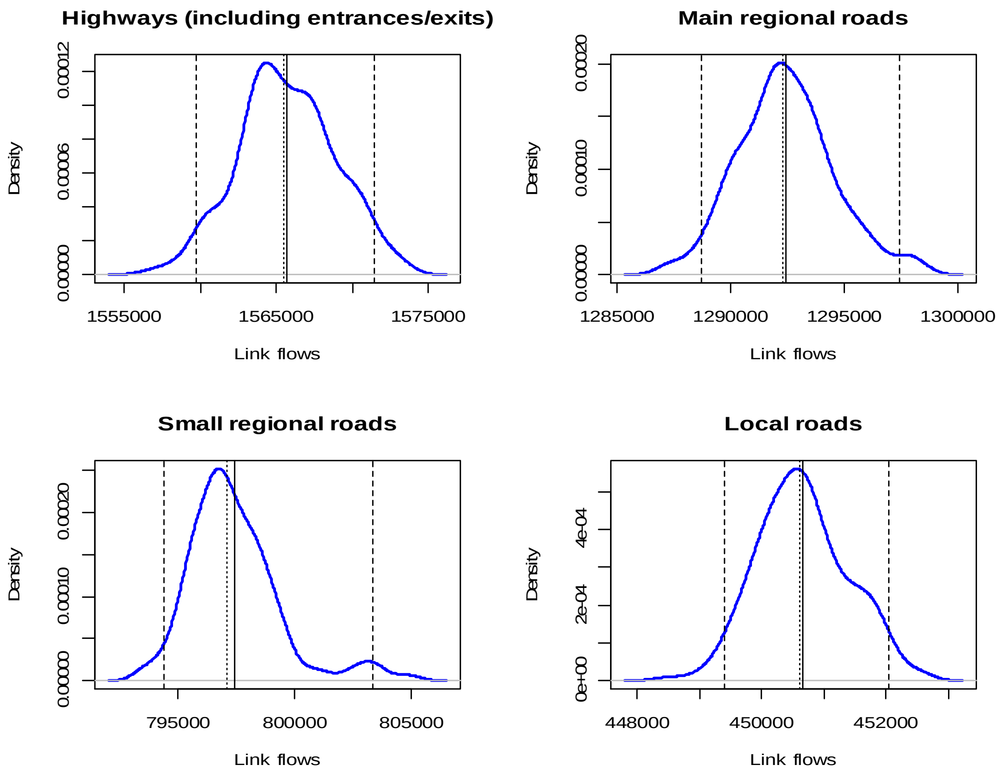

FIGURE 2 Kernel distributions from method 2 for traffic flows on highways, main regional roads, small regional roads and local roads. The solid, dotted and dashed vertical lines correspond to the mean, median and 95% interval of each distribution, respectively.

It is interesting to note that the V/C distributions of Figure 3 and consequently the corresponding link flow distributions resulting from DUE assignment are not necessarily close to normal distributions – for instance, bimodality is observed – in contrast to the distributions presented in Figure 2 which are aggregated link flows and in accordance to the central limit theorem closer to normal distributions.

Bimodality may be attributed to the iterative user equilibrium procedure itself. For instance, when the flows on a specific link and at a given iteration exceed a certain threshold – consequently leading to a high V/C ratio – and there exists an alternative link which has a cost which is close but lower, then in the following iteration a switch of flows will occur from the high-cost link to the low-cost link. This “switching” effect will eventually result to bimodal V/C distributions, as the ones presented in Figure 3.

