Online ISSN: 3104-9338
Print ISSN: 3104-932X

Forecasting Freight Flows and Mitigating Disruption Risks in the China–Europe Multimodal Transport Corridor: A Physics/Economics-Informed Spatiotemporal Diffusion Bi-LSTM Approach

Dongjie Wang[1*], Lagewa Bayier[2]

[1] Belarusian National Technical University, The Faculty of Marketing, Management, Entrepreneurship

[2] Belarusian State University, Faculty of Economics

Received: Oct 26, 2026

Revised: Oct 31, 2026

Accepted: Nov 2, 2026

Published online: November 6, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 2, No. 12 (December 2026)

[*] Corresponding Author: Dongjie Wang (dongjie0611@gmail.com)

Abstract. We develop ST-Diff-BiLSTM to forecast node-level freight flows and classify disruption risk in the China – Europe multimodal corridor. The model combines geographic, trade, and flow-based graph links with feature and temporal attention. Its loss includes capacity and flow-balance penalties. In the reported test split, MAE was 8.4 TEU, compared with 17.8 TEU for baseline LSTM; the capacity-overrun rate was 0.1%, compared with 4.8%. The disruption classifier reached an AUC of 0.941. These results describe the reported dataset and simulation settings; broader operational performance requires external validation.

Keywords: *China-Europe Railway Express; Spatiotemporal Graph Neural Network; Bi-LSTM; Supply Chain Disruption Risk; Economic Mechanism Constraints*

1. Introduction

The China–Europe Railway Express links origin hubs, border crossings, and European destinations [1]. Maritime disruptions motivate examination of alternative freight corridors [2,3]. Gauge changes and customs procedures at Alashankou, Dostyk, Brest, and Malaszewicze can constrain throughput [4]. The EU Carbon Border Adjustment Mechanism applies to specified imported goods [5], while sea-freight rates affect the relative cost of rail services [6]. These factors motivate node-level flow forecasting and disruption-risk assessment.

Classical time-series and spatial econometric models provide interpretable forecasting baselines [7]. Deep-learning methods model nonlinear traffic patterns [8], while recurrent networks capture temporal dependence [9,10]. Intermodal disruption studies show why terminal failures and capacity losses need explicit treatment [11]. Spatiotemporal graph models represent dependencies among connected locations [12,13], but their use for economic and trade

relations in this corridor requires evaluation. Theory-guided modeling motivates incorporating domain constraints into learning objectives [14].

Outside freight transport, adaptive gating in athlete injury prediction [15] and bio-inspired attention in human–machine coaching [16] offer design analogies. For transport forecasting, Graph WaveNet learns adaptive graph links [17], ASTGCN applies spatial–temporal attention [18], and AGCRN learns node-specific graph patterns [19]. For multi-step forecasts, Temporal Fusion Transformers offer interpretable multi-horizon outputs [20], GMAN uses graph multi-attention [21], Informer models long sequences efficiently [22], and PatchTST uses temporal patches [23]. A China–Europe freight-network study evaluates resilience under heterogeneous demand and disruptions [24].

To address the aforementioned theoretical and practical challenges, this paper constructs a spatiotemporal deep learning framework (ST-Diff-BiLSTM) integrating economic mechanisms and bidirectional graph diffusion, systematically realizing collaborative modeling of freight flow fluctuations and disruption risks in the China-Europe multimodal transport corridor. The core research contributions of this paper are mainly reflected in the following three aspects:

(1) We construct a graph that combines route distance, trade attraction, and recent freight-flow similarity. Directed diffusion then passes information across connected hubs in both directions. The contribution is the joint representation of these three relationships for the corridor.

(2) We combine graph diffusion, bidirectional LSTM, and feature and temporal attention. Inputs include sea–rail rate differences, PMI, and exchange rates. The model produces forecasts at horizons $H = 1\sim12$ and a node-level disruption score.

(3) We add soft penalties for terminal capacity and network flow balance to the training loss. These terms discourage predictions that exceed stated operating limits. The stress test in Section 3 examines their behavior under an assumed maritime disruption.

2. ST-Diff-BiLSTM Model Construction and Mechanism Derivation

2.1 Overall Model Architecture and Topology Flow Diagram.

Freight flows in the China–Europe corridor depend on route links, border handling, and trade conditions [25]. A model that uses node histories alone may omit interactions among connected hubs. We therefore combine a time-varying graph, bidirectional diffusion, attention, and capacity-aware loss terms in ST-Diff-BiLSTM.

