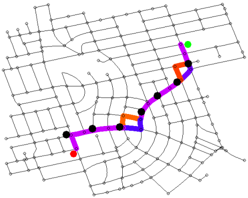
Predicting Travel Time on Road Networks

Dawn Woodard
Senior Data Science Manager of Dynamic Pricing
Uber

SIAM 2017



Sponsors & Collaborators

Chiwei Yan, Nikita Korolko, Brad Westgate, Galina Nogin, Shane Henderson,
David Matteson, Paul Koch, David Racz, Moises Goldszmidt, Eric Horvitz

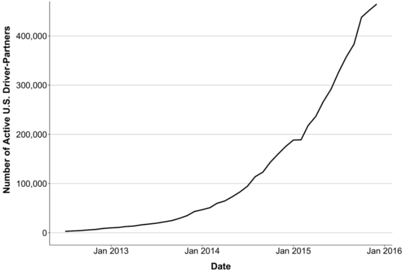
Outline

Outline

Hypergrowth

Rapid growth of ride-sharing platforms due to data-driven marketplace tech

- Efficient matching
- Calibrating demand with supply through pricing



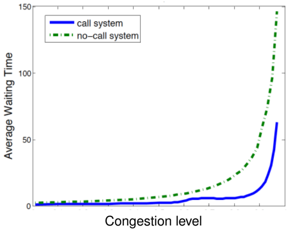
Source: Hall and Kreuger (2016)

Hypergrowth

Southern California Growth

Higher Driver Efficiency and Lower Rider Wait Times

Percent of miles driven with a passenger



Lower waiting time than street-hailing via intelligent dispatch

Source: Cramer and Krueger (2016); Feng, Kong, and Wang (2017)

Fundamentals of Ride-Sharing Market

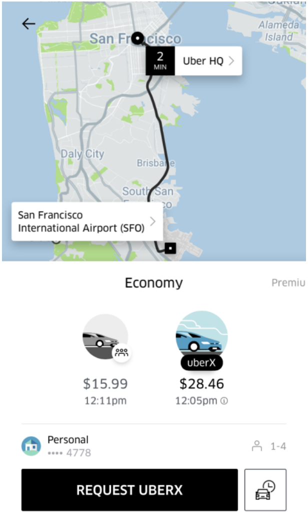
- **Two-sided market:**
  - Riders must be provided with both service and prices that are comparable or better than their alternatives.
  - Drivers must be able to plan on consistent earnings that are comparable or better than their alternatives.
- **Geographically interconnected:** drivers moving to one part of the city means they are not available elsewhere

Outline

Dynamic Pricing

How should a ride be priced, to calibrate supply and demand?

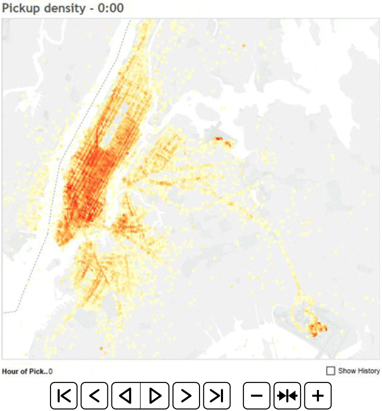
Price is optimized using a simplified ride-sharing model, using predicted demand and supply



Demand Forecast

Pricing requires prediction of demand and supply over time and geolocation

NYC taxi pickup data



Source: Daulton, Raman, Kindt

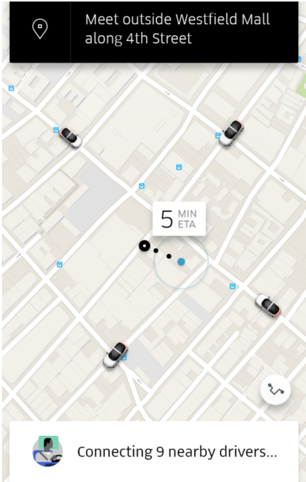
Outline

1. **Pricing and Matching in Ride-Sharing**
   - Dynamic Pricing
   - Matching
2. **Travel Time Reliability Prediction**
3. **Methods**
4. **Case Study**
5. **Summary**

Matching

How should riders be matched with open drivers?