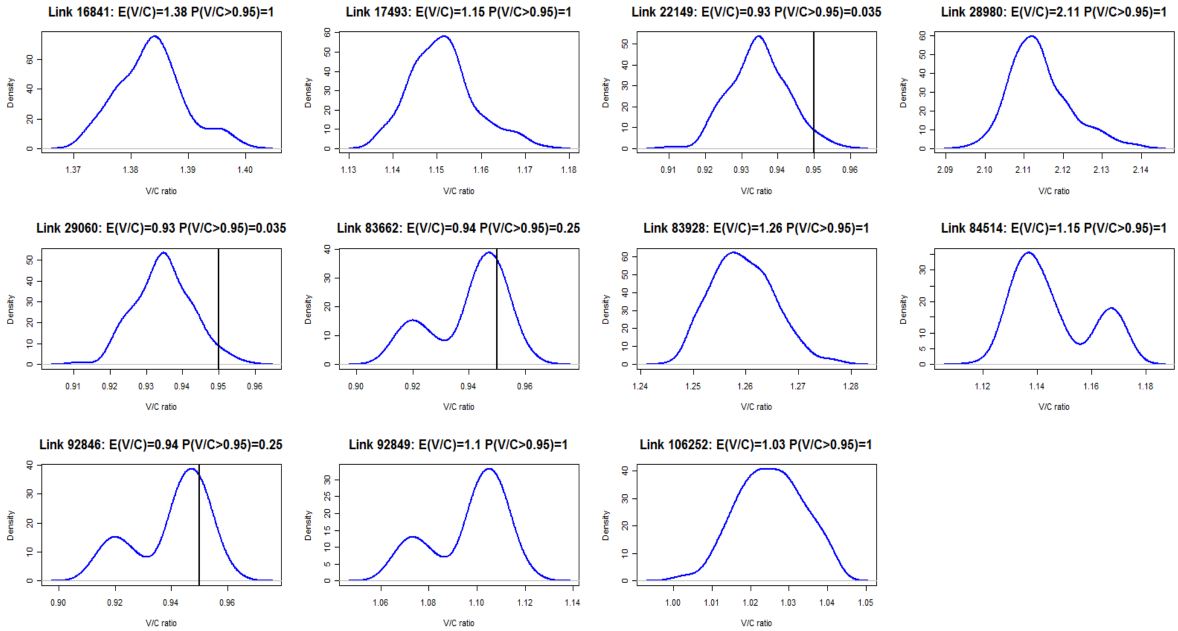

FIGURE 3 V/C kernel distribution estimates from method 2 for the 11 critical links which include or exceed the value of 0.95, highlighted by a vertical line in the distributions which include this value.

The results in Figure 3 show that seven out of the eleven critical links have a V/C value greater than 0.95 with probability 1. Visual examination of the distributions in Figure 3 additionally reveals that these seven links also exceed the value $c = 1$ with probability 1, except perhaps of link 106252 which has its minimum located near 1 and may therefore include smaller values than 1 with a low probability. Conclusively, even without taking into consideration the proportion of traffic related to non-Flemish residents and to other trip-purposes than work or school trips, congestion on these 7 links is almost certain. The highest V/C ratio is observed in highway link 28980 with an expectation of 2.11, while regional road link 16841 and highway link 83928 follow with expected V/C ratios equal to 1.38 and 1.26, respectively. For the remaining 4 links, the probability of exceeding a V/C ratio of 0.95 is lower, equal to 25% for links 83662, 92846 and equal to 3.5% for links 22149 and 29060. Not surprisingly, the 11 critical links are situated near the major municipal centers of Antwerp, Ghent and Bruges; the exact locations are presented in Figure 4.

As illustrated in Figure 4, five of the critical links belong to the wider municipal area of Antwerp, including link 28980 which is a segment of R1 highway ring in the north of Antwerp and has the highest expected V/C ratio. Links 22149 and 29060 are local outgoing road segments near highway ring R1, whereas link 16841, which has the second highest V/C ratio, is a segment of N1 regional road directing right to the center of Antwerp. Finally, link 17493 is a local road segment outside Antwerp, nevertheless very close to Antwerp airport situated south-east of the city.

Five critical links also appear in the municipal area of Ghent. Link 84514 to the north-east is a segment of N70 regional road very near the exit of the R4 highway ring with a direction to the center of Ghent, whereas links 83928, 83662, 92486 and 92849, near the center, are all road segments clustered around the end of the part of highway E17 that has a direction to the center of Ghent.

