Probabilistic Forecasting of Bus Travel Time with a Bayesian Gaussian Mixture Model

Xiaoxu Chen[1,4] Zhanhong Cheng[1,4] Jian Gang Jin[2]
Martin Trépanier[3,4] Lijun Sun[1,4]

[1]Department of Civil Engineering, McGill University
[2]School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University
[3]Department of Mathematics and Industrial Engineering, Polytechnique Montreal
[4]Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

November 7, 2022

Outline

1. **Introduction**
   - Background
   - Challenges
2. **Methodology**
   - Problem description
   - Augmented random variable
   - Bayesian multivariate Gaussian mixture model
   - Model inference: Gibbs sampling
   - Probabilistic forecasting
3. **Experiments**
   - Data and experiment settings
   - Forecasting performance
   - Interpreting mixture components
   - Predicted distribution
4. **Conclusion**

Background

- Bus travel time forecasting and its reliability/uncertainty are important.
- Passengers: make better travel plans.
  - Departure time
  - Route choice
  - Transport mode choice
- Bus agencies: design robust bus management strategies.
  - Bus timetable
  - Bus priority signal control
  - Bus bunching control
- Most studies mainly center on making point estimation (i.e, deterministic forecasting).

Probabilistic forecasting for bus travel time

Background

- Bus travel time forecasting and its reliability/uncertainty are important.
- Passengers: make better travel plans.
  - Departure time
  - Route choice
  - Transport mode choice
- Bus agencies: design robust bus management strategies.
  - Bus timetable
  - Bus priority signal control
  - Bus bunching control
- Most studies mainly center on making point estimation (i.e, deterministic forecasting).

Probabilistic forecasting for bus travel time

Background

- Bus travel time forecasting and its reliability/uncertainty are important.
- Passengers: make better travel plans.
  - Departure time
  - Route choice
  - Transport mode choice
- Bus agencies: design robust bus management strategies.
  - Bus timetable
  - Bus priority signal control
  - Bus bunching control
- Most studies mainly center on making point estimation (i.e, deterministic forecasting).

Probabilistic forecasting for bus travel time

Background

- Bus travel time forecasting and its reliability/uncertainty are important.
- Passengers: make better travel plans.
  - Departure time
  - Route choice
  - Transport mode choice
- Bus agencies: design robust bus management strategies.
  - Bus timetable
  - Bus priority signal control
  - Bus bunching control
- Most studies mainly center on making point estimation (i.e, deterministic forecasting).

Probabilistic forecasting for bus travel time

Challenges

Key point: construct the probability distribution for bus travel time.

- Complex correlations among different links (local and long-range correlations).
- Strong interactions between two adjacent buses (e.g, bus bunching).
- Bus travel time distributions are usually not normal and exhibit long-tailed and multimodal characteristics.
- Real-world data often have many missing/ragged values.

Challenges

Key point: construct the probability distribution for bus travel time.

- Complex correlations among different links (local and long-range correlations).
- Strong interactions between two adjacent buses (e.g, bus bunching).
- Bus travel time distributions are usually not normal and exhibit long-tailed and multimodal characteristics.
- Real-world data often have many missing/ragged values.

Challenges

Key point: construct the probability distribution for bus travel time.

- Complex correlations among different links (local and long-range correlations).
- Strong interactions between two adjacent buses (e.g, bus bunching).
- Bus travel time distributions are usually not normal and exhibit long-tailed and multimodal characteristics.
- Real-world data often have many missing/ragged values.

Challenges

Key point: construct the probability distribution for bus travel time.

- Complex correlations among different links (local and long-range correlations).
- Strong interactions between two adjacent buses (e.g, bus bunching).
- Bus travel time distributions are usually not normal and exhibit long-tailed and multimodal characteristics.
- Real-world data often have many missing/ragged values.

Problem Description

- **Bus link**: the directional road segment connecting two adjacent stops on a bus route.
- **Link travel time**: travel time of a bus link, including the dwell time.
- **Trip travel time**: sum of several link travel times.
- **Objective**: forecast the travel time of a bus on its upcoming links/trips and providing ETA distribution.

