A BAYESIAN MULTINOMIAL-POISSON SIMPLIFIED MODEL FOR NETWORK TRAFFIC INFERENCE BASED ON LINK COUNT DATA

Luis E. Carvalho, Boston University, `lecarval@math.bu.edu`
Carlos F. G. Loureiro, Universidade Federal do Ceará, `felipe@det.ufc.br`

ABSTRACT

Given a network where traffic counts are observed on a selected set of links, we consider the problem of estimating traffic intensities between origin-destination (OD) pairs in this network. We adopt a Bayesian approach and set independent Poisson prior distributions on the number of trips generated on a given time interval between each OD pair. Trips are further assigned to routes in the network according to a multinomial distribution where route choice proportions are defined by a SUE model. Given the complexity in the likelihood function, we derive a hybrid Markov chain Monte Carlo algorithm that samples asymptotically from the posterior distribution and thus enables the joint inference of mean OD pair and network flows. We discuss theoretical aspects of the model, including its relation to previously studied models and the adequacy of the assumed prior distribution, and practical issues such as computational efficiency. We illustrate our method on well known example networks, compare estimates with other methods, and finish by presenting directions for future research.

Keywords: OD matrix estimation, Bayesian statistics, posterior inference, MCMC methods.

1 INTRODUCTION

Consider a network with origin-destination (OD) pairs such that each pair generates trips during a certain time interval and trips are allowed to take more than one route between origin and destination. Under this static non-fixed routing setting, and during the same time period, we record traffic counts at some specific subset of links in the network and then wish to infer OD pair flows based on observed link counts. This problem is often called OD matrix estimation. If pair flows are further assumed to be random our interest then also lies in estimating distribution parameters or expected OD matrices.

These problems are highly under-specified for most networks of interest since the number of OD pairs greatly exceeds the number of counting links and so infinitely many flow configurations can lead to same link count observations. To remedy this situation, additional information, usually in the form of previously observed or subjectively defined OD matrices, is considered to constrain or guide the estimation procedure. Following Castillo et al. (2008) and Kolaczyk (2009, Chap. 9), we briefly review two broad classes of proposed methods for OD matrix estimation:

Optimization based methods. As the name suggests, these methods require the constrained optimization of a particular objective function. In least-squares (Robillard, 1975; Cascetta and Nguyen, 1988) and generalized least-squares (Cascetta, 1984; Bell, 1991) methods an approximate Gaussian measurement model is adopted and the sum of squares of the difference between predicted and previous OD pair flows is minimized. These approaches have the advantage of being convenient from a statistical point of view since they deliver unbiased estimates and a covariance that quantifies the uncertainty in estimating the flows. On a similar vein, entropy based methods (Van Zuylen and Willumsen, 1980; Zhang et al., 2003) seek to minimize the relative entropy between predicted and previous flows subject to constraints that encourage consistency to observed link counts.

Statistical based methods. While optimization based methods attempt to find OD flows that are consistent with observed link counts and similar to a previous estimate of flows, statistical based methods assume that flows are random and try to estimate parameters that specify flow distributions. A common approach is to assume Poisson OD flow distributions (Hazelton, 2000; Lo et al., 1996) or normal approximations (Vardi, 1996; Hazelton, 2001). A Bayesian approach, in which parameters can be taken to be random themselves, tries to incorporate prior information in a more principled, if not natural, way. Noteworthy references are Maher (1983), that combined a multivariate normal likelihood with a prior distribution on mean OD flows, and Tebaldi and West (1998) that also adopted Poissonian flows based on the frequentist model of Vardi (1996). As Hazelton (2001) points out, optimization and statistical based methods attempt to solve essentially different problems, namely of OD flow estimation and mean flow estimation.

Here we adopt a Bayesian perspective that is similar to the approach from Tebaldi and West (1998) but differs from it in that we assume an even more model-based design that accounts for measurement errors in link counts and unobservable route traffic. Our proposed statistical model assumes Poissonian OD pair flows, multinomial route preferences based on a user equilibrium model, and a simplified Poisson likelihood for link count errors. In Section 2 we present this model, extend it to allow mean flows, route probabilities, and mean link count errors to be random according to a hierarchical model structure, and derive sampling procedures that drive the OD pair traffic estimation. Next, in Section 3, we illustrate the proposed methodology with computational experiments in simple example networks. Finally, we discuss the model, our proposed estimation procedure, and some directions for future work in Section 4.

2 STATISTICAL MODEL

Let $\mathcal{O}$ be the collection of origin-destination pairs in a network $\mathcal{N}$, each pair $s \in \mathcal{O}$ with origin $o(s)$ and destination $d(s)$ generating N_s trips independently according to a Poisson distribution with mean η_s: $N_s \sim \texttt{Po}(\eta_s)$. It might be desirable to further model pair interdependencies among the Poisson means but we just assume that the η_s are known for simplicity for now. In Section 2.3 we discuss an extended hierarchical model where η_s is random.

For each pair $s \in \mathcal{O}$ we select a *non-exhaustive* number of routes connecting $o(s)$ to $d(s)$, the *route set* for s, $R(s)$. A trip from $o(s)$ to $d(s)$ has a probability $p_{s,r}$ of choosing route $r \in R(s)$, and a probability $p_{s,o} \doteq 1 - \sum_{r \in R(s)} p_{s,r}$ of not using any route in $R(s)$. Let us denote by $N_{s,r}$ the number of trips between $o(d)$ and $d(s)$ that use route $r \in R(s)$. Then, conditional on N_s, the route flows for OD pair s have a multinomial distribution: $((N_{s,r})_{r \in R(s)}, N_s - \sum_{r \in R(s)} N_{s,r}) \mid N_s \sim \texttt{MN}(N_s; \mathbf{p}_s)$, where $\mathbf{p}_s \doteq ((p_{s,r})_{r \in R(s)}, p_{s,o})$.

The motivation in designing the route set is to elect a minimal number of routes that have high traffic flows and hence high probabilities of being chosen by network users. The route probabilities are, for now, assumed to be known and determined according to a stochastic user equilibrium (SUE) model (Daganzo and Sheffi, 1977; Praskher and Bekhor, 2004); in the next section we present a more flexible model within which we allow $p_{s,r}$ to be random.

For a selected set of links $\mathcal{I} \subset \mathcal{N}$ we observe traffic counts $X = (X_i)_{i \in \mathcal{I}}$. Without measurement errors and given the route flows we should have

$$\sum_{s\in\mathcal{O}} \sum_{r\in R_i(s)} N_{s,r} = X_i, \tag{1}$$

where $R_i(s) = \{r \in R(s) : i \in r\}$ is the route subset that contains link i. To avoid ambiguities, we assume that a counting link belongs to at least one route in $R(s)$ for some OD pair $s \in \mathcal{O}$ and that no two routes contain exactly the same counting link set. Conversely, given the link counts X we say that a traffic flow configuration $S \doteq ((N_{s,r})_{r\in R(s)})_{s\in\mathcal{O}}$ is count consistent if it satisfies equation (1) for all $i \in \mathcal{I}$. As hinted before, since $|\cup_{s\in\mathcal{O}} R(s)| \gg |\mathcal{I}|$, there are usually many count consistent configurations whose enumeration is prohibitively expensive. It should be noted that in realistic network designs the counting links are usually placed along high probability routes in the route sets connecting OD pairs.

To incorporate measurement errors we assume that measurement discrepancies are independent with distribution $X_i - \sum_{s\in\mathcal{O}} \sum_{r\in R_i(s)} N_{s,r} \,|\, S \sim L_i(\xi_i)$. As an example, we could adopt a zero-mean Gaussian kernel with variance parameter ξ_i: $L_i(Y;\xi_i) \propto \exp(-Y^2/2\xi_i)$. We then say that a route flow configuration S is *weakly count consistent* with respect to link counts X, denoted by $S \in \mathcal{C}(X)$, if

$$L_i\left(X_i - \sum_{s\in\mathcal{O}} \sum_{r\in R_i(s)} N_{s,r}; \xi_i\right) > 0 \tag{2}$$

for all $i \in \mathcal{I}$. Note that if $L_i(0;\xi_i) > 0$ for all links then count consistency implies weak count consistency; in particular, in the absence of errors $L_i(Y;\xi_i) = I(Y = 0)$, where $I(\cdot)$ denotes the indicator function, and thus weak count consistency and count consistency become equivalent.

2.1 Model inference

According to our model specification the *likelihood* $\mathrm{P}(X \,|\, N)$ of N is given by

$$\mathrm{P}(X \,|\, N) = \sum_{S\in\mathcal{C}(X)} \mathrm{P}(X, S \,|\, N)$$

where

$$\begin{aligned}
\mathrm{P}(X, S \,|\, N) &= \mathrm{P}(X \,|\, S, N)\mathrm{P}(S \,|\, N) \\
&= \prod_{i\in\mathcal{I}} L_i\left(X_i - \sum_{s\in\mathcal{O}} \sum_{r\in R_i(s)} N_{s,r}; \xi_i\right) \\
&\quad \prod_{s\in\mathcal{O}} \frac{N_s!}{\prod_{r\in R(s)} N_{s,r}!(N_s - \sum_{r\in R(s)} N_{s,r})!} \\
&\quad \prod_{r\in R(s)} p_{s,r}^{N_{s,r}} \left(1 - \sum_{r\in R(s)} p_{s,r}\right)^{N_s - \sum_{r\in R(s)} N_{s,r}}.
\end{aligned} \tag{3}$$

A possible inferential approach is then to find $\arg\max_{\tilde{N}} P(X \,|\, \tilde{N})$, the maximum likelihood estimator. However, as previously noted, N is highly non-identifiable with many solutions yielding the same likelihood.

Alternatively, in a Bayesian statistical setting all the inference is driven by the *posterior*, conditional on data, distribution of the variable of interest. In our case N has posterior distribution

$$\mathrm{P}(N \,|\, X) = \frac{\mathrm{P}(X \,|\, N)\mathrm{P}(N)}{\sum_N \mathrm{P}(X \,|\, N)\mathrm{P}(N)},$$

according to Bayes' rule, where prior information is considered through $\mathrm{P}(N)$, termed the *prior* distribution of N,

$$\mathrm{P}(N) = \prod_{s \in \mathcal{O}} \frac{\eta_s^{N_s} e^{-\eta_s}}{N_s!}. \tag{4}$$

Estimates for N can then be simply obtained from the posterior mean $\mathrm{E}_{N \mid X}[N]$ or posterior mode $\arg\max_{\tilde{N}} \mathrm{P}(\tilde{N} \mid X)$.

The main difficulty in evaluating $\mathrm{P}(X \mid N)$, and hence $\mathrm{P}(N \mid X)$, is in summing over all consistent flow configurations in $\mathcal{C}(X)$ that satisfy equation (2), but this is not a new hurdle since it stems from the linear constraints on $N_{s,r}$ given X. Fortunately, we can resort to Markov chain Monte Carlo (MCMC) methods (Gelman et al., 2003; Givens and Hoeting, 2005) to sample from the joint posterior $\mathrm{P}(N, S \mid X)$ and obtain marginal posterior estimates for N, as we show next.

2.2 Posterior sampling

In a Gibbs sampler (Geman and Geman, 1984), a type of MCMC algorithm, we iteratively sample from the conditional distribution of one variable given the others until convergence to the desired, target distribution. In our case we wish to sample from $\mathrm{P}(N_s \mid N_{[-s]}, S, X)$ for each $s \in \mathcal{O}$, where $N_{[-s]} = (N_u)_{u \in \mathcal{O}, u \neq s}$, and $\mathrm{P}(N_{s,r} \mid N, S_{[-r]}, X)$ for each $s \in \mathcal{O}, r \in R(s)$, where $S_{[-r]} = ((N_{u,t})_{t \in R(u), t \neq r})_{u \in \mathcal{O}}$, until convergence to $\mathrm{P}(N, S \mid X)$.

Now, since from (3) and (4)

$$\mathrm{P}(N_s, N_{[-s]}, S, X) = \frac{\eta_s^{N_s} \Phi_{[-s]}(N_{[-s]}, S, X)}{(N_s - \sum_{r \in R(s) N_{s,r}})!} \left(1 - \sum_{r \in R(s)} p_{s,r}\right)^{N_s - \sum_{r \in R(s)} N_{s,r}},$$

where $\Phi_{[-s]}(\cdot)$ only contains terms that do not depend on N_s, we then have

$$\mathrm{P}(N_s \mid N_{[-s]}, S, X) = \frac{\mathrm{P}(N_s, N_{[-s]}, S, X)}{\sum_{N_s} \mathrm{P}(N_s, N_{[-s]}, S, X)} = \frac{(\eta_s p_{s,o})^{N_s - \sum_{r \in R(s)} N_{s,r}} e^{-\eta_s p_{s,o}}}{(N_s - \sum_{r \in R(s) N_{s,r}})!}$$

and so

$$N_s - \sum_{r \in R(s)} N_{s,r} \,\middle|\, N_{[-s]}, S, X \sim \mathtt{Po}\left[\eta_s \left(1 - \sum_{r \in R(s)} p_{s,r}\right)\right]. \tag{5}$$

Next, before finding $\mathrm{P}(N_{s,r} \mid S_{[-r]}, N, X)$, let us define $X_{\langle i,r \rangle} \doteq X_i - \sum_{u \in \mathcal{O}} \sum_{t \in R(u), t \neq r} N_{u,t}$, $p_{\langle s,r \rangle} \doteq 1 - \sum_{t \in R(s), t \neq r} p_{s,t}$, and $N_{\langle s,r \rangle} \doteq N_s - \sum_{t \in R(s), t \neq r} N_{s,t}$ to simplify the notation. Then, also from (3) and (4) we have

$$\mathrm{P}(N_{s,r}, S_{[-r]}, N, X) = \prod_{\{i \in \mathcal{I} : r \in R_i(s)\}} L_i(X_{\langle i,r \rangle} - N_{s,r}; \xi_i) \Psi_{[-r]}(S_{[-r]}, N, X) \frac{p_{s,r}^{N_{s,r}} (p_{\langle s,r \rangle} - p_{s,r})^{N_{\langle s,r \rangle} - N_{s,r}}}{N_{s,r}! (N_{\langle s,r \rangle} - N_{s,r})!}$$

where $\Psi_{[-r]}(\cdot)$ only includes terms that do not depend on $N_{s,r}$. Rearranging and scaling terms we obtain

$$\mathrm{P}(N_{s,r} \mid S_{[-r]}, N, X) \propto \frac{N_{\langle s,r \rangle}!}{N_{s,r}! (N_{\langle s,r \rangle} - N_{s,r})!} \left(\frac{p_{s,r}}{p_{\langle s,r \rangle}}\right)^{N_{s,r}} \left(\frac{p_{\langle s,r \rangle} - p_{s,r}}{p_{\langle s,r \rangle}}\right)^{N_{\langle s,r \rangle} - N_{s,r}} \prod_{\{i \in \mathcal{I} : r \in R_i(s)\}} L_i(X_{\langle i,r \rangle} - N_{s,r}; \xi_i). \tag{6}$$

To sample from $\mathrm{P}(N_{s,r} \mid S_{[-r]}, N, X)$ we would need to find the proportionality constant in (6) by summing the expression over the support of $N_{s,r}$,

$$\mathcal{S}_{s,r} \doteq \{0, \ldots, N_{\langle s,r \rangle}\} \cap \bigcap_{\{i \in \mathcal{I} \,:\, r \in R_i(s)\}} \Big\{ N_{s,r} : L_i(X_{\langle i,r \rangle} - N_{s,r}; \xi_i) > 0 \Big\},$$

the intersection of the binomial and likelihood supports. However such operation would incur in a significant computational burden. Alternatively, we employ a Metropolis-Hasting step (Hastings, 1970): we sample a candidate $N^*_{s,r}$ from a proposal distribution g, compute a Metropolis ratio

$$R(N_{s,r}, N^*_{s,r}) = \frac{\mathrm{P}(N^*_{s,r} \mid S_{[-r]}, N, X) g(N_{s,r})}{\mathrm{P}(N_{s,r} \mid S_{[-r]}, N, X) g(N^*_{s,r})},$$

and accept $N^*_{s,r}$ as a valid sample with probability $\min\{R(N_{s,r}, N^*_{s,r}), 1\}$. Since in this case we sample $N^*_{s,r}$ regardless of $N_{s,r}$, this routine implements an *independent* Metropolis-Hastings step.

Let

$$g(N_{s,r}) \propto \frac{N_{\langle s,r \rangle}!}{N_{s,r}!(N_{\langle s,r \rangle} - N_{s,r})!} \left(\frac{p_{s,r}}{p_{\langle s,r \rangle}} \right)^{N_{s,r}} \left(\frac{p_{\langle s,r \rangle} - p_{s,r}}{p_{\langle s,r \rangle}} \right)^{N_{\langle s,r \rangle} - N_{s,r}} I(N_{s,r} \in \mathcal{S}_{s,r}) \qquad (7)$$

and

$$w(N_{s,r}) \doteq \prod_{\{i \in \mathcal{I} \,:\, r \in R_i(s)\}} L_i(X_{\langle i,r \rangle} - N_{s,r}; \xi_i)$$

and thus $\mathrm{P}(N_{s,r} \mid S_{[-r]}, N, X) \propto g(N_{s,r}) w(N_{s,r})$ according to (6). We then use $g(\cdot)$ as proposal distribution, that is, we sample a candidate

$$N^*_{s,r} \,\Big|\, S_{[-r]}, N, X \sim \mathtt{Bin}\left(N_{\langle s,r \rangle}, \frac{p_{s,r}}{p_{\langle s,r \rangle}} \right)$$

until $N^*_{s,r}$ is in the support $\mathcal{S}_{s,r}$ and accept it based on the Metropolis ratio

$$R(N_{s,r}, N^*_{s,r}) = \prod_{\{i \in \mathcal{I} \,:\, r \in R_i(s)\}} \frac{w(N^*_{s,r})}{w(N_{s,r})} = \prod_{\{i \in \mathcal{I} \,:\, r \in R_i(s)\}} \frac{L_i(X_{\langle i,r \rangle} - N^*_{s,r}; \xi_i)}{L_i(X_{\langle i,r \rangle} - N_{s,r}; \xi_i)}, \qquad (8)$$

a product likelihood ratio. Note that, as an advantage, we do not need to compute the proportionality constant in (6) anymore.

Finally, putting all the steps together we obtain a hybrid MCMC algorithm—known as "Metropolis-within-Gibbs" (Gilks et al., 1995)—at Figure 1 for sampling from the joint posterior distribution of N and S. We can now estimate N_s, for some $s \in \mathcal{O}$, by approximating the posterior mean of N_s using the samples:

$$\mathrm{E}_{N_s \mid X}[N_s] \approx \frac{1}{n} \sum_{k=1}^{n} N_s^{(k)}$$

and similarly for $N_{s,r}$, $r \in R(s)$.

2.3 Extended hierarchical model

To better incorporate uncertainty into model parameters, since they are usually estimated or guessed from previous observations, we now adopt a simple hierarchical model with conjugate hyper-priors: for $s \in \mathcal{O}$, $\eta_s \sim \mathtt{Ga}(a_s, b_s)$ and $\mathbf{p}_s \sim \mathtt{Dir}(\pi_s)$, where $\mathtt{Ga}(a, b)$ denotes the gamma distribution with shape a and rate b, and $\mathtt{Dir}$ is the Dirichlet distribution.

1. *(Initialization)* Sample from prior until $S^{(1)} \in \mathcal{C}(X)$: for $s \in \mathcal{O}$ sample $N_s^{(1)} \sim \mathtt{Po}(\eta_s)$ and for $r \in R(s)$ sample $S^{(1)} \sim \mathtt{MN}(N_s^{(1)}, \mathbf{p}_s)$.

2. *(Gibbs sampling)* For $k = 1, \ldots, n$, with n large enough to guarantee convergence to $\mathrm{P}(N, S \,|\, X)$:

   (a) For $s \in \mathcal{O}$ sample according to (5):

   $$N_s^{(k+1)} - \sum_{r \in R(s)} N_{s,r}^{(k)} \,\Bigg|\, N_{[-s]}^{(k)}, S^{(k)}, X \sim \mathtt{Po}\left[\eta_s \left(1 - \sum_{r \in R(s)} p_{s,r}\right)\right].$$

   (b) *(Metropolis-Hastings step)* For $s \in \mathcal{O}$, $r \in R(s)$:

      i. Sample candidate $N_{s,r}^{*}$ from proposal in (7):

      $$N_{s,r}^{*} \,\Bigg|\, S_{[-r]}^{(k)}, N^{(k+1)}, X \sim \mathtt{Bin}\left(N_{\langle s,r\rangle}^{(k)}, \frac{p_{s,r}}{p_{\langle s,r\rangle}}\right)$$

      until $N_{s,r}^{*} \in \mathcal{S}_{s,r}^{(k)}$, where some of the route flows in $S_{[-r]}^{(k)}$ might have already been updated by a previous Metropolis-Hastings step.

      ii. Sample $U \sim \mathtt{U}[0,1]$ and compute, according to (8),

      $$R(N_{s,r}^{(k)}, N_{s,r}^{*}) = \prod_{\{i \in \mathcal{I} \,:\, r \in R_i(s)\}} \frac{L_i(X_{\langle i,r\rangle} - N_{s,r}^{*}; \xi_i)}{L_i(X_{\langle i,r\rangle} - N_{s,r}^{(k)}; \xi_i)}.$$

      iii. If $U \leq \min\{R(N_{s,r}^{(k)}, N_{s,r}^{*}), 1\}$ set $N_{s,r}^{(k+1)} = N_{s,r}^{*}$ (accept); otherwise set $N_{s,r}^{(k+1)} = N_{s,r}^{(k)}$ (reject).

Figure 1: Metropolis-within-Gibbs sampling algorithm for the joint posterior on N and S.

To apply the same treatment to measurement errors we should define hyper-priors for the parameters ξ_i. For computational simplicity—and to have a more homogeneous model—we settle on Poisson errors with mean ξ_i, $X_i - \sum_{s\in\mathcal{O}}\sum_{r\in R_i(s)} N_{s,r} \mid S \sim \mathtt{Po}(\xi_i)$, where ξ_i are random with conjugate distribution $\xi_i \sim \mathtt{Ga}(c_i, d_i)$, for $i \in \mathcal{I}$. As a consequence, $S \in \mathcal{C}(X)$ corresponds to count sub-consistent route flows that satisfy

$$\sum_{s\in\mathcal{O}}\sum_{r\in R_i(s)} N_{s,r} \le X_i$$

for all counting links.

