Travel Time Estimation for Ambulances using Bayesian Data Augmentation

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

Cornell University
School of Operations Research and Information Engineering

July 8, 2011

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 called the map-matching problem. We consider the unknown paths and travel times to be missing data, and simultaneously estimate them and the parameters of each road segment travel time distribution using Bayesian data augmentation. We also introduce two alternative estimation methods using GPS speed data that are simple to implement in practice.

We test the predictive accuracy of the three methods on a subregion of Toronto, using simulated data and data from Toronto EMS. All three methods perform well. Point estimates of ambulance trip durations from the Bayesian method outperform estimates from the alternative methods by roughly 5% in root mean squared error. Interval estimates from the Bayesian method for the Toronto EMS data are substantially better than interval estimates from the alternative methods. Map-matching estimates from the Bayesian method are robust to large GPS location errors, and interpolate well between widely spaced GPS points.

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

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. Ambulances are often assigned to a new emergency while away from their ambulance base [6], so the problem is more complicated than estimating response times from several fixed bases. Travel times also play a central role in locating bases and parking locations [3, 10, 12]. Travel times are variable, and recent EMS research has shown the importance of accounting for this uncertainty [7, 13]. In this paper, we estimate the distribution of ambulance travel times on each road segment (the section of road between neighboring intersections) in a city. This enables estimation of fastest paths in expectation between any two locations and simulation of travel times for any given path.

Available data are historic Global Positioning System (GPS) readings, stored during ambulance trips. Most EMS providers record this information; we use GPS data from Toronto EMS from 2007 and 2008. The GPS data contain locations, timestamps, speeds, vehicle and emergency incident IDs, and other information. In this dataset, GPS readings are stored every 200 meters (m) or 240 seconds (s), whichever comes first. The true GPS sampling rate is much higher, but this scheme is used to minimize data transmission and storage. This is standard practice across EMS providers, though the parameters used vary [16].

The GPS location and speed data are both subject to error. Location readings are particularly poor near tall buildings or in tunnels [5, 16, 24], as illustrated in Figure 1. Chen et al. [5] reported large GPS location errors in parts of Hong Kong with narrow streets and tall buildings, observing average errors of 27 meters, with some errors over 100 meters. Error is also present in GPS speed readings; Witte and Wilson [29] found GPS speed errors of roughly 5% on average, with largest error at high speeds and when few GPS satellites were visible.

A natural idea for estimating road segment travel times is to use the time between successive GPS readings. However, there is rarely more than one GPS point recorded on a given road

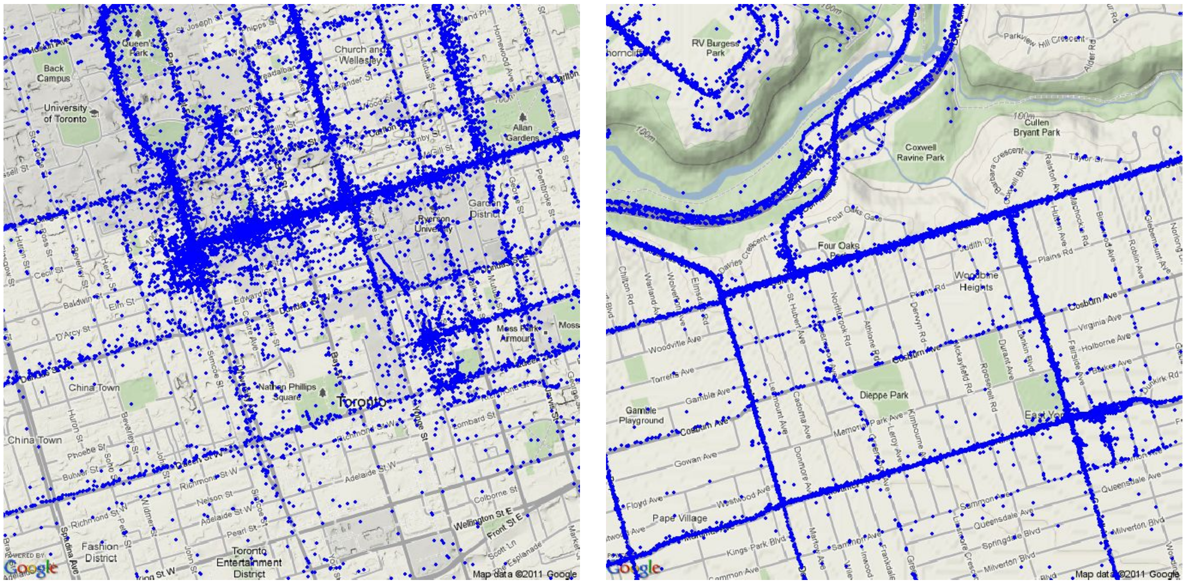

Figure 1: Ambulance GPS location readings in downtown Toronto (left) and northeast of downtown Toronto (right). The points scattered between roads are mainly GPS location errors; there is more error in downtown due to tall buildings.

segment per ambulance trip, so this is difficult. Instead, we first introduce local methods using the GPS speed data. Each GPS reading is mapped to the nearest road segment, and the mapped speeds are used to estimate the travel time for each segment. We call these local methods since they do estimation independently for each segment. This general approach is used by at least one EMS provider, although our estimation methods are different, and we obtain intervals and distributions of the travel time in addition to the mean travel time.

We introduce two local methods (Section 4). The first divides the road segment length by each GPS speed to create estimated travel times, and averages these time estimates. This is equivalent to calculating the harmonic mean of the speeds, an approach that is commonly used for estimating travel times with speed data recorded by loop detectors [18, 21, 27]. We give theoretical results supporting the use of this approach for ambulance GPS data (Appendix B). This approach also naturally yields interval and distribution estimates of the travel time. Our second local method assumes a lognormal distribution for the GPS speeds on each segment [1, 2], and calculates maximum likelihood (MLE) estimates of the parameters of this distribution. These can be used to obtain point, interval, or distribution estimates of the travel time.

A method that effectively combines the GPS speed and timestamp data for each ambulance trip should outperform these local methods. We propose such a method, modeling the ambulance trips together with the GPS readings (Sections 2 and 3). Using Bayesian data augmentation [25], we simultaneously estimate the path taken for each ambulance trip (solving the so-called map-matching problem [16]) and the distribution of travel times on each road segment. For a prior distribution on the path, we use a multinomial logit choice model [17]. As in the local MLE method, we assume a lognormal distribution for the travel time on each road segment. We do computation via Markov chain Monte Carlo methods. Since the number of road segments used in each trip is unknown and varies between possible paths, we use a reversible-jump Metropolis-Hastings step [11, 19] to produce posterior samples of the paths.

We compare the performance of the two local methods and the Bayesian method on a subregion of Toronto, using simulated data (Section 5) and historical data from Toronto EMS (Section 6). The Toronto EMS dataset is divided into "lights-and-sirens" (L-S) and "standard travel" (Std) ambulance trips. We compare out-of-sample ambulance trip duration estimates for each method. The Bayesian method outperforms the local methods; root mean squared error and mean absolute error for the Bayesian method are typically 5% lower than for the local methods. Interval estimates from the Bayesian method for the Toronto EMS data have dramatically better width and coverage than their counterparts from the local methods.

The two local methods perform similarly. We recommend the MLE method over the harmonic mean method if a simple-to-implement solution is desired. Interval estimates from the MLE method have better performance on the L-S data, the most important case, and do not require sampling from all the GPS data, so are more convenient if the dataset is large. Also, the MLE method is less dependent on correcting for zero-speed readings.

We also assess the time-dependence of the travel times in the Toronto EMS dataset, by applying our methods to the rush-hour and non-rush-hour data separately. This binning has little effect on predictive accuracy for the L-S data, though it improves performance of the

Bayesian method on the Std data.

