informs
https://pubsonline.informs.org/journal/trsc

TRANSPORTATION SCIENCE

Submitted to *Transportation Science*

Inductive Transportation Origin-Destination Demand Forecasting via Bayesian Destination Choice Modeling

Xiaoxu Chen
Department of Logistics and Operations Management, HEC Montréal, Montreal, QC H3T 2A7, Canada;
xiaoxu.chen@hec.ca

Yafeng Yin
Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan 48109, United States;
yafeng@umich.edu

Lijun Sun
Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada;
lijun.sun@mcgill.ca

Appendix A: Proof of Proposition 1 (Origin-Factor Gradient)

We derive the likelihood gradient directly in matrix form. For time t, write

$$\boldsymbol{G}_t = \sum_{s=1}^{R} w_{ts}\boldsymbol{u}_s\boldsymbol{v}_s^\top, \qquad \boldsymbol{H}_t = C(\boldsymbol{G}_t),$$

where $C(\cdot)$ denotes row-wise centering over feasible destinations. Let $\boldsymbol{\Lambda}_t$ be the matrix of destination-choice probabilities and define

$$\boldsymbol{Y}_t^+ = \sum_{m=1}^{M} \boldsymbol{Y}_t^m, \qquad \boldsymbol{n}_t^+ = \sum_{m=1}^{M} (N_{1t}^m, \ldots, N_{St}^m)^\top,$$
$$\boldsymbol{R}_t = \boldsymbol{Y}_t^+ - \operatorname{diag}(\boldsymbol{n}_t^+)\boldsymbol{\Lambda}_t.$$

The row-wise multinomial softmax likelihood satisfies

$$d\mathcal{L}_{\text{like},t} = \langle \boldsymbol{R}_t, d\boldsymbol{H}_t \rangle .$$

Because each row of $\boldsymbol{R}_t$ sums to zero, $\boldsymbol{R}_t$ is orthogonal to the row-wise constants introduced by the centering operator. Hence

$$\langle \boldsymbol{R}_t, d\boldsymbol{H}_t \rangle = \langle \boldsymbol{R}_t, d\boldsymbol{G}_t \rangle .$$

For an origin-factor perturbation $d\boldsymbol{u}_r$,

$$d\boldsymbol{G}_t = w_{tr}\, d\boldsymbol{u}_r \boldsymbol{v}_r^\top .$$

Therefore,

$$d\mathcal{L}_{\text{like}} = \sum_{t=1}^{T} w_{tr} \left\langle \boldsymbol{R}_t, d\boldsymbol{u}_r \boldsymbol{v}_r^\top \right\rangle = \left(\sum_{t=1}^{T} w_{tr} \boldsymbol{R}_t \boldsymbol{v}_r \right)^\top d\boldsymbol{u}_r .$$

Thus,

$$\nabla_{\boldsymbol{u}_r} \mathcal{L}_{\text{like}} = \sum_{t=1}^{T} w_{tr} \boldsymbol{R}_t \boldsymbol{v}_r .$$

Adding the Gaussian-prior differential

$$d\left(-\frac{1}{2}\boldsymbol{u}_r^\top \boldsymbol{K}_u^{-1}\boldsymbol{u}_r\right) = -\left(\boldsymbol{K}_u^{-1}\boldsymbol{u}_r\right)^\top d\boldsymbol{u}_r$$

gives the full posterior gradient.

Appendix B: Proof of Proposition 2 (Destination-Factor Gradient)

For a destination-factor perturbation $d\boldsymbol{v}_r$,

$$d\boldsymbol{G}_t = w_{tr}\boldsymbol{u}_r d\boldsymbol{v}_r^\top.$$

Using the same residual identity,

$$d\mathcal{L}_{\text{like}} = \sum_{t=1}^{T} w_{tr}\left\langle \boldsymbol{R}_t, \boldsymbol{u}_r d\boldsymbol{v}_r^\top\right\rangle = \left(\sum_{t=1}^{T} w_{tr}\left(\boldsymbol{R}_t\right)^\top \boldsymbol{u}_r\right)^\top d\boldsymbol{v}_r.$$

Therefore,

$$\nabla_{\boldsymbol{v}_r}\mathcal{L}_{\text{like}} = \sum_{t=1}^{T} w_{tr}\left(\boldsymbol{R}_t\right)^\top \boldsymbol{u}_r.$$

Adding

$$d\left(-\frac{1}{2}\boldsymbol{v}_r^\top \boldsymbol{K}_v^{-1}\boldsymbol{v}_r\right) = -\left(\boldsymbol{K}_v^{-1}\boldsymbol{v}_r\right)^\top d\boldsymbol{v}_r$$

gives the full posterior gradient.

Appendix C: Matrix Proof of Proposition 3 (Temporal-Factor Gradient)

For an arbitrary temporal-factor perturbation $d\boldsymbol{w}_r = (dw_{1r}, \ldots, dw_{Tr})^\top$,

$$d\boldsymbol{G}_t = \boldsymbol{u}_r \boldsymbol{v}_r^\top dw_{tr}.$$

Then

$$d\mathcal{L}_{\text{like}} = \sum_{t=1}^{T}\left\langle \boldsymbol{R}_t, \boldsymbol{u}_r\boldsymbol{v}_r^\top\right\rangle dw_{tr} = \sum_{t=1}^{T}\boldsymbol{u}_r^\top \boldsymbol{R}_t \boldsymbol{v}_r\, dw_{tr}.$$

Hence

$$\left[\nabla_{\boldsymbol{w}_r}\mathcal{L}_{\text{like}}\right]_t = \boldsymbol{u}_r^\top \boldsymbol{R}_t \boldsymbol{v}_r.$$

The prior differential

$$d\left(-\frac{1}{2}\boldsymbol{w}_r^\top \boldsymbol{K}_w^{-1}\boldsymbol{w}_r\right) = -\left(\boldsymbol{K}_w^{-1}\boldsymbol{w}_r\right)^\top d\boldsymbol{w}_r$$

completes the matrix derivation of the posterior gradient.

Appendix D: Proof of Proposition 4 (Log-Transform Jacobian)

We use the standard formula for a change of variables in a probability density. Let θ_j be a random variable with density $p\left(\theta_j\right)$ and let $\phi_j = h\left(\theta_j\right)$ be a one-to-one transformation, where $h\left(\theta_j\right) = \log\left(\theta_j\right)$. The inverse transformation is $\theta_j = h^{-1}\left(\phi_j\right) = \exp\left(\phi_j\right)$.

The probability density function for ϕ_j, which we denote $\tilde{p}\left(\phi_j\right)$, is given by:

$$\tilde{p}\left(\phi_j\right) = p\left(h^{-1}\left(\phi_j\right)\right)\left|\frac{dh^{-1}\left(\phi_j\right)}{d\phi_j}\right|. \tag{1}$$

We apply this formula to our specific case. The original density is the conditional posterior $p\left(\theta_j \mid \cdot\right)$. We can get the $\tilde{p}\left(\phi_j \mid \cdot\right)$ as:

$$\tilde{p}\left(\phi_j \mid \cdot\right) = p\left(\theta_j = \exp\left(\phi_j\right) \mid \cdot\right) \left| \frac{d}{d\phi_j} \exp\left(\phi_j\right) \right|. \tag{2}$$

The derivative of the inverse transformation is:

$$\frac{d}{d\phi_j} \exp\left(\phi_j\right) = \exp\left(\phi_j\right). \tag{3}$$

Since $\theta_j > 0$, the transformation $\phi_j = \log\left(\theta_j\right)$ is defined, and its inverse $\exp\left(\phi_j\right)$ is always positive. Therefore, the absolute value is redundant:

$$\left| \frac{d}{d\phi_j} \exp\left(\phi_j\right) \right| = \left| \exp\left(\phi_j\right) \right| = \exp\left(\phi_j\right). \tag{4}$$

Substituting this back, we get the density for ϕ_j:

$$\tilde{p}\left(\phi_j \mid \cdot\right) = p\left(\theta_j = \exp\left(\phi_j\right) \mid \cdot\right) \exp\left(\phi_j\right). \tag{5}$$

To obtain the target function for the slice sampler, we take the logarithm of this density:

$$\begin{aligned} \mathcal{L}\left(\phi_j\right) &= \log \tilde{p}\left(\phi_j \mid \cdot\right) \\ &= \log\left(p\left(\theta_j = \exp\left(\phi_j\right) \mid \cdot\right) \exp\left(\phi_j\right)\right) \\ &= \log p\left(\boldsymbol{\theta}_K\left(\phi_j\right) \mid \cdot\right) + \phi_j. \end{aligned} \tag{6}$$

Appendix E: Proof of Corollary 1

Fix an existing origin o and time interval t. Let $\mathcal{J}_o$ denote the original set of feasible destinations, and suppose that a new destination $s \notin \mathcal{J}_o$, with $s \neq o$, is added to the network. The expanded feasible destination set is therefore $\mathcal{J}_o^+ = \mathcal{J}_o \cup \{s\}$. The result is conditional on the latent utilities $\{G_{odt} : d \in \mathcal{J}_o\}$ and G_{ost}, which are held fixed when the new destination is introduced.

Before the network expansion, define the mean utility over the original feasible destinations as

$$\bar{G}_{ot} = \frac{1}{|\mathcal{J}_o|} \sum_{k \in \mathcal{J}_o} G_{okt},$$

and let the corresponding centered utility be

$$g_{odt} = G_{odt} - \bar{G}_{ot}, \qquad d \in \mathcal{J}_o.$$

The original destination-choice probability is

$$\lambda_{odt} = \frac{\exp(g_{odt})}{\sum_{k \in \mathcal{J}_o} \exp(g_{okt})}.$$

Because the same centering constant is subtracted from every feasible utility, it cancels from the softmax numerator and denominator. Hence,

$$\lambda_{odt} = \frac{\exp(G_{odt} - \bar{G}_{ot})}{\sum_{k \in \mathcal{J}_o} \exp(G_{okt} - \bar{G}_{ot})} = \frac{\exp(G_{odt})}{\sum_{k \in \mathcal{J}_o} \exp(G_{okt})}.$$

Define

$$Z_{ot} = \sum_{k \in \mathcal{J}_o} \exp(G_{okt}), \qquad a_{ost} = \exp(G_{ost}).$$

It follows that

$$\lambda_{odt} = \frac{\exp(G_{odt})}{Z_{ot}}, \qquad d \in \mathcal{J}_o.$$

After destination s is added, the mean utility over the expanded feasible set becomes

$$\bar{G}^+_{ot} = \frac{1}{|\mathcal{J}_o| + 1} \left(\sum_{k \in \mathcal{J}_o} G_{okt} + G_{ost} \right).$$

The expanded-network probabilities are obtained by applying the softmax transformation to the centered utilities $G_{odt} - \bar{G}^+_{ot}$. Again, the common centering constant cancels. Therefore, the probability assigned to the new destination is

$$\begin{aligned} \lambda^+_{ost} &= \frac{\exp(G_{ost} - \bar{G}^+_{ot})}{\sum_{k \in \mathcal{J}_o} \exp(G_{okt} - \bar{G}^+_{ot}) + \exp(G_{ost} - \bar{G}^+_{ot})} \\ &= \frac{\exp(G_{ost})}{\sum_{k \in \mathcal{J}_o} \exp(G_{okt}) + \exp(G_{ost})} \\ &= \frac{a_{ost}}{Z_{ot} + a_{ost}}. \end{aligned}$$

Similarly, for every existing destination $d \in \mathcal{J}_o$,

$$\begin{aligned} \lambda^+_{odt} &= \frac{\exp(G_{odt})}{Z_{ot} + a_{ost}} \\ &= \frac{Z_{ot}}{Z_{ot} + a_{ost}} \frac{\exp(G_{odt})}{Z_{ot}} \\ &= \frac{Z_{ot}}{Z_{ot} + a_{ost}} \lambda_{odt}. \end{aligned}$$

Since

$$1 - \lambda^+_{ost} = 1 - \frac{a_{ost}}{Z_{ot} + a_{ost}} = \frac{Z_{ot}}{Z_{ot} + a_{ost}},$$

we obtain

$$\lambda^+_{odt} = \left(1 - \lambda^+_{ost}\right) \lambda_{odt}, \qquad d \in \mathcal{J}_o.$$

Thus, all probabilities assigned to the existing destinations are reduced by the same proportional factor.

Finally, for any two existing destinations $d, k \in \mathcal{J}_o$,

$$\frac{\lambda^+_{odt}}{\lambda^+_{okt}} = \frac{\left(1 - \lambda^+_{ost}\right) \lambda_{odt}}{\left(1 - \lambda^+_{ost}\right) \lambda_{okt}} = \frac{\lambda_{odt}}{\lambda_{okt}}.$$

This proves the corollary.