Problem Description

- **Bus link**: the directional road segment connecting two adjacent stops on a bus route.
- **Link travel time**: travel time of a bus link, including the dwell time.
- **Trip travel time**: sum of several link travel times.
- **Objective**: forecast the travel time of a bus on its upcoming links/trips and providing ETA distribution.

Problem Description

- **Bus link**: the directional road segment connecting two adjacent stops on a bus route.
- **Link travel time**: travel time of a bus link, including the dwell time.
- **Trip travel time**: sum of several link travel times.
- **Objective**: forecast the travel time of a bus on its upcoming links/trips and providing ETA distribution.

Problem Description

- **Bus link**: the directional road segment connecting two adjacent stops on a bus route.
- **Link travel time**: travel time of a bus link, including the dwell time.
- **Trip travel time**: sum of several link travel times.
- **Objective**: forecast the travel time of a bus on its upcoming links/trips and providing ETA distribution.

Augmented Random Variable

- $\ell_{i,m}$: the link travel time of the i-th bus on the m-th link.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i = [\ell_{i,1}, \ell_{i,2}, \cdots, \ell_{i,n}]^\top$.
- $h_{i,m}$: the headway between the i-th bus pair at the m-th bus stop.
- Define an augmented random variable x to capture correlations between two adjacent buses.
- Specifically, bus i and its leading bus $i-1$ produce a sample of x:

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{\ell}_i \\ \boldsymbol{\ell}_{i-1} \\ \boldsymbol{h}_i \end{bmatrix} = [\ell_{i,1}, \cdots, \ell_{i,n}, \ell_{i-1,1}, \cdots, \ell_{i-1,n}, h_{i,1}, \cdots, h_{i,n}]^\top.$$

Augmented Random Variable

- $\ell_{i,m}$: the link travel time of the i-th bus on the m-th link.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i = [\ell_{i,1}, \ell_{i,2}, \cdots, \ell_{i,n}]^\top$.
- $h_{i,m}$: the headway between the i-th bus pair at the m-th bus stop.
- Define an augmented random variable x to capture correlations between two adjacent buses.
- Specifically, bus i and its leading bus $i-1$ produce a sample of x:

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{\ell}_i \\ \boldsymbol{\ell}_{i-1} \\ \boldsymbol{h}_i \end{bmatrix} = [\ell_{i,1}, \cdots, \ell_{i,n}, \ell_{i-1,1}, \cdots, \ell_{i-1,n}, h_{i,1}, \cdots, h_{i,n}]^\top.$$

Augmented Random Variable

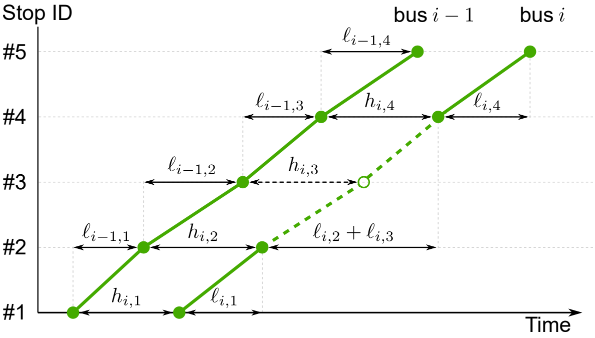
- $\ell_{i,m}$: the link travel time of the i-th bus on the m-th link.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i = [\ell_{i,1}, \ell_{i,2}, \cdots, \ell_{i,n}]^\top$.
- $h_{i,m}$: the headway between the i-th bus pair at the m-th bus stop.
- Define an augmented random variable $\boldsymbol{x}$ to capture correlations between two adjacent buses.
- Specifically, bus i and its leading bus $i-1$ produce a sample of $\boldsymbol{x}$:

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{\ell}_i \\ \boldsymbol{\ell}_{i-1} \\ \boldsymbol{h}_i \end{bmatrix} = [\ell_{i,1}, \cdots, \ell_{i,n}, \ell_{i-1,1}, \cdots, \ell_{i-1,n}, h_{i,1}, \cdots, h_{i,n}]^\top.$$

Augmented Random Variable

- The inherent relationship between link travel time and headway:

$$h_{i,m+1} - h_{i,m} + \ell_{i-1,m} - \ell_{i,m} = 0, \quad m = 1, \ldots, n-1.$$



Augmented Random Variable

- Ragged values are also constrains.
- Constrains can be summarized into linear equations:

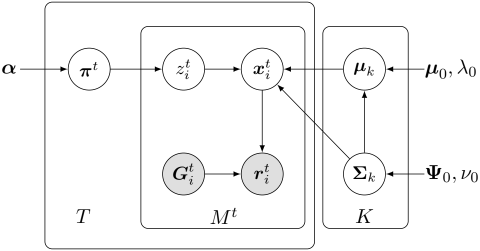
$$\underbrace{\begin{bmatrix} 1&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&0&0&1&0&0&0\\ 1&0&0&0&-1&0&0&0&-1&1&0&0\\ 0&1&0&0&0&-1&0&0&0&-1&1&0\\ 0&0&1&0&0&0&-1&0&0&0&-1&1 \end{bmatrix}}_{\mathbf{G}_i} \boldsymbol{x}_i = \begin{bmatrix} \ell_{i,1}\\ \ell_{i,2}+\ell_{i,3}\\ \ell_{i,4}\\ \ell_{i-1,1}\\ \ell_{i-1,2}\\ \ell_{i-1,3}\\ \ell_{i-1,4}\\ h_{i,1}\\ h_{i,2}-h_{i,1}+\ell_{i,1}-\ell_{i-1,1}\\ h_{i,3}-h_{i,2}+\ell_{i,2}-\ell_{i-1,2}\\ h_{i,4}-h_{i,3}+\ell_{i,3}-\ell_{i-1,3} \end{bmatrix} = \underbrace{\begin{bmatrix} r_{i,1}\\ r_{i,2}\\ r_{i,3}\\ r_{i,4}\\ r_{i,5}\\ r_{i,6}\\ r_{i,7}\\ r_{i,8}\\ 0\\ 0\\ 0 \end{bmatrix}}_{\boldsymbol{r}_i}$$

- $\mathbf{G}_i$ (*alignment matrix*) and $\boldsymbol{r}_i$ (*recording vector*) for bus i.
- Task 1: model $p(\boldsymbol{x})$ using historical $\{\mathbf{G}_i\}$ and $\{\boldsymbol{r}_i\}$.
- Task 2: use $p(\boldsymbol{x})$ and observed links to forecast upcoming links.

Bayesian Multivariate Gaussian Mixture Model

- **Data/sampling distribution**: $p^t\left(\boldsymbol{x}^t\right)=\sum_{k=1}^{K}\pi_k^t\mathcal{N}\left(\boldsymbol{x}^t\mid\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\right)$
- **Prior distributions**:

$$\boldsymbol{\pi}^t\sim\text{Dirichlet}\left(\boldsymbol{\alpha}\right)$$
$$\boldsymbol{\Sigma}_k\sim\mathcal{W}^{-1}\left(\boldsymbol{\Psi}_0,\nu_0\right)$$
$$\boldsymbol{\mu}_k\sim\mathcal{N}\left(\boldsymbol{\mu}_0,\frac{1}{\lambda_0}\boldsymbol{\Sigma}_k\right)$$
$$z_i^t\sim\text{Categorical}\left(\boldsymbol{\pi}^t\right)$$
$$\boldsymbol{x}_i^t\mid z_i^t=k\sim\mathcal{N}\left(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\right)$$
$$\boldsymbol{r}_i^t=\mathbf{G}_i^t\boldsymbol{x}_i^t$$