Finally, we assess the map-matching estimates from the Bayesian method. Path estimates are interpolated accurately between widely-separated GPS points and are robust to large GPS error. The posterior distribution is able to capture multiple high-probability paths when the true path is unclear from the GPS data. There is an interesting tradeoff between routes closer to the GPS points and routes estimated to be faster. The entire dataset of trips is used to produce the path estimate for each trip, rather than analyzing each trip in isolation. Many of these features are not found in state-of-the-art map-matching techniques [14, 15, 16, 24, 28].

Recent work on estimating ambulance travel time distributions has been done by Budge et al. [4] and Aladdini [1], using estimates based on travel distance, not GPS data. Neither of these papers considered travel times on individual road segments, the level of detail that we desire. Budge et al. [4] found that conditional on travel distance, the log of the travel times had a symmetric distribution, but with heavier tails than a normal. Thus, they modeled the log travel times using t distributions. Aladdini [1] found that travel times in Waterloo, Ontario, were well characterized by lognormal distributions. We hypothesize that Budge et al. found heavier tails because Aladdini separated trips by location, whereas Budge et al. did not.

2 Bayesian Formulation

2.1 Model

Consider a network of L directed road segments, called "arcs," and a set of M ambulance trips on this network. Assume that trips begin and end on known nodes (intersections) in the network, at known times (these are estimated for real data in Section 6). For trip i, define:

- $A_i = \left\{A_i^1, A_i^2, \ldots, A_i^{N_i}\right\}$, the unknown path (sequence of arcs), of unknown length N_i.
- $T_i = \left\{T_i\left(A_i^1\right), T_i\left(A_i^2\right), \ldots, T_i\left(A_i^{N_i}\right)\right\}$, the unknown travel times on each arc.
- d_i^1, d_i^2, the known nodes in the network where the trip begins and ends.

- $\mathcal{P}_i$, the set of possible paths in the network (with no repeated nodes) from d_i^1 to d_i^2.
- $\{X_i^l, Y_i^l, V_i^l, t_i^l\}_{l=1}^{r_i}$, the observed GPS readings, with coordinates X_i^l and Y_i^l, speeds V_i^l, and timestamps t_i^l, where r_i is the number of observations.

We use the following model:

- Each travel time $T_i(j)$ follows a lognormal distribution, independently across i and j. Specifically, $\log(T_i(j)) \sim N(\mu_j, \sigma_j^2)$, where the parameters μ_j and σ_j^2 are unknown. The expected travel time on arc j is $\theta(j) = \exp\left\{\mu_j + \sigma_j^2/2\right\}$.
- Each GPS time t_i^l is accurate, but the location $\left(X_i^l, Y_i^l\right)$ is observed with error. The location has a bivariate normal distribution [14, 16], centered at the true location, with known covariance matrix Σ (in practice we use a data-based estimate; see Appendix A).
- Each GPS speed V_i^l is also observed with error, following a lognormal distribution. Specifically, $\log\left(V_i^l\right) \sim N\left(\log\left(\hat{V}_i^l\right) - \zeta^2/2, \zeta^2\right)$, where $\hat{V}_i^l$ is the true speed at time t_i^l, and the variance parameter ζ^2 is unknown. The expected observed speed $E\left(V_i^l\right) = \hat{V}_i^l$.
- The GPS location and speed errors are independent of each other and the travel times.
- Ambulances travel at constant speed on each arc.

2.2 Prior Distributions

We wish to estimate the missing data $\{A_i, T_i\}_{i=1}^M$, travel time parameters $\left\{\mu_j, \sigma_j^2\right\}_{j=1}^L$, and GPS speed error parameter ζ^2. A Bayesian approach is a natural way to perform this estimation jointly. This requires specification of prior distributions for all unknowns.

For the prior distribution on each path A_i, conditional on $\left\{\mu_j, \sigma_j^2\right\}_{j=1}^L$, we use a multinomial logit choice model [17]. Let $a_i = \left\{a_i^1, \ldots, a_i^{n_i}\right\}$ be a possible route from d_i^1 to d_i^2. In the multinomial logit choice model, the probability that path a_i is selected is

$$\pi(a_i) = \frac{\exp\left\{-C \sum_{l=1}^{n_i} \theta\left(a_i^l\right)\right\}}{\sum_{A_i \in \mathcal{P}_i} \exp\left\{-C \sum_{l=1}^{N_i} \theta\left(A_i^l\right)\right\}}, \tag{1}$$

where $C > 0$ is a fixed hyperparameter. The fastest routes in expectation have the highest probability according to this model.

For the prior distributions on the GPS speed error parameter ζ^2 and the travel time distribution parameters μ_j and σ_j, for each arc j, we use

$$\mu_j \sim N(m_j, s^2), \qquad \sigma_j \sim \text{Unif}(b_1, b_2), \qquad \zeta \sim \text{Unif}(b_3, b_4), \tag{2}$$

independently, where m_j, s^2, b_1, b_2, b_3, and b_4 are fixed hyperparameters. We use the uniform prior on the standard deviations σ_j and ζ because we do not have much prior information for these parameters, except that we expect them to fall in the specified range. The normal prior is a standard choice for the location parameter of a lognormal distribution. The specification of all hyperparameters is described in Appendix A.

3 Bayesian Estimation Method

We use a Markov chain method [26] to obtain samples from the joint posterior distribution of all unknowns: paths A_i, travel times T_i, and parameters μ_j, σ_j^2, and ζ^2, given the observed GPS data and known start and end nodes and durations for each trip. In the Markov chain, each unknown quantity is updated in turn, conditional on the other unknowns. Each update is either a draw from the closed-form conditional posterior distribution or a Metropolis-Hastings (M-H) move; this guarantees convergence of the sample vector to the joint posterior distribution, and validity of the Monte Carlo estimates based on these samples [20, 26].

3.1 Markov Chain Initial Conditions

First we describe the initial conditions used for the Markov chain. 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 duration 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 the priors.

3.2 Updating the Paths

Next we describe the updating of each path A_i in the Markov chain. We update A_i using a reversible-jump M-H proposal [11, 19]. We propose a small change to the current path A_i, giving proposed sample A_i^*. Because the path changes, the travel times T_i must be updated, giving proposed sample T_i^*. The samples A_i^* and T_i^* are accepted with the appropriate M-H acceptance probability, detailed below. This is a reversible jump M-H move because the number of arcs in the path may change, changing the number of parameters in the model.

The proposal changes a contiguous subset of the path. The length of this subpath is limited to some maximum value k; k is specified in Section 3.5. The proposal works as follows.

1. With equal probability, choose a node d' from the path A_i, excluding the final node.
2. Let b be the number of nodes that follow d' in the path. With equal probability, choose an integer $a \in \{1, \ldots, \min(b, k)\}$. Denote the ath node following d' as d''. The subpath from d' to d'' is the section to be updated (the "current update section").
3. Collect the alternative 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"), obtaining the proposed path A_i^*.

Next, we propose new travel times T_i^* that are compatible with the new path A_i^*. Let $\{c_1, \ldots, c_m\} \subset A_i$ and $\{p_1, \ldots, p_n\} \subset A_i^*$ denote the arcs in the current and proposed update sections, noting that m and n will be different if the number of arcs has changed. For each $j \in A_i^* \setminus \{p_1, \ldots, p_n\}$, set $T_i^*(j) = T_i(j)$. Let $S_i = \sum_{l=1}^m T_i(c_l)$ be the total travel time of the current update section. We must have $\sum_{l=1}^n T_i^*(p_l) = S_i$ also, because the total duration of the trip is known. The travel times $T_i^*(p_1), \ldots, T_i^*(p_n)$ are proposed as follows.

- Draw $(r_1, \ldots, r_n) \sim \text{Dirichlet}\left(\alpha\theta(p_1), \ldots, \alpha\theta(p_n)\right)$, for a constant α (specified below). Set the proposed travel times $T_i^*(p_l) = r_l S_i$, for $l = 1, \ldots, n$.

