A Bayesian approach for modeling origin-destination matrices

Peer-reviewed author version

PERRAKIS, Konstantinos; KARLIS, Dimitris; COOLS, Mario; JANSSENS, Davy; VANHOOF, Koen & WETS, Geert (2012) A Bayesian approach for modeling origin-destination matrices. In: TRANSPORTATION RESEARCH PART A-POLICY AND PRACTICE, 46 (1), p. 200-212.

DOI: 10.1016/j.tra.2011.06.005
Handle: http://hdl.handle.net/1942/13279

A BAYESIAN APPROACH FOR MODELING ORIGIN-DESTINATION MATRICES

Konstantinos Perrakis, Dimitris Karlis, Mario Cools, Davy Janssens and Geert Wets*

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 9140
Email: konstantinos.perrakis@uhasselt.be

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

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{31, 28, 58}
Email: {mario.cools, davy.janssens, geert.wets}@uhasselt.be

Number of words: 6470
Number of Tables: 1
Number of Figures: 3
Total number of words: 6470 + 4*250 = 7470

Revised paper submitted: November 12, 2010

* Corresponding author

ABSTRACT

The majority of Origin Destination (OD) matrix estimation methods focus on situations where weak or partial information, derived from sample *travel surveys*, is available. Information derived from travel *census studies*, in contrast, covers the entire population of a specific study area of interest. In such cases where reliable historical data exist, statistical methodology may serve as a flexible alternative to traditional travel demand models by incorporating estimation of trip-generation, trip-attraction and trip-distribution in one model. In this research, a statistical Bayesian approach on OD matrix estimation is presented, where modeling of OD flows, derived from census data, is related only to a set of general explanatory variables. The assumptions of a Poisson model and of a Negative-Binomial model are investigated on a realistic application area concerning the region of Flanders on the level of municipalities. Problems related to the absence of closed-form expressions are bypassed with the use of a Markov Chain Monte Carlo algorithm, known as the *Metropolis-Hastings* algorithm. Additionally, a strategy is proposed in order to obtain predictions from the hierarchical, Poisson-Gamma structure of the Negative-Binomial model conditional on the posterior expectations of the mixing parameters. In general, Bayesian methodology reduces the overall uncertainty of the estimates by delivering *posterior distributions* for the parameters of scientific interest as well as *predictive distributions* for future OD flows. Predictive goodness-of-fit tests suggest a good fit to the data and overall results indicate that the approach is applicable on large networks, with relatively low computational and explanatory data-gathering costs.

1. INTRODUCTION

The OD matrix estimation problem is a well known problem in transportation analysis and a crucial part of transportation planning. The existence of different schools of thought has resulted to a diverse range of approaches dealing with the matter and therefore, OD estimation methods vary significantly with respect to the modeling assumptions adopted and the methodological tools utilized. Nevertheless, the selection of a specific OD estimation method does not only depend on the methodological or philosophical framework or the overall scope of research but also relies significantly on the amount and type of information which is available.

Information for OD flows usually originates from travel surveys but is rarely used for inferential purposes directly. As illustrated in Cools et al. (*1*), sample estimates of OD matrices derived from travel surveys are biased even for large sampling rates and therefore insufficient in delivering reliable estimates. Travel demand models, such as the four-step model (*2*), take into account trip productions and trip attractions derived from travel surveys and deliver more reliable OD estimates through gravity or entropy-maximization models during the trip distribution step. Activity-based models, which form another trend in transportation modeling (*3*), also use information from travel surveys in the model training phase. Finally, methods which rely on observed link traffic counts, use OD matrices derived from travel surveys as "prior" information in order to impose constraints and cope with the under-specification problem between link flows and OD pairs (*4*). The last category of methods constitutes the main body of existing research in OD matrix estimation and the relative literature is extensive. A recent classification and discussion is provided by Timms (*5*). Notable contributions within the Bayesian framework include the studies of Maher (*6*), Tebaldi and West (*7*), Li (*8*), Castillo et al. (*9*) and Hazelton (*10*).

In contrast to research focused in OD matrix estimation from link traffic counts and/or sample OD estimates, little or no research has been conducted for cases in which historical OD data from census studies exist. OD matrices derived from census studies refer to the population of a specific study area and therefore statistical methodology may be safely utilized without necessarily linking the estimation problem to traffic counts. In addition, in such cases statistical methodology may serve as an effective alternative to the widely used travel demand models by integrating the steps of trip generation, trip attraction and trip distribution into statistical models which deliver reliable parameter estimates and accurate predictions.

In this current study, a statistical approach is presented where modeling is focused directly on OD pairs derived from census data. The approach challenges some of the practical and also methodological issues involved in OD matrix estimation, issues mainly related to costs, extent of applicability and evaluation of uncertainty. Regarding cost-efficiency, the approach is in general not cost demanding since OD flows are explained only by means of general and easily obtainable explanatory variables. The extent of applicability is tested on a realistic study area, concerning the municipality network of the Northern, Dutch-speaking part of Belgium, namely the region of Flanders which consists of 308 zones. Finally, the main aim of the approach is to reduce the overall uncertainty of estimation. To this extend, two models are investigated, a Poisson model and a Negative Binomial model. In addition, the estimation is purely Bayesian and the Metropolis-Hastings algorithm, a Markov Chain Monte Carlo algorithm, is used in order to acquire samples from the joint *posterior* *distribution* of all parameters. Moreover, a strategy is suggested in order to obtain accurate *predictions* of OD flows from the corresponding hierarchical Poisson-Gamma structure of the Negative Binomial model.

As illustrated in the study, the proposed approach is applicable for networks of large dimensionality, while at the same time data-gathering and computational costs are low. In addition, Bayesian methodology reduces uncertainty over the randomness of OD flows in two key aspects; first information is provided for the entire posterior distributions of the parameters that influence OD flows and second prediction of future OD flows is similarly based on *predictive distributions* instead of predictive point estimates. The former is useful in obtaining a wider perspective over the factors that may help explain the generation and attraction of OD trips. The latter, in combination with the inherent hierarchical nature of OD matrices, facilitates transportation policy-making by providing *predictive scenarios* for traffic volumes over multiple levels of aggregation and for different types of trips. Evaluation of such scenarios by policy-makers reduces the uncertainty involved in decisions related to transport infrastructure.

2. DATA

2.1 OD Matrix

The OD matrix is derived from the 2001 Belgian census, which contains information about the departure/arrival times and locations of work and school trips for the 10,296,350 Belgian residents. The work and school trips are one-directional, from zone of origin to zone of destination. Thus, the OD matrix contains the number of daily going-to-work/school related trips for a normal weekday and for all travel modes. The area of concern in this study is not the entire country of Belgium but the region of Flanders with a population of 6,058,368 residents. Information is provided on a highly analytic level, that is, the municipality network of Flanders which consists of 308 zones. The resulting OD matrix contains 94,864 cells.