Finally, one critical link appears near the municipality of Bruges, that is link 106252. This link corresponds to a local road segment, very near N31 regional road, and with a direction towards the center of Bruges from the west.

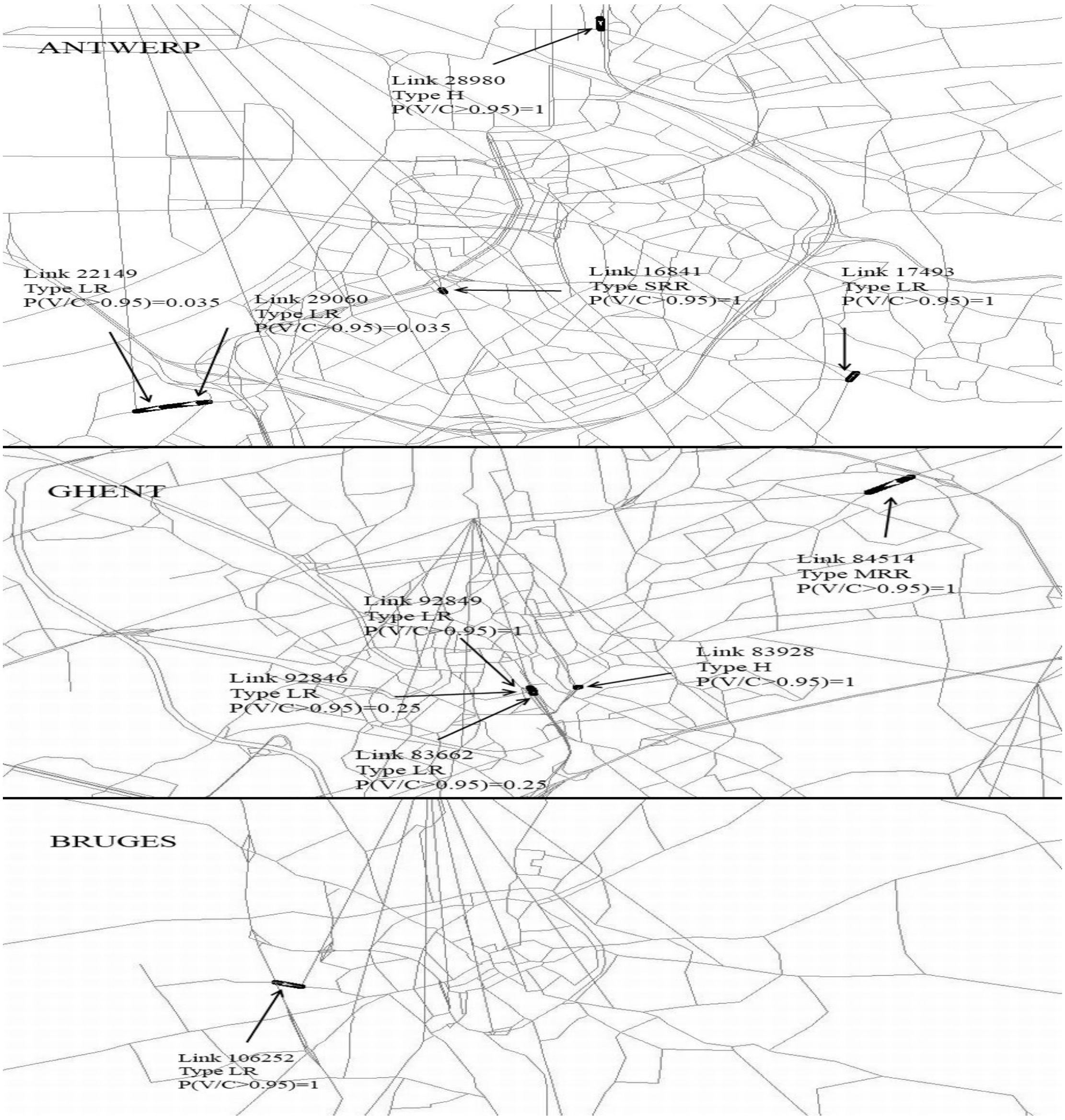

FIGURE 4 The critical links in Antwerp, Ghent and Bruges, the link type abbreviations H, MRR, SRR and LR correspond to Highways, Main Regional Roads, Small Regional Roads and Local Roads.

The analysis of critical links is based on the relatively conservative choice $c = 0.95$. Naturally, for smaller values of c the number of critical links increases, e.g. for the values 0.9, 0.85, 0.8, 0.75, 0.7 the number of critical links rises to 16, 20, 25, 47 and 69, respectively. In general, for situations where the exact proportion of traffic is not known, as in the application presented in this study, the choice of c is under the control of the researcher or policy-planner. In such cases, inference may be based on more than one values of c. For cases in which there is certainty that all the potential traffic or – at least – most of the potential traffic of a network is included in the analysis, then the value $c=1$ may be safely adopted.

5. CONCLUSIONS

This study focuses on utilizing Bayesian OD predictions as input in traffic assignment models, as a mean of incorporating and quantifying input-uncertainty. Two methods of inputting OD predictions are proposed. In the first method, an OD summary is calculated from the multiple OD predictions and is assigned to the corresponding network. Whereas, in the second method all OD predictions are assigned to the network individually.

In terms of computational expense, the first method is less demanding and suitable for delivering basic inferential tools, such as point and approximate interval estimates of link flows, with a limited number of assignments. On the other hand, method 2 results to more reliable interval estimates and also provides dispersion and distributional estimates which are not computable through method 1.

In general, method 2 provides greater inferential capabilities and gives rise to novel applications as is the identification of critical links explicitly in terms of probability estimates.

The proposed methods can be applied on short-term as well as long-term OD matrices and furthermore their use in not constrained by the selection of the assignment model. Therefore, the methods may be utilized for different travel-demand perspectives under deterministic and stochastic assignment models.

An application for the road network of Flanders was presented for the morning peak hour between 7 am and 8 am, under the assumption of a DUE model and by utilizing 200 predictive OD matrices. Traffic flows in Flanders were found to be denser around the major municipal centers of Antwerp, Ghent, Leuven and Bruges and on the highways which connect these cities with each other and also with Brussels. Eleven critical links were identified for a threshold value of 0.95, the majority of which belonging to Antwerp and Ghent.

Future research will focus further on implementing both methods under system optimum and also deterministic/stochastic user equilibrium assignment assumptions for various travel-demand scenarios. That would allow for a comparison between theoretical expectations and the actual performance of those models, under different experimental settings.