Model Inference: Gibbs Sampling

- Sample $\boldsymbol{\pi}_t$ from $p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right) \sim \text{Dirichlet}\left(M_1^t + \alpha_1, M_2^t + \alpha_2, \cdots, M_K^t + \alpha_K\right).$$

- Sample z_i^t from $p\left(z_i^t \mid \boldsymbol{\pi}_i^t, \boldsymbol{\mu}, \Sigma, \boldsymbol{x}_i^t\right)$.

$$p\left(z_i^t = k \mid \boldsymbol{\pi}^t, \boldsymbol{\mu}, \Sigma, \boldsymbol{x}_i^t\right) = \frac{\pi_k^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_k, \Sigma_k\right)}{\sum_{m=1}^K \pi_m^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_m, \Sigma_m\right)}.$$

- Sample $(\boldsymbol{\mu}_k, \Sigma_k)$ from $p\left(\boldsymbol{\mu}, \Sigma \mid \mathcal{X}_k, \Theta\right)$.

$$p\left(\boldsymbol{\mu}_k, \Sigma_k \mid \mathcal{X}_k, \Theta\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_k \mid \boldsymbol{\mu}_0^*, \frac{1}{\lambda_0^*}\Sigma_k\right)\mathcal{W}^{-1}\left(\Sigma_k \mid \Psi_0^*, \nu_0^*\right),$$

- Sample $\mathcal{X}$ from $p\left(\mathcal{X} \mid \boldsymbol{\mu}, \Sigma, \boldsymbol{z}, \mathcal{R}, \mathcal{G}\right)$.

$$\boldsymbol{x}_i^t \mid z_i^t = k \sim \mathcal{N}_{\mathcal{S}_i^t}\left(\boldsymbol{\mu}_k, \Sigma_k\right), \quad \mathcal{S}_i^t = \left\{\boldsymbol{x}_i^t \mid \mathbf{G}_i^t \boldsymbol{x}_i^t = \boldsymbol{r}_i^t\right\}.$$

Model Inference: Gibbs Sampling

- Sample $\boldsymbol{\pi}_t$ from $p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right) \sim \text{Dirichlet}\left(M_1^t + \alpha_1, M_2^t + \alpha_2, \cdots, M_K^t + \alpha_K\right).$$

- Sample z_i^t from $p\left(z_i^t \mid \boldsymbol{\pi}_i^t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x}_i^t\right)$.

$$p\left(z_i^t = k \mid \boldsymbol{\pi}^t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x}_i^t\right) = \frac{\pi_k^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{m=1}^K \pi_m^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\right)}.$$

- Sample $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathcal{X}_k, \Theta\right)$.

$$p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{X}_k, \Theta\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_k \mid \boldsymbol{\mu}_0^*, \frac{1}{\lambda_0^*}\boldsymbol{\Sigma}_k\right) \mathcal{W}^{-1}\left(\boldsymbol{\Sigma}_k \mid \boldsymbol{\Psi}_0^*, \nu_0^*\right),$$

- Sample $\mathcal{X}$ from $p\left(\mathcal{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{z}, \mathcal{R}, \mathcal{G}\right)$.

$$\boldsymbol{x}_i^t \mid z_i^t = k \sim \mathcal{N}_{\mathcal{S}_i^t}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad \mathcal{S}_i^t = \left\{\boldsymbol{x}_i^t \mid \mathbf{G}_i^t \boldsymbol{x}_i^t = \boldsymbol{r}_i^t\right\}.$$

Model Inference: Gibbs Sampling

- Sample $\boldsymbol{\pi}_t$ from $p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right) \sim \text{Dirichlet}\left(M_1^t + \alpha_1, M_2^t + \alpha_2, \cdots, M_K^t + \alpha_K\right).$$

- Sample z_i^t from $p\left(z_i^t \mid \boldsymbol{\pi}_i^t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x}_i^t\right)$.

$$p\left(z_i^t = k \mid \boldsymbol{\pi}^t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x}_i^t\right) = \frac{\pi_k^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{m=1}^K \pi_m^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\right)}.$$