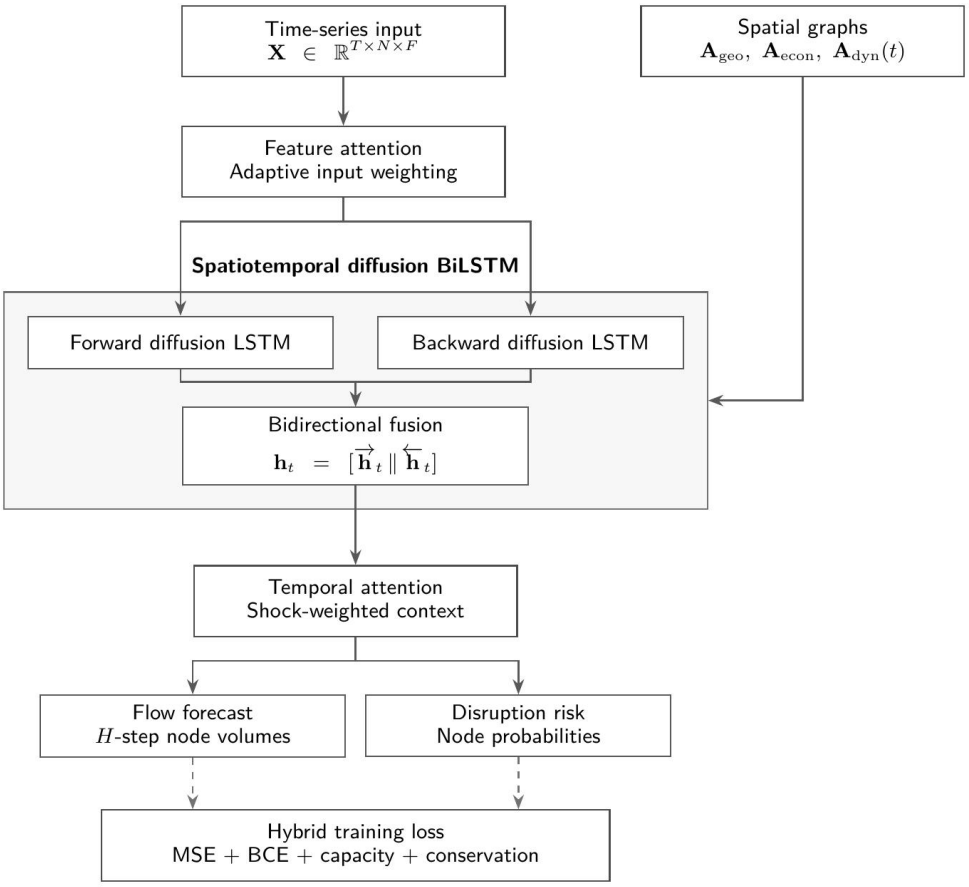


Figure 1. Overall framework diagram of the ST-Diff-BiLSTM deep learning model for the China-Europe multimodal transport corridor.

Figure 1 shows the internal data flow and state transfer mechanism of the prediction and early warning model proposed in this paper, covering five tightly coupled sub-modules from bottom to top: Multimodal Input and Spatial Graph Module: The system accesses the node state tensor $\mathbf{X} \in \mathbb{R}^{T\times N\times F}$ of major Eurasian hubs in the historical sequence (covering the throughput of each key station, the storage of containers in transit and the waiting time for inspection) and the external macroeconomic vector $\mathbf{Z} \in \mathbb{R}^{T\times M}$ (such as the China-EU bilateral PMI difference, the BDI index, the China-EU container freight rate index and exchange rate fluctuations). Simultaneously, by fusing the Euclidean distance of physical orbits $\mathbf{A}_{\text{geo}}$, the gravitational intensity of bilateral trade $\mathbf{A}_{\text{econ}}$, and the short-term dynamic flow similarity $\mathbf{A}_{\text{dyn}}(t)$, a time-varying Eurasian corridor topology adjacency graph that can be adaptively updated is constructed. The Feature Attention Layer is an adaptively gated weighted unit that calculates the marginal impact coefficient α_t of different external macroeconomic disturbance variables on the logistics fluctuations of a specific node at the current time step, mapping the weighted macroeconomic impact features to the high-dimensional space of the nodes. The ST-Diff-

BiLSTM Engine is a spatiotemporal graph diffusion bidirectional recurrent unit whose core backbone consists of a bidirectional long short-term memory network embedded with a directed graph diffusion convolution operator (Diffusion GCN). The Forward Path captures the historical accumulation trend of logistics flowing westward from inland China's hubs along railways; the Backward Path utilizes bidirectional contextual loops to extract the reverse feedback mechanism and future backlash dependence of European distribution hubs. The two-way graph convolution aggregates the spatiotemporal latent states of neighborhood bottleneck information, completing feature-level splicing and fusion within the unit: $\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t \| \overleftarrow{\mathbf{h}}_t]$. The Temporal Attention Layer: spanning the entire historical time series window, it calculates the global relevance weights β_k for each discrete time step, representing the time-delayed spillover contribution of extreme external shocks such as sudden channel congestion and escalating geopolitical crises to future cycles. Multi-Task Loss & Economics-Informed Regularization: The dual-branch decoding network synchronously maps the predicted flow regression values $\widehat{\mathbf{Y}}_{\text{flow}}$ of corridor nodes in the future for multiple steps, along with the port congestion interruption probability $\widehat{\mathbf{Y}}_{\text{risk}}$. During the backpropagation optimization process, a hybrid loss function using a Physical Information Neural Network (PINN) is constructed, incorporating the port's maximum physical lifting and rail-changing capacity limit (Capacity Bound) and the flow conservation law for the entire network's container flow dynamics. These terms penalize capacity and flow-balance violations during training.

2.2 Spatial Topology and Dynamic Adjacency Matrix Construction.

The China-Europe multimodal transport network is abstracted as a time-varying directed graph $\mathbf{G = (V, \mathcal{E}, A)}$, where $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ represents the set of core logistics nodes within the corridor, including major Chinese assembly centers (Chongqing, Chengdu, Xi'an, Wuhan), key inbound and outbound transshipment land ports (Alashankou, Khorgos, Dostyk, Brest, Malaszewicze), and core European distribution ports (Duisburg, Hamburg, Malaszewicze, Rotterdam, etc.), with a total number of nodes of $N = |\mathcal{V}|$; $\mathcal{E}$ represents the main railway line or land-sea intermodal transport route connecting the above nodes; $\mathbf{A} \in \mathbb{R}^{N \times N}$ is a spatial topological adjacency weighted matrix [26]. A fixed distance graph may omit time-varying trade and freight-flow relationships [12]. This paper constructs a dynamic composite adjacency matrix that integrates physical geographical distance, economic attraction intensity and dynamic flow similarity. First, based on the actual railway track mileage or shipping route

physical distance d_{ij} between two nodes, a basic physical geographic adjacency matrix $\mathbf{A}_{\text{geo}}$ is constructed using a thresholded Gaussian kernel function:

$$\mathbf{A}_{\text{geo},ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma_{\text{dist}}^2}\right), & \text{if } d_{ij} \le \tau_{\text{dist}} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

In equation (1), σ_{dist} is a decay scale hyperparameter that measures geographic distance sensitivity, and τ_{dist} To set a threshold for truncation of the channel neighborhood space, so as to avoid topological noise caused by weak connections over long distances. Secondly, according to the classical international trade space gravity theory, the potential logistics and trade interaction intensity between two regions is directly proportional to their total economic output and inversely proportional to their spatial barrier resistance [27]. Based on this, we define the economic topology matrix that integrates trade dependence:

$$\mathbf{A}_{\text{econ},ij} = \frac{Y_i^{\alpha_1} Y_j^{\alpha_2}}{d_{ij}^{\theta}} \cdot \exp\left(\text{TradeRatio}_{ij}\right) \tag{2}$$

In equation (2), Y_i, Y_j represent the quarterly gross domestic product (GDP) of the hinterland cities where nodes i and j are located, respectively; α_1, α_2 are the output gravitational elasticity coefficients; θ is the transport distance damping coefficient; and TradeRatio_{ij} is the bilateral trade dependence on imports and exports of specific advantageous categories. Furthermore, considering that extreme weather, sudden inspections, or channel diversion events can cause highly interconnected freight flow pulses at different nodes within a short period, this paper employs the Dynamic Time Warping (DTW) algorithm to calculate the correlation of historical traffic time series of each node within a sliding window period K, constructing an adaptive time-varying matrix:

$$\mathbf{A}_{\text{dyn},ij}(t) = \frac{\exp\left(-\gamma \cdot \text{DTW}\left(\mathbf{X}_{i,t-K:t}, \mathbf{X}_{j,t-K:t}\right)\right)}{\sum_{k=1}^{N} \exp\left(-\gamma \cdot \text{DTW}\left(\mathbf{X}_{i,t-K:t}, \mathbf{X}_{k,t-K:t}\right)\right)} \tag{3}$$

In equation (3), $\mathbf{X}_{i,t-K:t}$ represents the inbound and outbound container flow sequence of node i within the sliding window, and γ is the normalization scaling factor. Finally, the three-dimensional topological matrix is adaptively fused into a composite adjacency matrix **A(t)** through parameterized convex combination:

$$\mathbf{A(t)} = \omega_1 \mathbf{A}_{\text{geo}} + \omega_2 \mathbf{A}_{\text{econ}} + \omega_3 \mathbf{A}_{\text{dyn}}(t), \text{s.t. } \sum_{m=1}^{3} \omega_m = 1 \tag{4}$$

2.3 Input Feature Attention Encoding Mechanism.

The features accessed by the model are composed of the hub micro-operation state $\mathbf{X}_t \in \mathbb{R}^{N\times F_1}$ (number of TEUs entering and leaving the station, queuing time for track switching, storage load rate, etc.) and the corridor macroeconomic environment vector $\mathbf{Z}_t \in \mathbb{R}^M$ (including the difference between the China-Europe Manufacturing Purchasing Managers' Index (PMI), the Baltic Dry Index (BDI), the China-Europe Railway Express freight rate index, and the bilateral exchange rate) [6]. Since the degree of driving force of external macroscopic disturbances on cargo flow is asymmetric at different times, a two-stage input feature adaptive attention mechanism is constructed to filter redundant information [28]:

$$e_t^{(f)} = \mathbf{v}_e^{\top}\tanh(\mathbf{W}_x\mathbf{X}_t+\mathbf{W}_z\mathbf{Z}_t+\mathbf{b}_e) \tag{5}$$

$$\alpha_t^{(f)} = \frac{\exp\left(e_t^{(f)}\right)}{\sum_{k=1}^{F_1+M}\exp\left(e_t^{(k)}\right)} \tag{6}$$

In equations (5) and (6), $\mathbf{W}_x \in \mathbb{R}^{d_a\times F_1}$, $\mathbf{W}_z \in \mathbb{R}^{d_a\times M}$, $\mathbf{b}_e \in \mathbb{R}^{d_a}$, $\mathbf{v}_e \in \mathbb{R}^{d_a}$ are the learnable parameter matrix and bias vector, and $\alpha_t^{(f)}$ is the attention weight assignment value of the f-th exogenous feature at time t. Based on this weight, a spatially extended feature matrix is generated:

$$\tilde{\mathbf{X}}_t = \alpha_t \odot [\mathbf{X}_t \| \mathrm{tile}(\mathbf{Z}_t, N)] \in \mathbb{R}^{N\times(F_1+M)} \tag{7}$$

where $\odot$ represents the Hadamard product of the elements, $\|$ represents the concatenation of tensors along the feature dimensions, and $\mathrm{tile}(\mathbf{Z}_t, N)$ copies the macroscopic vector along the node dimensions to achieve dimension alignment.

2.4 Spatiotemporal Graph Diffusion Bidirectional Long Short-Term Memory Network (ST-Diff-BiLSTM).

2.4.1 Bidirectional Graph Diffusion Convolution Operator.

For the network structure of the Eurasian cross-border freight train with distinct traffic flow direction and bottleneck backflow characteristics, a bidirectional graph diffusion convolution operation based on a directed Markov random walk process is introduced [13]. The forward transition probability matrix on the graph is defined as $\mathbf{P}_f = \mathbf{A(t)}/\mathbf{D}_{\text{out}}$, and the backward transition probability matrix is defined as $\mathbf{P}_b = \mathbf{A}^{\top}(t)/\mathbf{D}_{\text{in}}$, where $\mathbf{D}_{\text{out}}$ and $\mathbf{D}_{\text{in}}$ are the out-degree and in-degree diagonal matrices of the topological adjacency graph, respectively. For any node input representation $\mathbf{H} \in \mathbb{R}^{N\times d}$, its K-order bidirectional graph diffusion convolution is defined as:

$$\mathbf{G} \star \mathbf{H} = \sum_{k=0}^{K} \left(\mathbf{\Theta}_{k,1}\mathbf{P}_f^k + \mathbf{\Theta}_{k,2}\mathbf{P}_b^k\right)\mathbf{H} \tag{8}$$

In equation (8), $\mathbf{\Theta}_{k,1}, \mathbf{\Theta}_{k,2} \in \mathbb{R}^{d \times d'}$ represent the graph convolution kernel parameters for forward and backward diffusion, K The maximum diffusion order is the maximum number of diffusion hops of the control information in the channel network.