Appendix F: Matrix Gradients and Fisher Preconditioner

Let $\mathbf{\Lambda}_t = \{\lambda_{odt}\}_{o,d=1}^{S}$ be the destination-choice probability matrix at time t, with zero diagonal. Define

$$\boldsymbol{Y}^+_t = \sum_{m=1}^{M} \boldsymbol{Y}^m_t, \qquad \boldsymbol{n}^+_t = \sum_{m=1}^{M} (N^m_{1t}, \ldots, N^m_{St})^\top,$$

and

$$\boldsymbol{R}_t = \boldsymbol{Y}^+_t - \operatorname{diag}(\boldsymbol{n}^+_t)\mathbf{\Lambda}_t. \tag{7}$$

Because $\boldsymbol{R}_t \mathbf{1} = \mathbf{0}$, the row-centering operation in the logits does not add an explicit correction term to the matrix gradients. The likelihood-gradient blocks used in the implementation are

$$\nabla_{\boldsymbol{u}_r} \mathcal{L}_{\text{like}} = \sum_{t=1}^{T} w_{tr} \boldsymbol{R}_t \boldsymbol{v}_r, \tag{8}$$

$$\nabla_{\boldsymbol{v}_r} \mathcal{L}_{\text{like}} = \sum_{t=1}^{T} w_{tr} \, (\boldsymbol{R}_t)^\top \boldsymbol{u}_r, \tag{9}$$

$$\left[\nabla_{\boldsymbol{w}_r} \mathcal{L}_{\text{like}}\right]_t = \boldsymbol{u}_r^\top \boldsymbol{R}_t \boldsymbol{v}_r. \tag{10}$$

The full posterior gradients subtract the GP prior precision terms, e.g.,

$$\nabla_{\boldsymbol{u}_r} \mathcal{L} = \nabla_{\boldsymbol{u}_r} \mathcal{L}_{\text{like}} - \boldsymbol{K}_u^{-1} \boldsymbol{u}_r,$$

$$\nabla_{\boldsymbol{v}_r} \mathcal{L} = \nabla_{\boldsymbol{v}_r} \mathcal{L}_{\text{like}} - \boldsymbol{K}_v^{-1} \boldsymbol{v}_r,$$

$$\nabla_{\boldsymbol{w}_r} \mathcal{L} = \nabla_{\boldsymbol{w}_r} \mathcal{L}_{\text{like}} - \boldsymbol{K}_w^{-1} \boldsymbol{w}_r.$$

For each origin-time pair, let $\boldsymbol{\lambda}_{ot} = (\lambda_{odt})_{d \in \mathcal{J}_o}$ and

$$\boldsymbol{F}_{ot} = \operatorname{diag}(\boldsymbol{\lambda}_{ot}) - \boldsymbol{\lambda}_{ot} \boldsymbol{\lambda}_{ot}^\top.$$

Let $\boldsymbol{P}_o \in \mathbb{R}^{(S-1) \times S}$ be the selection-centering matrix satisfying $\boldsymbol{P}_o \boldsymbol{v}_r = (v_{dr} - \bar{v}_{or})_{d \in \mathcal{J}_o}$. With $n_{ot}^+ = \sum_{m=1}^{M} N_{ot}^m$, the likelihood Fisher blocks are

$$\boldsymbol{\mathcal{I}}_{u_r} = \sum_{o=1}^{S} \sum_{t=1}^{T} n_{ot}^+ \, \boldsymbol{e}_o \boldsymbol{e}_o^\top \, (w_{tr} \boldsymbol{P}_o \boldsymbol{v}_r)^\top \boldsymbol{F}_{ot} \, (w_{tr} \boldsymbol{P}_o \boldsymbol{v}_r), \tag{11}$$

$$\boldsymbol{\mathcal{I}}_{v_r} = \sum_{o=1}^{S} \sum_{t=1}^{T} n_{ot}^+ (u_{or} w_{tr})^2 \boldsymbol{P}_o^\top \boldsymbol{F}_{ot} \boldsymbol{P}_o, \tag{12}$$

$$\boldsymbol{\mathcal{I}}_{w_r} = \sum_{o=1}^{S} \sum_{t=1}^{T} n_{ot}^+ \, \boldsymbol{e}_t \boldsymbol{e}_t^\top \, (u_{or} \boldsymbol{P}_o \boldsymbol{v}_r)^\top \boldsymbol{F}_{ot} \, (u_{or} \boldsymbol{P}_o \boldsymbol{v}_r). \tag{13}$$

The Fisher-preconditioned pMALA metric for a generic block $\boldsymbol{\beta}$ with prior covariance $\boldsymbol{C}_\beta$ is

$$\boldsymbol{M}_\beta(\boldsymbol{\beta}) = \left(\boldsymbol{C}_\beta^{-1} + \boldsymbol{\mathcal{I}}_\beta(\boldsymbol{\beta}) + \epsilon \boldsymbol{I} \right)^{-1}. \tag{14}$$

Appendix G: Algorithms for Model Inference and Probabilistic Forecasting

Algorithm 1 Fisher-preconditioned pMALA for a Tensor-Factor Block

Input: Current block $\boldsymbol{\beta}^{(i)} \in \{\boldsymbol{u}_r, \boldsymbol{v}_r, \boldsymbol{w}_r\}$, step size δ_β, jitter ϵ, prior covariance $\boldsymbol{C}_\beta \in \{\boldsymbol{K}_u, \boldsymbol{K}_v, \boldsymbol{K}_w\}$, current values of the remaining factors, and data $\boldsymbol{Y}$.

Output: New block state $\boldsymbol{\beta}^{(i+1)}$.

Compute $\boldsymbol{R}_t$ for $t = 1, \ldots, T$ using Eq. (7).

Compute the block-specific likelihood gradient $\boldsymbol{a}^{(i)} = \nabla_\beta \mathcal{L}_{\text{like}}$ using Eq. (8), (9), or (10).

Construct the corresponding likelihood Fisher block $\boldsymbol{\mathcal{I}}_\beta^{(i)}$ using Eq. (11), (12), or (13).

Compute full gradient $\boldsymbol{g}^{(i)} = \boldsymbol{a}^{(i)} - \boldsymbol{C}_\beta^{-1}\boldsymbol{\beta}^{(i)}$.

Set $\boldsymbol{G}^{(i)} = \boldsymbol{C}_\beta^{-1} + \boldsymbol{\mathcal{I}}_\beta^{(i)} + \epsilon \boldsymbol{I}$, $\boldsymbol{M}^{(i)} = \left(\boldsymbol{G}^{(i)}\right)^{-1}$, and compute $\boldsymbol{L}^{(i)} \left(\boldsymbol{L}^{(i)}\right)^\top = \boldsymbol{M}^{(i)}$.

Calculate proposal mean $\boldsymbol{\mu}^{(i)} = \boldsymbol{\beta}^{(i)} + \frac{\delta_\beta}{2}\boldsymbol{M}^{(i)}\boldsymbol{g}^{(i)}$.

Sample $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and propose $\boldsymbol{\beta}^* = \boldsymbol{\mu}^{(i)} + \sqrt{\delta_\beta}\boldsymbol{L}^{(i)}\boldsymbol{\xi}$.

At $\boldsymbol{\beta}^*$, recompute $\boldsymbol{a}^*$, $\boldsymbol{g}^*$, $\boldsymbol{\mathcal{I}}_\beta^*$, $\boldsymbol{G}^*$, $\boldsymbol{M}^*$, and $\boldsymbol{\mu}^* = \boldsymbol{\beta}^* + \frac{\delta_\beta}{2}\boldsymbol{M}^*\boldsymbol{g}^*$.

Calculate log-proposal densities:

- $\log q\left(\boldsymbol{\beta}^* \mid \boldsymbol{\beta}^{(i)}\right) = -\frac{1}{2}\log\left|\delta_\beta \boldsymbol{M}^{(i)}\right| - \frac{1}{2\delta_\beta}\left(\boldsymbol{\beta}^* - \boldsymbol{\mu}^{(i)}\right)^\top \boldsymbol{G}^{(i)}\left(\boldsymbol{\beta}^* - \boldsymbol{\mu}^{(i)}\right)$,
- $\log q\left(\boldsymbol{\beta}^{(i)} \mid \boldsymbol{\beta}^*\right) = -\frac{1}{2}\log\left|\delta_\beta \boldsymbol{M}^*\right| - \frac{1}{2\delta_\beta}\left(\boldsymbol{\beta}^{(i)} - \boldsymbol{\mu}^*\right)^\top \boldsymbol{G}^*\left(\boldsymbol{\beta}^{(i)} - \boldsymbol{\mu}^*\right)$.

Calculate full posterior log-ratio:

$\log \alpha = \left(\mathcal{L}\left(\boldsymbol{\beta}^*\right) - \mathcal{L}\left(\boldsymbol{\beta}^{(i)}\right)\right) + \left(\log q\left(\boldsymbol{\beta}^{(i)} \mid \boldsymbol{\beta}^*\right) - \log q\left(\boldsymbol{\beta}^* \mid \boldsymbol{\beta}^{(i)}\right)\right)$.

Draw $a \sim \text{Uniform}(0, 1)$.

if $\log(a) < \log \alpha$ **then**

  $\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^*$

else

  $\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^{(i)}$

end if

return $\boldsymbol{\beta}^{(i+1)}$.

Algorithm 2 Slice Sampling for Hyperparameter θ_j

Input: Current state $\theta_j^{(i)}$, posterior log-density $\mathcal{L}\left(\phi_j\right)$, current factors $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$.
Output: New state $\theta_j^{(i+1)}$.
Compute $\phi_j = \log\left(\theta_j^{(i)}\right)$.
Log-likelihood threshold: $\gamma \sim \text{Uniform}\,(0,1)$, $\log c = \mathcal{L}\left(\phi_j\right) + \log\gamma$.
Draw an initial sampling range: $\kappa \sim \text{Uniform}\,(0,\epsilon)$, $\phi_{\min} = \phi_j - \kappa$, $\phi_{\max} = \phi_{\min} + \epsilon$.
loop
    Propose new state: $\phi_j^* \sim \text{Uniform}\,(\phi_{\min}, \phi_{\max})$.
    if $\mathcal{L}\left(\phi_j^*\right) > \log c$ **then**
        $\theta_j^{(i+1)} = \exp\left(\phi_j^*\right)$.
        break
    else
        if $\phi_j^* < \phi_j$ **then:** $\phi_{\min} = \phi_j^*$ **else:** $\phi_{\max} = \phi_j^*$.
    end if
end loop
return $\theta_j^{(i+1)}$.

Algorithm 3 Centered and Non-centered Interweaving for GP Hyperparameters

Input: Centered factors $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$, current covariance hyperparameters $\boldsymbol{\theta}_K$, data $\boldsymbol{Y}$, and the slice sampler in Algorithm 2.

Output: Updated centered factors $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ and hyperparameters $\boldsymbol{\theta}_K$.

Construct current Cholesky factors $\boldsymbol{L}_u, \boldsymbol{L}_v, \boldsymbol{L}_w$ from $\boldsymbol{K}_u(\boldsymbol{\theta}_u)$, $\boldsymbol{K}_v(\boldsymbol{\theta}_v)$, and $\boldsymbol{K}_w(\boldsymbol{\theta}_w)$.

Whiten the current centered factors: $\boldsymbol{Z}_u = \boldsymbol{L}_u^{-1}\boldsymbol{U}$, $\boldsymbol{Z}_v = \boldsymbol{L}_v^{-1}\boldsymbol{V}$, and $\boldsymbol{Z}_w = \boldsymbol{L}_w^{-1}\boldsymbol{W}$.

for each hyperparameter θ_j in the origin spatial covariance block **do**

  Define $\phi_j = \log(\theta_j)$ and reconstruct factors by $\boldsymbol{U}(\phi_j) = \boldsymbol{L}_u\{\boldsymbol{\theta}_u(\phi_j)\}\boldsymbol{Z}_u$.

  Define slice target $\mathcal{L}_{\mathrm{NC}}(\phi_j) = \log p(\boldsymbol{Y} \mid \boldsymbol{U}(\phi_j), \boldsymbol{V}, \boldsymbol{W}) + \log p(\exp(\phi_j)) + \phi_j$.

  Update θ_j using Algorithm 2.

end for

Reconstruct $\boldsymbol{U} = \boldsymbol{L}_u(\boldsymbol{\theta}_u)\boldsymbol{Z}_u$ using the accepted origin spatial hyperparameters.

for each hyperparameter θ_j in the destination spatial covariance block **do**

  Define $\phi_j = \log(\theta_j)$ and reconstruct factors by $\boldsymbol{V}(\phi_j) = \boldsymbol{L}_v\{\boldsymbol{\theta}_v(\phi_j)\}\boldsymbol{Z}_v$.

  Define slice target $\mathcal{L}_{\mathrm{NC}}(\phi_j) = \log p(\boldsymbol{Y} \mid \boldsymbol{U}, \boldsymbol{V}(\phi_j), \boldsymbol{W}) + \log p(\exp(\phi_j)) + \phi_j$.