This gives a proposal that is reasonable (and thus likely to be accepted), because the expected value of the new travel time on arc p_l is (see [8])

$$E\left(T_i^*(p_l)\right) = S_i \frac{\theta(p_l)}{\sum_{h=1}^{n} \theta(p_h)},$$

so the total travel time on the current arcs is randomly distributed over the proposed arcs, weighted by the arc expected travel times. The constant α influences the variance of each component, but not the expected values. In our experience $\alpha = 1$ works well for our application; one can also tune α to obtain higher acceptance rates for a particular dataset [20].

The M-H acceptance probability for this reversible-jump proposal [11, 19] is

$$p_{\text{A}} = \min\left\{1, \frac{f\left(A_i^*, T_i^* \,\middle|\, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{L}, \zeta^2\right) q\left(A_i, T_i \,\middle|\, A_i^*, T_i^*, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{L}\right)}{f\left(A_i, T_i \,\middle|\, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{L}, \zeta^2\right) q\left(A_i^*, T_i^* \,\middle|\, A_i, T_i, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^{L}\right)} |J| \right\},$$

where f denotes the conditional posterior density, q denotes the proposal density, and $|J|$ denotes the Jacobian of the transformation between the parameter spaces corresponding to the current and proposed paths [11]. Expressions for f, q, and $|J|$ are given in Appendix C.

3.3 Updating the Trip Travel Times

The travel times $\{T_i\}_{i=1}^{M}$ are changed (by necessity) in the path proposal above, but that proposal has a low acceptance rate, so we also include another M-H update of only the travel times, to improve mixing of the Markov chain. The proposal works as follows.

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

Similarly to the path proposal above, 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' [8]. In our experience $\alpha' = 0.5$ is effective for our application; the value of α' can also be tuned to improve the acceptance rate for a particular dataset [20]. 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 $\mu_j \left| \sigma_j^2, \{A_i\, T_i\}_{i=1}^M \right. \sim N\left(\hat{\mu}_j, \hat{s}_j^2\right)$, 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 $I_j \subset \{1,\ldots,M\}$ 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 [26]. We propose $\sigma_j^{2*} = \sigma_j^2 \exp\{\epsilon\}$, where $\epsilon \sim N(0,\eta^2)$, with fixed variance η^2. Thus, $\sigma_j^{2*} \sim \text{Log-}N\left(\log(\sigma_j^2), \eta^2\right)$. The likelihood is $\ell\left(\sigma_j^2 \middle| \mu_j, \{T_i(j)\}_{i\in I_j}\right) = \prod_{i\in I_j} \text{Log-}N\left(T_i(j); \mu_j, \sigma_j^2\right)$, where Log-$N()$ denotes the lognormal density, because the travel times are independent. The M-H acceptance probability [26] is

$$p_\sigma = \min\left\{1, \frac{\sigma_j}{\sigma_j^*}\mathbf{1}_{\{\sigma_j^*\in[a,b]\}} \left(\frac{\prod_{i\in I_j}\text{Log-}N\left(T_i(j);\mu_j,\sigma_j^{2*}\right)}{\prod_{i\in I_j}\text{Log-}N\left(T_i(j);\mu_j,\sigma_j^2\right)}\right) \frac{\text{Log-}N\left(\sigma_j^2;\log\left(\sigma_j^{2*}\right),\eta^2\right)}{\text{Log-}N\left(\sigma_j^{2*};\log\left(\sigma_j^2\right),\eta^2\right)}\right\}.$$

To update ζ^2, we use another M-H step with a lognormal proposal, with different variance ν^2. The M-H acceptance probability can be calculated similarly.

3.5 Markov Chain Convergence

The Markov chain converges to the posterior distribution as long as the M-H transition kernels are reversible, aperiodic, and irreducible [26]. The proposals for ζ^2, σ_j^2, and T_i satisfy these requirements. The A_i kernel is aperiodic, and reversible because the current path has update section length $m \leq k$, so it is possible to transition from state $\{A_i^*, T_i^*\}$ to state $\{A_i, T_i\}$. Rarely, a path may be initialized with a repeat node (see Section 3.1), in which case the reverse transition is not allowed. However, this initial state is transient, so it can be neglected.

The A_i kernel is irreducible if for any path i, it is possible to move between any two paths in $\mathcal{P}_i$ in a finite number of iterations. For a given road network, the maximum update section length k can be set high enough to meet this criterion. On road networks with high connectivity, a low k is sufficient. For example, $k = 3$ is sufficient for a square grid. 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, this could occur with a highway parallel to a small road. If the small road intersects other small roads, with each intersection beginning a new arc, there could be many arcs of the small road alongside a single arc of the highway. Then, a very large k would be needed to allow transitions between the highway and the small road. If k were kept smaller, the Markov chain would not be irreducible. In this case, the chain converges to the conditional posterior distribution for 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 5 and 6, 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 [9], 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 (using the second half of each chain) 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.

4 Local Methods

Here we describe the two local methods outlined in Section 1. Each GPS reading is mapped to the nearest arc (both directions of travel are treated together). The problem of estimating each arc travel time distribution using the mapped GPS speeds is similar to the problem of estimating travel times using speed data recorded by loop detectors [18]. This problem has been well studied in the transportation research literature [21, 23, 30], and in this context it is standard to estimate travel times via the harmonic mean of the observed speeds (the "space-mean speed" [18, 21, 27]).

Let n_j be the number of GPS points mapped to arc j, L_j be the length of arc j, and the mapped speed observations $\left\{V_j^l\right\}_{l=1}^{n_j}$. In our first local method, the harmonic mean of the speeds $\left\{V_j^l\right\}_{l=1}^{n_j}$ is calculated, and converted to a travel time point estimate

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

This is equivalent to assuming constant speed, converting speed observations to travel time estimates $T_j^l = L_j/V_j^l$, and averaging these times. The empirical distribution of the estimated times $\left\{T_j^l\right\}_{l=1}^{n_j}$ 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 5mph. Results are fairly sensitive to this correction. If the speed threshold is lower, there are some significantly higher estimated times, and the mean travel time estimates are higher.

In Appendix B, we consider this travel time estimator $\hat{T}_j^H$ and its relation to the GPS sampling scheme. If GPS points are sampled by distance, $\hat{T}_j^H$ is unbiased and consistent. However, if GPS points are sampled by time, $\hat{T}_j^H$ overestimates the mean travel time. In the Toronto EMS dataset, samples are drawn every 200m or 240s, a combination of sampling-by-distance and sampling-by-time. However, the distance constraint is usually satisfied first (see Figure 6, where the sampled GPS points are regularly spaced). Thus, the travel time estimator

$\hat{T}_j^H$ is appropriate.

In our second local method, we assume $V_j^l \sim \text{Log-}N(m_j, s_j^2)$, independently across l, for unknown travel time parameters m_j and s_j^2. Again if the estimated travel time $T_j^l = L_j/V_j^l$, this implies $T_j^l \sim \text{Log-}N\left(\log(L_j) - m_j, s_j^2\right)$. We use the maximum likelihood estimators

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

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

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

This method provides a natural distribution estimate for the travel times via the estimated lognormal distribution for T_j^l. Correcting for zero-speed readings is again required, to avoid $\log(0)$, but the results are less dependent on the threshold, because the speeds are not inverted.

The MLE method is biased towards overestimating the travel times, because of extra variation in the observed speeds, caused by GPS speed errors [29] and departures from the assumption of constant speed on each road segment. Consider modeling this extra variation with a lognormal error. Let the observed speed $O_j^l = V_j^l \mathcal{E}$, where V_j^l is now the true speed at the GPS time, and $\mathcal{E} \sim \text{Log-}N\left(-\psi^2/2, \psi^2\right)$ (so $E(\mathcal{E}) = 1$). Then $O_j^l \sim \text{Log-}N\left(m_j - \psi^2/2, s_j^2 + \psi^2\right)$. The MLE method actually estimates $E\left(L_j/O_j^l\right)$ instead of $E\left(L_j/V_j^l\right) = E\left(T_j^l\right)$, causing overestimation, because $E\left(L_j/O_j^l\right) = E\left(T_j^l\right)\exp\left\{\psi^2\right\}$. There is no way to adjust for this bias if the parameters are estimated independently for each arc, as in our local methods. The speed variability s_j^2 and the error ψ^2 are not separately identified; only their sum is identified.