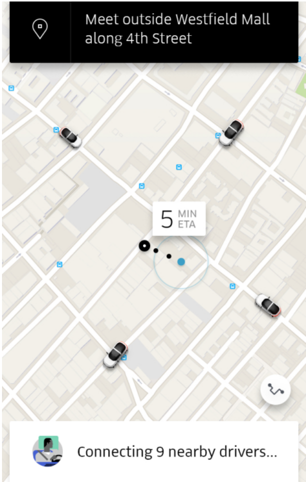
How should carpool riders be matched with each other and with drivers?



Matching

How should riders be matched with open drivers?

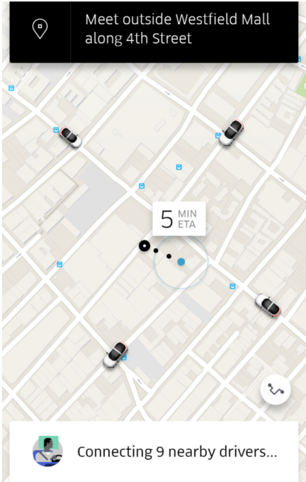
- Can be done efficiently by immediately dispatching the driver with the shortest pickup time
- Improved further by mechanisms like "Trip Swap"



Matching

How should riders be matched with open drivers?

- Can be done efficiently by immediately dispatching the driver with the shortest pickup time
- Improved further by mechanisms like "Trip Swap"

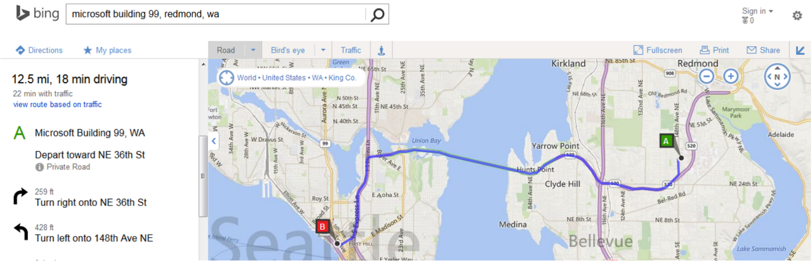


Trip Swap

Trip swap

Predicting Travel Time Reliability

Matching and pricing require prediction of travel time between two points



Predicting Travel Time Reliability

Deterministic predictions are never perfectly accurate, due to:

- Uncertainty in traffic light schedules
- Unexpected traffic and weather conditions
- Differences in driver behavior

Probabilistic prediction takes into account travel time uncertainty

- **Robust Matching**
  - Penalize the chance of a long pickup time or bad carpool match
  - Ex: Dispatch the driver with the lowest value of the 90th percentile of pickup time
- **Report travel time reliability** to a rider or driver
  - Range for travel time (example: 10-15 mins)
  - Percentile of travel time (example: 80th percentile)

Predicting Travel Time Reliability

Deterministic predictions are never perfectly accurate, due to:

- Uncertainty in traffic light schedules
- Unexpected traffic and weather conditions
- Differences in driver behavior

Probabilistic prediction takes into account travel time uncertainty

- **Robust Matching**
  - Penalize the chance of a long pickup time or bad carpool match
  - Ex: Dispatch the driver with the lowest value of the 90th percentile of pickup time
- **Report travel time reliability** to a rider or driver
  - Range for travel time (example: 10-15 mins)
  - Percentile of travel time (example: 80th percentile)

Predicting Travel Time Reliability

Deterministic predictions are never perfectly accurate, due to:

- Uncertainty in traffic light schedules
- Unexpected traffic and weather conditions
- Differences in driver behavior

Probabilistic prediction takes into account travel time uncertainty

- **Robust Matching**
  - Penalize the chance of a long pickup time or bad carpool match
  - Ex: Dispatch the driver with the lowest value of the 90th percentile of pickup time
- **Report travel time reliability** to a rider or driver
  - Range for travel time (example: 10-15 mins)
  - Percentile of travel time (example: 80th percentile)

Outline

1. Pricing and Matching in Ride-Sharing
   - Dynamic Pricing
   - Matching
2. **Travel Time Reliability Prediction**
3. Methods
4. Case Study
5. Summary

Travel Time Prediction

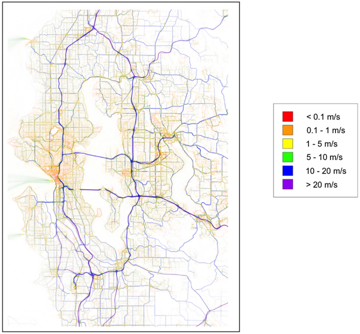
Companies that require travel time predictions:

Travel Time Prediction

Travel time prediction uses mobile phone GPS data

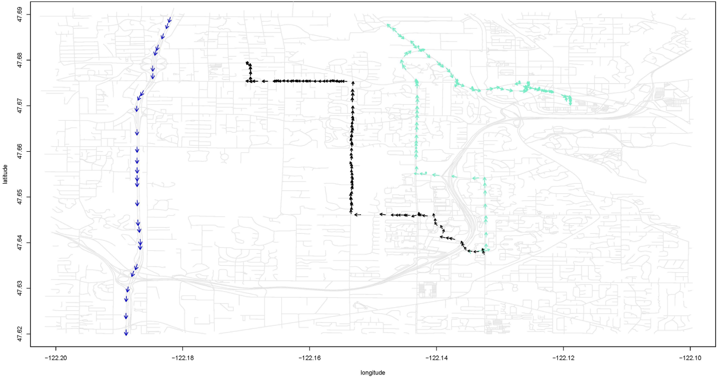
- Many companies now have access to user location data
- The only source of information about traffic & travel time that can achieve near-comprehensive coverage of the road network
- Increasing evidence that traffic conditions can be estimated accurately using only such data (Work et al. 2010)

Anonymized Windows phone GPS locations for the Seattle metropolitan region, colored by speed:



Mapping Services

Isolate vehicle trips as sequences of GPS points with high measured speed. Examples:



Predicting Travel Time Reliability

Goal

Using GPS data from vehicles traveling on the road network, predict the probability distribution of travel time on an arbitrary route in the network, at a given time.

Challenges:

- Large number of possible routes
- Small number of trips in the data that follow any particular route
- Dependence of the travel time on time of day, traffic, and other effects

Predicting Travel Time Reliability

Goal

Using GPS data from vehicles traveling on the road network, predict the probability distribution of travel time on an arbitrary route in the network, at a given time.

Challenges:

- Large number of possible routes
- Small number of trips in the data that follow any particular route
- Dependence of the travel time on time of day, traffic, and other effects

Predicting Travel Time Reliability

Goal

Using GPS data from vehicles traveling on the road network, predict the probability distribution of travel time on an arbitrary route in the network, at a given time.

Challenges:

- Large number of possible routes
- Small number of trips in the data that follow any particular route
- Dependence of the travel time on time of day, traffic, and other effects

Predicting Travel Time Reliability

Goal

Using GPS data from vehicles traveling on the road network, predict the probability distribution of travel time on an arbitrary route in the network, at a given time.

Challenges:

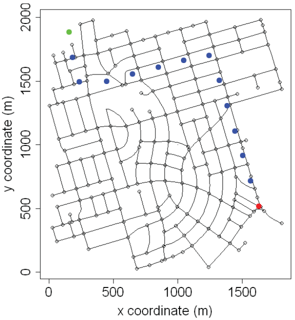
- Large number of possible routes
- Small number of trips in the data that follow any particular route
- Dependence of the travel time on time of day, traffic, and other effects

Outline

Data Processing

Snap the GPS trace to the road network. For each trip i this yields:

- Route $R_i = (R_{i,1}, \ldots, , R_{i,n_i})$ where $R_{i,k}$ is the kth link traversed
- Distance $d_{i,k}$ traversed on each link
- Time $T_{i,k}$ spent traversing link $R_{i,k}$



Predicting Travel Time Reliability

To accurately predict travel time reliability for commercial use, an approach must:

- **Give informed predictions** for parts of the road network with little data
- **Capture weekly cycles** in congestion levels
- **Be computationally efficient** even for large road networks & datasets
- **Accurately capture dependence** between the travel time on the road links in the route
  - Ex: If the speed is high on the first half of the trip, it is likely to be high on the second half

Predicting Travel Time Reliability

To accurately predict travel time reliability for commercial use, an approach must:

- **Give informed predictions** for parts of the road network with little data
- **Capture weekly cycles** in congestion levels
- **Be computationally efficient** even for large road networks & datasets
- **Accurately capture dependence** between the travel time on the road links in the route
  - Ex: If the speed is high on the first half of the trip, it is likely to be high on the second half

Predicting Travel Time Reliability

To accurately predict travel time reliability for commercial use, an approach must:

- **Give informed predictions** for parts of the road network with little data
- **Capture weekly cycles** in congestion levels
- **Be computationally efficient** even for large road networks & datasets
- **Accurately capture dependence** between the travel time on the road links in the route
  - Ex: If the speed is high on the first half of the trip, it is likely to be high on the second half

Predicting Travel Time Reliability

To accurately predict travel time reliability for commercial use, an approach must:

- **Give informed predictions** for parts of the road network with little data
- **Capture weekly cycles** in congestion levels
- **Be computationally efficient** even for large road networks & datasets
- **Accurately capture dependence** between the travel time on the road links in the route
  - Ex: If the speed is high on the first half of the trip, it is likely to be high on the second half

Predicting Travel Time Reliability

To accurately predict travel time reliability for commercial use, an approach must:

- **Give informed predictions** for parts of the road network with little data
- **Capture weekly cycles** in congestion levels
- **Be computationally efficient** even for large road networks & datasets
- **Accurately capture dependence** between the travel time on the road links in the route
  - Ex: If the speed is high on the first half of the trip, it is likely to be high on the second half

Statistical Modeling

"TRIP": Travel time Reliability Inference & Prediction

Model the travel time $T_{i,k}$ of trip i on link $R_{i,k}$ as

$$T_{i,k} = \frac{d_{i,k}}{E_i S_{i,k}}$$

Speed variability decomposed into trip-level variability and link-level variability:

- Trip effect E_i: due e.g. to traffic conditions affecting whole trip.

$$\log(E_i) \sim \mathcal{N}(0, \tau^2)$$

- Link effect $S_{i,k}$: due e.g. to local traffic conditions. Model it conditional on an unobserved congestion state $Q_{i,k} \in \{1, \ldots, \mathcal{Q}\}$:

$$\log(S_{i,k}) | Q_{i,k} \sim \mathcal{N}(\mu_{R_{i,k}, Q_{i,k}}, \sigma^2_{R_{i,k}, Q_{i,k}})$$

Statistical Modeling

"TRIP": Travel time Reliability Inference & Prediction

Model the travel time $T_{i,k}$ of trip i on link $R_{i,k}$ as

$$T_{i,k} = \frac{d_{i,k}}{E_i S_{i,k}}$$

Speed variability decomposed into trip-level variability and link-level variability:

- Trip effect E_i: due e.g. to traffic conditions affecting whole trip.

$$\log(E_i) \sim \mathcal{N}(0, \tau^2)$$

- Link effect $S_{i,k}$: due e.g. to local traffic conditions. Model it conditional on an unobserved congestion state $Q_{i,k} \in \{1, \ldots, \mathcal{Q}\}$:

$$\log(S_{i,k}) | Q_{i,k} \sim \mathcal{N}(\mu_{R_{i,k}, Q_{i,k}}, \sigma^2_{R_{i,k}, Q_{i,k}})$$