REFERENCES

(1) Cascetta, E. *Transportation Systems Analysis: Models and Applications, Second Edition*. Springer, New York, 2009.
(2) Thomas, R. *Traffic Assignment Techniques*. Avebury Technical, Aldershot, 1991.
(3) Patriksson, M. *Traffic Assignment Problems: Models and Methods. Topics in transportation*. V.N.U. Science Press, Utrecht, 1994.
(4) Florian, M. and D. Hearn. Network equilibrium models and algorithms. In *Handbooks in OR and MS*, Vol. 8, 1995, pp. 485-550.
(5) Cantarella, G. E. and E. Cascetta. Stochastic assignment to transportation networks: models and algorithms. In *Equilibrium and advanced transportation modelling*, Vol.5, 1998, pp. 87–107.
(6) de Jong, G., A. Daly, M. Pieters, S. Miller, R. Plasmeijer, and F. Hofman. Uncertainty in traffic forecasts: literature review and new results for The Netherlands. *Transportation*, Vol. 34, No. 4, 2006, pp. 375-395.
(7) Ettema, D. and H. Timmermans. Costs of travel time uncertainty and benefits of travel time information: Conceptual model and numerical examples. *Transportation Research Part C: Emerging Technologies*, Vol. 14, No. 5, 2006, pp. 335-350.
(8) Li, R. and G. Rose. Incorporating uncertainty into short-term travel time predictions. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 6, 2011, pp. 1006-1018.
(9) Kuhn, K. D. and S. M. Madanat. Model uncertainty and the management of a system of infrastructure facilities. *Transportation Research Part C: Emerging Technologies*, Vol. 13, No. 5-6, 2005, pp. 391-404.
(10) Ng, M., Z. Zhang, and S. T. Waller. The price of uncertainty in pavement infrastructure management planning: An integer programming approach. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 6, 2011, pp. 1326-1338.
(11) Matas, A., J. L. Raymond, and A. K. Ruiz. Traffic forecasts under uncertainty and capacity constraints. *Transportation*, in press, doi: 10.1007/s11116 -011-9325-1.
(12) Hazelton, M.L. Some remarks on stochastic user equilibrium. *Transportation Research Part B: Methodological*, Vol. 32, No. 2, 1998, pp. 101-108.
(13) Waller, S.T., J. Schofer, and A. Ziliaskopoulos. Evaluation with traffic assignment under demand uncertainty. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1771, No. 1, 2001, pp. 69-74.
(14) Ukkusuri, S. V. and S. T. Waller. Single-point approximations for traffic equilibrium problem under uncertain demand. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1964, No. 1, 2006, pp. 169-175.
(15) Gardner, L. M., A. Unnikrishnan, and S. Waller. Robust pricing of transportation networks under uncertain demand. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2085, No. 1, 2008, pp. 21-30.
(16) Gardner, L. M., A. Unnikrishnan, and S. T. Waller. Solution methods for robust pricing of transportation networks under uncertain demand. *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 5, 2010, pp. 656-667.
(17) Duthie, J. C., A. Unnikrishnan, and S. T. Waller. Influence of demand uncertainty and correlations on traffic predictions and decisions. *Computer Aided Civil and Infrastructure Engineering*, Vol. 26, No. 1, 2011, pp. 16-29.

(18) Ukkusuri, S. V., T. V. Mathew, and S. T. Waller. Robust transportation network design under demand uncertainty. *Computer Aided Civil and Infrastructure Engineering,* Vol. 22, No. 1, 2007, pp. 6-18.
(19) Ukkusuri S. V. and G. Patil. Multi-period transportation network design under demand uncertainty. *Transportation Research Part B: Methodological*, Vol. 43, No. 6, 2009, pp. 625-642.
(20) Perrakis, K., D. Karlis, M. Cools, D. Janssens, K. Vanhoof, and G. Wets. A Bayesian approach for modeling origin-destination matrices. *Transportation Research Part A: Policy and Practice,* Vol. 46, No. 1, 2012, pp. 200-212.
(21) Fernández, C., E. Ley, and M. F. J. Steel. Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, Vol. 100, No. 2, 2001, pp. 381-427.
(22) Ntzoufras, I. *Bayesian Modeling Using WinBUGS.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2009.
(23) Gamerman, D. and H.F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition.* Chapman and Hall/CRC, London, 2006.
(24) Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, Vol. 57, No. 1, 1970, pp. 97-109.
(25) Agresti, A. *Categorical Data Analysis, Second edition.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2002.
(26) Bureau of Public Roads. *Traffic assignment manual for application with a large, high speed computer*. U.S. Department of Commerce, 1964.
(27) Jenelius, E., T. Petersen, and L.G. Mattsson. Importance and exposure in road network vulnerability analysis. *Transportation Research Part A: Policy and Practice*, Vol. 40, No. 7, 2006, pp. 537-560.
(28) Taylor, M. A. P. Critical transport infrastructure in urban areas: impacts of traffic incidents assessed using accessibility-based network vulnerability analysis. *Growth and Change,* Vol. 39, No. 4, 2008, pp. 593-616.