

# A Bayesian Markov Mesh Regime-Switching Regression Model for Bus Travel Time Forecasting

**Xiaoxu Chen**

Department of Logistics and Operations Management,  
HEC Montréal

26th ISTTT, July 21, 2026

# Collaborators



**Xiaoxu Chen**

Logistics and Operations Management,  
*HEC Montréal*



**Martin Trépanier**

Mathematics and Industrial Engineering,  
*Polytechnique Montréal*



**Lijun Sun**

Department of Civil Engineering,  
*McGill University*

Chen, X., Trépanier, M., & Sun, L. (2026). A Bayesian Markov mesh regime-switching regression model for bus travel time forecasting. *Transportation Research Part E: Logistics and Transportation Review*, 211, 104872.

# When you take a bus, you want to know

Origin bus stop

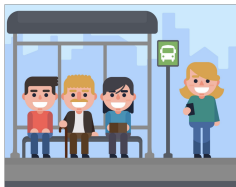


Travel time

Destination bus stop



Arrival time



- ? How long will the trip take?
- ? When will the bus arrive?
- ? **How reliable is this estimate?**

# When you take a bus, you want to know

Origin bus stop

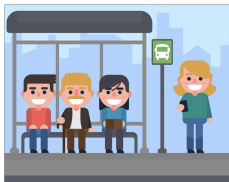


Travel time

Destination bus stop



Arrival time



- ? How long will the trip take?
- ? When will the bus arrive?
- ? How reliable is the estimate?

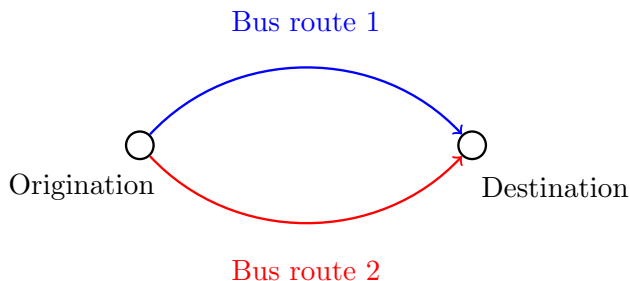
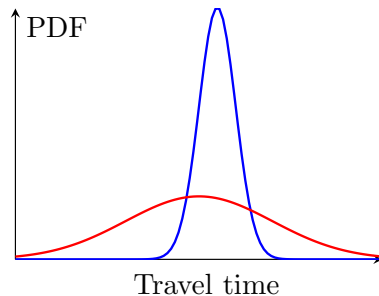
**Travel time forecasting**



**Uncertainty quantification**

# Probabilistic bus travel time forecasting

- Provide predictive **probability distributions** of trip travel time or arrival time.



Route 1 has a longer expected travel time but is far more reliable than Route 2.

