Travel Time Estimation for Ambulances using Bayesian Data Augmentation

Bradford S. Westgate*, Dawn B. Woodard, David S. Matteson, Shane G. Henderson

Cornell University

August 14, 2012

Abstract

Estimates of ambulance travel times on road networks are critical for effective ambulance base placement and real-time ambulance dispatching. We introduce new methods for estimating the distribution of travel times on each road segment in a city, using Global Positioning System (GPS) data recorded during ambulance trips. Our preferred method uses a Bayesian model of the ambulance trips and GPS data. Due to sparseness and error in the GPS data, the exact ambulance paths and travel times on each road segment are unknown. To estimate the travel time distributions using the GPS data, we must also estimate each ambulance path. This is known as the map-matching problem. We simultaneously estimate the unknown paths, travel times, and the parameters of each road segment travel time distribution using Bayesian data augmentation. We also introduce two alternative estimation methods based on GPS speed data that are simple to implement in practice.

We compare the predictive accuracy of the three methods to a recently-published method, using simulated data and data from Toronto EMS. In both cases, out-of-sample point and interval estimates of ambulance trip times from the Bayesian method outperform estimates from the alternative methods. We also construct probability-of-coverage maps, which are essential for ambulance providers. The Bayesian method gives more reasonable maps than the competing method. Finally, map-matching estimates from the Bayesian method interpolate well between sparsely recorded GPS readings and are robust to GPS location errors.

Keywords: Reversible jump, Markov chain, map matching, Global Positioning System, emergency medical services

*Corresponding author. Email: bsw62@cornell.edu

1 Introduction

Emergency medical service (EMS) providers prefer to assign the closest available ambulance to respond to a new emergency [6]. Thus, it is vital to have accurate estimates of the travel time of each ambulance to the emergency location. An ambulance is often assigned to a new emergency while away from its base [6], so the problem is more difficult than estimating response times from several fixed bases. Travel times also play a central role in positioning bases and parking locations [3, 12, 14]. Accounting for variability in travel times can lead to considerable improvements in EMS management [7, 15]. We introduce methods for estimating the distribution of travel times for arbitrary routes on a municipal road network, using historical trip durations and vehicle Global Positioning System (GPS) readings. This enables estimation of fastest paths in expectation between any two locations, as well as estimation of the probability an ambulance will reach its destination within a given time threshold.

Most EMS providers record ambulance GPS information; we use data from Toronto EMS from 2007-2008. The GPS data include locations, timestamps, speeds, and vehicle and emergency incident identifiers. Readings are stored every 200 meters (m) or 240 seconds (s), whichever comes first. The true sampling rate is higher, but this scheme minimizes data transmission and storage. This is standard practice across EMS providers, though the storage rates vary [19]. In related applications the GPS readings can be even sparser; Lou et al. [17] analyzed data from taxis in Tokyo in which GPS readings are separated by 1-2 km or more.

The GPS location and speed data are also subject to error. Location accuracy degrades in "urban canyons," where GPS satellites may be obscured and signals reflected [5, 19, 27]. Chen et al. [5] observed average location errors of 27 m in parts of Hong Kong with narrow streets and tall buildings, with some errors over 100 m. Location error is also present in the Toronto data; see Figure 1. Witte and Wilson [31] found GPS speed errors of roughly 5% on average, with largest error at high speeds and when few GPS satellites were visible.

Recent work on estimating ambulance travel time distributions has been done by Budge et al. [4] and Aladdini [1], using estimates based on total trip distance and time, not GPS data.

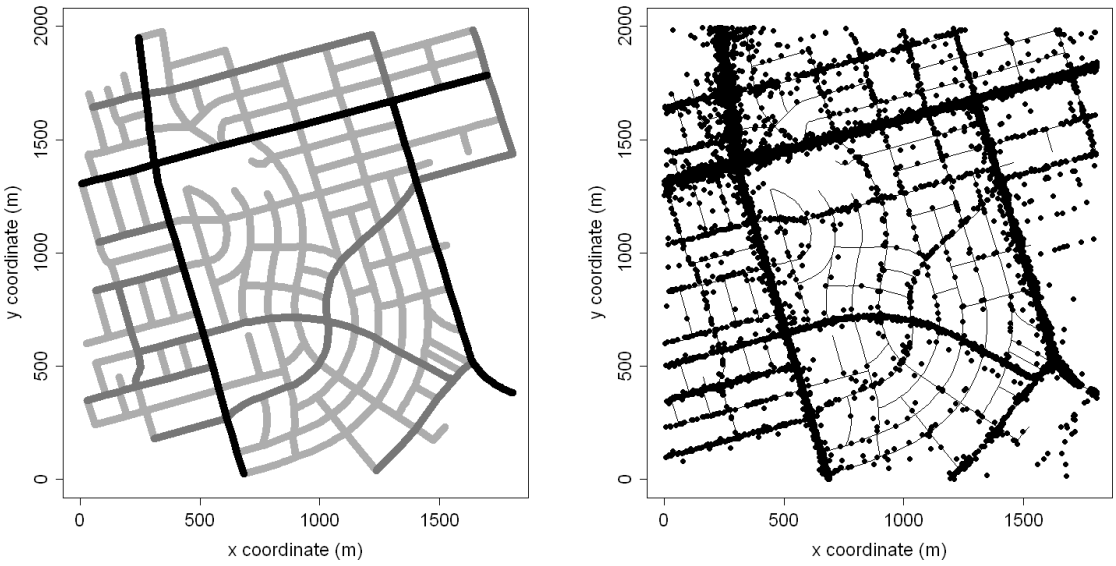

Figure 1: Left: A subregion of Toronto, with primary roads (black), secondary roads (gray) and tertiary roads (light gray). Right: GPS data on this region from the Toronto EMS "standard travel" dataset.

Budge et al. proposed modeling the log travel times using a t-distribution, where the median and coefficient of variation are functions of the trip distance (see Section 4.2). Aladdini found that the lognormal distribution provided a good fit for ambulance travel times between specific start and end locations [1]. Neither of these papers considered travel times on individual road segments. For this reason they cannot capture some desired features, such as faster response times to locations near major roads (see Section 7.4)

We first introduce two "local" methods using only the GPS locations and speeds (Section 4.1). Each GPS reading is mapped to the nearest road segment (the section of road between neighboring intersections), and the mapped speeds are used to estimate the travel time for each segment. We call these methods "local" because they do estimation independently for each segment. In the first, we use the harmonic mean of the mapped GPS speeds to create a point estimator of the travel time. We are the first to propose this estimator for mapped GPS data, though it is commonly used for estimating travel times using speed data recorded by loop detectors [22, 25, 30]. We give theoretical results supporting this approach in Appendix B. This method also yields natural interval and distribution estimates of the travel time. In our

second local method, we assume a parametric distribution for the GPS speeds on each segment, and calculate maximum likelihood estimates (MLEs) of the parameters of this distribution. These can be used to obtain point, interval, or distribution estimates of the travel time.

In Sections 2 and 3, we propose a more sophisticated method, modeling the data at the trip level. Whereas the local methods use only GPS data and the method of Budge et al. uses only the trip start and end locations and times, this method combines the two sources of information. We simultaneously estimate the path taken for each ambulance trip (solving the "map-matching problem" [19]) and the distribution of travel times on each road segment, using Bayesian data augmentation [28]. For computation, we introduce a reversible-jump Markov chain Monte Carlo method [13]. Although parameter estimation is more computationally intensive than for the other methods, prediction is very fast. Also, the parameter estimates are updated offline, so increased computation time for this stage is not an operational handicap.

We compare the predictive accuracy on out-of-sample trips for the Bayesian method, the local methods, and the method of Budge et al. on a subregion of Toronto, using simulated data and real data (Sections 6 and 7). Since the methods have some bias due in part to the GPS sampling scheme, we first use a correction factor to make each method approximately unbiased (Section 5). On simulated data, point estimates from the Bayesian method outperform the alternative methods by over 50% (in root mean squared error), relative to an "Oracle" method with the lowest possible error. On real data, point estimates from the Bayesian method again outperform the alternative methods. Interval estimates from the Bayesian method have dramatically better coverage than intervals from the local methods.

We also produce probability-of-coverage maps [4], showing the probability of traveling from a given intersection to any other intersection within a time threshold (Section 7.4). This is the performance standard in many EMS contracts; a typical EMS organization attempts to respond to, e.g., 90% of all emergencies within 9 minutes [8]. The estimates from the Bayesian method are more reasonable than those from the method of Budge et al.

Finally, we assess the map-matching solutions from the Bayesian method (Sections 6.3 and

7.5). Unlike most map-matching techniques that analyze each trip in isolation [16, 17, 18, 19], the entire dataset of trips is used to produce the path estimate for each trip. Also, the posterior distribution over paths can capture multiple high-probability paths when the true path is unclear from the GPS data. Our path estimates interpolate accurately between widely-separated GPS points and are robust to GPS error.

2 Bayesian Formulation

2.1 Model

Consider a network of J directed road segments, called "arcs," and a set of I ambulance trips on this network. Assume that each trip i starts and finishes on known nodes (intersections) d_i^s and d_i^f in the network, at known times t_i^s and t_i^f (so the total travel time $t_i^f - t_i^s$ is known). In practice, trips sometimes begin or end in the interior of a road segment; however, road segments are short enough that this is a minor issue (the median road segment length in the Toronto network is 111 m, the mean is 162 m, and the maximum is 4613 m). The data associated with each trip i consist of the observed GPS readings, indexed by $\ell \in \{1, \ldots, r_i\}$, and gathered at fixed times t_i^ℓ. GPS reading ℓ is the triplet $\left(X_i^\ell, Y_i^\ell, V_i^\ell\right)$, where X_i^ℓ and Y_i^ℓ are the measured geographic coordinates and V_i^ℓ is the measured speed. Denote $G_i = \left\{\left(X_i^\ell, Y_i^\ell, V_i^\ell\right)\right\}_{\ell=1}^{r_i}$.

The relevant unobserved variables for each trip i are the following:

1. The unknown path (sequence of arcs) $A_i = \{A_{i,1}, \ldots, A_{i,N_i}\}$ traveled by the ambulance from d_i^s to d_i^f. The path length N_i is also unknown.

2. The unknown travel times $T_i = (T_{i,1}, \ldots, T_{i,N_i})$ on the arcs in the path. We also use the notation $T_i(j)$ to refer to the travel time in trip i on arc j.

