Published in : Journal of the Royal Statistical Society. Series A Statistics in Society (2015)
Status : Postprint (Author's version)

Bayesian inference for transportation origin-destination matrices: the Poisson-inverse Gaussian and other Poisson mixtures

Konstantinos Perrakis[1], Dimitris Karlis[2], Mario Cools[3] and Davy Janssens[4]

[1]*University of Athens, Greece, and Hasselt University, Belgium*
[2]*Athens University of Economics and Business, Greece*
[3]*University of Liège, Belgium*
[4]*Hasselt University, Belgium*

Summary

Transportation origin-destination analysis is investigated through the use of Poisson mixtures by introducing covariate-based models which incorporate different transport modelling phases and also allow for direct probabilistic inference on link traffic based on Bayesian predictions. Emphasis is placed on the Poisson-inverse Gaussian model as an alternative to the commonly used Poisson-gamma and Poisson-log-normal models. We present a first full Bayesian formulation and demonstrate that the Poisson-inverse Gaussian model is particularly suited for origin-destination analysis because of its desirable marginal and hierarchical properties. In addition, the integrated nested Laplace approximation is considered as an alternative to Markov chain Monte Carlo sampling and the two methodologies are compared under specific modelling assumptions. The case-study is based on 2001 Belgian census data and focuses on a large, sparsely distributed origin-destination matrix containing trip information for 308 Flemish municipalities.

Keywords: Hierarchical Bayesian modelling; Integrated nested Laplace approximation; Origin-destination matrix; Overdispersion; Poisson mixtures

1. Introduction

In transportation analysis the *travel demand* within a geographical area, which is dividable into a given number of non-overlapping zones, is summarized by an origin-destination (OD) matrix which contains the *trips ox flows* that have occurred from each zone of that area to every other zone. Consider an area which can be divided into m zones and let T_{0d} denote the flows from zone of *origin o* to zone of *destination d*, where $o,d=1,2,...,m$. The OD matrix **T** is then

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_{12} & \dots & T_{1m} \\ T_{21} & T_{22} & \dots & T_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ T_{m1} & T_{m2} & \dots & T_{mm} \end{pmatrix}.$$

The elements T_{od}, for $o \neq d$, correspond to *interzonal* flows, whereas the elements across the main diagonal T_{oo} correspond to *intrazonal* flows. The marginal totals $T_{o\cdot} = \sum_d T_{od}$ and $T_{\cdot d} = \sum_o T_{od}$ are commonly referred to as *trip productions* and *trip attractions* respectively. In lexicographical order the matrix **T** can be represented as $\mathbf{y} = (y_1,y_2,...,y_n)^{\mathrm{T}} \equiv (T_{11}, T_{12},..., T_{mm})^{\mathrm{T}}$ *with* $n = m^2$. The inferential scope in OD modelling depends on several defining aspects such as spatial resolution, time resolution and classification by trip purpose. In addition, OD modelling is itself part of a larger inferential framework. Specifically, the traditional transportation modelling framework consists of a sequence of four modelling steps, namely (*i*) trip generation, (*ii*) trip distribution, (*iii*) modal split and (*iv*) traffic assignment.

Trip generation models are typically regression or cross-classification models which relate trip productions and trip attractions to socio-economic, location and land use variables. Trip distribution models balance trip productions and trip attractions, and distribute the trips to the cells of an OD matrix usually by using supplementary prior information in the form of an outdated OD matrix. Commonly used trip distribution models include gravity and direct demand models. The subsequent step of modal split entails disaggregating the OD matrix with respect to mode choice. Finally, traffic assignment involves allocating the n-m interzonal flows on a corresponding transport network consisting of all the available links which define the possible routes from zone of origin o to zone of destination d, for o, d = 1, 2,..., m and $o \neq d$. Interested readers are referred to Ortúzar and Willumsen (2001) for four-step modelling and to Thomas (1991) for traffic assignment.

In general, the four-step procedure remains widely accepted by transportation planners, so OD modelling up to the present is mainly based on trip generation and trip distribution principles. The first modern Bayesian approach to trip distribution, based on the gravity model, was discussed in West (1994). It is also worth noting that a different approach for OD estimation relies on information from link traffic data where the traffic assignment problem is actually inverted; see for example Tebaldi and West (1998) and Hazelton (2010) for Bayesian methods. The methodological framework is quite different under this approach and it is actually part of a broader literature on network tomography (e.g. Medina *et al.* (2002)). In this study we extend the methodology that was presented in Perrakis *et al.* (2012b) for OD modelling based on census data and Perrakis *et al.* (2012a) for traffic assignment inference through Bayesian predictions. Additional references concerning OD estimation from travel surveys and/or link traffic can be found in Perrakis *et al.* (2012b).

In particular, we investigate the performance of three Poisson mixture models, namely the Poisson-gamma (PG), Poisson-log-normal (PLN) and Poisson-inverse Gaussian (PIG) models. The PG model is the most commonly used and well-established model within the family of Poisson mixtures, whereas the PLN model remains up to the present the predominant alternative. The PIG model is the less known and less used model among the three, especially within the Bayesian framework. We present a first full Bayesian treatment of the PIG model and demonstrate that it has desirable properties both in its marginal and in its hierarchical forms. In addition, we consider the integrated nested Laplace approximation (INLA) framework (Rue *et al.*, 2009) as a potentially efficient alternative to Markov chain Monte Carlo (MCMC) methods for the PG and PLN models. The case-study focuses on a large-scale OD matrix, derived from the 2001 Belgian census study, containing trip information for 308 municipalities in the region of Flanders.

The paper is organized as follows. A literature review and Bayesian formulations for the three models in question are provided in Section 2. The OD matrix, the transport network of Flanders and the selection of explanatory variables are described in Section 3. Results are presented in Section 4. The paper ends with conclusions and considerations of future research in Section 5.

2. Poisson mixture models

With Poisson mixture models we assume that the OD flows y_i are independent and identically distributed Poisson realizations and that the rate of the Poisson distribution is $\lambda_i = \mu_i u_i$ for $i = 1,2,...,n$. The rate λ_i is split into two parts; μ_i is the part which is related to the vector of $p + 1$ unknown parameters $\beta = (\beta_0,\beta_1,..., \beta_p)^T$ and the set of explanatory variables $\mathbf{x}_i = (1, x_{i1},..., x_{ip})^T$ through the log-link function $\log(\mu_i) = \beta^T\mathbf{x}_i$, and μ_i is a random component— interpreted as a multiplicative random effect accounting for heterogeneity—which is attributed with a density $g_1(u_i)$. The Poisson mixture modelling formulation is summarized as

$$\begin{aligned} y_i &\sim \text{Pois}(\lambda_i), \qquad \lambda_i = \mu_i u_i, \\ \mu_i &= \exp(\boldsymbol{\beta}^T \mathbf{x}_i), \\ u_i &\sim g_1(u_i), \\ E(u_i) &= 1. \end{aligned}$$

The density g_1 is known as the mixing density and can be continuous, discrete or even a finite support distribution. The constraint on the expected value of the random component u_i ensures that the model is scale identifiable. Poisson mixtures are employed as overdispersed alternatives to the simple Poisson model which arises when the mixing density becomes degenerate. Alternatively, from a generalized linear mixed model perspective the above model can be expressed as

$$\begin{aligned} y_i &\sim \text{Pois}(\lambda_i), \qquad \log(\lambda_i) = \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i, \\ \varepsilon_i &\sim g_2(\varepsilon_i), \\ E(\varepsilon_i) &= 0, \end{aligned}$$

where ε_i is an additive random-error term. Here the constraint on the expected value ensures location identifiability. The two formulations are equivalent; however, the intercepts and the interpretations of marginal means are different owing to the identifiability constraints (Lee and Nelder, 2004). The Poisson likelihood is the conditional likelihood given the unobserved random-effect vector $\mathbf{u} = (u_1,u_2,...,u_n)^T$. Integration over $\mathbf{u}$ results in the marginal sampling likelihood, i.e. $p(\mathbf{y}|\boldsymbol{\mu}) = \int p(\mathbf{y}|\boldsymbol{\mu},\mathbf{u})\, g_1(\mathbf{u})\, d\mathbf{u}$. Frequentist inference usually focuses on the marginal structure under maximum likelihood (ML), restricted ML, quasi-likelihood and pseudolikelihood estimation procedures.

When the mixing density g_1 is a gamma distribution, we have the PG model which is the most frequently used Poisson mixture model because of the property that the resulting marginal likelihood is a negative binomial distribution. Properties and estimation procedures for negative binomial regression can be found in Lawless

(1987). The PG model is also included in the family of hierarchical generalized linear models that were introduced by Lee and Nelder (1996) who provided ML estimates for regression parameters as well as random effects based on the hierarchical likelihood. The PLN model arises when g_1 is a log-normal distribution. The resulting marginal distribution of this model, which is known simply as PLN (Shaban, 1988), does not have a closed form expression and thus numerical integration is needed for marginal estimation. Nevertheless, the PLN model is regularly used in practice owing to its distinct historical development as a generalized linear mixed model for count data based on the assumption that g_2 is a normal distribution (Breslow, 1984). Estimation of the model through Gaussian quadrature and the EM algorithm was handled in Aitkin (1996). An inverse Gaussian (IG) density for g_1 results in the PIG model which leads to a PIG marginal density. This distribution, unlike the PLN case, does have a closed form expression. The PIG model was first presented by Holla (1967). Information on ML estimation can be found in Dean *et al.* (1989) and in the references therein. The PIG model has been used in actuarial science (Willmot, 1987; Carlson, 2002) and in linguistics where the zero-truncated marginal form is of particular interest (e.g. Puig *et al.* (2009)).

A first consideration of all three models was presented in Chen and Ahn (1996). Later, Karlis (2001) provided a generally applicable EM algorithm for Poisson mixtures and compared the three models on a real data set. In Boucher and Denuit (2006), the performance of the three models was investigated from a random-effects *versus* fixed effects perspective on motor insurance claims. Finally, in Nikoloulopoulos and Karlis (2008) the models were compared with respect to distributional properties such as skewness and kurtosis under simulation experiments. This study illustrates some theoretical expectations, namely that the PLN and PIG models allow for longer right-hand tails and are thus more appropriate than the PG model for modelling highly positive skewed data.

From a Bayesian perspective, Poisson mixtures have a natural interpretation as hierarchical or multilevel models where the mixing distribution is considered as a first-level prior of which the parameters are assigned with a second-level prior or hyperprior. With respect to the equivalence between the multiplicative and additive forms, it is the choice of hyperprior which affects inferences about the intercept, depending on whether the $E(u_i) = 1$ or $E(\varepsilon_i) = 0$ constraint is imposed through the hyperprior. Bayesian applications of negative binomial modelling as well as hierarchical PG and PLN modelling can be found in Ntzoufras (2009) and in the references therein. Bayesian literature on PIG modelling is limited to the study of Font *et al.* (2013), which emphasizes a marginal, zero-truncated form of the model specifically suited for linguistic analysis.

In what follows, we present the hierarchical and marginal forms and properties of the three models, with emphasis placed on the PIG. MCMC sampling is based on a posterior factorization which is not common but is particularly convenient in our context given the large data size. Specifically, if we denote by ω the hyperparameter of the mixing prior of $\mathbf{u}$, then the joint posterior is $p(\boldsymbol{\mu}, \mathbf{u}, \omega|\mathbf{y}) = p(\mathbf{u}|\boldsymbol{\mu}, \omega, \mathbf{y})\, p(\boldsymbol{\mu}, \omega|\mathbf{y})$. Thus, for hierarchical inference one can use the marginal likelihood for sampling from $p(\boldsymbol{\mu}, \omega|\mathbf{y})$ and generate $\mathbf{u}$ subsequently from $p(\mathbf{u}|\boldsymbol{\mu}, \omega, \mathbf{y})$. As illustrated next, this is straightforward for the PG and PIG models.