A complication for both local methods is that some small residential arcs have no assigned GPS points in the Toronto EMS dataset (see Figure 3). In this case, we use a breadth-first search on the nearby arcs to find an arc with data, and use its assigned GPS speeds instead. Each arc has a "road class," associated with its size. We restrict the search to arcs of the same class, because we assume that nearby arcs of the same class have similar speed distributions.

We also considered modifications to the two methods. First, we required each arc to have at least N assigned GPS readings (for example, $N = 10$). If an arc had fewer readings, we performed the same breadth-first search, saving any readings, until a total of N had been found. Second, because the small residential roads tend to have little data, we considered pooling all the GPS data for these arcs. However, neither of these modifications consistently improved performance.

5 Simulation Experiments

We test the Bayesian and local methods in the area of Leaside, Toronto, shown in Figure 3. The region is almost 4 square kilometers. In this region, a value $k = 6$ (see Section 3.5) guarantees that the Markov chain is irreducible and thus valid. This region has four classes of road size. 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 (see Figure 3).

In this section, we present a set of simulation experiments, comparing the accuracy of the three methods in predicting durations of test trips with known paths. We also evaluate the map-matching estimates from the Bayesian method for example simulated paths.

5.1 Generating Simulated Data

Some aspects of the data-generating mechanism follow the assumptions of our various models, while others do not. The arc travel times are lognormal: $T_i(j) \sim \text{Log-}N(\mu_j, \sigma_j^2)$. To set the true parameters μ_j and σ_j^2, we uniformly generate a speed between 20-40mph. We set μ_j and σ_j^2 so that the arc length divided by the mean travel time equals this speed, but also to give the arcs a range of travel time variances. Specifically, we draw $\sigma_j \sim \text{Unif}\left(0.5\log\left(\sqrt{3}\right), 0.5\log(3)\right)$, which then defines the required value of μ_j. This range for σ_j is narrower than the prior range (see Appendix A), but still allows for substantial differences in arc travel time variances. Comparisons between the three methods are invariant to moderate changes in the σ_j range.

We simulate ambulance trips with true paths, travel times, and GPS readings. For each trip i, we uniformly choose start and end nodes. We construct each path A_i arc-by-arc. At each node, beginning at the start node, we uniformly choose an arc that lowers the expected time to the end node, until the end node is reached. This method differs from the Bayesian prior (see Section 2.2). For some trips, the initial path in the Markov chain (see Section 3.1) is quite different from the true path, testing the mixing of the Markov chain in the space $\mathcal{P}_i$.

We simulate datasets with two types of GPS data: "good" and "bad." The good GPS data is 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 use the longer along-path distance). The GPS locations are drawn from a bivariate normal distribution with $\Sigma = \left(\begin{smallmatrix} 100 & 0 \\ 0 & 100 \end{smallmatrix}\right)$ (see Appendix A). The GPS speeds are drawn from a lognormal distribution with $\zeta^2 = 0.004$, corresponding to a mean absolute error of 5% of speed, roughly the average results seen by Witte and Wilson [29]. In the bad GPS data, the GPS points are sampled every 500m. The GPS error is designed to be slightly higher than that seen by Chen et al. [5] and Witte and Wilson [29]. The constant $\Sigma = \left(\begin{smallmatrix} 1145 & 0 \\ 0 & 1145 \end{smallmatrix}\right)$, consistent with the average error seen by Chen et al. [5], if it is in one dimension (see Appendix A). The parameter $\zeta^2 = 0.0227$, corresponding to mean absolute error of 12% of speed.

5.2 Travel Time Prediction

We simulate ten good GPS datasets and ten bad GPS datasets, as defined in Section 5.1, each with a training set of 2000 trips and a test set of 2000 trips (in total, roughly the size of the Toronto EMS Std dataset). Taking the true path for each test trip as known, we calculate point and 95% predictive interval estimates for the trip durations using the three methods. We use an empirical Bayes (data-based) travel time prior distribution for the Bayesian method (see Appendix A).

We compare the predictive accuracy of the point estimates from the three methods via the root mean squared error (RMSE, in seconds), the mean absolute error (MAE, in seconds), the geometric mean of the ratios of the predicted durations to the true durations (Ratio), and the correlation between the predicted and true durations (Cor.). The Ratio metric assesses the bias in the point estimates, and the Cor. metric assesses the variance. We compare the interval estimates using the arithmetic mean width of the 95% predictive intervals (Width) and the percentage of 95% predictive intervals that contain the true trip duration (Cov. %). Table 1 gives means for all these metrics over the ten replications of good and bad GPS datasets.

Good GPS data (Avg. over ten datasets)						
Estimation method	RMSE (s)	MAE (s)	Cor.	Ratio	Width (s)	Cov. %
Bayesian	14.9	11.2	0.953	1.016	59.0	96.0
Local MLE	15.5	11.7	0.949	1.021	58.5	94.8
Local Harm.	15.5	11.7	0.949	1.020	57.9	94.4
Bad GPS data (Avg. over ten datasets)						
Estimation method	RMSE (s)	MAE (s)	Cor.	Ratio	Width (s)	Cov. %
Bayesian	15.2	11.4	0.950	1.015	61.4	96.2
Local MLE	16.2	12.3	0.944	1.035	63.5	94.8
Local Harm.	16.2	12.2	0.944	1.033	62.3	94.3

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

In both dataset types, the point estimates from the Bayesian method outperform the estimates from the local methods. For a given good GPS dataset, improvement is typically 3-5% in RMSE and MAE, with an average of roughly 4%. For a bad GPS dataset, improvement is typically 4-8% in RMSE and MAE, with an average of roughly 7%. The interval estimates are very similar for the good GPS data. For the bad GPS data, the Bayesian intervals are superior: slightly narrower (1-3%) with higher coverage percentage. The two local methods performed almost identically in all metrics. Performance of the Bayesian method only worsened by 2% on average in RMSE and MAE from good to bad GPS data, whereas performance of the local methods worsened more.

5.3 Map-Matched Path Results

Next we assess map-matching estimates from the Bayesian method. Figure 2 shows two example ambulance paths. The black points show the GPS locations, and the white nodes follow the true path taken. The starting node is circled in green and the ending node in red. Each arc is colored by the marginal posterior probability it is traversed in the path. Arcs with probability less than 1% are uncolored. The left-hand path is from a good GPS dataset (as defined in Section 5.1). 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.

