CSTP: DATA-EFFICIENT ROAD-CLOSURE PREDICTION WITH STRUCTURED INVENTORY DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Road-closure planning requires rapid predictions, even though comparable disruptions are relatively rare in real-world data. We present the Closure-aware Structured Traffic Predictor (CSTP), which predicts vehicle inventory on road segments using a road-network graph, a closure query, and a corresponding normal-traffic scenario. CSTP represents departures from normal traffic through a compact physical state. Queue storage captures vehicles retained on each segment, while learned estimates of congestion-wave propagation, inventory depletion, and junction interactions determine how congestion develops over time and across the network.

We learn road profiles and wave behavior from normal traffic. Separate closure scenarios calibrate only two response multipliers. Vehicle counts are computed from the resulting state, while learned prediction heads estimate speed and flow. We evaluate individual mechanisms using matched simulations and conduct grouped SUMO experiments in which test locations are excluded from parameter fitting and model selection. Relative to the Cell Transmission Model (CTM), CSTP reduces the mean absolute error of local vehicle counts by 28.8% and regional count error by 9.6%, with improvements in every tested local road group. Count error increases only slightly as the number of closure labels decreases, and CSTP retains the lowest core count error across the tested label budgets. PeMS incident records provide an additional real-world evaluation of speed and flow prediction. Repeated CPU tests show that CSTP is 5.2$\times$ faster at inference and uses 49% less peak memory than CTM combined with learned prediction heads.

1 INTRODUCTION

Traffic congestion imposes costs through lost travel time, unreliable journeys, excess fuel use, and emissions (Schrank et al., 2025; Barth & Boriboonsomsin, 2008). To address these costs, intelligent transportation planning relies on predictions to assess traveler guidance (Ben-Akiva et al., 2001), coordinated control (Hegyi et al., 2005), and work-zone plans (Sankar et al., 2006). For these decisions, forecasts must locate and quantify congestion accurately while remaining fast enough to compare alternative plans (Mallick et al., 2020). For road-closure planning, these requirements mean estimating how vehicles will accumulate and how the effects will spread across surrounding roads before choosing where and when to restrict traffic and putting a plan into practice.

Planning ahead requires *counterfactual prediction*: estimating an unobserved closure response under unchanged background conditions. Comparable events may be scarce at the queried location, time, and demand level (Fan et al., 2026; Li et al., 2024). A closure also changes possible movements, so relationships learned during normal operation may no longer hold (Xu et al., 2023; Schölkopf et al., 2021). Pearl (2009) explains why this matters: observational associations alone do not generally determine intervention outcomes. Predicting those outcomes therefore requires causal assumptions about the response to blocked movements. A causal structure makes these assumptions explicit: it specifies which variables interact and how interventions change their mechanisms. In traffic, the road network provides a natural starting structure: connectivity constrains interactions, while physical mechanisms govern upstream, downstream, and road junction effects.

Through these network interactions, a closure changes congestion by retaining vehicles upstream and reducing arrivals downstream (Daganzo, 1994; 1995). To quantify the resulting accumulation and depletion, we focus on *vehicle inventory*: the number of vehicles present on each road segment. These

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

counts provide an intuitive, direct measure of network accumulation, central to macroscopic descriptions of congestion (Geroliminis & Daganzo, 2008). Relative to road storage, they also reveal where accumulation approaches physical limits. By contrast, detector counts measure throughput rather than the vehicles present (Klein et al., 2006); speed and flow provide complementary information about their movement.

To predict these inventory changes, we start from a supplied *normal scenario*: road-level count, speed, and flow trajectories over the planning horizon assuming no closure. Together with a road graph and closure query, this reference defines the inputs, without requiring origin–destination (OD) demands, route records, or target-closure observations. Some related causal approaches, however, require OD demand and route information as core inputs (Xu et al., 2023). This leaves a practical gap in closure-inventory prediction when only a road graph and aggregate normal traffic are available.

We address this gap with the *Closure-aware Structured Traffic Predictor (CSTP)* for *what-if closure queries*. CSTP tracks added queues, inventory deficits, and propagation progress on the road network, representing bounded changes from normal inventory. In this state, *queue storage* captures retained vehicles, while *learned wave propagation*, *inventory depletion*, and *junction interactions* describe when and where effects spread. Learning road profiles and wave estimates from normal traffic leaves only two response multipliers to calibrate on separate closures for this inventory state. The state directly predicts vehicle counts. Separate heads predict speed and flow.

We first use matched simulations to determine whether the proposed mechanisms reproduce the effects of road closures. We then evaluate the complete model in terms of prediction accuracy, data requirements, and inference cost. Our main contributions are: **(a) Network-based prediction of closure-induced vehicle accumulation.** The model represents storage, deficits, and congestion propagation across connected road segments. Matched simulations and ablation studies examine the role of each mechanism (Sections 3–4, 6). **(b) Accurate count prediction with limited closure data.** Under a shared set of closure labels, CSTP achieves the lowest count mean absolute error (MAE) at every tested training budget. Training with only two groups increases MAE by 1.5% relative to training with all available groups (Section 5). **(c) Network-scale accuracy and computational efficiency.** CSTP achieves the lowest count MAE and RMSE in both evaluation regions. In repeated CPU tests, a two-hour network forecast takes a median of 2.43 s and uses 0.323 GiB of peak RAM. Additional speed, flow, and PeMS evaluations examine the model's extensions (Section 5).

2 Background and related work

Traffic dynamics and junction interactions. Kinematic-wave theory describes density and flow propagation (Lighthill & Whitham, 1955; Richards, 1956; Newell, 1993). CTM extends these principles to finite-storage road networks (Daganzo, 1994; 1995); node models allocate receiving capacity among competing movements (Tampère et al., 2011). These principles also regularize learned state estimation (Shi et al., 2021). CSTP builds on these established traffic mechanisms but organizes them as bounded changes to a supplied normal inventory trajectory. The distinction from our reference-anchored CTM comparator is not residual prediction alone: that comparator also adds a closure-minus-open correction to the normal reference. CTM obtains this correction by evolving cell inventories through demand–supply fluxes; CSTP evolves an explicit road-level response state, with road profiles and wave estimates learned from normal traffic and two physical response multipliers calibrated on closures. Our comparisons therefore test this representation and calibration strategy under shared aggregate inputs, rather than proposing a replacement for the underlying traffic laws.

Graph forecasting and incident-aware prediction. Graph models learn spatiotemporal dependence through diffusion, convolution, adaptive connectivity, or learned embeddings (Li et al., 2018; Yu et al., 2018; Wu et al., 2019; Liu et al., 2023). PDFormer also models propagation delays (Jiang et al., 2023), while IGSTGNN models the spatial influence and temporal decay of incidents (Fan et al., 2026). Our setting instead supplies a complete normal scenario and a specified closure, without traffic history from the target closure. We therefore adapt incident-aware prediction to these inputs, including a count-output IGSTGNN comparator.

Causal modeling and traffic interventions. A closure changes permitted movements. Structural interventions express such mechanism changes, while causal abstraction relates descriptions at different levels (Pearl, 2009; Beckers et al., 2020). For CSTP, this means specifying aggregate traffic

responses and testing them against vehicle-level simulations. Reusing the same response rules across closures further requires mechanism modularity (Schölkopf et al., 2021). Related treatment-sequence models address counterfactual outcomes and time-varying confounding (Bica et al., 2020; Melnychuk et al., 2022). In traffic, TraffNet predicts volume using trajectory-derived demand and routes (Xu et al., 2023); MSCT predicts speed responses to crash-treatment sequences (Li et al., 2024). Because their targets and inputs differ from closure inventory prediction, we adapt each model to the shared inputs and compare the outcomes it predicts (Appendix B). CSTP's speed and flow heads make these comparisons possible and allow evaluation on detector records.

3 CSTP: AN INVENTORY STATE FOR CLOSURE RESPONSES

CSTP expresses these traffic responses as a recurrent structural causal model (SCM) (Pearl, 2009): blocked vehicles queue, fewer reach downstream roads, and junction competition changes. Traffic assumptions specify the dependencies, while learning fits parameters and observation maps. Queues, deficits, and propagation progress describe changes from normal traffic, determining counts and features for speed and flow prediction.

3.1 CLOSURE INTERVENTION AND REFERENCE CONDITIONS

To isolate closure effects from demand changes, we match background conditions between restricted and normal operation. Let G describe roads and permitted connections. Intervention $a = (C, t_0, b)$ restricts road C from t_0. Boundary b specifies entry-only restriction, allowing initial occupants to leave, or a stopped road, also blocking discharge; $a = 0$ denotes normal operation. In SCM terms, $\mathrm{do}(A = a)$ sets the closure variable A to a, retaining the other response mechanisms (Xia et al., 2021). Under the same external conditions u (demand, initial conditions, and simulation randomness), the effect is

$$\mathbf{\Delta}_k(a; u) = \mathbf{Y}_k(a; u) - \mathbf{Y}_k(0; u), \tag{1}$$

Here $\mathbf{Y}_k$ contains road counts, speeds, and entering flows at step k, and supplied normal values $\mathcal{B} = (N^0, V^0, Q^0)$ provide the reference. Thus, the causal question is conditional on this reference and matched background conditions; transfer across closure locations requires the response mechanisms to remain applicable (Schölkopf et al., 2021).

3.2 FROM THE INTERVENTION TO TRAFFIC MECHANISMS

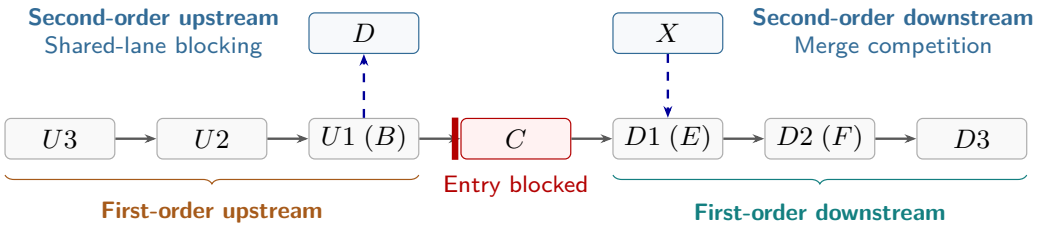

Figure 1: **Four fundamental local road structures.** Arrows follow traffic; entry into C is blocked. Dashed links add branch D or feeder X.

At step k, S_k^a records queues, propagation progress, inventory deficits, and extra merge departures for closure $a = (C, t_0, b)$. First, F_θ updates the preceding state S_{k-1}^a; H_θ then converts the updated state into predicted counts, speeds, and entering flows $\widehat{\mathbf{Y}}_k^a$ (Figure 1):

$$S_k^a = F_\theta(S_{k-1}^a; \mathcal{B}, G, M_k^a, k), \qquad \widehat{\mathbf{Y}}_k^a = H_\theta(S_k^a; \mathcal{B}, G, \Psi_k^a). \tag{2}$$

Both use road graph G, normal count/speed/flow reference $\mathcal{B}$, and fitted parameters θ, fixed during prediction. The movement mask M_k^a blocks entry into road C from onset t_0. Closure type b permits existing occupants to leave (entry-only) or blocks departure (stopped road). Speed/flow additionally use Ψ_k^a: road position relative to C and time since t_0. These updates act through the four local

structures in Figure 1. Upstream chains ($U1$–$U3$) accumulate queues, while downstream chains ($D1$–$D3$) lose arrivals: *first-order* effects along the route through C. Across routes, shared-lane branches (D) lose inflow through queue obstruction, whereas feeder X can discharge more as merge competition falls. These *second-order* effects connect through junctions $B = U1$ and $E = D1$.