  Update θ_j using Algorithm 2.

end for

Reconstruct $\boldsymbol{U} = \boldsymbol{L}_u(\boldsymbol{\theta}_u)\boldsymbol{Z}_u$ and $\boldsymbol{V} = \boldsymbol{L}_v(\boldsymbol{\theta}_v)\boldsymbol{Z}_v$ using the accepted spatial hyperparameters.

for each hyperparameter θ_j in the temporal covariance block **do**

  Define $\phi_j = \log(\theta_j)$ and reconstruct trial factors by $\boldsymbol{W}(\phi_j) = \boldsymbol{L}_w\{\boldsymbol{\theta}_w(\phi_j)\}\boldsymbol{Z}_w$.

  Define slice target $\mathcal{L}_{\mathrm{NC}}(\phi_j) = \log p(\boldsymbol{Y} \mid \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}(\phi_j)) + \log p(\exp(\phi_j)) + \phi_j$.

  Update θ_j using Algorithm 2.

end for

Reconstruct $\boldsymbol{W} = \boldsymbol{L}_w(\boldsymbol{\theta}_w)\boldsymbol{Z}_w$ using the accepted temporal hyperparameters.

return $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \boldsymbol{\theta}_K$.

Algorithm 4 Overall MCMC Sampling Algorithm for Model Estimation

Input: Observed tensors $\boldsymbol{Y}$, CP rank R, burn-in iterations I_1, sampling iterations I_2.
Output: Posterior samples $\left\{\boldsymbol{U}^{(i)}, \boldsymbol{V}^{(i)}, \boldsymbol{W}^{(i)}, \boldsymbol{\theta}_K^{(i)}\right\}_{i=I_1+1}^{I_1+I_2}$.
Initialize $\boldsymbol{U}^{(0)}, \boldsymbol{V}^{(0)}, \boldsymbol{W}^{(0)}, \boldsymbol{\theta}_K^{(0)}$.
for $i = 1$ to $I_1 + I_2$ **do**
  for $r = 1$ to R **do**
    Sample origin factor $\boldsymbol{u}_r^{(i)}$, $\boldsymbol{v}_r^{(i)}$, $\boldsymbol{w}_r^{(i)}$ using Fisher-preconditioned pMALA with Algorithm 1.
  end for
  Update covariance hyperparameters $\boldsymbol{\theta}_K^{(i)}$ using centered/non-centered interweaving with Algorithm 3; the one-dimensional log-scale updates inside this step use slice sampling with Algorithm 2.
end for
return $\left\{\boldsymbol{U}^{(i)}, \boldsymbol{V}^{(i)}, \boldsymbol{W}^{(i)}, \boldsymbol{\theta}_K^{(i)}\right\}_{i=I_1+1}^{I_1+I_2}$.

Algorithm 5 Posterior Predictive Inference for New Location s

Input: Posterior samples $\left\{\boldsymbol{U}^{(i)}, \boldsymbol{V}^{(i)}, \boldsymbol{W}^{(i)}, \boldsymbol{\theta}_K^{(i)}\right\}_{i=1}^{I_2}$, network information for new location s (to compute covariances), estimated departures $\{N_{st}\}_{t=1}^{T}$ from new location s.
Output: Predictive samples of O-D flows $\left\{\boldsymbol{y}_{st}^{(i)}\right\}_{i=1,t=1}^{I_2,T}$.
for $i = 1$ to I_2 **do**
  Compute $\boldsymbol{K}_u^{(i)}$, $\boldsymbol{K}_v^{(i)}$, $\boldsymbol{k}_{us}^{(i)}$, $\boldsymbol{k}_{vs}^{(i)}$, $k_{uss}^{(i)}$, $k_{vss}^{(i)}$ using the spatial kernel functions.
  Compute predictive variances $\left(\sigma_{us}^2\right)^{(i)}$ and $\left(\sigma_{vs}^2\right)^{(i)}$.
  for $r = 1$ to R **do**
    Compute predictive means $\left(\mu_r^{us}\right)^{(i)}$ and $\left(\mu_r^{vs}\right)^{(i)}$.
    Draw new factors $\left(u_r^s\right)^{(i)}$ and $\left(v_r^s\right)^{(i)}$ from the Gaussian predictive distributions.
  end for
  Construct $\boldsymbol{u}_s^{(i)} = \left[\left(u_1^s\right)^{(i)}, \ldots, \left(u_R^s\right)^{(i)}\right]^\top$ and $\boldsymbol{v}_s^{(i)} = \left[\left(v_1^s\right)^{(i)}, \ldots, \left(v_R^s\right)^{(i)}\right]^\top$.
  for $t = 1$ to T **do**
    Compute logits $G_{sdt}^{(i)} = \sum_{r=1}^{R} \left(u_r^s\right)^{(i)} v_{dr}^{(i)} w_{tr}^{(i)}$ for $d \in \{1, \ldots, S\}$.
    Compute probabilities $\boldsymbol{\lambda}_{st}^{(i)}$ using the masked softmax.
    Draw predictive counts: $\boldsymbol{y}_{st}^{(i)} \sim \text{Multinomial}\left(N_{st}, \boldsymbol{\lambda}_{st}^{(i)}\right)$.
  end for
end for
return All predictive samples $\left\{\boldsymbol{y}_{st}^{(i)}\right\}_{i=1,t=1}^{I_2,T}$.

Appendix H: Some Tables and Figures in the Experiments

Table 1 **Inductive bike O-D forecast performance under varying observed proportions/ranks.**

Observed proportion	Rank	logL (10^6)	MAE	RMSE	CRPS
10%	1	-2.087	0.022574	0.128420	0.021744
10%	2	-2.076	0.022258	0.128816	0.020776
10%	5	-2.035	0.022064	0.128837	0.020203
10%	10	-2.069	0.022070	0.129494	0.019905
10%	15	-2.063	0.021893	0.130349	0.019592
30%	1	-2.069	0.022486	0.128045	0.021557
30%	2	-2.012	0.022078	0.127885	0.020402
30%	5	-1.966	0.021694	0.127955	0.019665
30%	10	-1.946	0.021537	0.127113	0.019410
30%	15	-1.949	0.021519	0.127146	0.019368
50%	1	-2.059	0.022447	0.127984	0.021389
50%	2	-2.000	0.021978	0.127188	0.020160
50%	5	-1.948	0.021498	0.127633	0.019332
50%	10	-1.939	0.021379	0.127960	0.019256
50%	15	-1.925	0.021341	0.127515	0.019239
70%	1	-2.039	0.022385	0.126865	0.021352
70%	2	-1.953	0.021905	0.125528	0.020124
70%	5	-1.882	0.021336	0.123889	0.019196
70%	10	-1.871	0.021168	0.123435	0.019028
70%	15	-1.867	0.021081	0.123455	0.018926
90%	1	-2.029	0.022371	0.126780	0.021259
90%	2	-1.944	0.021873	0.125689	0.020029
90%	5	-1.870	0.021275	0.124456	0.019101
90%	10	-1.856	0.021094	0.124065	0.018920
90%	15	**-1.849**	**0.020993**	**0.123360**	**0.018812**

Table 2 Inductive metro O-D forecast performance with varying observed proportions/ranks.