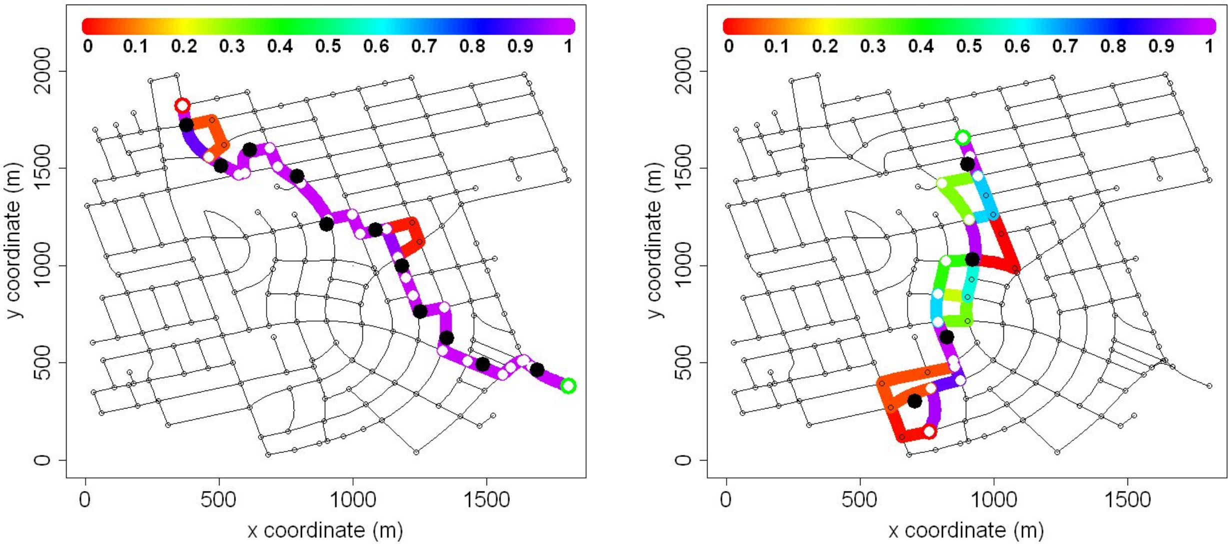

Figure 2: Map-matching estimates for two simulated trips, colored by the probability each arc is traversed.

The right-hand path is from a bad GPS dataset. Routes closer to the GPS points are preferred, because the GPS location likelihood increases (see Section 2.1), and routes with lower expected travel time are preferred, because the prior probability of the path increases (see Section 2.2). However, there is sometimes a tradeoff between these. For example, the last GPS point in the path has high location error. The correct shortest route has very high probability, even though there is an alternative route much closer to the GPS point, because the alternative route is much longer. Earlier in the path, there are two occasions where the GPS points are too sparse to define the route clearly. There are multiple routes with similar expected travel times, so each has substantial probability.

6 Analysis of Toronto EMS Data

In this section, we compare the Bayesian and local methods on the dataset from Toronto EMS.

6.1 Data

We use data provided by Toronto EMS from 2007 and 2008. The GPS data include locations, timestamps, speeds, and ambulance and emergency incident IDs. The priority of the emergency determines whether the ambulance travels at "lights-and-sirens" (L-S) or "standard" (Std) speed. We consider only the GPS points in the Leaside subregion of Toronto. The right plot in Figure 3 shows these GPS locations for the Std dataset. This dataset contains 3989 ambulance trips and almost 35,000 GPS points. On this region, the GPS location error is reasonably low; most of the points are close to a road, presumably the correct road. 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 (1932 trips), with a similar spatial distribution of points.

We use only the "to-scene" portion of each ambulance trip, and discard trips for which this cannot be identified, for ease of sorting the trips into the L-S and Std datasets. 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). Finally, most of the trips in the dataset do not begin or end in the subregion, they simply pass through. We assume that trips start and end at known nodes and times, so we must approximate these. We use the closest node to the first GPS location as the approximated start node, and the

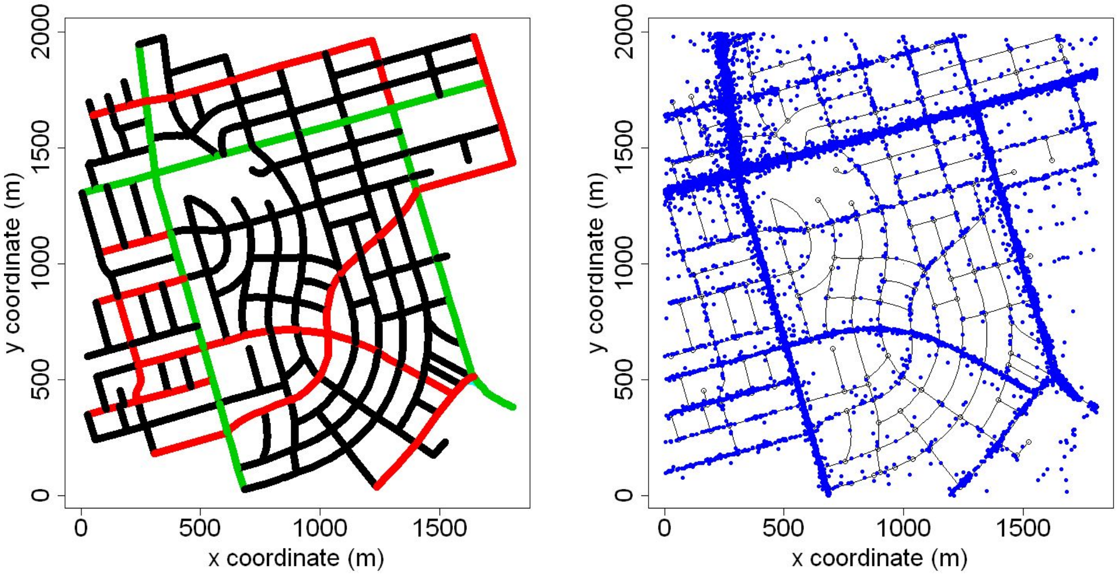

Figure 3: The subregion of Toronto (left), with primary roads (green), secondary roads (red) and tertiary roads (black). All GPS locations on this region from the standard travel dataset (right).

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.

Table 2 compares rush hour (7-9 AM and 4-6 PM weekdays), non-rush hour, and overall mean GPS speeds, for the L-S and Std datasets. L-S speeds are higher than Std speeds, and speeds decrease for both datasets during rush hour, with a greater difference in the Std data.

Mean GPS speed	All data	Rush hour	Non-rush hour
L-S	33.5 mph	31.5 mph	34.0 mph
Std	24.2 mph	21.7 mph	25.0 mph

Table 2: Mean observed GPS speeds for L-S and Std datasets.

6.2 Travel Speed Estimation

Toronto EMS has existing estimates of the travel times, which we use as the prior means as described in Appendix A. These estimates are different for L-S and Std trips, but are the same for the two travel directions on each road segment. We have also tested the Bayesian method with the data-based prior described in Appendix A and have observed almost as good performance.

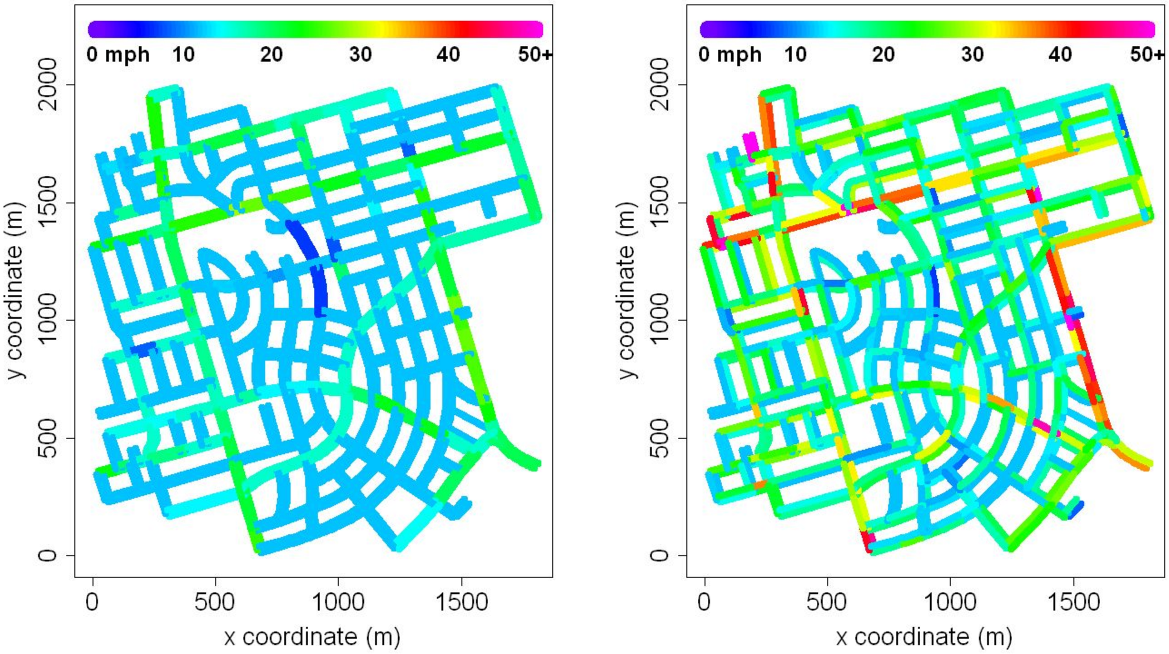

Figure 4: Prior (left) and posterior (right) speeds from the Bayesian method, for Toronto L-S data.