2.1. Poisson-gamma model

For the PG model we make the following likelihood and prior assumptions:

$$y_i|\boldsymbol{\beta}, u_i \sim \text{Pois}\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)u_i\},$$

$$\boldsymbol{\beta} \sim \mathbf{N}_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}_\beta) \qquad \boldsymbol{\Sigma}_\beta = n(\mathbf{X}'\mathbf{X})^{-1},$$

$$u_i \sim \text{gamma}(\theta, \theta),$$

$$\theta \sim \text{gamma}(a, a) \qquad a = 10^{-3}.$$

For the multivariate normal prior of the regression parameters we adopt the g-prior structure (Zellner, 1986) analogue of the benchmark prior that was discussed in Fernández *et al.* (2001) for normal linear models. The same unit information multivariate prior is also adopted for the PLN and PIG models. The gamma prior for u_i is defined in terms of shape and rate parameters which both equal θ, so that $E(u_i) = 1$ and $\text{var}(u_i) = \theta^{-1}$. The gamma hyperprior for dispersion parameter θ, with shape and rate equal to 0.001, is a commonly used diffuse prior (Ntzoufras, 2009). The joint posterior distribution of all parameters is $p(\beta, \mathbf{u}, \theta|\mathbf{y}) \propto p(\mathbf{y}|\beta, \mathbf{u})\, p(\beta)\, p(\mathbf{u}|\theta)\, p(\theta)$. The only full conditional which has a known form is that of the random effects which is a gamma distribution, namely $u_i|\beta,\theta,y_i \sim \text{gamma}(y_i + \theta, \mu_i + \theta)$ (Gelman and Hill, 2006). Therefore, MCMC sampling for the hierarchical model would require a Metropolis-within-Gibbs type of algorithm with Metropolis steps for the joint conditional of $\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{u}, \mathbf{y}$ or for the conditionals of $\boldsymbol{\beta}|\mathbf{u}, \mathbf{y}$ and $\boldsymbol{\theta}|\mathbf{u}, \mathbf{y}$. Alternatively, adaptive rejection sampling can also be used. Integration over **u** leads to a negative binomial marginal likelihood, i.e.

$$y_i|\boldsymbol{\beta},\theta \sim \text{NB}\{\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i),\theta\}.$$

Under this parameterization the marginal mean and variance are given by $E(\mathbf{y}|\beta) = \exp(\mathbf{X}\beta)$ and $\text{var}(\mathbf{y}|\beta, \theta) = \exp(\mathbf{X}\beta) + \exp(\mathbf{X}\beta)^2\theta^{-1}$, with the variance being a quadratic function of the mean. The posterior distribution now is $p(\boldsymbol{\beta}, \theta|\mathbf{y})$ oc $p(\mathbf{y}|\beta, \theta)\, p(\boldsymbol{\beta})\, p(\theta)$, which leads to the expression

$$p(\boldsymbol{\beta},\theta|\mathbf{y}) \propto \exp(\mathbf{y}^{\text{T}}\mathbf{X}\boldsymbol{\beta} - \tfrac{1}{2}\boldsymbol{\beta}^{\text{T}}\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\beta} - a\theta)\theta^{n\theta+a}\Gamma(\theta)^{-n}\prod_i [\Gamma(y_i+\theta)\{\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)+\theta\}^{-(y_i+\theta)}]. \quad (1)$$

The Metropolis-Hastings (MH) algorithm is used to sample from the joint posterior of $\beta, \theta|\mathbf{y}$. Once M posterior draws of β and θ are available, predictive inference from the hierarchical structure of the model is straightforward; we generate first $\mathbf{u}^{(m)} \sim \text{gamma}\{\mathbf{y} + \theta^{(m)}, \exp(\mathbf{X}\beta^{(m)}) + \theta^{(m)}\}$ and then $\mathbf{y}^{\text{pred}(m)} \sim \text{Pois}\{\exp(\mathbf{X}\beta^{(m)})\mathbf{u}^{(m)}\}$ for $m = 1, 2, ..., M$. As shown in Sections 4.4 and 4.5, predictions are used in posterior predictive checks and also for quantifying input uncertainty in deterministic traffic assignment modelling.

It is worth noting that recent developments (Martins and Rue, 2013) extend the initial INLA framework (Rue *et al.,* 2009) to applications on near Gaussian latent models. Therefore, we also consider the INLA as an alternative to MCMC sampling for the PG model; a comparison is presented in Section 4.1.

2.2. Poisson log-normal model

The assumptions for the Poisson log-normal model are as follows:

$$y_i|\boldsymbol{\beta},u_i \sim \text{Pois}\{\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)u_i\},$$
$$\boldsymbol{\beta} \sim \mathbf{N}_{p+1}(\mathbf{0},\boldsymbol{\Sigma}_{\beta}) \qquad \boldsymbol{\Sigma}_{\beta} = n(\mathbf{X}^{\text{T}}\mathbf{X})^{-1},$$
$$u_i \sim \text{LN}(-\sigma^2/2,\sigma^2),$$
$$\sigma^2 \sim \text{inv-gamma}(a,a) \qquad a = 10^{-3}.$$

Following the formulation of Lee and Nelder (2004) for scale identifiability, the prior distribution of u_i has location parameter equal to $-\sigma^2/2$ and scale σ^2, and so $E(u_i) = 1$ and $\text{var}(u_i) = \exp(\sigma^2) - 1$. The inverse gamma hyperprior for σ^2 is the common option for this model (Ntzoufras, 2009); for $a = 10^{-3}$ the distribution of σ^{-2} is a diffuse gamma distribution. The joint posterior distribution is $p(\beta, \mathbf{u}, \sigma^2|\mathbf{y})$ α $p(\mathbf{y}|\beta, \mathbf{u})\, p(\beta)\, p(\mathbf{u}|\sigma^2)\, p(\sigma^2)$. In this case, none of the full conditional distributions are of known form. MCMC sampling for the hierarchical PLN model is in general more convenient in its generalized linear mixed model form where the full conditional distribution of σ^2 is again an inverse gamma distribution, namely $\sigma^2|\mathbf{u}, \mathbf{y} \sim \text{inv-gamma}\{a + n/2, a + \Sigma_i \log(u_i)^2/2\}$. Thus, in the additive case sampling from the conditionals of β and **u** is possible with Metropolis steps or rejection sampling. Note that in the generalized linear mixed model form the corresponding prior for u_i must be specified as $\text{LN}(0, \sigma^2)$.

In the PLN model the marginal likelihood $p(\mathbf{y}|\beta, \sigma^2)$ is not known analytically; nevertheless the mean and variance of the PLN distribution are available and given by $E(\boldsymbol{y}\backslash\beta) = \exp(\mathbf{X}/3)$ and $\text{var}(\mathbf{y}|\beta, \sigma^2) = \exp(\mathbf{X}\beta) + \exp(\mathbf{X}\beta)^2\{\exp(\sigma^2) - 1\}$. As with the PG model, the variance is a quadratic function of the mean. The joint posterior density is $p(\beta, \sigma^2|\mathbf{y})$ α $p(\mathbf{y}|\beta, \sigma^2)\, p(\beta)\, p(\sigma^2)$, namely

$$p(\boldsymbol{\beta},\sigma^2|\mathbf{y}) \propto \prod_i \left(\int \exp\left[y_i\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i - \exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)u_i - \frac{\{\log(u_i)+\sigma^2/2\}^2}{2\sigma^2}\right]u_i^{y_i-1}\,\text{d}u_i\right)$$
$$\times \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{\text{T}}\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\beta} - \frac{a}{\sigma^2}\right)(\sigma^2)^{-(n/2+a+1)}. \quad (2)$$

We employ MH simulation to sample from thejoint posterior density of β and σ^2. The integral appearing in the unnormalized posterior can be evaluated through numerical integration, e.g. with Gauss-Hermite quadrature which is also frequently employed in frequentist practice for marginal estimation. Another alternative that is examined in this study is Monte Carlo (MC) integration from the log-normal prior of the random effect vector **u** within the Metropolis kernel, i.e., for a given MH iteration t and draws $\beta^{(t)}$, $\sigma^{2(t)}$ the above integral can be evaluated by generating first L draws $\{u_i^{(t,l)}, l = 1, 2, ..., L\}$ from $u_i^{(t,l)} \sim \text{LN}(-\sigma^{2(t)}/2, \sigma^{2(t)})$ and then by calculating the marginal probability as $p(y_i|\beta^{(t)}, \sigma^{2(t)}) = L^{-1}\sum_l p(y_i|\beta^{(t)}, u_i^{(t,l)})$.

A potentially efficient alternative to MCMC approaches for the PLN model is the INLA framework that was introduced in Rue *et al.* (2009). The INLA approach covers the family of Gaussian Markov random-field models and is based on efficient approximating schemes for the marginal posterior distributions. The PLN model is included in the family of Gaussian Markov random-field models as the random effects are normally distributed on additive scale. In Section 4.1 we compare the INLA with MCMC methods.

2.3. Poisson-inverse Gaussian model

For the hierarchical PIG we adopt the following assumptions:

$$y_i|\boldsymbol{\beta}, u_i \sim \text{Pois}\{\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)u_i\},$$

$$\boldsymbol{\beta} \sim \mathbf{N}_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}_\beta) \qquad \boldsymbol{\Sigma}_\beta = n(\mathbf{X}^{\text{T}}\mathbf{X})^{-1},$$

$$u_i \sim \text{IG}(1, \zeta),$$

$$\zeta \sim \text{gamma}(a, a) \qquad a = 10^{-3}.$$

The initial parameterization of Holla (1967) is used for the IG prior, with mean μ and shape ζ, specifically

$$p(u_i|\mu, \zeta) = \left(\frac{\zeta}{2\pi u_i^3}\right)^{1/2} \exp\left\{-\frac{\zeta(u_i-\mu)^2}{2\mu^2 u_i}\right\}. \tag{3}$$

For μ = 1 we have that *a priori* $E(u_i)$ = 1 and var(u_i) = ζ^{-1}. The IG distribution is a special case of the three-parameter generalized inverse Gaussian (GIG) distribution which is generally conjugate to the family of exponential distributions and was studied in detail in Jorgensen (1982). The probability distribution function of a GIG(λ,ψ,χ) distribution with parameters $\lambda \in \mathbf{R}$, $\chi, \psi > 0$ is given by

$$f(x) = \frac{(\psi/\chi)^{\lambda/2}}{2\,K_\lambda\{\sqrt{(\psi\chi)}\}} x^{\lambda-1} \exp\left\{-\frac{1}{2}(\psi x + \chi x^{-1})\right\}, \tag{4}$$

where K_λ is the modified Bessel function of the third kind with order λ. The IG distribution arises for $\lambda = -\frac{1}{2}$. Interestingly, the gamma distribution is also a special case of the GIG distribution for χ = 0. For shape parameter ζ we adopt the usual gamma hyperprior, similarly to the PG model. The posterior distribution now is $p(\beta, \mathbf{u}, \zeta|\mathbf{y})$ oc $p(\mathbf{y}|\beta, \mathbf{u})$ $p(\beta)$ $p(\mathbf{u}|\zeta)$ $p(\zeta)$ and it can be easily shown that the full conditionals of u and ζ are known distributions, namely

$$u_i|\boldsymbol{\beta}, \zeta \sim \text{GIG}\{y_i - \tfrac{1}{2}, 2\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i) + \zeta, \zeta\}$$

and

$$\zeta|\mathbf{u} \sim \text{gamma}\left\{a + n/2, a + \sum_i (u_i - 1)^2/2u_i\right\}.$$

Athreya (1986) was the first to note the specific conjugate relationship between the IG and Poisson distribution; see also Karlis (2001). Regarding simulation from the GIG distribution, random generators are readily available (e.g. Dagpunar (1988)). Thus, the hierarchical PIG model is actually simpler in terms of MCMC sampling in comparison with the PG and PLN models, since all that is needed is an MH step or rejection sampling algorithm for the conditional of β.