# Why are travel time distributions important?



## Passengers

Make better travel plans



Departure time



Bus route choice



Travel mode choice



## Bus agencies

Design robust bus management strategies



Bus scheduling

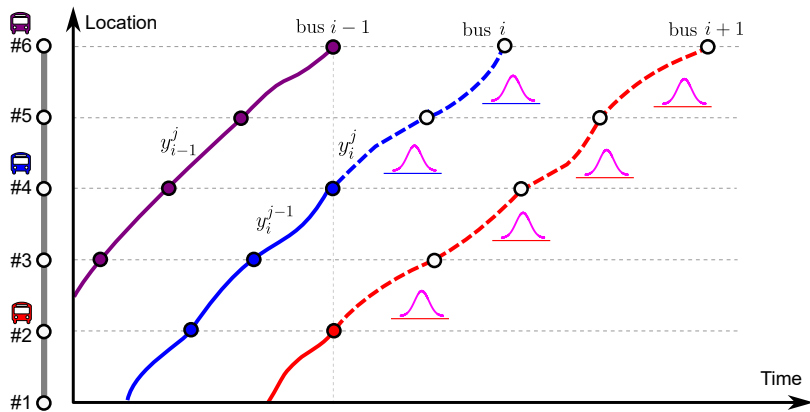


Transit signal priority



Bunching control

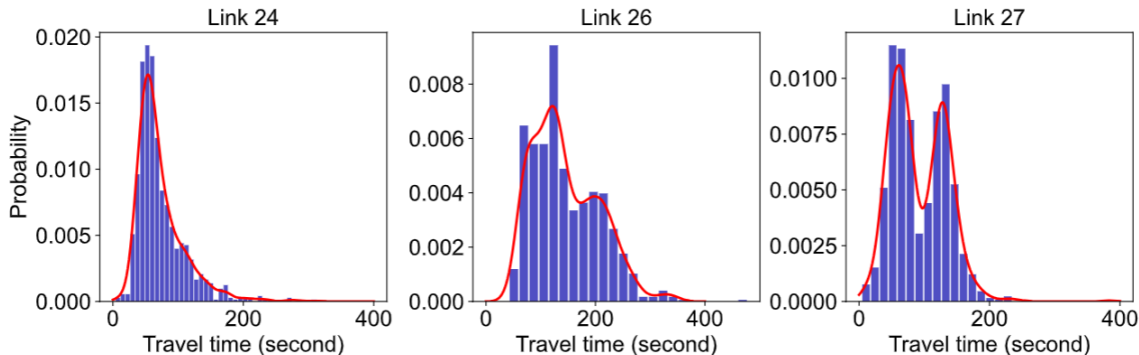
# Research problem



- $y_i^j$  denotes the link travel time of bus  $i$  on link  $j$ .
- Forecast the travel/arrival time **distributions** on downstream links/stops.

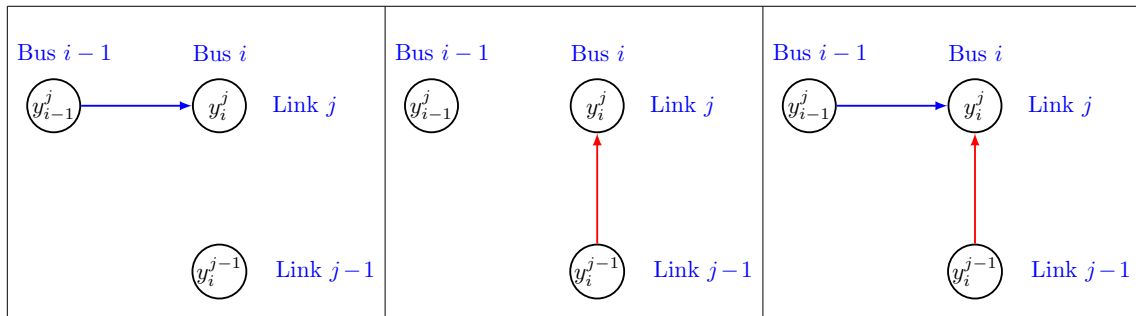
# Multimodal/skewed distributions of bus travel time

- Modeling should consider the complex distributions of bus travel time.



# Dependencies in bus link travel time

- (a) **Bus-to-bus**: driven by the preceding bus on the same link (e.g., bunching).
- (b) **Link-to-link**: driven by the same bus's previous link.
- (c) **Bus-link-to-bus-link**: both dependencies jointly.



(a) bus-to-bus

(b) link-to-link

(c) bus-link-to-bus-link

# Research gap

- Some probabilistic models capture only **one** dependency dimension:
  - link-to-link HMMs (e.g., Ma et al., 2017; Büchel & Corman, 2022)
- Bus-link Bayesian network with Gaussian likelihood (Büchel & Corman, 2022):
  - Unable to model the multimodal/skewed distributions
- Few models jointly consider 2-D dependency and multimodal/skewed distributions.

## References

Büchel, B., & Corman, F. (2022). Modeling conditional dependencies for bus travel time estimation. *Physica A*, 592, 126764.

Büchel, B., & Corman, F. (2022). What do we know when? modeling predictability of transit operations. *IEEE Trans. Intell. Transp. Syst.*, 23(9), 15684-15695.

Ma, Z., et al. (2017). Estimation of trip travel time distribution using a generalized Markov chain approach. *Transp. Res. Part C Emerging Technol.*, 74, 1-21.

# Markov mesh regime-switching regression

- Let  $y_i^j$  be the travel time of bus  $i$  on link  $j$ , and  $z_i^j$  its latent state.
- Conditional on state  $z_i^j = k$ :

$$y_i^j \mid y_{i-1}^j, y_i^{j-1}, z_i^j = k \sim \mathcal{N}(\theta_k y_{i-1}^j + \gamma_k y_i^{j-1} + \mu_k, \sigma_k^2)$$