- Sample $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathcal{X}_k, \Theta\right)$.

$$p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{X}_k, \Theta\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_k \mid \boldsymbol{\mu}_0^*, \frac{1}{\lambda_0^*}\boldsymbol{\Sigma}_k\right) \mathcal{W}^{-1}\left(\boldsymbol{\Sigma}_k \mid \boldsymbol{\Psi}_0^*, \nu_0^*\right),$$

- Sample $\mathcal{X}$ from $p\left(\mathcal{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{z}, \mathcal{R}, \mathcal{G}\right)$.

$$\boldsymbol{x}_i^t \mid z_i^t = k \sim \mathcal{N}_{\mathcal{S}_i^t}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad \mathcal{S}_i^t = \left\{\boldsymbol{x}_i^t \mid \mathbf{G}_i^t \boldsymbol{x}_i^t = \boldsymbol{r}_i^t\right\}.$$

Model Inference: Gibbs Sampling

- Sample $\boldsymbol{\pi}_t$ from $p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}^t \mid \boldsymbol{z}^t, \boldsymbol{\alpha}\right) \sim \text{Dirichlet}\left(M_1^t + \alpha_1, M_2^t + \alpha_2, \cdots, M_K^t + \alpha_K\right).$$

- Sample z_i^t from $p\left(z_i^t \mid \boldsymbol{\pi}_i^t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x}_i^t\right)$.

$$p\left(z_i^t = k \mid \boldsymbol{\pi}^t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{x}_i^t\right) = \frac{\pi_k^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)}{\sum_{m=1}^K \pi_m^t \mathcal{N}\left(\boldsymbol{x}_i^t \mid \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\right)}.$$

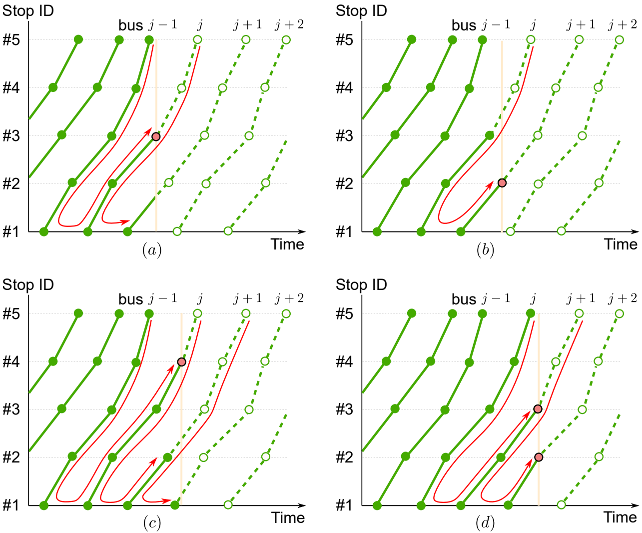
- Sample $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathcal{X}_k, \Theta\right)$.

$$p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{X}_k, \Theta\right) \sim \mathcal{N}\left(\boldsymbol{\mu}_k \mid \boldsymbol{\mu}_0^*, \frac{1}{\lambda_0^*}\boldsymbol{\Sigma}_k\right) \mathcal{W}^{-1}\left(\boldsymbol{\Sigma}_k \mid \boldsymbol{\Psi}_0^*, \nu_0^*\right),$$

- Sample $\mathcal{X}$ from $p\left(\mathcal{X} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{z}, \mathcal{R}, \mathcal{G}\right)$.

$$\boldsymbol{x}_i^t \mid z_i^t = k \sim \mathcal{N}_{\mathcal{S}_i^t}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad \mathcal{S}_i^t = \left\{\boldsymbol{x}_i^t \mid \mathbf{G}_i^t \boldsymbol{x}_i^t = \boldsymbol{r}_i^t\right\}.$$