An important feature of OD matrices is their inherent hierarchical structure. An OD matrix may be aggregated on different levels according to different geographical and/or municipal classifications. For the region of Flanders, there are several levels of aggregation that may be of interest; from the analytic level of municipalities to the more general levels of cantons, districts, arrondissements and finally provinces. The hierarchical structure of Flanders is represented below; on the higher level of municipalities the OD matrix has 308 zones and 94864 OD pairs whereas on the lower level of provinces there are only 5 zones and 25 possible OD pairs, in between we find the levels of cantons, districts and arrondissements. The downward direction of the arrows implies that each lower level is the result of an aggregation on the immediately higher level. Therefore, having an OD estimate on a high level of analysis is immediately advantageous, since it leads to direct OD estimates for all the lower levels, whereas the opposite is not true.

Another characteristic of OD matrices is that the flows are usually inflated across the main diagonal. The cells in the main diagonal correspond to “internal” trips; these are the trips that are made within the same zone where there is no distinction between origin and destination.

Municipalities 308×308 (94,864 cells)
↓
Cantons 103×103 (10,609 cells)
↓
Districts 52×52 (2,704 cells)
↓
Arrondissements 22×22 (484 cells)
↓
Provinces 5×5 (25 cells)

As expected, given the size of the matrix on municipality-level, the OD flows are sparsely distributed. Approximately, 63% of the cells in the matrix are zero-valued. In addition the data are clearly over-dispersed, since the mean of the OD flows equals 36.23 while the standard deviation is much larger, equal to 949.47. Finally, the cells across the main diagonal correspond to approximately 43% of the total OD flows of the matrix and the maximum value which is equal to 222,149 is observed in the diagonal cell belonging to Antwerp, the capital and largest municipality of Flanders.

2.2 Explanatory Variables

The selection of the explanatory variables is a combination of variables that can be derived immediately from the hierarchical structure of the OD matrix and of continuous explanatory variables. The second category consists of variables such as employment ratios, population densities, relative length of road networks, perimeter lengths of municipalities and yearly traffic in highways and provincial/municipal roads. The set of explanatory variables is listed below.

[1] **dum.prov**: dummy variable for internal-province trips
[2] **dum.arron**: dummy variable for internal-arrondissement trips
[3] **dum.dist**: dummy variable for internal-district trips
[4] **dum.cant**: dummy variable for internal-canton trips
[5] **dum.munic**: dummy variable for internal-municipality trips
[6] **munic.cant**: number of municipalities between the cantons of origin and destination
[7] **munic.dist**: number of municipalities between the districts of origin and destination
[8] **munic.arron**: number of municipalities between the arrondissements of origin and destination
[9] **munic.prov**: number of municipalities between the provinces of origin and destination
[10] **empl.o**: employment ratio of origin-zone
[11] **empl.d**: employment ratio of destination-zone
[12] **pop.dens.o**: population density of origin-zone (thousand inhabitants per square km)
[13] **pop.dens.d**: population density of destination-zone (thousand inhabitants per square km)
[14] **road.length.o**: length of road network relative to surface of origin-zone (km per square km)
[15] **road.length.d**: length of road network relative to surface of destination-zone (km per square km)
[16] **perim.o**: perimeter of origin-zone (in km's)
[17] **perim.d**: perimeter of destination-zone (in km's)
[18] **HWT.o**: km's driven per year in highway roads of origin-zone (in millions)
[19] **HWT.d**: km's driven per year in highway roads of destination-zone (in millions)
[20] **PMT.o**: km's driven per year in provincial and municipal roads of origin-zone (in millions)
[21] **PMT.d**: km's driven per year in provincial and municipal roads of destination-zone (in millions)

The variables which are extracted directly from the hierarchical structure of the OD matrix are [1]-[9]. In particular, variables [1]-[5] are dummy variables indicating whether a trip is internal or not for each level of aggregation, respectively. These variables are multiplied by 100 so that they correspond to a difference of one hundred trips. Variables [6]-[9] correspond to the total number of municipalities belonging to the specific cantons, districts, arrondissements and provinces of each OD pair. The rest [10]-[21], are the external explanatory variables, which come in pairs, since they relate to origin as well as destination. Finally, variables [6]-[21] are transformed in logarithmic scale, so that the multiplicative interpretation of the models presented next remains on natural scale.

The set of the explanatory variables is in general simple, costless and also easy to obtain. As mentioned, part of the explanatory variables is immediately derived by the structure of the OD matrix. Variables related to populations, surfaces and perimeters are usually available in transportation research centers and institutes. Finally, variables related to length of road networks were obtained by the Belgian governmental website for statistics (*11*).

3. MODELS

In this section, a brief description of the Poisson and Poisson-Gamma likelihood assumptions is presented along with the selection of the corresponding prior distributions. Expressions for the posterior distributions are then derived from the application of Bayes' theorem. For computational and notational convenience the OD flows are represented as a vector. Let n denote the data size and p the number of explanatory variables. In addition, let $\mathbf{y} = (y_1, y_2, ..., y_n)^T$ denote the vector of OD flows, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^T$ the vector of unknown parameters and $\mathbf{X}$ the design matrix of dimensionality $n \times (p+1)$, containing the intercept and the p explanatory variables, with $\mathbf{x}_i = (x_{i0}, x_{i1}, x_{i2}, ..., x_{ip})^T$ being the i-th row of $\mathbf{X}$ related to OD flow y_i and $i = 1, 2, ...n$.

3.1 The Poisson Model

The likelihood assumption is that the OD flows are independently Poisson distributed, that is $y_i \mid \boldsymbol{\beta} \sim Pois(\mu_i)$ for $i = 1, 2, ...n$, where μ_i is the Poisson mean for y_i, related to the explanatory variables through the log-link function $\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$. The log-link function implies the assumption that the effects of the explanatory variables are linear to the log-mean of y_i. Consequently, the effects are exponential on natural scale, since $\mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)$. The complete likelihood is given by

$$p(\mathbf{y} \mid \boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\exp\left[-\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right] \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)^{y_i}}{y_i!}. \tag{1}$$

Poisson regression is a common option when modeling count data and it is frequently used in practice. Nevertheless, Poisson models usually do not perform well in cases of over-dispersed data, since a strong restriction of Poisson modeling is that the mean is equal to the variance of the data, that is $E\left(y_i \mid \boldsymbol{\beta}\right) = Var\left(y_i \mid \boldsymbol{\beta}\right) = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)$. Properties and estimation procedures for Poisson regression can be found in Agresti (*12*) and McCullagh and Nelder (*13*), Bayesian applications are presented in Ntzoufras (*14*).

A flat non-informative prior with mean located at 0 and close-to-infinite variance is assigned for parameter vector $\boldsymbol{\beta}$. Specifically, the multivariate normal prior $\boldsymbol{\beta} \sim \mathbf{N_{p+1}}(\mathbf{0}, \boldsymbol{\Sigma_\beta})$, with $\boldsymbol{\Sigma_\beta} = n \times \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \times 10^3$, which is one of the "benchmark" priors suggested in Fernández et al. (*15*). This prior distribution has the form

$$p(\boldsymbol{\beta}) = \frac{1}{\left(2\pi\right)^{(p+1)/2} \left|\boldsymbol{\Sigma_\beta}\right|^{1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\Sigma_\beta}^{-1} \boldsymbol{\beta}\right). \tag{2}$$

By applying the Bayes' theorem, the posterior distribution of $\boldsymbol{\beta} \mid \mathbf{y}$ is proportional to $p(\boldsymbol{\beta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}) p(\boldsymbol{\beta})$. From expressions (1) and (2) the resulting posterior distribution is

$$p(\boldsymbol{\beta} \mid \mathbf{y}) \propto \prod_{i=1}^{n} \left[\exp\left[-\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right] \left[\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right]^{y_i} \right] \times \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\Sigma_\beta}^{-1} \boldsymbol{\beta}\right). \tag{3}$$

Sampling directly from the posterior distribution is not feasible, since expression (3) does not result in a known distributional form.

3.2 The Poisson-Gamma Model

The Poisson-Gamma model is a *mixed Poisson regression* model, where the mixing density is assumed to be a Gamma distribution. Mixed Poisson models incorporate over-dispersion and are frequently used as alternatives to the simple Poisson model (*16*). The likelihood assumption is $y_i \mid \boldsymbol{\beta}, u_i \sim Pois(\mu_i u_i)$, for $i = 1, 2, \ldots n$, where μ_i is again the part of the Poisson mean related to the explanatory variables through the log-link function $\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ and $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$ is a vector of *random deviations* or *random intercepts* distributed as $u_i \mid \theta \sim Gamma(\theta, \theta)$ with $\theta > 0$, so that $E\left(u_i\right) = 1$. The Poisson likelihood is the *conditional likelihood* of **y** given the vector **u**; the complete conditional likelihood is given by

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}) = \prod_{i=1}^{n} \frac{\exp\left[-\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) u_i\right] \left[\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) u_i\right]^{y_i}}{y_i!}. \tag{4}$$

From a Bayesian perspective the Poisson-Gamma model is an *hierarchical* model, since the mixing distribution is regarded as a 1st level prior distribution for **u** and parameter θ is then assigned a 2nd level prior distribution (*14*).

Alternatively, one may work with the *marginal* form of the model by integrating over the mixing density; the integration $p\left(\mathbf{y} \mid \boldsymbol{\beta}, \theta\right) = \int p\left(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{u}\right) p\left(\mathbf{u} \mid \theta\right) d\mathbf{u}$ results to a Negative-

Binomial marginal likelihood, that is $y_i \mid \boldsymbol{\beta}, \theta \sim NB(\mu_i, \theta)$, with $\mu_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)$ for $i = 1, 2, \ldots n$. The complete marginal likelihood then, is

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \theta) = \prod_{i=1}^{n} \frac{\Gamma\left(y_i + \theta\right)}{\Gamma\left(\theta\right) y_i !} \frac{\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)^{y_i} \theta^{\theta}}{\left[\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) + \theta\right]^{y_i + \theta}}. \tag{5}$$

The mean of the data in this case is $E\left(y_i \mid \boldsymbol{\beta}, \theta\right) = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)$, while the variance is $Var\left(y_i \mid \boldsymbol{\beta}, \theta\right) = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right) + \left[\exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right]^2 \theta^{-1}$. Note that the variance now is a quadratic function of the mean. Thus, Negative-Binomial regression incorporates over-dispersion, since the assumed variance always exceeds the assumed mean. Information for the Negative-Binomial model can be found in Agresti (*12*) and McCullagh and Nelder (*13*). A general Expectation-Maximization (EM) algorithm for obtaining Maximum Likelihood (ML) estimates for mixed Poisson models, with emphasis on the Poisson-Gamma case, is provided by Karlis (*16*). Within the Bayesian framework, Ntzoufras (*14*) presents descriptions and applications for both the hierarchical and the marginal formulations of the model.

By means of Bayesian methodology, one might choose to fit either the hierarchical or the marginal form of the model. In both cases, the estimates for the parameters of main scientific interest, $\boldsymbol{\beta}$ and θ, will be the same due to the equivalence of the two models. The hierarchical Poisson-Gamma model provides additional information over the posterior distribution of **u** but it also requires estimation of the full set of parameters $\boldsymbol{\beta}, \mathbf{u}$ and θ. The marginal Negative-Binomial model on the other hand is simpler to fit, since estimation is restricted to the reduced set of parameters $\boldsymbol{\beta}$ and θ. Due to the large size of the OD matrix, fitting the hierarchical model in our case would prove to be a very difficult task which would require estimating all of the u_i's that correspond to the 94864 random intercepts. Instead, we choose to work with the simpler Negative-Binomial distribution. As we will see in section 5.2, information over the vector **u** is not completely lost and prediction from the hierarchical structure is still feasible conditional on the posterior expectation of **u**.

Independent and non-informative priors are adopted for parameters $\boldsymbol{\beta}$ and θ. For parameter vector $\boldsymbol{\beta}$, the same multivariate normal distribution defined in expression (2) is used. Regarding parameter θ, a $Gamma(a, a)$ distribution, with $a = 10^{-3}$, as presented in Ntzoufras (*14*) is chosen. The prior of θ is given by

$$p(\theta) = \frac{a^a}{\Gamma(a)} \theta^{a-1} \exp\left(-a\theta\right). \tag{6}$$

Under the parameterization in expression (6) $E(\theta) = a / a$ and $Var(\theta) = a / a^2$. Thus, for $a = 10^{-3}$ the prior distribution of θ is a flat, non-informative distribution with mean equal to 1 and variance equal to 1000.

The joint posterior distribution of $\boldsymbol{\beta}, \theta \mid \mathbf{y}$ is now proportional to $p(\boldsymbol{\beta}, \theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \theta) p(\boldsymbol{\beta}) p(\theta)$, which leads to expression

$$p(\boldsymbol{\beta},\theta\,|\,\mathbf{y}) \propto \prod_{i=1}^{n}\left[\frac{\Gamma(y_i+\theta)}{\Gamma(\theta)}\frac{\exp\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)^{y_i}\theta^{\theta}}{\left[\exp\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)+\theta\right]^{y_i+\theta}}\right]\times\exp\left(-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}\right)\times\theta^{a-1}\times\exp\left(-a\theta\right). \quad (7)$$

Inference from the posterior distribution is again not straightforward, since expression (7) does not have a closed form solution. In the following section, we describe a Markov Chain Monte Carlo method known as the *Metropolis-Hastings* algorithm, which is utilized in order to generate samples from the posterior distributions in expressions (3) and (7).

