

<https://doi.org/10.1038/s44333-026-00094-4>

# Assessing subway ridership resilience under extreme weather with vine copula modeling

Check for updates

Yan Guo<sup>1</sup>, Brian Yueshuai He<sup>1</sup>✉, Joseph Y. J. Chow<sup>2</sup>, Zhiya Su<sup>3</sup> & Omar Wani<sup>2</sup>

Extreme weather poses increasing challenges to urban transit systems, yet the resilience of subway ridership under such conditions remains insufficiently understood. This study develops an hour-specific vine copula framework for New York City subway ridership modeling, decomposing high-dimensional inter-station relationships into bivariate components while preserving non-linear and asymmetric dependencies. The methodology captures time-varying dependencies, generates realistic ridership distributions under diverse weather conditions, and enables quantitative assessment of ridership resilience to extreme events. Validation demonstrates strong performance, with 83 percent of scenarios achieving Kullback–Leibler divergence below 0.15. Results show a dynamic dependence structure across stations that varies under different environmental conditions. Results indicate that Manhattan core stations exhibit higher ridership resilience, whereas outer borough stations are more vulnerable. Heavy precipitation produces the most severe peak-hour impacts, while extreme cold primarily reduces off-peak ridership. For example, heavy precipitation during peak hours leads to a median 19.3 percent decline in Penn Station ridership (95% CI [−19.6, −3.4]), whereas extreme heat during off-peak hours reduces Broadway/Jackson Heights ridership by a median 14.8 percent (95% CI [−31.4, −12.7]). This framework provides a data-driven foundation for assessing ridership resilience and guiding climate adaptation and equitable transit investment in metropolitan systems.

Underground metro or subway systems serve as the backbone of urban transportation networks due to their high efficiency, large capacity, and operational reliability. These advantages make them essential for sustaining high-volume passenger movement in densely populated metropolitan areas. However, subway ridership demonstrates substantial spatial and temporal variation, shaped by heterogeneous urban contexts such as land use diversity, population density gradients, employment center distributions, and geographic constraints<sup>1,2</sup>. Importantly, ridership patterns across stations are not independent; instead, they exhibit complex and often nonlinear dependencies driven by geographic proximity, overlapping catchment areas, network connectivity, and coordinated temporal rhythms across diurnal, weekly, and seasonal cycles<sup>3</sup>. These inter-station dependencies are particularly critical during extreme weather events, when ridership responses differ due to variations in infrastructure robustness, surrounding built environments, and exposure to meteorological stressors<sup>4,5</sup>. Capturing and modeling these spatiotemporal dependencies, especially under perturbations from extreme conditions, is essential for accurate demand forecasting, resilient network design, and effective emergency preparedness.

Failing to account for such interdependencies may lead to misinformed operational decisions and increased vulnerability during periods of disruption or surging demand.

Subway ridership modeling has evolved significantly from traditional aggregate-level approaches to sophisticated station-level modeling techniques. Early methods relied on professional judgment, route comparisons, and regional travel demand models, but these approaches proved inadequate for capturing the spatial heterogeneity inherent in transit networks<sup>6,7</sup>. Linear regression models using ordinary least squares (OLS) became widely adopted for establishing relationships between ridership and influencing factors such as land use characteristics and transit accessibility<sup>1,2,8</sup>. However, these global modeling approaches assume uniform relationships across all stations, failing to account for spatial variations that characterize real-world transit systems<sup>9</sup>. Most existing research on transit ridership forecast modeling has concentrated on direct demand modeling using regression- or machine-learning-based methods. These approaches often treat transit stations as independent analytical units, thereby overlooking the

<sup>1</sup>Department of Civil and Environmental Engineering University of Louisville, Louisville, KY, USA. <sup>2</sup>Tandon School of Engineering New York University, Brooklyn, NY, USA.

<sup>3</sup>The University of Hong Kong, Hong Kong, China. ✉e-mail: [yueshuai.he@louisville.edu](mailto:yueshuai.he@louisville.edu)

complex multivariate dependencies that characterize real-world urban transit networks<sup>3,9</sup>.

To address these limitations, geographically weighted regression (GWR) models emerged to capture spatially varying relationships, showing improved prediction accuracy over traditional OLS approaches<sup>3,9–13</sup>. However, GWR models are susceptible to multicollinearity among local coefficients and fail to adequately model the complex dependencies between explanatory variables across different locations, which can undermine interpretability and robustness<sup>10,11</sup>. Recent advances in deep learning have shown promise for ridership prediction through sophisticated architectures including spatial-temporal frameworks, graph convolutional networks, and dynamic spatial-temporal representation learning approaches<sup>14–17</sup>. While promising, these methods often require extensive training data, are computationally intensive, and exhibit limited interpretability, factors that hinder their practical application in contexts where understanding the underlying dependency structures is critical. The black-box nature of deep learning models makes it difficult to extract meaningful insights about specific dependency structures between stations, particularly under varying operational and environmental conditions. These constraints are particularly problematic in the context of extreme weather analysis, where historical data on rare events are inherently sparse, making it difficult for data-hungry models to generalize effectively.

The transportation literature has traditionally approached weather-ridership relationships through individual station analysis, employing regression-based methodologies to quantify how meteorological variables affect ridership volumes at discrete locations. This paradigm conceptualizes transit stations as independent analytical units responding to weather stimuli in isolation, generating elasticity coefficients for aggregate system responses. Studies have documented negative relationships between precipitation and ridership demand, with varying sensitivities across station types and temporal periods<sup>4,5,18,19</sup>. However, this approach fundamentally overlooks the interconnected nature of urban transit networks, where passenger flows create complex interdependencies that evolve dynamically with changing conditions. Moreover, traditional correlation-based approaches<sup>20–22</sup> are ill-suited for capturing the non-linear, asymmetric, and temporally evolving dependency structures inherent in interconnected transit systems. This shortcoming becomes especially evident when dependencies vary under different operational and meteorological conditions. Such conditions demand more flexible dependency modeling techniques that can move beyond static, linear assumptions.

To address these limitations, this study proposes a copula-based framework for modeling spatio-temporal dependencies in station-level subway ridership, with a particular focus on assessing demand resilience under extreme weather conditions. Copula methods, grounded in Sklar's theorem<sup>23</sup>, enable the separation of joint distributions into marginal components and dependence structures, thereby facilitating flexible modeling of interdependencies while preserving the distributional characteristics of individual variables<sup>24,25</sup>. A Copula is a multivariate cumulative distribution function<sup>23</sup> where each marginal distribution is uniform on [0,1]. Formally, for random variables  $X_1, X_2, \dots, X_n$ , with cumulative distribution functions  $F_1(x), F_2(x), \dots, F_n(x)$ , their joint distribution can be expressed as:

$$F(X_1, X_2, \dots, X_n) = C_\theta(F_1(x), F_2(x), \dots, F_n(x)) \quad (1)$$

where  $C_\theta$  is the copula function that encapsulates the dependency structure between the variables.  $\theta$  represents the parameter vector governing the copula. This formulation allows the study to model dependencies regardless of the specific marginal distributions involved. The cornerstone of copula theory is Sklar's theorem<sup>23</sup>, which establishes that for any multivariate distribution function, there exists a unique copula that connects the individual marginal distributions to form the joint distribution. This means copulas are ideal for modeling the dependency relationships between subway station ridership patterns while preserving each station's individual ridership characteristics.

Copula-based modeling has been extensively applied across diverse disciplines, demonstrating its effectiveness in capturing complex dependencies that traditional statistical methods cannot adequately represent. Applications span financial risk management for systemic risk simulation and portfolio management, hydrology for environmental dependencies and flood risk assessment, machine learning for pattern recognition, and health sciences for disease comorbidity analysis<sup>26–28</sup>. In transportation research, copula-based methodology has gained traction for modeling joint decision processes across various applications. Studies have demonstrated copula effectiveness in analyzing combined travel choices, such as simultaneous mode selection and trip chaining decisions<sup>29</sup>, with comparative analyses showing superior model performance relative to traditional nested logit structures<sup>30</sup>. Researchers have applied copula frameworks to temporal decision interdependencies, including telecommuting behavior patterns<sup>31</sup> and coordinated timing-destination choices<sup>32,33</sup>. The methodology has extended to activity-duration relationships<sup>34</sup>, freight logistics applications examining shipment characteristics and modal selection<sup>35,36</sup>, and bus operational reliability modeling<sup>37</sup>. Recent developments have applied copulas to analyze bike-sharing and subway demand interdependencies<sup>38</sup> and joint mode change and choice decisions under transportation demand management policies<sup>39</sup>, with vine copulas specifically being used to model degradable segment capacities of metro networks for resilient bus network design<sup>40</sup>. These diverse applications demonstrate copulas' fundamental advantage in capturing complex behavioral interdependencies through flexible dependency structures while avoiding restrictive distributional assumptions that limit conventional choice modeling approaches.

The copula framework offers a powerful tool for transforming raw observations into relative usage levels, thereby enabling a more nuanced analysis of how interdependencies evolve under varying conditions. Its inherent sensitivity to tail dependencies allows for the detection of strong correlations that emerge during extreme events, an essential feature for modeling urban transit systems exposed to disruptions such as severe weather. This makes copulas particularly well-suited for capturing the complex, nonlinear, and time-varying relationships among subway stations, which are often shaped by both structural network properties and external environmental factors<sup>41,42</sup>. Such capabilities offer a rigorous foundation for evaluating variability and demand resilience across stations and time under diverse weather scenarios and provide a statistically grounded and interpretable framework for understanding subway demand dynamics and quantifying resilience under uncertain and extreme conditions<sup>40</sup>. However, despite the growing body of copula-based transportation research and interest in these methods, limited attention has been devoted to modeling inter-station ridership dependencies within transit networks under varying environmental conditions. Few studies have explored multivariate dependencies across multiple subway stations simultaneously. This represents a critical gap in the literature, one in which advanced frameworks such as vine copulas can offer greater flexibility and interpretability in modeling high-dimensional dependency structures under diverse operational and weather-related scenarios.

As city subway systems are complex and extensive, involving multiple stations with interdependent ridership patterns influenced by various operational and environmental factors, it is necessary to build joint probability distributions for multivariate ridership data. Bedford and Cooke<sup>41</sup> and Joe et al.<sup>42</sup> introduced a more advanced and flexible alternative method of constructing dependence structures called Vine Copula. At least  $n(n-1)/2$  bivariate copulas with free specification can be established between  $n$  given stations under this flexible structure. Vine copulas decompose high-dimensional dependencies into manageable sequences of bivariate copula relationships, significantly enhancing modeling flexibility for complex systems<sup>24</sup>. Bedford and Cooke<sup>41</sup> developed Regular vine (R-vine) structures as graphical modeling tools and introduced two specialized architectures: Canonical vines (C-vines) that organize dependencies around central hub variables, and Drawable vines (D-vines) that arrange variables in sequential path structures. Aas et al.<sup>25</sup> extended this methodology by providing comprehensive mathematical formulations, demonstrating that C-vines

accommodate star-shaped dependency networks while D-vines excel in modeling sequential relationships, establishing vine copulas as versatile tools for diverse multivariate modeling applications.

To contextualize the focus of this study, it is important to clarify how demand resilience fits within the broader resilience literature. Resilience refers to the ability of systems to resist, absorb, accommodate, and recover from hazards in a timely and efficient manner<sup>43</sup>. In infrastructure contexts, resilience encompasses multiple interconnected dimensions: vulnerability reflects the degree of performance degradation during disruption, survivability describes how gracefully systems transition from normal to disrupted states, response captures immediate actions taken to maintain functionality, and recovery measures the ability to return to original conditions<sup>44</sup>. In railway transport systems, resilience research has developed along two complementary dimensions. Supply-side studies focus on infrastructure robustness and operational performance, quantifying system degradation through service cancellations, capacity reductions, and disruption durations<sup>45,46</sup>. Demand-side research examines passenger behavioral responses, how travelers adjust their mobility patterns under adverse conditions, measured through changes in ridership volumes, passenger delays, and economic impacts<sup>47,48</sup>. This study focuses specifically on demand-side resilience, characterizing how ridership patterns respond to extreme weather rather than supply-side factors such as service disruptions or operational recovery.

The framework is demonstrated using the New York City (NYC) subway system as a case study. As one of the largest and most complex urban transit systems in the world, with approximately 3.4 million daily riders<sup>49</sup>, NYC offers a rich empirical setting for modeling and validating high-dimensional dependencies. Its extensive network, pronounced temporal ridership variations, heterogeneous station characteristics ranging from dense commercial hubs to residential neighborhoods, and exposure to diverse weather conditions collectively create a rigorous testbed for assessing demand resilience. While the analysis focuses on NYC, the methodology is broadly applicable to other urban transit systems with similar ridership heterogeneity and spatio-temporal dependencies.

The key contributions of this study are threefold. First, we develop an hour-specific vine copula-based modeling framework that decomposes high-dimensional inter-station dependencies into structured, tractable bivariate copulas, enabling flexible representation of nonlinear, asymmetric, and tail-dependent relationships that vary across different hours of the day. Second, we introduce a data synthesis method capable of generating realistic system-level ridership distributions by conditioning on observed historical patterns, leveraging the theoretical tail-dependency properties of vine copulas to model network-wide dependency structures under both typical and extreme weather conditions. Third, we demonstrate the effectiveness of the proposed framework through comprehensive validation and resilience assessment, showing how different extreme weather types differentially impact ridership patterns across peak and off-peak periods and across different station locations. Specifically, our work enables data-driven decision-making support through model-based distributional foundations for resilience assessment, where modeling the dependence structure between multiple stations and time periods during extreme weather conditions allows for better estimation of network vulnerability and compound weather risks. Our analysis quantifies demand-side vulnerability by characterizing how travelers adjust subway usage under adverse conditions, which informs capacity planning and passenger communication but should be interpreted alongside supply-side operational data for comprehensive resilience planning.

## Results

### Vine structure

Given the scale of the New York City subway system with 472 stations and the computational complexity of high-dimensional vine copula models, this study focuses on 10 strategically selected stations based on three criteria: ridership prominence<sup>2</sup>, spatial and functional non-redundancy, and ref. 3 borough-level representativeness. The selected stations (Fig. 1) include

Times Square, Penn Station, Grand Central, Union Square, Fulton St, Columbus Circle, Broadway/Jackson Heights, Flushing-Main St, Chambers St/WTC, and Atlantic Av-Barclays Center, distributed across Manhattan (seven stations), Queens (two stations), and Brooklyn (one station).

Figure 2 displays the R-vine Tree 1 structure at 7–8 AM and 8–9 AM as an example. Red edges indicate negative correlations, green edges indicate positive correlations, and line width represents the absolute value of Kendall's tau (wider lines correspond to stronger dependencies). Dashed lines represent within-hour relationships between stations, while solid lines represent across-hour relationships. High tau values indicate stronger bivariate dependencies: a positive tau value means that when ridership at one station increases, the connected station's ridership also tends to increase, while a negative tau value means the connected station's ridership tends to decrease. The results show that vine copula structures vary substantially across different hours, with differences in node connectivity patterns, selected copula families, and the strength and direction of dependencies between connected stations.

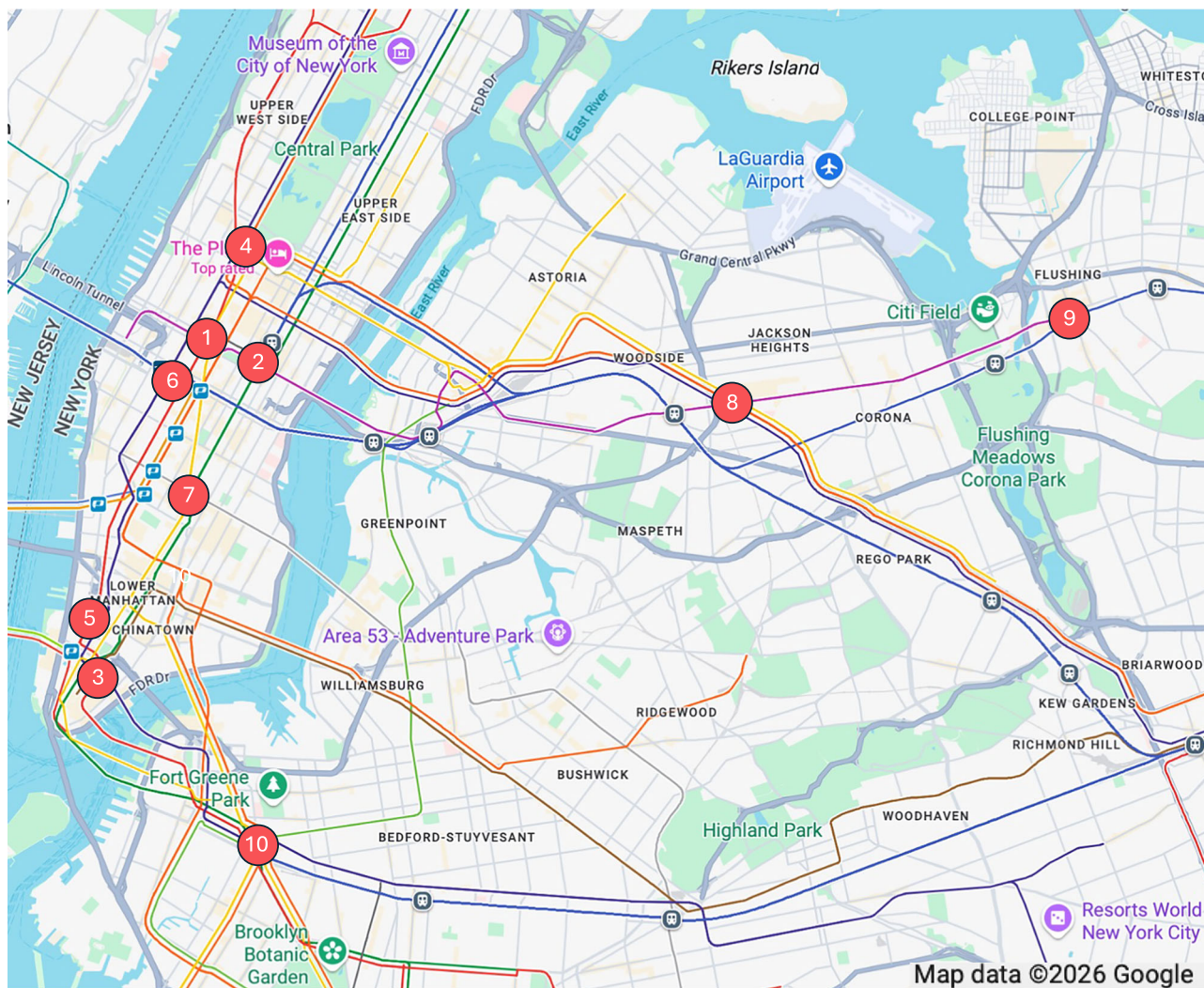
The Tree 1 structure differs between the two time periods, yet some consistent patterns emerge. For instance, within-hour relationships at 8 AM show similarities across both configurations. In both the 7–8 AM and 8–9 AM models, Columbus Circle at 8 AM connects with Atlantic Av-Barclays Center at 8 AM and Grand Central at 8 AM. Additionally, World Trade Center at 8 AM connects with Fulton Street at 8 AM, and Penn Station at 8 AM maintains connections in both plots. However, differences also appear. In the 7–8 AM structure, Union Square at 8 AM connects with Columbus Circle at 8 AM, whereas this connection is absent in the 8–9 AM structure. These variations show that while certain core spatial dependencies persist across adjacent time windows, the overall network structure adapts to reflect hour-specific ridership patterns.

Figure 2 also shows differences in network centrality. Grand Central at 8 AM has 4 edges in the 7–8 AM Tree 1 structure but 5 edges in the 8–9 AM structure, indicating that Grand Central at 8 AM becomes more interconnected during the 8–9 AM period. Stations with more edges serve as network hubs, meaning their ridership changes are more likely to influence or be influenced by changes at other stations. Table 1 identifies stations with four or more edges in Tree 1 across different hours, indicating which stations serve as network hubs during specific time periods. Grand Central demonstrates the most consistent hub behavior, with high centrality in 10 different hours throughout the day, including both morning (hours 7–8) and evening (hours 20–23) peaks. This sustained centrality suggests Grand Central functions as a persistent coordination point for network-wide ridership patterns. In contrast, other major stations like Times Square and Union Square show more time-specific hub behavior, with Times Square exhibiting higher centrality during off-peak hours and Union Square during afternoon periods. Notably, several stations including Chambers St/WTC, Penn Station, and Broadway/Jackson Heights never reach hub status ( $\geq 4$  edges), suggesting their ridership patterns are less central to the overall network dependency structure.

### Validation

After selecting optimal vine structures for each hour, synthetic data were generated for model validation using the ensemble averaging approach from Eq. 9. A total of 12 scenarios were selected, systematically varying temperature and precipitation conditions across AM peak (8–9 AM), PM peak (5–6 PM), and off-peak hours (12–1 PM) as shown in Fig. 3. Kullback–Leibler (KL) divergence quantifies distributional differences between observed and synthetic ridership patterns, where values below 0.2 indicate highly similar distributions. Mean Absolute Percentage Error (MAPE) evaluates point prediction accuracy, with lower values indicating better performance.

The vine copula model demonstrates strong performance across diverse weather conditions. Overall, 83% of scenarios achieve KL divergence below 0.15, indicating synthetic ridership distributions closely match observed patterns (Tables 2 and 3). MAPE maintains consistent accuracy with 91.7% of scenarios below 10% (overall mean: 6.83%). Performance



**Fig. 1 | Ten strategically selected NYC subway stations span three boroughs to ensure spatial and functional diversity.** Stations are distributed across Manhattan (Times Square-42nd St, Grand Central-42nd St, Fulton St, 59th St-Columbus Circle, Chambers St/World Trade Center (WTC)/Park Pl/Cortlandt, 34th St-Penn Station, 14th St-Union Square), Queens (74th St-Broadway/Jackson Heights-Roosevelt Av,

Flushing-Main St), and Brooklyn (Atlantic Av-Barclays Center), collectively representing major transportation hubs, commercial centers, and residential neighborhoods. Stations are referred to by shortened names (e.g., Times Square, Grand Central, Penn Station) in subsequent text and figures. Map data ©2026 Google.

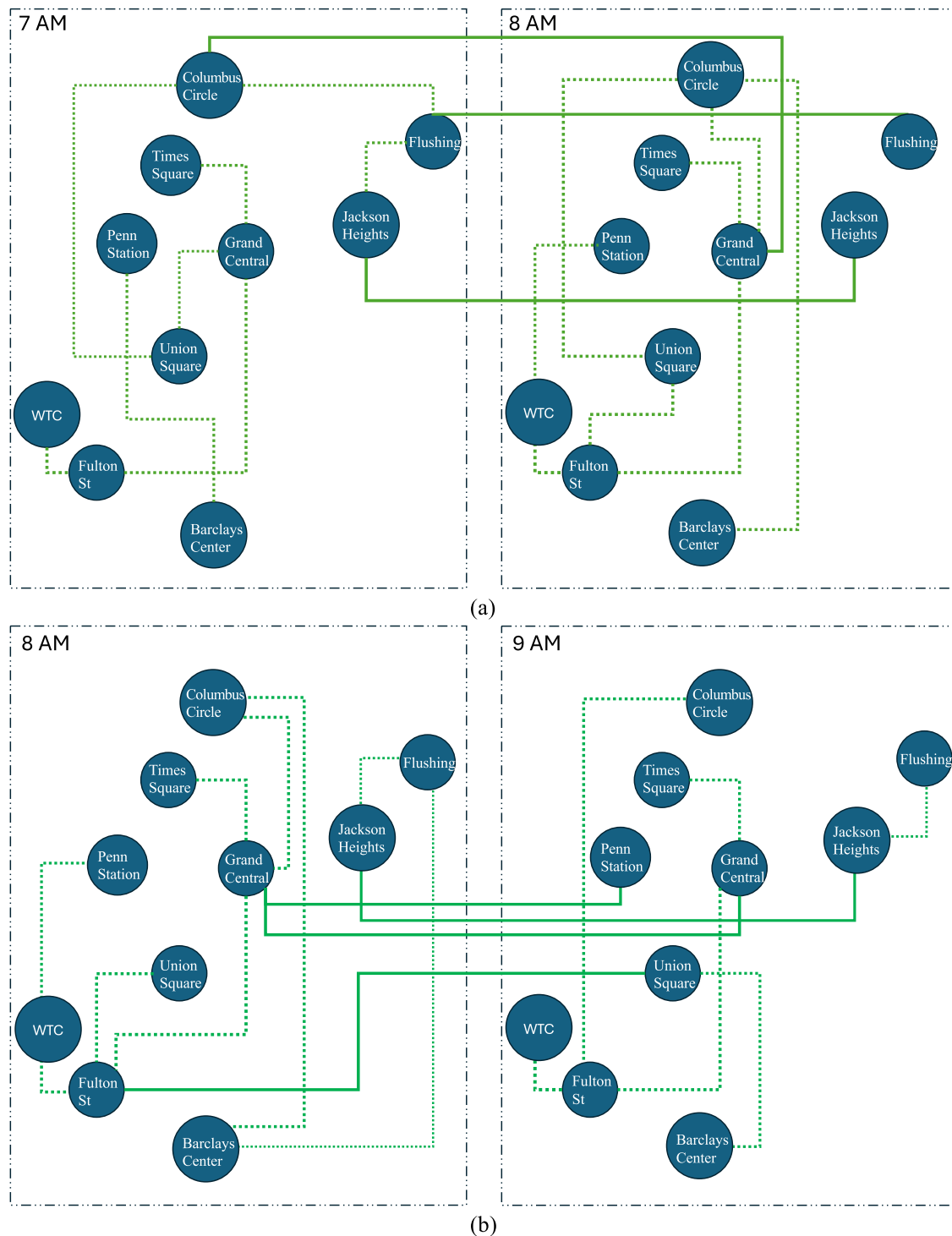
varies systematically across time periods: AM peak scenarios achieve the best fit (mean KL: 0.091, MAPE: 6.43%), and followed by PM peak (KL: 0.108, MAPE: 6.83%). Spatial patterns show Manhattan core stations (Chambers St/WTC, Fulton St, Penn Station) maintaining consistently excellent performance, while outer borough stations (Broadway/Jackson Heights, Flushing-Main St) exhibit higher KL divergence but comparable MAPE, suggesting the model captures mean ridership well but faces challenges replicating complete distributional characteristics. These results demonstrate the model’s capability to capture dynamic dependence structures across stations that vary under different environmental conditions.

**Ridership resilience assessment**

All reported ridership changes reflect realized demand under observed operating conditions and therefore represent conditional ridership resilience, capturing the combined effects of traveler response and contemporaneous service adjustments during extreme weather. To assess demand-side resilience, ridership patterns under extreme weather are compared against baseline conditions, quantifying how travelers adjust their subway usage in response to adverse weather across the network. Resilience assessment was performed across six conditions to compare ridership changes under different extreme weather conditions during both peak and

off-peak hours. The analysis examines three types of extreme weather, extreme cold (below  $-10\text{ }^{\circ}\text{C}$ ), extreme heat (above  $38\text{ }^{\circ}\text{C}$ ), and heavy precipitation (above  $5\text{ mm/h}$ ), each tested during both peak and off-peak periods as shown in Table 4. For each scenario, the vine copula model generates conditional synthetic ridership distributions that preserve both inter-station and temporal dependency structures. Extreme-weather distributions are synthesized by conditioning on observed meteorological states and lagged ridership, while baseline distributions are generated under normal weather conditions with other temporal characteristics held constant. This design isolates demand-side responses to weather stress from routine diurnal variation. Demand-side resilience is quantified using distributional change metrics rather than point estimates, allowing both the magnitude and uncertainty of ridership responses to be assessed. Median percentage change reflects central demand shifts, while predictive intervals characterize variability and stability across the network. This distribution-based framing recognizes that resilience is inherently probabilistic and heterogeneous across stations.

Although extreme weather observations are sparse, the copula-based approach reconstructs plausible demand responses by leveraging dependency structures learned from the full dataset, including that theoretically capture coordinated behavioral responses under rare conditions. The



**Fig. 2 | R-vine tree structures show time-varying inter-station dependencies during morning peak hours.** Structure of R-vine Tree 1 of 7–8 AM (a) and at 8–9 AM (b), Edge color: green = positive dependency, red = negative dependency. Line width = dependency strength ( $|\tau|$ ). Dashed = within-hour; solid = across-hour.

While certain core dependencies persist, Columbus Circle and Grand Central at 8 AM, overall connectivity patterns, copula families, and dependency strengths differ substantially between the two periods.

resulting uncertainty bounds therefore represent model-based predictive intervals rather than purely empirical estimates. While vine copulas are well suited for resilience analysis in low-frequency, high-impact contexts, the representation of tail behavior under sustained extremes remains an assumption embedded in the modeling architecture rather than a property directly validated with extensive observations.

The resilience estimates presented below represent model-based projections rather than empirical averages of repeatedly observed extremes. The

extreme weather scenarios in Table 4 have limited observational support: sustained extreme heat ( $>38\text{ }^{\circ}\text{C}$  for consecutive hours) appears only three times during the study period, sustained extreme cold ( $<-10\text{ }^{\circ}\text{C}$ ) four times, and heavy precipitation ( $>5\text{ mm/h}$  for consecutive hours) once. Given this scarcity, the vine copula framework integrates these few direct observations with dependence structures learned from the broader dataset, using tail-dependence behavior to conditionally generate network-wide responses under rare conditions. The resulting distributions represent model-

informed counterfactual system responses to sustained extremes, not descriptive summaries of frequently observed historical events. Uncertainty intervals therefore reflect variability across model-generated realizations rather than sampling variability from multiple past occurrences. These estimates are intended to support forward-looking resilience assessment under plausible but infrequently observed climate stressors and should be interpreted in light of the model assumptions and conditioning framework.

Table 5 presents the ridership change percentiles under extreme weather conditions across all stations for peak and off-peak hours, respectively. The results show substantial and systematic ridership reductions under adverse weather conditions, with all median values negative across all scenarios. However, the magnitude, variability, and predictability of impacts vary considerably across weather types, time periods, and station locations.

To examine whether the dependency network structure relates to resilience outcomes, network centrality metrics from the vine copula were linked to vulnerability patterns across six weather scenarios. Network centrality was measured through<sup>1</sup> degree, the number of direct dependency connections in Tree 1 of each bi-hour copula structure that each condition represents, and ref.<sup>2</sup> strength, the sum of absolute dependency coefficients (Kendall's tau) across these connections of each bi-hour copula structure that each condition represents. Resilience was characterized by median ridership decline (vulnerability magnitude) and confidence interval width (97.5th minus 2.5th percentile, capturing prediction uncertainty) for the six scenarios presented in Table 4. Relationships were assessed using Pearson correlation coefficient (r), which measures linear association strength, and Spearman's rank correlation coefficient (ρ), a non-parametric measure robust to non-linear relationships and outliers. Negative correlations would indicate that stations with higher network centrality experience smaller ridership declines or narrower confidence intervals (greater resilience and predictability), while positive correlations would suggest that highly

connected stations are more vulnerable or exhibit greater uncertainty. This approach tests whether network integration systematically predicts vulnerability magnitude and uncertainty across weather conditions and time periods.

As shown in Table 6, across the six weather scenarios, network centrality metrics showed varying associations with resilience measures, with

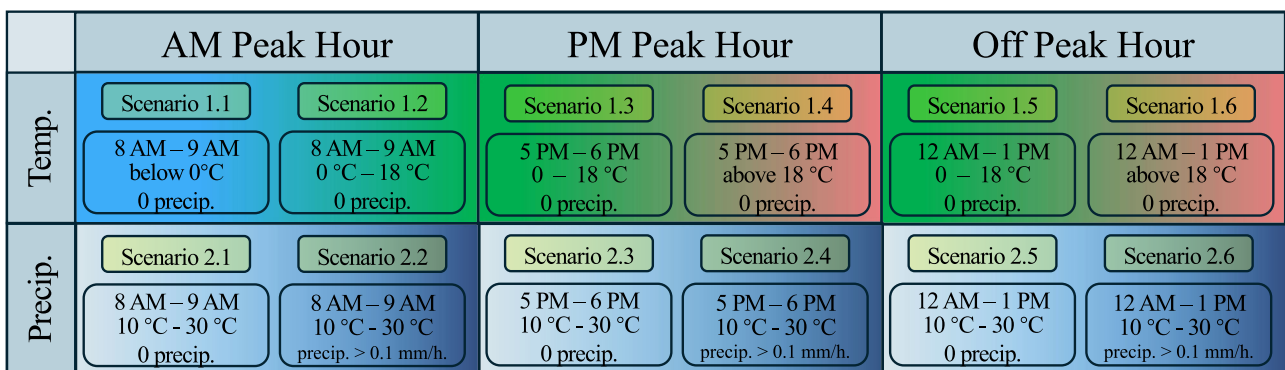
**Table 2 | KL divergence result of all scenarios**

| Temperature KL Divergence   |        |        |        |        |       |       |
|-----------------------------|--------|--------|--------|--------|-------|-------|
| Station Name                | 1.1    | 1.2    | 1.3    | 1.4    | 1.5   | 1.6   |
| Times Square                | 0.097  | 0.123  | 0.070  | 0.060  | 0.076 | 0.130 |
| Grand Central               | 0.078  | 0.092  | 0.079  | 0.067  | 0.084 | 0.127 |
| Fulton St                   | 0.040  | 0.111  | 0.074  | 0.067  | 0.085 | 0.124 |
| Columbus Circle             | 0.078  | 0.084  | 0.072  | 0.146  | 0.091 | 0.247 |
| Chambers St/WTC             | 0.060  | 0.081  | 0.086  | 0.087  | 0.095 | 0.148 |
| Penn Station                | 0.098  | 0.067  | 0.076  | 0.069  | 0.099 | 0.127 |
| Union Square                | 0.054  | 0.077  | 0.076  | 0.098  | 0.120 | 0.145 |
| Broadway/Jackson Heights    | 0.216  | 0.168  | 0.181  | 0.223  | 0.124 | 0.207 |
| Flushing-Main St            | 0.190  | 0.151  | 0.150  | 0.191  | 0.131 | 0.198 |
| Atlantic Av-Barclays Center | 0.108  | 0.071  | 0.114  | 0.147  | 0.189 | 0.193 |
| Precipitation KL Divergence |        |        |        |        |       |       |
| Station Name                | 2.1    | 2.2    | 2.3    | 2.4    | 2.5   | 2.6   |
| Times Square                | 0.1-12 | 0.0-57 | 0.0-71 | 0.1-01 | 0.080 | 0.123 |
| Grand Central               | 0.0-69 | 0.0-66 | 0.0-73 | 0.1-06 | 0.094 | 0.118 |
| Fulton St                   | 0.0-71 | 0.0-61 | 0.0-69 | 0.0-63 | 0.090 | 0.079 |
| Columbus Circle             | 0.0-59 | 0.0-88 | 0.0-99 | 0.0-76 | 0.145 | 0.107 |
| Chambers St/WTC             | 0.0-54 | 0.0-49 | 0.0-70 | 0.0-75 | 0.100 | 0.083 |
| Penn Station                | 0.0-68 | 0.0-80 | 0.0-61 | 0.1-00 | 0.090 | 0.104 |
| Union Square                | 0.0-64 | 0.0-36 | 0.0-74 | 0.0-75 | 0.107 | 0.134 |
| Broadway/Jackson Heights    | 0.1-60 | 0.1-47 | 0.2-38 | 0.2-39 | 0.149 | 0.119 |
| Flushing-Main St            | 0.1-62 | 0.1-02 | 0.1-55 | 0.2-45 | 0.111 | 0.107 |
| Atlantic Av-Barclays Center | 0.0-62 | 0.0-41 | 0.1-01 | 0.1-12 | 0.150 | 0.123 |

**Table 1 | Network Hub Stations (≥ 4 edges) by Hour in First Tree Structure**

| Station Name                | Hours (≥ 4 Edges in Tree 1)        |
|-----------------------------|------------------------------------|
| Times Square                | 1, 3, 4, 12, 13                    |
| Grand Central               | 0, 1, 7, 8, 10, 15, 20, 21, 22, 23 |
| Fulton St                   | 7, 8, 9, 14                        |
| Columbus Circle             | 7                                  |
| Union Square                | 14, 15, 16, 19                     |
| Flushing-Main St            | 5, 20                              |
| Atlantic Av-Barclays Center | 6                                  |

Hours (0–23) when station has ≥4 edges in Tree 1 (hub status). More hub hours indicate greater network centrality and connectivity.



**Fig. 3 | Twelve validation scenarios systematically vary temperature and precipitation conditions across three time periods.** Scenarios span AM peak (8–9 AM), PM peak (5–6 PM), and off-peak (12–1 PM) hours, crossing three weather types: cold,

hot, and precipitation, each compared against a normal weather baseline (10–30 °C, 0 mm/h). This design enables comprehensive evaluation of model performance across diverse operational and meteorological conditions. precip.: precipitation.

**Table 3 | MAPE Result of All Scenarios**

| Temperature MAPE (%)        |      |      |      |      |       |       |
|-----------------------------|------|------|------|------|-------|-------|
| Station Name                | 1.1  | 1.2  | 1.3  | 1.4  | 1.5   | 1.6   |
| Times Square                | 8.57 | 8.07 | 4.34 | 5.69 | 5.49  | 3.83  |
| Grand Central               | 5.82 | 8.63 | 6.32 | 8.32 | 7.46  | 4.85  |
| Fulton St                   | 6.86 | 6.29 | 7.78 | 7.21 | 6.18  | 5.05  |
| Columbus Circle             | 6.88 | 5.70 | 5.28 | 8.69 | 10.41 | 14.79 |
| Chambers St/WTC             | 8.32 | 7.89 | 7.75 | 6.98 | 7.60  | 6.96  |
| Penn Station                | 6.79 | 6.31 | 6.41 | 7.36 | 8.34  | 5.90  |
| Union Square                | 6.77 | 6.20 | 5.69 | 8.27 | 6.22  | 6.22  |
| Broadway/Jackson Heights    | 5.74 | 6.22 | 5.72 | 6.09 | 6.22  | 3.53  |
| Flushing-Main St            | 6.37 | 4.50 | 6.18 | 6.33 | 5.68  | 4.86  |
| Atlantic Av-Barclays Center | 8.57 | 5.42 | 5.91 | 7.96 | 5.43  | 5.85  |
| Precipitation MAPE (%)      |      |      |      |      |       |       |
| Station Name                | 2.1  | 2.2  | 2.3  | 2.4  | 2.5   | 2.6   |
| Times Square                | 7.49 | 6.87 | 4.26 | 5.43 | 4.50  | 5.51  |
| Grand Central               | 7.94 | 6.98 | 5.42 | 8.55 | 6.53  | 6.67  |
| Fulton St                   | 5.79 | 5.28 | 6.17 | 7.87 | 6.01  | 6.52  |
| Columbus Circle             | 5.30 | 6.54 | 6.14 | 7.85 | 12.06 | 14.58 |
| Chambers St/WTC             | 6.88 | 6.26 | 6.12 | 9.10 | 7.47  | 8.12  |
| Penn Station                | 6.77 | 6.06 | 5.23 | 6.84 | 6.67  | 6.92  |
| Union Square                | 5.79 | 5.91 | 5.44 | 8.53 | 6.29  | 8.51  |
| Broadway/Jackson Heights    | 6.08 | 5.97 | 6.28 | 9.39 | 4.56  | 5.87  |
| Flushing-Main St            | 4.67 | 4.03 | 6.61 | 9.92 | 4.89  | 8.00  |
| Atlantic Av-Barclays Center | 4.84 | 5.74 | 5.76 | 8.05 | 6.07  | 6.89  |

**Table 4 | Resilience assessment scenario description**

|          | Peak Hour                    |   | Non-Peak Hour                |   |   |   |
|----------|------------------------------|---|------------------------------|---|---|---|
| Temp.    | 1                            | 8 AM – 9 AM                                 | 4                            | 10 AM – 11 AM                                   |   |   |
|          |                              | Extreme Cold                                |                              | Baseline  | Extreme Cold                                | Baseline  |
|          |                              | Below -10 °C4 Observations                  | 0 °C – 30 °C467 Observations |   | Below -10 °C3 Observations                  | 0 °C – 30 °C491 Observations                    |
|          | 2                            | 4 PM – 5 PM                                 | 5                            | 12 AM – 1 PM                                    |   |   |
| Baseline |                              | Extreme Heat                                |                              | Baseline  | Extreme Heat                                |   |
|          | 0 °C – 30 °C486 Observations | Above 38 °C3 Observations                   |                              | 0 °C – 30 °C481 Observations                    | Above 38 °C1 Observations                   |   |
| Precip.  | 3                            | 4 PM – 5 PM                                 | 6                            | 12 AM – 1 PM                                    |   |   |
|          |                              | Baseline                                    |                              | Precipitation                                   | Baseline                                    | Precipitation                                   |
|          |                              | 0 °C – 30 °C with 0 precip.452 Observations |                              | 0 °C – 30 °C with >5 mm/h precip.1 Observations | 0 °C – 30 °C with 0 precip.438 Observations | 0 °C – 30 °C with >5 mm/h precip.2 Observations |

several notable negative correlations observed (Table 6). Degree centrality exhibited the strongest relationship during extreme cold non-peak hours, negatively correlated with median ridership decline (Pearson  $r = -0.76$ ,  $p = 0.01$ ; Spearman  $\rho = -0.62$ ,  $p = 0.05$ ). Similar negative associations emerged for extreme heat non-peak (Pearson  $r = -0.61$ ,  $p = 0.06$ ; Spearman  $\rho = -0.66$ ,  $p = 0.04$ ) and heavy precipitation non-peak (Pearson  $r = -0.56$ ,  $p = 0.09$ ; Spearman  $\rho = -0.58$ ,  $p = 0.09$ ). Strength metrics showed comparable patterns, with consistent negative associations with confidence interval width, suggesting that well-connected stations exhibit more predictable responses under stress. Peak-hour scenarios generally showed weaker correlations ranging from  $-0.12$  to  $-0.41$ .

While most correlations do not reach conventional levels of statistical significance ( $p < 0.05$ ) due to the small sample size ( $n = 10$ ), the directional

consistency across multiple scenarios and metrics supports a plausible mechanism: stronger inter-station dependencies may reflect demand redundancy and substitution capacity, which buffers ridership shocks under extreme weather. This analysis is exploratory and intended to show structural patterns rather than establish causal relationships. A larger station set and expanded temporal coverage would be required to formally test these mechanisms.

Comparing the same station across different extreme weather scenarios shows distinct patterns in weather sensitivity that differ markedly between peak and off-peak periods. During peak hours, heavy precipitation produces the most severe impacts with median reductions of  $-7.80\%$  to  $-28.87\%$  (as presented in Fig. 4), substantially exceeding extreme heat ( $-3.42\%$  to  $-9.24\%$ ) and extreme cold ( $-0.99\%$  to  $-2.41\%$ ). Precipitation also exhibits the widest confidence intervals, with some stations showing 97.5th

**Table 5 | Resilience assessment result across all conditions**

| $\Delta R_{i,t}$ (%) percentile | Extreme Cold (Peak) |       |       | Extreme Heat (Peak) |       |       | Heavy Precip. (Peak) |        |       |
|---------------------------------|---------------------|-------|-------|---------------------|-------|-------|----------------------|--------|-------|
|                                 | 0.025               | 0.500 | 0.975 | 0.025               | 0.500 | 0.975 | 0.025                | 0.500  | 0.975 |
| Times Square                    | -3.00               | -1.72 | 1.75  | -9.05               | -8.38 | -5.00 | -24.54               | -23.89 | 4.62  |
| Grand Central                   | -1.31               | -1.09 | 0.04  | -3.78               | -3.42 | -2.63 | -8.06                | -7.81  | 1.11  |
| Fulton St                       | -3.74               | -2.00 | 3.91  | -9.97               | -9.24 | -5.33 | -25.30               | -24.71 | -1.89 |
| Columbus Circle                 | -3.43               | -2.41 | 3.57  | -8.59               | -7.22 | -4.74 | -29.51               | -28.87 | 1.39  |
| Chambers St/WTC                 | -2.79               | -1.89 | 2.78  | -7.83               | -6.06 | -4.79 | -22.13               | -21.50 | -4.22 |
| Penn Station                    | -2.33               | -1.81 | 2.05  | -9.13               | -8.27 | -5.45 | -19.62               | -19.25 | -3.35 |
| Union Square                    | -1.61               | -0.99 | 0.39  | -4.36               | -3.90 | -2.98 | -10.58               | -10.24 | 1.51  |
| Broadway/Jackson Heights        | -2.94               | -2.00 | 1.27  | -8.09               | -6.92 | -6.48 | -21.69               | -20.35 | 1.37  |
| Flushing-Main St                | -3.55               | -2.37 | 5.05  | -9.43               | -7.86 | -6.00 | -26.86               | -26.36 | 0.14  |
| Atlantic Av-Barclays Center     | -3.01               | -1.80 | 3.23  | -8.85               | -7.03 | -5.31 | -21.73               | -21.38 | -0.98 |

| $\Delta R_{i,t}$ (%) percentile | Extreme Cold (Peak) |        |       | Extreme Heat (Peak) |        |        | Heavy Precip. (Peak) |        |        |
|---------------------------------|---------------------|--------|-------|---------------------|--------|--------|----------------------|--------|--------|
|                                 | 0.025               | 0.500  | 0.975 | 0.025               | 0.500  | 0.975  | 0.025                | 0.500  | 0.975  |
| Times Square                    | -8.19               | -7.68  | -6.60 | -8.70               | -6.90  | -6.20  | -8.38                | -6.63  | -5.65  |
| Grand Central                   | -4.98               | -4.13  | -3.07 | -9.92               | -9.28  | -8.77  | -10.47               | -9.45  | -9.16  |
| Fulton St                       | -9.22               | -8.75  | -7.58 | -15.78              | -13.82 | -12.67 | -16.65               | -14.42 | -13.10 |
| Columbus Circle                 | -7.97               | -6.53  | -4.82 | -9.71               | -8.05  | -7.08  | -9.72                | -7.96  | -7.15  |
| Chambers St/WTC                 | -7.51               | -6.50  | -5.13 | -9.27               | -7.70  | -6.18  | -9.09                | -7.41  | -6.05  |
| Penn Station                    | -9.50               | -8.62  | -7.70 | -11.74              | -10.94 | -9.93  | -12.66               | -11.35 | -10.35 |
| Union Square                    | -7.12               | -4.12  | -2.39 | -5.65               | -5.14  | -3.95  | -6.03                | -5.19  | -4.34  |
| Broadway/Jackson Heights        | -6.79               | -6.18  | -5.00 | -31.35              | -14.84 | -12.68 | -31.03               | -17.74 | -13.92 |
| Flushing-Main St                | -11.24              | -10.14 | -8.67 | -14.33              | -11.98 | -9.10  | -14.20               | -11.88 | -8.12  |
| Atlantic Av-Barclays Center     | -8.93               | -8.01  | -7.00 | -11.36              | -9.55  | -6.95  | -11.41               | -9.05  | -7.03  |

$\Delta R_{i,t}$  (%) percentile represents percentage ridership change under extreme weather vs. baseline (0–30°C, 0 mm/h precipitation), calculated using quantile-matching where sorted extreme and baseline distributions are compared rank-wise (Eq. 10). Percentiles (2.5th, 50th, 97.5th) from 100 synthetic samples represent the distribution of impacts. Negative values indicate ridership reductions.

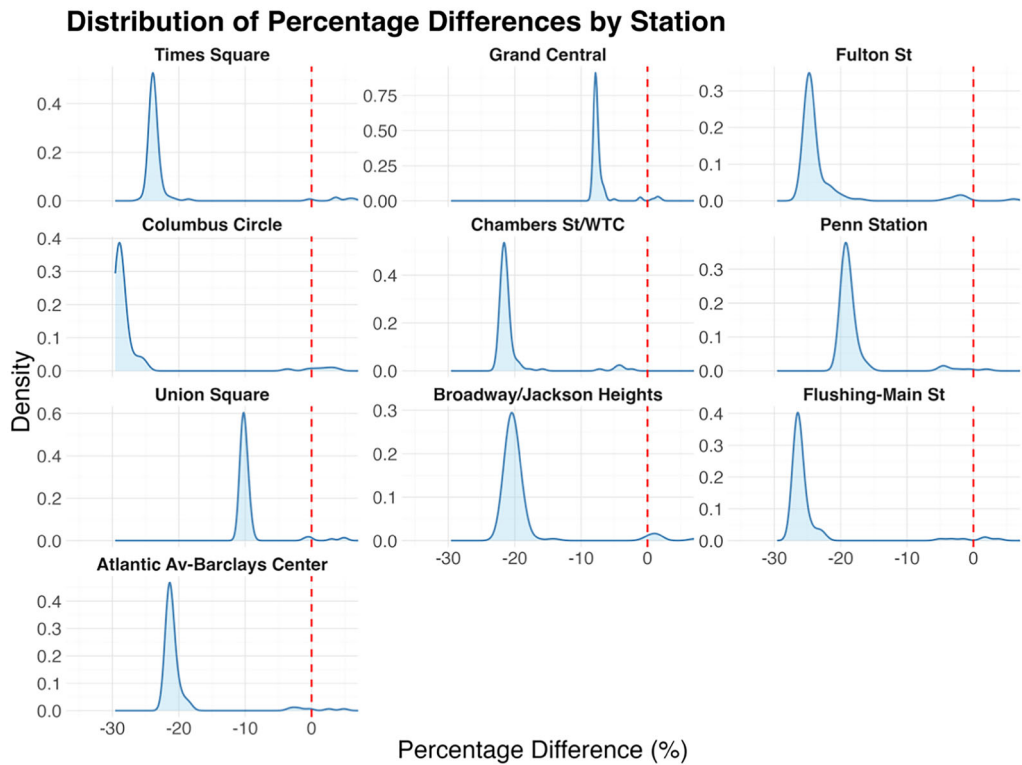
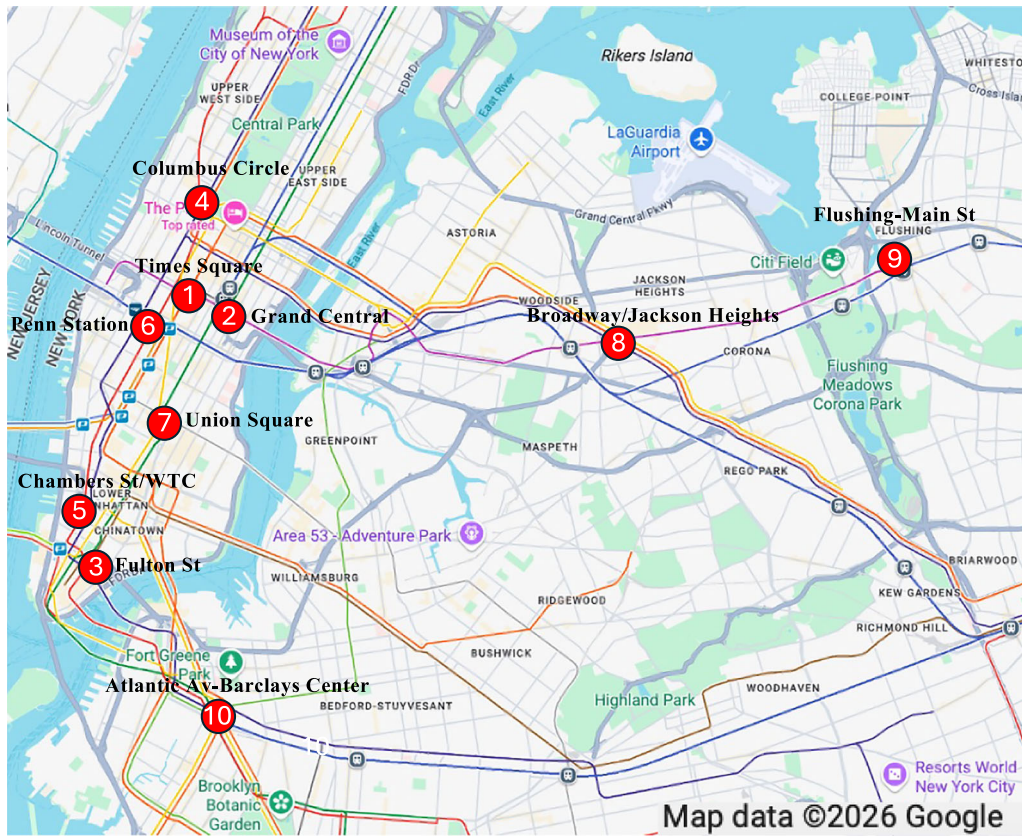
**Table 6 | Hub centrality effects on resilience result**

|                        | Degree            |                     |                         |                     |
|------------------------|-------------------|---------------------|-------------------------|---------------------|
|                        | Median $\Delta R$ | $\Delta R$          | $\Delta R$ 95% Interval |                     |
| <b>Conditions</b>      | Pearson $r$ (p)   | Spearman $\rho$ (p) | Pearson $r$ (p)         | Spearman $\rho$ (p) |
| Extreme Cold Peak      | -0.25 (0.48)      | -0.37 (0.31)        | -0.41 (0.24)            | -0.36 (0.31)        |
| Extreme Heat Peak      | -0.402 (0.25)     | -0.21 (0.58)        | -0.12 (0.65)            | 0.09 (0.96)         |
| Heavy Previp. Peak     | -0.35 (0.32)      | -0.26 (0.47)        | -0.27 (0.45)            | -0.27 (0.45)        |
| Extreme Cold Non-Peak  | -0.76 (0.01)      | -0.62 (0.05)        | -0.47 (0.17)            | -0.21 (0.56)        |
| Extreme Heat Non-Peak  | -0.61 (0.06)      | -0.66 (0.04)        | -0.50 (0.14)            | -0.63 (0.12)        |
| Heavy Previp. Non-Peak | -0.56 (0.09)      | -0.58 (0.09)        | -0.54 (0.1)             | -0.56 (0.14)        |
|                        | Strength          |                     |                         |                     |
| Extreme Cold Peak      | -0.26 (0.47)      | -0.23 (0.52)        | -0.41 (0.24)            | -0.35 (0.32)        |
| Extreme Heat Peak      | -0.347 (0.33)     | -0.17 (0.64)        | -0.17 (0.75)            | -0.19 (0.61)        |
| Heavy Previp. Peak     | -0.18 (0.62)      | -0.32 (0.38)        | -0.25 (0.49)            | -0.29 (0.42)        |
| Extreme Cold Non-Peak  | -0.61 (0.12)      | -0.59 (0.07)        | -0.50 (0.14)            | -0.27 (0.45)        |
| Extreme Heat Non-Peak  | -0.45 (0.21)      | -0.58 (0.08)        | -0.50 (0.15)            | -0.64 (0.05)        |
| Heavy Previp. Non-Peak | -0.41 (0.23)      | -0.50 (0.15)        | -0.53 (0.11)            | -0.58 (0.19)        |

percentiles above 0%, suggesting high uncertainty and diverse rider responses, some delay or cancel trips while others maintain their commute patterns, leading to high variability in aggregate impacts. During off-peak hours, impacts become more uniform across all weather types (medians: -4% to -14%) with consistently negative confidence intervals, suggesting that discretionary off-peak travel is similarly suppressed by any extreme weather condition. Comparing peak versus off-peak patterns shows that

extreme cold impacts off-peak hours more severely (medians: -4% to -10%) than peak hours (medians: -1% to -2%), while precipitation impacts peak hours more severely (medians: -20% to -29%) than off-peak hours (medians: -6% to -14%), with extreme heat showing similar impacts across both periods.

These patterns show that ridership responses to extreme weather are strongly modulated by both temporal context and the underlying nature of



**Fig. 4 | Heavy precipitation produces severe and spatially heterogeneous ridership declines across stations during peak hours. a** Location of the 10 selected stations. **b** Distribution of percentage ridership change per station under heavy precipitation during peak hours; the red dashed line indicates zero percentage

change. Median declines range from  $-7.81\%$  at Grand Central to  $-28.87\%$  at Columbus Circle, with several stations showing 97.5th percentile values above zero, reflecting diverse behavioral responses among peak-hour travelers. Map data ©2026 Google.

travel demand. Extreme cold leads to more pronounced reductions during off-peak hours, reflecting the predominance of discretionary trips in these periods. Off-peak travelers, engaged in activities such as shopping, recreation, or social visits, possess greater flexibility in timing and mode choice and are more likely to postpone or forgo trips in cold conditions. In contrast, peak-hour commuters tend to exhibit habitual, schedule-driven travel behavior, maintaining subway usage despite thermal discomfort. The relatively smaller reductions during peak periods thus reflect the inelasticity of essential work and school trips to cold weather.

Extreme heat produces ridership reductions that are comparable between peak and off-peak hours, with slightly greater declines observed during peak periods. Unlike cold weather, which primarily discourages discretionary off-peak trips, heat affects travelers more uniformly because thermal discomfort persists throughout the day and across trip purposes. High ambient temperatures elevate perceived travel costs for both commuters and discretionary travelers, particularly in older subway stations or lines with limited ventilation and air conditioning. The slightly larger reductions during peak hours likely stem from compounded discomfort caused by crowding and heat accumulation within stations and train cars, which are most severe during rush periods. Furthermore, heat-related service slowdowns, such as reduced train speeds or longer headways, can amplify travel burdens during already congested peak periods. Consequently, while extreme heat broadly suppresses ridership across all hours, its marginally greater impact during peak times underscores the interaction between thermal discomfort, crowding, and constrained system capacity.

Precipitation, however, demonstrates the opposite pattern, with stronger ridership declines during peak hours. This outcome arises because peak periods coincide with the highest passenger volumes and most constrained access conditions. Precipitation exacerbates first- and last-mile challenges, such as longer walking times, limited shelter availability, and crowding at station entrances, amplifying perceived and actual inconvenience for commuters. Moreover, the subway system operates near capacity during peak hours, leaving little tolerance for weather-induced slowdowns, increased dwell times, or congestion at platforms. As a result, precipitation disproportionately disrupts peak-hour travel, while off-peak riders, who experience less crowding and have more flexibility, are less affected.

Examining performance across different stations within the same weather scenario shows distinct spatial patterns in weather vulnerability. Grand Central and Union Square demonstrate the greatest resilience across all scenarios, with peak impacts ranging from  $-1.09\%$  to  $-10.24\%$ . In contrast, Columbus Circle exhibits extreme vulnerability specifically to precipitation during peak hours (median:  $-28.87\%$ ), Flushing-Main St shows consistently high impacts across all conditions (peak:  $-2.37\%$  to  $-26.36\%$ ; off-peak:  $-8.67\%$  to  $-11.98\%$ ), and Fulton St and Penn Station show off-peak vulnerability (medians:  $-14.42\%$  and  $-11.35\%$ ). Broadway/Jackson Heights presents extreme variance during off-peak heat and precipitation scenarios, with 2.5th percentile reaching  $-31\%$  despite median impacts of only  $-14\%$  to  $-17\%$ . This wide lower tail indicates model-generated distributions with nontrivial probability mass assigned to severe but infrequent outcomes, consistent with a stress-testing interpretation rather than repeated empirical occurrence.

Manhattan core stations (Grand Central, Union Square, Chambers St/WTC) generally exhibit greater resilience compared to outer borough stations (Broadway/Jackson Heights, Flushing-Main St, Atlantic Av-Barclays Center), likely due to shorter travel distances, better connectivity, and indoor station environments that provide weather protection. However, this pattern is more nuanced than simple geographic dichotomy. While the two most resilient stations are Manhattan core hubs (Grand Central at  $-7.80\%$  and Union Square at  $-10.24\%$ ), the most vulnerable station is also in Manhattan (Columbus Circle at  $-28.87\%$ ), demonstrating that location alone does not determine resilience. Moreover, controlling for ridership level shows no consistent Manhattan advantage: Flushing-Main St showed  $-26.36\%$  impact compared to Columbus Circle at  $-28.87\%$ . These comparisons suggest vulnerability reflects station-specific infrastructure

characteristics, such as platform configuration, covered access, and ventilation, and network position, particularly multimodal connectivity at Grand Central and Penn Station rather than borough location alone. These findings indicate that network resilience under extreme weather depends on both infrastructure characteristics and behavioral patterns of different rider populations, with spatial patterns best interpreted as characterizing high-ridership station behavior rather than system-wide generalizations.

## Discussion

Modern urban transit systems exhibit intricate interdependencies shaped by temporal variation, spatial configuration, weather dynamics, and network topology. Traditional modeling approaches that treat stations as independent entities fail to capture these complex, multivariate relationships<sup>1–5</sup>. This study addresses this gap by developing a spatio-temporal vine copula-based framework to model inter-station ridership dependencies using New York City subway data. The proposed approach decomposes high-dimensional dependency structures into tractable bivariate components, effectively preserving non-linear, asymmetric, and time-varying relationships across the network.

This research contributes to the literature in three key aspects. First, it introduces an hour-specific vine copula framework that captures dynamic inter-station dependencies, showing that ridership correlations are strongest within the same hour or between consecutive hours, accounting for over 95% of all significant dependencies. Second, an ensemble averaging synthesis approach is proposed to generate realistic ridership distributions under diverse weather conditions, providing a basis for demand resilience assessment. The vine copula's theoretical capacity to model tail dependencies enables reconstruction of plausible ridership distributions under extreme weather conditions even when empirical data are scarce, though this tail behavior represents a design feature of the copula framework rather than an empirically validated property. Third, comprehensive validation across 12 weather scenarios demonstrates strong model performance, with 83% of cases achieving Kullback–Leibler divergence below 0.15 and 92% achieving MAPE below 10%.

Key findings show systematic patterns in ridership resilience across spatial and temporal dimensions. Manhattan core stations, including Grand Central, Fulton St, and Penn Station, demonstrate superior resilience with consistently low distributional differences and prediction errors, indicating well-captured spatial-temporal dependencies and regular commute patterns. These stations' excellent performance reflects their roles as major transfer hubs with multiple line connections and indoor environments that provide weather protection. In contrast, outer borough stations Broadway/Jackson Heights and Flushing-Main St exhibit higher distributional divergence while maintaining comparable point prediction accuracy, suggesting that while the model captures mean ridership well, the complete distributional characteristics, potentially including heavier tails, multimodality, and greater variability, present additional modeling challenges due to longer travel distances, transfer dependencies, and more heterogeneous passenger demographics.

Weather-related resilience patterns show differential vulnerabilities across conditions and time periods. Extreme cold shows larger impacts during off-peak hours compared to peak hours, as peak commuters maintain travel patterns due to necessity while off-peak travelers avoid discretionary trips. Conversely, precipitation impacts peak hours more severely than off-peak hours, suggesting that precipitation creates compounding challenges during high-volume periods. Among all extreme weather types during peak hours, heavy precipitation produces the most severe impacts, substantially exceeding both extreme heat and extreme cold. These findings demonstrate that network resilience depends not only on infrastructure characteristics such as station connectivity and weather protection but also on land use patterns, travel purposes, and the distinct behavioral responses of different rider populations to varying extreme weather conditions.

This research provides practical value for resilient transportation planning by offering actionable insights into service coordination, emergency preparedness, and resource allocation. By characterizing demand-side responses to extreme weather, the framework informs capacity planning and passenger communication strategies, though comprehensive

**Table 7 | Kendall’s tau comparison between full joint conditioning synthesis and ensemble averaging synthesis**

| Station A     | Station B     | Full Joint Kendall’s tau | Ensemble Averaging Kendall’s tau | Absolute Difference |
|---------------|---------------|--------------------------|----------------------------------|---------------------|
| Time Square   | Grand Central | 0.329                    | 0.288                            | 0.04                |
| Grand Central | Fulton St     | 0.264                    | 0.423                            | 0.16                |
| Fulton St     | Time Square   | 0.307                    | 0.291                            | 0.02                |

Kendall’s tau measures association strength between station pairs (range: -1 to +1). Comparison based on 100 synthetic samples using full joint conditioning versus ensemble averaging for 3 stations. Mean absolute difference of 0.072 indicates mild dependency smoothing, with all correlation signs preserved.

**Table 8 | Robustness to Ensemble Averaging Result**

| Category                                    | Metric                                   | Ensemble Averaging        | Condition on Grand Central | Condition on Jackson Height |
|---|--|---------------------------|----------------------------|-----------------------------|
| Weather Hierarchy (Peak Hour)               | Extreme Cold Mean Median $\Delta R$ (%)  | -1.81                     | -2.43                      | -8.55                       |
|   | Extreme Heat Mean Median $\Delta R$ (%)  | -6.83                     | -2.58                      | -3.59                       |
|   | Heavy Precip. Mean Median $\Delta R$ (%) | -20.44                    | -21.78                     | -15.21                      |
|   | Ranking (Absolute Value)                 | Precip. > Heat > Cold     | Precip. > Heat > Cold      | Precip. > Cold > Heat       |
| Spatial Patterns (Heavy Precip., Peak Hour) | Manhattan Mean Median $\Delta R$ (%)     | -19.46                    | -20.73                     | -14.91                      |
|   | Outer Borough Mean Median $\Delta R$ (%) | -22.70                    | -24.23                     | -15.4                       |
|   | Ranking (Absolute Value)                 | Manhattan < Outer Borough | Manhattan < Outer Borough  | Manhattan < Outer Borough   |

$\Delta R$  (%) represents percentage ridership change under extreme weather vs. baseline (0-30°C, 0 mm/h precipitation), calculated using quantile-matching where sorted extreme and baseline distributions are compared rank-wise. The higher the absolute value, the greater the weather impact on ridership patterns.

resilience planning requires integration with supply-side operational data on service disruptions and recovery trajectories. The synthetic data generation capability is especially valuable for rare but high-impact events where observed data are limited, allowing transit agencies to simulate system-wide responses to compound weather scenarios and test intervention strategies before implementation. Through resilience assessment using these synthetic distributions, weather-related findings highlight the need for differentiated planning strategies: precipitation requires enhanced peak-hour service capacity and operational redundancy, while extreme cold necessitates off-peak service adjustments and passenger comfort measures. The identification of vulnerable stations, such as Columbus Circle’s precipitation sensitivity and outer borough stations’ consistent challenges enables targeted infrastructure improvements, enhanced weather protection, and focused resource deployment during adverse conditions.

The ensemble averaging synthesis strategy represents a necessary computational approximation when full joint conditioning becomes infeasible for high-dimensional networks. To assess whether this averaging introduces artifacts through dependency smoothing, we conducted validation comparing ensemble averaging against full joint conditioning using a 3-station subset for the hours 8 and 9 prediction task.

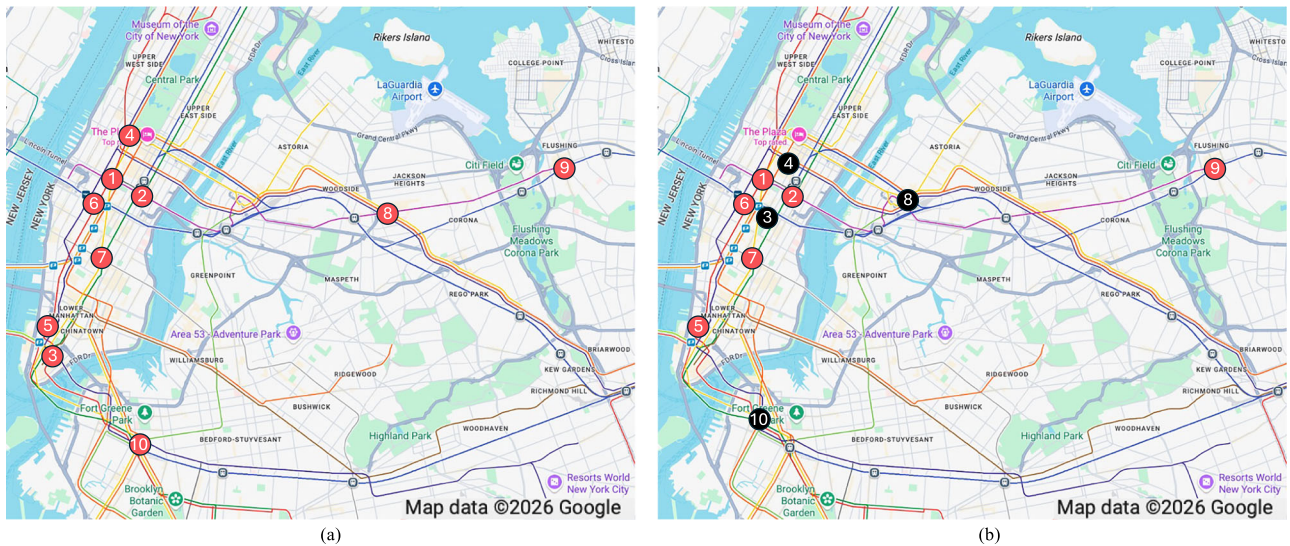
Full joint conditioning simultaneously conditions on all hour 8 stations to predict hour 9 ridership, representing the theoretically ideal approach that preserves the complete dependency structure. In contrast, ensemble averaging conditions on each station separately and averages the resulting predictions, reducing complexity while leveraging established principles that averaging reduces prediction variance and mitigates single-station conditioning bias. To quantify dependency preservation, we compared Kendall’s tau coefficients which measure rank correlation strength and direction between station pairs across 100 synthetic samples from both methods. Results shown mean absolute tau difference of 0.073 as shown in Table 7, indicating mild deviation in dependency strength between the two approaches. Critically, all correlation signs were preserved, indicating that fundamental co-movement patterns remain intact despite moderate smoothing of dependency magnitudes. Stations exhibiting positive dependencies under full joint conditioning maintained positive dependencies under ensemble averaging, though with attenuated strength. These findings demonstrate the approximation quality is adequate for our analytical

purposes. While the method does not perfectly replicate full joint conditioning, it maintains the essential dependency structure while enabling analysis of the 10-station network where full joint conditioning would be entirely infeasible.

To assess whether ensemble averaging introduces systematic bias, the peak-hour resilience assessment was re-run using single-station conditioning. Two anchoring cases were examined: Grand Central (highest network centrality, 10 hub hours) and Jackson Heights (zero hub hours). The single station results largely preserve the primary resilience signals observed under ensemble conditioning as shown in Table 8. In all cases, the spatial contrast remains stable, with Manhattan stations consistently showing greater resilience than outer-borough stations under heavy precipitation, and precipitation producing the most pronounced ridership declines relative to thermal extremes.

However, single-station conditioning introduces substantial variability in magnitude and secondary ordering. For example, the relative ranking of extreme heat versus extreme cold reverses under Jackson Heights conditioning, and absolute effects vary notably across anchors (e.g., extreme cold ranging from -1.81% under ensemble conditioning to -8.55% under Jackson Heights). This instability reflects the fact that single-station conditioning propagates network responses from a single temporal anchor, making estimates sensitive to station-specific dynamics and sampling noise. Ensemble conditioning, by contrast, aggregates information across multiple feasible station states and therefore functions as a variance-reduction estimator of network response. This pooling mitigates anchor dependence while preserving core structural signals, consistent with the Kendall’s  $\tau$  diagnostics (Table 7) showing retained dependency directionality with moderate smoothing of magnitudes.

Importantly, extreme weather represents a system-wide perturbation rather than a station-specific shock. Ensemble conditioning therefore aligns more closely with the research interest: network-level resilience, by approximating the expected response across plausible system states, whereas single-station conditioning is better interpreted as a sensitivity diagnostic reflecting localized propagation scenarios. The robustness results indicate that ensemble averaging stabilizes estimates without altering the principal resilience conclusions, while single-station synthesis remains more stochastic due to anchor dependence.



**Fig. 5 | Original and alternative station configurations enable sensitivity analysis of key findings to station selection. a** Original 10-station selection spanning Manhattan, Queens, and Brooklyn. **b** Alternative selection retaining six original stations (red markers) and replacing four with Herald Square, Rockefeller Center, Queens Plaza/Queensboro Plaza, and Jay St-MetroTech (black markers). The alternative set preserves geographic diversity while testing whether weather hierarchy and spatial vulnerability patterns hold under a different network configuration. Map data ©2026 Google.

Queens Plaza/Queensboro Plaza, and Jay St-MetroTech (black markers). The alternative set preserves geographic diversity while testing whether weather hierarchy and spatial vulnerability patterns hold under a different network configuration. Map data ©2026 Google.

**Table 9 | Station selection sensitivity analysis results**

|  | Original Set              | Alternative Set           |
|--|---------------------------|---------------------------|
| <b>Weather Hierarchy (Peak Hour)</b>               |                           |                           |
| Extreme Cold Mean Median $\Delta R$ (%)            | -1.81                     | -3.15                     |
| Extreme Heat Mean Median $\Delta R$ (%)            | -6.83                     | -4.90                     |
| Heavy Precip. Mean Median $\Delta R$ (%)           | -20.44                    | -19.86                    |
| Ranking (Absolute Value)                           | Precip. > Heat > Cold     | Precip. > Heat > Cold     |
| <b>Spatial Patterns (Heavy Precip., Peak Hour)</b> |                           |                           |
| Manhattan Mean Median $\Delta R$ (%)               | -19.46                    | -18.84                    |
| Outer Borough Mean Median $\Delta R$ (%)           | -22.70                    | -24.31                    |
| Ranking (Absolute Value)                           | Manhattan < Outer Borough | Manhattan < Outer Borough |

$\Delta R$  (%) represents percentage ridership change under extreme weather vs. baseline (0–30 °C, 0 mm/h precipitation), calculated using quantile-matching where sorted extreme and baseline distributions are compared rank-wise. The higher the absolute value, the greater the weather impact on ridership patterns.

To assess whether key findings are robust to station selection, a sensitivity analysis was conducted using an alternative 10-station set. The alternative selection begins with the top 5 highest-ridership stations: Times Square, Grand Central, Penn Station, Union Square, and Herald Square (previously excluded due to proximity to Penn Station). Given the spatial proximity between Fulton St and Chambers St/WTC in the original selection, only Chambers St/WTC was retained. The remaining stations include Rockefeller Center (replacing Columbus Circle), Queens Plaza/Queensboro Plaza (replacing Jackson Heights), Jay St-MetroTech (replacing Atlantic Av-Barclays Center), and Flushing-Main St as shown in Fig. 5.

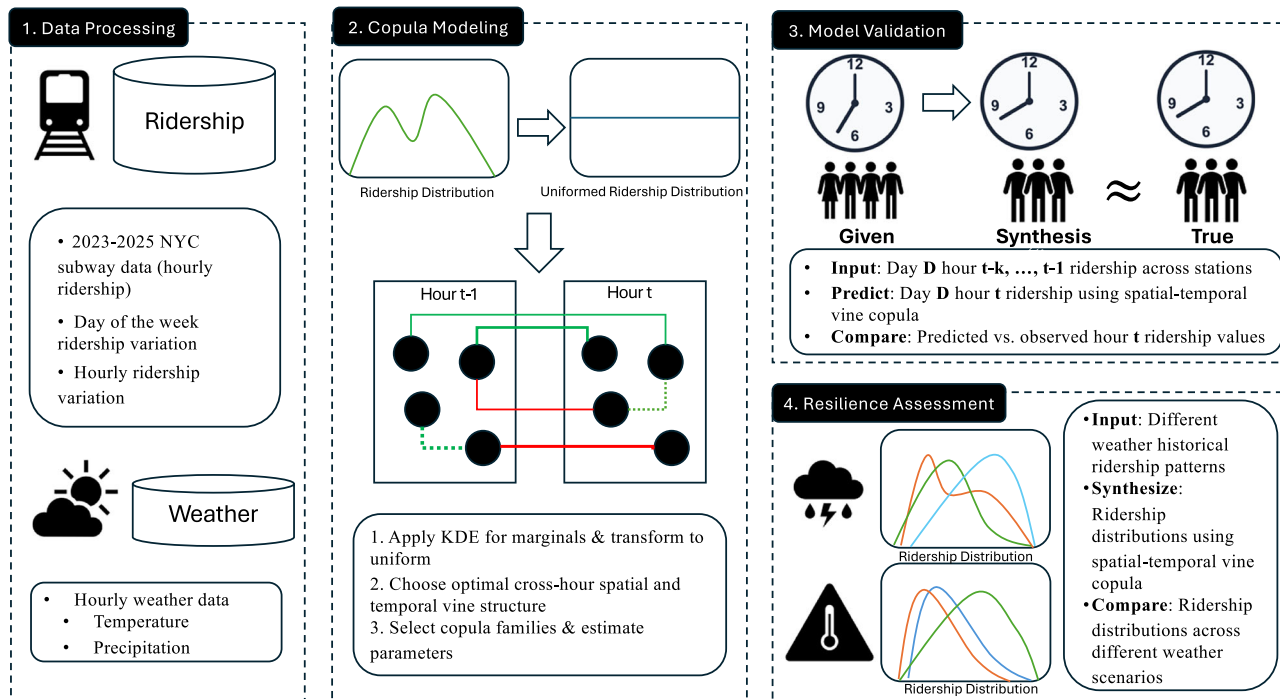
The sensitivity analysis using an alternative 10-station configuration demonstrates that key qualitative findings are robust to station selection. The weather hierarchy during peak hours remains consistent across both station sets, with heavy precipitation producing the most severe impacts (original: -20.9%, alternative: -19.86%), substantially exceeding extreme heat (original: -7.91%, alternative: -4.9%) and extreme cold (original: -2.77%, alternative: -3.15%) as shown in Table 9. The Precipitation > Heat

> Cold ranking is preserved despite substituting four stations. Similarly, spatial vulnerability patterns remain consistent, with outer borough stations showing systematically higher impacts than Manhattan stations under heavy precipitation during peak hours. While absolute magnitudes vary moderately between station sets reflecting genuine heterogeneity in station characteristics, the preservation of both weather hierarchies and spatial rankings across different network configurations indicates that these patterns represent system-wide behavioral responses rather than artifacts of the specific stations analyzed. This robustness supports confidence in the generalizability of findings to other high-ridership stations within the NYC subway system.

This work acknowledges several limitations. Due to the exponential growth in model complexity with network size, we limited our analysis to 10 strategically chosen stations. This selection introduces potential biases that affect generalizability. High-ridership stations typically benefit from superior infrastructure maintenance, more frequent service, and greater operational priority during disruptions, meaning our resilience estimates likely represent upper bounds compared to peripheral, lower-volume stations. The natural Manhattan concentration in top-ridership stations limits outer-borough representation. Consequently, findings characterize resilience patterns among the system’s highest-volume hubs rather than complete network behavior, and vulnerability estimates may underestimate impacts at peripheral stations with limited infrastructure investment and single-line service. As computational capabilities advance, scaling this framework to the full subway network and to other cities can enhance generalizability and uncover broader mobility insights.

The weather data approach uses a single citywide reference point (40.7128°N, 74.0060°W), which does not capture potential microclimate variations such as temperature differences between waterfront locations (e.g., Atlantic Avenue-Barclays Center) and inland stations, or localized precipitation intensity variations. However, this limitation is unlikely to substantially affect findings because the analysis focuses on region-wide extreme weather thresholds (temperature below -10 °C, above 38 °C, precipitation exceeding 5 mm/h) that broadly impact transit demand across the metropolitan area rather than hyperlocal meteorological gradients.

Finally, the analysis cannot disentangle demand-side behavioral responses from supply-side operational changes under extreme weather. Weather-related conditions may trigger service reductions, speed restrictions, signal failures, or capacity constraints that coincide with travelers’ decisions to reduce subway use. Rather than attempting to causally isolate



**Fig. 6 | The vine-copula-based framework integrates four components to model and assess subway ridership resilience.** The pipeline proceeds from data processing, disaggregating station-level ridership into hourly intervals, through copula modeling using KDE-estimated marginals and hour-specific vine copulas, followed

by validation comparing observed and synthetic distributions via KL divergence and MAPE, and concluding with resilience assessment that quantifies ridership distributional changes under extreme weather conditions relative to baseline. NYC: New York City; KDE: Kernel Density Estimation.

these effects, the framework intentionally conditions on observed operations, modeling realized ridership outcomes under stress as experienced by users and operators. Consequently, estimated ridership declines reflect the combined behavioral and operational consequences of extreme weather, rather than pure traveler response or intrinsic system reliability. This conditioning may bias estimates upward during hazards that more frequently disrupt service, such as heavy precipitation, while temperature-related impacts may more closely reflect demand-side behavior. Findings should therefore be interpreted as measures of conditional ridership resilience, capturing how subway usage materializes under extreme weather given prevailing operational responses, rather than as causal estimates of behavioral avoidance or supply reliability alone.

Future research directions include incorporating additional contextual variables such as service disruptions, special events, and demographic factors; extending the framework to predict cascading failures and system-wide disruptions; and developing real-time decision support tools that integrate vine copula models with operational control systems. This modeling approach can be applied to other urban transit systems, supporting data-informed planning and decision-making in increasingly complex and unpredictable operating environments while advancing the broader goal of building climate-resilient transportation infrastructure.

**Methods**

To capture the complex, non-linear dependencies among subway ridership across stations and time, we propose a methodological framework based on vine copulas, as illustrated in Fig. 6. The framework synthesizes station-level ridership under varying conditions, enabling the assessment of ridership resilience to extreme weather events. It consists of four key components: data processing, copula modeling, model validation, and resilience assessment. During data processing, station-level ridership is disaggregated into hourly intervals to reflect temporal dynamics. In the copula modeling phase, hour-specific marginals are estimated using kernel density estimation (KDE) to capture localized demand variations, followed by vine copulas to flexibly characterize inter-station and temporal dependencies. Model

**Table 10 | Sample Pairwise Mann–Whitney *U* test result at 8:00 AM of Grand Central**

| W Statistics | Monday    | Tuesday   | Wednesday | Thursday  |
|--------------|-----------|-----------|-----------|-----------|
| Friday       | 15370 *** | 17563 *** | 17544 *** | 17397 *** |
| Saturday     | 15824 *** | 17956 *** | 17822 *** | 17688 *** |
| Sunday       | 16022 *** | 18090 *** | 17955 *** | 17820 *** |

\*\*\*indicate statistical significance at the 0.001 levels.

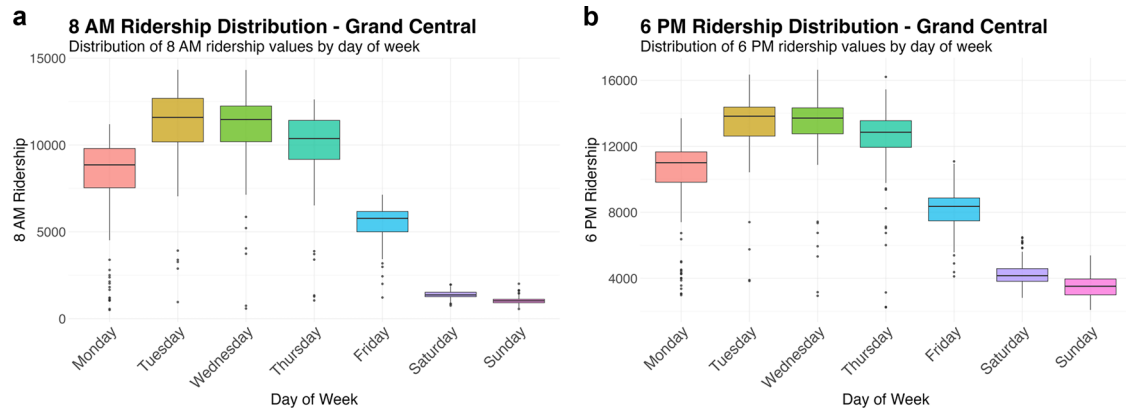
validation evaluates the accuracy of the fitted dependency structures and the reliability of synthetic data by comparing observed and simulated patterns. Finally, resilience assessment leverages the validated copula model to analyze how ridership correlations shift under stress scenarios, quantifying the robustness and adaptability of demand distribution across the network.

**Subway ridership data**

The Metropolitan Transportation Authority (MTA) Subway Hourly Ridership dataset<sup>49</sup> provides comprehensive ridership data for New York City’s subway system, capturing hourly patterns at the station complex level through OMNY contactless payments and MetroCard swipes across various fare classes. All entry types are aggregated into one total ridership value per hour for each station regardless of payment method or fare classification. The dataset is open-sourced and publicly available through the MTA’s official data portal.

The research spans January 2023 through August 2025, selected to ensure data quality and behavioral consistency as ridership patterns stabilized following pandemic disruptions<sup>50</sup>. This timeframe eliminates exceptional pandemic-related volatility from 2020–2022 recovery periods that could skew ridership patterns.

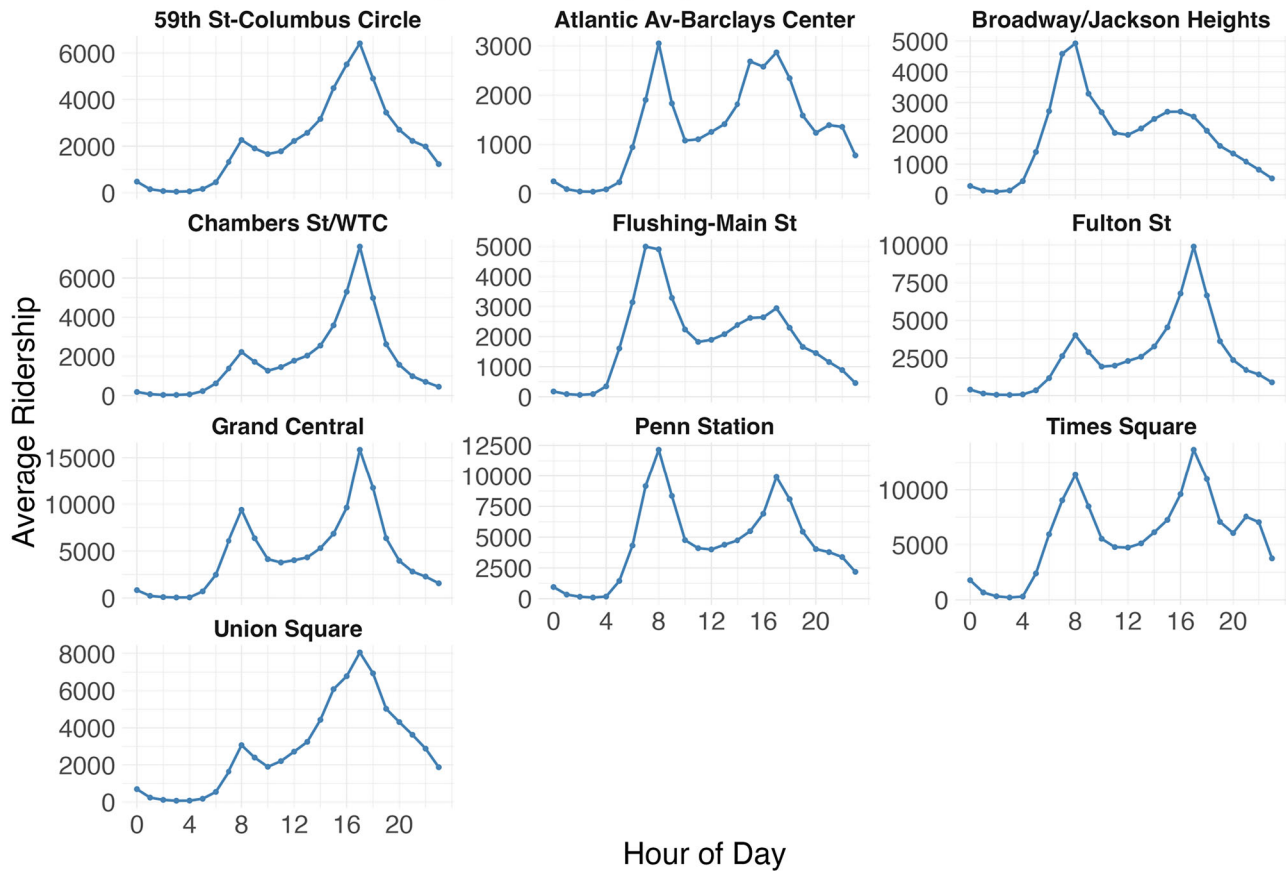
Given the extensive network of 472 subway stations and the computational complexity of high-dimensional vine copula models, this study focuses on 10 strategically selected stations following a systematic selection



**Fig. 7 | Grand Central exhibits different ridership patterns across days, with weekdays and weekends forming two distinct groups.** Boxplot distributions at 8 AM (a) and 6 PM (b) show that Monday through Thursday ridership follows a stable

pattern, while Friday and weekend ridership diverges significantly, as confirmed by Mann–Whitney *U* tests ( $p < 0.001$ ). These systematic differences justify treating weekday and weekend ridership separately in the vine copula modeling framework.

### Average Ridership by Hour for Each Station



**Fig. 8 | Hourly ridership profiles differ markedly across stations, reflecting diverse catchment areas and travel purposes.** Commercial hubs such as Fulton St peak in the afternoon, while residential-serving stations like Flushing-Main St peak earlier in the morning. Penn Station’s rush begins at 8 AM compared to Flushing-

Main St’s 7 AM peak. These pronounced temporal and spatial heterogeneities underscore the necessity of hour-specific marginal modeling within the vine copula framework.

process. First, the top 10 highest-ridership stations from 2024 MTA data were identified, which are heavily concentrated in Manhattan. Second, spatial redundancy was assessed: Herald Square was excluded due to its spatial proximity to Penn Station and line redundancy with other selected stations. Third, to enhance outer-borough representation, Atlantic Av-Barclays Center, the highest-ridership Brooklyn station and major outer-borough transfer hub, was added. The final selection comprises Times

Square St, Penn Station, Grand Central, Union Square, Fulton St, Columbus Circle, Broadway/Jackson Heights, Flushing-Main St, Chambers St/WTC, and Atlantic Av-Barclays Center, distributed across Manhattan (seven stations), Queens (two stations), and Brooklyn (one station), ensuring functional diversity spanning transportation hubs, entertainment centers, business districts, transfer points, and intermodal connections.

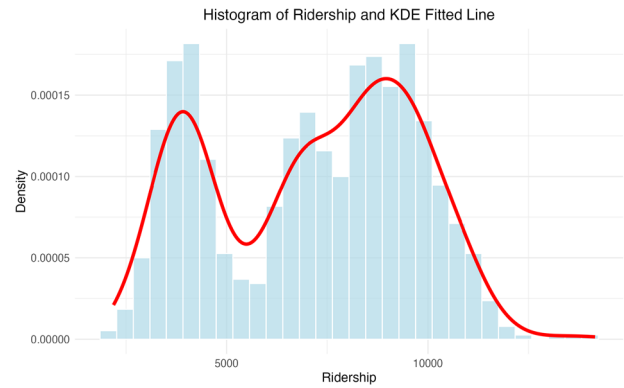
Considering the variation in travel purposes and heterogeneity among populations, subway ridership may not follow a uniform pattern across different days of the week. Since vine copulas model the dependence structure conditional on the marginal distributions, pooling heterogeneous days could obscure true inter-station dependencies and bias the estimated correlations. To ensure that the copula captures representative dependency structures, it is therefore important to differentiate day types when systematic differences exist. To test this, we conducted Mann–Whitney  $U$  tests on daily ridership distributions from Monday through Friday as well as weekends. The Mann–Whitney  $U$  test is a nonparametric method that evaluates whether two independent samples are drawn from the same distribution, making it particularly suitable for skewed and heavy-tailed ridership data without assuming normality. Table 10 shows sample results from the pairwise Mann–Whitney  $U$  test at 8:00 AM for Grand Central, comparing Monday–Thursday with Friday–Sunday. The large  $W$  statistics and  $p$ -values are all less than 0.001 indicate that Friday and weekend ridership exhibits distinct characteristics that diverge from typical Monday–Thursday patterns. Figure 7 displays boxplot distributions for Grand Central station during the morning rush hour (8:00 AM) and evening peak (6:00 PM), exemplifying these differences. Since the analysis confirmed that Friday ridership significantly differs from Monday–Thursday patterns, we treated them separately rather than grouping them into a unified weekday category. Consequently, this study focused on Monday–Thursday ridership patterns, excluding holidays, which represent core business-day behavior and capture the temporal consistency inherent in mid-week transit demand. While Friday and weekend patterns can also be modeled using the same vine copula approach, their distinct statistical characteristics warrant independent analysis to ensure accurate dependency modeling.

Instead of modeling a station’s daily ridership as a single distribution, we estimate hour-specific marginals to better capture diurnal variation and local demand dynamics. As shown in Fig. 8 for NYC stations, subway ridership exhibits distinct temporal patterns that vary significantly by station type and location. Commercial hubs like Fulton St peak in the afternoon as commuters return home, while residential areas experience morning peaks as residents begin their daily commutes. Peak timing also varies substantially across stations: Penn Station’s morning rush begins at 8 AM, while Flushing-Main St’s peak occurs at 7 AM, reflecting different catchment areas, demographic compositions, and travel purposes. These temporal variations demonstrate pronounced heterogeneity in hourly ridership patterns across the subway network. Hourly aggregation enables the copula framework to capture both intra-hour dependencies and inter-hour transitions: within each hour, vine copulas characterize correlations across stations, while across consecutive hours they capture the temporal continuity of ridership evolution. This design avoids the smoothing effect of daily aggregation, which would obscure sharp demand transitions, and instead provides a more realistic representation of ridership dynamics. The observed diurnal fluctuations, consistent with findings by Yu et al.<sup>51</sup> and Gu and Ye<sup>32</sup>, underscore the necessity of hour-specific modeling approaches to accurately capture the spatial-temporal dependencies inherent in urban transit systems.

**Weather information**

Historical weather data for New York City were obtained from the Open-Meteo Historical Weather API<sup>53</sup>, which provides comprehensive meteorological records through reanalysis datasets combining observations from weather stations, aircraft, buoys, radar, and satellite systems. The data offer 9-kilometer spatial resolution and hourly temporal resolution, providing fine-scale weather details suitable for urban transit analysis.

Weather variables were retrieved using a single reference point centered on New York City (40.7128°N, 74.0060°W). This citywide approach is appropriate given that the 10 study stations span approximately 15 kilometers north-south and 8 kilometers east-west, falling largely within 1–2 grid cells at the 9-kilometer resolution. Moreover, the extreme weather events of interest, temperature below



**Fig. 9 | Times Square ridership at 9:00 AM follows a right-skewed, unimodal distribution well captured by kernel density estimation.** The histogram (blue) displays observed ridership frequency, while the fitted KDE curve (red) provides a smooth continuous approximation that avoids the discrete jumps of empirical CDFs. The close alignment between histogram and KDE confirms the suitability of this nonparametric approach for transforming station-level ridership to uniform marginals required for vine copula modeling. KDE: Kernel Density Estimation.

–10 °C, above 38 °C, or precipitation exceeding 5 mm/h, typically affect the entire metropolitan region simultaneously rather than exhibiting substantial station-level variation.

Weather data were matched to ridership at hourly resolution, with each date-hour combination in the ridership dataset paired with the corresponding hourly weather observation. For scenario selection requiring sustained conditions, both consecutive hours were required to meet the specified threshold (e.g., bi-hourly precipitation >5 mm/h) to ensure ridership patterns reflect sustained weather rather than transient events. The Open-Meteo Historical Weather API<sup>53</sup> provided complete records for the study period (July 2020–December 2024) with no missing values, eliminating the need for imputation.

**Vine copula modeling**

Assume  $R_{i,t}$  represents the ridership distribution of station  $i$  at hour  $t$ . The first step of copula modeling is to convert the hourly marginal distribution to a uniform distribution (Eq.2):

$$U_{i,t} = \hat{F}_{i,t}(R_{i,t}), \tag{2}$$

where  $\hat{F}_{i,t}(R_{i,t})$  is the cumulative distribution function.

Given the need for hour-specific modeling to capture heterogeneous dependency structures, Kernel Density Estimation (KDE) was chosen for uniform data transformation due to its advantages over alternatives:

$$\hat{f}_{i,t}(R) = \frac{1}{n * h} \sum_{j=1}^n K\left(\frac{R - r_{i,t,j}}{h}\right), \tag{3}$$

where  $\hat{f}_{i,t}(R)$  is the estimated probability density for station  $i$  at hour  $t$ ,  $r_{i,t,j}$  is the observed ridership at station  $i$ , hour  $t$  and day  $j$ ,  $n$  is the number of observations at hour  $t$ , and  $h$  is the smoothing parameter, and  $K(\cdot)$  is the Gaussian kernel function.

Unlike empirical cumulative distribution functions, which have discrete jumps, KDE provides smooth, continuous probability transformations with derivatives crucial for gradient-based copula parameter estimation. KDE effectively models complex ridership distributions without restrictive parametric assumptions, including multimodality from rush hours and diverse travel patterns, asymmetric shapes due to capacity limits, and irregular tails from extreme events. This ensures that vine copula analysis reflects genuine network dependencies rather than artifacts from transformation. Figure 9 illustrates KDE’s fitting performance on Times Square ridership at 9:00 AM, highlighting its importance for accurate copula modeling.

**Table 11 | Edge Connections by Temporal Lag in First Tree (k = 2)**

| Hour       | 0-2 | 1-3 | 2-4 | 3-5 | 4-6 | 5-7 | 6-8 | 7-9 | 8-10 | 9-11 | 10-12 |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|------|------|-------|
| Within 2 h | 27  | 25  | 25  | 28  | 27  | 26  | 27  | 29  | 29   | 29   | 27    |
| Across 3 h | 2   | 4   | 4   | 1   | 2   | 3   | 2   | 0   | 0    | 0    | 2     |

| Hour       | 11-13 | 12-14 | 13-15 | 14-16 | 15-17 | 16-18 | 17-19 | 18-20 | 19-21 | 20-22 | 21-23 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Within 2 h | 26    | 29    | 28    | 27    | 27    | 29    | 28    | 29    | 29    | 26    | 27    |
| Across 3 h | 3     | 0     | 1     | 2     | 2     | 0     | 1     | 0     | 0     | 3     | 2     |

**Table 12 | Likelihood Value of Each Bi-Hour Vine Structure**

| Hour  | C-Vine      | R-Vine       | D-Vine |
|-------|-------------|--------------|--------|
| 0-1   | 5204        | <b>5307</b>  | 4964   |
| 1-2   | 3468        | <b>3481</b>  | 3276   |
| 2-3   | 2374        | <b>2416</b>  | 2327   |
| 3-4   | <b>2253</b> | 2248         | 2195   |
| 4-5   | 3202        | <b>3231</b>  | 3032   |
| 5-6   | <b>5229</b> | 5226         | 4971   |
| 6-7   | <b>7920</b> | 7893         | 7449   |
| 7-8   | 9616        | <b>9642</b>  | 9136   |
| 8-9   | 9349        | <b>9377</b>  | 9020   |
| 9-10  | 7456        | <b>7528</b>  | 7272   |
| 10-11 | 6760        | <b>6762</b>  | 6438   |
| 11-12 | 7242        | <b>7274</b>  | 6756   |
| 12-13 | 7511        | <b>7589</b>  | 7151   |
| 13-14 | 7655        | <b>7728</b>  | 7283   |
| 14-15 | 8417        | <b>8495</b>  | 7719   |
| 15-16 | 8195        | <b>8345</b>  | 7734   |
| 16-17 | 9495        | <b>9620</b>  | 8948   |
| 17-18 | 10120       | <b>10377</b> | 9728   |
| 18-19 | 8107        | <b>8195</b>  | 7684   |
| 19-20 | <b>5827</b> | 5788         | 5562   |
| 20-21 | 5325        | <b>5356</b>  | 5032   |
| 21-22 | 4194        | <b>4198</b>  | 4021   |
| 22-23 | <b>3952</b> | 3909         | 3749   |

The bold values are likelihood values that are the highest across C-Vine, R-Vine, and D-Vine, indicating better performance with corresponding vine structures.

As the Vine Copulas can capture both spatial relationships (across stations) and temporal relationships (across time), we define  $k$  as the temporal span in hours, allowing the model to capture dependencies across multiple time intervals. The joint distribution of ridership across all  $n$  stations over  $k + 1$  consecutive hours ( $t - k, t - k - 1, \dots, t - 1, t$ ) can be expressed using Sklar’s theorem<sup>14</sup> and Eq. 1 as

$$F^{t-k:t}(R_{1,t-k}, \dots, R_{n,t-k}, \dots, R_{1,t}, \dots, R_{n,t}) = C_{\theta}^{t-k:t}(U_{1,t-k}, \dots, U_{n,t-k}, \dots, U_{1,t}, \dots, U_{n,t}), \tag{4}$$

where  $F^{t-k:t}(\cdot)$  is the joint cumulative distribution function of ridership across all stations from hour  $t - k$  to hour  $t$ .  $R_{i,s}$  is the ridership at station  $i$  during hour  $s$  where  $i \in \{1, 2, \dots, n\}$  and  $s \in \{t - k, t - k + 1, \dots, t\}$ .  $U_{i,s}$  is the uniform [0,1] transformed ridership values using  $\hat{F}_{i,t}(R_{i,t})$  and  $\hat{F}_{i,t}(R_{i,s})$  is the cumulative distribution function derived from the kernel density estimate.  $C_{\theta}^{t-k:t}(\cdot)$  is the vine copula function parameterized by  $\theta$ , capturing the spatial and temporal dependencies among all  $n$  stations across the  $k + 1$  time periods.

To establish the appropriate temporal span for the vine copula model, we conducted empirical testing to identify how far back in time

dependencies remain meaningful. Each hour’s ridership at each station was transformed to uniform marginals using KDE, enabling construction of spatial-temporal vine copulas with different lag configurations. Since ridership correlations may weaken over longer periods, we examined  $k = 2$  (three consecutive hours:  $t - 2, t - 1$  and  $t$ ), which included 30 variables with 29 edges in the first tree. Analysis showed that more than 95% of the strongest dependencies occurred within the same hour (between different stations at hour  $t$ ) or across two consecutive hours (between stations at hours  $t - 1$  and  $t$ ), with only a small proportion of connections across three hours (between stations at hours  $t - 2$  and  $t$ ) as shown in Table 11. This weak cross-temporal relationship beyond consecutive hours indicates that dependencies beyond lag-1 are minimal. Based on these findings, we selected  $k = 1$  for the spatial-temporal vine copula framework, focusing on consecutive hour pairs.

Vine copulas employ a Regular Vine structure (R-vine) as a general framework for decomposing high-dimensional dependencies into bivariate components. R-vines encompass various structural configurations, including two well-known special cases: Canonical Vine (C-vine) and Drawable Vine (D-vine), among others. C-vine structures assume a central node that mediates dependencies throughout the network, with each tree layer having a center component, making them particularly suitable for NYC subway networks where major hub stations like Times Square or Grand Central could serve as primary coordination points for ridership patterns across the system. In modeling ridership with temporal lag  $k$ , a C-vine would model how a central station’s ridership across hour  $t - k$  through  $t$  influences the spatial-temporal dependencies of all other stations in the network.

D-vine structures organize nodes in a sequential chain where each tree forms a path connecting all nodes through adjacent dependencies, making them appropriate for NYC subway networks with clear spatial or operational ordering relationships such as stations along subway lines or geographic corridors. With temporal lag  $k$ , D-vines can capture dependencies along both spatial sequences (station-to-station along lines) and temporal sequences (hour-to-hour patterns).

R-vine structures offer maximum flexibility by allowing arbitrary dependency patterns without requiring a central coordinating node or sequential ordering, accommodating complex network topologies such as NYC’s intricate subway system with multiple interconnected lines and transfer points. This flexibility is particularly valuable when modeling  $k + 1$  consecutive hours across  $n$  stations, as it can capture non-standard dependency patterns that may emerge in the network. However, the vast number of potential R-vine configurations has historically limited practical implementation. To address this challenge, Dissmann et al.<sup>54</sup> developed an automated selection framework that integrates Maximum Spanning Tree algorithms with R-vine copulas, streamlining the structure selection process and enabling efficient high-dimensional modeling.

Parameter estimation for vine copula structures follows sequential tree-by-tree maximum likelihood estimation, beginning with Tree 1 bivariate parameters from transformed uniform data, with subsequent trees requiring conditional transformations generating pseudo-observations through h-functions for higher-order dependency estimation. Table 12 shows the likelihood values of each different vine structure for most hours with relatively large ridership amounts during the day, with the highest likelihood values shown in bold. The table demonstrates that R-vine structures capture dependencies better than C-vine and D-vine

configurations in most cases, though likelihood values are not substantially different across vine types for certain hours. The likelihood-ratio based test proposed by Vuong<sup>55</sup> comparing C-vine, D-vine, and R-vine copulas confirms that the general differences for pairwise models are not statistically significant in some hours between C-vine and R-vine structures. The final model selection is based on the likelihood values, with different vine structures being selected for different hours to optimize the fit for each time period's specific ridership dependency patterns.

Implementation involves two phases: family selection and parameter estimation for the subway ridership vine copula across  $k + 1$  consecutive hours. Each bivariate relationship between station-hour pairs (e.g., Station  $i$  at hour  $s$  with Station  $j$  at hour  $s'$ , where  $s, s' \in \{t - k, t - k + 1, \dots, t\}$ ) requires optimal copula identification from diverse random distribution families using Akaike Information Criterion (AIC) to balance fit against complexity<sup>54</sup>. Parameter estimation follows sequential tree-by-tree maximum likelihood estimation, beginning with Tree 1 bivariate parameters from ridership data transformed to uniform marginals via hour-specific KDE. Subsequent trees require conditional transformations generating pseudo-observations through:

$$F(R_{i,s}|v) = \frac{\partial C_{xv_j|v_{-j}}(F(R_{i,s}|v_{-j}), F(v_j|v_{-j})|\theta)}{\partial F(v_j|v_{-j})} \quad (5)$$

where  $C_{ijk}$  is a bivariate copula distribution function,  $v_i$  is one arbitrarily chosen component of  $v$  and  $v_{-j}$  denotes the  $v$ -vector, excluding this component. These conditional transformations convert the original variables into pseudo-observations that serve as input data for Tree 2 parameter estimation<sup>16,36</sup>.

### Copula-based ridership synthesis and modal validation

For ridership synthesis, the model assumes that prior ridership at hour  $t - k, t - k + 1, \dots, t - 1$  is observed across all stations. Conditional on these observed values, the spatial temporal vine copula structure enables synthesis of ridership distributions for all stations at hour  $t$ .

Suppose the ridership at station  $i \in \{1, 2, \dots, n\}$  at hour  $s \in \{t - k, t - k + 1, \dots, t - 1\}$  is observed as  $R_{i,s} = r_{i,s}^*$ , which is then transformed into a uniform using the KDE-estimated marginal CDFs:  $u_{i,s}^* = \hat{F}_{i,s}(r_{i,s}^*)$ . When conditioning on station  $i$ 's observed ridership at hour  $t - k$  through  $t - 1$  to synthesize ridership at all target station  $j \in \{1, 2, \dots, n\}$  at hour  $t$ , the conditional joint density is defined as:

$$f^{t-k:t}(R_{1,t}, \dots, R_{n,t} | R_{i,s} = r_{i,s}^*) = \frac{c_0^{t-k:t}(u_{i,t-k}^*, \dots, u_{i,t-1}^*, \hat{F}_{1,t}(R_{1,t}), \dots, \hat{F}_{n,t}(R_{n,t}))}{C^{t-k:t-1}(u_{i,t-k}^*, \dots, u_{i,t-1}^*)} \times j = 1n \prod \hat{f}_{j,t}(R_{j,t}), \quad (6)$$

where  $c_0^{t-k:t}$  is the joint copula density of all uniform values (both observed conditioning values and current hour variables),  $C^{t-k:t-1}(\cdot)$  is the marginal copula density across all stations for hour  $t - k$  through  $t - 1$ , and  $\hat{f}_{j,t}(R_{j,t})$  is the KDE-estimated marginal density at station  $j$  at hour  $t$

Synthetic ridership values of station  $j$  at hour  $t$  given the ridership of station  $i$  at hour  $t - k$  through  $t - 1$  are generated by sampling the corresponding uniform variable  $\tilde{U}_{j,t}$  from the conditional joint distribution:

$$\tilde{U}_{j,t} \sim c_0^{t-k:t} \left( U_{j,t} | U_{i,s} = u_{i,s}^*, i, j \in \{1, 2, \dots, n\}, s \in \{t - k, \dots, t - 1\} \right) \quad (7)$$

This conditional sampling leverages the hierarchical structure of the vine copula at hour  $t$  given the known ridership at hour  $t - k$ , through  $t - 1$ , ensuring that the sampled uniform  $\tilde{U}_{j,t}$  respect the dependence on the observed  $u_{i,s}^*$ . Finally, the synthetic uniforms are transformed back to the ridership scale using inverse KDE-estimated marginal CDFs:

$$\tilde{R}_{j,t} = \hat{F}_{j,t}^{-1}(\tilde{U}_{j,t}), j \in \{1, \dots, n\} \quad (8)$$

While conditioning on all stations simultaneously would capture the complete joint distribution, this approach becomes computationally prohibitive for large networks. Therefore, this study employs an ensemble averaging strategy where each station  $j \in \{1, \dots, n\}$ , the model conditionals on observed ridership at station  $i \in \{1, \dots, n\}$  across hour  $t - k$  through  $t - 1$  (i.e.,  $R_{i,t-k} = r_{i,t-k}^*, \dots, R_{i,t-1} = r_{i,t-1}^*$ ) and generate synthesis ridership for all stations at hour  $t$ . This process produced  $n$  sets of synthetic ridership values:  $\{\tilde{R}_{j,t}^i\}$  condition on station  $i$ . The final synthetic ridership is obtained by average across all conditioning stations:

$$\bar{R}_{j,t} = \frac{1}{n} \sum_{i=1}^n \tilde{R}_{j,t}^i, i, j \in \{1, 2, \dots, n\} \quad (9)$$

This ensemble approach reduces computational complexity from exponential to linear in the number of stations while leveraging temporal information from each station across the network. The averaging strategy leverages established principles from ensemble methods, which is averaging reduces prediction variance and mitigates biases that could arise from selecting a single conditioning station. While this approximation does not perfectly preserve the complete joint distribution, it provides a tractable alternative when full joint conditioning becomes computationally infeasible. Small-scale validation (3 stations) comparing ensemble averaging against full joint conditioning demonstrated that the method maintains point prediction accuracy while introducing mild dependency smoothing (mean absolute tau difference of 0.072, preserving correlation signs). This approximation is computationally necessary as full joint conditioning is infeasible for the 10-station network.

The vine copula model enables conditional synthesis of subway ridership, whereby knowledge of one station's ridership at hours  $t - k$  through  $t - 1$  facilitates prediction of network-wide ridership at hour  $t$ . The validation process evaluates the model's distributional fidelity under specific weather conditions by selecting  $N$  days characterized by weather patterns. For each target date, the model iteratively conditions on each individual station's observed ridership across hours  $t - k$  through  $t - 1$  and generates hour  $t$  predictions for all selected stations in the network. This process is repeated for each of the selected stations as conditioning variables, producing sets of network-wide hour  $t$  predictions. These prediction sets are then averaged according to Eq.9 to create a single consensus forecast for all stations at hour  $t$ . To construct synthetic distributions, this entire ensemble process is repeated  $N$  times, each time drawing new random samples from the fitted copula, resulting in  $N$  averaged realizations for each scenario at each station. These synthetic hour  $t$  distributions are then statistically compared against observed hour  $t$  distributions from days with matching weather conditions using Kullback–Leibler (KL) divergence to assess distributional similarity and Mean Absolute Percentage Error (MAPE) to evaluate point prediction accuracy.

### Ridership resilience assessment

The vine copula framework enables the assessment of ridership resilience under extreme weather conditions by synthesizing ridership distributions and quantifying deviations from baseline patterns. Rather than modeling system operations, this analysis evaluates how station-level ridership demand responds to environmental stressors relative to typical conditions.

A key advantage of the vine copula approach lies in its ability to reconstruct plausible ridership distributions even when extreme event data are scarce. By leveraging the dependency structure learned from historical observations, including tail dependencies that capture co-movement among stations during rare but impactful conditions, the model can conditionally synthesize ridership patterns that represent realistic network-wide responses under stress.

The assessment process begins by generating synthetic ridership distributions under two contrasting scenarios, extreme weather and baseline conditions. For the extreme scenarios, the model conditions on historical

records where<sup>1</sup> temperature falls below  $-10^{\circ}\text{C}$ ,<sup>2</sup> temperature exceeds  $38^{\circ}\text{C}$ , or 3 sustained precipitation reaches at least 5 mm/h for two consecutive hours. For each extreme scenario, the model generates  $N$  synthetic ridership realizations following the ensemble averaging methodology from Eq. 9. The baseline scenario represents normal weather conditions, temperatures within the comfort range ( $0$ – $30^{\circ}\text{C}$ ) and no precipitation, while maintaining all other temporal and spatial factors consistent with the extreme cases.

To quantify ridership resilience, both extreme and normal scenario distributions are sorted independently by magnitude. The paired rank-wise differences are then used to compute percentage changes in ridership as:

$$\Delta R_{i,t}^{(m)} = \frac{R_{i,t}^{\text{extreme},(m)} - R_{i,t}^{\text{baseline},(m)}}{R_{i,t}^{\text{baseline},(m)}} * 100\%, m = 1, 2, \dots, N., \quad (10)$$

where  $R_{i,t}^{\text{extreme},(m)}$  and  $R_{i,t}^{\text{baseline},(m)}$  represent the  $m$ -th ranked (sorted) synthetic ridership values at station  $i$  and hour  $t$  under extreme and baseline weather conditions, respectively. This quantile-matching approach ensures that comparable ridership levels are compared: the lowest extreme weather ridership is compared to the lowest baseline weather ridership, the median to the median, and so on.

This produces a distribution of  $N$  percentage change values  $\{\Delta R_{i,t}^1, \Delta R_{i,t}^2, \dots, \Delta R_{i,t}^N\}$  across the full range of ridership outcomes. These percentage changes are then sorted to obtain the ordered distribution, from which confidence intervals are derived. The 95% confidence interval is calculated as the 2.5th and 97.5th percentiles of the sorted change distribution, while the median (50th percentile) provides a robust central estimate. The quantile-based comparison shows how extreme weather impacts ridership across the entire distribution, from low-ridership scenarios to high-ridership scenarios, rather than just comparing mean values, thereby capturing the full range of system vulnerability and resilience.

## Data availability

The datasets generated and/or analysed during the current study are available in the MTA Subway Hourly Ridership: 2020–2024 repository, [https://data.ny.gov/Transportation/MTA-Subway-Hourly-Ridership-2020-2024/wujg-7c2s/about\\_data](https://data.ny.gov/Transportation/MTA-Subway-Hourly-Ridership-2020-2024/wujg-7c2s/about_data).

Received: 5 November 2025; Accepted: 2 March 2026;

Published online: 01 April 2026

## References

- Kuby, M., Barranda, A. & Upchurch, C. Factors influencing light-rail station boardings in the United States. *Transport. Res. Part A Policy Pract.* **38**, 223–247 (2004).
- Sohn, K. & Shim, H. Factors generating boardings at Metro stations in the Seoul metropolitan area. *Cities*. **27**, 358–368 (2010).
- He, Y., Zhao, Y. & Tsui, K. L. An adapted geographically weighted LASSO (Ada-GWL) model for predicting subway ridership. *Transportation*. **48**, 1185–1216 (2021).
- Singhal, A., Kamga, C. & Yazici, A. Impact of weather on urban transit ridership. *Transport. Res. Part A Policy Pract.* **69**, 379–391 (2014).
- Zhou, M. et al. Impacts of weather on public transport ridership: Results from mining data from different sources. *Transport. Res. Part C Emerg. Technol.* **75**, 17–29 (2017).
- Kain, J. F. & Liu, Z. Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transport. Res. Part A Policy Pract.* **33**, 601–624 (1999).
- Boyle, D. K. *Fixed-route Transit Ridership Forecasting and Service Planning Methods* (Vol. 66) (Transportation Research Board, 2006).
- Sung, H. & Oh, J. T. Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*. **28**, 70–82 (2011).
- Cardozo, O. D., García-Palomares, J. C. & Gutiérrez, J. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Appl. Geogr.* **34**, 548–558 (2012).
- Chow, L. F., Zhao, F., Liu, X., Li, M. T. & Ubaka, I. Transit ridership model based on geographically weighted regression. *Transport. Res. Rec.* **2006**, 105–114 (1972).
- Goetzke, F. Network effects in public transit use: evidence from a spatially autoregressive mode choice model for New York. *Urban Stud.* **45**, 407–427 (2008).
- Zhao, F. & Park, N. Using geographically weighted regression models to estimate annual average daily traffic. *Transport. Res. Rec.* **1879**, 99–107 (2004).
- Kim, J. & Zhang, M. Determining transit's impact on Seoul commercial land values: An application of spatial econometrics. *Int. Real. Estate Rev.* **8**, 1–26 (2005).
- Liu, L. et al. Dynamic spatial-temporal representation learning for traffic flow prediction. *IEEE Trans. Intell. Transport. Syst.* **21**, 2663–2674 (2020).
- Chen, P., Fu, X. & Wang, X. A graph convolutional stacked bidirectional unidirectional-LSTM neural network for metro ridership prediction. *IEEE Trans. Intell. Transport. Syst.* **23**, 6950–6962 (2021).
- Yao, H., Tang, X., Wei, H., Zheng, G. & Li, Z. Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction. *Proc. AAAI Conf. