Conditional forecasting of bus travel time and passenger occupancy with Bayesian Markov regime-switching vector autoregression

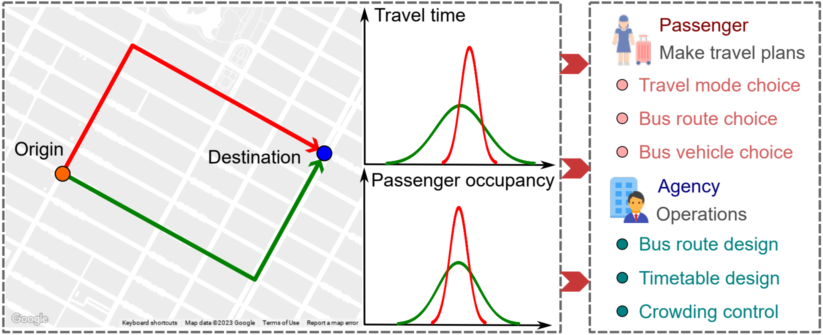
Xiaoxu Chen[1] Zhanhong Cheng[1] Alexandra Schmidt[2] Lijun Sun[1]

[1]Department of Civil Engineering, McGill University
[2]Department of Epidemiology, Biostatistics and Occupational Health, McGill University

September 6, 2024

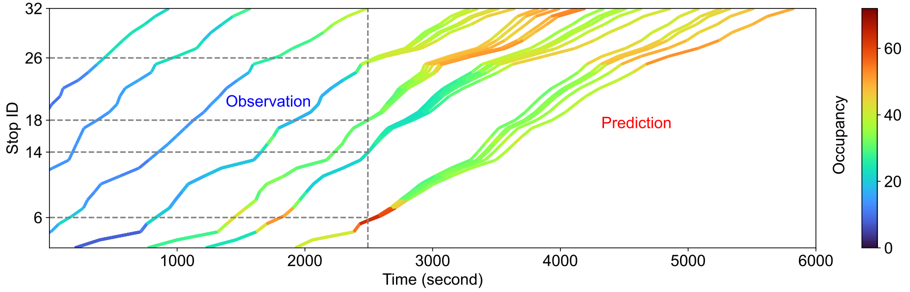
Motivation

- Bus travel time/passenger occupancy forecasting:
  - Deterministic models vs. Probabilistic models
- Why probabilistic forecasting models? reliability/uncertainty



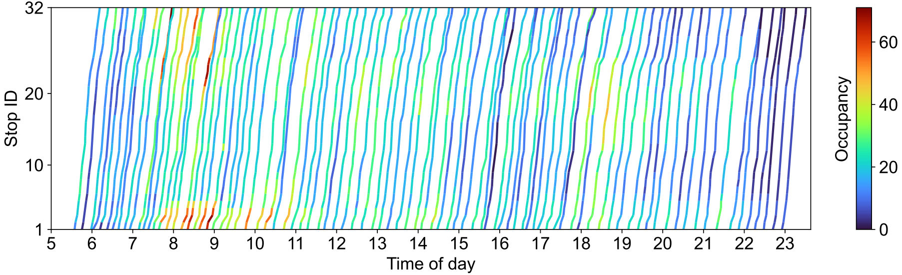
Research question

- Make probabilistic forecasting of travel time and passenger occupancy of buses on their downstream links at any time.



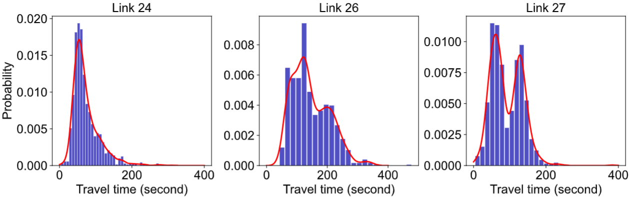
Empirical data analysis

- Bus time-space trajectories with passenger occupancy.
  - Temporal patterns, e.g., peak hours and off-peak hours.
  - Interactions between adjacent buses, e.g., bus bunching.

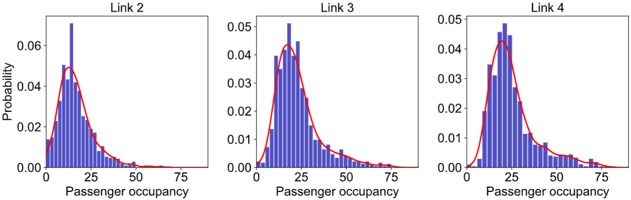


Empirical data analysis

- Distributions of link travel time and passenger occupancy.