We model the observed and unobserved variables $\{A_i, T_i, G_i\}_{i=1}^I$ as follows. Conditional on A_i, each element $T_{i,k}$ of the vector T_i follows a lognormal distribution with parameters

$\mu_{A_{i,k}}, \sigma^2_{A_{i,k}}$, independently across i and k. Specifically,

$$T_{i,k}|A_i \sim \mathcal{LN}\left(\mu_{A_{i,k}}, \sigma^2_{A_{i,k}}\right) = \frac{1}{T_{i,k}\sqrt{2\pi\sigma^2_{A_{i,k}}}} \exp\left(-\frac{\left(\log T_{i,k} - \mu_{A_{i,k}}\right)^2}{2\sigma^2_{A_{i,k}}}\right). \tag{1}$$

In the literature, ambulance travel times between specific locations have been observed and modeled to be lognormal [1, 2]. While Budge et al. [4] found log t-distributions to be a better fit, we hypothesize that this was because they did not condition on the trip location. Denote the expected travel time on each arc $j \in \{1, \ldots, J\}$ by $\theta(j) = \exp\left(\mu_j + \sigma_j^2/2\right)$. We use a multinomial logit choice model [20] for the path A_i, with likelihood

$$f(A_i) = \frac{\exp\left(-C\sum_{k=1}^{N_i}\theta\left(A_{i,k}\right)\right)}{\sum_{a_i \in \mathcal{P}_i}\exp\left(-C\sum_{k=1}^{n_i}\theta\left(a_{i,k}\right)\right)}, \tag{2}$$

where $C > 0$ is a fixed constant and $\mathcal{P}_i$ is the set of possible paths from d_i^s to d_i^f (with no repeated nodes) in the network, and $a_i = \{a_{i,1}, \ldots, a_{i,n_i}\}$ indexes the paths in $\mathcal{P}_i$. This model captures the fact that the fastest routes in expectation have the highest probability.

We assume that ambulances travel at constant speed on a single arc in a given trip. This is a necessary approximation since there is typically at most one GPS reading on any arc in a given trip, and thus very little information in the data regarding changes in speed on individual arcs. Thus, the true location and speed of the ambulance at time t_i^ℓ are deterministic functions $\text{loc}\left(A_i, T_i, t_i^\ell\right)$ (short for "location") and $\text{speed}\left(A_i, T_i, t_i^\ell\right)$ of A_i and T_i. Conditional on A_i, T_i, the measured location $\left(X_i^\ell, Y_i^\ell\right)$ is assumed to have a bivariate normal distribution (a standard assumption; see [16, 19]) centered at $\text{loc}\left(A_i, T_i, t_i^\ell\right)$, with known covariance matrix Σ. Similarly, the measured speed V_i^ℓ is assumed to have a lognormal distribution with expected value equal to $\text{speed}\left(A_i, T_i, t_i^\ell\right)$ and variance parameter ζ^2:

$$\left(X_i^\ell, Y_i^\ell\right) \Big| A_i, T_i \sim N_2\left(\text{loc}\left(A_i, T_i, t_i^\ell\right), \Sigma\right), \tag{3}$$

$$\log V_i^\ell \Big| A_i, T_i \sim N\left(\log \text{speed}\left(A_i, T_i, t_i^\ell\right) - \frac{\zeta^2}{2}, \zeta^2\right). \tag{4}$$

We assume independence between all the GPS speed and location errors. Combining Equations 1-4, we obtain the likelihood

$$f\left(\left\{A_i, T_i, G_i\right\}_{i=1}^{I} \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}, \zeta^2\right) = \prod_{i=1}^{I}\left[f(A_i)\prod_{k=1}^{N_i}\mathcal{LN}\left(T_{i,k}; \mu_{A_{i,k}}, \sigma^2_{A_{i,k}}\right)\right.$$
$$\left.\prod_{\ell=1}^{r_i}\left[N_2\left(\left(X_i^\ell, Y_i^\ell\right); \mathrm{loc}\left(A_i, T_i, t_i^\ell\right), \Sigma\right) \times \mathcal{LN}\left(V_i^\ell; \log \mathrm{speed}\left(A_i, T_i, t_i^\ell\right) - \frac{\zeta^2}{2}, \zeta^2\right)\right]\right]. \quad (5)$$

In practice we use data-based choices for the constants Σ and C (see Appendix A). The unknown parameters in the model are the arc travel time parameters $\left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}$ and the GPS speed error parameter ζ^2.

2.2 Prior Distributions

To complete the model, we specify prior distributions for the unknown parameters. We use

$$\mu_j \sim N\left(m_j, s^2\right), \qquad \sigma_j \sim \mathrm{Unif}\left(b_1, b_2\right), \qquad \zeta \sim \mathrm{Unif}\left(b_3, b_4\right), \quad (6)$$

independently, where m_j, s^2, b_1, b_2, b_3, b_4 are fixed hyperparameters. A normal prior is a standard choice for the location parameter of a lognormal distribution. We use uniform priors on the standard deviations σ_j and ζ [9]. The prior ranges $[b_1, b_2]$ and $[b_3, b_4]$ are set to be wide enough to capture all remotely plausible parameter values. The prior mean for μ_j depends on j, while the other hyperparameters do not, because there are often existing road speed estimates that can be used in the specification of m_j. Prior information regarding the values s^2, b_1, b_2, b_3, b_4 is more limited. We use a combination of prior information and the data to specify all hyperparameters, as described in Appendix A.

3 Bayesian Computational Method

We use a Markov chain method to obtain samples $\left(\zeta^{2(\ell)}, \left\{\mu_j^{(\ell)}, \sigma_j^{(\ell)}\right\}_{j=1}^{J}, \left\{A_i^{(\ell)}, T_i^{(\ell)}\right\}_{i=1}^{I}\right)$ from the joint posterior distribution of all unknowns [23, 29]. Each unknown quantity is updated in turn, conditional on the other unknowns, via either a draw from the closed-form conditional posterior distribution or a Metropolis-Hastings move (M-H). Estimation of any desired function $g\left(\zeta^2, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}\right)$ of the unknown parameters is done via Monte Carlo, taking $\hat{g} = \frac{1}{M}\sum_{\ell=1}^{M} g\left(\zeta^{2(\ell)}, \left\{\mu_j^{(\ell)}, \sigma_j^{2(\ell)}\right\}_{j=1}^{J}\right)$.

3.1 Markov Chain Initial Conditions

To initialize each path A_i, select the "middle" GPS reading, reading number $\lfloor r_i/2 \rfloor + 1$. Find the nearest node in the road network to this GPS location, and route the initial path A_i through this node, taking the shortest-distance path to and from the middle node. To initialize the travel time vector T_i, distribute the known trip time across the arcs in the path A_i, weighted by arc length. Finally, to initialize ζ^2 and each μ_j and σ_j^2, draw from their priors.

3.2 Updating the Paths

Updating the path A_i may also require updating the travel times T_i, since the number of arcs in the path may change. Since this changes the dimension of the vector T_i, we update (A_i, T_i) using a reversible-jump M-H move [13]. Calling the current values $\left(A_i^{(1)}, T_i^{(1)}\right)$, we propose new values $\left(A_i^{(2)}, T_i^{(2)}\right)$ and accept them with the appropriate probability, detailed below.

The proposal changes a contiguous subset of the path. The length (number of arcs) of this subpath is limited to some maximum value K; K is specified in Section 3.5. Precisely:

1. With equal probability, choose a node d' from the path $A_i^{(1)}$, excluding the final node.
2. Let $a^{(1)}$ be the number of nodes that follow d' in the path. With equal probability, choose an integer $w \in \{1, \ldots, \min\left(a^{(1)}, K\right)\}$. Denote the wth node following d' as d''. The subpath from d' to d'' is the section to be updated (the "current update section").

3. Consider all possible routes of length up to K from d' to d''. With equal probability, propose one of these routes as a change to the path (the "proposed update section"), giving the proposed path $A_i^{(2)}$.

Next we propose travel times that are compatible with $A_i^{(2)}$. Let $\{c_1, \ldots, c_m\} \subset A_i^{(1)}$ and $\{p_1, \ldots, p_n\} \subset A_i^{(2)}$ denote the arcs in the current and proposed update sections, noting that m and n may be different. Recall that $T_i(j)$ denotes the travel time of trip i on arc j. For each arc $j \in A_i^{(2)} \setminus \{p_1, \ldots, p_n\}$, set $T_i^{(2)}(j) = T_i^{(1)}(j)$. Let $S_i = \sum_{\ell=1}^{m} T_i^{(1)}(c_\ell)$ be the total travel time of the current update section. We must have $\sum_{\ell=1}^{n} T_i^{(2)}(p_\ell) = S_i$ also, because the total travel time of the trip is known. The travel times $T_i^{(2)}(p_1), \ldots, T_i^{(2)}(p_n)$ are proposed by drawing $(r_1, \ldots, r_n) \sim \text{Dirichlet}\left(\alpha\theta(p_1), \ldots, \alpha\theta(p_n)\right)$ for a constant $\alpha > 0$ (specified below), and setting $T_i^{(2)}(p_\ell) = r_\ell S_i$ for $\ell \in \{1, \ldots, n\}$. The expected value of the proposed travel time on arc p_ℓ is $E\left(T_i^{(2)}(p_\ell)\right) = S_i \frac{\theta(p_\ell)}{\sum_{k=1}^{n} \theta(p_k)}$. Therefore, the expected values of the proposed times are weighted by the arc travel time expected values [10]. The constant α controls the variance of each component $T_i^{(2)}(p_\ell)$. In our experience $\alpha = 1$ works well; one can also tune α to obtain a desired acceptance rate for a particular dataset [23, 24].

Let $N_i^{(j)}$ be the number of edges in the path $A_i^{(j)}$ for $j \in \{1, 2\}$, and let $a^{(2)}$ be the number of nodes that follow d' in the path $A_i^{(2)}$. We accept the proposed values $(A_i^{(2)}, T_i^{(2)})$ with probability equal to the minimum of one and

$$\frac{f_i\left(A_i^{(2)}, T_i^{(2)}, G_i \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}, \zeta^2\right)}{f_i\left(A_i^{(1)}, T_i^{(1)}, G_i \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}, \zeta^2\right)} \quad \times$$
$$\frac{N_i^{(1)} \min(a^{(1)}, K)}{N_i^{(2)} \min(a^{(2)}, K)} \frac{\text{Dir}\left(\frac{T_i^{(1)}(c_1)}{S_i}, \ldots, \frac{T_i^{(1)}(c_m)}{S_i}; \alpha\theta(c_1), \ldots, \alpha\theta(c_m)\right)}{\text{Dir}\left(\frac{T_i^{(2)}(p_1)}{S_i}, \ldots, \frac{T_i^{(2)}(p_n)}{S_i}; \alpha\theta(p_1), \ldots, \alpha\theta(p_n)\right)} S_i^{n-m} \tag{7}$$

where f_i is the contribution of trip i to Equation 5 and $\text{Dir}(x; y)$ denotes the Dirichlet density with parameter vector y, evaluated at x. In Appendix C we show that this move is valid since it is reversible with respect to the conditional posterior distribution of (A_i, T_i).

3.3 Updating the Trip Travel Times

To update the realized travel time vector $T_i(j)$, we use the following M-H move. Given current travel times $T_i^{(1)}$, we propose travel times $T_i^{(2)}$.

1. With equal probability, choose a pair of distinct arcs j_1 and j_2 in the path A_i. Let $S_i = T_i^{(1)}(j_1) + T_i^{(1)}(j_2)$.
2. Draw $(r_1, r_2) \sim \text{Dirichlet}\left(\alpha'\theta(j_1), \alpha'\theta(j_2)\right)$. Set $T_i^{(2)}(j_1) = r_1 S_i$ and $T_i^{(2)}(j_2) = r_2 S_i$.

Similarly to the path proposal (Section 3.2), this proposal randomly distributes the travel time over the two arcs, weighted by the expected travel times $\theta(j_1)$ and $\theta(j_2)$, with variances controlled by the constant α' [10]. In our experience $\alpha' = 0.5$ is effective for our application. The M-H acceptance probability may be calculated in a similar manner as in Appendix C.

3.4 Updating the Parameters μ_j, σ_j^2, and ζ^2

To update each μ_j, we sample from the full conditional posterior distribution, which is available in closed form. We have $\mu_j \left| \sigma_j^2, \{A_i, T_i\}_{i=1}^I \sim N\left(\hat{\mu}_j, \hat{s}_j^2\right)\right.$, where

$$\hat{s}_j^2 = \left[\frac{1}{s^2} + \frac{n_j}{\sigma_j^2}\right]^{-1}, \qquad \hat{\mu}_j = \hat{s}_j^2 \left[\frac{m_j}{s^2} + \frac{1}{\sigma_j^2}\sum_{i \in I_j} \log T_i(j)\right],$$

the index set $I_j \subset \{1, \ldots, I\}$ indicates the subset of trips using arc j, and $n_j = |I_j|$.

To update each σ_j^2, we use a local M-H step [29]. We propose $\sigma_j^{2*} \sim \mathcal{LN}(\log \sigma_j^2, \eta^2)$, having fixed variance η^2. Using Equations 1 and 6, the M-H acceptance probability is

$$p_\sigma = \min\left\{1, \frac{\sigma_j}{\sigma_j^*}\mathbf{1}_{\{\sigma_j^* \in [b_1, b_2]\}} \left(\frac{\prod_{i \in I_j} \mathcal{LN}\left(T_i(j); \mu_j, \sigma_j^{2*}\right)}{\prod_{i \in I_j} \mathcal{LN}\left(T_i(j); \mu_j, \sigma_j^2\right)}\right) \frac{\mathcal{LN}\left(\sigma_j^2; \log\left(\sigma_j^{2*}\right), \eta^2\right)}{\mathcal{LN}\left(\sigma_j^{2*}; \log\left(\sigma_j^2\right), \eta^2\right)}\right\}.$$

To update ζ^2, we use another M-H step with a lognormal proposal, with different variance ν^2. The acceptance probability may be calculated similarly. The proposal variances η^2, ν^2 are tuned to achieve an acceptance rate of approximately 23% [24].

3.5 Markov Chain Convergence

The transition kernel for updating the path A_i is irreducible, and hence valid [29], if it is possible to move between any two paths in $\mathcal{P}_i$ in a finite number of iterations, for all i. For a given road network, the maximum update section length K can be set high enough to meet this criterion. However, the value of K should be set as low as possible, because increasing K tends to lower the acceptance rate. If there is a region of the city with sparse connectivity, the required value of K may be impractically large. For example, if a highway is parallel to a minor road, there could be many arcs of the minor road alongside a single arc of the highway. Then, a large K would be needed to allow transitions between the highway and the minor road. If K is kept smaller, the Markov chain is not irreducible. In this case, the chain converges to the posterior distribution restricted to the closed communicating class in which the chain is absorbed. If this class contains much of the posterior mass, as might arise if the initial path follows the GPS data reasonably closely, then this should be a good approximation.

In Sections 6 and 7, we apply the Bayesian method to simulated data and data from Toronto EMS, on a subregion of Toronto with 623 arcs. Each Markov chain was run for 50,000 iterations (where each iteration updates all parameters), after a burn-in period of 25,000 iterations. We calculated Gelman-Rubin diagnostics [11], using two chains, for the parameters ζ^2, μ_j, and σ_j^2. Results from a typical simulation study were: potential scale reduction factor of 1.06 for ζ^2, of less than 1.1 for μ_j for 549 arcs (88.1%), between 1.1-1.2 for 43 arcs (6.9%), between 1.2-1.5 for 30 arcs (4.8%), and less than 2 for the remaining 1 arc, with similar results for the parameters σ_j^2. These results indicate no lack of convergence.

Each Markov chain run for these experiments takes roughly 2 hours on a 3.2 GHz workstation. By contrast, the calculation of the local method estimates (Section 4.1) for all arcs takes roughly 0.1 s. However, once parameter estimation is done (in practice estimates are updated infrequently and offline), prediction for new routes and generation of our graphical displays is virtually instantaneous. If parameter estimation for the Bayesian method is computationally impractical for the entire city, it can be divided into several regions and estimated in parallel.

The full Toronto road network has roughly 110 times as many arcs as the test region, and the full Toronto EMS dataset has roughly 80 times as many ambulance trips.

4 Comparison Methods

4.1 Local Methods

Here we detail the two "local" methods outlined in Section 1. Each GPS reading is mapped to the nearest arc (both directions of travel are treated together). Let n_j be the number of GPS points mapped to arc j, L_j the length of arc j, and $\left\{V_j^k\right\}_{k=1}^{n_j}$ the mapped speed observations. We assume constant speed on each arc, as in the Bayesian method (Section 2.1). Thus, let $T_j^k = L_j/V_j^k$ be the travel time associated with observed speed V_j^k.

In the first local method, we calculate the harmonic mean of the speeds $\left\{V_j^k\right\}_{k=1}^{n_j}$, and convert to a travel time point estimate

$$\hat{T}_j^H = \frac{L_j}{n_j} \sum_{k=1}^{n_j} \frac{1}{V_j^k}.$$

This is equivalent to calculating the arithmetic mean of the associated travel times T_j^k. The empirical distribution of the associated times $\left\{T_j^k\right\}_{k=1}^{n_j}$ can be used as a distribution estimate. Because readings with speed 0 occur in the Toronto EMS dataset, we set any reading with speed below 5 miles per hour (mph) equal to 5 mph. This harmonic mean estimator is well-known in the transportation research literature (where it is called the "space mean speed") in the context of estimating travel times using speed data recorded by loop detectors [22, 25, 30].

In Appendix B, we consider this travel time estimator $\hat{T}_j^H$ and its relation to the GPS sampling scheme. We show that if GPS points are sampled by distance (for example, every 100 m), $\hat{T}_j^H$ is an unbiased and consistent estimator for the true mean travel time. However, if GPS points are sampled by time (for example, every 30 s), $\hat{T}_j^H$ overestimates the mean travel time. The Toronto EMS dataset uses a combination of sampling-by-distance and sampling-

by-time. However, the distance constraint is usually satisfied first (see Figure 5, where the sampled GPS points are regularly spaced). Thus, the travel time estimator $\hat{T}_j^H$ is appropriate.

In the second local method, we assume $V_j^k \sim \mathcal{LN}(m_j, s_j^2)$, independently across k, for unknown travel time parameters m_j and s_j^2. This distribution on the travel speed implies a related lognormal distribution on the travel time. Specifically, $T_j^k \sim \mathcal{LN}\left(\log(L_j) - m_j, s_j^2\right)$. We use the maximum likelihood estimators

$$\hat{m}_j = \frac{1}{n_j}\sum_{k=1}^{n_j} \log\left(V_j^k\right), \qquad \hat{s}_j^2 = \frac{1}{n_j}\sum_{k=1}^{n_j}\left(\log\left(V_j^k\right) - \hat{m}_j\right)^2$$

to estimate m_j and s_j^2. Our point travel time estimator is

$$\hat{T}_j^{\text{MLE}} = E\left(T_j | \, \hat{m}_j, \hat{s}_j^2\right) = \exp\left(\log(L_j) - \hat{m}_j + \frac{\hat{s}_j^2}{2}\right).$$

This second local method also provides a natural distribution estimate for the travel times via the estimated lognormal distribution for T_j^k. Correcting for zero-speed readings is again done by thresholding, to avoid $\log(0)$.

Some small residential arcs have no assigned GPS points in the Toronto EMS dataset (see Figure 1). In this case, we use a breadth-first search [21] to find the closest arc in the same "road class" that has assigned GPS points. The road classes are described in Section 6; by restricting our search to arcs of the same class we ensure that the speeds are comparable.

4.2 Method of Budge et al.

Budge et al. [4] introduced a travel time distribution estimation method relying on trip distance. Since the exact path traveled is usually unknown, the length of the shortest distance path between the start and end locations is used as a surrogate for this distance. The method relies on the model $t_i = m(d_i)\exp(c(d_i)\epsilon_i)$, where t_i and d_i are the total time and distance for trip i, ϵ_i follows a t-distribution with τ degrees of freedom, and $m(\cdot)$ and $c(\cdot)$ are unknown functions. In their preferred method, they assume parametric expressions for the functions

$m(\cdot)$ and $c(\cdot)$, and estimate the parameters using maximum likelihood.

We implemented this parametric method and compared it to a related binning method. In the binning method, we divide the ambulance trips into bins by trip distance, and fit a separate t-distribution to the log travel times for each bin. We then linearly interpolate between the quantiles of the travel time distributions for adjacent bins, to generate a travel time distribution estimate for a trip of any distance. On simulated data, the parametric and binning methods perform very similarly, while on real data the binning method outperforms the parametric method. Thus, we report only results of the binning method in Sections 6-7.

5 Bias Correction

We use a bias correction factor to make each method approximately unbiased, because we have found that this improves performance for all methods. There are several reasons why the methods result in biased estimates, some inherent to the methods themselves and some due to sampling characteristics of the GPS data. One source of bias is the inspection paradox in the GPS data, discussed at length in Appendix B. The Bayesian method is also biased because of the difference in path estimation from the training to the test data. On the training data, the Bayesian method uses the GPS data to estimate a solution to the map-matching problem. On the test data, the estimated fastest path between the start and end nodes is used, instead of the GPS data (to imitate the prediction scenario where the route is not known beforehand). This leads to underestimation of the true travel times.

The bias correction factor for each method is calculated in the following manner. We divide the set of trips from each dataset randomly into training, validation, and test sets. We fit the methods on the training data, calculate a bias correction factor on the validation data, and predict the travel times for the trips in the test data. The data are split into 50% training and 50% validation and test. We then use a cross-validation approach: divide the validation/test data into ten sets, use nine sets for the validation data, the tenth for the test data, and repeat

for all ten cases. For a given validation set of n trips, where the estimated trip travel times are $\{\hat{t}_i\}_{i=1}^n$ and the true travel times are $\{t_i\}_{i=1}^n$, the bias correction factor is

$$b = \frac{1}{n}\left(\sum_{i=1}^{n} \log \hat{t}_i - \sum_{i=1}^{n} \log t_i\right)$$

Subtracting this factor from the log estimates on the test data makes each method unbiased on the log scale. We calculate the bias correction on the log scale because it is more robust to travel times outliers.

6 Simulation Experiments

Next we test the Bayesian method, local methods, and the method of Budge et al. on simulated data. We compare the accuracy of the four methods for predicting travel times of test trips. We simulate ambulance trips on the road network of Leaside, Toronto, shown in Figure 1 (roughly 4 square kilometers). This region has four road classes; we define the highest-speed class to be "primary" arcs, the two intermediate classes to be "secondary" arcs, and the lowest-speed class to be "tertiary" arcs (Figure 1). In the Leaside region, a value $K = 6$ (see Section 3.5) guarantees that the Markov chain is valid.

6.1 Generating Simulated Data

We simulate data as follows. We generate ambulance trips with true paths, travel times, and GPS readings. For each trip i, we uniformly choose start and end nodes. We then construct the true path A_i arc-by-arc. At each node, beginning at the start node, we uniformly choose an adjacent arc out of those that lower the expected time to the end node. We repeat this process until the end node is reached. This method differs from the Bayesian prior (see Section 2.1), and can lead to a wide variety of paths traveled between two nodes.

The arc travel times are lognormal: $T_{i,k} \sim \mathcal{LN}(\mu_{A_{i,k}}, \sigma^2_{A_{i,k}})$. To set the true travel time parameters $\left\{\mu_j, \sigma_j^2\right\}$ for arc j, we uniformly generate a speed between 20-40 mph. We set μ_j

and σ_j^2 so that the arc length divided by the mean travel time equals this speed. To give the arcs a range of travel time variances, we draw $\sigma_j \sim \text{Unif}\left(0.5\log\left(\sqrt{3}\right), 0.5\log(3)\right)$. These two constraints define the required value of μ_j. The range for σ_j is narrower than the prior range (see Appendix A), but still generates a wide variety of arc travel time variances. Comparisons between the estimation methods are invariant to moderate changes in the σ_j range.

We simulate datasets with two types of GPS data: "good" and "bad." The "good" GPS datasets are designed to mimic the conditions of the Toronto EMS dataset. Each GPS point is sampled at a travel distance of 250m after the previous point (straight-line distance is 200m in the Toronto EMS data, but we simulate data via the longer along-path distance). The GPS locations are drawn from a bivariate normal distribution with $\Sigma = \left(\begin{smallmatrix} 100 & 0 \\ 0 & 100 \end{smallmatrix}\right)$. The GPS speeds are drawn from a lognormal distribution with $\zeta^2 = 0.004$, which gives a mean absolute error of 5% of speed, approximately the average result seen by Witte and Wilson [31].

The "bad" GPS datasets are designed to be sparse and have GPS error consistent with the high error results seen by Chen et al. [5] and Witte and Wilson [31]. GPS points are sampled every 1000m, which is still more frequent than the rate observed in the Tokyo taxi data [17]. The constant $\Sigma = \left(\begin{smallmatrix} 465 & 0 \\ 0 & 465 \end{smallmatrix}\right)$, which gives mean distance of 27 m between the true and observed locations, the average error seen in Hong Kong by Chen et al. [5]. The parameter $\zeta^2 = 0.01575$, corresponding to mean absolute error of 10% of speed, which is approximately the result from low-quality GPS settings tested by Witte and Wilson [31].

6.2 Travel Time Prediction

We simulate ten good GPS datasets and ten bad GPS datasets, as defined in Section 6.1, each with a training set of 2000 trips and a validation/test set of 2000 trips. Taking the true path for each test trip as known and using the cross-validation approach of Section 5 to estimate bias correction factors, we calculate point and 95% predictive interval estimates for the test set travel times using the four methods. To obtain a "gold standard" for performance, we implement an "Oracle" method. In this method, the true travel time parameters $\left\{\mu_j, \sigma_j^2\right\}$

for each arc j are known. The true expected travel time for each test trip is used as a point estimate. This implies that the Oracle method has the lowest possible root mean squared error (RMSE) for realized travel time estimation.

We compare the predictive accuracy of the point estimates from the four methods via the RMSE (in seconds), the RMSE of the log predictions relative to the true log times ("RMSE log"), and the mean absolute bias on the log scale over the test sets of the cross-validation procedure ("Bias M.A."). We calculate metrics on the log scale because the residuals on the log scale are much closer to normally distributed. On the original scale, there are several outlying trips in the Toronto EMS data (Section 7) with very large travel times that heavily influence the metrics. The bias metric measures how well the bias correction works. If $k \in \{1, \ldots, 10\}$ indexes the cross-validation test sets, where test set k has n_k trips with true travel times $t_i^{(k)}$ and estimates $\hat{t}_i^{(k)}$, for $i \in \{1, \ldots, n_k\}$, then

$$\text{Bias (M.A.)} = \frac{1}{10} \sum_{k=1}^{10} \left| \frac{1}{n_k} \left(\sum_{i=1}^{n_k} \log \hat{t}_i^{(k)} - \sum_{i=1}^{n_k} \log t_i^{(k)} \right) \right|. \tag{8}$$

We compare the interval estimates using the the percentage of 95% predictive intervals that contain the true travel time ("Cov. %") and the geometric mean (arithmetic mean on the log scale) width of the 95% predictive intervals ("Width"). Table 1 gives arithmetic means for these metrics over the ten good and bad simulated datasets.

In both dataset types, the point estimates from the Bayesian method greatly outperform the estimates from the local methods and the method of Budge et al. The Bayesian estimates closely approach the Oracle estimates, especially in the good GPS datasets. In the good datasets, the Bayesian method has 70% lower error than the local methods in RMSE on the log scale, and 78% lower error than Budge et al., after eliminating the unavoidable error of the Oracle method. In the bad datasets, the Bayesian method outperforms the local methods by 70% and Budge et al. by 56% in log scale RMSE, relative to the Oracle method. The method of Budge et al. outperforms the local methods on the bad GPS data, while the reverse holds

Good GPS data (Mean over ten datasets)					
Estimation method	RMSE (s)	RMSE log	Bias (M.A.)	Cov. %	Width (s)
Oracle	15.9	0.183	0.010	-	-
Bayesian	16.1	0.187	0.010	95.8	57.2
Local MLE	16.8	0.196	0.010	94.4	56.8
Local Harm.	16.8	0.196	0.010	94.0	56.2
Budge et al.	17.3	0.201	0.011	96.2	67.2
Bad GPS data (Mean over ten datasets)					
Estimation method	RMSE (s)	RMSE log	Bias (M.A.)	Cov. %	Width (s)
Oracle	16.4	0.183	0.012	-	-
Bayesian	16.9	0.191	0.013	96.1	60.4
Local MLE	18.1	0.209	0.014	92.3	57.8
Local Harm.	18.1	0.209	0.014	90.9	55.5
Budge et al.	17.9	0.201	0.013	96.2	68.2

Table 1: Out-of-sample trip travel time estimation performance on simulated data.

for the good GPS data. The bias is low for all methods.

The Bayesian method also outperforms the other methods in interval estimates. For the good GPS data, the interval estimates from the Bayesian and local methods are similar, while the estimates from the method of Budge et al. are substantially wider, with slightly higher coverage percentage. For the bad GPS data, the intervals from the Bayesian method have higher coverage percentage than the intervals from the local methods, and the intervals from the method of Budge et al. are again wider, with no corresponding increase in coverage percentage.

6.3 Map-Matched Path Results

Next we assess map-matching estimates from the Bayesian method for representative paths, shown in Figure 2. The GPS locations are shown in white. The starting node is marked with a cross, the ending node with an X. Each arc is shaded in gray according to the marginal posterior probability that it is traversed in the path. Arcs with probability less than 1% are unshaded. The left-hand path is from a good GPS dataset (as defined in Section 6.1). The Bayesian method easily identifies the correct path. Every correct arc has close to 100% probability, and only two incorrect detours have probability above 1%. This is typical performance for simulated trips with good GPS data. The right-hand path is from a bad GPS dataset. The

sparsity in GPS readings makes the path very uncertain. Near the beginning of the path, there are five routes with similar expected travel times, and the GPS readings do not distinguish between them, so each has close to 20% posterior probability. Near the end of the path, there are two routes with roughly 50% probability. The Bayesian method is very effective at identifying alternative routes when the true path is unclear.

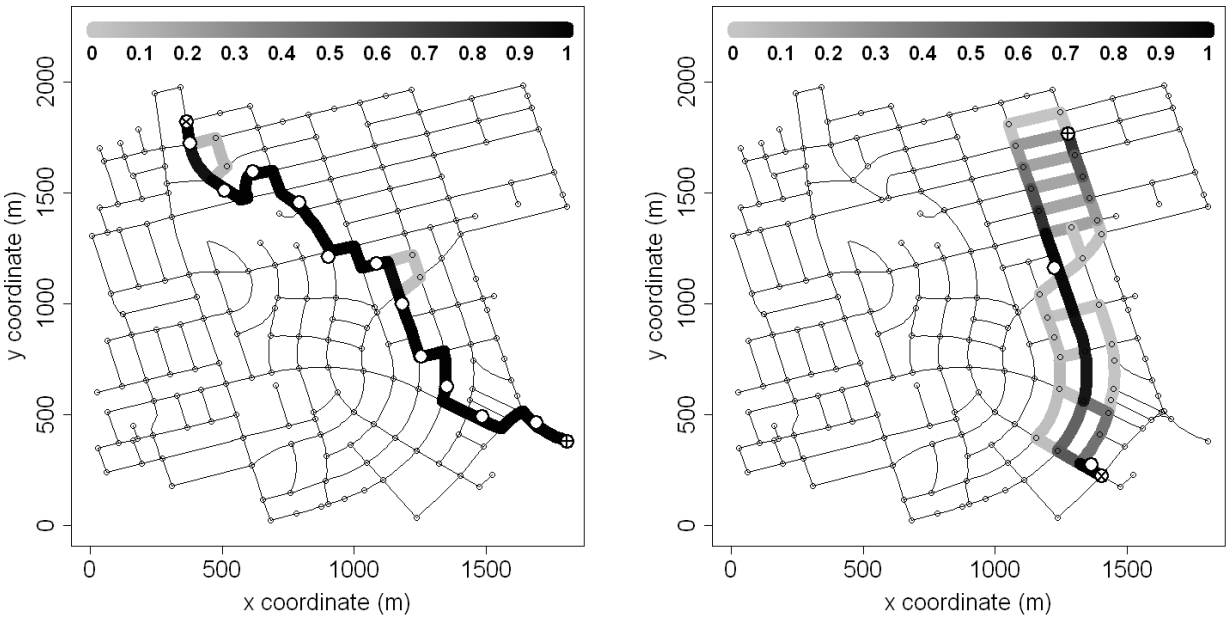

Figure 2: Map-matching estimates for two simulated trips, colored by the probability each arc is traversed.

7 Analysis of Toronto EMS Data

Next we compare all four methods on the Toronto EMS data.

7.1 Data

The Toronto data consist of GPS data and trip information for ambulance trips with one of two priority levels: "lights-and-sirens" (L-S) or "standard travel" (Std). We address these separately, again focusing on the Leaside subregion of Toronto. The right plot in Figure 1 shows the GPS locations for the Std dataset. This dataset contains 3989 ambulance trips

and almost 35,000 GPS points. The primary roads tend to have a large amount of data, the secondary roads a moderate amount, and the tertiary roads a small amount. The L-S dataset is smaller (1930 trips), with a similar spatial distribution of points.

We use only the portion of trips where the ambulance was driving to scene in response to an emergency call, and discard trips for which this cannot be identified. We also discard some trips (roughly 1%) that would impair estimation: for example, trips where the ambulance turned around or where the ambulance stopped for a long period, not at a stoplight or in traffic. Finally, most of the trips in the dataset do not begin or end in the subregion, they simply pass through, so we define start and end nodes and times as follows. We use the closest node to the first GPS location as the approximated start node, and the time of the first GPS reading as the start time. Similarly, we use the last GPS reading for the end node. This produces some inaccuracy of estimated travel times on the boundary of the region. This could be fixed by applying our method to overlapping regions and discarding estimates on the boundary.

7.2 Arc Travel Time Estimates

Here we report the travel time estimates from the Bayesian method. Toronto EMS has existing estimates of the travel times, which we use to set the prior $\{m_j\}_{j=1}^{J}$ hyperparameters (Appendix A). These estimates are different for L-S and Std trips, but are the same for the two travel directions of parallel arcs. We have also tested the Bayesian method with the data-based hyperparameters described in Appendix A and have observed similar performance. Figure 3 shows prior and posterior speed estimates (length divided by mean travel time) from the Bayesian method on the L-S dataset. Each arc is shaded in gray based on its speed estimate, so most roads have two shades in the right hand plot, corresponding to travel in each direction.

The posterior speed estimates from the Bayesian method are reasonable; primary arcs tend to have high speed estimates, and estimated speeds for consecutive arcs on the same road are typically similar. Arcs heading into major intersections (intersections between two primary or secondary roads, as shown in Figure 1) are often slower than the reverse arcs. In

the corresponding figure for Std data (not shown), the slowdown into major intersections is even more severe. This effect cannot be captured by the local methods, because both travel directions have the same estimates. Arcs with little data usually have posterior estimated speeds close to the prior estimates. For most arcs the posterior estimate of the speed is higher than the prior estimate, suggesting that the existing road speed estimates used to specify the prior are underestimates.

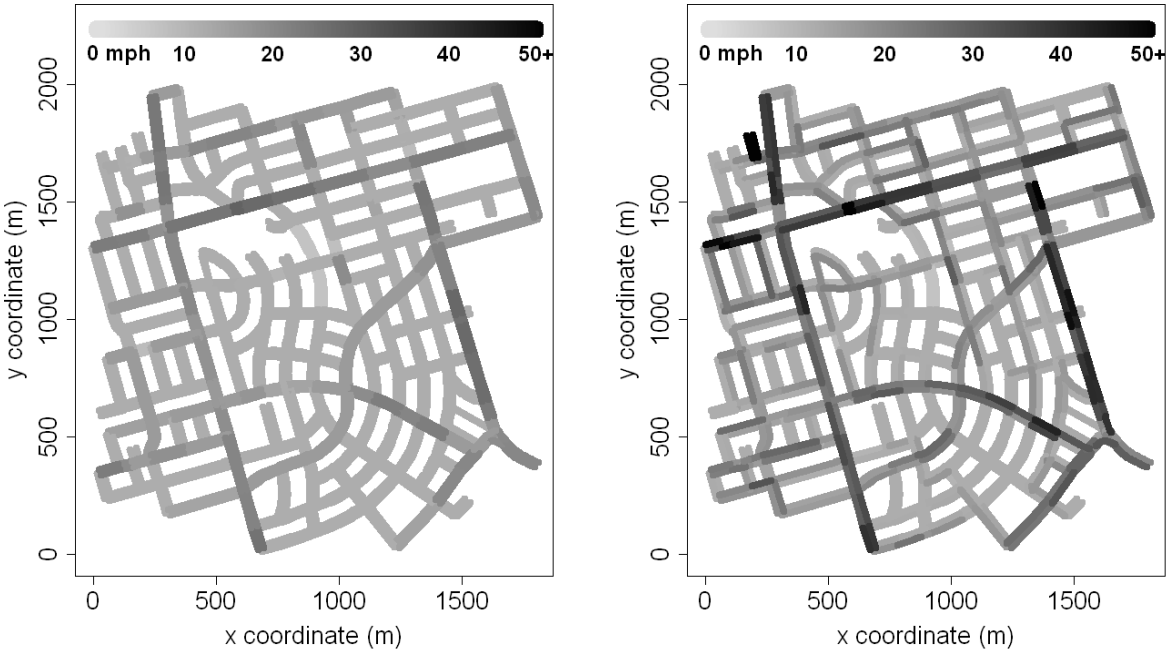

Figure 3: Prior (left) and posterior (right) speeds from the Bayesian method, for Toronto L-S data.

There are a few arcs that have poor estimates from the Bayesian method. For example, parallel black arcs in the top-left corner have poor estimates due to edge effects. Also, some short interior arcs have unrealistically high estimates, likely because there are few GPS points on these arcs. This undesirable behavior could be reduced or eliminated by using a random effect prior distribution [10] for roads in the same class, which would have the effect of pooling the available data.

7.3 Travel Time Prediction

We compare the known travel time of each trip in the test data with the point and 95% interval predictions from each method. Unlike the simulated test data in Section 6, the true paths are not known. For the Bayesian and local methods, we assume the path taken is the expected fastest path, using the mean travel time estimates for each method. This measures the ability of the methods to estimate both the fastest path and the travel time distributions accurately.

As in Section 6.2, we use the cross-validation approach of Section 5 to estimate bias correction factors. We repeat this five times, resampling random training and validation/test sets, and give arithmetic means of the performance metrics over the five replications in Table 2. We again compare the point estimates from the three methods on the test data using RMSE, RMSE log, and Bias (M.A.), and compare the interval estimates using Width and Cov. %. Because the true travel time distributions are unknown, we cannot use the Oracle method as in Section 6.2. However, we still wish to estimate "gold standard" performance, so we implement an "Estimated Oracle" method, in which we assume that the parametric model and MLE estimates from the Local MLE method are the truth. We simulate realized travel times on the fastest path for each test trip (in expectation, as estimated by the Local MLE method), and compare these to the expected value point estimates used by that method. To avoid simulation error, we use Monte Carlo estimates from 1000 simulated travel times.

For the L-S data, the Bayesian method outperforms the method of Budge et al. and the local methods, suggesting that it is effectively combining trip information with GPS information. The Bayesian method is roughly 6% better in log scale RMSE, after subtracting the error from the Estimated Oracle method. The method of Budge et al. and the local methods perform similarly. The bias correction is successful at eliminating bias (there is 2-3% bias remaining).

The Bayesian method substantially outperforms the local methods in interval estimates. The Bayesian intervals have much higher coverage percentage than the intervals from the local methods. The method of Budge et al. has higher coverage percentage than the Bayesian method; however, the intervals are also substantially wider. The intervals from the MLE

L-S data (Mean over five replications)					
Estimation method	RMSE (s)	RMSE log	Bias (M.A.)	Cov. %	Width (s)
Est. oracle	14.9	0.168	0.018	-	-
Bayesian	37.8	0.332	0.025	85.8	75.0
Local MLE	38.4	0.342	0.027	73.3	55.0
Local Harm.	38.5	0.343	0.028	77.5	75.2
Budge et al.	39.8	0.342	0.028	94.5	122.3
Std data (Mean over five replications)					
Estimation method	RMSE (s)	RMSE log	Bias (M.A.)	Cov. %	Width (s)
Est. oracle	35.2	0.191	0.018	-	-
Bayesian	126.8	0.465	0.025	73.0	141.8
Local MLE	129.0	0.480	0.025	58.4	118.6
Local Harm.	129.0	0.480	0.025	64.8	142.8
Budge et al.	127.9	0.475	0.026	94.3	370.8

Table 2: Out-of-sample trip travel time estimation performance on Toronto EMS data.

method are narrow and have low coverage percentage. Therefore, the Local MLE method does not adequately account for travel time variability, suggesting that the Estimated Oracle method may underestimate the baseline error. If so, the Bayesian method outperforms the other methods by an even larger amount, relative to the baseline error.

For the Std data, the Bayesian method outperforms the local methods by roughly 5% in RMSE on the log scale, and outperforms the method of Budge et al. by 3.5%, again relative to the Estimated Oracle error. Point estimates from the method of Budge et al. slightly outperform the local methods. Interval estimation is less successful for the Bayesian and local methods than for the L-S data, probably because the Std travel times have more unaccounted sources of variability than the L-S travel times, such as variation from to traffic and time of day. The intervals from Budge et al. have high coverage percentage, but are so wide as to have little practical use. The median travel time in the Std dataset is 190 seconds, and a typical interval estimate for a Std travel time is [70, 440].

This region and dataset are favorable for the method of Budge et al. The travel speeds are similar across most roads in this test region, which mitigates the main weakness of the Budge et al. method, namely its inability to distinguish between fast and slow roads. Also, the test region is relatively small in area. In fact, several start/end node pairs are repeated in this

dataset over one hundred times. Therefore the Budge et al. method does not suffer because of using the same travel time distribution estimates regardless of location.

7.4 Response Within Time Threshold

Next we estimate the probability an ambulance completes its trip within a certain time threshold. These probabilities are critical for EMS providers (see Section 1). In Figure 4, we assume that an ambulance begins at the node marked with a black "X" and estimate the probability it reaches each other node in 150 seconds, following the fastest path (in expectation). For the Bayesian method, these probabilities are calculated by simulating travel times from the posterior distribution of each arc in the route, and using Monte Carlo approximation as described in Section 3. The left figure depicts probabilities from the Bayesian method, and the right figure depicts probabilities from the method of Budge et al.

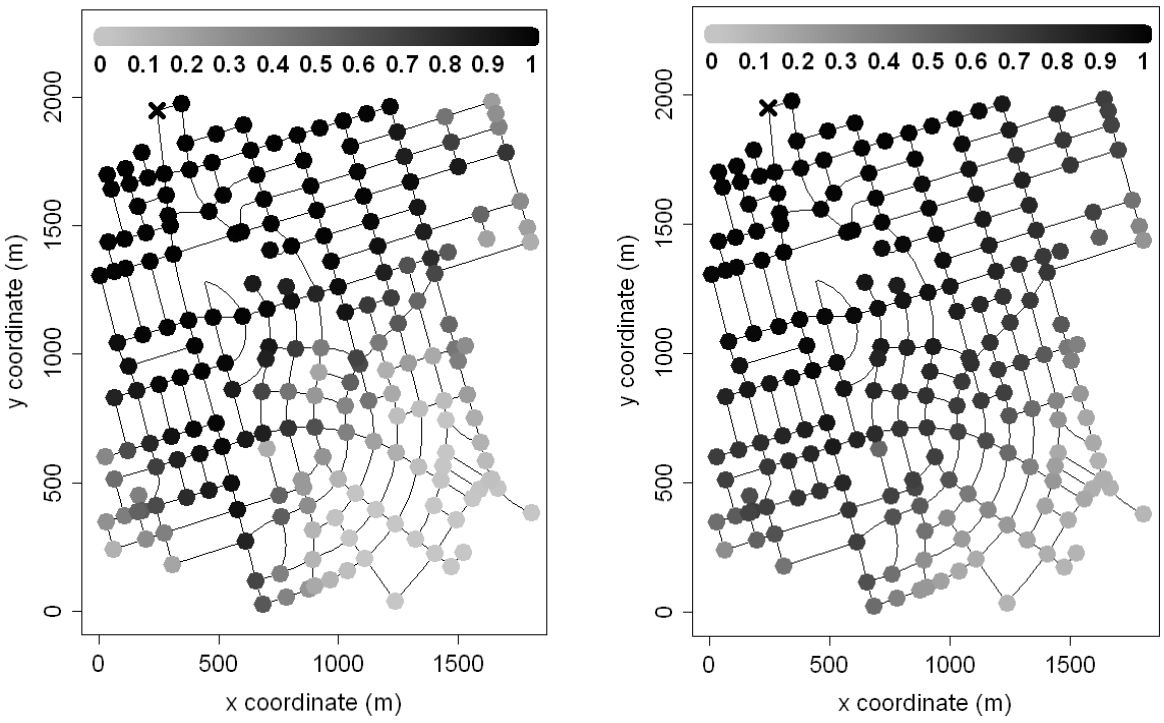

Figure 4: Estimates of probability of reaching each node in 150 seconds, Bayesian method (left), Budge et al. method (right), from the location marked "X."

The probabilities are high for nodes close to the start node and decrease for nodes further away. The probabilities from the Bayesian method appear more reasonable than from the method of Budge et al. Nodes on main roads tend to have higher probabilities from the Bayesian method (for example, traveling south from the start node), whereas nodes on minor roads far from the start node have lower probabilities from the Bayesian method (see the bottom-right in each plot). This is because the method of Budge et al. does not take into account the different speeds of different roads.

7.5 Map-Matched Path Accuracy

Finally, we assess map-matching estimates from the Bayesian method, for the Toronto L-S data. Figure 5 shows two example ambulance paths from the L-S dataset. The GPS locations are shown in white; the first reading is marked with a cross, the last with an X. As in Section 6.3, each arc is shaded by its marginal posterior probability, if that is greater than 1%. In the left-hand path, there are two occasions where the path is not precisely defined by the GPS readings. On both occasions, most of the posterior probability ($\approx 90\%$) is given to a route following the main road, which is estimated to be faster. The final two GPS readings appear to have location error. However, the fastest path is still given roughly 100% posterior probability, instead of a detour that would be slightly closer to the second-to-last GPS reading. In the right-hand path, for an unknown reason, there is a large gap between GPS points. Almost all the posterior probability is given to the fastest route, following a primary and then secondary road. This illustrates the robustness of the Bayesian method to sparse GPS data.

8 Conclusions

We proposed a Bayesian method to estimate the travel time distribution on any route in a road network, using sparse and error-prone GPS data. We simultaneously estimated the vehicle paths and the parameters of the travel time distributions. We also introduced two

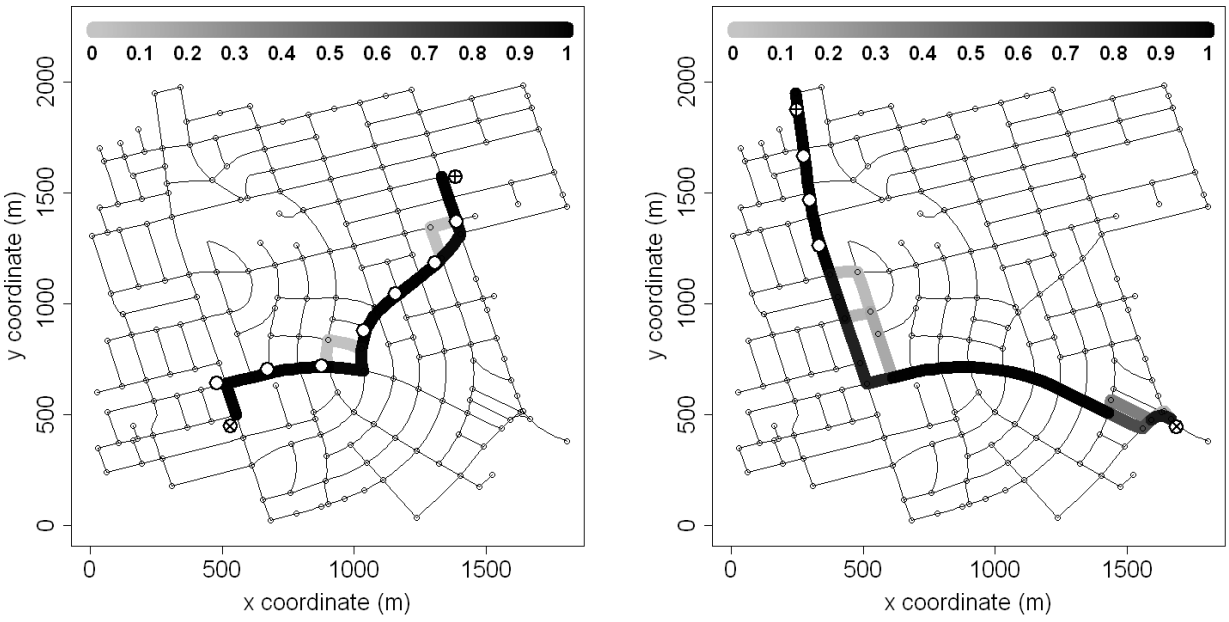

Figure 5: Map-matching estimates for two Toronto L-S trips, colored by the probability each arc is traversed.

"local" methods based on mapping each GPS reading to the nearest road segment. The first method used the harmonic mean of the GPS speeds; the second performed maximum-likelihood estimation for a lognormal distribution of travel speeds on each segment.

We compared these three methods to an existing method from Budge et al [4]. In simulations, the Bayesian method greatly outperformed the local methods and the method of Budge et al. in estimating out-of-sample trip travel times, for both point and interval estimates. The estimates from the Bayesian method remained excellent even when the GPS data had high error. On the Toronto EMS data, the Bayesian method again outperformed the competing methods in out-of-sample prediction, and provided more reasonable estimates of the probability of completing a trip within a time threshold than the method of Budge et al.

We plan to extend the Bayesian method to capture the time-varying nature of travel times. For instance, speeds decrease during rush hour. Applying the methods of this paper separately to rush hour and non-rush hour improves performance on "standard travel" Toronto data, although it has little effect on performance for "lights-and-sirens" data. We expect that a

more sophisticated approach that smooths across time of day will have better success.

We are currently investigating a number of other extensions. First, we are developing methods to approximate or modify the Bayesian solutions to obtain efficient computation on very large networks. Second, we are experimenting with ways to share information across roads, to improve estimates on infrequently-used roads. Third, we are incorporating dependence between arc travel times within each trip, arising due to traffic congestion effects or a driver's speed preference, for example. Finally, we are investigating the use of turn penalties to capture the fact that a left turn can require more time than a right turn.

A Constants and Hyperparameters

There are several constants and hyperparameters to be specified in the Bayesian model. To set the GPS position error covariance matrix Σ, we calculate the minimum distance from each GPS location in the data to the nearest arc. Assuming that the error is radially symmetric, that the vehicle was on the nearest arc when it generated the GPS point, and approximating that arc locally by a straight line, this minimum distance should equal the absolute value of one component of the 2-dimensional error, i.e. the absolute value of a random variable $\mathcal{E}_1 \sim N(0, \sigma^2)$, where $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. Since $E(|\mathcal{E}_1|) = \sigma\sqrt{2/\pi}$, we take $\hat{\sigma} = \hat{E}(|\mathcal{E}_1|)\sqrt{\pi/2}$, where $\hat{E}(|\mathcal{E}_1|)$ is the mean minimum distance of each GPS point to the nearest arc in the data. In the Toronto EMS datasets, we have $\hat{E}(|\mathcal{E}_1|) = 8.4$ m for the L-S data and 7.7 m for the Std data, yielding $\Sigma_{\text{L-S}} = \left(\begin{smallmatrix} 111.6 & 0 \\ 0 & 111.6 \end{smallmatrix}\right)$ and $\Sigma_{\text{Std}} = \left(\begin{smallmatrix} 92.7 & 0 \\ 0 & 92.7 \end{smallmatrix}\right)$. In the simulated data, a typical dataset has $\hat{E}(|\mathcal{E}_1|) = 7.3$ m for "good" GPS data and 14.1 m for "bad" GPS data, yielding $\Sigma_{\text{Good}} = \left(\begin{smallmatrix} 84 & 0 \\ 0 & 84 \end{smallmatrix}\right)$, and $\Sigma_{\text{Bad}} = \left(\begin{smallmatrix} 312 & 0 \\ 0 & 312 \end{smallmatrix}\right)$.

The hyperparameters b_1, b_2, s^2, and m_j control the prior distributions on the travel time parameters μ_j and σ_j^2 (Equation 6). We set b_1 and b_2 by estimating the possible range in travel time variation for a single arc. Some arcs have very consistent travel times (for example an arc with little traffic and no major intersections at either end). We estimate that such an

arc could have travel time above or below the median time by a factor of 1.1. Taking this range to be a two standard deviation σ_j interval (so that $1.1\exp(\mu_j) = \exp(\mu_j + 2\sigma_j)$) yields $\sigma_j \approx 0.0477$. Other arcs have very variable travel times (for example an arc with substantial traffic). We estimate that such an arc could have travel time above or below the median time by a factor of 3.5, corresponding to $\sigma_j \approx 0.6264$. Thus, we set $b_1 = 0.0477$ and $b_2 = 0.6264$.

We assume we have an initial travel time estimate τ_j for each arc j (for example, in Section 7 we use previous estimates from Toronto EMS). We expect this estimate to be typically correct within a factor of two. Thus, we specify m_j and s^2 so that the prior distribution for $E[T_{i,j}]$ is centered at τ_j and has a two standard deviation interval from $\tau_j/2$ to $2\tau_j$. This gives

$$\begin{aligned}
\tau_j &= E\left(\exp\left(\mu_j + \sigma_j^2/2\right)\right) \\
&= \exp\left(m_j + s^2/2\right) E\left(\exp\left(\sigma_j^2/2\right)\right), \\
\frac{\tau_j}{2} &= \exp\left(m_j + s^2/2 - 2s\right) E\left(\exp\left(\sigma_j^2/2\right)\right), \\
2\tau_j &= \exp\left(m_j + s^2/2 + 2s\right) E\left(\exp\left(\sigma_j^2/2\right)\right),
\end{aligned}$$

where the final equation is redundant. Therefore,

$$m_j = \log\left(\frac{\tau_j}{E\left(\exp\left(\sigma_j^2/2\right)\right)}\right) - \frac{s^2}{2}, \qquad\qquad s = \frac{\log(2)}{2}.$$

When τ_j is not available, as in our simulation study, one can use the following data-based choice for τ_j: find the harmonic mean GPS speed reading in the entire dataset and convert this speed to a travel time for each road.

Results are very insensitive to the hyperparameters b_3 and b_4, as long as the interval $[b_3, b_4]$ does not exclude regions of high likelihood. This is because the entire dataset is used to estimate ζ^2 (unlike for the parameters σ_j^2). We fix $b_3 = 0$ and $b_4 = 0.5$. For observed GPS speed V_i^ℓ, suppose the true speed at that moment is v. By Equation 4, $V_i^\ell \sim \mathcal{LN}\left(\log(v) - \zeta^2/2, \zeta^2\right)$. If

$\zeta = 0.5$, we estimate by simulation that

$$\frac{E\left(\left|V_i^\ell - v\right|\right)}{v} \approx 0.4,$$

which is much higher than any mean absolute error observed by Witte and Wilson [31]. Thus, it is not realistic that the error could be greater than this.

The hyperparameter C governs the multinomial logit choice model prior distribution on paths. While the results of the Bayesian method are generally insensitive to moderate changes in the other hyperparameters, changes in the value of C do have a noticeable effect, so we obtain a careful data-based estimate. Equation 2 implies that the ratio of the probabilities of two possible paths depends on their difference in expected travel time. For example, let $C = 0.1$ and consider paths $\tilde{a}_i$ and $\dot{a}_i$ from d_i^s to d_i^f, where the expected travel time of $\tilde{a}_i$ is 10 seconds less than the expected travel time of $\dot{a}_i$. Then path $\tilde{a}_i$ is $e \approx 2.72$ times more likely.

We specify C using the principle that for a trip of average travel time, a driver is ten times less likely to choose a path that has 10% longer travel time. If $\bar{T}$ is the average travel time, then by Equation 2, this requires

$$0.1 = \frac{\exp\left(-C\left(1.1\bar{T}\right)\right)}{\exp\left(-C\bar{T}\right)} = \exp\left(-0.1C\bar{T}\right), \tag{9}$$

giving $C = -\log(0.1)/\left(0.1\bar{T}\right)$. For our simulated data, this yields $C_{\text{Sim}} = 0.24$.

On the real data, we make a small adjustment to pool information across the L-S and Std datasets. Observing that the route choices are very similar in visual inspection of these datasets, we ensure that the prior distribution on the route taken between two fixed locations is the same for the L-S and Std datasets. To do this, we combine all the L-S and Std data to calculate an overall mean L_1 trip length L_1^{Tor} (change in x coordinate plus change in y coordinate) for the Toronto EMS data, which is $L_1^{\text{Tor}} = 1378.8$m. Let $L_1^{\mathcal{D}}$ and $T^{\mathcal{D}}$ be the mean L_1 length and mean trip time for each dataset $\mathcal{D}$. We estimate a weighted mean time $T_W^{\mathcal{D}} = T^{\mathcal{D}} L_1^{\text{Tor}} / L_1^{\mathcal{D}}$ for dataset $\mathcal{D}$ for a trip of length L_1^{Tor}, and use the time $T_W^{\mathcal{D}}$ to set C by

Equation 9. This yields $C_{\text{L-S}} = 0.211$ and $C_{\text{Std}} = 0.110$.

B Harmonic Mean Speed and GPS Sampling

When estimating road segment travel times via speed data from GPS readings, as in the local methods of Section 4.1, it is critical whether the GPS readings are sampled by distance or by time. Sampling-by-distance could mean recording a GPS point every 100m, and sampling-by-time could mean recording a GPS point every 30s, for example. As discussed in Sections 1 and 4.1, most EMS providers use a combination of distance and time sampling. If both constraints are satisfied frequently (unlike in the Toronto EMS dataset, where most points are sampled by distance), this could create a problem for estimating travel times via these speeds.

In the transportation research literature, where sampling is done by distance (because speeds are recorded at loop detectors at fixed locations on the road), it is well known that the harmonic mean of the observed speeds (the "space mean speed") is appropriate for estimating travel times [22, 25, 30]. Under a simple probabilistic model of sampling-by-distance, without assuming constant speed, we confirm that the harmonic mean speed gives an unbiased and consistent estimator of the mean travel time. However, we also show that if the sampling is done by time, the harmonic mean is biased towards overestimating the mean travel time.

Consider a set of n ambulance trips on a single road segment. For convenience, let the length of the road segment be 1. Let the travel time on the segment for ambulance i be T_i, and assume that the T_i are iid with finite expectation. Let $x_i(t)$ be the position function of ambulance i, conditional on T_i, so $x_i(0) = 0$ and $x_i(T_i) = 1$. Assume that $x_i(t)$ is continuously differentiable, with derivative $v_i(t)$, the velocity function, and that $v_i(t) > 0$ for all t. Each trip samples one GPS point. Let V_i^o be the observed GPS speed for the ith ambulance.

First, consider sampling-by-distance. For trip i, draw a random location $\xi_i \sim \text{Unif}(0, 1)$ at which to sample the GPS point. This is different from the example of sampling-by-distance above. However, if the sampling locations are not random, we cannot say anything about

the observed speeds in general (the ambulances might briefly speed up significantly where the reading is observed, for example). Assuming that the ambulance trip started before this road segment, it is reasonable to model sampling-by-distance with a uniform random location.

Conditional on T_i, $x_i(\cdot)$ is a cumulative distribution function, with support $[0, T_i]$, density $v_i(\cdot)$, and inverse $x_i^{-1}(\cdot)$. Thus, $\tau_i = x_i^{-1}(\xi_i)$, the random time of the GPS reading, has distribution function $x_i(\cdot)$ and density $v_i(\cdot)$, by the probability integral transform. The observed speed $V_i^o = v_i(\tau_i)$, so the GPS reading is more likely to be sampled when the ambulance has high speed than when it has low speed. This is called the inspection paradox (see e.g. [26]). Mathematically,

$$E(V_i^o | T_i) = E(v_i(\tau_i) | T_i) = \int_0^{T_i} v_i(t) v_i(t) dt \geq \frac{\left(\int_0^{T_i} v_i(t) dt\right)^2}{\int_0^{T_i} 1^2 dt} = \frac{1}{T_i},$$

by the Cauchy-Schwarz inequality, with strict inequality unless $v_i(\cdot)$ is constant. However, if we draw a uniform time $\phi_i \sim U(0, T_i)$, then

$$E(v_i(\phi_i) | T_i) = \int_0^{T_i} v_i(t) \frac{1}{T_i} dt = \frac{1}{T_i}. \tag{10}$$

The inspection paradox has a greater impact in the Toronto Std data than in the L-S data, because ambulance speed varies more in standard travel.

Consider estimating the mean travel time $E(T_i)$ via the estimator $\hat{T}^H = 1/\bar{V}_H^o$, where $\bar{V}_H^o$ is the harmonic mean observed speed. We have

$$\begin{aligned} E\left(\hat{T}^H\right) &= E\left(E\left(\hat{T}^H \middle| \{T_i\}_{i=1}^n\right)\right) = E\left(\frac{1}{n}\sum_{i=1}^n E\left(\frac{1}{v_i(\tau_i)} \middle| T_i\right)\right) \\ &= E\left(\frac{1}{n}\sum_{i=1}^n \int_{t=0}^{T_i} \frac{1}{v_i(t)} v_i(t) dt\right) = E\left(\frac{1}{n}\sum_{i=1}^n T_i\right) = E(T_i), \end{aligned}$$

and so it is unbiased. Moreover, it is consistent as $n \to \infty$, by the Law of Large Numbers.

Next, suppose the sampling is instead done by time. To model this, let $\tau_i \sim \text{Unif}(0, T_i)$ be

a random time to sample the GPS point for ambulance i. In this case, we have

$$\begin{aligned} E\left(\hat{T}^H\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{v_i(\tau_i)}\middle| T_i\right)\right) \\ &\geq E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{E\left(v_i(\tau_i)|\, T_i\right)}\right) \\ &= E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\frac{1}{T_i}}\right) = E(T_i), \end{aligned}$$

by Jensen's Inequality and Equation 10. Again, the inequality is strict unless $v_i(\cdot)$ is constant.

C Reversibility of the Path Update

The path $A_i = (A_{i,1}, \ldots, A_{i,N_i})$ takes values in the finite set $\mathcal{P}_i$. Conditional on A_i, the vector T_i takes values on the simplex $\mathcal{X}_{N_i} \triangleq \left\{T_i \in \mathbb{R}^{N_i} : T_{i,j} > 0, \sum_{j=1}^{N_i} T_{i,j} = t_i^f - t_i^s\right\}$, where $t_i^f - t_i^s$ is the known total travel time of trip i. For the reference measure on $\mathcal{X}_{N_i}$ we use $(N_i - 1)$-dimensional Lebesgue measure on the first $N_i - 1$ elements of the vector. Then

$$(A_i, T_i) \in \mathcal{C} \triangleq \bigcup_{A \in \mathcal{P}_i} \{A\} \times \mathcal{X}_{\text{len}(A)}$$

where $\text{len}(A)$ is the number of arcs in $A \in \mathcal{P}_i$. We claim that the move for (A_i, T_i) is reversible with respect to the conditional posterior density of (A_i, T_i) given the GPS data $G = \{G_{i'}\}_{i'=1}^{I}$, the parameters, and the paths and travel times $A_{[-i]}, T_{[-i]}$ for all other trips:

$$\begin{aligned} \nu(A_i, T_i) &\triangleq \pi\left(A_i, T_i \;\middle|\; G, A_{[-i]}, T_{[-i]}, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}, \zeta^2\right) \\ &\propto f_i\left(A_i, T_i, G_i \;\middle|\; \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{J}, \zeta^2\right). \end{aligned} \tag{11}$$

Since the dimension of the unknown vector T_i depends on A_i, one can consider this to be a case of model uncertainty as in Green [13], where the model index k corresponds to the value of $A_i \in \mathcal{P}_i$. Our context, which has an uncertain route for each trip, is slightly different

from the context of Green [13], which has a single uncertain model index k and corresponding parameter vector $\theta^{(k)}$. However, their argument can still be used to show reversibility of a move for (A_i, T_i) conditional on $A_{[-i]}, T_{[-i]}$ and the parameters $\{\mu_j, \sigma_j^2\}_{j=1}^J, \zeta^2$.

Conditional on $A_i^{(1)}$ and $A_i^{(2)}$, we show that our move from $T_i^{(1)} \in \mathcal{X}_{\text{len}\left(A_i^{(1)}\right)}$ to $T_i^{(2)} \in \mathcal{X}_{\text{len}\left(A_i^{(2)}\right)}$ satisfies the "dimension-matching" condition of Green [13] Section 3.3. To do this we need a bijection between an augmented vector $\left(T_i^{(1)}, u^{(1)}\right)$ and the corresponding augmented vector $\left(T_i^{(2)}, u^{(2)}\right)$, for some $u^{(1)}$ and $u^{(2)}$. This holds by taking $u^{(1)} \triangleq \left(T_i^{(2)}(p_1), \ldots, T_i^{(2)}(p_n)\right)$ and $u^{(2)} \triangleq \left(T_i^{(1)}(c_1), \ldots, T_i^{(1)}(c_m)\right)$ and remembering that $u^{(1)}$ is drawn independently of $T_i^{(1)}$. Define the bijection $h\left(T_i^{(1)}, u^{(1)}\right) \triangleq \left(T_i^{(2)}, u^{(2)}\right)$ that simply rearranges the elements of the vector $\left(T_i^{(1)}, u^{(1)}\right)$. The absolute value of the Jacobian of such a transformation is one (since that of the identity transform is one, and since rearranging the elements corresponds to permuting the rows of the Jacobian, which only changes the sign of the determinant). Although for notational convenience we have included the redundant final elements of the vectors $u^{(1)}$, $u^{(2)}$, $T_i^{(1)}$, and $T_i^{(2)}$, the dimension-matching is on the non-redundant elements of the vectors; in the notation of Green [13], $n_1 = N_i^{(1)} - 1$, $m_1 = n - 1$, $n_2 = N_i^{(2)} - 1$, and $m_2 = m - 1$.

For a dimension-matching move, the acceptance probability that ensures reversibility with respect to a density $\nu(A_i, T_i)$ is given by Equation 7 of Green [13]. It is equal to the absolute value of the Jacobian, times $\frac{\nu\left(A_i^{(2)}, T_i^{(2)}\right)}{\nu\left(A_i^{(1)}, T_i^{(1)}\right)}$, times the ratio of the proposal density of the reverse move relative to that of the proposed move. The probability of proposing a move to $A_i^{(2)}$, given that the current state is $\left(A_i^{(1)}, T_i^{(1)}\right)$, is $\frac{1}{N_i^{(1)} \min\{a^{(1)}, K\}}$ divided by the number of paths of length $\leq K$ from d' to d''. The probability of attempting the reverse move is $\frac{1}{N_i^{(2)} \min\{a^{(2)}, K\}}$ divided by the number of paths of length $\leq K$ from d' to d''. We propose $T_i^{(2)}$ by drawing the subvector $T_i^{(2)}(j) : j \in \{p_1, \ldots, p_n\}$ according to the density $\frac{1}{S_i^{n-1}}$ $\text{Dir}\left(\frac{T_i^{(2)}(p_1)}{S_i}, \ldots, \frac{T_i^{(2)}(p_n)}{S_i}; \alpha\theta(p_1), \ldots, \alpha\theta(p_n)\right)$ on the simplex $\{T_i \in \mathbb{R}^n : T_{i,j} > 0, \sum_{j=1}^n T_{i,j} = S_i\}$, with respect to $(n-1)$-dimensional Lebesgue measure. The reverse move, from $T_i^{(2)} \in \mathcal{X}_{\text{len}\left(A_i^{(2)}\right)}$ to $T_i^{(1)} \in \mathcal{X}_{\text{len}\left(A_i^{(1)}\right)}$, proposes $T_i^{(1)}$ by drawing the subvector $T_i^{(1)}(j) : j \in \{c_1, \ldots, c_m\}$ according to the density $\frac{1}{S_i^{m-1}}$ $\text{Dir}\left(\frac{T_i^{(1)}(c_1)}{S_i}, \ldots, \frac{T_i^{(1)}(c_m)}{S_i}; \alpha\theta(c_1), \ldots, \alpha\theta(c_m)\right)$.

Plugging these quantities into Equation 7 of Green [13] and using our Equation 11 gives the acceptance probability in our Equation 7.

Acknowledgements

We would like to thank the referees and Associate Editor for their careful reading of the paper and helpful comments. We would also like to thank The Optima Corporation and Dave Lyons of Toronto EMS. This research was partially supported by the National Science Foundation under Grant CMMI-0926814.

References

[1] K. Aladdini. EMS response time models: A case study and analysis for the region of Waterloo. Master's thesis, University of Waterloo, 2010.

[2] R. Alanis, A. Ingolfsson, and B. Kolfal. A Markov Chain model for an EMS system with repositioning. 2010. Working paper.

[3] L. Brotcorne, G. Laporte, and F. Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147:451–463, 2003.

[4] S. Budge, A. Ingolfsson, and D. Zerom. Empirical analysis of ambulance travel times: The case of Calgary emergency medical services. *Management Science*, 56:716–723, 2010.

[5] W. Chen, Z. Li, M. Yu, and Y. Chen. Effects of sensor errors on the performance of map matching. *The Journal of Navigation*, 58:273–282, 2005.

[6] S.F. Dean. Why the closest ambulance cannot be dispatched in an urban emergency medical services system. *Prehospital and Disaster Medicine*, 23:161–165, 2008.

[7] E. Erkut, A. Ingolfsson, and G. Erdoğan. Ambulance location for maximum survival. *Naval Research Logistics (NRL)*, 55:42–58, 2008.

[8] J.J. Fitch. *Prehospital Care Administration: Issues, Readings, Cases.* St. Louis: Mosby-Year Book, 1995.

[9] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:515–533, 2006.

[10] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis.* London: Chapman & Hall, 2004.

[11] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472, 1992.

[12] J.B. Goldberg. Operations research models for the deployment of emergency services vehicles. *EMS Management Journal*, 1:20–39, 2004.

[13] P.J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[14] S.G. Henderson. Operations research tools for addressing current challenges in emergency medical services. In *Wiley Encyclopedia of Operations Research and Management Science.* New York: Wiley, 2010.

[15] A. Ingolfsson, S. Budge, and E. Erkut. Optimal ambulance location with random delays and travel times. *Health Care Management Science*, 11:262–274, 2008.

[16] J. Krumm, J. Letchner, and E. Horvitz. Map matching with travel time constraints. In *Society of Automotive Engineers (SAE) 2007 World Congress*, 2007.

[17] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang. Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 352–361. ACM, 2009.

[18] F. Marchal, J. Hackney, and K.W. Axhausen. Efficient map matching of large Global Positioning System data sets: Tests on speed-monitoring experiment in Zurich. *Transportation Research Record: Journal of the Transportation Research Board*, 1935:93–100, 2005.

[19] A.J. Mason. Emergency vehicle trip analysis using GPS AVL data: A dynamic program for map matching. In *Proceedings of the 40th Annual Conference of the Operational Research Society of New Zealand. Wellington, New Zealand*, pages 295–304, 2005.

[20] D. McFadden. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, pages 105–142. New York: Academic Press, 1973.

[21] N.J. Nilsson. *Artificial Intelligence: A New Synthesis*. San Francisco: Morgan Kaufmann, 1998.

[22] H. Rakha and W. Zhang. Estimating traffic stream space mean speed and reliability from dual- and single-loop detectors. *Transportation Research Record: Journal of the Transportation Research Board*, 1925:38–47, 2005.

[23] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 2004.

[24] G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16:351–367, 2001.

[25] F. Soriguera and F. Robuste. Estimation of traffic stream space mean speed from time aggregations of double loop detector data. *Transportation Research Part C: Emerging Technologies*, 19:115–129, 2011.

[26] W.E. Stein and R. Dattero. Sampling bias and the inspection paradox. *Mathematics Magazine*, 58:96–99, 1985.

[27] S. Syed. Development of map aided GPS algorithms for vehicle navigation in urban canyons. Master's thesis, University of Calgary, 2005.

[28] M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–540, 1987.

[29] L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22:1701–1728, 1994.

[30] J.G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers*, 2:325–378, 1952.

[31] T.H. Witte and A.M. Wilson. Accuracy of non-differential GPS for the determination of speed over ground. *Journal of Biomechanics*, 37:1891–1898, 2004.