Marginally we have that $y_i|\beta, \zeta \sim \text{PIG}\{\exp(\beta^{\text{T}}\mathbf{x}_i), \zeta\}$ for i = 1, 2,..., n with likelihood function given by

$$p(y_i|\boldsymbol{\beta}, \zeta) = K_{y_i-1/2}[\sqrt{\{2\zeta\phi(\mathbf{x}_i)\}}]\left(\frac{2\zeta}{\pi}\right)^{1/2} \frac{\exp\{\zeta/\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)\}}{y_i!}\{2\phi(\mathbf{x}_i)\zeta^{-1}\}^{1/2(y_i-1/2)}, \tag{5}$$

where

$$\phi(\mathbf{x}_i) = 1 + \frac{\zeta}{2\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)}. \tag{6}$$

The marginal mean and variance are $E(\mathbf{y}|\beta) = \exp(\mathbf{X}\beta)$ and $\text{var}(\mathbf{y}|\beta, \zeta) = \exp(\mathbf{X}\beta) + \exp(\mathbf{X}\beta)^3\zeta^{-1}$. The variance is thus a cubic function of the mean in the PIG model, allowing for greater overdispersion. The posterior distribution $p(\beta, \zeta|\mathbf{y})$ α $p(\mathbf{y}|\beta, \zeta)$ $p(\beta)$ $p(\zeta)$ can be expressed as

$$p(\boldsymbol{\beta}, \zeta|\mathbf{y}) \propto \prod_i \left[K_{y_i-1/2}[\sqrt{\{2\zeta\phi(\mathbf{x}_i)\}}]\exp\left\{\frac{\zeta}{\exp(\boldsymbol{\beta}^{\text{T}}\mathbf{x}_i)}\right\}\{2\phi(\mathbf{x}_i)\zeta^{-1}\}^{1/2(y_i-1/2)}\right]$$
$$\times \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{\text{T}}\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta} - a\zeta\right)\zeta^{n/2+a-1}. \tag{7}$$

Samples from the posterior of β and ζ can be obtained through MH simulation from the joint posterior. As with the PG model, when M posterior draws of β and ζ are available, predictive inference from the hierarchical structure of the PIG model is possible by generating first $\mathbf{u}^{(m)} \sim \mathrm{GIG}\{\mathbf{y} - \frac{1}{2}, 2\exp(X\beta^{(m)}) + \zeta^{(m)}, \zeta^{(m)}\}$ and then $\mathbf{y}^{\mathrm{pred}(m)} \sim \mathrm{Pois}\{\exp(\mathbf{X}\beta^{(m)})\mathbf{u}^{(m)}\}$ for $m = 1, 2, ..., M$.

3. Data

3.1. Origin-destination matrix and the transport network of Flanders

The OD matrix was derived from the 2001 Belgian census study and contains information about the departure and arrival locations for work- and school-related trips of the approximately 10 million Belgian residents. The recorded work or school trips refer to a normal weekday for all possible travel modes and are one directional, from zone of origin to zone of destination. The study area is not the entire country of Belgium, but the northern, Dutch-speaking region of Flanders which roughly accounts for 60% of the total population and 44% of the country's surface area. From an administrative viewpoint Flanders is divided into five provinces, 22 *arrondissements,* 52 districts, 103 cantons and 308 municipalities. Our analysis is implemented on the municipal level at which the OD matrix contains 94864 cells.

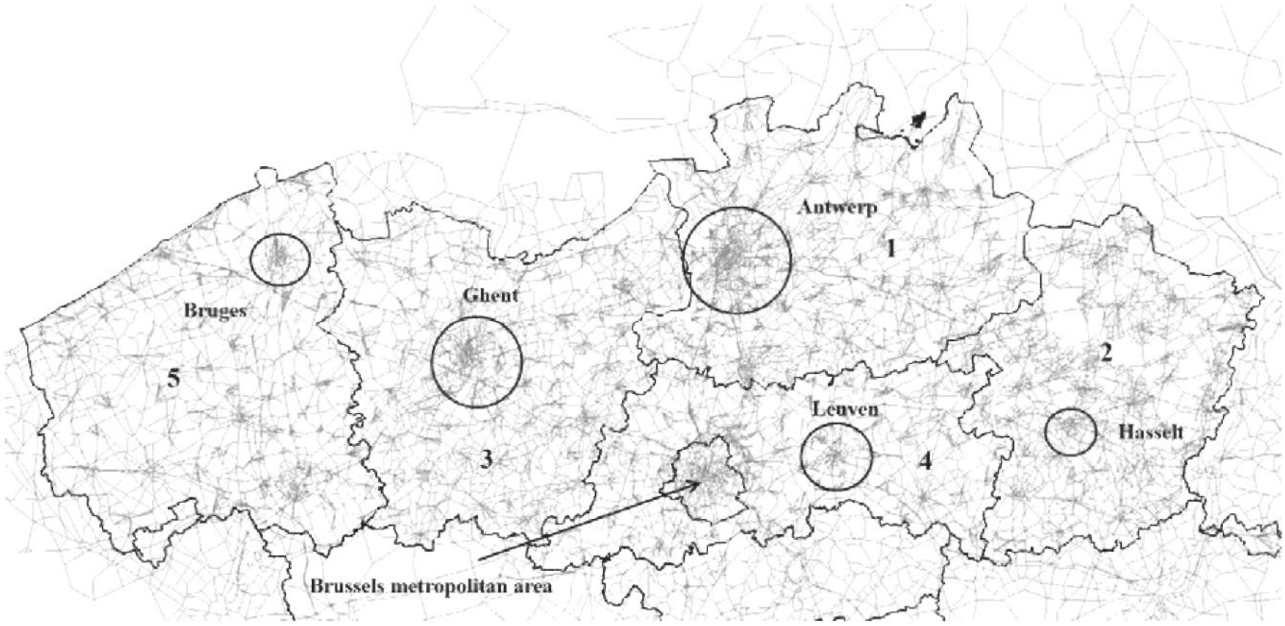
The OD flows on municipality zonal level are sparsely distributed and extremely overdispersed with large outlying observations. Approximately 63%o of the observations are zero valued, with an overall mean of 38.47 and a standard deviation of 960.47. All the zero-valued observations belong to interzonal flows off the main diagonal. The mean and standard deviation of interzonal flows are equal to 18.48 and 156.67 respectively. The maximum value is observed in the diagonal cell which corresponds to the intrazonal flows occurring in Antwerp—the largest Flemish municipality—and is equal to 211 681. In general, the majority of trips correspond to intrazonal flows with the counts on the main diagonal accounting for approximately 51 % of the total number of trips. The mean for intrazonal flows is 6064.32, and the standard deviation is 15516.64.

The road network of Flanders with the corresponding borders of the five Flemish provinces, Antwerp, Limburg, East Flanders, Flemish Brabant and West Flanders, is presented in Fig. 1. The circled areas indicate the capital municipality of each province; the size of each circle is a relative representation of population size. Antwerp is the most populated capital, followed by Ghent, Leuven, Bruges and Hasselt. Brussels metropolitan area, which is also marked on the map, is not included in the analysis as it is a separate administrative centre. Overall, the network runs a total length of 65296.72 km and contains 97450 links which can be categorized into highways (8.58% including entrance-exit road segments), main regional roads (15.49%), small regional roads (21.1%), local municipal roads (52.91%) and walk or bicycle paths (1.92%).

3.2. Explanatory variables

The set of explanatory variables consists of six categorical variables and 12 discrete or continuous variables. The first five categorical variables capture the effects of intrazonal flows measured in differences of 100 trips. Thus, these dummy variables take the value 100 if the trips are intrazonal in municipalities, DM, cantons, DC, districts, DD, *arrondissements,* DA, and provinces, DP, and 0 otherwise. These predictors capture individual effects; for instance, for intrazonal flows in the main diagonal the municipality predictor DM will equal 100, whereas DC, DD, DA and DP will equal 0. The sixth categorical predictor, DE, is associated with the effect of higher education institutes in destination zones; it takes the value of 1 if the destination zone supports a college and/or a university and 0 otherwise. The set of covariates includes four discrete-valued variables which contain the total number of neighbouring municipalities on canton, MC, district, MD, *arrondissement,* MA, and province, MP, levels for each corresponding OD pair. The rest of the covariates are continuous. Specifically, we include employment rate ER, population density PD (thousands of inhabitants per square kilometre), relative length of road networks, RL (road length in kilometres per surface area in square kilometres), perimeter length PL in kilometres, car ownership ratio CR, yearly traffic on highways, HT, and on provincial or municipal roads,

Fig. 1. *Road network of Flanders and the five Flemish provinces of 1, Antwerp, 2, Limburg, 3, East Flanders, 4, Flemish Brabant, and 5, West Flanders, with corresponding capitals Antwerp, Hasselt, Ghent, Leuven and Bruges*



PMT, in kilometres, and finally distance *D* in kilometres. All covariates are used on a logarithmic scale. Distance, of course, is zero for intrazonal municipality flows and to use the logarithm it is set equal to 0.1, a value which for most practical purposes refers to negligible distance (100 m). Furthermore, owing to the particularity of the OD problem variables ER, PD, RL, PL, CR, HT and PMT come in pairs, i.e. each is used twice, one time for the origin zone and one time for the destination zone. The arguments for employing the continuous variables in pairs are as follows:

(a) preliminary research revealed that it is better to use information for origin and destination zones separately rather than the average, for instance, between origin and destination zones;
(b) having separate parameter estimates for origin and destination zones allows for elementary comparison with trip production and trip attraction modelling;
(c) using pairs on a logarithmic scale and including distance provides an alternative interpretation of the Poisson mixture log-linear models as stochastic gravity, direct demand models.

Most of the continuous variables were transformed to ratios relative to populations or surface areas. The specific transformations were chosen to maintain reasonable interpretations, but also to solve multicollinearity problems which were evidently present in raw variables. Analysis based on variance inflation factors indicated no serious multicollinearity problems for the transformed variables with the highest variance inflation factor value being equal to 3.877.

4. Results

We start this section with a comparison between MCMC and INLA estimates for the PLN and PG models on an OD matrix of smaller scale. The full analysis for the entire of Flanders, including posterior and predictive inference based on MCMC sampling, is presented next. Details concerning MH implementation are presented in Appendix A.

4.1. Comparing Markov chain Monte Carlo and integrated nested Laplace approximation approaches

The comparison that is presented here concerns a 10 × 10 OD matrix containing the flows between the 10 largest (in terms of population) Flemish municipalities. The rationale in choosing a smaller OD matrix is to evaluate how well can INLAs approximate marginal posterior distributions under relatively small samples. The categorical predictors are not meaningful to use in this case; therefore, we use only employment rate, population density, length of road networks, highway traffic, provincial or municipal traffic and distance as covariates.

The reasons for considering a smaller OD are the following. First, the INLA is based on the assumption of conditional independence for the Gaussian latent random field which means that the inverse covariance matrix is sparsely distributed allowing for fast and efficient Cholesky decomposition. In general, the assumption of conditional independence becomes stronger as the dimensionality of the random field increases. Therefore, we find it interesting to evaluate INLAs on a smaller random field (i.e. 100 random effects, instead of 94 864, plus the regression parameters). Second, the INLA estimates the marginal posteriors either through Gaussian or through Laplace approximations based on Taylor's expansions around the posterior modes. Such approximations generally perform well when the sample size is large and the marginal posteriors are usually well centered near the posterior mode. Thus, we are interested in testing the INLA on a smaller subset of the data.

Table 1. *Posterior means and standard deviations (in parentheses) from an MH sample of 20000 draws and from the three INLA approaches for the PLN model†*

Parameter	*MH estimate*	*INLA estimates*		
		Gaussian	*Simplified Laplace*	*Laplace*
β_0 intercept	-1.031 (2.288)	-1.022 (2.623)	-1.061 (2.623)	-1.061 (2.623)
β_1 ER (o)	0.763 (0.865)	0.749 (0.985)	0.755 (0.985)	0.755 (0.985)
β_2 ER (d)	1.826 (0.883)	1.802 (0.979)	1.824 (0.979)	1.824 (0.979)
β_3 PD (o)	0.406 (0.420)	0.391 (0.476)	0.401 (0.476)	0.401 (0.476)
β_4 PD (d)	1.414 (0.425)	1.401 (0.477)	1.420 (0.477)	1.420 (0.477)
β_5 RL (o)	0.697 (0.779)	0.684 (0.892)	0.689 (0.891)	0.689 (0.891)
β_6 RL (d)	-0.060 (0.805)	-0.048 (0.895)	-0.057 (0.894)	-0.057 (0.894)
β_7 HT (o)	-0.297 (0.154)	-0.291 (0.180)	-0.291 (0.180)	-0.291 (0.178)
β_8 HT (d)	0.194 (0.156)	0.181 (0.180)	0.187 (0.180)	0.187 (0.180)
β_9 PMT (o)	0.891 (0.225)	0.897 (0.260)	0.901 (0.260)	0.901 (0.260)
β_{10} PMT (d)	0.886 (0.220)	0.889 (0.260)	0.890 (0.260)	0.890 (0.259)
β_{11} D	-1.129 (0.048)	-1.131 (0.057)	-1.135 (0.057)	-1.135 (0.057)
τ $(1/\sigma^2)$	0.989 (0.139)	0.906 (0.139)	0.906 (0.139)	0.906 (0.139)

†'o' refers to origin effects and 'd' to destination effects.

Table 2. *Posterior means and standard deviations (in parentheses) from an MH sample of 20000 draws and from the three INLA approaches for the PG model†*

Parameter	*MH estimate*	*INLA estimates*		
		Gaussian	*Simplified Laplace*	*Laplace*
β_0 intercept	-2.013 (2.488)	-2.147 (2.496)	-2.087 (2.496)	-2.052 (2.484)
β_1 ER (o)	1.081 (0.925)	1.039(0.918)	1.016(0.918)	1.058(0.913)
β_2 ER (d)	1.542(0.895)	1.486(0.901)	1.473(0.901)	1.515(0.897)
β_3 PD (o)	0.507 (0.428)	0.512(0.441)	0.526 (0.441)	0.513 (0.439)
β_4 PD (d)	1.282(0.440)	1.271 (0.434)	1.289(0.434)	1.272(0.432)
β_5 RL (o)	0.765 (0.785)	0.718(0.781)	0.757 (0.781)	0.763 (0.776)
β_6 RL (d)	0.078 (0.823)	-0.033 (0.810)	0.012(0.810)	0.022 (0.805)
β_7 HT (o)	-0.431 (0.180)	-0.434(0.177)	-0.436 (0.177)	-0.440 (0.176)
β_8 HT (d)	0.056(0.188)	0.077(0.171)	0.073 (0.171)	0.068 (0.170)
β_9 PMT (o)	1.082(0.246)	1.099(0.243)	1.087(0.243)	1.095(0.242)
β_{10} PMT (d)	1.163(0.232)	1.166(0.232)	1.156(0.232)	1.161 (0.231)
β_{11} D	-1.184(0.079)	-1.168(0.081)	-1.168(0.081)	-1.181 (0.082)
θ	1.070(0.123)	1.070(0.142)	1.070(0.142)	1.070(0.142)

†'o' refers to origin effects and 'd' to destination effects.

For comparison, we change the prior assumption for the intercept and the regression parameters. Specifically, instead of using the unit information prior $\beta \sim \mathbf{N}_{12}(\mathbf{0}, \Sigma_\beta)$ with $\Sigma_\beta = n\,(\mathbf{X}^T\mathbf{X})^{-1}$, we assume independent normal priors with a large variance, namely $\beta \sim \mathbf{N}_{12}\,(\mathbf{0}, \mathbf{I}_{12}\sigma^2)$ with $\sigma^2 = 10^3$. For fitting the PLN model through the INLA, we used the R-INLA package (www.r-inla.org). We consider the three INLA approximating strategies, namely the Gaussian, the simplified Laplace and the Laplace approximations for marginal posterior distributions; see Rue *et al.* (2009) for details. In addition, as mentioned in Section 2.1, recent developments extend the INLA to near Gaussian latent models (Martins and Rue, 2013). The gamma prior is included in the available options of the R-INLA package, which gives us the opportunity to compare MCMC and INLA approaches for the PG model as well.

Posterior means and standard deviations from MH samples of 20 000 draws and from the three INLA approximations for the PLN and PG models are presented in Table 1 and Table 2 respectively. The PLN intercept corresponds to that of the additive model formulation. As seen, the posterior PLN means from the MCMC and INLA approaches agree in general, especially under the simplified Laplace and Laplace approximations. The precision estimates slightly differ; the MCMC estimate is closer to the corresponding ML estimate which is 1.036. Also, the standard deviations from the MCMC method are overall lower. Concerning the PG model, the Laplace approximation seems to provide more accurate estimates which are closer to the MCMC estimates in comparison with the Gaussian and simplified Laplace approaches. The standard deviations are virtually the same for this model. Marginal posterior distribution estimates for the regressors of the PG model under the Laplace approximation are shown in Fig. 2; as seen the estimates approximate particularly well the histograms derived from MCMC sampling. The corresponding figures for the PLN model (which are not presented here) are equivalent.

In conclusion, we find that the INLA provides a fast and efficient alternative to MCMC sampling under specific prior assumptions, which makes it a potentially promising tool for OD modelling on large-scale networks. It is worth mentioning that INLA run times were 2.64 s for the PG model and 2.14 s for the PLN model. In contrast, 21000 MH iterations (we used the first 1000 as burn-in) required 14.82 s for the PG model and approximately 21 min for the PLN model owing to additional numerical integration within the MCMC algorithm. Thus, the INLA is particularly useful for the PLN model. Arguably, in medium-sized examples like this, using MCMC sampling for the hierarchical data-augmented PLN model could be more efficient than MCMC sampling with numerical integration for the marginal likelihood. We tested this by using WinBUGS (Spiegelhalter *et al.*, 2003); nevertheless, the sampler failed to converge even after 200 000 iterations. All computations were performed on a standard 64-bit laptop with 2.20 GHz central processor unit and 4 Gbytes of random-access memory using R version 3.0.1 (R Core Team, 2013). The R code that was used for the INLA and the subset of the OD data are available from http://wileyonlibrary.com/journal/rss-datasets

Despite these advantages, we find that the R-INLA package is still restrictive with respect to certain aspects and requires further development which will allow for more general modelling frameworks. In particular, the package does not yet fully support multivariate prior assumptions such as g-prior structures (Zellner, 1986) for regression coefficients. Moreover, it would be interesting to include further distributional options covering near Gaussian latent fields, such as the IG prior that is considered in this paper.

Fig. 2. *Histograms of PG regression parameters from 20000 posterior draws and estimates of the posterior marginals (———) from the INLA method*

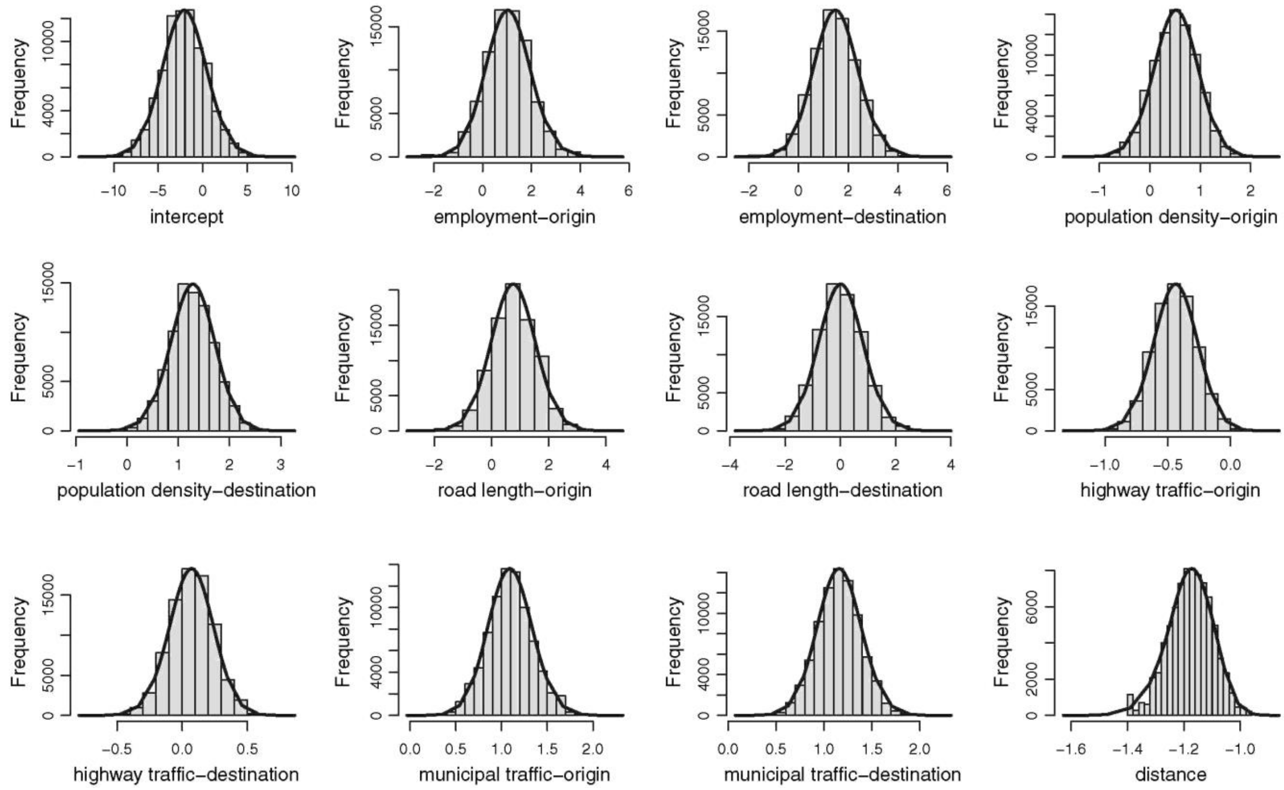


Table 3. *Posterior means and 95% credible intervals for regression and dispersion parameters and the values of AIC, BIC, the marginal DIC and the hierarchical DIC†*

Parameter	*Results for PG model*		*Results for PLN model*		*Results for PIG model*	
	Mean	*95% credible interval*	*Mean*	*95% credible interval*	*Mean*	*95% credible interval*
β_0	4.027	(3.214, 4.869)	6.104	(5.230, 6.989)	6.847	(6.028, 7.675)
β_1 DP	0.005	(0.005, 0.005)	0.005	(0.005, 0.006)	0.006	(0.005, 0.006)
β_2 DA	0.007	(0.007, 0.008)	0.007	(0.007, 0.008)	0.008	(0.007, 0.008)
β_3 DD	0.008	(0.008, 0.009)	0.008	(0.008, 0.009)	0.009	(0.008, 0.009)
β_4 DC	0.008	(0.007, 0.009)	0.006	(0.006, 0.007)	0.006	(0.006, 0.007)
β_5 DM	-0.082	(-0.083, -0.080)	-0.086	(-0.087,-0.084)	-0.084	(-0.086, -0.083)
β_6 DE	0.424	(0.388, 0.460)	0.535	(0.496, 0.574)	0.536	(0.497, 0.581)
β_7 MC	0.473	(0.435, 0.510)	0.461	(0.422, 0.501)	0.450	(0.411, 0.490)
β_8 MD	-0.494	(-0.542, -0.445)	-0.441	(-0.491,-0.391)	-0.442	(-0.489, -0.392)
β_9 MA	-0.088	(-0.124, -0.055)	-0.188	(-0.225,-0.149)	-0.210	(-0.249, -0.167)
β_{10} MP	-0.491	(-0.636, -0.345)	-0.737	(-0.888, -0.589)	-0.783	(-0.924, -0.636)
β_{11} ER(o)	-1.062	(-1.207, -0.918)	-0.482	(-0.629, -0.334)	-0.240	(-0.383, -0.105)
β_{12} ER (d)	0.326	(0.194, 0.462)	0.492	(0.345, 0.641)	0.608	(0.460, 0.759)
β_{13} PD (o)	0.505	(0.477, 0.533)	0.499	(0.470, 0.528)	0.500	(0.474, 0.529)
β_{14} PD (d)	0.577	(0.548, 0.606)	0.626	(0.594, 0.658)	0.631	(0.595, 0.662)
β_{15} RL(o)	-0.315	(-0.359, -0.272)	-0.318	(-0.365, -0.272)	-0.334	(-0.380, -0.289)
β_{16} RL(d)	0.280	(0.236, 0.322)	0.267	(0.220, 0.315)	0.265	(0.219, 0.310)
β_{17} PL (o)	1.253	(1.208, 1.298)	1.289	(1.241, 1.338)	1.283	(1.238, 1.327)
β_{18} PL (d)	0.430	(0.385, 0.475)	0.500	(0.452, 0.549)	0.509	(0.459, 0.559)
β_{19} CR (o)	3.454	(3.149, 3.762)	3.413	(3.095, 3.731)	3.520	(3.227, 3.831)
β_{20} CR (d)	-1.465	(-1.768, -1.180)	-1.255	(-1.577, -0.939)	-1.081	(-1.386, -0.752)
β_{21} HT (o)	0.010	(0.007, 0.014)	0.011	(0.008, 0.015)	0.010	(0.007, 0.014)
β_{22} HT (d)	0.052	(0.049, 0.056)	0.050	(0.047, 0.054)	0.050	(0.046, 0.053)
β_{23} PMT (o)	0.270	(0.250, 0.289)	0.278	(0.257, 0.299)	0.275	(0.254, 0.294)
β_{24} PMT (d)	0.869	(0.850, 0.888)	0.876	(0.853, 0.898)	0.870	(0.849, 0.891)
β_{25} *D*	-2.906	(-2.927, -2.885)	-2.984	(-3.007, -2.960)	-2.936	(-2.957, -2.915)
θ	0.965	(0.947, 0.983)		—		—
σ^2		—	1.065	(1.043, 1.086)		—
ζ		—		—	0.377	(0.359, 0.399)
AIC		281519.3		279364.7		278468.9
BIC		281774.7		279620.1		278724.3
DIC (marginal)		281492.4		279337.7		278441.4
DIC (hierarchical)		224141.4		—		224146.1

†'o' refers to origin effects and 'd' to destination effects.

4.2. Posterior inference for Flanders

Posterior means and 95% credible intervals based on 4000 posterior draws are presented in Table 3. The corresponding INLA Gaussian estimates for the PG and PLN models are presented in Table 4. The INLA estimates are slightly different because of the different prior assumption, namely $\beta \sim \mathbf{N}_{26}(\mathbf{0}, \mathbf{I}_{26}\sigma^2)$ with $\sigma^2 = 10^3$. Nevertheless, the overall conclusions discussed next are also supported by the INLA estimates. Details of MH implementation and a comparison of MCMC and INLA run times can be found in Appendix A. In general, the posterior means of the PLN and PIGmodels are more similar. For instance, parameters β_0, β_6, β_9, β_{10}, β_{14} and β_{18} of the PG model are substantially different from the corresponding estimates of the other two models, especially the intercept estimate. However, parameters β_{11}, β_{12} and β_{20} differ across models.

The parameters β_1-β_5 of the categorical variables are all positive except for the last parameter for intrazonal municipality trips. The positive effects of β_1-β_4 are to be expected, since the OD flows are generally larger in diagonal blocks of cells of the OD matrix corresponding to intrazonal flows for the various administrative levels. The negative sign of β_5 is not expected but it might be explained as simply counterbalancing the absence of the strong negative effect of distance which is set almost equal to 0 for intrazonal municipality trips. Parameter β_6 is positive, which leads to the consistent interpretation that destination zones which support a college or a university are more likely to attract trips than zones without a college or university.

Table 4. *INLA estimates for the PG and PLN models using the Gaussian approximation for the entire data set†*

Parameter	*Results for PG model*		*Results for PLN model*	
	Mean	*95% credible interval*	*Mean*	*95% credible interval*
β_0	3.605	(2.903, 4.307)	5.974	(5.106, 6.847)
β_1 DP	0.005	(0.005, 0.005)	0.005	(0.005, 0.005)
β_2 DA	0.007	(0.006, 0.007)	0.007	(0.006, 0.007)
β_3 DD	0.008	(0.007, 0.008)	0.008	(0.008, 0.009)
β_4 DC	0.008	(0.007, 0.008)	0.007	(0.006, 0.008)
β_5 DM	-0.082	(-0.084, -0.081)	-0.078	(-0.080, -0.076)
β_6 DE	0.435	(0.404, 0.466)	0.509	(0.471, 0.547)
β_7 MC	0.459	(0.427, 0.491)	0.431	(0.392, 0.470)
β_8 MD	-0.482	(-0.523, -0.442)	-0.424	(-0.472, -0.376)
β_9 MA	-0.079	(-0.109, -0.048)	-0.177	(-0.214, -0.139)
β_{10} MP	-0.447	(-0.569, -0.324)	-0.699	(-0.848, -0.551)
β_{11} ER(o)	-1.037	(-1.163, -0.911)	-0.458	(-0.603, -0.313)
β_{12} ER (d)	0.310	(0.192, 0.428)	0.449	(0.302, 0.596)
β_{13} PD (o)	0.495	(0.471, 0.519)	0.469	(0.440, 0.498)
β_{14} PD (d)	0.558	(0.533, 0.583)	0.591	(0.560, 0.622)
β_{15} RL(o)	-0.315	(-0.353, -0.278)	-0.299	(-0.344, -0.254)
β_{16} RL(d)	0.287	(0.250, 0.324)	0.253	(0.208, 0.298)
β_{17} PL (o)	1.272	(1.232, 1.312)	1.212	(1.165, 1.259)
β_{18} PL (d)	0.442	(0.402, 0.482)	0.479	(0.429, 0.528)
β_{19} CR (o)	3.401	(3.140, 3.663)	3.183	(2.871, 3.495)
β_{20} CR (d)	-1.549	(-1.813, -1.286)	-1.177	(-1.496, -0.859)
β_{21} HT (o)	0.010	(0.007, 0.012)	0.010	(0.007, 0.014)
β_{22} HT (d)	0.051	(0.048, 0.054)	0.047	(0.043, 0.050)
β_{23} PMT (o)	0.264	(0.246, 0.281)	0.263	(0.242, 0.283)
β_{24} PMT (d)	0.867	(0.850, 0.884)	0.824	(0.802, 0.846)
β_{25} *D*	-2.912	(-2.930, -2.893)	-2.807	(-2.830, -2.785)
θ	0.969	(0.950, 0.986)		—
σ^2		—	1.020	(1.000, 1.050)

†'o' refers to origin effects and 'd' to destination effects.

Parameters β_7-β_{10} quantify the influence of the total number of surrounding municipalities on the levels of cantons, districts, *arrondissements* and provinces respectively. This effect is in general not straightforward to predict; nevertheless the parameter estimates provide some insights. On the small-scale level of cantons parameter β_7 has a positive sign, whereas, on the large-scale levels of districts, *arrondissements* and provinces—where the total number of municipalities increases and a spead-out of trips is more likely—the corresponding parameters β_8, β_9 and β_{10} are negative. This implies that the effect changes from positive to negative when exceeding a specific radius threshold of distance. Recent transportation studies discuss similar ideas such as the *neighbourhood effect* concept that was investigated in more detail by Sohn and Kim (2010).

Regarding the continuous variables used in pairs, the more general explanatory variables have parameters with positive signs, namely population density (β_{13},β_{14}), perimeter length (β_{17},β_{18}) and kilometres driven on highways (β_{21},β_{22}) and provincial or municipal roads (β_{23},β_{24}). The uniformly positive effects for origin and destination zones do not come as a surprise, since we would expect these four variables to be positively correlated with trip production (origin zones) as well as trip attraction (destination zones). In contrast, the parameters of employment rate (β_{11},β_{12}), relative length of road network (β_{15},β_{16}) and car ownership ratio (β_{19},β_{20}) have opposite signs for origin and destination effects.

In transportation studies employment rate is commonly associated with trip attraction models (see for example Yao and Morikawa (2005)). In accordance, the posterior estimate of employment rate is positive for destination zones and negative for origin zones, which leads to the rational interpretation that zones with high employment rates are more likely to attract trips rather than to generate trips. The relative length of road networks is associated with the concept of *accessibility* (see for example Odoki *et al.* (2001)), a concept which is present primarily in trip attraction studies. In general, a larger relative length in the network will decrease the friction of travel (e.g. distance and time) significantly, and thus increase accessibility. The posterior mean is positive for destination zones and negative for origin zones. Consistently, this implies that zones with high levels of accessibility are more likely to attract trips than low accessible zones. Conversely, high accessible zones are less likely to produce trips than low accessible zones. A possible explanation for the negative origin effect is that high levels of accessibility within a zone might encourage intrazonal trips and reduce outgoing trips. Car ownership is traditionally used as an explanatory variable with positive effect in trip production models. In agreement, the posterior mean for car ownership is positive for origin zones, which means that zones with high car ownership ratios also have high trip production rates. The estimate is negative for destination zones, implying that high car ownership ratios are negatively correlated with trip attraction. The negative destination effect may be attributed to congestion issues.

Distance with parameter β_{25} is the final variable. Distance is a key variable in gravity-type and direct demand models, since it is directly related to the costs of the deterrence function that is used within the trip distribution step. In our model, distance has a negative posterior mean which accords with the basic deterrent gravitational assumption of trip distribution models. Furthermore, on the basis of the posterior mean over standard deviation ratio, distance is the most significant explanatory variable in all the models.

Table 3 also includes the values of the Akaike information criterion AIC, Bayesian information criterion BIC and marginal or hierarchical deviance information criterion DIC. The posterior mean of the deviance is used for the calculation of AIC and BIC. The three criteria provide more support for the PLN and PIG models, which provides a justification for the similarity of the posterior estimates from the two models. Furthermore, all three criteria indicate that the PIG distribution is the most appropriate marginal sampling distribution. The hierarchical DIC is calculated on the basis of reduced samples of 500 draws, owing to memory limitations given the large dimensionality of the data-augmented space. In addition, sampling the random effects of the PLN model is relatively complicated and time consuming; therefore we focus on the PG and PIG models for the remainder of this paper. On the basis of the hierarchical DIC it is difficult to distinguish which hierarchical model is more appropriate for predictive purposes, since the differences between the PG and PIG models are marginal.

The random effects present some dissimilarities between the two models. The range of the PG random effects is from 3.81×10^{-8} to 40.61 (from -17.83 to 3.71 on a log-scale), whereas the PIG random effects range from 9.48 $\times 10^{-3}$ to 132.35 (from -4.66 to 4.89 on a log-scale). Because of the GIG posterior distribution, the PIG random effects exhibit a longer right-hand tail than the PG random effects which are gamma distributed. On a logarithmic scale the PIG random effects are relatively more symmetrical near 0, whereas the PG random effects have a longer left-hand tail.

Fig. 3. *Histograms of dispersion parameters (a)-(c) θ and (d)-(f) ζ under gamma hyperpriors with (a), (d) a = 0.001, (b), (e) a = 0.1 and (c), (f) a= 1*

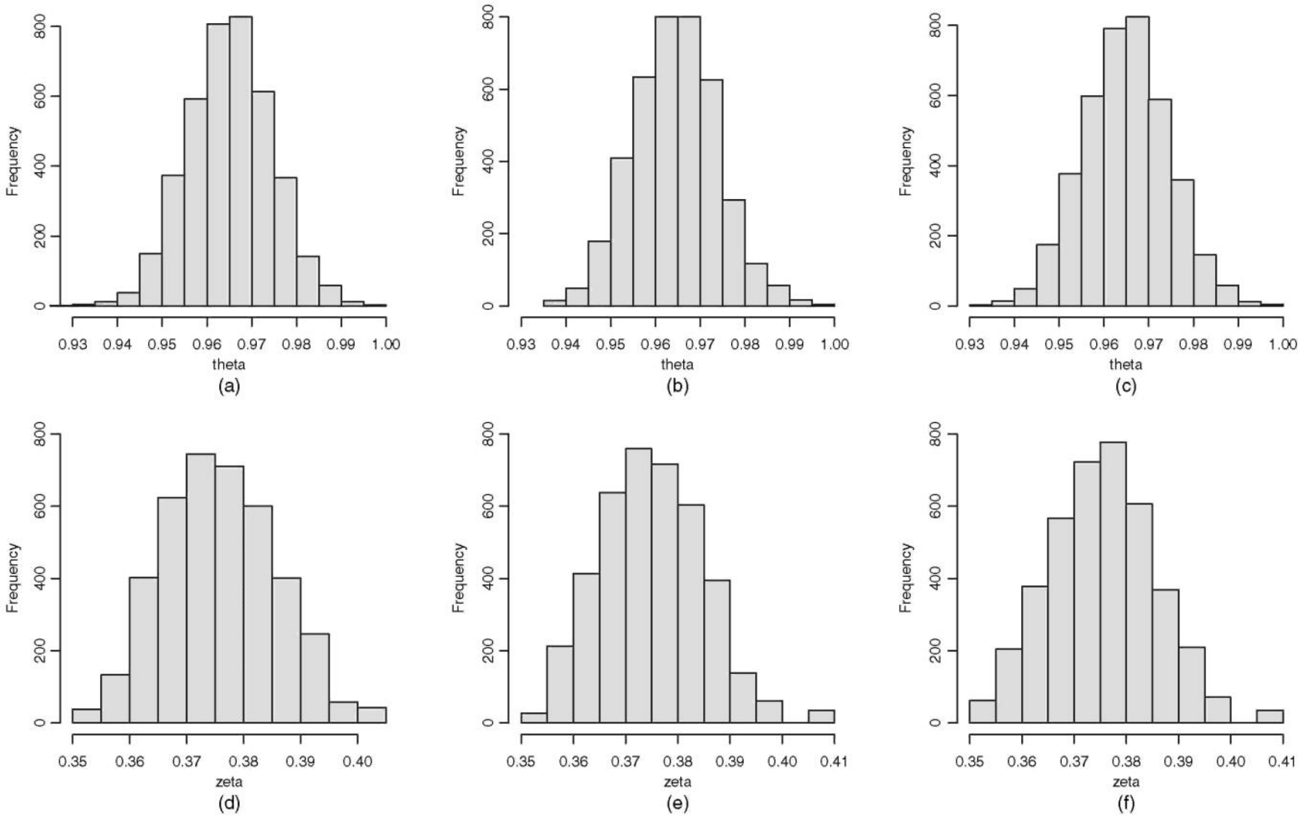


Table 5. *Bayesian p-values for the absolute distance, squared distance and deviance test quantities from 500 posterior draws of the hierarchical PG and PIG models*

Test quantity	*Formula*	*Results for PG model*	*Results for PIG model*
Absolute distance	$\Sigma\{\mathbf{y} - E(\mathbf{y}\|\beta, \mathbf{u})\}$	0.278	0.440
Squared distance	$\Sigma\{\mathbf{y} - E(\mathbf{y}\|\beta, \mathbf{u})\}^2$	0.532	0.488
Deviance	$-2 \log\{p(\mathbf{y}\|\beta, \mathbf{u})\}$	0.996	0.648

4.3. Sensitivity analysis for hyperpriors

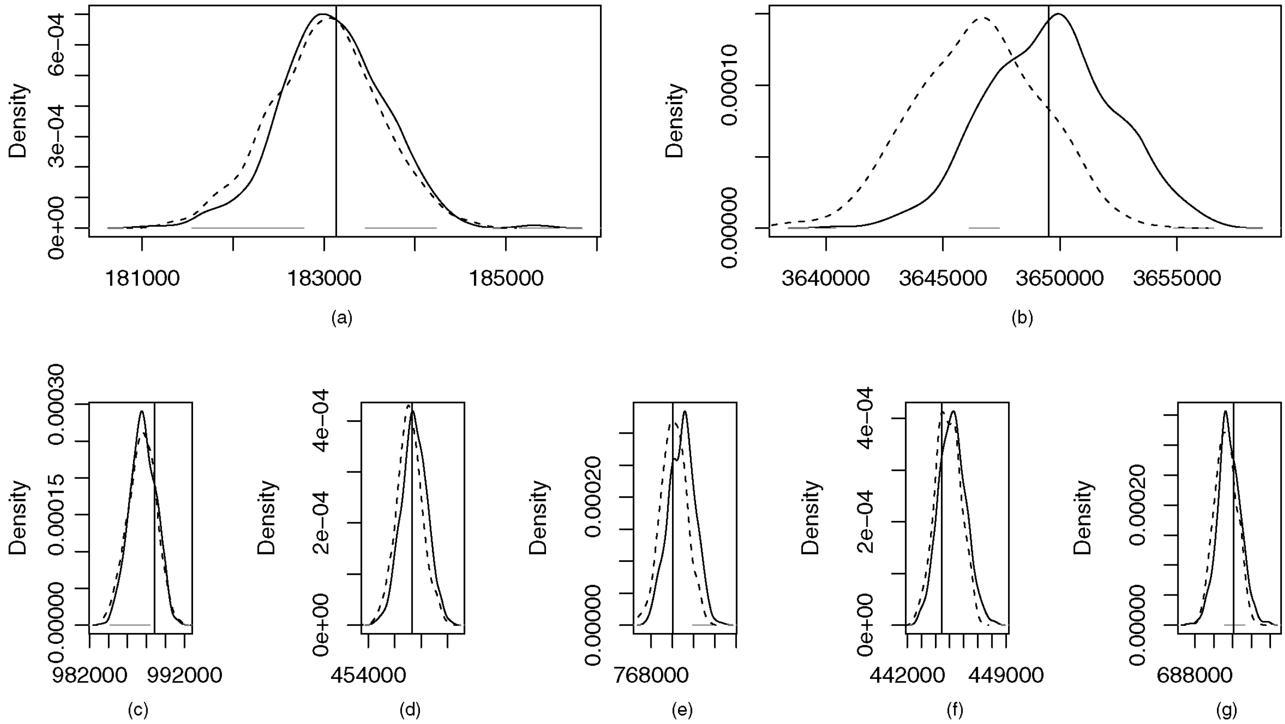
In this section we perform a sensitivity analysis for parameter a of the gamma hyperpriors assigned to parameters θ and ζ of the PG and PIG models. Parameter σ^2 is not included in the analysis, because of the substantial time which is required for MH simulation from the PLN model. Histograms of 4000 posterior draws of parameters θ and ζ for values of a equal to 0.001 (the initial value), 0.1 and 1 are presented in Fig. 3. As seen, the posterior distributions are not influenced by hyperparameter a. We can note a slight change in the right-hand tail of the posterior distribution of ζ for a equal to 0.1 and 1; nonetheless this does not affect posterior inferences. The results are in line with the discussion in Gelman (2006), since in our case the random effects are observational and therefore we would not expect to have the sensitivity problems that arise in grouped random-effects settings for small numbers of groups.

4.4. Posterior predictive checks for origin-destination flows

For overall goodness of fit, we employ posterior predictive checks (Meng, 1994) for the absolute and squared distances (with respect to the expected values) and for the hierarchical deviance. The absolute distance is more sensitive to small deviations, whereas squared distance assigns more penalty to large deviations. Each test quantity is calculated for observed and predicted data over the 500 posterior draws.

The test quantities with the corresponding Bayesian p-values are presented in Table 5. In general, both models provide satisfactory Bayesian p-values for squared distances, which are close to the ideal value of 0.5. Predictions from the PIG model seem to replicate better the observed data for small deviations from the expected values and also with respect to the Poisson distributional assumption. Note that the aim here is not model comparison, but examination of the characteristics of predictions.

***Fig.** 4. Observed quantities (|) and kernel estimates of predictive distributions for going-to-work or school trips from the PG model (———) and the PIG model (- - -) for (a) incoming trips to the city of Antwerp (PG p-value 0.46; PIG p-value 0.43), (b) all trips to Flanders (PG p-value 0.52; PIG p-value 0.14), and intrazonal trips for the five Flemish provinces (c) Antwerp (PG p-value 0.21; PIG p-value 0.18), (d) Limburg (PG p-value 0.57; PIG p-value 0.41), (e) East Flanders (PG p-value 0.74; PIG p-value 0.52), (f) Flemish Brabant (PG p-value 0.76; PIG p-value 0.66) and (g) West Flanders (PG p-value 0.33; PIG p-value 0.22)*



An interesting feature of OD modelling is that the administrative structure allows for various aggregations of observed and replicated data with respect to administrative levels and also types of trip. From a statistical perspective, the aggregated distributions can be compared with the observed aggregated values, thus resulting in Bayesian *p*-values for case-specific tests. Examples of such tests for incoming trips to the municipality of Antwerp, total trips for Flanders and intrazonal trips for the five Flemish provinces are presented in Fig. 4. In general, all *p*-values are within acceptable limits. From a transportation planning perspective, such predictions are particularly useful for policy evaluation.

4.5. Predictive inference for link flows

For traffic assignment we utilize the deterministic user equilibrium (DUE) model which is based on Wardrop's first principle (Wardrop, 1952), which is also known as the *equilibrium principle.* In short, DUE assignment uses an iterative process to reach a convergent solution in which travellers cannot reduce their travel times by switching routes. At each iteration link capacity restraints and link-flow-dependent travel times are taken into account to calculate link flows. As *link performance function* we adopt the common Bureau of Public Roads formulation (Bureau of Public Roads, 1964) which relates link travel times to volume-over-capacity, *VIC,* ratios, specifically $t = t_f\{1 + \alpha(\upsilon/c)^{\beta}\}$, where t is the link travel time, t_f is link free-flow travel time, υ is link volume (flow), c is link capacity and α and β are calibration parameters which are set equal to their historical values of 0.15 and 4 respectively. If we denote by **A** the DUE assignment operator, we execute 500 individual assignments from the predictive ODs of each model and obtain 500 corresponding link flow or link volume vectors **v**, i.e. $\mathbf{A}\mathbf{y}^{\text{pred}(m)} = \mathbf{v}^{(m)}$ for $m = 1, 2, ..., 500$, where $\mathbf{v}^{(m)} = (\upsilon_1^{(m)}, \upsilon_2^{(m)}, ..., \upsilon_l^{(m)})^{\mathrm{T}}$ and l is the total number of network links. For the Flemish network l is equal to 97450. The assignments concern the morning peak hour interval between 7 a.m. and 8 a.m. for a normal weekday.

The mean state of the Flemish network under DUE assignment and OD predictions from the PIG model is presented in Fig. 5. By 'mean state' it is meant that the 500 link flow vectors were averaged first and then visualized. To make Fig. 5 simpler to comprehend only volumes and *V*/*C*-ratios for highway links are highlighted. The main findings are as follows. Higher *V*/*C*-ratios (between 0.5 and 0.75) are observed in specific segments on or near the highways rings of Antwerp (R1) and Ghent (R4). Relatively high *V*/*C*-ratios (between 0.25 and 0.5) also occur on the northern part of highway ring R0 around Brussels, on highway E40 near Leuven, highway E313 which connects Antwerp with Hasselt and to a lesser degree on highways E17 and E19 which connect Antwerp with Ghent and Brussels respectively. The corresponding visualization map based on PG predictions is not presented as it is almost identical to Fig. 5, with differences being difficult to spot on a global scale.

An interesting application is the identification of congested links on the network. Congestion identification is related to *critical link* identification, which is customarily a subject of *vulnerability analysis* and relies significantly on traffic assignment procedures (see for example Jenelius *et al.* (2006)). Through our approach congested links are evaluated directly in terms of probability estimates. As congested links we define those links on which the *V*/*C*-ratio exceeds a certain threshold value t with a certain probability $P(V/C > t)$. As a conservative choice and in order not to overestimate the number of critical links a threshold value of 0.95 is adopted, based on the assumption that the majority of trips taking place between 7 a.m. and 8 a.m. are either work- or school-related trips. For t = 0.95 congestion is identified in 11 links which all belong to large Flemish municipalities; five in Antwerp, five in Ghent and one link in Bruges. The *V*/*C*-distributions from both models are presented in Fig. 6.

Certain remarks can be made, based on Fig. 6, regarding *V*/*C*-distributions and consequently link flow distributions from DUE assignment. First, the choice of statistical model does not seem to affect individual *V*/*C*-distributions as the corresponding distributions are very similar. Second, individual *V*/*C*- and link distributions are not necessarily close to normal distributions— for instance bimodalities are observed—in contrast with aggregated distributions (e.g. link flows for highways—not presented here) which converge to normality in accordance with the central limit theorem. Third, the bimodalities may be attributed to the iterative user equilibrium procedure; when the flows on a specific link and at a given iteration exceed a certain threshold—leading to a high *V*/*C*-ratio—and there is an alternative link with a marginally lower cost, 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 in bimodal distributions like those observed in Fig. 6.

Seven out of the 11 links have a *V*/*C*-value greater than 0.95 with probability 1. Visual examination of the distributions in Fig. 6 additionally reveals that these seven links also exceed the value t = 1 with probability 1, except perhaps link 106252 which has its minimum near 1 and may therefore include smaller values than 1 with a low probability. The remaining four links have lower *V*/*C*-ratios and exceed the value 0.95 with a probability lower than 1. The expected values and the corresponding probabilities for the 11 congested links are presented in Table 6.

Assignment with PG predictions results in slightly higher probabilities for links 22149, 29060, 83662 and 92846. We also note that if the analysis was based on the expected values congestion would not have been identified on those four links.

Fig. 5. *Mean visualization of highway flows and V/C-ratios for going-to-work or school trips between 7 a.m. and 8 a.m. in Flanders under DUE assignment and PIG OD predictions*

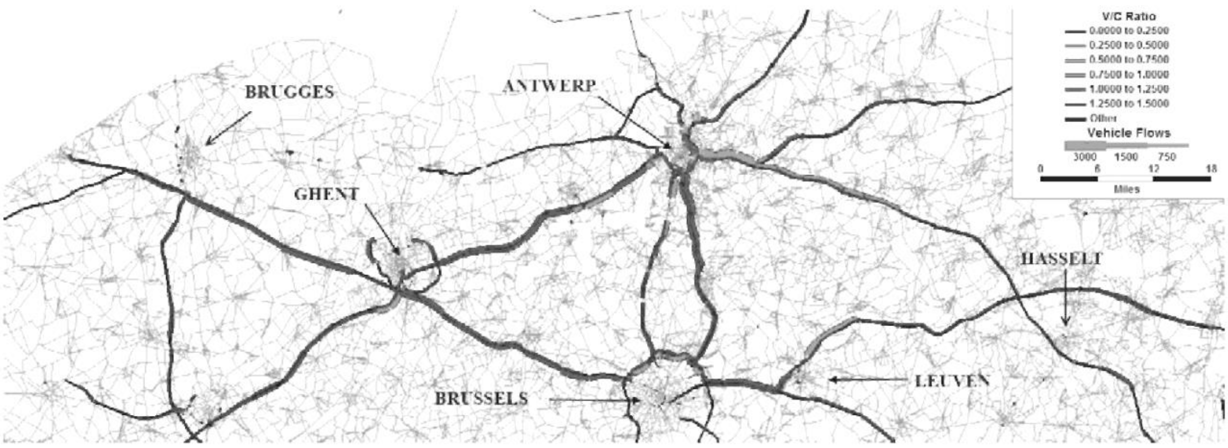


Fig. 6. *Kernel estimates of the PG (———) and PIG (- - -) V/C-distributions of the 11 congested links which either include or exceed the threshold value of 0.95 (|) in the distributions which include this value: (a) link 18641, small regional roads, Antwerp; (b) link 17493, local roads, Antwerp; (c) link 22149, local roads, Antwerp; (d) link 28980, highways, Antwerp; (e) link 29060, local roads, Antwerp; (f) link 83662, local roads, Ghent; (g) link 83928, highways, Ghent; (h) link 84514, main regional roads, Ghent; (i) link 92846, local roads, Ghent; (j) link 92849, local roads Ghent; (k) link 106252, local roads, Bruges*

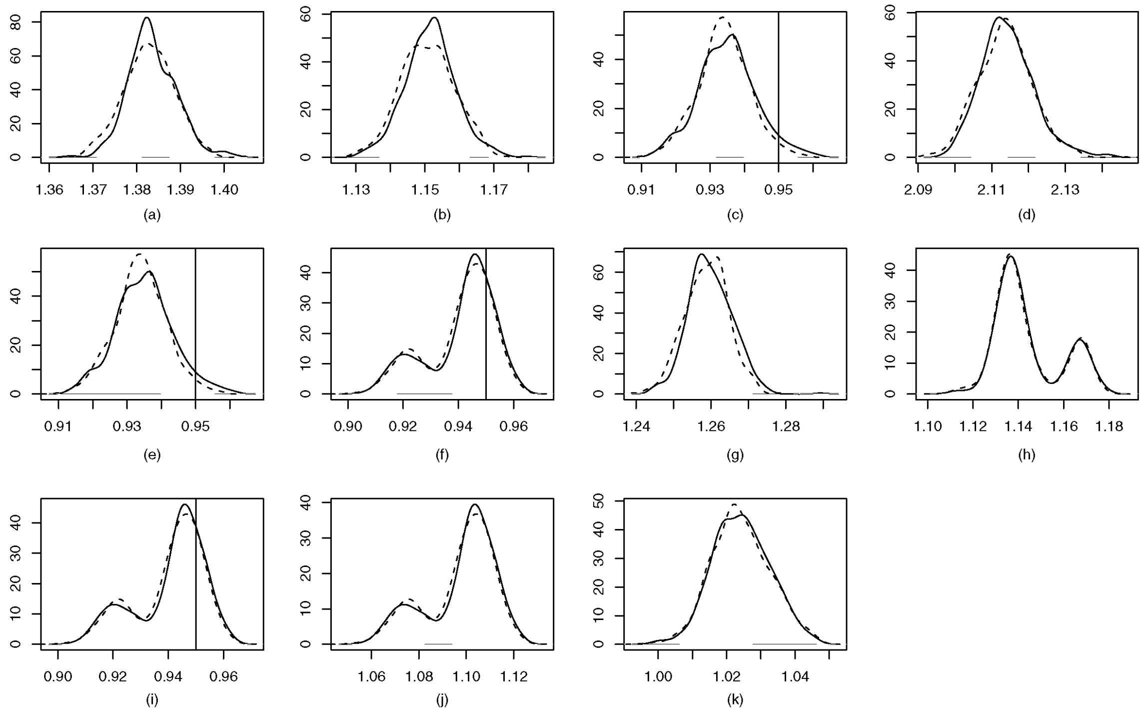


Table 6. *Expected V/C-ratios and probabilities of exceeding a V/C of 0.95 for the 11 congested links under DUE assignment and PG and PIG predictions*

Congested link	*Link type*	*Results for PG model*		*Results for PIG model*	
		E (V/C)	*P (V/C > 0.95)*	*E (V/C)*	*P (V/C > 0.95)*
16841	Small regional road	1.384	1	1.383	1
17493	Local road	1.152	1	1.151	1
22149	Local road	0.935	0.046	0.934	0.022
28980	Highway	2.114	1	2.114	1
29060	Local road	0.935	0.046	0.934	0.022
83662	Local road	0.941	0.236	0.941	0.208
83928	Highway	1.260	1	1.259	1
84514	Main regional road	1.144	1	1.144	1
92846	Local road	0.941	0.236	0.941	0.208
92849	Local road	1.098	1	1.097	1
106252	Local road	1.024	1	1.024	1

5. Discussion

In this paper we investigated the use of Poisson mixtures in OD modelling as a viable alternative to traditional transportation models. The advantages of the approach proposed are that (*i*) it incorporates the steps of trip generation and trip distribution in statistical models which provide a wider inferential scope and (*ii*) it allows for probabilistic inference on link traffic and congestion, conditional on the assignment model.

At the same time, the approach may be viewed as a statistical, direct demand, gravity model, thus retaining a strong relationship with traditional transportation models.

The case-study focused on a large, sparsely distributed and overdispersed OD matrix derived from the 2001 Belgian travel census covering the region of Flanders. In particular, we considered the PG, PLN and PIG models as alternative modelling options. The PIG model—a model that is not as popular as its competing alternatives—

provided the best marginal fit and resulted in consistent short-term predictions. Given the convenient distributional properties of the PIG model, we recommend its use when analysing large-scale OD matrices. In addition, we investigated the performance of INLA compared with MCMC methods for the PG and PLN models and found that the INLA approach can provide fast and accurate approximations. From this point of view, the INLA is particularly suited for the PLN, which proved to be the most cumbersome model to work with by using MCMC sampling. Further development of the R-INLA package, in terms of prior extensions, will make it a useful tool for large-scale OD analysis.