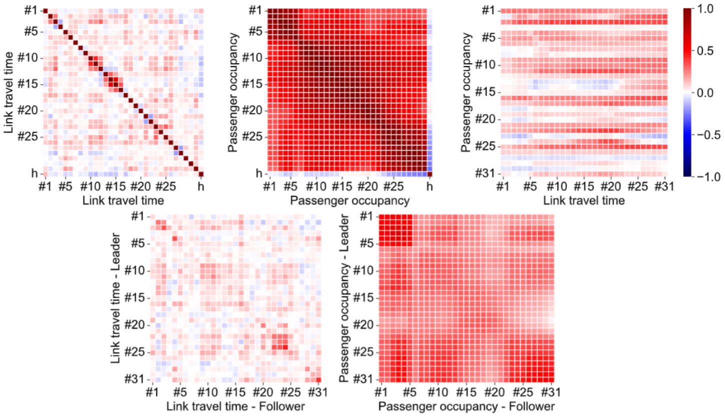
(a) Travel time distributions of some links.



(b) Passenger occupancy distributions of some links.

Empirical data analysis

- Correlation and cross-correlation matrices of associated variables.



Research gap

1. Most previous studies focus on deterministic forecasting.
2. Complex distributions of travel time and passenger occupancy.
3. Strong interactions between travel time and passenger occupancy.
4. Complex link correlations of travel time/passenger occupancy.
5. Interactions/correlations between adjacent buses along a bus route.

Research gap

1. Most previous studies focus on deterministic forecasting.
2. Complex distributions of travel time and passenger occupancy.
3. Strong interactions between travel time and passenger occupancy.
4. Complex link correlations of travel time/passenger occupancy.
5. Interactions/correlations between adjacent buses along a bus route.

Research gap

1. Most previous studies focus on deterministic forecasting.
2. Complex distributions of travel time and passenger occupancy.
3. Strong interactions between travel time and passenger occupancy.
4. Complex link correlations of travel time/passenger occupancy.
5. Interactions/correlations between adjacent buses along a bus route.

Research gap

1. Most previous studies focus on deterministic forecasting.
2. Complex distributions of travel time and passenger occupancy.
3. Strong interactions between travel time and passenger occupancy.
4. Complex link correlations of travel time/passenger occupancy.
5. Interactions/correlations between adjacent buses along a bus route.

Research gap

1. Most previous studies focus on deterministic forecasting.
2. Complex distributions of travel time and passenger occupancy.
3. Strong interactions between travel time and passenger occupancy.
4. Complex link correlations of travel time/passenger occupancy.
5. Interactions/correlations between adjacent buses along a bus route.

Random variable

- $\ell_{i,m}^{(d)}$: travel time of the i-th bus on the m-th link on the d-th day.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i^{(d)} = \left[\ell_{i,1}^{(d)}, \ell_{i,2}^{(d)}, \cdots, \ell_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $f_{i,m}^{(d)}$: occupancy of the i-th bus on the m-th link on the d-th day.
- Link occupancy vector of bus i: $\boldsymbol{f}_i^{(d)} = \left[f_{i,1}^{(d)}, f_{i,2}^{(d)}, \cdots, f_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $h_i^{(d)}$: headway of i-th bus pair at origin stop on d-th day.
- Define a random variable, $\boldsymbol{y}_i^{(d)}$ as

$$\boldsymbol{y}_i^{(d)} = \left[\boldsymbol{\ell}_i^{(d)\top}, \boldsymbol{f}_i^{(d)\top}, h_i^{(d)}\right]^\top \in \mathbb{R}^{2n+1}.$$

Random variable

- $\ell_{i,m}^{(d)}$: travel time of the i-th bus on the m-th link on the d-th day.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i^{(d)} = \left[\ell_{i,1}^{(d)}, \ell_{i,2}^{(d)}, \cdots, \ell_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $f_{i,m}^{(d)}$: occupancy of the i-th bus on the m-th link on the d-th day.
- Link occupancy vector of bus i: $\boldsymbol{f}_i^{(d)} = \left[f_{i,1}^{(d)}, f_{i,2}^{(d)}, \cdots, f_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $h_i^{(d)}$: headway of i-th bus pair at origin stop on d-th day.
- Define a random variable, $\boldsymbol{y}_i^{(d)}$ as

$$\boldsymbol{y}_i^{(d)} = \left[{\boldsymbol{\ell}_i^{(d)}}^\top, {\boldsymbol{f}_i^{(d)}}^\top, h_i^{(d)}\right]^\top \in \mathbb{R}^{2n+1}.$$

Random variable

- $\ell_{i,m}^{(d)}$: travel time of the i-th bus on the m-th link on the d-th day.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i^{(d)} = \left[\ell_{i,1}^{(d)}, \ell_{i,2}^{(d)}, \cdots, \ell_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $f_{i,m}^{(d)}$: occupancy of the i-th bus on the m-th link on the d-th day.
- Link occupancy vector of bus i: $\boldsymbol{f}_i^{(d)} = \left[f_{i,1}^{(d)}, f_{i,2}^{(d)}, \cdots, f_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $h_i^{(d)}$: headway of i-th bus pair at origin stop on d-th day.
- Define a random variable, $\boldsymbol{y}_i^{(d)}$ as

$$\boldsymbol{y}_i^{(d)} = \left[\boldsymbol{\ell}_i^{(d)\top}, \boldsymbol{f}_i^{(d)\top}, h_i^{(d)}\right]^\top \in \mathbb{R}^{2n+1}.$$

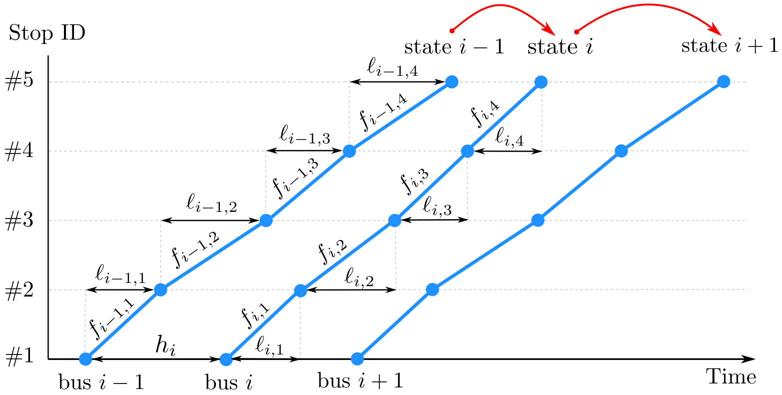
Random variable

- $\ell_{i,m}^{(d)}$: travel time of the i-th bus on the m-th link on the d-th day.
- Link travel time vector of bus i: $\boldsymbol{\ell}_i^{(d)} = \left[\ell_{i,1}^{(d)}, \ell_{i,2}^{(d)}, \cdots, \ell_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $f_{i,m}^{(d)}$: occupancy of the i-th bus on the m-th link on the d-th day.
- Link occupancy vector of bus i: $\boldsymbol{f}_i^{(d)} = \left[f_{i,1}^{(d)}, f_{i,2}^{(d)}, \cdots, f_{i,n}^{(d)}\right]^\top \in \mathbb{R}^n$.
- $h_i^{(d)}$: headway of i-th bus pair at origin stop on d-th day.
- Define a random variable, $\boldsymbol{y}_i^{(d)}$ as

$$\boldsymbol{y}_i^{(d)} = \left[\boldsymbol{\ell}_i^{(d)\top}, \boldsymbol{f}_i^{(d)\top}, h_i^{(d)}\right]^\top \in \mathbb{R}^{2n+1}.$$

Problem analysis

- Gaussian mixture model vs. Hidden Markov model



Markov regime-switching vector autoregressive

- **Data generation process**:

$$\boldsymbol{\pi}_k \mid \boldsymbol{\alpha} \sim \text{Dirichlet}(\boldsymbol{\alpha})$$

$$\boldsymbol{\Sigma}_k \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}_0, \nu_0)$$

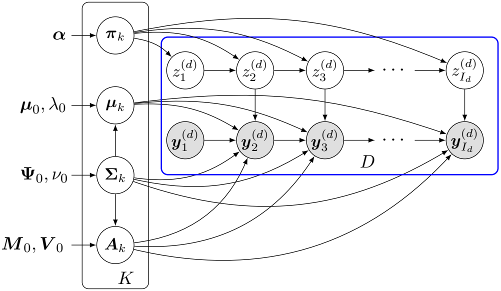
$$\boldsymbol{\mu}_k \sim \mathcal{N}\left(\boldsymbol{\mu}_0, \frac{1}{\lambda_0}\boldsymbol{\Sigma}_k\right)$$

$$\boldsymbol{A}_k \sim \mathcal{MN}(\boldsymbol{M}_0, \boldsymbol{\Sigma}_k, \boldsymbol{V}_0)$$

for $i = 2, \ldots, I_d$

$$z_i^{(d)} \sim \text{Categorical}\left(\boldsymbol{\pi}_{z_{i-1}^{(d)}}\right)$$

$$\boldsymbol{y}_i^{(d)} \mid \boldsymbol{y}_{i-1}^{(d)}, z_i^{(d)} = k \sim \mathcal{N}\left(\boldsymbol{A}_k \boldsymbol{y}_{i-1}^{(d)} + \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$



Model inference: MCMC sampling

- Sample state transition probability $\boldsymbol{\pi}_k$ from $p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right) = \text{Dirichlet}\left(|M_{k,1}| + \alpha_1, \cdots, |M_{k,K}| + \alpha_K\right).$$

- Sample state sequence $\boldsymbol{z}_{1:I_d}^{(d)}$ from $p\left(\boldsymbol{z}_{1:I_d}^{(d)} \mid \boldsymbol{y}_{1:I_d}^{(d)}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{A}\right)$.

- Sample mean and covariance $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{Y}_k, \Theta, \boldsymbol{A}_k\right)$.

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

- Sample coefficient matrix $\boldsymbol{A}_k$ from $p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$.

$$p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \mathcal{MN}\left(\boldsymbol{A}_k \mid \boldsymbol{M}_0^*, \boldsymbol{\Sigma}_k, \boldsymbol{V}_0^*\right).$$

Model inference: MCMC sampling

- Sample state transition probability $\boldsymbol{\pi}_k$ from $p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right) = \text{Dirichlet}\left(|M_{k,1}| + \alpha_1, \cdots, |M_{k,K}| + \alpha_K\right).$$

- Sample state sequence $\boldsymbol{z}_{1:I_d}^{(d)}$ from $p\left(\boldsymbol{z}_{1:I_d}^{(d)} \mid \boldsymbol{y}_{1:I_d}^{(d)}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{A}\right)$.

- Sample mean and covariance $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{Y}_k, \Theta, \boldsymbol{A}_k\right)$.

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

- Sample coefficient matrix $\boldsymbol{A}_k$ from $p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$.

$$p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \mathcal{MN}\left(\boldsymbol{A}_k \mid \boldsymbol{M}_0^*, \boldsymbol{\Sigma}_k, \boldsymbol{V}_0^*\right).$$

Model inference: MCMC sampling

- Sample state transition probability $\boldsymbol{\pi}_k$ from $p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right) = \text{Dirichlet}\left(|M_{k,1}| + \alpha_1, \cdots, |M_{k,K}| + \alpha_K\right).$$

- Sample state sequence $\boldsymbol{z}_{1:I_d}^{(d)}$ from $p\left(\boldsymbol{z}_{1:I_d}^{(d)} \mid \boldsymbol{y}_{1:I_d}^{(d)}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{A}\right)$.

- Sample mean and covariance $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{Y}_k, \Theta, \boldsymbol{A}_k\right)$.

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

- Sample coefficient matrix $\boldsymbol{A}_k$ from $p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$.

$$p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \mathcal{MN}\left(\boldsymbol{A}_k \mid \boldsymbol{M}_0^*, \boldsymbol{\Sigma}_k, \boldsymbol{V}_0^*\right).$$

Model inference: MCMC sampling

- Sample state transition probability $\boldsymbol{\pi}_k$ from $p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right)$.

$$p\left(\boldsymbol{\pi}_k \mid \boldsymbol{z}^k, \boldsymbol{\alpha}\right) = \text{Dirichlet}\left(|M_{k,1}| + \alpha_1, \cdots, |M_{k,K}| + \alpha_K\right).$$

- Sample state sequence $\boldsymbol{z}_{1:I_d}^{(d)}$ from $p\left(\boldsymbol{z}_{1:I_d}^{(d)} \mid \boldsymbol{y}_{1:I_d}^{(d)}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{A}\right)$.

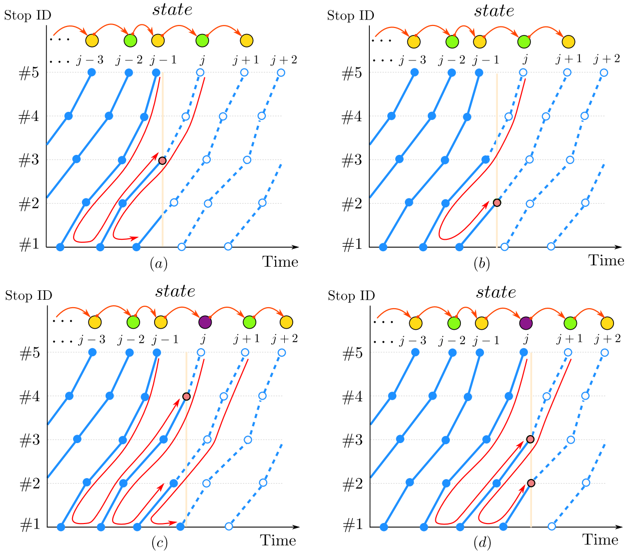
- Sample mean and covariance $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ from $p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \mid \mathcal{Y}_k, \Theta, \boldsymbol{A}_k\right)$.

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

- Sample coefficient matrix $\boldsymbol{A}_k$ from $p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$.

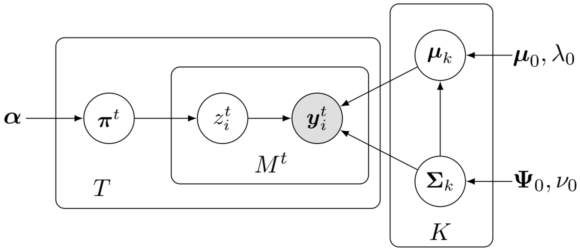
$$p\left(\boldsymbol{A}_k \mid \mathcal{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \mathcal{MN}\left(\boldsymbol{A}_k \mid \boldsymbol{M}_0^*, \boldsymbol{\Sigma}_k, \boldsymbol{V}_0^*\right).$$

Probabilistic forecasting



Experiment settings

- Performance metrics: RMSE, MAE, CRPS
- Models in comparison:
  - Bayesian Gaussian mixture model



  - BGMM-S, BGMM-J, MSAR-S, MSAR-J

Forecasting performance

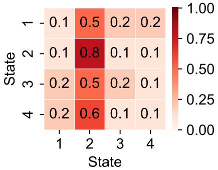
Table 1: Performance of probabilistic forecasting of link travel time, passenger occupancy, and trip travel time.

		Link travel time (sec)			Passenger occupancy (pax)			Trip travel time (sec)		
		RMSE	MAE	CRPS	RMSE	MAE	CRPS	RMSE	MAE	CRPS
BGMM-S	$K=1$	64.31	51.32	32.44	10.05	8.13	6.20	258.83	215.61	175.18
	$K=5$	54.51	42.67	29.60	9.16	7.34	5.87	246.89	201.36	157.85
	$K=10$	47.41	36.82	27.44	8.00	6.89	5.21	236.14	189.39	144.83
	$K=20$	45.73	35.54	25.91	7.81	6.70	5.43	218.65	176.94	132.44
	$K=30$	**40.53**	**31.49**	**20.59**	7.67	6.60	5.37	**199.49**	**161.29**	**115.78**
	$K=40$	43.96	33.97	22.28	**6.94**	**5.95**	**4.95**	212.09	172.35	125.93
BGMM-J	$K=1$	47.23	35.71	27.25	8.97	7.01	5.45	221.09	184.14	135.29
	$K=5$	44.63	34.35	25.69	8.64	6.92	5.28	213.53	174.15	131.21
	$K=10$	38.94	30.28	19.96	6.46	5.27	4.26	183.32	149.51	106.41
	$K=20$	36.87	28.32	19.65	6.16	5.04	3.96	180.88	143.22	103.60
	$K=30$	24.16	17.28	17.63	5.76	4.69	3.84	170.43	112.47	79.89
	$K=40$	**18.25**	**13.02**	**14.35**	**4.53**	**3.60**	**3.57**	**164.22**	**102.23**	**72.36**
	$K=50$	18.41	13.06	14.64	4.60	3.74	3.75	166.80	105.52	73.99
MSAR-S	$K=1$	62.87	49.48	31.15	10.23	8.53	6.32	257.11	214.04	174.87
	$K=5$	51.86	40.46	25.32	7.65	6.46	5.25	235.50	194.67	143.99
	$K=10$	42.61	32.73	21.74	6.89	5.87	4.73	205.99	169.55	120.36
	$K=20$	**34.45**	**25.96**	**19.59**	**6.35**	**5.39**	**4.27**	**195.90**	**156.32**	**109.32**
	$K=30$	39.09	30.28	20.01	6.58	5.61	4.53	205.60	157.92	112.60
MSAR-J	$K=1$	48.32	35.99	27.45	9.20	7.03	5.48	222.28	184.37	135.60
	$K=5$	39.38	30.36	20.39	6.43	5.45	4.82	197.60	157.43	114.00
	$K=10$	30.16	22.57	18.25	5.27	4.41	3.90	190.47	141.01	105.90
	$K=20$	18.35	13.38	14.36	4.86	4.07	3.79	164.47	103.55	72.60
	$K=30$	**16.11**	**11.66**	**12.14**	**3.48**	**2.92**	**3.07**	**137.13**	**83.48**	**57.98**
	$K=40$	17.02	12.57	13.37	4.50	3.98	3.71	153.35	95.72	63.06