Observed proportion	Rank	logL (10^6)	MAE	RMSE	CRPS
10%	1	-148.717	8.4009	23.7554	7.2945
10%	2	-147.944	8.2723	23.6477	7.1934
10%	5	-149.648	8.3659	24.1015	7.3290
10%	10	-148.320	8.2014	24.8707	7.1675
10%	15	-150.149	8.4064	24.6396	7.3545
10%	20	-150.178	8.4189	24.3917	7.3633
30%	1	-146.642	7.8128	21.7549	6.7623
30%	2	-145.239	7.5330	21.1748	6.5115
30%	5	-144.533	7.2463	21.0766	6.2950
30%	10	-144.598	7.1568	21.4866	6.2350
30%	15	-144.740	7.1667	21.5719	6.2470
30%	20	-145.229	7.2181	21.5942	6.2974
50%	1	-145.916	7.7158	21.9070	6.6988
50%	2	-144.349	7.4116	21.3580	6.4346
50%	5	-143.971	7.1818	23.6866	6.2871
50%	10	-142.938	6.8255	24.5503	5.9697
50%	15	-142.730	6.7415	24.0499	5.8909
50%	20	-142.229	6.6422	22.8594	5.7899
70%	1	-142.329	7.0743	20.3997	6.0725
70%	2	-139.870	6.4902	19.2653	5.5425
70%	5	-135.932	5.4306	17.0432	4.5726
70%	10	-134.781	4.9498	17.0001	4.1509
70%	15	-134.532	4.8292	16.5844	4.0400
70%	20	-134.376	4.7289	16.4594	3.9512
90%	1	-141.144	6.7820	19.7189	5.7967
90%	2	-138.360	6.1439	18.2477	5.2116
90%	5	-133.795	4.7994	14.9143	3.9736
90%	10	-131.569	3.8913	12.8672	3.1448
90%	15	-131.233	3.7382	12.1666	3.0053
90%	20	**-130.988**	**3.5930**	**11.8154**	**2.8754**

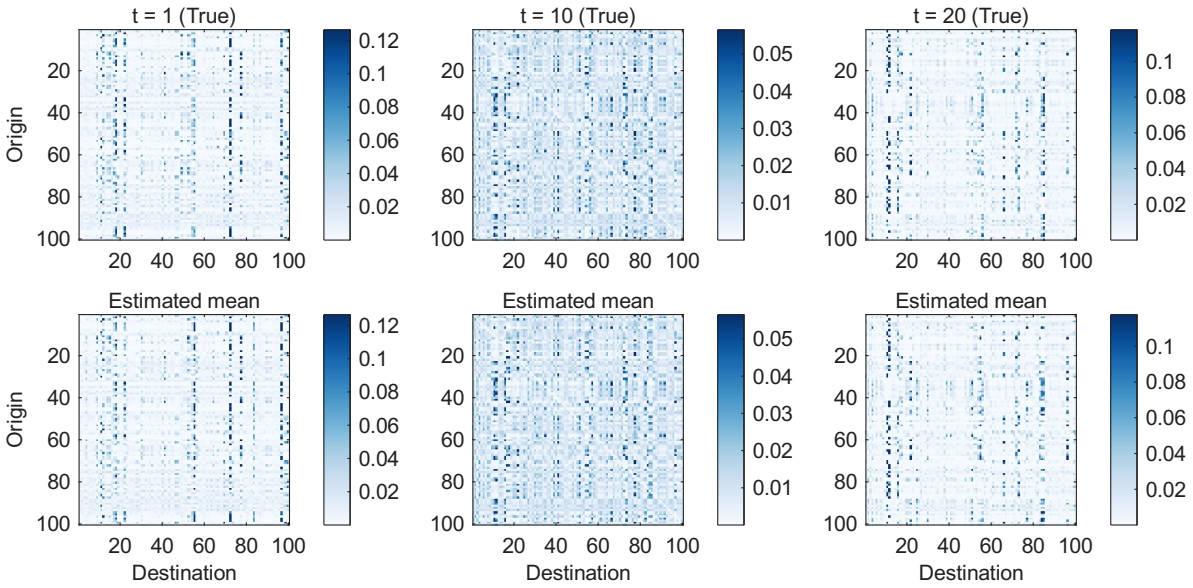

Figure 1 Visual comparison of O-D assignment probability matrices.

Note. The top row displays the ground truth probabilities for $t = 1, 10, 20$, while the bottom row displays the corresponding estimated posterior means. The color intensity represents the probability magnitude.

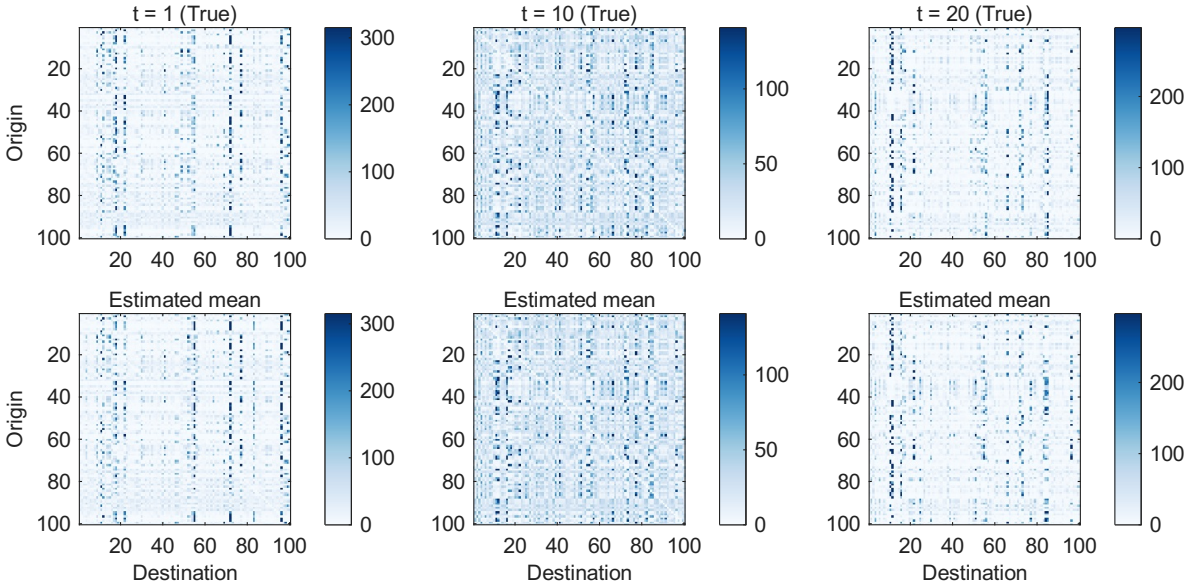

Figure 2 Visual comparison of O-D matrices.