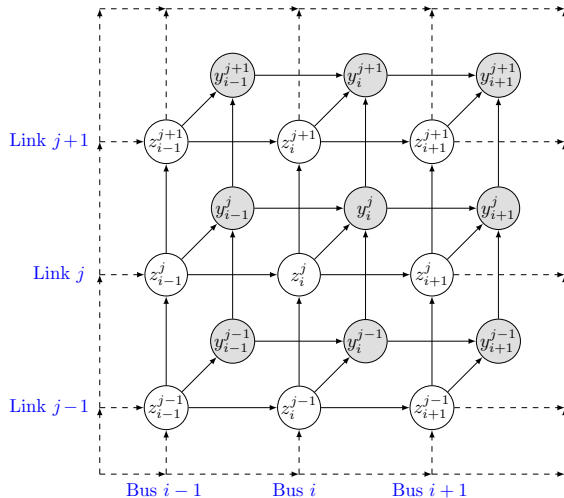
- $\theta_k$ : **bus-to-bus** coefficient (preceding bus, same link).
- $\gamma_k$ : **link-to-link** coefficient (same bus, previous link).
- $\mu_k, \sigma_k^2$ : state-specific baseline and residual variance.
- Mixing over states  $k$  yields a **multimodal, skewed** marginal distribution.

# Markov mesh regime-switching regression

- State  $z_i^j$  depends on **two** parents: the preceding bus  $z_{i-1}^j$  and the previous link  $z_i^{j-1}$ :

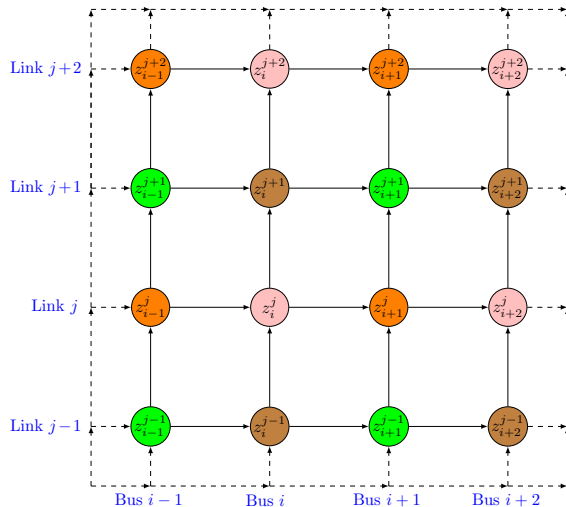
$$z_i^j \mid z_{i-1}^j, z_i^{j-1} \sim \text{Cat} \left( \pi_{z_{i-1}^j, z_i^{j-1}} \right)$$

- Transition tensor  $\pi \in \mathbb{R}^{K \times K \times K}$ .
- Conjugate priors: Dirichlet on  $\pi_{k,k'}$ ; inverse-Gamma on  $\sigma_k^2$ ; Gaussian on  $\mu_k, \theta_k, \gamma_k$ .



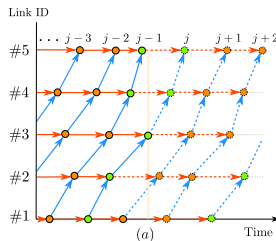
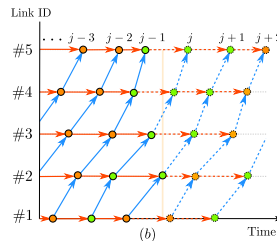
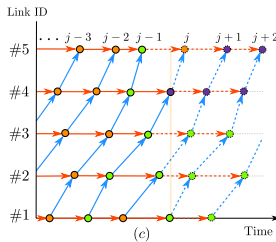
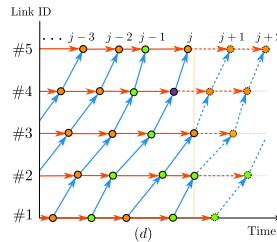
# Model inference: block Gibbs sampling

- With conjugate priors, we derive closed-form Gibbs updates for  $\pi$ ,  $\mu_k$ ,  $\sigma_k^2$ ,  $\theta_k$ ,  $\gamma_k$ .
- Each  $z_i^j$  depends on 2 parents and 3 children in the mesh; sequential updates are slow.
- Block Gibbs:** partition the mesh into 4 conditionally-independent color blocks and update each block jointly.



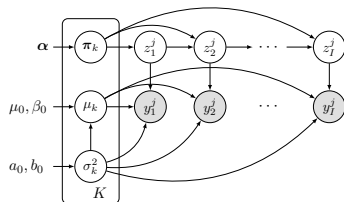
# Probabilistic forecasting procedure

- Use posterior parameter samples from training.
- Infer hidden states of observed links, then forecast upcoming links via the observation model.
- Forecasts update in real time as each new link is observed.

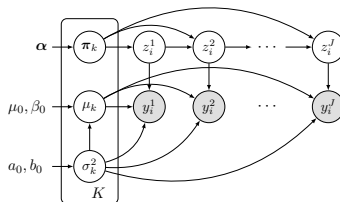


# Experiment settings

- Performance metrics: RMSE, MAE, CRPS
- Baseline models:

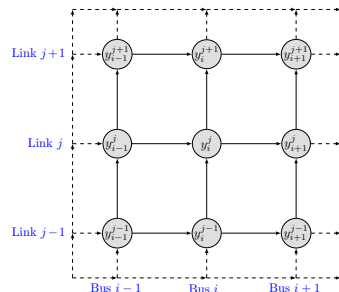


HMM (bus-to-bus)



HMM (link-to-link)

Ma et al. (2017); Büchel & Corman (2022)



Bayesian network  
(bus-link-to-bus-link, single state)

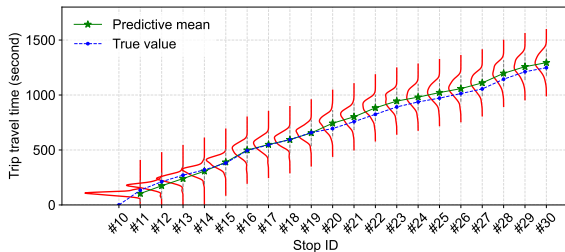
Büchel & Corman (2022)

# Forecasting results

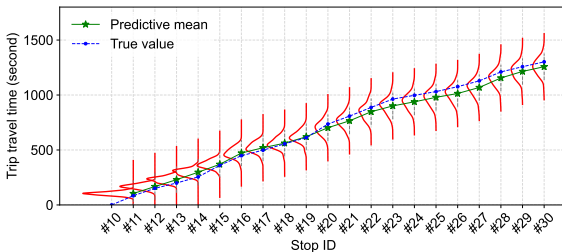
- Modeling 2-D dependency improves the accuracy.
- With more states, the forecasting is more accurate.
- The proposed Markov mesh regime-switching regression model outperforms all baselines.

|                                       |         | Link travel time (sec) |       |      | Trip travel time (sec) |       |       |
|---------------------------------------|---------|------------------------|-------|------|------------------------|-------|-------|
|                                       |         | RMSE                   | MAE   | CRPS | RMSE                   | MAE   | CRPS  |
| Bayesian network                      |         | 15.01                  | 10.41 | 7.58 | 72.15                  | 59.10 | 42.06 |
| HMM (link-to-link)                    | $K = 1$ | 15.26                  | 10.57 | 7.71 | 111.45                 | 91.83 | 67.73 |
|                                       | $K = 2$ | 14.87                  | 11.00 | 8.29 | 72.90                  | 58.45 | 42.52 |
|                                       | $K = 4$ | 16.34                  | 10.97 | 8.35 | 88.04                  | 69.95 | 51.58 |
|                                       | $K = 6$ | 16.24                  | 10.17 | 7.31 | 71.17                  | 53.24 | 40.34 |
| HMM (bus-to-bus)                      | $K = 1$ | 15.23                  | 10.54 | 7.72 | 112.47                 | 92.54 | 68.12 |
|                                       | $K = 2$ | 15.29                  | 11.32 | 8.56 | 73.56                  | 61.46 | 42.85 |
|                                       | $K = 4$ | 15.55                  | 10.26 | 7.73 | 86.04                  | 66.87 | 49.30 |
|                                       | $K = 6$ | 15.15                  | 10.40 | 7.40 | 67.85                  | 53.63 | 39.00 |
| Markov Mesh<br>regime-switching model | $K = 1$ | 14.94                  | 10.39 | 7.55 | 71.85                  | 58.33 | 41.59 |
|                                       | $K = 2$ | 14.68                  | 10.95 | 7.86 | 65.35                  | 52.46 | 37.70 |
|                                       | $K = 4$ | 14.63                  | 10.30 | 7.55 | 70.73                  | 54.22 | 39.59 |
|                                       | $K = 6$ | 14.12                  | 9.99  | 7.23 | 61.20                  | 45.47 | 34.21 |

# Probabilistic forecast examples



(a) Observation 1



(b) Observation 2

- Predictive means (green) closely follow actual travel times (blue).
- Forecast uncertainty grows with the number of downstream links.

# Summary

- Bayesian Markov mesh regime-switching regression model for probabilistic bus travel time forecasting.
- Two-dimensional mesh jointly captures **bus-to-bus** and **link-to-link** dependence.
- Regime-switching regression captures multimodal, skewed travel time distributions.
- Efficient block-Gibbs MCMC for both inference and forecasting.
- Outperforms HMM and Bayesian-network baselines.

# Thank you!

Questions and discussion welcome!

Xiaoxu Chen    `xiaoxu.chen@hec.ca`