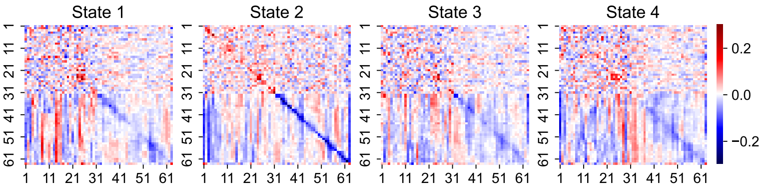
Best results are highlighted in bold fonts.

Interpreting analysis

- Estimated **transition matrix.**

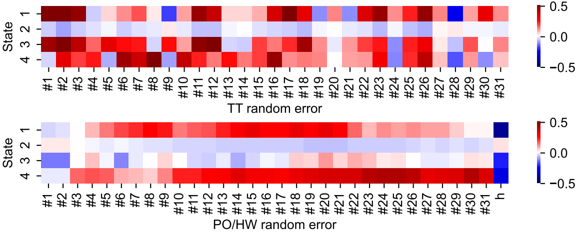


- Estimated **coefficient matrix.**

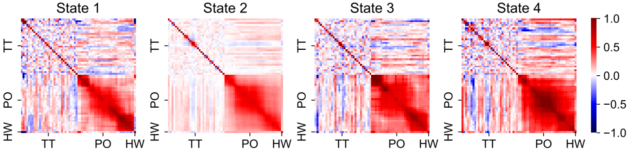


Interpreting analysis

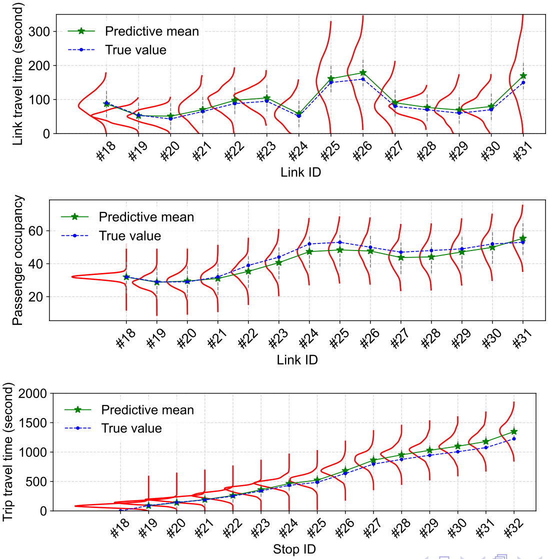
- Estimated **mean vectors** of the random error term.



- Estimated **covariance matrices** of the random error term.



Predicted distribution



Conclusion

- We propose a Bayesian Markov regime-switching vector autoregressive model for probabilistic forecasting of bus travel time and passenger occupancy.
- Our approach can capture/address:
  - correlations between travel time and passenger occupancy
  - relationship between adjacent buses
  - multimodality/skewness of travel time/ occupancy distributions
- The proposed model is evaluated on a real-world dataset and results show it performs well.

Q&A

Thank You!

Contact Email:
xiaoxu.chen@mail.mcgill.ca