Note. The top row displays the true O-D matrices for $t = 1, 10, 20$, while the bottom row displays the corresponding estimated posterior O-D means. The color intensity represents the O-D flow.

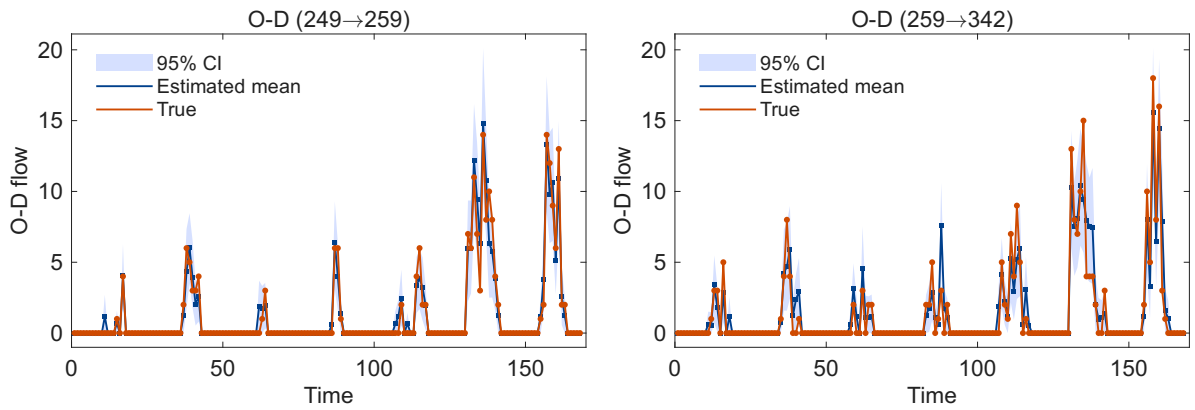

Figure 3 Estimated bike O-D flows with uncertainty quantification.

Note. The shaded regions display posterior predictive 95% credible intervals, the blue line is the posterior predictive mean, and the red points/line are the observed validation flows.

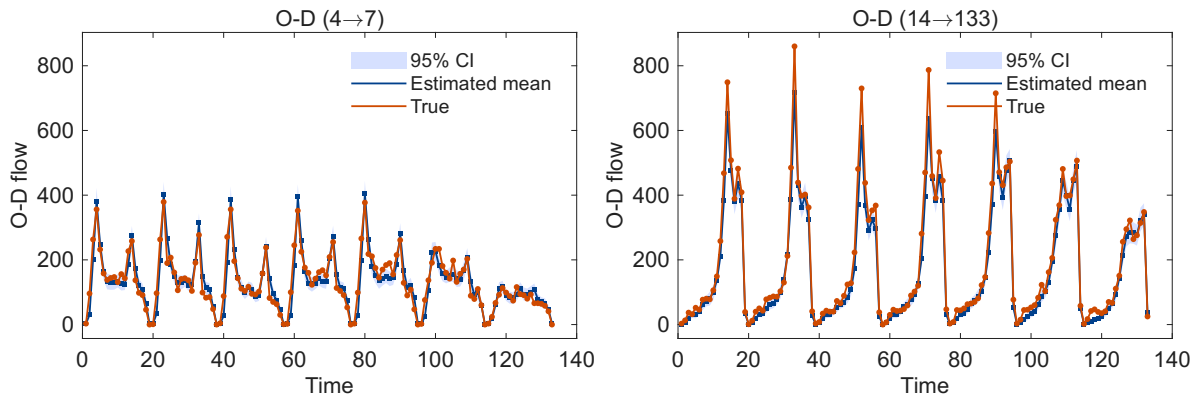

Figure 4 Estimated metro O-D flows with uncertainty quantification

Note. The shaded regions display posterior predictive 95% credible intervals, the blue line is the posterior predictive mean, and the red points/line are the observed validation flows.

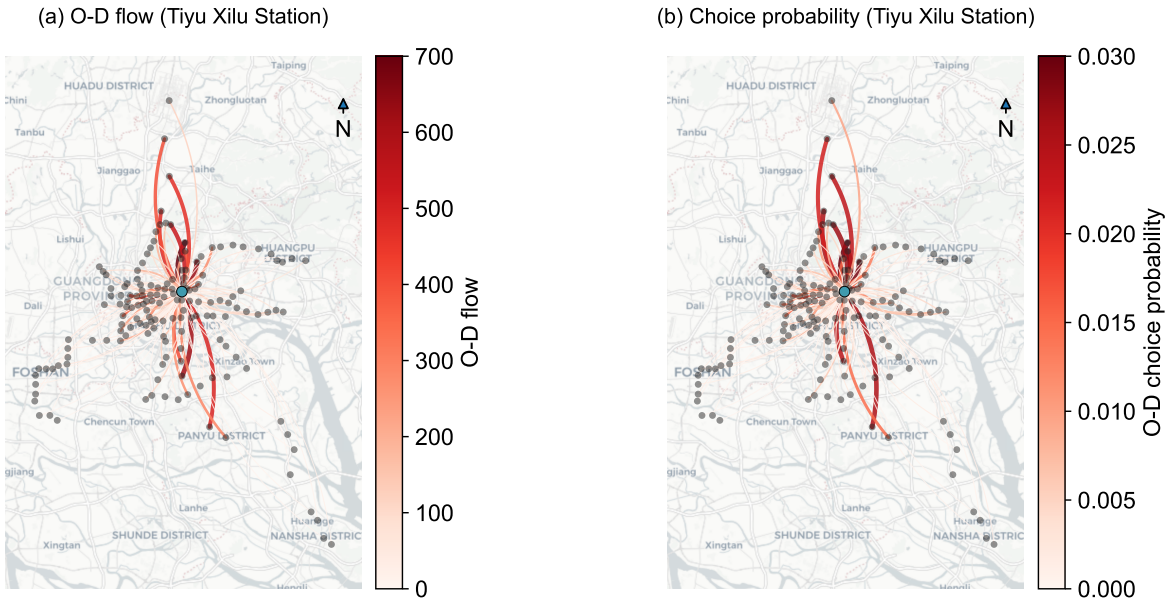

Figure 5 **True O-D flow and the estimated O-D choice probabilities for Tiyu Xilu Station.**

Note. The left panel shows aggregate validation O-D flows from the selected origin, and the right panel shows the estimated destination-choice probabilities.