4. METROPOLIS-HASTINGS IMPLEMENTATION

Markov Chain Monte Carlo (MCMC) methods are frequently used within the Bayesian framework and are mainly employed in situations where the posterior distribution is not of known form. The basic idea of MCMC is to initiate a Markov process from a specific starting point and then iterate the process over a sufficient period of time. Due to the properties of Markov processes, the resulting chain eventually converges to a stationary distribution which is also the "target" posterior distribution. Once this is accomplished, an initial part of the chain is discarded as part of the so-called "burn-in" period of the chain, which is the period that the Markov chain has not yet reached convergence. The final result of MCMC is a dependent sample from the posterior distribution, from which one may acquire summaries for any posterior quantity of interest. Analytic information over the theoretical background and applications of various MCMC algorithms can be found in Gamerman and Lopes (*17*) and Gilks et al. (*18*).

Among the different types of MCMC methods, the Metropolis-Hastings (M-H) algorithm is the most general method. The M-H algorithm is an iterative method, which requires initially, specification of *proposal distributions* and of *starting values* for all parameters included in a given model. The iterative procedure follows; at each iteration draws of parameters are generated first from the proposal distributions, the draws are then accepted or rejected according to a certain *transition* or *acceptance probability*. An extensive description of the M-H algorithm is provided by Chib and Greenberg (*19*).

In particular, an *independence-chain* M-H algorithm is utilized where the location and scale parameters of the proposal distribution remain fixed. The large data size results to considerable time-consuming calculations and independence-chain M-H simulation proves to perform faster than *random-walk-chain* M-H or other types of *Metropolis-within-Gibbs* algorithms. The choice for the proposal distribution of parameter $\boldsymbol{\beta}$, common in both the Poisson and the Negative-Binomial model, is a multivariate normal distribution, $q(\boldsymbol{\beta}) \sim \mathbf{N}_{\mathbf{p+1}}(\tilde{\boldsymbol{\beta}},\tilde{\mathbf{V}}_{\boldsymbol{\beta}})$, where $\tilde{\boldsymbol{\beta}}$ is the ML estimate of $\boldsymbol{\beta}$ and $\tilde{\mathbf{V}}_{\boldsymbol{\beta}}$ is the estimated covariance matrix of $\boldsymbol{\beta}$. For parameter θ of the Negative-Binomial model, the proposal distribution is defined as $q(\theta) \sim Gamma(\tilde{a},\tilde{b})$, where parameters $\tilde{a}$ and $\tilde{b}$ are set to satisfy $\tilde{a}/\tilde{b} = \tilde{\theta}$ and $\tilde{a}/\tilde{b}^2 = Var\left(\tilde{\theta}\right)$ with $\tilde{\theta}$ being the ML estimate of θ. Having specified the proposal distributions, the M-H algorithm for each model proceeds as presented below.

To simulate a M-H sample of size N for the Poisson model:

1) Set initial value $\boldsymbol{\beta}^0$
2) For iterations $t = 1, 2, ..., N$:
   a. Generate $\boldsymbol{\beta}^*$ from the proposal $q(\boldsymbol{\beta})$
   b. Calculate the transition probability $a_{MH} = \min\left[\dfrac{p(\boldsymbol{\beta}^* \mid \mathbf{y}) q(\boldsymbol{\beta}^{t-1})}{p(\boldsymbol{\beta}^{t-1} \mid \mathbf{y}) q(\boldsymbol{\beta}^*)}, 1\right]$
   c. Generate a uniform random number u from $U(0,1)$
   d. Set $\boldsymbol{\beta}^t = \begin{cases} \boldsymbol{\beta}^* & \text{, if } u \le a_{MH} \\ \boldsymbol{\beta}^{t-1} & \text{, if } u > a_{MH} \end{cases}$

To simulate a M-H sample of size N for the Negative-Binomial model:

1) Set initial values $\boldsymbol{\beta}^0$ and θ^0
2) For iterations $t = 1, 2, ..., N$:
   a. Generate $\boldsymbol{\beta}^*$ from the proposal $q(\boldsymbol{\beta})$ and θ^* from the proposal $q(\theta)$
   b. Calculate the transition probability $a_{MH} = \min\left[\dfrac{p(\boldsymbol{\beta}^*, \theta^* \mid \mathbf{y}) q(\boldsymbol{\beta}^{t-1}) q(\theta^{t-1})}{p(\boldsymbol{\beta}^{t-1}, \theta^{t-1} \mid \mathbf{y}) q(\boldsymbol{\beta}^*) q(\theta^*)}, 1\right]$
   c. Generate a uniform random number u from $U(0,1)$
   d. Set $(\boldsymbol{\beta}^t, \theta^t) = \begin{cases} (\boldsymbol{\beta}^*, \theta^*) & \text{, if } u \le a_{MH} \\ (\boldsymbol{\beta}^{t-1}, \theta^{t-1}) & \text{, if } u > a_{MH} \end{cases}$

After certain preliminary tests, 5000 iterations for the Poisson model and 21000 iterations for the Negative-Binomial model were used in the final M-H runs, with resulting acceptance ratios of 95% and 57%, respectively. The first 1000 iterations were discarded as the "burn-in" part for both models. Convergence checks were based on the methods of Raftery and Lewis (*20*), Geweke (*21*) and Heidelberger and Welch (*22*). The sample of the Poisson model passed all the diagnostics, but due to memory limitations in calculations every 4th iteration was kept, resulting to a final sample of size 1000. Regarding the Negative-Binomial model, the diagnostic of Raftery and Lewis (*20*) indicated autocorrelation problems. In order to break the strong autocorrelations, every 40th draw of the sample was kept. For the final sample of 500 draws, all lag 1 autocorrelations were below 0.05.

5. RESULTS

In this section, results from the Poisson and Negative-Binomial regressions are summarized. Posterior summaries, model comparison and plots of the posterior distributions are presented first. A strategy for the Negative-Binomial model is suggested next, which allows to obtain predictions from the corresponding Poisson predictive distribution. Several goodness-of-fit tests are applied on the predictions and finally examples of predictive inference are presented.

5.1 Posterior Inference

The results presented in this section, apply to the exponential parameters, $B_j = \exp(\beta_j)$ for $j = 0,1,2,...21$. The effect of these parameters on the mean OD flows is multiplicative on natural scale and therefore interpretation is straightforward. For instance, posterior means greater than 1 correspond to an increasing multiplicative effect, whereas posterior means less than 1 have a decreasing multiplicative effect.

Posterior means, standard deviations and 95% probability intervals for parameters B_j and parameter θ are summarized in Table 1.

TABLE 1 Posterior Means, Standard Deviations, 95% Probability Intervals and the Values of DIC for the Poisson and Negative-Binomial Models

Parameter	Poisson			Negative-Binomial		
	Mean	SD	95% P.I.	Mean	SD	95% P.I.
B_0 **; intercept**	39.788	1.2305	(37.485-42.148)	0.1440	0.0949	(0.0307-0.3954)
B_1 **; dum.prov**	1.0301	0.0001	(1.0300-1.0301)	1.0268	0.0003	(1.0263-1.0273)
B_2 **; dum.arron**	1.0391	0.0001	(1.0391-1.0392)	1.0349	0.0004	(1.0342-1.0357)
B_3 **; dum.dist**	1.0413	0.0001	(1.0412-1.0413)	1.0423	0.0007	(1.0411-1.0436)
B_4 **; dum.kant**	1.0494	0.0001	(1.0494-1.0495)	1.0552	0.0008	(1.0537-1.0568)
B_5 **; dum.munic**	1.0733	0.0001	(1.0733-1.0734)	1.0855	0.0014	(1.0832-1.0885)
B_6 **; munic.kant**	0.8689	0.0015	(0.8657-0.8716)	0.7057	0.0210	(0.6627-0.7487)
B_7 **; munic.dist**	1.2729	0.0023	(1.2687-1.2777)	1.0486	0.0421	(0.9655-1.1289)
B_8 **; munic.arron**	0.6325	0.0008	(0.6308-0.6341)	1.0561	0.0329	(0.9935-1.1208)
B_9 **; munic.prov**	0.1528	0.0009	(0.1511-0.1545)	0.3222	0.0355	(0.2585-0.3986)
B_{10} **; empl.o**	0.7191	0.0052	(0.7084-0.7294)	7.3389	0.7637	(5.9655-8.8454)
B_{11} **; empl.d**	2.2170	0.0142	(2.1906-2.2444)	6.3666	0.5890	(5.3556-7.5903)
B_{12} **; pop.dens.o**	1.3304	0.0022	(1.3261-1.3349)	2.2587	0.0472	(2.1761-2.3598)
B_{13} **; pop.dens.d**	2.5036	0.0051	(2.4938-2.5136)	3.2724	0.0631	(3.1517-3.3987)
B_{14} **; road.length.o**	0.7478	0.0019	(0.7441-0.7515)	0.9361	0.0294	(0.8800-0.9964)
B_{15} **; road.length.d**	0.9144	0.0029	(0.9090-0.9201)	1.2809	0.0423	(1.1980-1.3676)
B_{16} **; perim.o**	1.5712	0.0041	(1.5633-1.5789)	5.9521	0.2311	(5.5425-6.3852)
B_{17} **; perim.d**	3.0781	0.0098	(3.0588-3.0975)	4.1485	0.1451	(3.8740-4.4447)
B_{18} **; HWT.o**	1.0013	0.0002	(1.0009-1.0017)	0.9730	0.0025	(0.9683-0.9776)
B_{19} **; HWT.d**	1.0203	0.0002	(1.0198-1.0208)	1.0149	0.0026	(1.0093-1.0198)
B_{20} **; PMT.o**	1.0217	0.0012	(1.0195-1.0242)	0.9184	0.0145	(0.8905-0.9443)
B_{21} **; PMT.d**	2.0998	0.0036	(2.0928-2.1065)	1.5790	0.0225	(1.5323-1.6196)
θ **; theta**	–			0.2047	0.0015	(0.2016-0.2074)
DIC	3,620,498			329,157.4		

Statistical significance may be checked directly upon examination of the 95% posterior probability intervals. Regarding parameters B_j of the Poisson model, none of the corresponding posterior intervals includes the value of 1, consequently all parameters have significant effects. In the Negative-Binomial model parameters B_7 and B_8 do not seem to have a significant effect. The rest of the regression parameters are significant. For the case of dispersion parameter θ of the Negative-Binomial model, the posterior interval does not support the value of zero, therefore parameter θ is also significant. Based on the posterior means of regression parameters B_j, the parameters that seem to have a greater impact, especially in the Negative-Binomial model, are B_{10}, B_{11}, B_{12}, B_{13}, B_{16} and B_{17}, which correspond to the effects of employment ratio, of population density and of perimeter length for the zones of origin and destination, respectively. Finally, parameter B_{21} corresponding to the effect of yearly traffic in provincial/municipal roads of destination zones is also strongly influential in both models.

In addition to posterior point estimates and intervals presented in Table 1, direct examination of the posterior distribution often provides extra information and a more comprehensive view regarding the random nature of parameters. Kernel smoothed estimates of the 23 posterior distributions for the parameters of the Negative-Binomial model are presented in Figure 1.

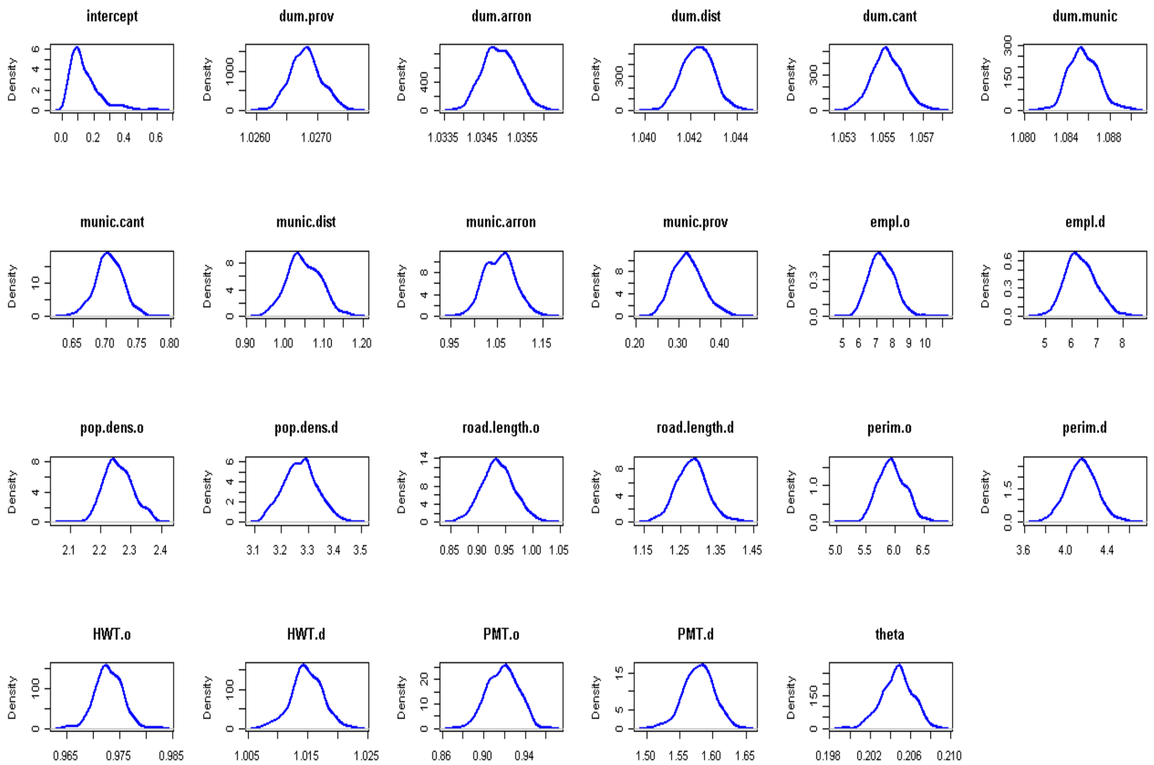

FIGURE 1 Kernel posterior distribution estimates for the parameters of the Negative-Binomial model.

Model comparison is based on the Deviance Information Criterion (DIC), introduced by Spiegelhalter et al. (*23*). The DIC is a model selection criterion, useful in determining the best model within a specific group of models. Based on the DIC support is given to the model with the lowest resulting value. The DIC values for the two models are also shown in Table 1, indicating that the value of the Negative-Binomial model is much lower than the corresponding value of the Poisson model. Consequently, according to the DIC, the Negative-Binomial model clearly outperforms the simple Poisson model. Evidently, the latter does not provide a good fit to the data due to the strong presence of over-dispersion. This is in accordance with the finding that parameter θ, which accounts for the extra variability, is statistically significant.

5.2 Prediction

According to a lemma provided by Sapatinas (*24*), if $y \mid \mu, u \sim Pois(\mu u)$ and u has a probability function $G(\cdot)$, i.e. $u \sim G(u)$, then, posterior expectations of u can be derived from the formula

$$E\left(u^r \mid y, \mu\right) = \frac{(y+r)!}{\mu^r y!} \frac{p_G(y+r)}{p_G(y)}, \tag{8}$$

where $p_G(\cdot)$ is the probability function of the corresponding mixed Poisson distribution. Expression (8) holds for all cases of mixed Poisson models. The formula is also utilized by Karlis (*16*) in a general EM algorithm for mixed Poisson models.

In our context, the mixed Poisson distribution corresponds to the Negative-Binomial distribution, denoted previously as $p(\mathbf{y} \mid \boldsymbol{\beta}, \theta)$ and given in expression (5). It is then possible, given formula (8), to obtain a sample of posterior expectations of **u**; let *(l)* be an indicator for the 500 MCMC draws, then, by setting in (8) $r=1$ and by "plugging-in" the MCMC draws $\boldsymbol{\beta}^{(l)}$, $\theta^{(l)}$, for $l = 1,2,...500$, we obtain posterior expectations of **u** as follows

$$\mathbf{u}_{\mathbf{EXP}}^{(l)} = E\left(\mathbf{u} \mid \mathbf{y}, \boldsymbol{\beta}, \theta\right)^{(l)} = \frac{(\mathbf{y}+1)!}{\exp(\mathbf{X}\boldsymbol{\beta}^{(l)})\mathbf{y}!} \frac{p(\mathbf{y}+1 \mid \boldsymbol{\beta}^{(l)}, \theta^{(l)})}{p(\mathbf{y} \mid \boldsymbol{\beta}^{(l)}, \theta^{(l)})}. \tag{9}$$

Now, predictions of OD flows can be generated from the Poisson distribution conditional on $\boldsymbol{\beta}$ and $\mathbf{u}_{\mathbf{EXP}}$; for each $\boldsymbol{\beta}^{(l)}$ and $\mathbf{u}_{\mathbf{EXP}}^{(l)}$, with $l = 1,2,...500$, we generate one predictive dataset $\mathbf{y}^{pred(l)}$ from

$$\mathbf{y}^{pred(l)} \sim Pois(\boldsymbol{\beta}^{(l)} \mathbf{u}_{\mathbf{EXP}}^{(l)}). \tag{10}$$

Each one of the 500 $\mathbf{y}^{pred}$'s, consists of one predictive OD matrix for Flanders. Predictions from the Poisson distribution, unlike predictions from the Negative-Binomial distribution, take into account the specific random intercept of each OD flow. The proximity of these predictions with respect to the original dataset is investigated next.

5.3 Goodness-of-fit

In order to evaluate the goodness-of-fit of the Negative-Binomial model, several measures of fit are considered. A measure frequently used within the transportation field is initially calculated. Bayesian methodology enhances the information provided by the measure, since the outcome is once again a distribution estimate rather than a point estimate. Evaluation of the fit is then supplemented by statistical tests based on *Bayesian p-values.*

The distance between the predictive datasets and the initial dataset is assessed by the Mean Absolute Percentage Error (MAPE) measure, which corresponds to an average percentage of deviation from the initial dataset. By definition, the calculation of MAPE cannot include the zero-valued cells of the OD matrix. Nevertheless, in large OD matrices, small or even medium deviations from zero-valued cells are usually not influential. If we denote with m the total number of cells which are not zero and with k an indicator $k = 1, 2, ...m$ for $y_k > 0$, then, we obtain 500 corresponding MAPE values from

$$MAPE^{(l)} = \sum_{k=1}^{m} \left| \frac{y_k - y_k^{pred(l)}}{y_k} \right| \Big/ m \, ,$$

for $l = 1, 2, ..., 500$. The resulting mean value of MAPE is 0.45, with a minimum of 0.445 and a maximum of 0.459. The mean MAPE seems relatively high, corresponding to a 45% deviation from the initial dataset. Nevertheless, this value is slightly misleading due to the fact that MAPE is also highly influenced from small deviations in low-valued cells. Excluding categories of low-valued cells in the calculation of MAPE, reveals that the mean value decreases drastically; the value of the mean MAPE for OD flows greater than 10 is decreased to 0.134 and for OD flows which are greater than 20 the corresponding value becomes 0.1. Finally, for OD flows greater than 50 the mean is 0.067, with a minimum of 0.065 and a maximum of 0.07. These results are summarized in the plots of Figure 2; as we observe in plot (c) the mean of MAPE is decreasing steadily and the deviations from the initial dataset become almost negligible for medium and large valued cells.

According to MAPE the Negative-Binomial models performs well for prediction of medium and large OD flows. The 6.7% deviation for OD flows greater than 50 is already small. Yet, MAPE is not very informative concerning the fit of the model in low-valued cells, since small deviations, which may not be significant in practical terms have a high influence in the calculation of the measure. A direct way of evaluating the fit in low-valued cells is to simply calculate the absolute differences between the initial and the predictive datasets. Plot (d) in Figure 2 is a histogram with a summary of the average absolute differences for OD flows equal to or less than 50. Note that the differences are not large; the mean equals 0.68, 50% are equal to or less than 0.18, 75% are equal to or less than 0.79 and the maximum absolute difference is 19.28.

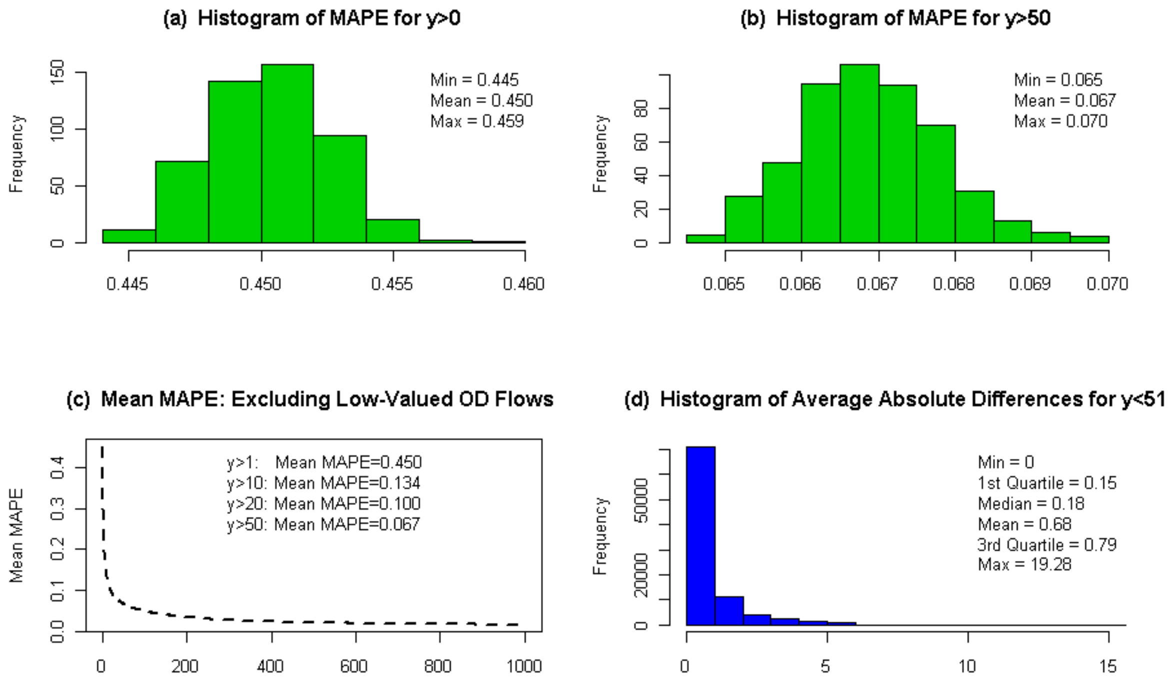

FIGURE 2 Histogram of MAPE (a), histogram of MAPE for OD flows greater than 50 (b), plot of the mean values of MAPE resulting by excluding low-valued cells (c) and histogram of the average absolute differences for OD flows equal or less than 50 (d).

.

In addition to the previous analysis, two extra measures of discrepancy between the predictions of the model and the data are considered; the *absolute distances* and the *squared distances* of the initial and the predictive data from the corresponding expected values of the model. In Bayesian terms, the measures are identified as *test quantities* which are evaluated by means of Bayesian p-values. A Bayesian p-value should ideally equal 0.5, extreme values very close to 0 or 1 suggest failure of a model in the specific aspect that is investigated by the test quantity (*25*). The Bayesian p-value was initially defined by Rubin *(26*), several examples for the use of test quantities and interpretation of Bayesian p-values are presented in Gelman et al. (*25*). Following the terminology used by Gelman et al. (*25*) we denote the two test quantities as

$$\text{Absolute-Distance: } T_1(\mathbf{y},\boldsymbol{\beta},\theta) = \sum_{i=1}^{n} \left| y_i - E(y_i \mid \boldsymbol{\beta},\boldsymbol{\theta}) \right|$$

$$\text{Squared-Distance: } T_2(\mathbf{y},\boldsymbol{\beta},\theta) = \sum_{i=1}^{n} \left(y_i - E(y_i \mid \boldsymbol{\beta},\boldsymbol{\theta}) \right)^2.$$

The resulting Bayesian p-value is 0 for the Absolute-Distance quantity, indicating a bad fit, and 0.488 for the Squared-Distance quantity which actually suggests a very good fit. The result at first glance seems contradictive, nevertheless it is in accordance with the previous findings. The Absolute-Distance is a strict measure which assigns more penalty to small deviations, while the Squared-Distance measure gives more weight to large deviations from the data. Like MAPE, the Absolute-Distance measure is influenced by small deviations, especially in low-valued cells. Given the size of the data, the cumulative effect of these deviations appears to be statistically significant under certain strict measures, yet in practical terms the overall effect is not significant. In our case, the Squared-Distance measure seems a more suitable test quantity for evaluating goodness-of-fit.

5.4 Predictive Inference

The 500 datasets generated from the predictive distribution in expression (10) may now be used in various types of predictions of traffic volumes. As mentioned in section 2.1, modeling on the level of municipalities allows for prediction on other levels of aggregation as well. For instance, predictions for OD flows between districts can be derived directly as summations of the predictions for OD flows between municipalities. Thus, predictive inference is not necessarily restricted on the level of municipalities; it can be applied on any other hierarchical level, such as the levels of cantons, districts, arrondissements and provinces. In addition, prediction may also be focused on specific types of traffic volumes that might be of interest, e.g. strictly in-coming trips, strictly out-coming trips or just internal trips.

In Figure 3, applications of prediction on different levels of aggregation and for different types of trips are demonstrated. The applications correspond to predictions for the total number of in-coming, going-to-work/school trips from all other municipalities to the capital of Flanders, Antwerp, predictions for the total number of going-to-work/school trips that occur daily in the whole region of Flanders and finally predictions for the daily internal going-to-work/school trips that take place in each one of the five Flemish provinces.

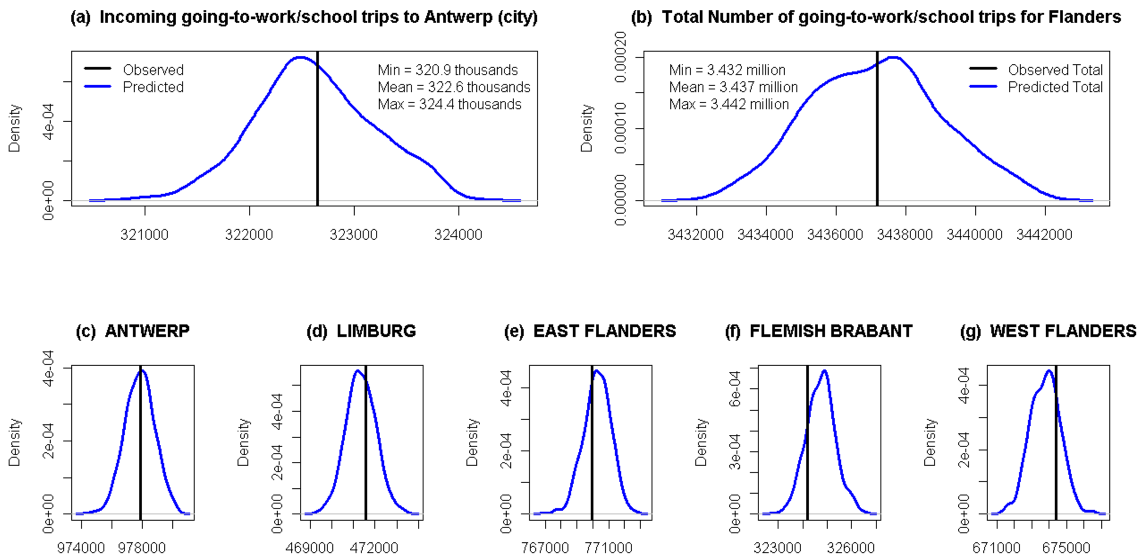

FIGURE 3 Going-to-work/school trip predictive distributions for incoming trips to Antwerp (a), for total number of trips in Flanders (b) and for internal trips within each of the five Flemish provinces; Antwerp (c), Limburg (d), East Flanders (e), Flemish Brabant (f) and West Flanders (g). The vertical black lines indicate the corresponding observed quantities.

Similar predictive distributions can be derived for any case of specific OD flows that might be of particular interest. It is worth noting, that these predictions also serve as further goodness-of-fit tests, since in every case there is a corresponding observed quantity to compare with. In the applications above, the observed quantities are represented with vertical black lines. As illustrated in Figure 3, all observed quantities are well within high-density regions of the corresponding predictive distributions, an indication that the predictions are not extreme with respect to the initial data.

In general, the predictive distributions provide all the necessary information concerning the variability of future traffic flows. The predictive effects may be examined under different assumptions; one might choose to infer based on conservative summaries such as the predictive mean or median, or one might be interested in examining the effect of more extreme summaries such as the 99^{th} percentile or the maximum value. These alternative options reduce overall uncertainty and may serve as predictive scenarios for transportation policy-makers, e.g. in decisions concerning infrastructure expansion.

6. CONCLUSIONS AND DISCUSSION

In this paper, OD matrix estimation from census data was investigated from a Bayesian modeling perspective. Applications of a Poisson model and of a Negative-Binomial model were presented for the municipality network of Flanders. All of the regression parameters of the Poisson model and most of the parameters of the Negative-Binomial model including the dispersion parameter proved to be statistically significant. Model comparison based on the DIC indicated that Negative-Binomial regression is a more suitable choice than simple Poisson regression due to the great degree of over-dispersion present in OD flows. Finally, predictions were obtained from the corresponding hierarchical structure of the Negative-Binomial model, conditional on the posterior expectation of the mixing parameters. The proximity of these predictions with respect to the initial data was evaluated according to several measures of discrepancy. The overall fit was found to be satisfactory.

A novel application emerges as a direct extension of the proposed methodology. The application entails using the predictive output of a certain model as input to a specific *assignment* method. That would allow for predictions on the level of *link flows* and also provide the opportunity to additionally compare *observable* link flows with respect to the corresponding predictive distributions.

Future research may focus further on the selection of explanatory variables. The choice of explanatory variables used, should be viewed as a first attempt and not as a concluding proposition. Expanding the models, by including appropriate explanatory variables that influence the generation and attraction of trips, is a matter of ongoing research. For instance, variables related to distances and coordinates proved to be highly significant in experiments of smaller scale and will be included in future results.

Uncertainty over model choice also provides space for further investigation. The class of mixed Poisson distributions, results to several potential models that might be reasonable candidates for OD matrix modeling. The widely used Poisson-Log Normal model, for example, appearing more frequently in the relative literature as a Poisson model with normally distributed random effects, is a possible alternative to the Poisson-Gamma model. A less known alternative belonging to the same class, is the Poisson-Inverse Gaussian regression model.

Finally, it is arguable that the proposed methodology may serve as an effective alternative to the traditional four-step transportation model for cases in which historical OD data exist. From this point of view the methodology may be seen as a joint trip generation, trip attraction and trip distribution method which integrates the first two phases of a four-step model in one statistical model with wider predictive capabilities.

REFERENCES

(1) Cools, M., E. Moons, and G. Wets. Assessing Quality of Origin-Destination Matrices Derived from Activity and Travel Surveys: Results from a Monte Carlo Experiment. Forthcoming in *Transportation Research Record: Journal of the Transportation Research Board*, 2010.

(2) Hensher, D.A. and K.J. Button. *Handbook of transport modelling*. Elsevier Ltd., London, 2000.

(3) Henson, K., K. Goulias, and R. Golledge. An assessment of activity-based modeling and simulation for applications in operational studies, disaster preparedness, and homeland security. *Transportation Letters: The International Journal of Transportation Research*, vol. 1, 2009, pp. 19-39.

(4) Abrahamsson, T. *Estimation of Origin-Destination Matrices Using Traffic Counts – A Literature Survey.* IIASA Interim Report IR-98-021, 1998.

(5) Timms, P. A philosophical context for methods to estimate origin destination trip matrices using link counts. *Transport Reviews: A Transnational Transdisciplinary Journal*, Vol. 21, No. 3, 2001, pp. 269-301.

(6) Maher, M.J. Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. *Transportation Research Part B: Methodological*, Vol. 17, 1983, pp. 435-447.

(7) Tebaldi, C. and M. West. Bayesian Inference on Network Traffic Using Link Count Data. *Journal of the American Statistical Association*, Vol. 93, 1996, pp. 557-576.

(8) Li, B. Bayesian Inference for Origin-Destination Matrices of Transport Networks Using the EM Algorithm. *Technometrics*, Vol. 47, 2005, pp. 399-408.

(9) Castillo, E., J.M. Menéndez, and S. Sánchez-Cambronero. Predicting traffic flow using Bayesian networks. *Transportation Research Part B: Methodological*, Vol. 42, 2008, pp. 482-509.

(10) Hazelton, M.L. Bayesian inference for network-based models with a linear inverse structure. *Transportation Research Part B: Methodological*, Vol. 44, 2010, pp. 674-685.

(11) FOD Economie, *FOD Economie - SPF Economie - Statbel.* www.statbel.fgov.be. Accessed February 22, 2010.

(12) Agresti, A. *Categorical Data Analysis, Second edition.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2002.

(13) McCullagh, P. and J.A. Nelder. *Generalized Linear Models, Second Edition.* Chapman and Hall/CRC, London, 1989.

(14) Ntzoufras, I. *Bayesian Modeling Using WinBUGS.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2009.

(15) Fernández, C., E. Ley, and M.F.J. Steel. Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, Vol. 100, 2001, pp. 381-427.

(16) Karlis, D. A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, Vol. 1, No. 1, 2001, pp. 305-318.

(17) Gamerman, D. and H.F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition.* Chapman and Hall/CRC, London, 2006.

(18) Gilks, W., S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics.* Chapman and Hall/CRC, London, 1995.

(19) Chib, S. and E. Greenberg. Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, Vol. 49, No. 4, 1995, pp. 335-327
(20) Raftery, A.E. and S. Lewis. How Many Iterations in the Gibbs Sampler? *In Bayesian Statistics*, Vol. 4, 1992, pp. 763-773.
(21) Geweke, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *In Bayesian Statistics*, Vol. 4, 1992, pp. 169-193.
(22) Heidelberger, P. and P.D. Welch. Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, Vol. 31, 1983, pp. 1109-1144.
(23) Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A.V.D. Linde. Bayesian measures of model complexity and fit. *Journal Of The Royal Statistical Society Series B*, Vol. 64, No. 4, 2002, pp. 583-639.
(24) Sapatinas, T. Identifiability of mixtures of power-series distributions and related characterizations. *Annals of the Institute of Statistical Mathematics*, Vol. 47, No. 3, 1995, pp. 447-459.
(25) Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis, Second Edition.* Chapman & Hall, London, 2003.
(26) Rubin, D.B. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, Vol. 12, No. 4, 1984, pp. 1151-1172.