Probabilistic Forecasting



Data and Experiment Settings

- Bus in-out-stop record data in Guangzhou, China
- Weekdays from December 1st, 2016 to December 31st, 2016
- Perform **data standardization** (z-score normalization)
- **Performance metrics**: RMSE, MAPE, LogS, CRPS
- Models in comparison:
  - **Model A**: $\boldsymbol{x}_i = [\boldsymbol{l}_i]$
  - **Model B**: $\boldsymbol{x}_i = \left[{\boldsymbol{\ell}_i}^\top, {\boldsymbol{\ell}_{i-1}}^\top\right]^\top$
  - **Model C**: $\boldsymbol{x}_i = \left[{\boldsymbol{\ell}_i}^\top, {\boldsymbol{\ell}_{i-1}}^\top, {\boldsymbol{h}_i}^\top\right]^\top$

Forecasting Performance

Table 1 **Performance of different models for link travel time conditional forecasting.**

		Observed links											
		5 links				10 links				15 links			
		RMSE	MAPE	CRPS	LogS	RMSE	MAPE	CRPS	LogS	RMSE	MAPE	CRPS	LogS
Model A	K = 1	33.9	0.1439	15.48	-4.495	31.8	0.1274	14.54	-4.413	27.9	0.1151	13.03	-4.367
	K = 2	33.8	0.1436	14.90	-4.553	32.1	0.1275	14.13	-4.698	27.9	0.1175	12.55	-4.418
	K = 5	34.1	0.1430	14.51	-4.456	32.6	0.1252	13.53	-4.855	29.6	0.1200	11.79	-4.288
Model B	K = 1	33.5	0.1369	15.02	-4.451	29.7	0.1142	13.26	-4.342	30.3	0.1179	13.32	-4.344
	K = 2	33.7	0.1442	14.86	-4.434	29.3	0.1171	12.89	-4.303	31.1	0.1233	13.07	-4.297
	K = 5	34.5	0.1387	14.51	-4.411	29.7	0.1148	12.34	-4.261	31.9	0.1245	12.12	-4.220
Model C	K = 1	33.0	0.1341	14.49	-4.422	29.3	0.1139	12.62	-4.306	31.9	0.1187	12.78	-4.273
	K = 2	**29.7**	**0.1252**	**13.11**	**-4.334**	**22.0**	0.0989	10.26	**-4.164**	**17.0**	0.0918	**7.93**	**-3.970**
	K = 5	30.3	0.1253	13.19	-4.341	22.1	**0.0986**	**10.22**	-4.171	17.1	**0.0874**	7.97	-3.990

Best results are highlighted in bold fonts.

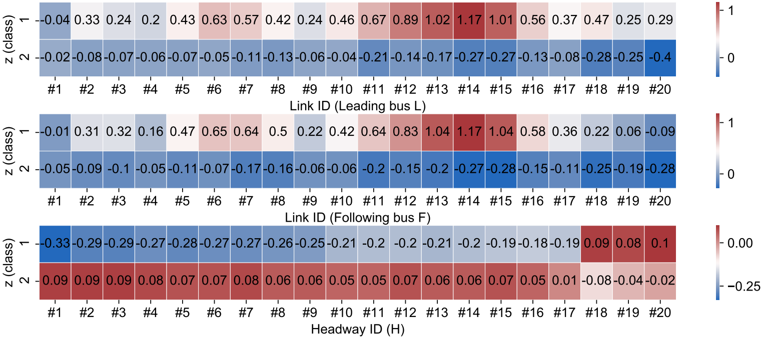
Table 2 **Performance of different models for trip travel time conditional forecasting.**