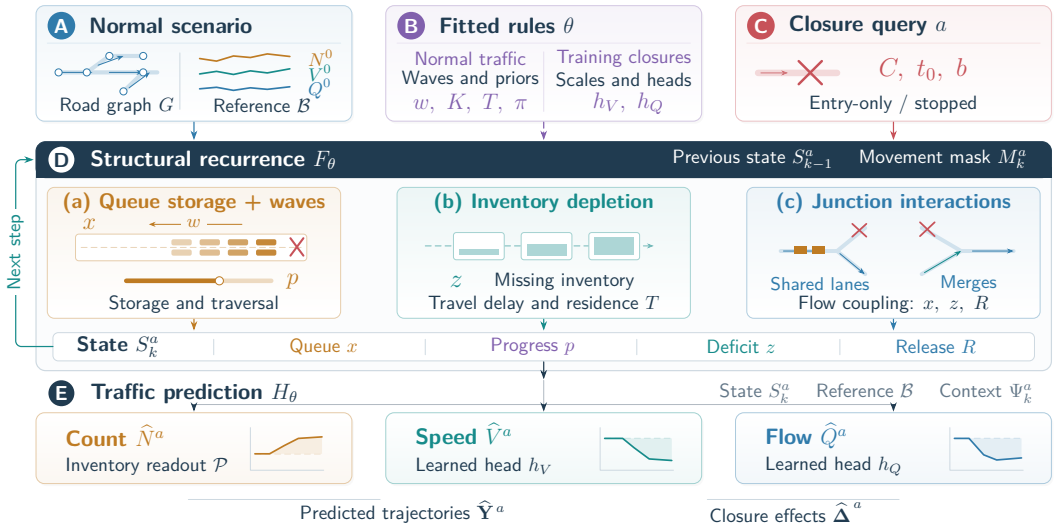

Figure 2: **CSTP implementation framework.**

Figure 2 shows how CSTP uses the road network, normal traffic reference, and closure query to predict traffic changes. Queue storage and wave propagation describe upstream buildup, while inventory depletion and junction interactions capture effects downstream and on connected roads. The resulting state supports count, speed, and flow prediction. We explain these components below.

Queue storage: where blocked vehicles accumulate. The additional queue x_e tracks vehicles retained upstream of C; finite storage limits accumulation and connects blockage to spillback. With step length δ and road storage K_e, available space is $A_{e,k} = [K_e - N^0_{e,k}]_+$:

$$x_{e,k} = \mathrm{clip}_{[0,A_{e,k}]}(x_{e,k-1} + \delta\,(d_{e,k} - r_{e,k}))\,. \tag{3}$$

Retained blocked demand d and queue release r are rates in vehicles/s. Normal throughput, blocked movement shares, available alternatives, and a shared demand cap determine these rates (Appendix A). As storage fills, blockage can reach the preceding road. Clipping enforces bounds; $[\cdot]_+$ denotes the positive part.

Learned wave propagation: when blockage spreads. Storage must be coupled with queue motion to predict arrival time. CSTP therefore learns speed magnitude w_e from occupancy o_e and positive count growth s^+_e, allowing traffic-dependent motion. This *learned wave law* advances progress p_e across the road, defining the *traversal delay*:

$$\begin{aligned} w_e &= \mathrm{clip}_{[w_{\min},w_{\max}]}\big(\exp\{\alpha_e + \beta_o o_e + \beta_s \log(1+s^+_e)\}\big)\,, \\ p_{e,k} &= \min\{1, p_{e,k-1} + \chi_{e,k} w^{\mathrm{app}}_e \delta / L^*_e\}\,. \end{aligned} \tag{4}$$

Normal queue movements fit intercepts α_e and slopes β_o, β_s. Here χ marks active travel, L^*_e is bounded road length, and w^{app} uses a scale calibrated on separate closures. Spillback then requires completed traversal ($p = 1$) and sufficient occupancy: storage and motion jointly set arrival time.

Inventory depletion: how lost arrivals move downstream. While upstream queues grow, lost arrivals deplete downstream inventory as vehicles leave. Missing arrivals m^{in} increase deficit z; missing departures m^{out} reduce it. CSTP therefore uses a reservoir update and a single-inflow corridor approximation:

$$z_{e,k} = \mathrm{clip}_{[0,N^0_{e,k}]}\big[z_{e,k-1} + \delta(m^{\mathrm{in}}_{e,k} - m^{\mathrm{out}}_{e,k})\big]\,,\ z_e(t) \quad = N^0_e(t)\,\mathrm{clip}_{[0,1]}((t - t_0 - d_e)/T_e) \tag{5}$$

Residence time T_e gives $m^{\text{out}} = \min(z_{e,k-1}/T_e, \bar{q}^0_{e,k})$. Boundary b determines when discharge from C is lost; d_e includes initial drainage for entry-only restrictions. The benchmark uses the reservoir update. In the controlled corridor tests, the finite-residence approximation ends at active merges, where other inflow can replace lost arrivals.

Junction interactions: why connected branches also respond. Beyond these direct effects, junction flows couple other routes to the queue and deficit updates. Where movements $B \to \{C, D\}$ share a lane, a queue for C obstructs traffic bound for D, adding blocked fraction

$$\beta_{BD,k} = \min\Big\{\pi_{BD}\ell_{BD}(1 - e^{-x_{B,k-1}/q_*}),\ 1 - b_{B,k}\Big\}. \tag{6}$$

Here π_{BD} is movement share, ℓ_{BD} lane overlap, and q_* queue scale. The resulting lost flow $\bar{q}^0_{B,k}\beta_{BD,k}$ adds blockage on B and missing arrivals on D. This second-order upstream effect vanishes without a queue or shared lanes. By contrast, at the downstream merge $C \to E \leftarrow X$, the coupling releases waiting traffic: lost inflow L^{lost}_E from C frees space for vehicles on X. Extra discharge is bounded by that inflow, spare capacity H_X, and remaining waiting stock:

$$v_{X,k} = \min\left\{L^{\text{lost}}_{E,k}, H_X, \frac{[N^0_{X,k} - M_X - R_{X,k-1}]_+}{\delta}\right\}, \qquad R_{X,k} = R_{X,k-1} + \delta v_{X,k}. \tag{7}$$

Here M_X estimates moving vehicles and R_X records extra departures, which lower inventory on X and restore arrivals to E. This rule applies where waiting stock can be estimated. Thus, both junction rules alter flows in the existing queue and depletion updates.

3.3 Count prediction and speed/flow extensions

Combining these mechanisms gives road inventory: on open roads, counts add queues to normal inventory and subtract deficits and merge departures. Two *learned output heads*, h_V and h_Q, extend the state to speed and flow corrections:

$$\begin{aligned}
\widehat{N}^a_{e,k} &= \mathcal{P}_{e,k}\left(N^0_{e,k} + x_{e,k} - z_{e,k} - R_{e,k}\right), \\
\widehat{V}^a_{e,k} &= \text{clip}_{[0,V^{\max}_{e,k}]}\big(B^V_{e,k} + \lambda_V v^*_e h_V(\phi_{e,k})\big), \\
\widehat{Q}^a_{e,k} &= \left[Q^0_{e,k} + \lambda_Q \ell_e h_Q(\phi_{e,k})\right]_+.
\end{aligned} \tag{8}$$

Here $\mathcal{P}$ enforces nonnegative counts and storage bounds, while the heads use features ϕ from reference traffic, physical state, and closure context. The *physical-state features* comprise predicted count, its change from normal, their storage ratios, and time since wave activation. B^V is normal speed, v^* reference speed, ℓ lane count, and λ_V, λ_Q fitted strengths. The heads therefore use the inventory state without changing its counts. The bounds and readouts do not enforce exact network conservation. Section 6 tests the roles of the state features and heads by removing them separately.

Normal traffic supplies road estimates and wave supervision; closure labels calibrate the wave-speed and blocked-demand-cap multipliers. Heads have separate parameters. Section 6 isolates physical calibration by fixing both multipliers to one. Appendix B records fitting and development provenance; Appendix A gives the remaining updates.

Subtracting the reference then gives the closure effect, $\widehat{\boldsymbol{\Delta}}^{\,a}_k = \widehat{\mathbf{Y}}^{\,a}_k - \mathcal{B}_k$, as in Equation (1). Its causal interpretation depends on the specified mechanisms and matched background. We therefore use matched simulations to test correspondence to vehicle-level effects (Beckers et al., 2020), and observed incidents to assess predictive accuracy. Sections 4 and 6 examine individual responses and their contributions to the complete predictor; Table 11 distinguishes the configurations.

4 Controlled tests of closure-response mechanisms

Section 3 links a closure to storage, propagation, and junction responses. We therefore test whether these mechanisms reproduce arrival timing and count effects in the four road structures of Figure 1. Each test pairs SUMO runs with and without blocked entry into C, keeping departures and routes identical. Their count difference gives the causal effect in Equation (1), across varied demand, initial

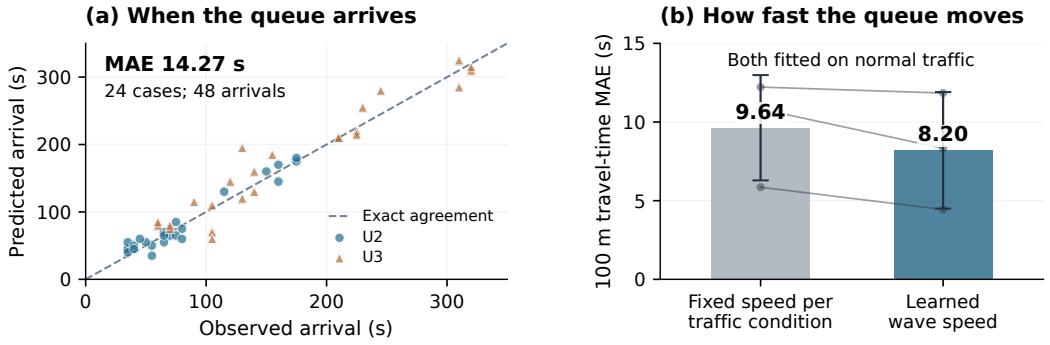

Figure 3: **Propagation timing and learned wave speed.** (a) Predicted versus SUMO arrival times at successive upstream roads $U2/U3$; the dashed line marks perfect agreement. (b) Learned versus fixed wave speeds under standard traffic, lines connect the same run.

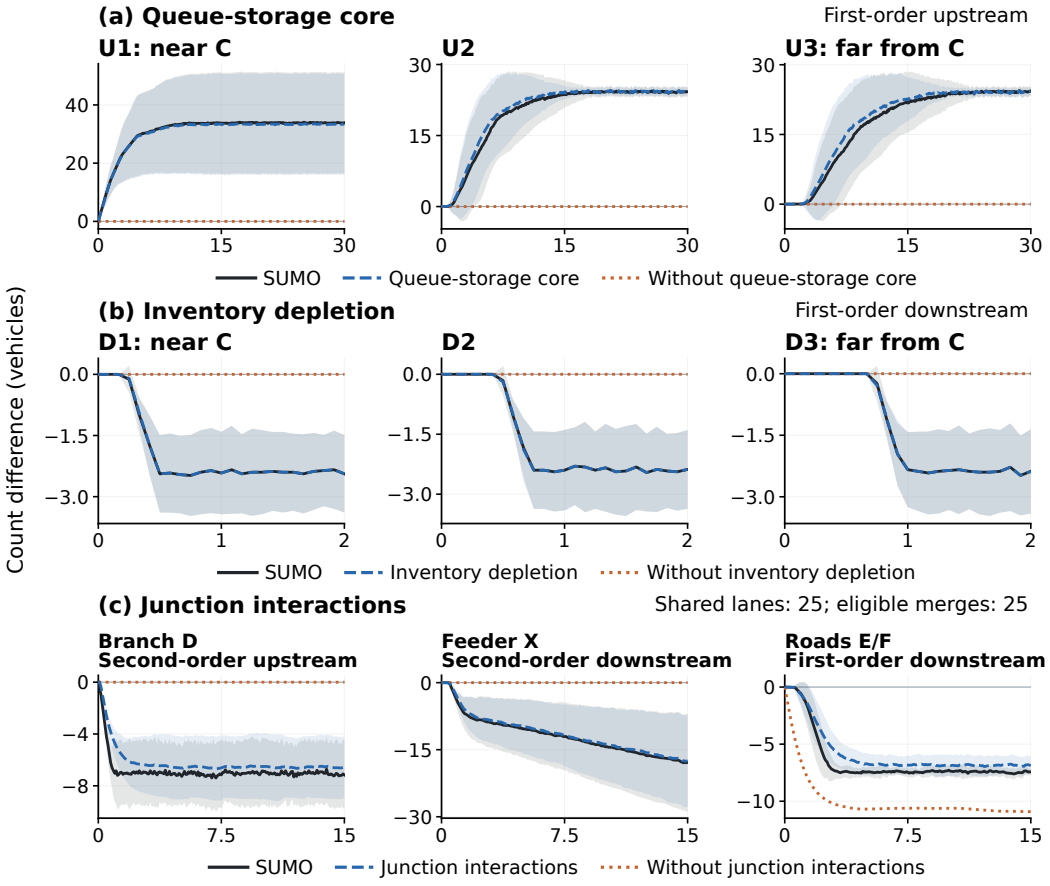

Figure 4: **Causal effects on vehicle counts.** (a) Upstream storage, (b) downstream depletion, and (c) shared-lane and merge interactions. Black: SUMO; blue: component on; orange: off. Curves show case means and bands one standard deviation for SUMO and the enabled component.

queues, and driving behavior. To isolate components, normal queue trajectories directly supervise the timing study, and the corridor study uses the finite-residence refinement. Sections 5 and 6 then evaluate the mechanisms together in the benchmark, using count-onset wave supervision and reservoir depletion. Appendix E gives the settings and measurement protocols.

Causal effect propagation time. Storage and wave propagation jointly predict arrival at $U2/U3$. Complete arrival MAE is 14.27 s for the learner (Figure 3a), versus 12.71 s for condition-specific constants; selection used calibration count MAE (Appendix E). To isolate wave estimation, independent normal-traffic tests compare the learned law with condition-specific fixed speeds. (For each traffic condition, the fixed wave speed is obtained by dividing the traversal distance by the median time taken by the queue tail to move that distance in normal training runs) Learning reduces traversal-time MAE from 9.64 to 8.20 s (14.9%; Figure 3b). This gain concerns direct traversal estimates; it does not extend to complete arrival timing after storage gating.

Causal effect magnitude. Component removals then test the magnitude of upstream, downstream, and junction responses (Figure 4). Positive count effects mean extra vehicles; negative effects mean fewer. **(a) Queue-storage core: first-order upstream effects.** For upstream chains, the storage update predicts accumulation on $U1$–$U3$, reducing count MAE from 22.76 to 0.82 vehicles (Figure 4a). **(b) Inventory depletion: first-order downstream effects.** The same blockage reduces downstream arrivals while existing vehicles depart. Accordingly, inventory depletion reproduces count decreases at $D1$–$D3$, reducing MAE from 1.72 to 0.04 vehicles (Figure 4b). **(c) Junction interactions: second-order effects across routes.** Beyond the restricted route, shared-lane blocking suppresses arrivals to D; modeling it reduces count MAE from 6.86 to 0.64 vehicles. By contrast, at the merge, lower inflow from C releases waiting traffic on X and partly restores arrivals to E/F. Including this competition lowers MAE from 11.88 to 0.87 on X and from 3.50 to 0.98 on E/F (Figure 4c). These tests isolate each response; Section 6 measures its contribution to the complete predictor.

5 Predictive accuracy and computational efficiency

After testing individual mechanisms, we evaluate the complete predictor across closure locations: inventory accuracy, data requirements, inference cost, and speed and flow prediction. The Nagoya SUMO benchmark evaluates 16 stopped-road closures on a 1,969-road network using grouped cross-validation. All methods receive the same road graph and normal traffic reference, without observations from the closure being predicted. We evaluate two-hour forecasts within eight-hop neighborhoods and a fixed 765-road network core selected by connectivity to reduce dilution from weakly affected roads. Vehicle count is the primary outcome, while speed and flow assess the model's extensions. We report MAE and root mean squared error (RMSE), with mean $\pm$ sample standard deviation across road groups.Anchored CTM provides the baseline vehicle count. CTM+heads adds two separately trained models using CSTP's architecture, to predict speed and flow from CTM states. The original CTM count predictions remain unchanged. Alongside these physical baselines, IGSTGNN provides a neural count comparison, while TraffNet and MSCT provide flow and speed comparisons, respectively. Because their original input requirements differ from our task, we adapt them to the common aggregate inputs. We evaluate these adapted implementations under the shared input setting; their training details are provided in Appendix B. The 16 SUMO cases involve critical roads whose closure substantially affects network congestion, while results for 10 additional non-critical closures are reported in Table 21.

Count accuracy across closure locations. CSTP has the lowest count MAE and RMSE in both scopes (Table 1). Relative to CTM, count MAE falls by **28.8% locally** and **9.6% across the core**; CTM+heads retains the same count predictions. Core count MAE is also 25.1% below IGSTGNN. Local count gains hold in every test road group (Appendix B.1).

Additional speed and flow prediction. CSTP's largest gains are in vehicle-count prediction. Its speed and flow predictions remain competitive. Its regional speed/flow MAE and local flow RMSE remain within 3% of CTM+heads. Although local speed/flow improvement intervals include zero (Appendix B.1), the mean advantages persist under reference and alignment checks (Appendices C.1 and G). The main benefit is therefore improved inventory prediction alongside comparable performance on the other traffic outputs.

Real-world detector-based evaluation of the extensions. To test transfer beyond simulation, we use ten Caltrans PeMS records spanning seven dates and a 2,944-link District 4 network (Caltrans, n.d.). CSTP transfers its SUMO-fitted response and heads, learning propagation from prior normal traffic. Methods share queries and normal references. CSTP has the lowest mean local speed/flow MAE and RMSE: flow MAE is **27.5% lower** than CTM and speed MAE 10.9% lower than MSCT (Table 2).

Table 1: **SUMO count accuracy and speed/flow extensions.** Mean ± sample SD across road groups; lower is better. Bold marks the lowest mean; dashes denote unsupported outputs. Neural rows use aggregate-input adaptations. CTM+heads uses CTM-only eligibility (Appendix C). Count is the primary metric; speed and flow are secondary extensions.

	Count (veh/road)		Speed (m/s)		Flow (veh/min/road)	
Method	MAE	RMSE	MAE	RMSE	MAE	RMSE
(a) Near closures: case-specific eight-hop neighborhoods						
CTM	14.27±7.21	43.76±36.79	2.88±1.05	5.72±1.75	11.68±3.80	19.97±5.57
CTM+heads	14.27±7.21	43.76±36.79	2.43±0.75	4.56±1.04	9.45±3.76	**15.36**±5.26
IGSTGNN	28.70±11.30	79.76±35.02	6.92±2.37	9.32±2.39	12.96±4.73	21.62±6.70
TraffNet	—	—	—	—	14.74±4.08	23.48±6.33
MSCT	—	—	3.94±1.31	5.83±1.91	—	—
CSTP	**10.16**±6.87	**36.70**±36.86	**2.21**±0.90	**4.32**±1.06	**9.32**±3.90	15.81±5.39
(b) Regional coverage: fixed 765-road core						
CTM	4.71±0.39	16.57±5.08	3.08±0.08	5.07±0.16	15.72±0.22	24.22±0.24
CTM+heads	4.71±0.39	16.57±5.08	**3.00**±0.07	4.89±0.10	**15.44**±0.32	**23.73**±0.40
IGSTGNN	5.69±0.97	25.65±7.00	3.29±0.16	5.28±0.29	15.52±0.21	23.84±0.23
TraffNet	—	—	—	—	15.58±0.16	23.83±0.28
MSCT	—	—	3.01±0.11	**4.41**±0.17	—	—
CSTP	**4.26**±0.26	**14.81**±5.24	3.02±0.08	4.93±0.12	15.52±0.29	23.93±0.31

Table 2: **PeMS speed/flow extensions.** MAE (mean ± sample SD across seven date groups).

Method	Speed (mph)	Flow (veh/5 min)
CTM	21.94±3.04	187.78±70.92
IGSTGNN	21.38±3.38	226.69±132.06
TraffNet	—	234.67±133.73
MSCT	19.08±2.39	—
CSTP	**17.00**±2.52	**136.20**±24.94

Table 3: **Closure-data efficiency.** Core count MAE versus training-group budget.

Method	2	4	All
CSTP	**4.27**	**4.25**	**4.21**
CTM	4.71	4.71	4.71
Graph neural	5.46	5.14	4.84
Direct trees	4.94	4.81	4.66
Residual trees	4.87	4.61	4.37

Table 4: **CPU inference cost.** Median time and peak RAM.

Method	Time (s)	RAM (GiB)
CTM	12.56	0.505
IGSTGNN	56.15	0.801
TraffNet	8.42	0.702
MSCT	43.21	0.645
CSTP	**2.42**	**0.316**

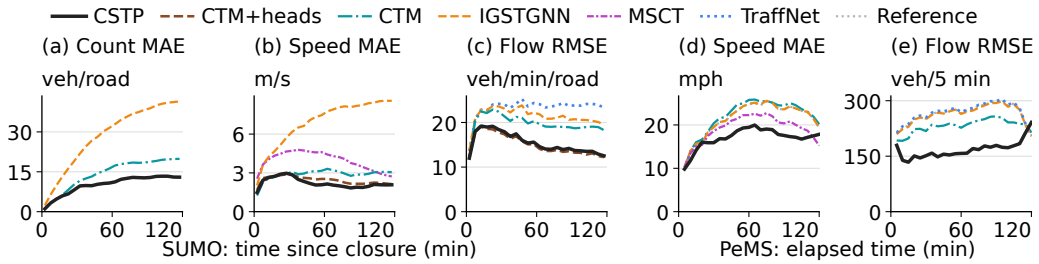

Figure 5: **Temporal accuracy.** (a–c) SUMO eight-hop errors; (d,e) PeMS incident errors.

This evaluates retrospective speed/flow, while inventory outcomes are unavailable. Appendix F details collection checks, query assumptions, and results. The change in prediction error over time for SUMO closures and PeMS incidents is shown in Figure 5.

Closure-data efficiency. To assess closure-data (traffic observations from training closure cases) requirements, we compare CSTP with CTM and count-prediction controls based on GraphSAGE and boosted trees. Each method receives the same sampled count observations from two, four, or all available training groups. This complements the traffic-model benchmark in Table 1 with a focused comparison of vehicle-count prediction under limited closure supervision. CSTP has the lowest core count MAE at every budget (Table 3). Its two-group MAE is 4.27, only **1.5% above** the all-data value of 4.21, and below even the all-data controls. Local MAE rises by 3.7% from all-data to two-group training. CSTP therefore retains count accuracy with few closure groups under this protocol. Appendix C.2 gives all cases, subset ranges, and controls.

Computational efficiency. Closure planning also requires fast prediction. A two-hour forecast takes a median **2.43 s** and uses **0.323 GiB** peak RAM in repeated CPU tests, including the heads: **5.2× faster** and **49% less peak RAM** than independent CTM+heads (Appendix G). Table 4 retains the original once-per-case comparison.

Overall, CSTP combines strong vehicle-count accuracy and competitive speed and flow prediction with limited closure training data. Its fast inference and low memory use further support practical, repeated what-if comparisons of road-closure plans.

6 Ablation study

To explain these results, we remove components individually and jointly on the same closures and two-hour horizon as Section 5. Table 5 links each removal to its equation and reports core/eight-hop errors. Count error evaluates the inventory prediction directly, while speed and flow errors show how the state and heads affect their predictions. The final row combines all removals. Appendix D specifies fitting and controls.

Table 5: **Individual and joint component ablations.** MAE, mean ± sample SD across the benchmark's road groups. Rows identify the component removed and its Section 3 equation; bold marks full CSTP. Count is in vehicles, speed in m/s, and flow in vehicles/min. Dashes mark the count-only calibration control.

	Core network			**8-hop neighborhoods**		
Removed component	Count	Speed	Flow	Count	Speed	Flow
None (full CSTP)	**4.26±0.26**	**3.02±0.08**	**15.52±0.29**	**10.16±6.87**	**2.21±0.90**	**9.32±3.90**
Queue storage (3)	5.69±0.98	3.25±0.11	15.93±0.14	28.55±11.27	5.13±1.40	13.77±3.45
Learned wave law (4)	4.30±0.23	3.01±0.08	15.52±0.30	10.56±6.75	2.21±0.92	9.23±4.00
Traversal delay (4)	4.49±0.44	3.03±0.08	15.50±0.30	10.61±6.71	2.21±0.91	9.15±3.94
Inventory depletion (5)	4.30±0.28	3.01±0.08	15.59±0.28	10.46±6.78	2.20±0.89	9.54±4.10
Junction interactions (6–7)	4.28±0.27	3.01±0.07	15.60±0.27	10.52±6.82	2.26±0.93	9.62±4.14
Closure calibration (Sec. 3.3)	5.00±0.68	—	—	17.89±7.61	—	—
Physical-state features (8)	4.26±0.26	3.06±0.08	15.59±0.26	10.16±6.87	2.76±0.83	10.15±3.51
Learned output heads (8)	4.26±0.26	3.35±0.17	16.09±0.21	10.16±6.87	7.26±2.55	16.09±4.34
All components	5.70±0.98	3.35±0.17	16.09±0.21	28.67±11.29	7.26±2.55	16.09±4.34

Queue storage supplies the main count gain. Removing storage raises core count MAE/RMSE by **33.6%/73.2%** and speed/flow MAE by $7.8\%/2.7\%$; all three MAEs worsen in every road group. Against joint removal, full CSTP lowers count MAE by **25.3% across the core** and **64.6% locally**.

Propagation, depletion, and junctions provide smaller count gains. Removing traversal delay raises core count MAE by 5.3%; replacing the learned wave law with a constant raises it by 1.0%. Removing inventory depletion raises local count/flow MAE by $2.9\%/2.4\%$. Similarly, junction removal raises local count, speed, and flow MAE by **3.6%, 2.3%, and 3.2%**, driven here by shared-lane blocking. These count contributions are smaller than storage's, and individual removals can improve speed or flow. The separate tests in Section 4 isolate wave timing and merge release.

Physical states support the speed/flow extensions. Removing physical-state features from ϕ and refitting h_V, h_Q raises local speed/flow MAE by **24.8%/9.0%**; core increases are $1.3\%/0.4\%$, with losses in every road group. Removing the heads instead restores normal speed/flow, raising core MAE by $11.2\%/3.6\%$ with physical counts retained. Removing either the state features or the learned heads increases prediction error.

Separating normal learning from closure calibration. Fixing both physical response multipliers to one isolates the effect of calibration: core/local count MAE is 5.00/17.89, still **12.4%/37.6% below** the normal reference. Count predictions therefore improve on the normal reference even without fitting these multipliers; calibration reduces error further. This control uses the same historical closure data as the main model (Appendix C.2) but omits the speed and flow heads.

7 Conclusion

CSTP uses network structure and causal traffic mechanisms to predict road-closure effects from aggregate normal traffic, without OD or route data. Experiments show strong vehicle-count accuracy with limited closure supervision, alongside competitive speed and flow prediction. Its fast inference and low memory use further support repeated what-if comparisons, making the framework practical for predicting traffic under road-closure scenarios. Our primary contribution is predictive accuracy; causal interpretation is conditional on the specified mechanisms and matched background conditions.

AI USE STATEMENT

AI assisted with refining the mechanistic formulation; literature searching; revising the manuscript; and preparing figures, tables, and references. Numerical results were computed by the recorded programs from the simulation and observed-traffic data. Recorded verification checks cover input and model hashes, fitting-group membership, reproduction of archived scores, prediction coverage, and independent metric calculations. The authors remain responsible for the final text, claims, code, and results.

REPRODUCIBILITY STATEMENT

The appendices specify the model variants, interventions, fitting and selection procedures, scoring masks, aggregation, and timing protocol. The accompanying experiment record supports reproduction with code, frozen protocols, configuration and prediction hashes, fitted controls, and case-level metrics. It distinguishes current benchmark results from controlled mechanism tests and historical development records.

REFERENCES

Matthew Barth and Kanok Boriboonsomsin. Real-world carbon dioxide impacts of traffic congestion. *Transportation Research Record*, 2058(1):163–171, 2008. doi: 10.3141/2058-20. URL https://journals.sagepub.com/doi/10.3141/2058-20.

Sander Beckers, Frederick Eberhardt, and Joseph Y. Halpern. Approximate causal abstractions. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 115 of *Proceedings of Machine Learning Research*, pp. 606–615, 2020. URL https://proceedings.mlr.press/v115/beckers20a.html.

Moshe Ben-Akiva, Michel Bierlaire, Didier Burton, Haris N. Koutsopoulos, and Rabi Mishalani. Network state estimation and prediction for real-time traffic management. *Networks and Spatial Economics*, 1:293–318, 2001. doi: 10.1023/A:1012883811652. URL https://link.springer.com/article/10.1023/A:1012883811652.

Ioana Bica, Ahmed M. Alaa, James Jordon, and Mihaela van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/2002.04083.

Caltrans. Performance measurement system (PeMS) data source. California Department of Transportation, n.d. URL https://dot.ca.gov/programs/traffic-operations/mpr/pems-source. Accessed 20 September 2026.

Carlos F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4): 269–287, 1994. doi: 10.1016/0191-2615(94)90002-7.

Carlos F. Daganzo. The cell transmission model, part II: Network traffic. *Transportation Research Part B: Methodological*, 29(2):79–93, 1995. doi: 10.1016/0191-2615(94)00022-R.

Lixiang Fan, Bohao Li, Tao Zou, Junchen Ye, and Bowen Du. Incident-guided spatiotemporal traffic forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 243–254, 2026. doi: 10.1145/3770854.3780215.

Nikolas Geroliminis and Carlos F. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9): 759–770, 2008. doi: 10.1016/j.trb.2008.02.002. URL https://doi.org/10.1016/j.trb.2008.02.002.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Andreas Hegyi, Bart De Schutter, and Hans Hellendoorn. Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies*, 13(3):185–209, 2005. doi: 10.1016/j.trc.2004.08.001. URL https://www.dcsc.tudelft.nl/~bdeschutter/pub/rep/03_016.pdf.

Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang. PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 4365–4373, 2023. doi: 10.1609/aaai.v37i4.25556. URL https://ojs.aaai.org/index.php/AAAI/article/view/25556.

Lawrence A. Klein, Milton K. Mills, and David R. P. Gibson. Traffic detector handbook: Third edition—volume I. Technical Report FHWA-HRT-06-108, Federal Highway Administration, October 2006. URL https://www.fhwa.dot.gov/publications/research/operations/its/06108/06108.pdf.

Shuang Li, Ziyuan Pu, Nan Zhang, Duxin Chen, Lu Dong, Daniel J. Graham, and Yinhai Wang. MSCT: Addressing time-varying confounding with marginal structural causal transformer for counterfactual post-crash traffic prediction. *arXiv preprint arXiv:2407.14065*, 2024. URL https://arxiv.org/abs/2407.14065.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1707.01926.

Michael James Lighthill and Gerald Beresford Whitham. On kinematic waves. II. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178):317–345, 1955. doi: 10.1098/rspa.1955.0089. URL https://doi.org/10.1098/rspa.1955.0089.

Hangchen Liu, Zheng Dong, Renhe Jiang, Jiewen Deng, Jinliang Deng, Quanjun Chen, and Xuan Song. Spatio-temporal adaptive embedding makes vanilla transformer SOTA for traffic forecasting. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 4125–4129, 2023. doi: 10.1145/3583780.3615160. URL https://arxiv.org/abs/2308.10425.

Tanwi Mallick, Prasanna Balaprakash, Eric Rask, and Jane Macfarlane. Graph-partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting. *Transportation Research Record*, 2674(9):473–488, 2020. doi: 10.1177/0361198120930010. URL https://arxiv.org/abs/1909.11197.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In *Proceedings of the International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 15293–15329, 2022. URL https://proceedings.mlr.press/v162/melnychuk22a.html.

G. F. Newell. A simplified theory of kinematic waves in highway traffic, part I: General theory. *Transportation Research Part B: Methodological*, 27(4):281–287, 1993. doi: 10.1016/0191-2615(93)90038-C.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009. doi: 10.1017/CBO9780511803161. URL https://doi.org/10.1017/CBO9780511803161.

Paul I. Richards. Shock waves on the highway. *Operations Research*, 4(1):42–51, 1956. doi: 10.1287/opre.4.1.42. URL https://doi.org/10.1287/opre.4.1.42.

Param Sankar, Krista Jeannotte, Jeffrey P. Arch, Marcelino Romero, and James E. Bryden. Work zone impacts assessment: An approach to assess and manage work zone safety and mobility impacts of road projects. Technical Report FHWA-HOP-05-068, Federal Highway Administration, Office of Operations, August 2006. URL https://ops.fhwa.dot.gov/wz/resources/final_rule/wzi_guide/index.htm.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Towards causal representation learning. *arXiv preprint arXiv:2102.11107*, 2021. URL https://arxiv.org/abs/2102.11107.

David Schrank, Luke Albert, Kartikeya Jha, and Bill Eisele. 2025 urban mobility report. Technical report, Texas A&M Transportation Institute, August 2025. URL https://tti.tamu.edu/documents/mobility-report-2025.pdf.

Rongye Shi, Zhaobin Mo, and Xuan Di. Physics-informed deep learning for traffic state estimation: A hybrid paradigm informed by second-order traffic models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 540–547, 2021. doi: 10.1609/aaai.v35i1.16132. URL https://ojs.aaai.org/index.php/AAAI/article/view/16132.

Chris M. J. Tampère, Ruben Corthout, Dirk Cattrysse, and Lambertus H. Immers. A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, 45(1):289–309, 2011. doi: 10.1016/j.trb.2010.06.004.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph WaveNet for deep spatial-temporal graph modeling. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2019. URL https://arxiv.org/abs/1906.00121.

Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://arxiv.org/abs/2107.00793.

Ming Xu, Qiang Ai, Ruimin Li, Yunyi Ma, Geqi Qi, Xiangfu Meng, and Haibo Jin. TraffNet: Learning causality of traffic generation for what-if prediction. *arXiv preprint arXiv:2303.15954*, 2023. URL https://arxiv.org/abs/2303.15954.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2018. doi: 10.24963/ijcai.2018/505. URL https://arxiv.org/abs/1709.04875.

A IMPLEMENTATION OF THE MECHANISM STATE

These implementation rules complete Equations (3)–(8) for the benchmark. Appendix E specifies controlled variants.

Queue storage and movement relief. Storage is $K_e = \max\{0.13L_e\ell_e, 1.05N^0_{e,0.99}, 1\}$, using normal 99th-percentile inventory $N^0_{e,0.99}$. Movement shares scale with downstream mean normal count divided by free-flow travel time. Smoothed, saturation-capped count/travel-time proxies supply throughput $\bar{q}^0$; they are distinct from measured entering flow. In Equation (3),

$$d_{e,k} = \eta_k[\bar{q}^0_{e,k} b_{e,k} - a_{e,k}]_+, \quad r_{e,k} = r^{\text{base}}_{e,k} + r^{\text{alt}}_{e,k}, \quad \eta_k = \min\left\{1, \frac{C^{\text{cap}}_k}{\sum_e [\bar{q}^0_{e,k} b_{e,k} - a_{e,k}]_+}\right\}. \tag{9}$$

The factor is one for zero denominator; $r^{\text{base}} = 0$. Alternatives absorb new blocked demand a and release existing queues through r^{alt}, subject to spare capacity. Relief is represented in aggregate; receiver-specific diverted inflows and overflow are not propagated. Farther-upstream blockage combines route-weighted gates one shortest-path order closer to the closure, with minimum movement share 0.005 and a cap of one.

Wave learning and update order. Wave-speed learning uses count-onset supervision, $\widetilde{w}_j = L_j/(t^{\text{on}}_i - t^{\text{on}}_j)$ when downstream onset precedes upstream onset. Road intercepts shrink toward road-type/lane-group priors; sparse roads use group/global fallbacks. Normal fitting uses 398,785 observations: 8,184 local fits, 3,309 group fallbacks, and 27 global fallbacks. During prediction, wave features use normal count plus the updated queue before deficit subtraction; occupancy is clipped to $[0, 1.5]$, and positive growth is in vehicles/minute. Waves are bounded to 0.05–15 m/s. Progress is active with at least three extra vehicles or at most three remaining storage spaces; road length is at least 1 m. Together, the preceding queue and progress set the gate:

$$g_{e,k} = \max\left\{g_{e,k-1}, \mathbf{1}[p_{e,k-1} \geq 1] \operatorname{clip}_{[0,1]}\left(\frac{x_{e,k-1}/A_{e,k} - 0.75}{0.95 - 0.75}\right)\right\}. \tag{10}$$

For $A \leq 10^{-6}$ vehicles, completed active traversal gives candidate gate one. Progress advances after the queue update and affects the next gate.

Inventory depletion and junctions. With $T_e = \max\{L_e/\max(v^{\lim}_e, 0.1), 2\}$ seconds and incoming roads $\mathcal{D}^{(C)}_e$ one order closer to C, the reservoir uses

$$m^{\text{out}}_{e,k} = \min\{z_{e,k-1}/T_e, \bar{q}^0_{e,k}\}, \quad m^{\text{out}}_{C,k} = \bar{q}^0_{C,k}, \quad m^{\text{in}}_{e,k} = \sum_{j \in \mathcal{D}^{(C)}_e} \pi_{je} m^{\text{out}}_{j,k}. \tag{11}$$

Deficits start at zero and propagate up to eight connection orders. All departure rates precede arrival and deficit updates. Shared-lane obstruction uses the previous queue with scale $q_* = 1$ vehicle and is rebuilt at each step. Merge candidates require one outgoing movement, positive normal discharge, and waiting stock above one vehicle. A source with normal count reaching $0.75K$ is guarded because external waiting stock is unobserved. Calibration uses normal counts during 06:45–07:00 and permitted lane connections. Movement estimates are proportionally capped by receiving throughput. Merge release is stock-bounded and paired with receiving-flow compensation; changing normal stock can reverse this correction within the normal-discharge bound. The final projections and readouts impose road-level bounds without a global conservation constraint. Closed-road counts retain initial occupants; closed-road speed and entering flow are zero.

Learned output heads. The speed and flow heads use targets $(V^{\text{obs}} - B^V)/v^*$ and $(Q^{\text{obs}} - Q^0)/\ell$, with absolute/squared-error loss, respectively. Each head has 100 boosting iterations, learning rate 0.07, at most 15 leaves, and minimum leaf size 120. Its 24 inputs cover normal traffic, road attributes, closure context, and the five physical-state features defined in Section 3.3; the source lists their order and normalization. Corrections apply after onset within eight directed upstream/downstream connections or when predicted count change exceeds 0.05 vehicles. Other roads retain normal references. Speed is capped at $\max(B^V, 1.5v^*)$; heads do not feed back into counts.

B SUMO EVALUATION DESIGN AND BASELINE ADAPTATIONS

Shared task and inputs. The benchmark contains 16 stopped-road closures. A topology-selected 1,969-road input subnetwork contains a fixed 765-road eight-hop evaluation core and 1,204 sur-

rounding buffer roads. All methods receive the same graph, normal (no-closure) count/speed/flow horizon, and closure query; prediction uses no target-closure observations. Local evaluation uses undirected eight-hop neighborhoods. Shared finite-target masks cover 1,440 five-second endpoints, 07:00:05–09:00. Speed/flow labels use an assumed +5 s interval-start offset; Appendix G reports alignment sensitivity. Appendix H extends the comparison to ten non-critical closures in a separate region.

Grouped fitting and model selection. Four-fold grouped cross-validation excludes each test road group from parameter fitting and within-fold model selection. Each fold tests four cases and fits the other 12 (seed 20260921). Subdivisions of the same base road stay together, giving 13 groups. Two training groups select neural epochs and residual strengths before refitting on all outer-training cases. Earlier component development also used this archive.

CSTP selects wave multiplier $\{1, 1.75, 2.5\}$ and demand-cap multiplier $\{10, 25, 50\}$ by training-group core count MAE. CTM selects wave speed $\{2.5, 5, 7.5\}$ m/s and capacity multiplier $\{0.8, 1, 1.2\}$ using equally weighted, reference-normalized output MAEs. Normal-derived profiles, turns, sources, and graph/path proxies remain fixed. Final CSTP adds junction rules calibrated from normal traffic to the fixed grouped fits; forecasts and ablations are then recomputed without retuning. Table 6 lists the selected parameters and refit epochs.

Table 6: Fold-specific calibration and validation-selected neural refit epochs. CSTP wave/cap and CTM capacity are multipliers; MSCT E/D is encoder/decoder epochs.

	CSTP		CTM		Refit epochs		
Fold	Wave	Cap	Wave (m/s)	Capacity	IGSTGNN	TraffNet	MSCT E/D
1	2.5	25	2.5	1.2	2	2	16/10
2	1.75	25	2.5	0.8	2	6	10/10
3	1.75	25	2.5	1.2	2	2	16/8
4	1.75	50	2.5	1.2	2	22	16/8

CTM. The reference-anchored solver uses 5,390 cells (at most 100 m), a CFL-compliant 0.02907 s step, triangular demand/supply, fixed turns, proportional merges, FIFO diverges, and source queues. Normal profiles supply speed, storage, capacity, turning, and demand proxies. Closure zeros sending/receiving capacities. To match the reference, anchoring adds the closure-minus-open simulation change to the supplied normal reference and clips below at zero. Appendix C adds matched learned readouts.

Neural adaptations. IGSTGNN uses the released five-layer, 32-unit backbone to predict twelve five-second residual steps from normal covariates, attributes, and closure inputs. TraffNet retains path GRUs, OD attention, route choice, demand-weighted aggregation, and temporal prediction (commit `18eccf9`), with reduced CPU widths/scoring depth. Its 2,183 synthetic paths cover all roads; 171 origin bundles have alternatives. Each bundle has at most three shortest travel-time paths, destinations six hops away where possible, and at most 24 roads per path. Closure-containing paths are disabled and feasible probabilities renormalized; normal flow supplies demand without path-volume pretraining. Paper-based MSCT uses a two-layer, 32-unit, two-head transformer, dropout 0.1, and an LSTM treatment pathway. Normal-reference controls support domain-confusion loss (weight 0.1); encoder weights freeze during decoder fitting.

Neural weights are initialized anew in each fold. TraffNet/MSCT use twelve normal-history samples and no-closure reference values at prediction times; MSCT predicts one encoder step and eleven feedback steps. Inner validation selects the refit epochs in Table 6, using each model's archived optimizer, sampling procedure, epoch ceiling, and early-stopping rule. IGSTGNN uses CPU bfloat16; others use float32. Residual strengths $\{0, 0.25, 0.5, 1\}$ minimize count/speed MAE and flow RMSE. The benchmark compares aggregate-input adaptations on the shared task. The independent CTM+heads comparator and matched-mask control in Appendix C further assess the physical-state representation with matched head architectures.

B.1 CASE COVERAGE, AGGREGATION, AND BROADER SCORING

All 16 cases use common finite road–time masks. Case MAEs and RMSEs are averaged within road groups, then equally across the 13 groups. Sample SD describes road-group variation; each fit uses one initialization.

Table 7: Core count MAE for every out-of-fold closure (vehicles/road). The entire test road group is excluded from fitting and inner selection. F denotes fold; IGST denotes IGSTGNN. Bold marks the lowest error; Ref. is normal traffic.

F	Test closure	CSTP	CTM	IGST	Ref.	F	Test closure	CSTP	CTM	IGST	Ref.
1	171574775	4.550	**4.344**	5.748	5.821	3	1152816905	**4.057**	4.314	4.600	4.554
1	435409944#3	**4.587**	5.033	6.480	6.563	3	171574786#1	**4.068**	4.362	6.328	6.304
1	435409944#4	**4.454**	4.850	6.401	6.488	3	755555177	**4.521**	4.829	5.306	5.253
1	48378603#3	**4.170**	4.710	5.066	5.143	3	833392258#6	**4.089**	4.419	5.385	5.346
2	127067570#6	**3.973**	4.971	4.101	4.121	4	1069206791#1	**4.068**	4.304	6.431	6.468
2	50224247#3	**3.958**	4.533	5.901	5.945	4	288587879	**4.754**	5.633	7.785	7.825
2	50224247#4	**3.967**	4.685	5.960	6.014	4	44685053#1	**4.170**	5.217	5.789	5.873
2	639674357#4	**4.311**	4.669	4.828	4.849	4	44685053#2	**4.510**	5.114	6.292	6.335

Scoring beyond the core. Scoring all 1,969 input roads, including the 1,204 buffers, gives count MAE 3.72 ± 0.29 for CSTP, 4.02 ± 0.39 for CTM, and 4.26 ± 0.57 for IGSTGNN: reductions of 7.4% and 12.7%, respectively. Table 8 gives the complete original-model comparison; local and core scopes remain primary.

Table 8: Original benchmark models on all 1,969 input roads: group mean $\pm$ sample SD. Flow is vehicles/minute/road; count is vehicles/road and speed m/s. Bold marks the lowest mean.

	Count		Speed		Flow	
Method	MAE	RMSE	MAE	RMSE	MAE	RMSE
Reference	4.32 ± 0.58	17.36 ± 4.46	3.79 ± 0.15	5.79 ± 0.22	17.10 ± 0.14	26.79 ± 0.19
CTM	4.02 ± 0.39	13.00 ± 3.30	3.60 ± 0.07	5.56 ± 0.11	16.87 ± 0.18	26.52 ± 0.17
IGSTGNN	4.26 ± 0.57	17.35 ± 4.45	3.66 ± 0.14	5.57 ± 0.23	16.37 ± 0.17	25.78 ± 0.17
TraffNet	—	—	—	—	$\mathbf{16.30 \pm 0.10}$	$\mathbf{25.50 \pm 0.16}$
MSCT	—	—	$\mathbf{3.32 \pm 0.09}$	$\mathbf{4.64 \pm 0.17}$	—	—
CSTP	$\mathbf{3.72 \pm 0.29}$	$\mathbf{11.41 \pm 3.57}$	3.55 ± 0.05	5.47 ± 0.09	16.75 ± 0.22	26.36 ± 0.23

Consistency across road groups. Table 9 compares final CSTP with independent CTM+heads. Percentile intervals resample the 13 paired road groups 20,000 times (seed 20260925), keeping each group's cases and the fitted models fixed. Local count error improves in every group; omitting any one group gives a 27.4–32.5% reduction. The table reports output-specific intervals and win counts for every output and scoring area.

Table 9: Paired MAE reductions relative to independent CTM+heads. Positive values favor CSTP. Intervals are conditional road-group bootstrap summaries.

Scope	Output	MAE reduction (%)	95% interval (%)	Group wins
Eight-hop	Count	28.79	$[21.85, 36.80]$	13/13
Eight-hop	Speed	8.78	$[-2.52, 20.69]$	8/13
Eight-hop	Flow	1.38	$[-1.43, 4.53]$	9/13
Core	Count	9.61	$[6.20, 12.76]$	12/13
Core	Speed	-0.50	$[-1.55, 0.59]$	5/13
Core	Flow	-0.54	$[-1.07, -0.11]$	4/13

Temporal summaries. Figure 5 averages five-minute errors within cases, then within road groups (SUMO) or incident dates (PeMS). SUMO bins use midpoints; PeMS uses interval ends.

C MODEL CONTROLS AND REFERENCE ROBUSTNESS

Independent CTM+heads comparator. CTM+heads adds learned speed and flow models while retaining CTM counts. Each fold uses its selected CTM parameters and the same 24-feature, 100-tree architecture, losses, normalization, inner groups, and strength grid as CSTP. CTM-derived counts,

occupancy ratios, and activation replace the corresponding CSTP features; activation begins at three excess vehicles. Training samples up to 80,000 positions per case from CTM-only eligibility with a deterministic seed. Fitting and selection exclude test road groups. Both fitting and inference use CTM states only, and Appendix G times this complete predictor.

Matched-mask representation control. To compare the information supplied by the two physical states, we refit both sets of heads on identical samples from the union of their input-defined masks (suffix U). Finite-target masks, head settings, and within-fold strength selection are shared. This control uses both states to define the sampling mask; Table 1 instead uses the independently executable CTM+heads comparator. Final CSTP, its ablations, and PeMS retain their original fitted heads.

Table 10: Matched-mask refits (U): mean ± sample SD across road groups. Both arms refit heads on the same rows; CSTP here is the representation control, distinct from final CSTP in Table 1. Count predictions are unchanged.

		Speed (m/s)		Flow (veh/min/road)	
Scope	State	MAE	RMSE	MAE	RMSE
Eight-hop	CTM+heads-U	2.39 ± 0.72	4.52 ± 1.03	9.50 ± 3.74	15.35 ± 5.22
Eight-hop	CSTP (refit, U)	2.05 ± 0.69	3.89 ± 0.84	9.31 ± 3.73	15.30 ± 5.13
Core	CTM+heads-U	3.00 ± 0.07	4.89 ± 0.09	15.42 ± 0.31	23.69 ± 0.41
Core	CSTP (refit, U)	2.97 ± 0.06	4.85 ± 0.10	15.41 ± 0.32	23.69 ± 0.40

With training rows and readouts matched, CSTP gives 14.2%/2.0% lower local speed/flow MAE; counts are independent of heads. Table 11 maps the experimental configurations.

Table 11: Configurations used in each evaluation. Controlled studies isolate mechanisms; Sections 5–6 evaluate the complete predictor.

Study	Boundary and movement	State configuration and evidence
Final SUMO benchmark	Stopped road; aggregate alternative-movement relief	Count-onset wave learning, reservoir depletion, both junction rules, frozen speed/flow heads; joint accuracy and component removals.
Corridor magnitude	Entry-only; fixed routes	Wave delay off to isolate storage; finite residence with initial drainage tests downstream depletion.
Junction magnitude	Entry-only; fixed routes	Count-onset wave fits and delay retained; shared-lane and merge rules test cross-route count effects.
Wave confirmation	Entry-only; fixed routes	Direct queue-position supervision; features frozen during traversal; arrival and traversal-time tests.
PeMS transfer	Supplied stopped-road query; benchmark movement rules	Frozen SUMO response/heads, normal-only wave and junction calibration; observed speed/flow prediction.

C.1 Normal-reference robustness

To assess sensitivity to the supplied normal reference, we perturb its timing, speed, and count/flow values while holding fitted predictors, cases, scoring masks, and road-group weights fixed.

Timing and speed perturbations. We compare final CSTP, CTM+heads-U, and the normal reference under five settings: unchanged inputs, a 15-minute advance/delay of all three outputs, and $\pm 10\%$ normal speed. A shift s replaces each value at t by its value at $\mathrm{clip}_{[05:00,09:00]}(t+s)$; endpoints are held constant. Positive s advances the pattern. Speed arms multiply finite values and preserve missing locations. Transformations also enter normal-only movement and junction calibration. Closure parameters, heads, road profiles, and wave estimates remain fixed. CTM recomputes normal-only sources/turns after temporal shifts; speed arms reuse identical count-derived states. Both readouts and their input-defined gates use the perturbed reference.

Under these perturbations, CSTP retains lower local MAE than CTM+heads-U in every perturbed setting: count gains are 27.0–28.9%, speed gains 6.1–10.8%, and flow gains 1.4–3.1%. Temporal shifts change CSTP's absolute errors, with increases of up to 9.6% locally and 16.3% in core count MAE; the local comparative advantage is retained.

Coherent count/flow bias. The count/flow test scales normal inputs (including junction calibration) by 0.9 or 1.1. Speed, response parameters, heads, profiles, and waves stay fixed. Local CSTP MAE changes by at most 2.6%; the upward bias increases core count and flow MAE by 8.2% and 5.5%. Table 13 retains the complete results.

Table 12: Reference-error stress test: group-mean MAE ± sample SD. Count is vehicles/road, speed m/s, and flow vehicles/minute/road. All methods receive identical perturbed inputs. Bold marks the lowest mean in each setting.

		765-road core			Eight-hop neighborhood		
Reference input	Method	Count	Speed	Flow	Count	Speed	Flow
Unchanged	Reference	5.71 ± 0.98	3.38 ± 0.17	16.13 ± 0.22	28.85 ± 11.36	7.71 ± 2.76	16.78 ± 4.63
	CTM+heads-U	4.71 ± 0.39	$\mathbf{3.00 \pm 0.07}$	$\mathbf{15.42 \pm 0.31}$	14.27 ± 7.21	2.39 ± 0.72	9.50 ± 3.74
	CSTP	$\mathbf{4.26 \pm 0.26}$	3.02 ± 0.08	15.52 ± 0.29	$\mathbf{10.16 \pm 6.87}$	$\mathbf{2.21 \pm 0.90}$	$\mathbf{9.32 \pm 3.90}$
Advance 15 min	Reference	6.39 ± 0.94	3.66 ± 0.17	17.46 ± 0.22	29.22 ± 11.30	7.89 ± 2.74	17.75 ± 4.98
	CTM+heads-U	5.40 ± 0.37	$\mathbf{3.29 \pm 0.08}$	$\mathbf{16.77 \pm 0.37}$	15.15 ± 7.43	2.66 ± 0.80	10.30 ± 3.86
	CSTP	$\mathbf{4.96 \pm 0.25}$	3.29 ± 0.08	16.80 ± 0.35	$\mathbf{10.77 \pm 7.05}$	$\mathbf{2.37 \pm 0.91}$	$\mathbf{9.99 \pm 3.88}$
Delay 15 min	Reference	6.28 ± 0.94	3.70 ± 0.16	17.33 ± 0.19	29.18 ± 11.42	7.94 ± 2.73	17.53 ± 4.68
	CTM+heads-U	5.21 ± 0.36	$\mathbf{3.31 \pm 0.07}$	$\mathbf{16.60 \pm 0.39}$	14.47 ± 7.23	2.64 ± 0.70	10.21 ± 3.71
	CSTP	$\mathbf{4.79 \pm 0.20}$	3.32 ± 0.08	16.70 ± 0.35	$\mathbf{10.57 \pm 6.91}$	$\mathbf{2.43 \pm 0.91}$	$\mathbf{10.07 \pm 3.89}$
Speed -10%	Reference	5.71 ± 0.98	3.62 ± 0.16	16.13 ± 0.22	28.85 ± 11.36	7.43 ± 2.48	16.78 ± 4.63
	CTM+heads-U	4.71 ± 0.39	$\mathbf{3.25 \pm 0.06}$	$\mathbf{15.42 \pm 0.31}$	14.27 ± 7.21	2.50 ± 0.75	9.50 ± 3.74
	CSTP	$\mathbf{4.26 \pm 0.26}$	3.28 ± 0.08	15.52 ± 0.29	$\mathbf{10.16 \pm 6.87}$	$\mathbf{2.34 \pm 0.95}$	$\mathbf{9.31 \pm 3.90}$
Speed $+10\%$	Reference	5.71 ± 0.98	3.89 ± 0.18	16.13 ± 0.22	28.85 ± 11.36	8.42 ± 2.85	16.78 ± 4.63
	CTM+heads-U	4.71 ± 0.39	$\mathbf{3.44 \pm 0.07}$	$\mathbf{15.42 \pm 0.31}$	14.27 ± 7.21	2.58 ± 0.75	9.50 ± 3.74
	CSTP	$\mathbf{4.26 \pm 0.26}$	3.47 ± 0.08	15.52 ± 0.29	$\mathbf{10.16 \pm 6.87}$	$\mathbf{2.42 \pm 0.93}$	$\mathbf{9.32 \pm 3.90}$

Table 13: CSTP under coherent count/flow bias. MAE, group mean ± sample SD; units match Table 12. A factor of 1.0 is unchanged.

	765-road core			Eight-hop neighborhood		
Count/flow factor	Count	Speed	Flow	Count	Speed	Flow
0.9	4.21 ± 0.18	3.02 ± 0.07	15.19 ± 0.29	10.34 ± 6.93	2.27 ± 0.91	9.19 ± 3.89
1.0	4.26 ± 0.26	3.02 ± 0.08	15.52 ± 0.29	10.16 ± 6.87	2.21 ± 0.90	9.32 ± 3.90
1.1	4.61 ± 0.32	3.02 ± 0.08	16.37 ± 0.32	10.42 ± 7.07	2.21 ± 0.90	9.38 ± 4.07

No model is retuned or perturbation setting selected using the perturbed test errors.

C.2 Closure-data efficiency and response-multiplier calibration

We compare closure-data requirements on the same held-out cases using identical sampled count supervision. Following a two-budget pilot, the protocol was fixed for the expanded comparison; outer test road groups remain excluded from fitting.

Shared supervision and selection. Two/four-group budgets use three nested, outcome-blind hash rankings (seeds 20260925–20260927); all-data fits use nine/ten groups and are shared across schedules. Within budgets, methods share 96 times per case, one per 75-second interval, on all core roads except the closed road. Existing nine-candidate physical grids select count MAE, averaging cases within groups then groups equally. CTM here selects count alone; Table 1 uses its multi-output selection. At the full training budget, methods use the same sampled labels, whereas the headline CSTP count model retains its dense calibration. The selected CTM settings give identical scored counts across budgets.

Count-trained controls. The direct and residual histogram-boosted tree controls use absolute error, 200 iterations, 31 leaves, minimum leaf size 40, rate 0.07, and $L_2 = 10$. Their 25 shared inputs describe normal traffic, geometry, closure context, and cumulative normal inflows, without CSTP states. A directed GraphSAGE-style residual control (Hamilton et al., 2017) adds two 64-wide layers with separate upstream/downstream means, LayerNorm, residual connections, and dropout 0.05 on the full graph, without road-ID embeddings. This is the "Graph neural" row in the tables, a separate count-trained control from IGSTGNN.

The neural control uses AdamW with rates $\{10^{-3}, 3\times10^{-4}\}$, decay 10^{-5}, clipping 5. Epochs update once/group using eight times, rotating subdivisions and cycling through each case's samples. Inner training groups select rate, epoch, and strength $\{0, 0.25, 0.5, 1\}$ before all-group refitting: validation

every six epochs, initial maximum 120, patience six checks after epoch 36. Fourteen settings selecting epoch ≥ 108 were extended to 240 before test scoring; three final selections use epoch 240. Initialization is fixed, and all outputs use the same nonnegativity and stopped-road bounds.

Table 14: Data-budget results: count MAE/RMSE, mean across road groups after averaging subset schedules. Brackets give the range of schedule-level MAEs, not confidence intervals. Full training uses one fit per outer fold.

		Core network			8-hop neighborhoods		
Groups	Method	MAE	RMSE	MAE range	MAE	RMSE	MAE range
2	CSTP	**4.27**	14.92	[4.24, 4.30]	**10.45**	37.79	[9.99, 10.74]
2	CTM	4.71	16.57	[4.71, 4.71]	14.27	43.76	[14.27, 14.27]
2	Graph neural	5.46	24.99	[5.44, 5.51]	27.19	76.99	[25.59, 28.10]
2	Direct trees	4.94	21.83	[4.92, 4.95]	22.30	65.02	[21.46, 22.76]
2	Residual trees	4.87	19.82	[4.82, 4.92]	20.77	58.07	[19.84, 21.31]
2	Normal reference	5.70	25.66	[5.70, 5.70]	28.67	79.81	[28.67, 28.67]
4	CSTP	**4.25**	14.65	[4.25, 4.26]	**10.07**	35.86	[10.01, 10.12]
4	CTM	4.71	16.57	[4.71, 4.71]	14.27	43.76	[14.27, 14.27]
4	Graph neural	5.14	21.29	[5.06, 5.27]	21.05	60.92	[18.41, 22.40]
4	Direct trees	4.81	20.96	[4.78, 4.85]	20.56	60.16	[19.96, 21.05]
4	Residual trees	4.61	18.24	[4.55, 4.65]	18.60	52.54	[18.00, 19.36]
4	Normal reference	5.70	25.66	[5.70, 5.70]	28.67	79.81	[28.67, 28.67]
All	CSTP	**4.21**	14.21	—	**10.07**	36.51	—
All	CTM	4.71	16.57	—	14.27	43.76	—
All	Graph neural	4.84	19.32	—	17.52	51.27	—
All	Direct trees	4.66	20.29	—	19.77	58.02	—
All	Residual trees	4.37	17.57	—	17.12	49.93	—
All	Normal reference	5.70	25.66	—	28.67	79.81	—

Without response-multiplier calibration. Setting both response multipliers to one leaves normal-derived waves unscaled and uses normal closure-path demand as the cap. The same CSTP mechanism rules and normal-derived junction calibration remain, and speed/flow heads are omitted. This count-only control isolates response calibration within the model described in Appendix B. All 560 fitted-model forecasts and 16 fixed-multiplier forecasts are saved before scoring.

D COMPLETE BENCHMARK ABLATIONS

Component ablations retain all 16 cases, masks, group weights, and boundaries from Appendix B. Physical removals keep parameters and heads fixed, passing altered states through the same feature and eligibility rules. The resulting error changes measure each component's role in the fitted predictor, including its effect on the heads. Table 15 adds RMSEs to the main-text MAEs.

Exact component removals. Queue-storage removal sets blocked-flow retention to zero. Learned-wave removal uses an unscaled 1 m/s constant and retains the fold multiplier; traversal removal bypasses delay while keeping storage gating. Inventory-depletion removal zeros deficits and their count subtraction, retaining the closure's missing-flow source. Junction removal disables shared-lane blockage and merge release. Output-head removal restores normal speed/flow on open roads. The joint removal combines all physical removals and removes both heads; counts retain the normal reference, projection, and closed-road convention. No physical-state features are evaluated when heads are absent.

Refitted physical-state feature control. To test the information supplied by the physical state, we delete the five features defined in Section 3.3. Grouped refits retain sampled rows, losses, 100 iterations, and inner groups; strengths $\{0, 0.25, 0.5, 1\}$ select one in every fold. Both full and reduced heads were fitted before junction activation and receive final-model states and identical eligibility at inference.

Junction responses in the benchmark. Every case contains eligible shared-lane movements (4–51 directly connected core branches). Shared-branch count MAE falls from 2.43 to 2.21 (9.0%), improving 12 of 13 groups. Relative to removing both junction rules, local count/speed/flow MAE falls by 3.43%/2.21%/3.14%. These benchmark gains come from shared-lane obstruction: the merge

Table 15: Complete core/eight-hop RMSEs (group mean ± sample SD). Units match the main ablation table: count in vehicles, speed in m/s, flow in vehicles/minute. The first row is full CSTP.

	Core			Eight-hop		
Removed component	Count	Speed	Flow	Count	Speed	Flow
None (full CSTP)	14.81±5.24	4.93±0.12	23.93±0.31	36.70±36.86	4.32±1.06	15.81±5.39
Queue storage	25.65±7.01	5.25±0.19	24.26±0.19	79.78±35.12	7.51±1.83	19.00±5.10
Learned wave law	15.24±5.25	4.93±0.11	23.93±0.31	38.29±36.65	4.33±1.11	15.73±5.41
Traversal delay	16.48±6.06	4.95±0.12	23.92±0.30	37.76±36.51	4.33±1.08	15.74±5.41
Inventory depletion	14.86±5.22	4.93±0.10	24.04±0.28	36.95±36.74	4.31±1.05	16.22±5.78
Junction interactions	14.98±5.02	4.94±0.10	24.05±0.28	38.47±36.80	4.39±1.11	16.25±5.79
Response multipliers (Sec. 3.3)	21.02±6.39	—	—	58.24±34.76	—	—
Physical-state features	14.81±5.24	4.99±0.12	23.98±0.30	36.70±36.86	5.07±1.14	16.30±5.30
Learned output heads	14.81±5.24	5.45±0.29	24.68±0.31	36.70±36.86	9.74±2.51	25.68±6.68
All components	25.66±7.01	5.45±0.29	24.68±0.31	79.81±35.11	9.74±2.51	25.68±6.68

term is inactive under the normal-data waiting-stock screen. Appendix E evaluates merge competition in controlled layouts.

E Controlled mechanism and wave experiments

Controlled tests isolate storage, depletion, and junction responses; Appendix D evaluates the complete predictor. Paired SUMO runs use identical departures and fixed routes. Entry into C is blocked while initial occupants drain, and fixed routes disable alternative-movement relief. Prediction receives the paired normal horizon, so open-road count error equals count-effect error. Corridor tests disable wave delay to isolate storage and use finite residence with initial drainage for depletion, ending at active merges. Junction tests retain count-onset fits and delay. Table 11 maps each experiment to its boundary conditions and model configuration.

The magnitude cohorts contain 50 corridors, 25 shared-lane cases, and 25 normal-screened merges with seeded departures. Scores weight cases equally within layouts/windows. Deficit removal also disables finite residence; the closure source and separate merge-release stock remain active.

Table 16: Count MAE (vehicles), case mean ± sample SD. In (a), upstream, downstream, and junction windows are 30, 2, and 15 min. Panel (b) excludes isolated road Z. Rows name removals; the first row is the reference.

(a) Mechanism-specific roads/windows

Removed component	$U1$–$U3$	$D1$–$D3$	D	X	E/F
None (reference)	0.82±0.34	0.04±0.01	0.64±0.14	0.87±0.48	0.98±0.38
Queue storage	22.76±4.07	0.04±0.01	6.86±2.34	0.87±0.48	0.98±0.38
Inventory depletion	0.82±0.34	1.72±0.63	6.86±2.34	0.87±0.48	6.55±0.39
Finite residence	0.82±0.34	0.49±0.15	0.64±0.14	0.87±0.48	0.98±0.38
Shared-lane blocking	0.82±0.34	0.04±0.01	6.86±2.34	0.87±0.48	0.98±0.38
Merge competition	0.82±0.34	0.04±0.01	0.64±0.14	11.88±7.22	3.50±2.31
Both junction rules	0.82±0.34	0.04±0.01	6.86±2.34	11.88±7.22	3.50±2.31

(b) All open roads, 15 min

Removed component	Corridor	Branch	Merge
None (reference)	1.03±0.31	5.49±1.35	4.99±1.32
Queue storage	10.72±3.74	19.66±1.29	16.08±0.94
Inventory depletion	2.03±0.53	8.93±1.68	7.22±1.22
Finite residence	1.20±0.33	5.49±1.35	4.99±1.32
Shared-lane blocking	1.03±0.31	8.16±1.71	4.99±1.32
Merge competition	1.03±0.31	5.49±1.35	8.20±2.90
Both junction rules	1.03±0.31	8.16±1.71	8.20±2.90

Wave traversal and arrival. The traversal-supervised timing study uses 24 confirmation closures that vary driving behavior, demand, initial queues, and seed. Arrival is six consecutive 5 s samples with at least three extra vehicles at $U2/U3$; their errors average within cases. Normal queue-position traversals supervise the learner. Features use the preceding 60 s at activation and freeze during

traversal. Scale 3.0 and alternative settings are selected on 12 separate calibration closures by 15 min $U2/U3$ count MAE.

Table 17: All wave controls on the same 24 confirmation cases. Arrival MAE is seconds; count MAE is vehicles (case mean ± sample SD). The first row is the calibration-selected model.

Wave configuration	Arrival	Count (15 min)	Count (30 min)
Selected learner (frozen features)	14.27 ± 8.77	2.72 ± 2.14	2.91 ± 2.40
No traversal delay	14.48 ± 9.61	2.74 ± 2.13	2.92 ± 2.39
One calibrated constant	22.29 ± 18.25	2.80 ± 2.20	2.95 ± 2.42
Condition-specific constants	12.71 ± 7.66	2.69 ± 2.16	2.90 ± 2.41
No occupancy/growth features	13.54 ± 9.00	2.72 ± 2.15	2.91 ± 2.40
Dynamic features	11.67 ± 7.36	2.71 ± 2.14	2.90 ± 2.40
Earlier count-onset learner	13.65 ± 8.04	2.73 ± 2.13	2.91 ± 2.40

Direct traversal is evaluated on 74 accepted 100 m queue-tail movements in three separate normal runs (27/12/35 by condition), weighted equally by run. For each condition, the fixed-speed baseline divides 100 m by the median traversal time in normal training runs. The learned law lowers unscaled traversal-time MAE from 9.64 ± 3.35 to 8.20 ± 3.70 s (14.9%). Table 17 reports arrival and count errors after storage gating, separating these endpoints from direct traversal estimation. Table 5 evaluates the count-onset wave learner used in the final benchmark predictor.

F PeMS INCIDENT PROTOCOL AND AGGREGATE RESULTS

The supplied District 4 collection contains 2,944 links and 2,858 origin detectors, distinct from the standard PEMS04 split. Ten records span seven dates: Bay Bridge (16 November 2023), A-1144255/A-1144248 (22 February 2021), A-1141333/A-1141334 (24 February 2021), A-1076650/A-1076682 (20 April 2021). A-1095204 (5 April 2021), A-1087730 (10 April 2021), A-1087216 (11 April 2021). Same-date records share observations and are grouped for scoring.

The PeMS collection contains 10 incident records, with each record corresponding to one distinct incident. Incidents occurring on the same date may share detector observations, so they are retained individually but aggregated by date for scoring. Thus, all 10 incidents are included in the evaluation, while the reported mean ± SD is computed over the 7 date groups rather than treating shared same-date observations as independent.

Table 18: **Complete PeMS metrics.** Date-group mean ± sample SD. Both scopes retain all 10 incidents. Same-date incidents share observations and are aggregated into date groups before scoring; the reported SD is across date groups. Speed is in mph; flow is in vehicles per 5 min. Bold marks the lowest mean within each scope.

(a) Primary neighborhoods

	Speed (mph)		Flow (veh/5 min)	
Method	MAE	RMSE	MAE	RMSE
Reference	21.32±3.41	28.04±4.86	228.64±132.68	271.06±139.54
CTM	21.94±3.04	29.72±4.07	187.78±70.92	236.82±87.63
IGSTGNN	21.38±3.38	28.05±4.77	226.69±132.06	268.74±138.61
TraffNet	—	—	234.67±133.73	274.30±141.52
MSCT	19.08±2.39	24.90±4.04	—	—
CSTP	**17.00**±2.52	**22.82**±4.78	**136.20**±24.94	**175.00**±23.10

(b) Wider regions

	Speed (mph)		Flow (veh/5 min)	
Method	MAE	RMSE	MAE	RMSE
Reference	12.85±6.31	21.32±6.90	104.19±45.00	160.05±43.84
CTM	13.02±6.57	21.83±7.23	100.08±45.51	153.37±44.06
IGSTGNN	12.88±6.31	21.30±6.85	102.97±43.59	158.13±43.10
TraffNet	—	—	109.91±46.84	163.37±45.81
MSCT	12.47±5.68	19.66±6.64	—	—
CSTP	**11.41**±4.84	**19.57**±4.91	**90.79**±39.35	**134.51**±38.08

G INFERENCE COST, SCORING CHECKS, AND REPRODUCIBILITY

Inference workload. Timing covers 1,440 five-second predictions on 1,969 roads with preloaded normal horizons and fitted models. It includes closure computation, eligibility, readouts, and reconstruction; reference generation, training, loading, scoring, and serialization are outside the timed interval. CTM reuses its all-open setup and computes every closure forecast. Peak RAM includes libraries, inputs, models, and temporaries. Runs use an AMD EPYC 9V74 with four logical CPUs and four threads.

Table 4 uses one run per case after three-block neural warmup: IGSTGNN uses CPU bfloat16 and predicts three outputs; TraffNet/MSCT use float32 and predict flow/speed, respectively. The repeated comparison runs each case three times in fresh processes, alternating method order across folds. CTM+heads computes only its own rollout, eligibility, and heads. Each case contributes its median repeat time; peak RAM is sampled every 10 ms.

Table 19: Repeated inference: all 16 cases, three repeats each. Median time and maximum RAM.

	Time (s)		RAM (GiB)	
Method	Median	Case SD	Peak	Case SD
CTM+heads	12.51	0.28	0.637	0.000
CSTP	2.43	0.49	0.323	0.004

For this inference workload, CSTP is 5.2× faster than CTM+heads with 49% lower peak RAM. All 96 timed forecasts reproduce the saved predictions within the recorded numerical tolerance. At fixed degree, physical updates are approximately linear in road–time positions; CSTP heads evaluate 1.57–6.76% of positions.

Timestamp sensitivity. We assess label alignment by rescoring fixed forecasts at 0/5/10 s on common endpoints 07:00:05–08:59:55 and the shared finite-label intersection. The benchmark's +5 s offset is assumed because the exporter convention is undocumented. Fitting alignment stays unchanged; no offset is selected by error.

Table 20: Label-alignment sensitivity: group-mean MAE on common finite support. These scores use a stricter mask than Table 1.

		Speed (m/s)		Flow (veh/min/road)	
Scope	Label offset	CSTP	CTM+heads	CSTP	CTM+heads
Eight-hop	0 s	2.09	2.34	9.10	9.22
Eight-hop	5 s	2.07	2.31	9.10	9.26
Eight-hop	10 s	2.14	2.38	9.59	9.68
Core	0 s	2.97	2.95	15.54	15.46
Core	5 s	2.91	2.90	15.65	15.57
Core	10 s	3.05	3.04	16.67	16.58

CSTP retains lower local MAE at every offset (speed: 10.1–10.5%; flow: 0.9–1.8%); CTM+heads retains its core advantage.

Reproducibility. The experiment record retains code, configurations, fold assignments, fitted models, forecasts, and case-level metrics. Independent scoring checks the shared masks and aggregation; saved ablation states verify component removals. Reference and timing checks reproduce the archived predictions.

H NON-CRITICAL ROAD CLOSURE CASES

Ten additional closures extend the comparison beyond the original region. The data provider designates these as non-critical cases. The same topology-based rule gives a 615-road eight-hop core within a 1,235-road twelve-hop input subnetwork. Local and regional metrics follow Table 1 on this new region.

We repeat four-fold cross-validation over ten base-road groups. Each fold tests two or three closures, fits seven or eight, and reserves two training groups for model selection. Neural weights are initialized anew, and normal-derived CTM turns and path proxies are recomputed. Architectures, selection grids, epoch limits, and sampling follow Appendix B. CSTP uses the final junction-enabled mechanism; CTM+heads uses CTM states only. Each test group is excluded from fitting and selection. We use these ten cases to evaluate the established architectures and selection procedures after freezing the design choices, they were not used to select the architecture, mechanism set, feature definitions, or selection grids, and were evaluated only after those choices were frozen.

All methods receive the normal horizon and closure query. Pre-onset counts match the reference in all ten cases. Scoring retains the 07:00–09:00 query, 1,440 five-second endpoints, assumed $+5$ s speed/flow alignment, and shared finite-target masks. The same stopped-road rule applies to every case, including one with increasing closed-road counts despite zero recorded entering flow; its prediction error remains scored. Metrics average equally across the ten groups, with sample SD describing group variation.

Table 21: **Non-critical road closure cases.** Mean $\pm$ sample SD over ten held-out road groups under separate four-fold cross-validation. Units, output coverage, and error definitions match Table 1. Neural rows use aggregate-input adaptations. Bold marks the lowest mean; dashes denote unsupported outputs.

	Count (veh/road)		Speed (m/s)		Flow (veh/min/road)	
Method	MAE	RMSE	MAE	RMSE	MAE	RMSE
(a) Near closures: case-specific eight-hop neighborhoods						
CTM	7.33 ± 3.34	15.92 ± 9.63	3.61 ± 1.08	5.87 ± 1.07	14.38 ± 3.49	25.85 ± 6.03
CTM+heads	7.33 ± 3.34	15.92 ± 9.63	3.67 ± 0.63	5.43 ± 0.75	13.57 ± 3.32	22.52 ± 6.03
IGSTGNN	7.27 ± 3.34	15.76 ± 7.57	6.09 ± 1.43	7.94 ± 1.39	13.72 ± 3.37	24.04 ± 7.74
TraffNet	—	—	—	—	14.65 ± 3.58	25.74 ± 6.16
MSCT	—	—	4.55 ± 0.66	5.62 ± 0.77	—	—
CSTP	$\mathbf{4.01}\pm1.47$	$\mathbf{8.07}\pm3.39$	$\mathbf{3.37}\pm0.55$	$\mathbf{5.21}\pm0.45$	$\mathbf{12.48}\pm4.82$	$\mathbf{21.53}\pm5.73$
(b) Regional coverage: fixed 615-road core						
CTM	4.21 ± 1.16	10.13 ± 5.48	4.37 ± 0.12	6.22 ± 0.13	18.17 ± 0.46	30.27 ± 0.63
CTM+heads	4.21 ± 1.16	10.13 ± 5.48	4.36 ± 0.10	6.12 ± 0.14	18.04 ± 0.50	29.71 ± 1.02
IGSTGNN	$\mathbf{3.35}\pm0.39$	7.87 ± 1.90	4.61 ± 0.19	6.30 ± 0.28	16.42 ± 0.38	28.98 ± 0.63
TraffNet	—	—	—	—	30.02 ± 0.73	
MSCT	—	—	$\mathbf{3.82}\pm0.11$	$\mathbf{4.79}\pm0.18$	—	—
CSTP	3.40 ± 0.21	$\mathbf{6.58}\pm0.92$	4.35 ± 0.07	6.15 ± 0.10	$\mathbf{16.20}\pm0.33$	$\mathbf{28.15}\pm0.49$

Table 21 shows a consistent local inventory benefit: CSTP reduces count MAE by 45.3% relative to CTM and 44.9% relative to IGSTGNN, with lower error in all ten road groups against both methods. CSTP also has the lowest local count RMSE, local speed MAE/RMSE, and core count RMSE. The full table reports all outputs in both scoring areas, while Table 22 gives CSTP's case-level results.

Table 22: Case coverage and CSTP MAE for the non-critical cohort. Each row uses its outer-fold prediction. Count is vehicles/road, speed is m/s, and flow is vehicles/min/road. The evaluation record contains all methods' case-level MAE and RMSE.

		Eight-hop neighborhood			615-road core		
Closed road	Fold	Count	Speed	Flow	Count	Speed	Flow
-977099627#3	4	2.18	2.57	10.02	3.21	4.36	18.71
163165902#10	2	3.73	3.76	18.86	3.41	4.33	17.91
26213877#32	2	3.82	3.66	18.61	3.18	4.34	17.90
274251569#8	2	4.62	2.57	16.40	3.57	4.35	17.39
406705728#5	3	4.50	3.05	13.97	3.37	4.39	17.89
420949097#11	4	2.69	2.86	10.25	3.43	4.49	18.14
51870023#3	1	2.58	3.74	10.09	3.24	4.37	18.55
608793058#5	1	6.65	3.66	11.36	3.46	4.35	18.35
648464071#6	4	6.02	3.80	13.83	3.87	4.19	17.92
766921409#14	3	3.26	4.05	13.78	3.20	4.32	18.61