Future research directions concerning transportation issues are many. First, the set of covariates that was used in this study is by no means conclusive. As pointed out by one referee the models could improve in terms of capturing representation of activities in destination zones. This can be achieved by including the number of workplaces and shopping facilities as predictors. This type of information was not available and could not be included in the current analysis. Second, the issue of modal split which was not pursued here can be potentially incorporated in the modelling approach proposed. A third issue concerns dynamic modelling of short-term OD matrices, e.g. analysis of OD matrices on hourly intervals. A fourth category of issues is related to a series of traffic assignment comparative studies between the DUE model, which was utilized here, and other assignment models such as the stochastic user equilibrium model, conditional on Bayesian predictions.

From a statistical perspective, discrete random effects could have been considered as an alternative approach for clustering purposes. This approach was not pursued here as the focus of this study was on modelling and capturing the heterogeneity per OD pair. Finally, the PIG model can be of potential value to any other count data analysis problem under the presence of overdispersion. From this point of view, it will be interesting to consider zero-inflated model extensions and also to compare with other mixing or prior distributional designs.

Acknowledgements

The authors thank the Associate Editor and two referees for their useful comments and suggestions. Moreover, we thank the INLA support team, particularly, Professor Håvard Rue for his suggestions and technical help concerning implementation of the INLA.

Appendix A: Metropolis-Hastings simulation

We utilize MH simulation on the marginal structures to bypass sampling 94 864 random effects at each MCMC iteration. Although sampling u in a Gibbs-like fashion is straightforward for the hierarchical PG and PIG models, memory limitations would require discarding u at the end of each iteration. MH sampling for the marginal PG and PIG structures is far more efficient with β, θ and ζ being easy to sample, whereas u can be generated subsequently as described in Sections 2.1 and 2.3. The PLN model is more problematic since an additional Metropolis step or rejection sampling is required for the hierarchical structure, which is an obvious burden for 94 864 random effects. In contrast, simulation for the marginal PLN structure requires numerical or MC integration within the MCMC algorithm and—in addition—vector u is not easy to sample subsequently.

In particular, we employ an independence chain MH algorithm where the location and scale of the proposals are fixed (see for example Chib and Greenberg (1995)) to the corresponding ML estimates. For regression vector β a multivariate normal proposal is used, i.e. $q(\beta) = \mathbf{N}_{p+1}(\beta^{\text{ML}}, \mathbf{V}^{\text{ML}})$ with β^{ML} being the ML estimate of β and $\mathbf{V}^{\text{ML}}$ the estimated variance-covariance matrix of β^{ML} for each model. For the dispersion parameters θ, σ^2 and ζ we used the following gamma proposals: $q(\theta)$ = gamma(a_{PG}, b_{PG}), $q(\sigma^2)$ = gamma(a_{PLN}, $b_{\text{PNLN}}$and $q(\zeta)$ = gamma(a_{PIG}, b_{PIG}) with proposal parameters set to satisfy the conditions $a_{\text{PG}}/b_{\text{PG}} = \theta^{\text{ML}}$, $a_{\text{PG}}/b_{\text{PG}}^2 = \text{var}(\theta^{\text{ML}})$, $a_{\text{PLN}}/b_{\text{PLN}} = \sigma^{2\text{ML}}$, $a_{\text{PLN}}/b_{\text{PLN}}^2 = \text{var}(\sigma^{2\text{ML}})$, $a_{\text{PIG}}/b_{\text{PIG}} = \zeta^{\text{ML}}$ and $a_{\text{PIG}}/b_{\text{PIG}}^2 = \text{var}(\zeta^{\text{ML}})$. Regarding probability calculations from the PLN distribution we implemented both numerical and MC integration. Results showed that the MC sample L should be preferably 2000 to obtain stable estimates, similar to the estimates from numerical integration, and numerical integration was already twice as fast as MC integration with a sample of 200. Therefore, numerical integration was preferred.

We utilized five independent MH chains of size 4200 and discarded the first 200 iterations as burn-in, resulting in posterior samples of 20 000 draws. The 10th, 30th, 50th, 70th and 90th percentile points of the proposal distributions were used as starting values. The resulting acceptance ratios were 72% for the PG model, 67% for the PLN model and 33% for the PIG model, on average. The multichain diagnostics of Gelman and Rubin (1992) and Brooks and Gelman (1998) were used to asses convergence. All univariate potential scale reduction factors were very close to 1 for all three models. The multivariate potential scale reduction factor for the PG, PLN and PIG models were 1.01, 1.01 and 1.06 respectively. Finally, to reduce the computational burden of the subsequent analysis the posterior samples were thinned by an interval of 5, resulting in final posterior samples of 4000 draws.

Implementation of MCMC sampling was done in R version 2.8.2 on a 64-bit Windows server 2003 R2 with 32 Gbytes of random-access memory. The simulations for the PG and PIG models required approximately 1 and 2.4 h respectively, whereas the PLN model required 3.6 days owing to numerical integration.

The INLA models based on the prior assumption $\beta \sim \mathbf{N}_{26}(\mathbf{0}, \mathbf{I}_{26}\sigma^2)$, with $\sigma^2 = 10^3$, were fitted remotely in R version 3.0.1 through the Linux server maintained by the INLA support team. The PG model required approximately 2.2 h and the PLN model about 2.5 h. The Gaussian approximation was used for both models.

References

Aitkin, A. (1996) A general maximum likelihood analysis of overdispersion in generalized linear models. *Statist. Comput.,* 6, 251-262.
Athreya, K. B. (1986) Another conjugate family for the normal distribution. *Statist. Probab. Lett.,* 4, 61-64.
Boucher, J.-P. and Denuit, M. (2006) Fixed versus random effects in Poisson regression models for claim counts: a case study with motor insurance. *Astin Bull.,* 1, 285-301.
Breslow, N. E. (1984) Extra-Poisson variation in log-linear models. *Appl. Statist.,* 33, 38-44.
Brooks, S. P. and Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *J. Computnl Graph. Statist.*, 7, 434-455.
Bureau of Public Roads (1964) *Traffic Assignment Manual for Application with a Large, High Speed Computer.* Washington: US Department of Commerce Bureau of Public Roads.
Carlson, M. (2002) Assessing microdata disclosure risk using the Poisson-inverse Gaussian distribution. *Statist. Transn,* 5, 901-925.
Chen, J. J. and Ahn, H. (1996) Fitting mixed Poisson regression models using quasi-likelihood methods. *Biometr. J.,* 38, 81-96.
Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *Am. Statistn,* 49, 327-335. Dagpunar, J. (1988) *Principles of Random Variate Generation.* New York: Oxford University Press.
Dean, C., Lawless, J. F. and Willmot, G. E. (1989) A mixed Poisson-inverse-Gaussian regression model. *Can. J. Statist.*, 17, 171-181.
Fernandez, C., Ley, E. and Steel, M. F J. (2001) Benchmark priors for Bayesian model averaging. *J. Econmetr,* 100, 381-427.
Font, M., Puig, X. and Ginebra, J. (2013) A Bayesian analysis of frequency count data. *J. Statist. Computn Simuln,* 83, 229-246.
Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Baysn Anal,* 1, 515-533.
Gelman, A. and Hill, J. (2006) *Data Analysis using Regression and Multilevel/Hierarchical Models,* 1 st edn. New York: Cambridge University Press.
Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.,* 7, 457-511.
Hazelton, M. L. (2010) Bayesian inference for network-based models with a linear inverse structure. *Transprtn Res.* B, 44, 674-685.
Holla, M. S. (1967) On a Poisson-inverse Gaussian distribution. *Metrika,* 11, 115-121.
Jenelius, E., Petersen, T. and Mattsson, L. G. (2006) Importance and exposure in road network vulnerability analysis. *Transprtn Res.* A, 40, 537-560.
Jorgensen, B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution.* New York: Springer.
Karlis, D. (2001) A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statist. Modllng,* 1, 305-318.
Lawless, J. F. (1987) Negative binomial and mixed Poisson regression. *Can. J. Statist.,* 15, 209-225.
Lee, Y. and Nelder, J. A. (1996) Hierarchical generalized linear models (with discussion). *J. R. Statist. Soc.* B, 58, 619-678.
Lee, Y. and Nelder, J. A. (2004) Conditional and marginal models: another view. *Statist. Sci.,* 19, 219-228.
Martins, T. G. and Rue, H. (2013) Extending INLA to a class of near-Gaussian latent models. *Preprint arXiv:1210.1434v2.*
Medina, A., Taft, N., Salamatian, K., Bhattacharyya, S. and Diot, C. (2002) Traffic matrix estimation. In *Proc. Conf. Applications, Technologies, Architectures and Protocols for Computer Communications.* New York: Association for Computing Machinery Press.
Meng, X.-L. (1994) Posterior predictive p-values. *Ann. Statist.,* 22, 1142-1160.
Nikoloulopoulos, A. K. and Karlis, D. (2008) On modeling count data: a comparison of some well-known discrete distributions. *J. Statist. Computn Simuln,* 78, 437-457.
Ntzoufras, I. (2009) *Bayesian Modeling using WinBUGS,* 1st edn. New York: Wiley.
Odoki, J. B., Kerali, H. R. and Santorini, F. (2001) An integrated model for quantifying accessibility-benefits in developing countries. *Transportn Res.* A, 35, 601-623.
Ortúzar, J. de Dios and Willumsen, L. G. (2001) *Modelling Transport,* 3rd edn. Chichester: Wiley.
Perrakis, K., Cools, M., Karlis, D., Janssens, D., Kochan, B., Bellemans, T. and Wets, G. (2012a) Quantifying input-uncertainty in traffic assignment models. In *Proc. 91st A. Meet. Transportation Research Board.* Washington DC: Transportation Research Board.
Perrakis, K., Karlis, D., Cools, M., Janssens, D., Vanhoof, K. and Wets, G. (2012b) A Bayesian approach for modeling origin-destination matrices. *Transportn Res.* A, 46, 200-212.
Puig, X., Ginebra, J. and Perez-Casany, M. (2009) Extended truncated inverse-Gaussian Poisson model. *Statist. Modllng*, 9, 151-171.
R Core Team (2013) *R: a Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.
Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc.* B, 71, 319-392.
Shaban, S. A. (1988) Poisson-lognormal distributions. In *Log-normal Distributions: Theory and Applications* (eds E. L. Crow and K. Shimizu), pp. 195-210. New York: Dekker.
Sohn, K. and Kim, D. (2010) Zonal centrality measures and the neighborhood effect. *Transprtn Res.* A, 44, 733-744.
Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003) *WinBUGS User Manual, Version 1.4.* Cambridge: Medical Research Council Biostatistics Unit. (Available from http://www.mrc-bsu.cam.ac.uk/bugs/ winbugs/manuall4.pdf.)
Tebaldi, C. and West, M. (1998) Bayesian inference on network traffic using link count data (with discussion). *J. Am. Statist. Ass.,* 93, 557-576.
Thomas, R. (1991) *Traffic Assignment Techniques.* Aldershot: Avebury Technical.
Wardrop, J. (1952) Some theoretical aspects of road traffic research. *Proc. Inst. Civ. Engrs* II, 1, 352-362.

West, M. (1994) Statistical inference for gravity models in transportation flow forecasting. *Discussion Paper 94-40.* Institute of Statistics and Decision Sciences, Duke University, Durham.

Willmot, G E. (1987) The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scand. Act. J.*, 3-4, 113-127.

Yao, E. and Morikawa, T. (2005) A study of an integrated intercity travel demand model. *Transprtn Res.* A, 39, 367-381.

Zellner, A. (1986) On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti* (eds P. Goel and A. Zellner), pp. 233-243. Amsterdam: North-Holland.