Under this extended model we can now address the issue of estimation of mean OD pair flows. As in the previous section, we now look to devise a sampling scheme for computing posterior means $\mathrm{E}_{\eta_s \mid X}[\eta_s]$ as estimates for η_s. But it is now straightforward, due to conjugacy, to obtain, for $s \in \mathcal{O}$ and $i \in \mathcal{I}$,

$$\begin{aligned}
\eta_s \mid \eta_{[-s]}, \xi, \mathbf{p}, S, N, X &\sim \mathtt{Ga}(a_s + N_s, b_s + 1),\\
\mathbf{p}_s \mid \eta, \xi, \mathbf{p}_{[-s]}, S, N, X &\sim \mathtt{Dir}\left((\pi_{s,r} + N_{s,r})_{r\in R(s)}, \pi_{s,o} + N_s - \sum_{r\in R(s)} N_{s,r}\right),\\
\xi_i \mid \eta, \xi_{[-i]}, \mathbf{p}, S, N, X &\sim \mathtt{Ga}\left(c_i + X_i - \sum_{s\in\mathcal{O}}\sum_{r\in R_i(s)} N_{s,r}, d_i + 1\right),
\end{aligned} \tag{9}$$

where $\eta \doteq (\eta_s)_{s\in\mathcal{O}}$, $\xi \doteq (\xi_i)_{i\in\mathcal{I}}$, $\mathbf{p} \doteq (\mathbf{p}_s)_{s\in\mathcal{O}}$, and we use $\eta_{[-s]}$ to denote the exclusion of OD pair s from $(\eta_s)_{s\in\mathcal{O}}$ and similarly for $\mathbf{p}_{[-s]}$ and $\xi_{[-i]}$ as before.

The conditional distributions in (9) can then be casted as steps in an extended version of the hybrid Gibbs sampler in Figure 1. Since N_s is conditionally independent of ξ_i and X given S, η_s, and $\mathbf{p}_s$, we do not depend on the choice of L_i to sample N_s similarly to (5):

$$N_s - \sum_{r\in R(s)} N_{s,r} \,\Bigg|\, \eta, \xi, \mathbf{p}, S, N_{[-s]}, X \sim \mathtt{Po}\left[\eta_s\left(1 - \sum_{r\in R(s)} p_{s,r}\right)\right]. \tag{10}$$

We still require a Metropolis-Hastings step to sample $N_{s,r}$, but given the specific choice of L_i we now know that the support $\mathcal{S}_{s,r}$ goes from 0 to

$$\min\left\{N_{\langle s,r\rangle}, \min_{\{i\in\mathcal{I}\,:\, r\in R_i(s)\}} X_{\langle i,r\rangle}\right\}$$

and, adopting the same truncated binomial proposal $g(\cdot)$ in (7), the "importance" weights

$$w(N_{s,r}) \doteq \prod_{\{i\in\mathcal{I}\,:\, r\in R_i(s)\}} \frac{\xi_i^{X_{\langle i,r\rangle} - N_{s,r}}}{(X_{\langle i,r\rangle} - N_{s,r})!}$$

yield the Metropolis ratio

$$R(N_{s,r}, N^*_{s,r}) = \prod_{\{i\in\mathcal{I}\,:\, r\in R_i(s)\}} \frac{w(N^*_{s,r})}{w(N_{s,r})} = \prod_{\{i\in\mathcal{I}\,:\, r\in R_i(s)\}} \xi_i^{N_{s,r} - N^*_{s,r}} \frac{(X_{\langle i,r\rangle} - N_{s,r})!}{(X_{\langle i,r\rangle} - N^*_{s,r})!}. \tag{11}$$

The new algorithm is listed in Figure 2.

3 CASE STUDIES

In this section we show two application examples for simple networks adapted from (Hazelton, 2000) and (Lo et al., 1996) respectively.

1. *(Initialization)* Sample from prior until $S^{(1)} \in \mathcal{C}(X)$:
   (a) *(Hyper-priors)* For $s \in \mathcal{O}$ sample $\eta_s^{(1)} \sim \mathtt{Ga}(a_s, b_s)$ and $\mathbf{p}_s^{(1)} \sim \mathtt{Dir}(\pi_s)$, and for $i \in \mathcal{I}$ sample $\xi_i^{(1)} \sim \mathtt{Ga}(c_i, d_i)$.
   (b) For $s \in \mathcal{O}$ sample $N_s^{(1)} \sim \mathtt{Po}(\eta_s^{(1)})$ and for $r \in R(s)$ sample $S^{(1)} \sim \mathtt{MN}(N_s^{(1)}, \mathbf{p}_s^{(1)})$.
2. *(Gibbs sampling)* For $k = 1, \ldots, n$, with n large enough to guarantee convergence to $\mathrm{P}(\eta, \xi, \mathbf{p}, N, S \,|\, X)$:
   (a) *(Hyper-priors)* Sample according to (9): for $s \in \mathcal{O}$,

$$\eta_s^{(k+1)} \,|\, \cdot \sim \mathtt{Ga}(a_s + N_s^{(k)}, b_s + 1),$$
$$\mathbf{p}_s^{(k+1)} \,\Bigg|\, \cdot \sim \mathtt{Dir}\left((\pi_{s,r}^{(k)} + N_{s,r}^{(k)})_{r \in R(s)}, \pi_{s,o}^{(k)} + N_s^{(k)} - \sum_{r \in R(s)} N_{s,r}^{(k)} \right),$$

   and, for $i \in \mathcal{I}$,

$$\xi_i^{(k+1)} \,\Bigg|\, \cdot \sim \mathtt{Ga}\left(c_i + X_i - \sum_{s \in \mathcal{O}} \sum_{r \in R_i(s)} N_{s,r}^{(k)}, d_i + 1 \right).$$

   (b) For $s \in \mathcal{O}$, sample according to (10),

$$N_s^{(k+1)} - \sum_{r \in R(s)} N_{s,r}^{(k)} \,\Bigg|\, \cdot \sim \mathtt{Po}\left[\eta_s^{(k+1)} \left(1 - \sum_{r \in R(s)} p_{s,r}^{(k+1)} \right) \right].$$

   (c) *(Metropolis-Hastings step)* For $s \in \mathcal{O}$, $r \in R(s)$:
      i. Sample candidate $N_{s,r}^*$ from proposal (7) until $N_{s,r}^* \in \mathcal{S}_{s,r}$:

$$N_{s,r}^* \,\Bigg|\, \cdot \sim \mathtt{Bin}\left(N_{\langle s,r \rangle}^{(k)}, \frac{p_{s,r}^{(k+1)}}{p_{\langle s,r \rangle}^{(k+1)}} \right).$$

      ii. Sample $U \sim \mathtt{U}[0,1]$ and compute, according to (11),

$$R(N_{s,r}^{(k)}, N_{s,r}^*) = \prod_{\{i \in \mathcal{I} \,:\, r \in R_i(s)\}} {\xi_i^{(k+1)}}^{N_{s,r}^{(k)} - N_{s,r}^*} \frac{(X_{\langle i,r \rangle} - N_{s,r}^{(k)})!}{(X_{\langle i,r \rangle} - N_{s,r}^*)!}.$$

      iii. If $U \le \min\{R(N_{s,r}^{(k)}, N_{s,r}^*), 1\}$ set $N_{s,r}^{(k+1)} = N_{s,r}^*$ (accept); otherwise set $N_{s,r}^{(k+1)} = N_{s,r}^{(k)}$ (reject).

Figure 2: Metropolis-within-Gibbs sampling algorithm for the joint posterior on N, S, and hyper-parameters η, ξ, and $\mathbf{p}$ under the extended hierarchical model.

Example 1

Consider the three-node network depicted in Figure 3 with three OD pairs: $1 \sim [1 \to 2]$, $2 \sim [1 \to 3]$, and $3 \sim [2 \to 3]$. Let us assume that the OD pairs have Poisson mean parameters $\eta_1^{(0)} = 70$, $\eta_2^{(0)} = 100$, and $\eta_3^{(0)} = 80$ estimated from similar networks. Observed link counts are $X = (120, 130, 50)$, and let us make $\xi_i^{(0)} = X_i\sigma^2$ for $i = 1, 2, 3$, where $\sigma = 0.2$ as estimated from previous observations.

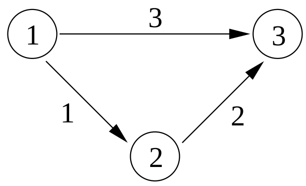

Figure 3: Network for Example 1 from (Hazelton, 2000).

Given the simplicity of the network it is reasonable to assume that only a small fraction of users does not use any of the links, and so we suppose that each OD pair s has $p_{s,o} \doteq \delta = 0.01$. OD pair 2 has two possible routes: one straight from node 1 to node 3 and other passing through node 2. We assume that route $1 \to 2 \to 3$ has probability $\theta(1-\delta)$. Table 1 summarizes the network parameters. For the experiments, let us assume that both routes connecting pair 2 have equal probability of being selected, and thus $\theta = 0.5$.

Table 1: Network parameters for Example 1.

Route r	OD pair s	Links	Prob. $p_{s,r}$
$1 \to 2$	1	1	$1-\delta$
$1 \to 2 \to 3$	2	$1, 2$	$(1-\delta)\theta$
$1 \to 3$	2	3	$(1-\delta)(1-\theta)$
$2 \to 3$	3	2	$1-\delta$

For an extended model we fix a parameter ρ for the gamma rates and set $a_s = \eta_s^{(0)}\rho$, $b_s = \rho$ for $s \in \mathcal{O}$ and $c_i = \xi_i^{(0)}\rho$, $d_i = \rho$ for $i \in \mathcal{I}$. Since $\mathtt{Ga}(a\rho, \rho)$ has mean a and variance a/ρ, parameter ρ in the extended model controls for how informative the estimates $\eta_s^{(0)}$ and $\xi_i^{(0)}$ are. Route probabilities have Dirichlet prior distribution with parameters $\pi_s = \mathbf{p}_s/\delta$, simply scaled so that $\pi_{s,o} = 1$.

We now run three scenarios: base model, extended model with $\rho = 0.05$ (non-informative), and extended model with $\rho = 5$ (informative). In all cases we run the algorithms in Figures 1 and 2 for the base and extended models respectively for $100{,}000$ iterations and keep the last $10{,}000$ assuming the limiting behavior of the chain has then prevailed. Table 2 and Figures 4 and 5 summarize our results.

Table 2: Posterior mean and standard deviation (in parentheses) of N and η estimated from $10{,}000$ iterations of the hybrid Gibbs sampler for Example 1.

Model/Param.	N_1	N_2	N_3
Base	70.26 (4.75)	94.82 (4.89)	79.96 (4.94)
Ext., $\rho = 0.05$	62.41 (6.76)	104.98 (7.25)	72.49 (7.87)
Ext., $\rho = 5$	69.47 (5.27)	95.47 (5.42)	79.48 (5.17)
	η_1	η_2	η_3
Ext., $\rho = 0.05$	62.83 (10.13)	104.71 (12.15)	72.94 (11.29)
Ext., $\rho = 5$	69.89 (3.52)	99.28 (4.21)	79.94 (3.74)

We first note that a small δ has the effect of estimating close to (strongly) consistent route flows. Since $\eta_s^{(0)}$ admits a count consistent flow where $N_{2,1} = N_{2,2} = 50$ and $N_s = \eta_s^{(0)}$, $s =$

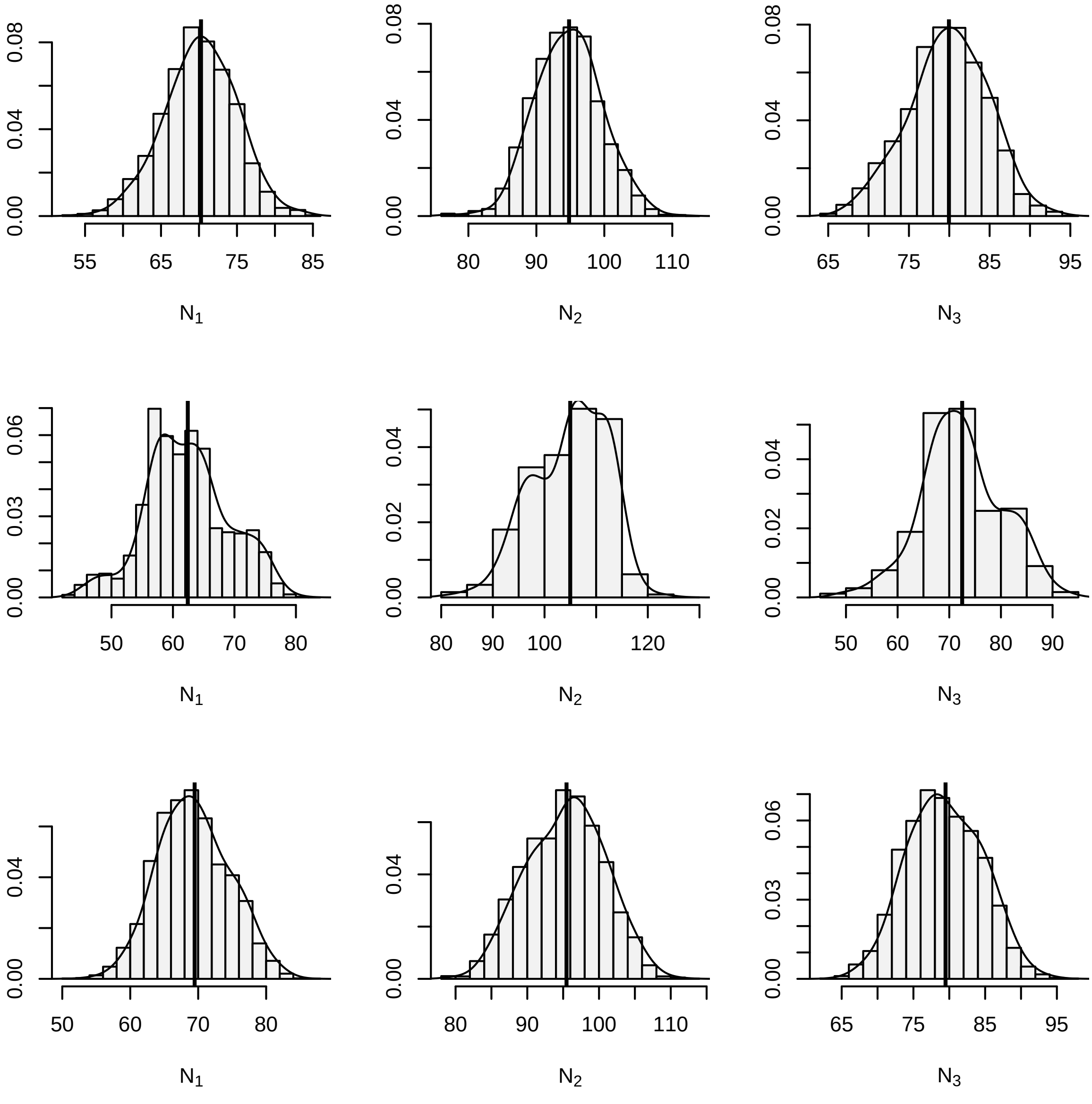

Figure 4: Estimated posterior distribution of N from base model (top), extended models with $\rho = 0.05$ (middle) and $\rho = 5$ (bottom) for Example 1. Bold vertical lines indicate the posterior mean.

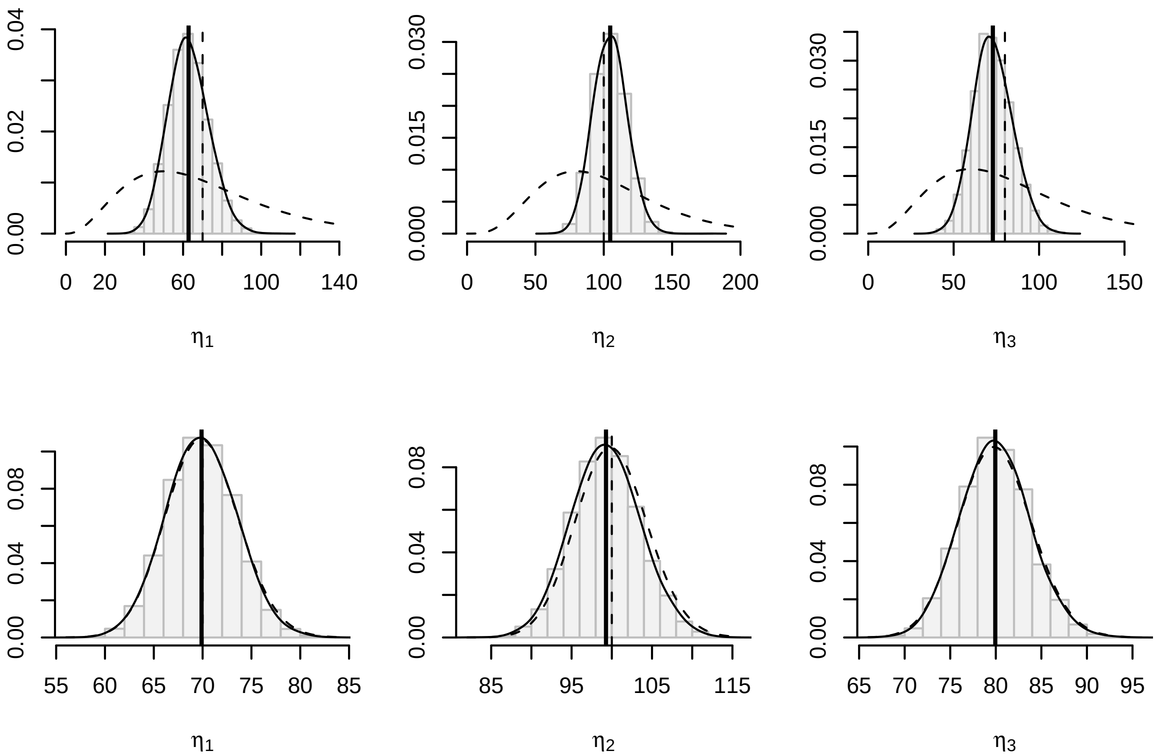

Figure 5: Estimated posterior distribution of η from extended models with $\rho = 0.05$ (top) and $\rho = 5$ (bottom) for Example 1. Bold vertical lines indicate the posterior mean. Dashed curves and vertical lines show prior distribution and prior mean.

$1, 2, 3$, we obtain estimates that are reasonably expected under the base and informative scenarios. Albeit different from the sampling-based approach in Hazelton (2000), these results are comparable to his "encouraging" simulation estimates obtained from likelihood maximization of a multivariate normal approximation. We point out that even though each N_s has a prior Poisson, the posterior distribution is less rigid with a much smaller variance given the influence of data. Such influence is also observed for mean parameters in the extended models; in particular, as the top panel in Figure 5 shows, the observed link counts greatly affect the posterior distribution of η_s.

The seemingly poorer results for the non-informed scenario should not come as a surprise: as we have less trust in our previous estimates more flow configurations become likely, that is, we face more severe identifiability issues. Less informative priors yield higher variance posteriors, as Table 2 demonstrates. The middle panel in Figure 4 hints at two modal flow configurations: one already corroborated by the base and informative cases, $N = (70, 100, 80)$, and other with $N = (60, 110, 70)$ and thus more likely unbalanced route flows in OD pair 2. However, this latter estimate becomes more plausible given that the posterior distribution of η_s is centered closer to it.

Example 2

The network in Figure 6, also discussed by Hazelton (2000), is taken from Cascetta and Nguyen (1988) and Lo et al. (1996). Here we take $\mathcal{O} = \{1, 3, 4, 6\}^2$ and consider all routes connecting each OD pair that are not cyclic. We assign route choice probabilities according to a logit model on route travel costs c_r, the sum of costs over links in route r, with parameter $\theta = 0.2$. We again follow Cascetta and Nguyen (1988) and assume that each link in route r has uniform cost of 5 units. As in the previous example, we set $p_{s,o} = \delta = 0.01$, and so

$$p_{s,r} = \frac{\exp(-\theta c_r)}{\sum_{t \in R(s)} \exp(-\theta c_t)} (1 - \delta).$$

Dirichlet priors are also taken as $\pi_s = \mathbf{p}_s / \delta$ to ensure $\pi_{s,o} = 1$.

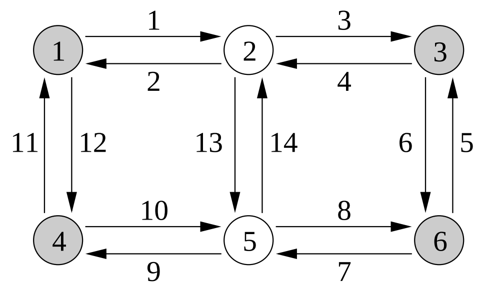

Figure 6: Network for Example 2 from (Cascetta and Nguyen, 1988). Origin/destination nodes are identified in gray.

Previous estimates for mean OD flows $\eta_s^{(0)}$ are listed in Table 3. For the base model we simply adopt $\eta_s^{(0)}$ as mean parameters. To generate link count data, we set $\sigma = 0.1$, simulate OD flows and route flows $\tilde{N}_{s,r}$ and sample, for $i = 1, \ldots, 14$,

$$X_i - \sum_{s \in \mathcal{O}} \sum_{r \in R_i(s)} \tilde{N}_{s,r} \,\Bigg|\, \texttt{Po}\left(\sum_{s \in \mathcal{O}} \sum_{r \in R_i(s)} \tilde{N}_{s,r} \sigma^2 \right),$$

obtaining $X = (810, 935, 859, 839, 539, 722, 746, 655, 642, 701, 715, 518, 370, 508)$. As before, we define a precision parameter ρ for how informative $\eta_s^{(0)}$ are and set priors $\eta_s \sim \texttt{Ga}(\eta_s^{(0)} \rho, \rho)$ for $s \in \mathcal{O}$ and $\xi_i \sim \texttt{Ga}(X_i \sigma^2 \rho, \rho)$ for $i \in \{1, \ldots, 14\}$.

Table 3: Previously estimated mean OD pair flows $\eta_s^{(0)}$ for network in Example 2.

Origin / Destination	1	3	4	6
1	—	500	250	250
3	500	—	250	250
4	250	500	—	250
6	500	250	250	—

The same three scenarios from Example 1 are simulated now: base model and extended non-informed ($\rho = 0.05$) and informed ($\rho = 5$) models. Table 4 and Figures 7 and 8 summarize the last $10{,}000$ samples from a simulation run of $100{,}000$ iterations. Similarly to the previous example, the posterior distributions for N_s under the base and informed models agree well with previous observations, while estimates under the non-informative setting exhibit larger variances. Given the higher complexity of this network, however, we should expect an increased dependence on prior specifications to alleviate identifiability problems in inference. In fact, as Figure 8 illustrates, posterior distributions for η_s are more disperse, and thus more sensible, depending on ρ.

Table 4: Posterior mean and standard deviation (in parentheses) of N and η estimated from $10{,}000$ iterations of the hybrid Gibbs sampler for Example 2.

OD pair s	N_s			η_s	
$[o(s) \to d(s)]$	Base	Ext., $\rho = 0.05$	Ext., $\rho = 5$	Ext., $\rho = 0.05$	Ext., $\rho = 5$
$1 \sim [1 \to 3]$	482.92 (15.39)	515.39 (25.44)	479.69 (18.39)	514.82 (32.97)	496.66 (9.68)
$2 \sim [1 \to 4]$	235.29 (10.61)	227.75 (15.99)	234.87 (10.98)	228.69 (21.16)	247.44 (6.68)
$3 \sim [1 \to 6]$	239.56 (12.17)	164.09 (17.94)	240.55 (11.82)	168.17 (21.21)	248.42 (6.77)
$4 \sim [3 \to 1]$	517.33 (17.27)	525.26 (14.83)	512.74 (18.15)	524.12 (26.40)	502.03 (9.61)
$5 \sim [3 \to 4]$	240.4 (10.98)	310.52 (20.72)	247.62 (15.57)	307.48 (26.36)	249.67 (7.00)
$6 \sim [3 \to 6]$	247.7 (12.88)	198.47 (18.21)	244.47 (12.76)	201.09 (22.15)	249.04 (6.74)
$7 \sim [4 \to 1]$	250.35 (11.44)	196.92 (22.51)	254.04 (13.42)	199.21 (25.51)	250.65 (6.87)
$8 \sim [4 \to 3]$	476.21 (13.55)	479.77 (27.41)	484.19 (14.40)	480.72 (33.79)	497.36 (9.47)
$9 \sim [4 \to 6]$	243.82 (10.82)	290.17 (26.67)	244.11 (14.95)	288.49 (30.62)	249.00 (6.93)
$10 \sim [6 \to 1]$	508.98 (14.47)	508.31 (18.59)	505.13 (17.54)	508.11 (28.34)	500.87 (9.53)
$11 \sim [6 \to 3]$	248.35 (10.33)	240.13 (25.10)	245.38 (14.04)	240.74 (28.15)	249.16 (6.91)
$12 \sim [6 \to 4]$	242.94 (13.18)	171.24 (29.02)	247.79 (14.84)	175.03 (30.47)	249.65 (6.85)

4 DISCUSSION

In this paper we have taken a Bayesian approach to the problem of estimating OD pair flows and mean OD pair flows in traffic networks based on link count data. Extending the work of Tebaldi and West (1998) and Vardi (1996), we adopted a model that considers count measurement errors and allows for unobserved route flows. The main contribution is then a methodology for OD matrix estimation under stochastic routing that incorporates prior network information within a sound statistical framework. To this end, we have developed a sampling scheme based on a hybrid MCMC Gibbs sampler that mixes independent Metropolis-Hastings steps. The proposed model and estimation methodology were illustrated in two simple network examples.

Besides the advantage of being a more principled approach, an statistical treatment of OD matrix estimation benefits from a long experience, as raised by Castillo et al. (2008), and encompasses a large set of tools including credibility intervals and hypothesis tests to quantify variability and test estimates. For instance, the assumption of unobservable routes makes the model arguably more appropriate for large-scale networks where, even though an OD pair is connected by a large number of routes, most of them have a very low probability of being chosen. However, from a broader perspective, it is a common and encouraged practice within Bayesian settings to validate model

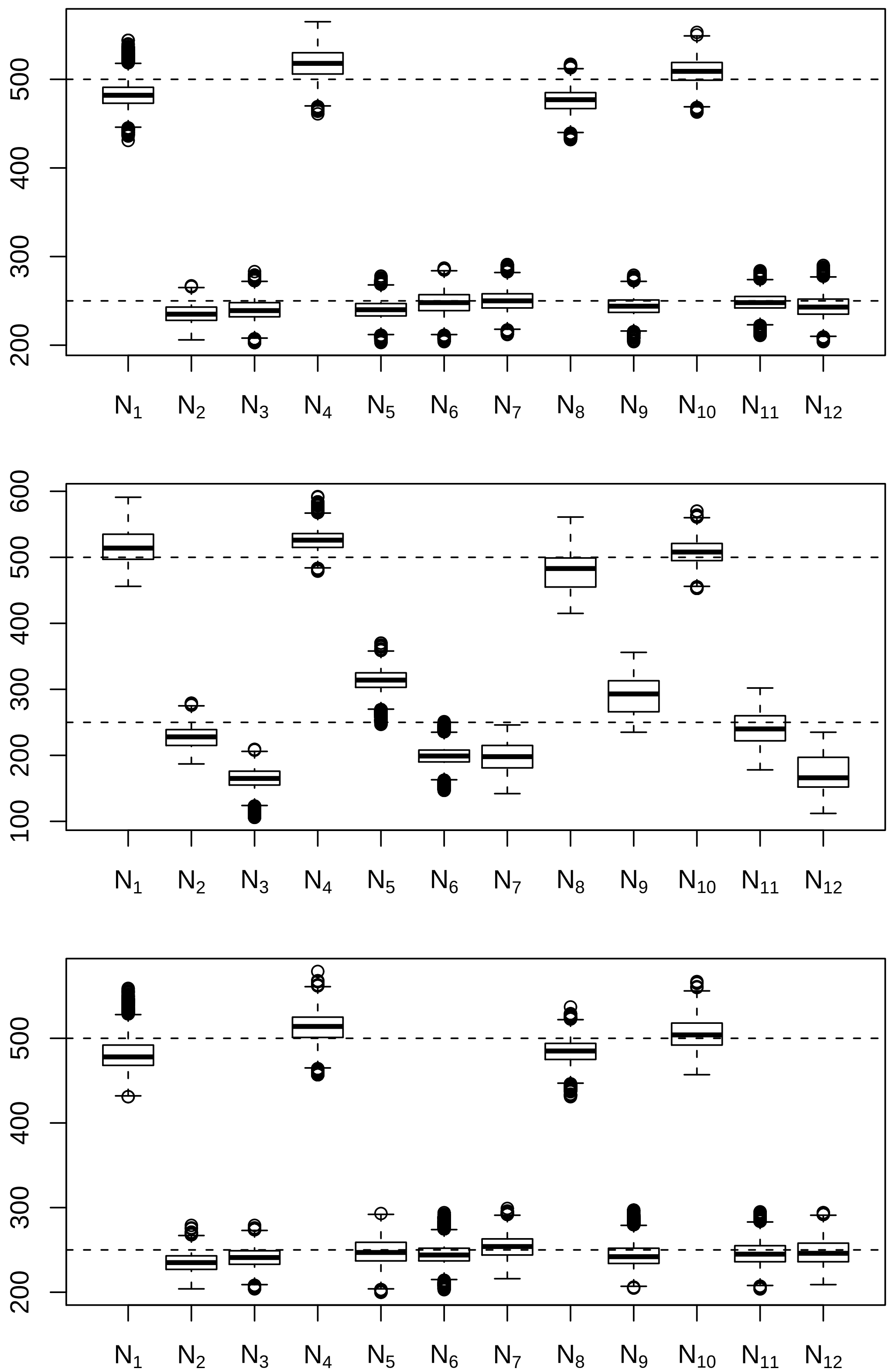

Figure 7: Boxplots of estimated posterior distribution of N from base model (top), extended models with $\rho = 0.05$ (middle) and $\rho = 4$ (bottom) for Example 2.

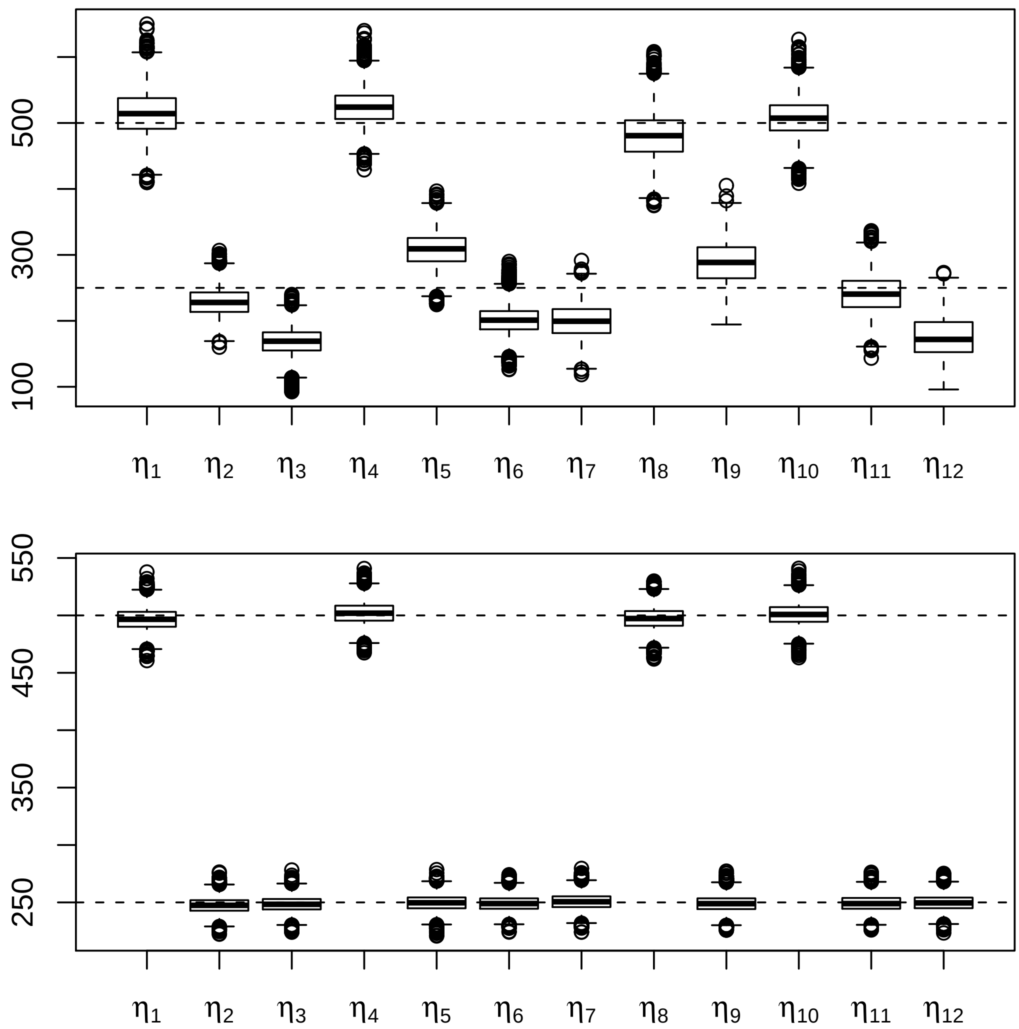

Figure 8: Boxplots of estimated posterior distribution of η from extended models with $\rho = 0.05$ (top) and $\rho = 5$ (bottom) for Example 2.

assumptions using posterior predictive checks. All these statistical tools are easy to implement and promptly obtainable from our sampling routine.

The main caveat in MCMC methods is the assessment of convergence to the target distribution. Even for small networks such as those in Section 3 the estimation problem might become severely ill-posed given the number of possible solutions that induce the same posterior probability. It is then natural to expect a large number of iterations until convergence is assured, since high correlation across variables due to consistency constraints can induce auto-correlation in the chain of samples. This difficulty can be present in a different way for the optimization-based methods reviewed in Section 1: optimization routines might require many different starting points to avoid local optima. Hazelton (2000) reports such experience when using a generalized least squares method.

The model can be enhanced in many ways. Even though we have not discussed models in which $p_{s,o} = 0$ and/or where link count observations are free of errors, it is relatively straightforward to modify the Metropolis-within-Gibbs sampler to accommodate these specifications. In general, it might be possible to sample more efficiently by devising a better proposal for the Metropolis-Hastings step—not necessarily independent—according to the count likelihoods L_i.

The methodology can also be extended to more comprehensive models that consider, for example, dynamic routing, in which case network flows and observed link counts are indexed by time, and joint SUE parameter inference. We are currently planning on applying the proposed model to a large-scale, real case scenario and intend to later report our findings on model adequacy, methodology results, and possible alternative solutions.

REFERENCES

Bell, M. (1991). The estimation of origin-destination matrices by constrained generalized least squares. *Transportation Research Part B*, 25(1):13–23.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B*, 18(4–5):289–299.

Cascetta, E. and Nguyen, S. (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B*, 22(6):437–455.

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

Daganzo, C. and Sheffi, Y. (1977). On stochastic models of traffic assignment. *Transportation Science*, 11:253–274.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall, second edition.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gilks, W., Richardson, S., and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, first edition.

Givens, G. H. and Hoeting, J. A. (2005). *Computational Statistics*. Wiley-Interscience, first edition.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hazelton, M. L. (2000). Estimation of origin-destination matrices from link flows on uncongested networks. *Transportation Research Part B*, 34:549–566.

Hazelton, M. L. (2001). Inference for origin-destination matrices: Estimation, prediction and reconstruction. *Transportation Research Part B*, 35:667–676.

Kolaczyk, E. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer, first edition.

Lo, H., Zhang, N., and Lam, W. H. K. (1996). Estimation of an origin-destination matrix with random link choice proportions: A statistical approach. *Transportation Research Part B*, 30:309–324.

Maher, M. (1983). Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. *Transportation Research Part B*, 20(6):435–447.

Praskher, J. and Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: A review. *Transportation Reviews*, 24:437–463.

Robillard, P. (1975). Estimating the O-D matrix from observed link volumes. *Transportation Research*, 9:123–128.

Tebaldi, C. and West, M. (1998). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, 93(442):557–573.

Van Zuylen, H. and Willumsen, L. (1980). Network tomography: the most likely trip matrix estimated from traffic counts. *Transportation Research Part B*, 14(3):291–293.

Vardi, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 91:365–377.

Zhang, Y., Roughan, M., Lund, C., and Donoho, D. (2003). An information-theoretic approach to traffic matrix estimation. In *Proc. SIGCOMM'03*.