Figure 4 shows prior and posterior speed estimates (length divided by mean travel time) from the Bayesian method on the L-S dataset. Each arc is colored based on its speed estimate, so most roads have two colors, corresponding to the two travel directions.

The posterior speed estimates from the Bayesian method are reasonable; primary arcs tend to have high speed estimates, and estimated speeds for successive arcs on the same road are typically similar. Arcs heading into major intersections (primary and secondary roads in Figure 3) are often slower than the reverse arcs. This effect cannot be captured by the local methods, because both travel directions have the same estimates. Arcs with little data tend to be close to the prior distribution. The prior appears to underestimate the true speeds, because arcs with a large amount of data generally have faster, more reasonable posterior estimates. This is a desirable property for the prior, because incorrectly estimating high speeds for arcs with little data would adversely affect fastest path estimation, which is our goal.

There are a few arcs that have poor estimates from the Bayesian method. For example, parallel pink 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 for these roads [8], which has the effect of pooling the available data.

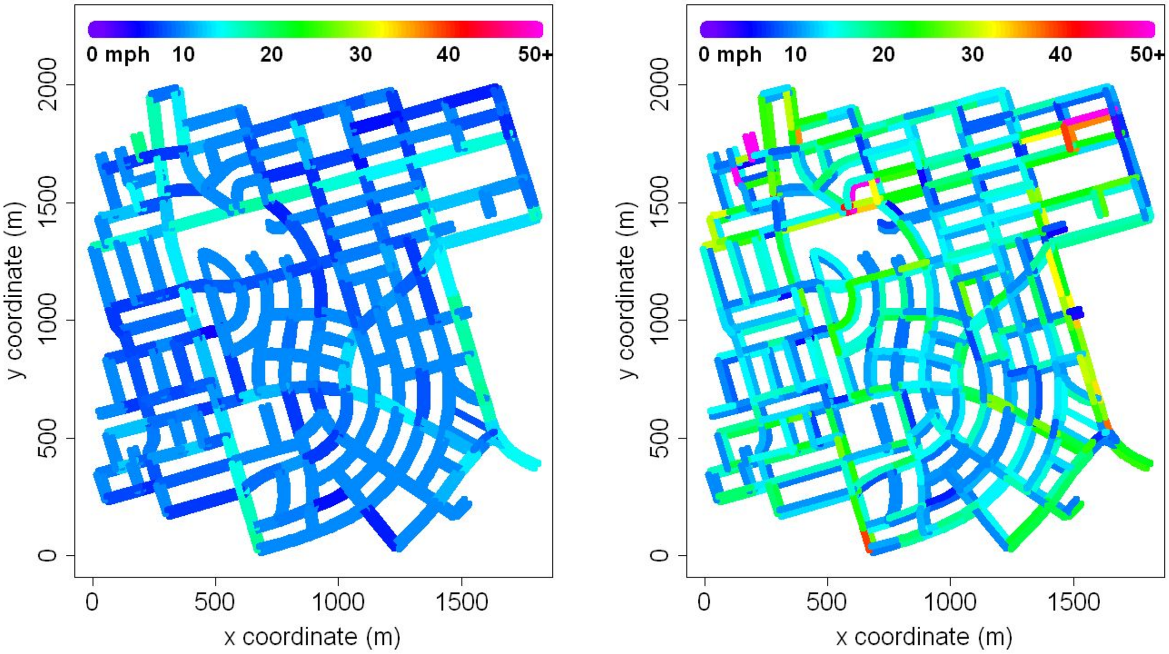

Figure 5: Prior (left) and posterior (right) speeds from the Bayesian method, for Toronto Std data.

Figure 5 shows the same two plots for the Std data. Again, the posterior estimates are generally reasonable. Successive arcs typically have similar speeds, and the posterior speeds for primary roads are typically faster than the prior speeds. Roads entering large intersections are now substantially slower, because ambulances must wait at stoplights in standard travel. For example, see the intersection at location (500, 600). There are again some arcs with poor estimates, such as the curved pink arc at (600, 1500). There is no data on this arc (except for some GPS errors), and some trips have mistakenly been estimated to leave the main road and use it instead. Again, this could be addressed via a random effect prior distribution [8].

6.3 Travel Time Prediction

Next we evaluate the accuracy of travel time predictions from the Bayesian and local methods. We divide the set of trips from each dataset randomly into two halves to create training and test sets, fit the models to the training data, and predict the travel times for the trips in the test data. We also check whether the predictive accuracy is improved by fitting the model separately to rush-hour and non-rush hour training data.

We compare the known duration of each trip in the test data with the point and 95% interval predictions from each method. Unlike the simulated test data in Section 5, the true paths are not known. We use point estimates for the path taken; for each estimation method, we assume the path taken is the the expected fastest path, using the mean travel time estimates for each arc. This measures the ability of the methods to estimate both the fastest path and the travel time distributions accurately.

As in Section 5.2, we compare the point estimates from the three methods on the test data using RMSE, MAE, Cor. and Ratio, and compare the interval estimates using Width and Cov. %. Mean results from using each half of the dataset as the training data are given in Table 3.

For the L-S data, the Bayesian point estimates perform better than the local method estimates in all metrics. RMSE is slightly lower, while MAE is roughly 5% lower. The Bayesian estimates have less bias and higher correlation to the true values. The interval estimates from the Bayesian method are far superior, having much higher coverage percentage while being narrower than the intervals from the harmonic method. The intervals from the Bayesian method are wider than the intervals from the MLE method, because they account for uncertainty in the travel time parameters. Partitioning the data into time bins does not substantially change performance.

L-S data (Avg. over the two test sets)						
Estimation method	RMSE (s)	MAE (s)	Cor.	Ratio	Width (s)	Cov. %
Bayesian (1 time bin)	41.1	23.2	0.759	1.01	76.0	85.7
Local MLE (1 time bin)	41.4	24.5	0.752	1.04	54.5	66.3
Local Harm. (1 time bin)	41.5	25.0	0.752	1.06	96.2	62.3
Bayesian (2 time bins)	41.0	23.4	0.759	1.02	77.0	85.0
Local MLE (2 time bins)	41.6	24.2	0.752	1.02	51.9	64.4
Local Harm. (2 time bins)	41.6	24.7	0.752	1.04	91.3	61.9
Std data (Avg. over the two test sets)						
Estimation method	RMSE (s)	MAE (s)	Cor.	Ratio	Width	Cov. %
Bayesian (1 time bin)	130.1	74.6	0.607	0.91	158.1	72.6
Local MLE (1 time bin)	136.3	78.8	0.597	0.87	112.5	55.8
Local Harm. (1 time bin)	134.8	77.8	0.598	0.89	172.0	66.4
Bayesian (2 time bins)	125.4	70.7	0.641	0.93	158.3	73.9
Local MLE (2 time bins)	135.1	77.6	0.625	0.85	105.8	54.3
Local Harm. (2 time bins)	133.6	76.6	0.625	0.87	163.4	67.0

Table 3: Out-of-sample trip travel time estimation performance on Toronto EMS data.

For the Std data, all estimation methods are less successful. The Bayesian estimates again outperform both local methods in all metrics. Improvement is roughly 4-5% in RMSE and MAE. Again, the interval estimates show substantial improvement. In this case, partitioning the data into time bins decreases RMSE and MAE by 3-5% and substantially increases correlation. We believe this is because the travel time distributions in the two time bins differ more in the Std data than in the L-S data, and the Std dataset is also larger. The local methods only improve by 1-2%, indicating that these methods may be more prone to overfitting.

The two local methods perform comparably. If a simple-to-implement solution is desired, we recommend the MLE method. Interval estimates from this method perform substantially better for the L-S data, which is the more important case, and are simpler to obtain, because they do not require sampling from all the GPS data. Also, estimates are less sensitive to the correction for zero-speed readings (see Section 4).

6.4 Map-Matched Path Accuracy

Finally, we assess map-matching estimates from the Bayesian method. Figure 6 shows two example ambulance paths from the L-S dataset. The first GPS point is colored green, the last red, and the others black. As in Section 5.3, each arc is colored by its marginal posterior probability. Arcs with less than 1% probability are uncolored. In the left-hand path, the ambulance has taken a surprising route, but the posterior estimates seem correct, except that the assumed start node is incorrect (an example of edge error; see Section 6.1). Most paths in the dataset have almost 100% probability on all the (presumably) correct arcs, as in this example. 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 (see Section 5.3), following a primary and then secondary road. This illustrates the robustness of the Bayesian method to poor GPS data.

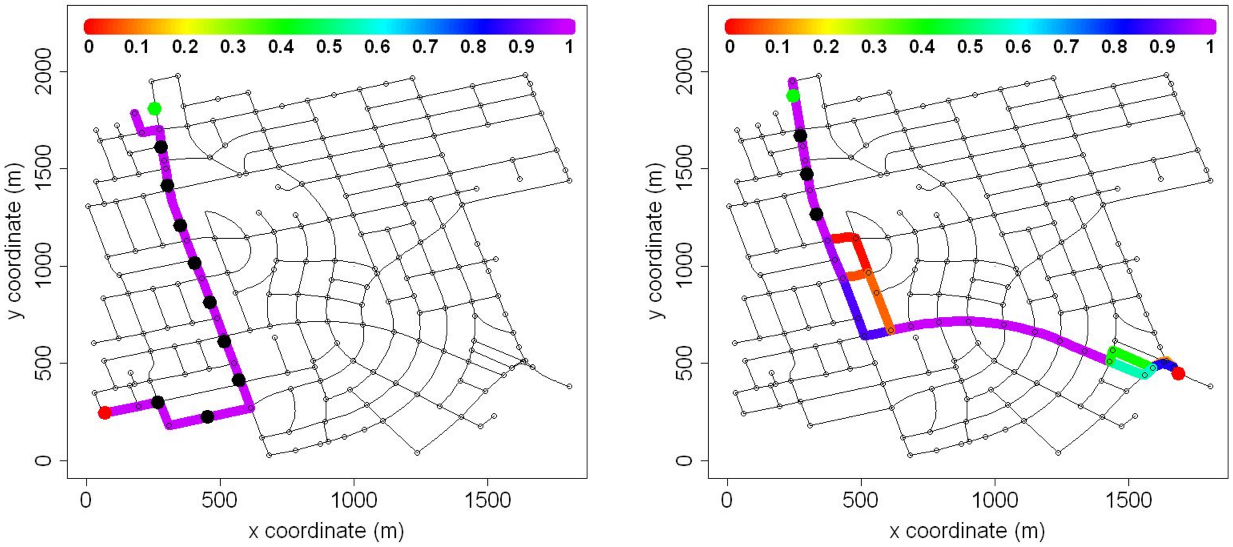

Figure 6: Map-matching estimates for two Toronto L-S trips, colored by the probability each arc is traversed.

7 Conclusions

We proposed a Bayesian method to estimate ambulance travel time distributions on each road segment in a city, using sparse and error-prone GPS data. We simultaneously estimated the ambulance paths and the parameters of the travel time distributions on each segment. We also introduced two 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 applied the three methods to simulated data and data from Toronto EMS, in a subregion of Toronto. In simulations, the Bayesian method outperformed the local methods in estimating out-of-sample trip durations, 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 provided reasonable travel time estimates for the road segments in the network, and again outperformed the competing methods in out-of-sample prediction.

We also applied the estimation methods independently to rush hour and non-rush hour L-S and Std data. This binning had little effect on predictive accuracy for any of the methods on the L-S data, but did improve accuracy on the Std data, especially for the Bayesian method. One could also consider smooth estimation of the travel times across the different times of day, using functional data analysis approaches. This could be effective in regions of a city where travel times vary more substantially between rush and non-rush hour.

Regarding other possible extensions, the lognormal travel time assumption of the Bayesian method could be weakened to a mixture of lognormals. This could capture the different travel time distributions at intersections, depending on the light cycle and direction the ambulance turns. Additionally, random effect modeling of travel time parameters within a road class could provide more smoothing [8]. Thirdly, the GPS location covariance Σ could be made a function of location, accounting for the larger error downtown.

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 the error is radially symmetric and the nearest arc is correct (and straight), 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 $\Sigma = \left(\begin{smallmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{smallmatrix}\right)$. Thus, we take $\sigma = \hat{E}(|\mathcal{E}_1|)\sqrt{\pi/2}$ where $\hat{E}(|\mathcal{E}_1|)$ is the average minimum distance of each GPS point to the nearest arc. In the Toronto EMS datasets, we have mean minimum distances 8.4m in the L-S data and 7.7m in the Std data. In the simulated data, we have roughly 7.3m for "good" GPS data and 18.7m for "bad" GPS data. These give $\Sigma_{\text{L-S}} = \left(\begin{smallmatrix} 111.6 & 0 \\ 0 & 111.6 \end{smallmatrix}\right)$, $\Sigma_{\text{Std}} = \left(\begin{smallmatrix} 92.7 & 0 \\ 0 & 92.7 \end{smallmatrix}\right)$, $\Sigma_{\text{Good}} = \left(\begin{smallmatrix} 84 & 0 \\ 0 & 84 \end{smallmatrix}\right)$, and $\Sigma_{\text{Bad}} = \left(\begin{smallmatrix} 550 & 0 \\ 0 & 550 \end{smallmatrix}\right)$.

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 (see Equation (2)). We assume we have an initial travel time estimate τ_j for each arc j. When this 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. The hyperparameter m_j is set so that $E\left(T_i(j)\right) = \tau_j$. Given the choices for s^2, b_1, and b_2 (specified below), this requires

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

The variance s^2 is set by estimating the possible range of the mean travel time about the prior estimate τ_j. An interval $\{m_j \pm 2s\}$ of two standard deviations in μ_j increases and decreases $E\left(T_i(j) \,\middle|\, \mu_j, \sigma_j^2\right)$ by a factor of e^{2s}. We set s^2 so that this factor equals two ($s = 0.5\log(2)$), because we expect our prior estimate typically to be correct to within a factor of two.

The hyperparameters b_1 and b_2 are set by estimating the possible range in travel time variation for a single arc. Given μ_j, 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 significant 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 $b_2 = 0.6264$.

Results are very insensitive to the hyperparameters b_3 and b_4, so 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 $b_4 = 0.5$; the latter corresponds to mean GPS speed error of 55%. 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 (1) 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 $\hat{a}_i$ from d_i^1 to d_i^2, where the expected travel time of $\tilde{a}_i$ is 10 seconds less than the expected travel time of $\hat{a}_i$. 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 duration, a driver is ten times less likely to choose a path that has 10% longer travel time. However, we make a small adjustment to ensure that the prior distribution on the route taken between two fixed locations is roughly the same for each dataset, because the route choices appear to be the same from visual inspection of the L-S and Std data. 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 duration for each dataset $\mathcal{D}$. We estimate a weighted mean duration $T_W^{\mathcal{D}} = T^{\mathcal{D}} L_1^{\text{Tor}} / L_1^{\mathcal{D}}$ for dataset $\mathcal{D}$ for a trip of length L_1^{Tor}, and use the duration $T_W^{\mathcal{D}}$ to set C

by the above principle. This yields $C_{\text{L-S}} = 0.211$, $C_{\text{Std}} = 0.110$, and $C_{\text{Sim}} = 0.208$.