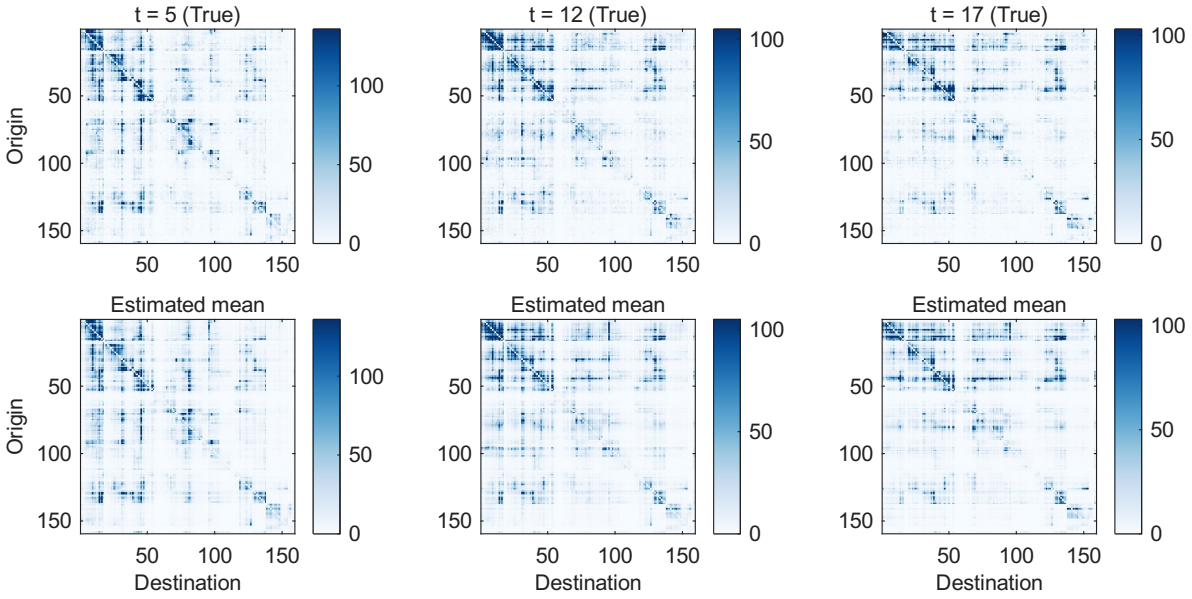

Figure 6 **Visual comparison of estimated and true metro O-D matrices.**

Note. The top row displays the true O-D matrices for $t = 5, 12, 17$, while the bottom row displays the corresponding estimated posterior O-D means. The color intensity represents the O-D flow on log scale.

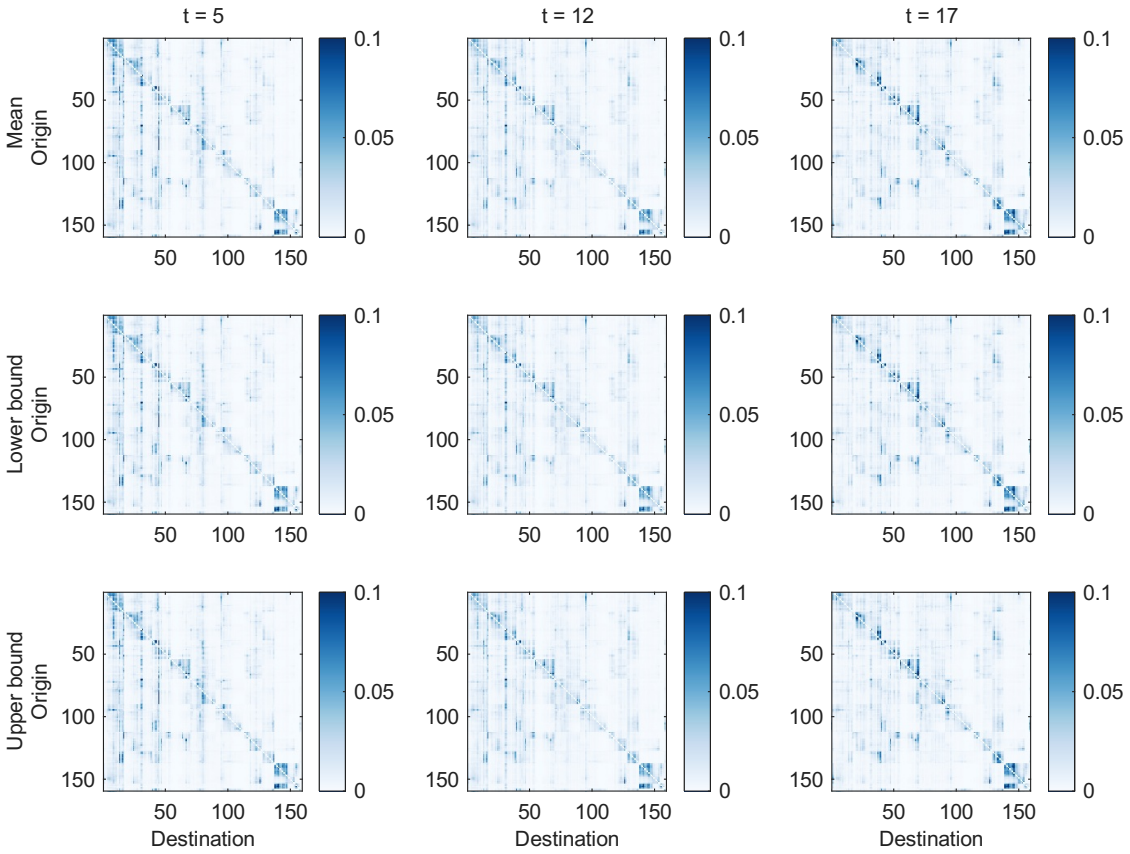

Figure 7 Estimated metro O-D assignment probability matrices.

Note. The top row displays the estimated posterior mean of probabilities for $t = 5, 12, 17$, while the second and the last rows respectively display the corresponding estimated lower and upper bounds of 95% CIs. The color intensity represents the probability magnitude.