		Observed links											
		5 links				10 links				15 links			
		RMSE	MAPE	CRPS	LogS	RMSE	MAPE	CRPS	LogS	RMSE	MAPE	CRPS	LogS
Model A	K = 1	187.8	0.0789	110.23	-6.921	132.2	0.0860	77.42	-6.411	71.5	0.0863	43.87	-5.845
	K = 2	188.4	0.0790	105.97	-6.751	136.3	0.0877	76.14	-6.492	70.4	0.0855	41.70	-5.764
	K = 5	190.1	0.0790	106.96	-6.712	137.2	0.0876	73.86	-6.320	72.9	0.0878	37.24	-5.544
Model B	K = 1	177.4	0.0760	102.13	-6.696	119.9	0.0762	68.51	-6.272	70.9	0.0884	42.10	-5.780
	K = 2	182.1	0.0801	105.01	-6.709	117.7	0.0770	67.24	-6.254	73.6	0.0911	41.71	-5.720
	K = 5	185.1	0.0786	102.37	-6.704	117.0	0.0740	63.98	-6.180	74.0	0.0908	36.45	-5.584
Model C	K = 1	171.6	0.0713	96.51	-6.594	115.9	0.0729	64.17	-6.151	75.6	0.0909	40.41	-5.647
	K = 2	**149.5**	**0.0686**	**83.79**	**-6.443**	87.2	0.0651	48.46	-5.865	36.0	**0.0619**	19.42	-4.943
	K = 5	151.6	0.0694	84.77	-6.502	**86.1**	**0.0641**	**47.83**	**-5.850**	**35.8**	0.0625	**19.38**	**-4.931**

Best results are highlighted in bold fonts.

Interpreting Mixture Components

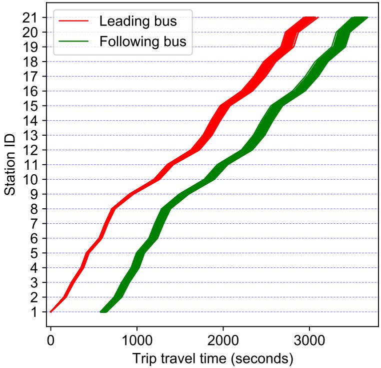
- The estimated **mean vector** (standardization)

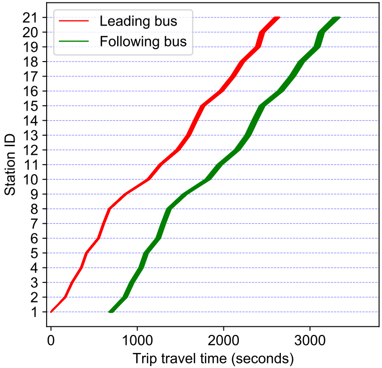


Significant differences in some links and many headways

Interpreting Mixture Components

- Distribution of the estimated trajectory

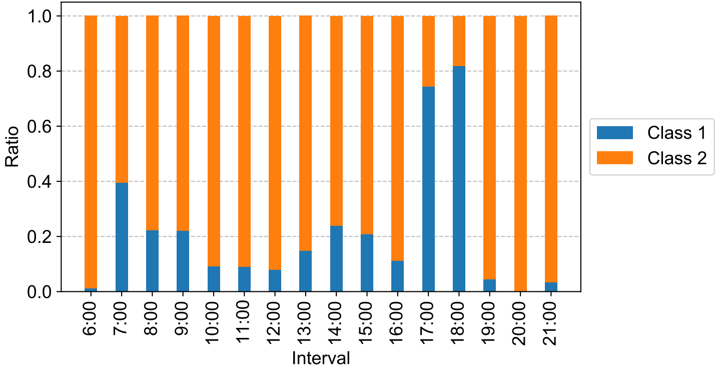




Class 1: longer link/trip travel times, shorter headways, larger variances

Interpreting Mixture Components

- **Component distribution** for different **intervals**

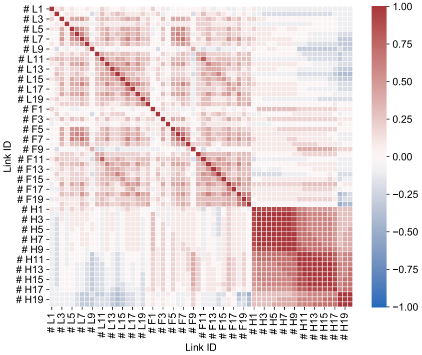


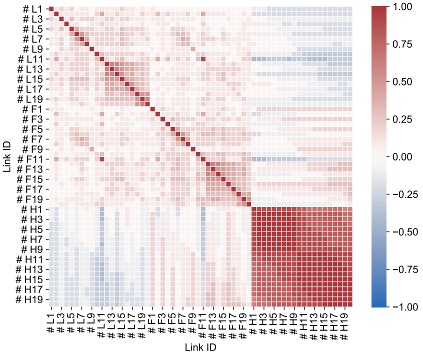
Class 1: dominant for **afternoon peak** hours