Interestingly, the value of C influences an overall bias in the Bayesian method. If C decreases, there is a general decrease in predicted travel times. We believe this is because if C is lower, the preference for faster paths is weaker, so alternative slower paths have relatively higher posterior probabilities. Alternative paths tend to be longer in distance. Since the total trip duration is known, these longer paths require faster estimated travel speeds. This leads to a general reduction in travel time estimates.

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, 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, 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 [18, 21, 27]. 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. [22]). 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

$$E(v_i(\phi_i)\,|T_i\,) = \int_0^{T_i} v_i(t)\frac{1}{T_i}dt = \frac{1}{T_i}. \tag{3}$$

In particular this implies that the speeds summarized in Table 2 are biased high. 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

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

by Jensen's Inequality and (3). Again, the inequality is strict unless $v_i(\cdot)$ is constant.

C Calculations for Updating the Paths

Here we calculate the ratios of posteriors f, proposals q, and Jacobian $|J|$ from Section 3.2. First, for the ratio of posteriors,

$$\frac{f\left(A_i^*, T_i^* \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L, \zeta^2\right)}{f\left(A_i, T_i \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L, \zeta^2\right)} = \frac{\pi\left(A_i^* \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)}{\pi\left(A_i \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)} \frac{\ell\left(T_i^* \middle| A_i^*, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)}{\ell\left(T_i \middle| A_i, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)} \frac{g\left(A_i^*, T_i^* \middle| \zeta^2\right)}{g\left(A_i, T_i \middle| \zeta^2\right)},$$

where $\pi(\cdot)$ is the prior probability of the path, $\ell(\cdot)$ is the product of the travel time likelihoods for each arc in the path, and $g(\cdot)$ is the product of the GPS location and speed likelihoods for all GPS readings in the path. For the ratio of priors, the denominator of (1) cancels, because

A_i^* and A_i have the same start node and end node. Thus,

$$\frac{\pi\left(A_i^* \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)}{\pi\left(A_i \middle| \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)} = \frac{\exp\left\{-C\sum_{l=1}^n \theta(p_l)\right\}}{\exp\left\{-C\sum_{l=1}^m \theta(c_l)\right\}}.$$

Only the arcs in the update section are changed, so the ratio of travel time likelihoods is

$$\frac{\ell\left(T_i^* \middle| A_i^*, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)}{\ell\left(T_i \middle| A_i, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)} = \frac{\prod_{l=1}^n \text{Log-}N\left(T_i^*(p_l); \mu_{p_l}, \sigma_{p_l}^2\right)}{\prod_{l=1}^m \text{Log-}N\left(T_i(c_l); \mu_{c_l}, \sigma_{c_l}^2\right)}.$$

For the ratio of GPS likelihoods, consider a single GPS reading $\left(X_i^l, Y_i^l, V_i^l, t_i^l\right)$. Given the state $\{A_i, T_i\}$, and the assumption of constant speed on each arc, infer the ambulance position $\left(\hat{X}_i^l, \hat{Y}_i^l\right)$ and speed $\hat{V}_i^l$ at time t_i^l. Let l_1 be the index of the first GPS reading in the update section, and l_2 be the index of the last. Then, letting N_2 denote the bivariate normal density,

$$\frac{g\left(A_i^*, T_i^* \left|\zeta^2\right.\right)}{g\left(A_i, T_i \left|\zeta^2\right.\right)} = \frac{\prod_{h=l_1}^{l_2} N_2\left(\left(X_i^h, Y_i^h\right); \left(\hat{X}_i^{h*}, \hat{Y}_i^{h*}\right), \Sigma\right) \text{Log-}N\left(V_i^h; \log\left(\hat{V}_i^{h*}\right) - \frac{\zeta^2}{2}, \zeta^2\right)}{\prod_{h=l_1}^{l_2} N_2\left(\left(X_i^h, Y_i^h\right); \left(\hat{X}_i^h, \hat{Y}_i^h\right), \Sigma\right) \text{Log-}N\left(V_i^h; \log\left(\hat{V}_i^h\right) - \frac{\zeta^2}{2}, \zeta^2\right)}.$$

Next we calculate the ratio of proposals. In Part 1 of the proposal in Section 3.2, the node d' is chosen with probability $1/N_i$, where N_i is the number of arcs in A_i. In Part 2, the node d'' is chosen with probability $1/\min(b, k)$. In Part 3, the number of routes of length up to k between d' and d'' is the same for the reverse proposal, so this probability cancels. Finally, the ratio of travel time proposals can be calculated easily. Letting Dir denote the Dirichlet density, we have

$$\frac{q\left(A_i, T_i \middle| A_i^*, T_i^*, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)}{q\left(A_i^*, T_i^* \middle| A_i, T_i, \left\{\mu_j, \sigma_j^2\right\}_{j=1}^L\right)} = \frac{N_i \min(b, k)}{N_i^* \min(b^*, k)} \frac{\text{Dir}\left(\frac{T_i(c_1)}{S_i}, \ldots, \frac{T_i(c_m)}{S_i}; \alpha\theta(c_1), \ldots, \alpha\theta(c_m)\right)}{\text{Dir}\left(r_1, \ldots, r_n; \alpha\theta(p_1), \ldots, \alpha\theta(p_n)\right)} S_i^{n-m}.$$

Finally, to calculate the Jacobian $|J|$, define random variables $U_l = r_l S_i$, for $l \in \{1, \ldots, n-1\}$ (emphasizing that the space of travel time proposals has dimension $n-1$). To take the same role for the reverse proposal, define $W_l = T_i(c_l)$, for $l \in \{1, \ldots, m-1\}$. Thus, we have a transformation between two spaces of dimension $m+n-1$, with Jacobian

$$|J| = \begin{vmatrix} \frac{\partial T_i(c_1)}{\partial T_i^*(p_1)} & \cdots & \frac{\partial T_i(c_m)}{\partial T_i^*(p_1)} & \frac{\partial U_1}{\partial T_i^*(p_1)} & \cdots & \frac{\partial U_{n-1}}{\partial T_i^*(p_1)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial T_i(c_1)}{\partial T_i^*(p_n)} & \cdots & \frac{\partial T_i(c_m)}{\partial T_i^*(p_n)} & \frac{\partial U_1}{\partial T_i^*(p_n)} & \cdots & \frac{\partial U_{n-1}}{\partial T_i^*(p_n)} \\ \frac{\partial T_i(c_1)}{\partial W_1} & \cdots & \frac{\partial T_i(c_m)}{\partial W_1} & \frac{\partial U_1}{\partial W_1} & \cdots & \frac{\partial U_{n-1}}{\partial W_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial T_i(c_1)}{\partial W_{m-1}} & \cdots & \frac{\partial T_i(c_m)}{\partial W_{m-1}} & \frac{\partial U_1}{\partial W_{m-1}} & \cdots & \frac{\partial U_{n-1}}{\partial W_{m-1}} \end{vmatrix} = 1,$$

by cofactor expansion.

Acknowledgements

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

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

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

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

[12] 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.

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

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

[15] 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.

[16] 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.

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

[18] 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.

[19] S. Richardson and P.J. Green. On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59:731–792, 1997.

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

[21] 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.

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

[23] L. Sun, J. Yang, and H. Mahmassani. Travel time estimation based on piecewise truncated quadratic speed trajectory. *Transportation Research Part A: Policy and Practice*, 42:173–186, 2008.

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

[25] 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.

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

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

[28] C.E. White, D. Bernstein, and A.L. Kornhauser. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8:91–108, 2000.

[29] 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.

[30] C.H. Wu, J.M. Ho, and D.T. Lee. Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*, 5:276–281, 2004.