Statistical Modeling

"TRIP": Travel time Reliability Inference & Prediction

Model the travel time $T_{i,k}$ of trip i on link $R_{i,k}$ as

$$T_{i,k} = \frac{d_{i,k}}{E_i S_{i,k}}$$

Speed variability decomposed into trip-level variability and link-level variability:

- Trip effect E_i: due e.g. to traffic conditions affecting whole trip.

$$\log(E_i) \sim \mathcal{N}(0, \tau^2)$$

- Link effect $S_{i,k}$: due e.g. to local traffic conditions. Model it conditional on an unobserved congestion state $Q_{i,k} \in \{1, \ldots, \mathcal{Q}\}$:

$$\log(S_{i,k})|Q_{i,k} \sim \mathcal{N}(\mu_{R_{i,k},Q_{i,k}}, \sigma^2_{R_{i,k},Q_{i,k}})$$

Statistical Modeling

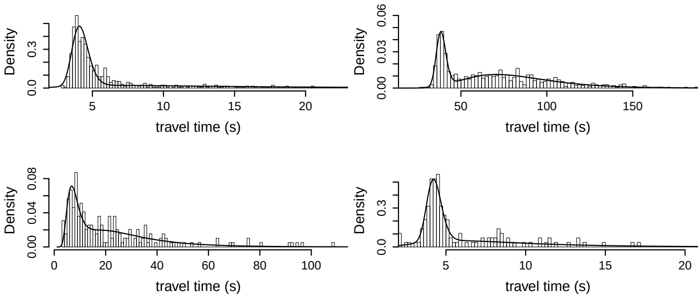
A Markov model for the congestion states $Q_{i,k}$:

$$\text{Pr}(Q_{i,1} = q) = p^{(0)}_{R_{i,1},b_{i,1}}(q)$$
$$\text{Pr}(Q_{i,k} = q | Q_{i,k-1} = \tilde{q}) = p_{R_{i,k},b_{i,k}}(\tilde{q}, q)$$

Captures weekly cycles in congestion levels, and dependence of congestion across links of trip

Statistical Modeling

Yields a normal mixture model for log travel time on a link, capturing the heavy right skew and multimodality in the data:



4 links with the most data: histogram = training data, curve = predicted density

Computation

- **Maximum a posteriori (MAP) parameter estimation**; i.e., maximize the density of

$$\theta = (\{\mu_{j,q}, \sigma^2_{j,q}, p^{(0)}_{j,b}, p_{j,b}\}, \tau^2, \{\log E_i\})$$

conditional on the data $\{\log T_{i,k}\}$

- **Computation by Expectation Conditional Maximization:**
  - Closed-form updates
  - Estimation time: 15-36 mins on a single processor (Seattle data)
  - Prediction time: 17 ms for single trip (fast enough for commercial mapping services)

Computation

- **Maximum a posteriori (MAP) parameter estimation**; i.e., maximize the density of

$$\theta = (\{\mu_{j,q}, \sigma^2_{j,q}, p^{(0)}_{j,b}, p_{j,b}\}, \tau^2, \{\log E_i\})$$

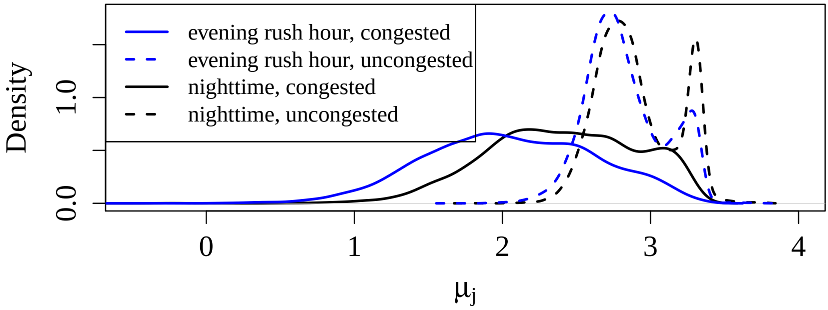
  conditional on the data $\{\log T_{i,k}\}$

- **Computation by Expectation Conditional Maximization:**
  - Closed-form updates
  - Estimation time: 15-36 mins on a single processor (Seattle data)
  - Prediction time: 17 ms for single trip (fast enough for commercial mapping services)

Outline

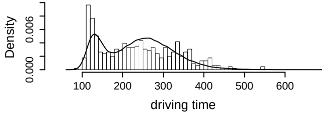
Seattle Case Study

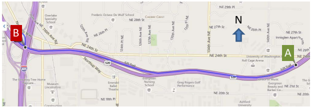
Distribution of estimated speed parameter over roads (network links):

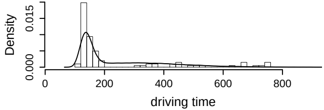


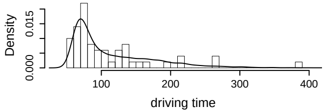
Seattle Case Study

Three routes in the road network. Histogram = travel times from test data (PM rush hour); Curve = predictive density







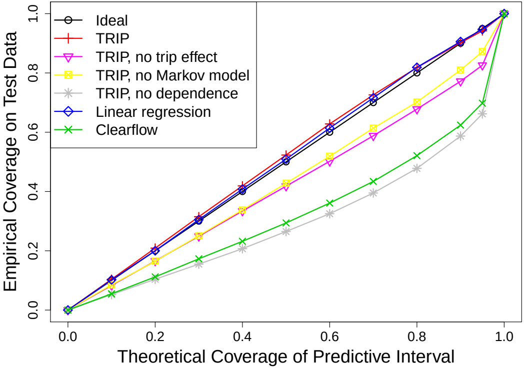


Comparisons

1. **Versions of our method lacking** one or both of the **dependencies**
2. **Microsoft's prediction method ("Clearflow"):**
   - Used in Bing Maps
   - Models distribution of travel time on each link based on:
     - Traffic measurements from roadway sensors
     - Speed limit, road class
     - Proximity to schools, shopping areas, stadiums
     - ...
3. **Regression-based methods:**
   - Regression of trip travel time on route distance, time of week, speed limit, etc.

Seattle Case Study

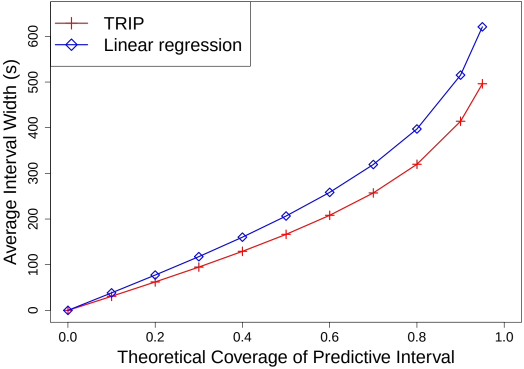
Coverage of predictive intervals on test data (35,190 trips on network of 221,980 links):



⇒ Methods that assume independence across links underpredict variability

Seattle Case Study

Avg. width of predictive intervals on test data, for methods with accurate coverage:



⇒ Interval predictions from TRIP are 19-21% narrower

Seattle Case Study

Performance of deterministic predictions:

	TRIP	TRIP, no trip effect	TRIP, no Markov model	TRIP, no dependence	Clearflow	Linear regression
On all test data:						
% error	10.1	9.6	10.0	9.8	10.4	12.8
% error w/ bias correction	9.5	9.3	9.4	9.3	9.7	12.8

$\Rightarrow$ Deterministic predictions from TRIP are slightly better than Clearflow.

$\Rightarrow$ Linear regression does poorly

Outline

Summary

- Methods for **probabilistic prediction of travel times in a road network, using mobile phone GPS data**
- Yields far better interval predictions than Clearflow, and slightly better deterministic predictions
- Application: **matching and pricing for ride-sharing**

Marketplace @ Uber

- statisticians, economists, operations researchers, ML scientists...
- developing Uber's marketplace decision systems...
  - dynamic pricing
  - dispatch & carpool matching
- and the inputs that feed into those systems:
  - predicted demand and supply
  - predicted travel times

Contact:
dawn@uber.com
`people.orie.cornell.edu/woodard`

Marketplace @ Uber

- statisticians, economists, operations researchers, ML scientists...
- developing Uber's marketplace decision systems...
  - dynamic pricing
  - dispatch & carpool matching
- and the inputs that feed into those systems:
  - predicted demand and supply
  - predicted travel times

Contact:
dawn@uber.com
`people.orie.cornell.edu/woodard`

Marketplace @ Uber

- statisticians, economists, operations researchers, ML scientists...
- developing Uber's marketplace decision systems...
  - dynamic pricing
  - dispatch & carpool matching
- and the inputs that feed into those systems:
  - predicted demand and supply
  - predicted travel times

Contact:
dawn@uber.com
`people.orie.cornell.edu/woodard`

Motivation for Model

- Autocorrelation of travel times within a trip is high, decreasing with distance

  $\Rightarrow$ Markov model for $Q_{i,k}$

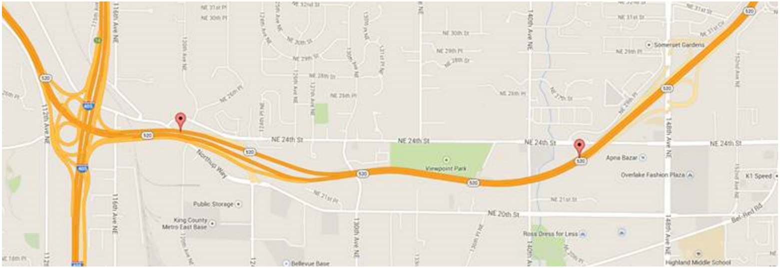
- Correlation of travel times for co-located vehicles is not consistently high
  - Due to: lining up to take an exit or turn, HOV lanes, ...

- "Congestion level is a property of the trip, not just the roads driven"

  $\Rightarrow$ $Q_{i,k}$ depends on the trip

Motivation for Model

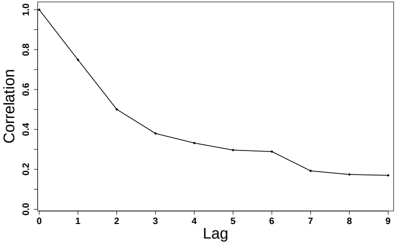
Example: sequence of 10 links on highway 520 West:



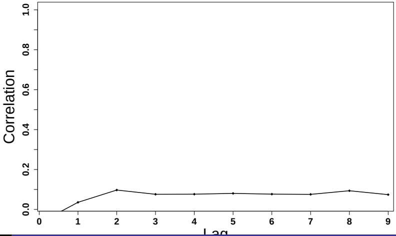
Motivation for Model

"Congestion level is a property of the trip, not just the roads driven"

Correlation of log(travel time) of first link with other links, **within same trip**:



Correlation of log(travel time) of first link with other links, **different trips**:



Motivation for Model

Reasons: HOV lanes, lining up to take an exit, . . .

Statistical Modeling

So the **median travel time** in time bin 0 is as follows, where $s_{c(j)}$ is the unknown speed parameter for links having road class $c(j)$:

$$b + \sum_{j \in R_i} \underbrace{d_{ij}/s_{c(j)}}_{\text{baseline travel time on link } j}$$

Intercept b captures, e.g., time to get up to speed at the beginning of the trip (Kolesar et al. 1975).

Toronto Case Study

We also investigated taking into account uncertainty in the routes driven by vehicles in the training data when fitting the travel time model

Statistical Methods

One-stage estimation:

- estimate all unknowns $\{R_i\}, \theta, \phi$ using the posterior distribution

$$\pi\left(\{R_i\}, \theta, \phi \,\middle|\, \{G_i\}, \{T_i\}\right) \propto \pi(\theta)\pi(\phi) \prod_i \left[f(R_i|g(\theta))f(G_i|R_i, \phi)f(T_i|R_i, \theta)\right]$$

Two-stage estimation:

- Obtain rough estimates $\hat{g}(\theta)$ of relevant summaries of travel time
- Obtain route estimates $\{\hat{R}_i\}$ by maximizing the route posterior $\pi(\{R_i\}|\{G_i\}, \hat{g}(\theta)) \propto \int \pi(\phi) \prod_i [f(R_i|\hat{g}(\theta))f(G_i|R_i, \phi)]\, d\phi$
- Conditional on the travel times T_i and estimated routes $\hat{R}_i$, obtain the posterior distribution of θ: $\pi(\theta|\{T_i, \hat{R}_i\}) \propto \pi(\theta) \prod_i f(T_i|\hat{R}_i, \theta)$

Route Modeling

- **Multinomial logit choice model for the route**:

  $$f(R_i|\theta) \propto \exp\{-C \times \mathsf{E}(T_i|R_i, \theta)\}$$

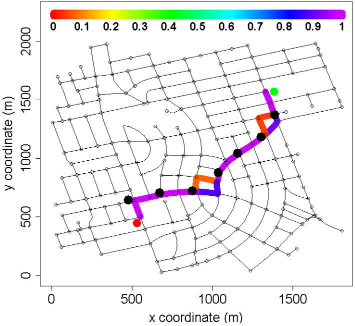
  for fixed $C > 0$

- In two-stage estimation we need an estimate of $\mathsf{E}(T_i|R_i, \theta)$ for the first stage
  - take the speed on each link to be the geometric mean of measured speeds from GPS readings closest to that link

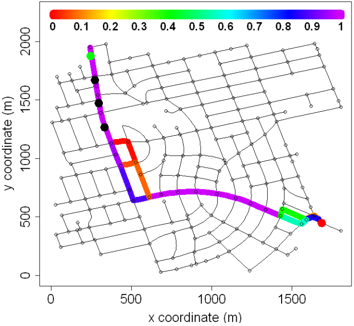
- **Model** $f(G_i|R_i, \phi)$ **for the GPS data:** assume that the distance of each measured location to the path is exponentially distributed.

Toronto Data, Route Estimation Results

Probability that each link was traversed, for two ambulance trips:



Trip w/ low GPS error



Trip w/ large GPS gap

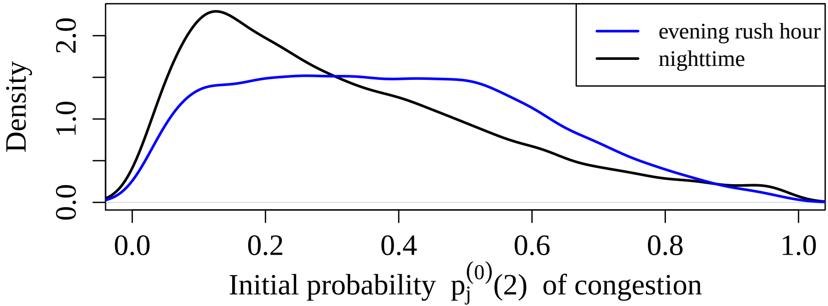
Application to Toronto EMS

Driving time predictive performance (out-of-sample) on a Toronto subregion:

Method	RMSE (s)	RMSE log	Cov. %	Width (s)
One-stage estimation (link-based model)	37.8	.332	85.8	75.0
Two-stage estimation	38.1	.331	91.3	90.3

Seattle Case Study

Distribution of estimated congestion probability over roads (network links):



Seattle Case Study

Bias of deterministic predictions:

	TRIP	TRIP, no trip effect	TRIP, no Markov model	TRIP, no dependence	Clearflow	Linear regression
On all test data:	.030	.014	.028	.024	.033	-.005
On parts of network with little data:	.108	.102	.105	.101	.066	.077

$\Rightarrow$ Bias is low overall ($< 3.4\%$) for all the methods, but higher on parts of network with little data.