Class 2: dominant for **off-peak** and **morning peak** hours

Interpreting Mixture Components

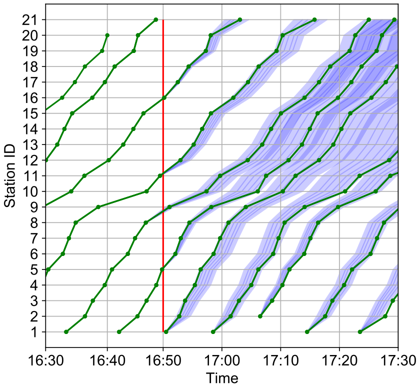
- **Correlation matrices** for different components



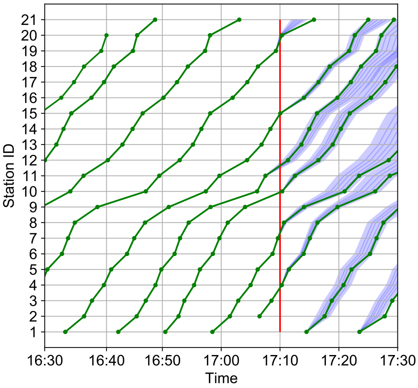


Class 1: leading bus and the following bus could be more correlated

Predicted Distribution

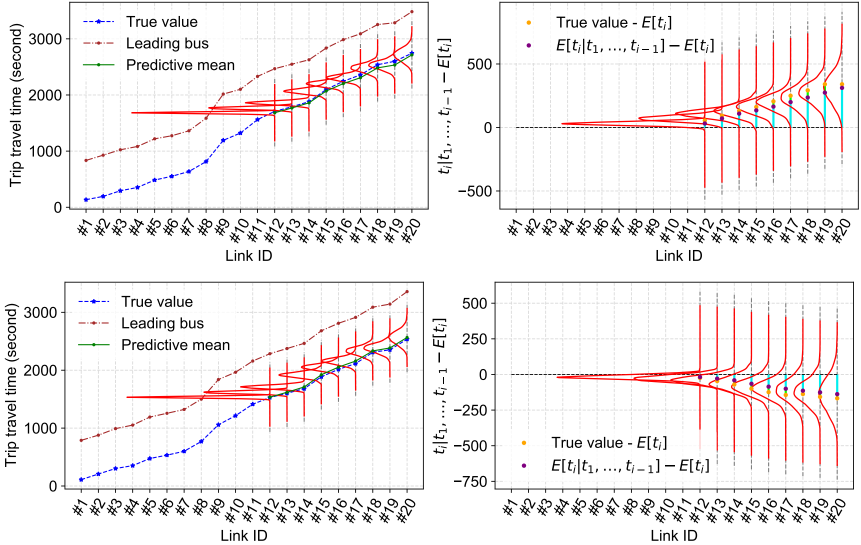


(a) Probabilistic forecasting for 16:50



(b) Probabilistic forecasting for 17:10

Predicted Distribution



Conclusion

- Our approach can capture/handle:
  - link travel time correlations of a bus route
  - interactions between adjacent buses
  - multimodality of bus travel time distribution
  - missing/ragged values in data
- We develop a Bayesian hierarchical framework to capture travel time patterns in different periods of a day.
- The proposed model is evaluated on a real-world dataset and results show it performs well.

Questions?

Thank You!

For more information

xiaoxu.chen[at]mail.mcgill.ca

https://arxiv.org/pdf/2206.06915.pdf