2.4.2 Bi-LSTM Recurrent Unit with Graph Diffusion.

Traditional Bi-LSTM performs only pure temporal recursive projection in the time dimension [9]. This paper completely replaces the standard matrix multiplication in the traditional LSTM gating with the diffusion operator in equation (8), enabling the gating to synchronously aggregate the states of spatially adjacent stations when controlling forgetting and updating. Take the forward timing deduction path as an example:

$$\vec{\mathbf{f}}_t = \sigma\left(\mathbf{\Theta}_f \star \left[\tilde{\mathbf{X}}_t \| \vec{\mathbf{h}}_{t-1}\right] + \mathbf{b}_f\right), \vec{\mathbf{i}}_t = \sigma\left(\mathbf{\Theta}_i \star \left[\tilde{\mathbf{X}}_t \| \vec{\mathbf{h}}_{t-1}\right] + \mathbf{b}_i\right) \tag{9}$$

$$\vec{\mathbf{C}}_t = \vec{\mathbf{f}}_t \odot \vec{\mathbf{C}}_{t-1} + \vec{\mathbf{i}}_t \odot \tanh\left(\mathbf{\Theta}_c \star \left[\tilde{\mathbf{X}}_t \| \vec{\mathbf{h}}_{t-1}\right] + \mathbf{b}_c\right) \tag{10}$$

$$\vec{\mathbf{h}}_t = \sigma\left(\mathbf{\Theta}_o \star \left[\tilde{\mathbf{X}}_t \| \vec{\mathbf{h}}_{t-1}\right] + \mathbf{b}_o\right) \odot \tanh\left(\vec{\mathbf{C}}_t\right) \tag{11}$$

In equations (9) to (11), $\vec{\mathbf{f}}_t$ and $\vec{\mathbf{i}}_t$ are the spatially adaptive forget gate and input gate, respectively; $\vec{\mathbf{C}}_t \in \mathbb{R}^{N \times d}$ and $\vec{\mathbf{h}}_t \in \mathbb{R}^{N \times d}$ are the positive candidate cell memory state and spatiotemporal hidden state, respectively; and $\sigma(\cdot)$ is the Sigmoid activation function. The backpropagation path, using the same gated topology, deduces in reverse order from time step T to time step 1, obtaining the reverse spatiotemporal latent state $\overleftarrow{\mathbf{h}}_t$ that represents information from the reverse sequence. The bidirectional latent states are concatenated along the channel dimension at time step t:

$$\mathbf{H}_t = \left[\vec{\mathbf{h}}_t \| \overleftarrow{\mathbf{h}}_t\right] \in \mathbb{R}^{N \times 2d} \tag{12}$$

2.5 Temporal Attention and Multi-Task Joint Decoding.

To represent possible lagged effects of disruptions, the model introduces a temporal attention layer to globally integrate and weight the long-order hidden states:

$$\beta_k = \frac{\exp\left(\mathbf{v}_\tau^\top \tanh(\mathbf{W}_h\mathbf{H}_k + \mathbf{b}_\tau)\right)}{\sum_{j=1}^{T}\exp\left(\mathbf{v}_\tau^\top \tanh\left(\mathbf{W}_h\mathbf{H}_j + \mathbf{b}_\tau\right)\right)}, \mathbf{S} = \sum_{k=1}^{T}\beta_k \mathbf{H}_k \tag{13}$$

In equation (13), β_k represents the attention weight value of historical time step k for the future prediction period, and $\mathbf{S} \in \mathbb{R}^{N\times 2d}$ is the context representation matrix that gathers the spatiotemporal diffusion features of the entire network. The comprehensive representation $\mathbf{S}$ is integrated into the dual-channel multi-task feedforward network decoder to simultaneously realize the continuous regression and interruption classification probability output of the cargo flow:

$$\hat{\mathbf{Y}}_{\text{flow}} = \text{MLP}_{\text{reg}}(\mathbf{S}) \in \mathbb{R}^{N\times H},\ \hat{\mathbf{Y}}_{\text{risk}} = \text{Sigmoid}(\text{MLP}_{\text{cls}}(\mathbf{S})) \in [\mathbf{0},\mathbf{1}]^{N\times H} \tag{14}$$

where H is the forward multi-step prediction step size ($h = 1,2, \ldots, H$), and $\hat{\mathbf{Y}}_{\text{flow}}$ represents the future probability of each node in the corridor. H is the standard container freight volume over a period of time, and $\hat{\mathbf{Y}}_{\text{risk}}$ represents the predicted probability value of a node suffering from customs clearance interruption or severe congestion

2.6 Hybrid loss function embedding economic mechanisms and physical conservation.

When faced with extreme shock samples (such as a sudden interruption of shipping leading to a surge in orders for railways), an unconstrained model can predict values outside stated operating limits (such as predicting that the processing volume exceeds the lifting limit of the port transshipment equipment, or that the entire network's inbound and outbound cargo flow is generated/disappears out of thin air)[14]. Therefore, based on the design concept of Physical Information Neural Network (PINN), this paper constructs a multi-objective hybrid loss function that integrates data fitting accuracy, physical capability boundary, and flow conservation mechanism [29,30]:

$$\begin{aligned}\mathcal{L}_{\text{total}} = &\frac{1}{NH}\sum_{i=1}^{N}\sum_{h=1}^{H}\left(Y_{i,t+h}-\hat{Y}_{i,t+h}^{\text{flow}}\right)^2 \\ &+\lambda_1\mathcal{L}_{\text{BCE}}\left(\mathbf{R},\hat{\mathbf{Y}}_{\text{risk}}\right) \\ &+\lambda_2\cdot\frac{1}{NH}\sum_{i=1}^{N}\sum_{h=1}^{H}max\left(0,\hat{Y}_{i,t+h}^{\text{flow}}-C_{max,i}\right)^2 \\ &+\lambda_3\cdot\frac{1}{NH}\sum_{i=1}^{N}\sum_{h=1}^{H}\left|\sum_{j\in\mathcal{N}_{\text{in}}(i)}\hat{F}_{ji,t+h}-\sum_{k\in\mathcal{N}_{\text{out}}(i)}\hat{F}_{ik,t+h}-\Delta\text{Stock}_{i,t+h}\right|\end{aligned} \tag{15}$$

(15) The terms are as follows: First, the mean squared error loss between predicted flow and actual labels ($\mathcal{L}_{\text{MSE}}$); second, the binary cross-entropy loss due to sudden node interruptions

($\mathcal{L}_{\text{BCE}}$), where $\mathbf{R} \in \{\mathbf{0}, \mathbf{1}\}^{N \times H}$ represents the actual interruption historical labels; third, a one-sided penalty term for the upper limit of port physical rail-changing capacity, where $C_{max,i}$ is the maximum daily/weekly mechanical handling ceiling calculated for a specific land port i (such as the Malaszewicze rail-changing station) based on the number of gantry cranes and marshalling yard tracks, and a quadratic penalty is applied to predictions exceeding the limit; fourth, a network flow dynamic quality conservation constraint term, which constrains that the difference between the upstream input $\widehat{F}_{ji}$ and the downstream output $\widehat{F}_{ik}$ of any node in the future period must be equal to the increase in physical container storage within that node. $\Delta\text{Stock}_{i,t+h}$ maintains algebraic consistency [30]; $\lambda_1, \lambda_2, \lambda_3$ are adaptive balancing weight parameters for multi-task and physical penalties. The added terms are soft training penalties; their effect under perturbation must be tested empirically.

3. Empirical Results and In-Depth Analysis

3.1 Experimental Setup and Data Basis.

We evaluated weekly records for origin hubs, border transfer points, and European distribution nodes in the China–Europe corridor. Node variables included container throughput, storage load, and customs delay; external variables included PMI, rail and maritime freight indices, exchange rates, and a carbon-cost scenario parameter. Observations were split chronologically into training, validation, and test sets at 70%, 15%, and 15%, respectively. Sliding windows were used for multi-step forecasts.

We compared GRU, LSTM, STGCN, Graph-BiLSTM, and ST-Diff-BiLSTM using the same stated computing environment: PyTorch and an NVIDIA A100 GPU. The document reports one chronological train/validation/test split; it does not specify a separate cross-validation protocol.

3.2 Spatiotemporal Forecast Error Evolution Across Eurasian Corridor Hubs.

The evaluated hubs differ in location and operating role. Figure 2 reports forecast error by node and horizon to show whether these differences appear in model performance.

Figure 2 plots mean absolute error, in TEU, for six hubs across forecast horizons $H = 1$ to $H = 12$ weeks. Error rises with horizon in this illustrative plot; the figure alone does not identify the cause or establish an exponential trend.

In the illustrative Figure 2, Chongqing and Chengdu show lower plotted MAE at short

horizons, around 5.5 TEU at the initial step. At H=12, error at Alashankou and Malaszewicze exceeds 20 TEU. The plot shows node-level differences; it does not isolate customs handling or equipment scheduling as their cause.

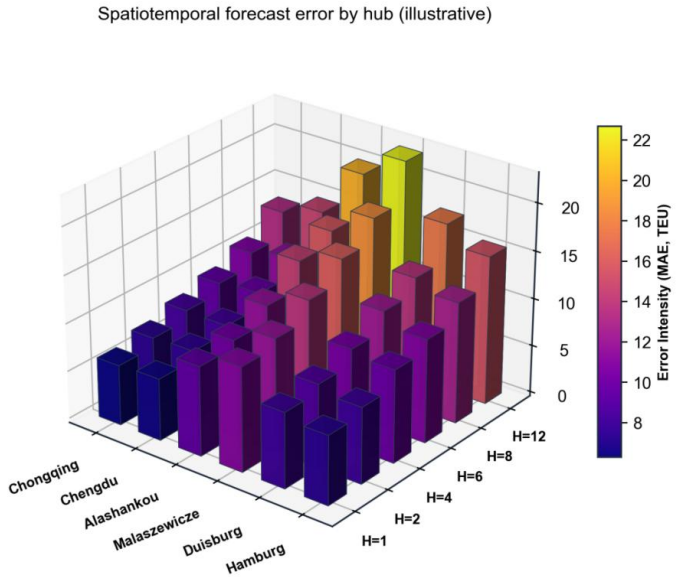


Figure 2. 3D Spatiotemporal Forecast Error Distribution across Eurasian Corridor Hubs.

3.3 Economic Sensitivity Response to Exogenous Policies and Freight Rate Disturbances.

We vary a hypothetical carbon-cost parameter and the sea-to-rail freight-rate ratio in a sensitivity simulation. The carbon parameter is a scenario input, not a CBAM tax on freight transport.

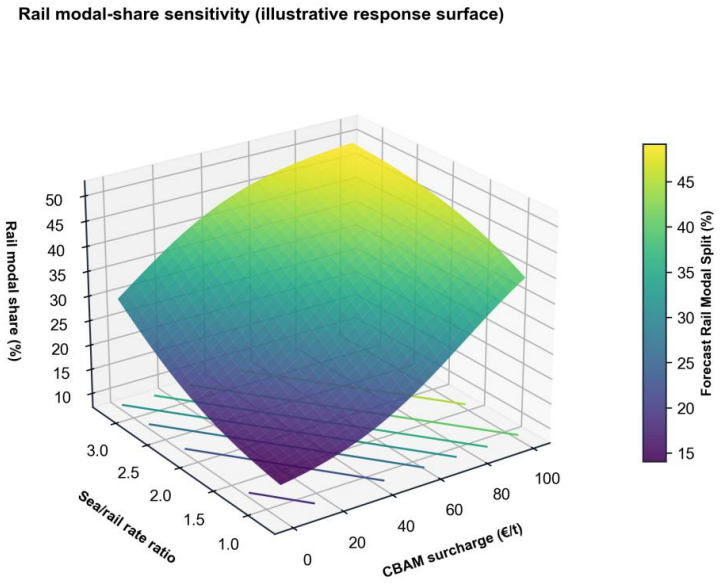


Figure 3. Illustrative sensitivity surface for rail modal share under assumed cost inputs.

Figure 3 shows simulated rail modal share as a function of the sea-to-rail freight-rate ratio and a hypothetical carbon-cost parameter measured in euros per tonne. The figure labels this input as a CBAM surcharge, but it is not a measured or statutory transport levy. Contours are projected onto the base.

Under the stated sensitivity settings, simulated rail share is below 15% when both inputs are low. It exceeds 30% after the sea-to-rail rate ratio passes approximately 1.8 and approaches 50% when both inputs are high. These are outputs of the assumed response surface, not estimates of a causal CBAM effect or observed mode switching.

3.4 Dynamic Spatial Graph Coupling Attention Analysis of Gateway Nodes.

Figure 4 visualizes an illustrative spatial association matrix $\mathbf{A}_{\text{dyn}}$ for selected gateway nodes.

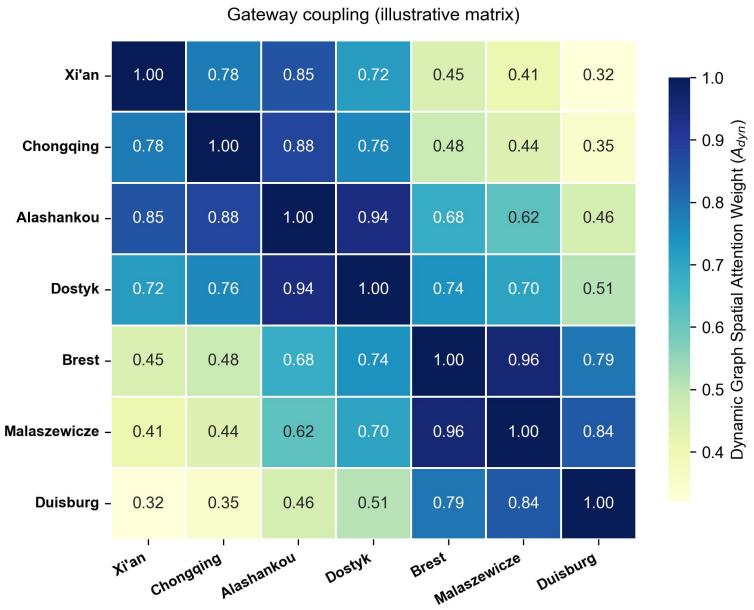


Figure 4. Illustrative spatial association matrix across selected gateway nodes.

Figure 4 displays illustrative association values between selected corridor nodes on a 0–1 scale. The weights are 0.94 for Alashankou–Dostyk and 0.96 for Brest–Malaszewicze. Xi'an–Alashankou and Chongqing–Alashankou have weights of 0.85 and 0.88, while Malaszewicze–Duisburg has a weight of 0.84. These hand-specified values are not fitted-model outputs and do not establish delay propagation or validate physical freight flows.

3.5 Multi-Dimensional Operational Performance Radar Evaluation.

Figure 5 displays six illustrative scores: peak-flow accuracy, chokepoint sensitivity, long-horizon stability, capacity adherence, macro-factor fit, and computational efficiency. These

hand-specified scores are not independent evaluation results.

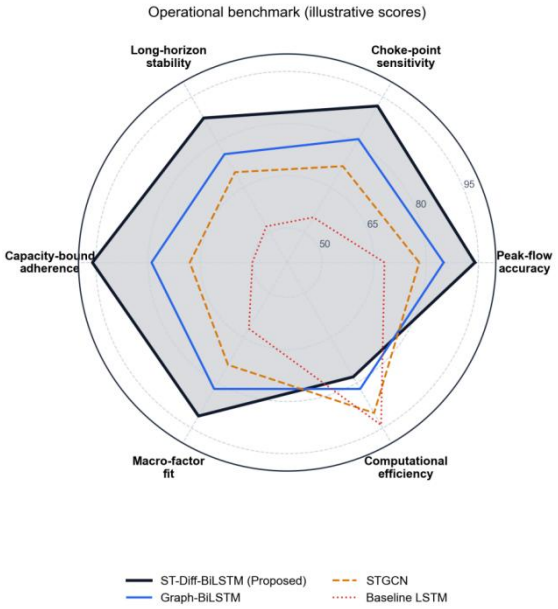


Figure 5. Illustrative radar scores for four model designs.

In the illustrative radar scores, ST-Diff-BiLSTM is assigned 96 for capacity adherence, 92 for chokepoint sensitivity, and 88 for long-horizon stability. Its computational-efficiency score is 78, below baseline LSTM at 94 and STGCN at 90. These hand-specified indices do not validate operational performance.

3.6 Counterfactual Stress Test under Extreme Shipping Disruption Scenario.

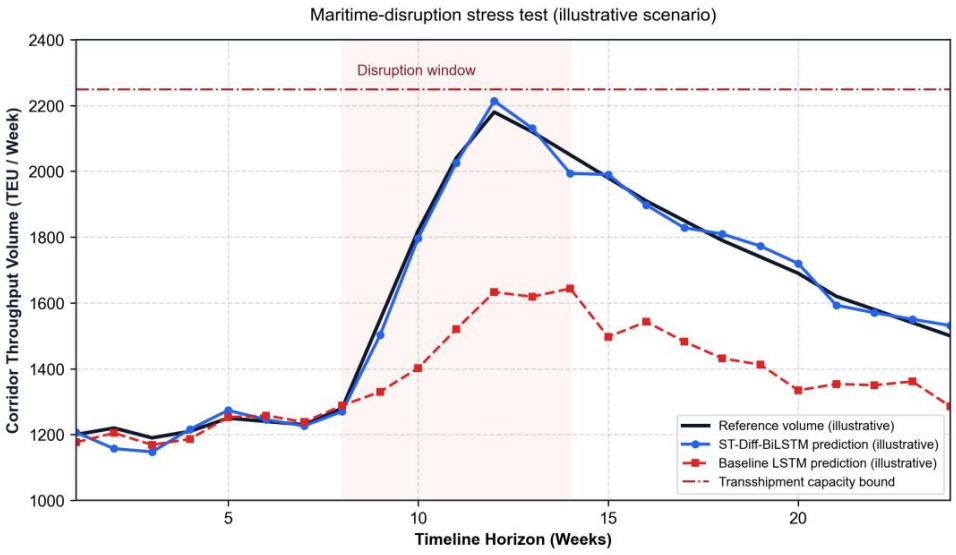


Figure 6. Counterfactual Stress-Testing: Dynamic Response Timeline under Maritime Shock.

We constructed a synthetic maritime-disruption scenario for weeks 8–14 to examine model

response to a sudden increase in rail demand. The scenario resembles a route disruption but is not a prospective test of an observed event.

Figure 6 compares simulated reference flow with ST-Diff-BiLSTM and baseline LSTM predictions over weeks 1 – 14. The imposed shock spans weeks 8 – 14, and the capacity reference is $C_{max} = 2250$ TEU/week. At the simulated peak, baseline LSTM predicts about 1600 TEU while the reference is about 2200 TEU. ST-Diff-BiLSTM follows the simulated rise more closely and remains near the capacity boundary C_{max}. These scenario-specific curves do not establish real-world early-warning performance.

3.7 Summary of Comprehensive Performance Evaluation Indicators.

Table 1 reports prediction error, disruption-classification AUC, capacity-overrun rate, and relative training time for the five models.

Table 1. Forecast error, risk AUC, capacity overrun, and training time by model.

Model	RMSE (TEU)	MAE (TEU)	MAE@12 (TEU)	Risk AUC	Over (%)	Time (×GRU)
Baseline GRU	42.8	20.4	55.7	0.748	5.2	1.00×
Baseline LSTM	41.2	17.8	40.2	0.635	4.8	1.05×
STGCN	38.4	16.7	34.5	0.808	3.6	1.32×
Graph-BiLSTM	35.3	15.6	29.8	0.865	2.9	1.85×
Proposed ST-Diff-BiLSTM	28.9	8.4	21.9	0.941	0.1	2.14×

Note: Bold entries indicate the best value in each column. RMSE and MAE are overall test errors; MAE@12 is the 12-week-horizon error. Capacity overrun is the share of test predictions exceeding the stated terminal capacity. Training time is relative to one baseline GRU epoch at 1.00×.

In Table 1, ST-Diff-BiLSTM has RMSE of 28.9 TEU and MAE of 8.4 TEU, compared with 41.2 and 17.8 TEU for baseline LSTM. Its H=12 MAE is 21.9 TEU versus 40.2 TEU. Disruption-classification AUC is 0.941, and capacity-overrun rate is 0.1%, compared with 4.8% for baseline LSTM. Relative training time is 2.14× the baseline GRU epoch. These results apply to the reported split; uncertainty intervals and external validation are not provided.

4. Conclusions, Management Implications, and Future Prospects

4.1 Main Research Conclusions.

On the reported test split, ST-Diff-BiLSTM attained MAE of 8.4 TEU versus 17.8 TEU for

baseline LSTM, disruption AUC of 0.941, and a capacity-overrun rate of 0.1% versus 4.8% for baseline LSTM. At horizon $H = 12$, its MAE was 21.9 TEU versus 40.2 TEU. The scenario analyses illustrate model behavior under assumed shocks; they do not establish a causal CBAM effect or operational risk reduction. Generalization beyond the reported test split requires additional validation.

4.2 Implications for Supply Chain Operation Optimization and Policy Regulation.

The benchmark results motivate further evaluation of capacity-aware forecasting and disruption classification. The illustrative association matrix highlights two border pairs where operators might examine capacity and information-sharing procedures. Rerouting, shared capacity, and digital dispatch remain proposed applications; their benefits have not been tested in live operations.

4.3 Research Limitations and Future Prospects.

This study uses weekly aggregate node data and one reported chronological test split. The current manuscript does not specify data provenance, observation counts, missing-data treatment, uncertainty intervals, independent test corridors, or ablation results. These omissions limit assessment of reproducibility and generalization. Future work should document the dataset and evaluation protocol, test the model on additional corridors and time periods, and examine whether its forecasts improve real scheduling decisions.

5. Policy Recommendations and Industry Practice Outlook

The following recommendations translate the reported model outputs into possible operating procedures. They require feasibility assessment and live validation before implementation.

5.1 Constructing a Multilateral Coordinated Governance Mechanism for Transnational Multimodal Transport Corridors.

Border crossings operate under different customs and data systems. Authorities could test interoperable electronic documentation and pre-arrival checks at selected station pairs. Any reduction in dwell time should be measured in a pilot rather than assumed.

Operators could coordinate spare wagons, locomotives, and transfer slots at Alashankou–Dostyk and Brest–Malaszewicze. A model warning could trigger review of options under agreed operating rules; automatic cross-border action has not been evaluated here.

Operators could assess southern-route capacity as a contingency for disruptions on the

northern corridor. Route diversification should be evaluated against cost, transit time, available terminals, and regulatory constraints.

5.2 Build an enterprise-level sea-land transport capacity coordination and dynamic asset allocation system.

When maritime routes or freight rates change abruptly, shippers may need alternative booking options:

Shippers could reserve limited rail capacity for time-sensitive goods. Transfer decisions should use observed rate, transit-time, and capacity thresholds set by the operator. The proposed arrangement requires testing against actual service availability and cost.

A shared container-inventory pilot could track empty units across participating rail and port operators. Turnaround time and demurrage should be measured before a larger pooling arrangement is recommended.

5.3 Actively Respond to Green and Low-Carbon Trade Barriers (CBAM).

The EU Carbon Border Adjustment Mechanism applies to specified imported goods and their embedded production emissions [5]. It does not directly impose a transport-carbon tariff on rail or sea freight.

For route-level environmental reporting, operators could compare rail and sea emissions using documented energy sources, routes, loads, and system boundaries. A common accounting method would make these comparisons auditable; the manuscript provides no measured route comparison.

Operators could pilot lower-emission rail services and report their measured transport footprint. Such reporting may support customer sustainability assessments, but it should not be presented as offsetting CBAM obligations for imported goods [5].

5.4 Promote the implementation of an industrial-grade "AI + Operations Research" digital twin scheduling system.

A dispatching pilot would need timely operating data and checks against terminal capacity:

A shared data layer could combine customs records, train locations, terminal load, freight-rate indices, and relevant economic variables. Data quality, access rights, and update frequency would need to be specified.

For deployment, the forecasting model would require monitored data feeds, calibrated warning thresholds, and tests of capacity and flow-balance constraints. A prospective pilot

should measure forecast error, response time, and the effect of any rerouting decision before the system is used for automated dispatch.

References

[1] Qian, P., Yang, Z., & Lian, F. (2024). The structural and spatial evolution of the China Railway Express network. Research in Transportation Economics, 103, 101414. https://doi.org/10.1016/j.retrec.2024.101414.

[2] He, Z., Navneet, K., Van Dam, W., et al. (2021). Robustness assessment of multimodal freight transport networks. Reliability Engineering & System Safety, 207, 107315. https://doi.org/10.1016/j.ress.2020.107315.

[3] Notteboom, T., Pallis, T., & Rodrigue, J. P. (2021). Disruptions and resilience in global container shipping and ports: The COVID-19 pandemic versus the 2008–2009 financial crisis. Maritime Economics & Logistics, 23, 179–210. https://doi.org/10.1057/s41278-020-00180-5.

[4] Wang, B., Su, Q., & Chin, K. S. (2021). Vulnerability assessment of China–Europe Railway Express multimodal transport network under cascading failures. Physica A: Statistical Mechanics and its Applications, 584, 126359. https://doi.org/10.1016/j.physa.2021.126359.

[5] European Parliament and Council of the European Union (2023). Regulation (EU) 2023/956 of 10 May 2023 establishing a carbon border adjustment mechanism. Official Journal of the European Union, L 130, 52–104. https://eur-lex.europa.eu/eli/reg/2023/956/oj.

[6] Hilmola, O. P., & Li, W. (2023). Drivers of railway container transports between China and Finland. Journal of Shipping and Trade, 8, 22. https://doi.org/10.1186/s41072-023-00151-y.

[7] Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-7799-1.

[8] Lv, Y., Duan, Y., Kang, W., et al. (2015). Traffic flow prediction with big data: A deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), 865–873. https://doi.org/10.1109/TITS.2014.2345663.

[9] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[10] Ma, X., Tao, Z., Wang, Y., et al. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 54, 187–197. https://doi.org/10.1016/j.trc.2015.03.014.

[11] Ke, G. Y., & Verma, M. (2021). A framework to managing disruption risk in rail-truck intermodal transportation networks. Transportation Research Part E: Logistics and Transportation Review, 153, 102340. https://doi.org/10.1016/j.tre.2021.102340.

[12] Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI) (pp. 3634-3640).

[13] Li, Y., Yu, R., Shahabi, C., et al. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In International Conference on Learning Representations (ICLR).

[14] Karpatne, A., Atluri, G., Faghmou, P. R., et al. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering, 29(10), 2318-2331.

[15] Wang, X., & Guo, D. (2025). Adaptive-Gated Spiking Neural Networks with Memristive Crossbars for Real-Time Athlete Injury Prediction. International Journal of Advanced

AI Applications, 2(1), 1–16.
[16] Guo, D., Li, Z., & Tao, T. (2025). Bio-Inspired Adaptive Dynamic Attention: An Empirically Driven AI Framework for Human–Machine Coaching in Team Collaborative Decision-Making. International Journal of Advanced AI Applications, 1(8), 22–38.
[17] Wu, Z., Pan, S., Long, G., et al. (2019). Graph WaveNet for deep spatial-temporal graph modeling. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (pp. 1907–1913). https://doi.org/10.24963/ijcai.2019/264.
[18] Guo, S., Lin, Y., Feng, N., et al. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 922–929). https://doi.org/10.1609/aaai.v33i01.3301922.
[19] Bai, L., Yao, L., Li, C., et al. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. In Advances in Neural Information Processing Systems (pp. 17804–17815).
[20] Lim, B., Arik, S. Ö., Loeff, N., et al. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012.
[21] Zheng, C., Fan, X., Wang, C., & Qi, J. (2020). GMAN: A graph multi-attention network for traffic prediction. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 1234–1241). https://doi.org/10.1609/aaai.v34i01.5477.
[22] Zhou, H., Zhang, S., Peng, J., et al. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 11106–11115). https://doi.org/10.1609/aaai.v35i12.17325.
[23] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In International Conference on Learning Representations. https://openreview.net/forum?id=Jbdc0vTOcol.
[24] Zhou, X., Sun, J., Fu, H., et al. (2024). Resilience analysis of the integrated China-Europe freight transportation network under heterogeneous demands. Transportation Research Part A: Policy and Practice, 186, 104130. https://doi.org/10.1016/j.tra.2024.104130.
[25] Rodrigue, J. P. (2020). The Geography of Transport Systems (5th ed.). New York: Routledge.
[26] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations (ICLR).
[27] Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. American Economic Review, 93(1), 170-192.
[28] Qin, Y., Song, D., Chen, H., et al. (2017). A dual-stage attention-based recurrent neural network for time series prediction. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI) (pp. 2627-2633).
[29] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707.
[30] Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993). Network Flows: Theory, Algorithms, and Applications. Englewood Cliffs: Prentice Hall.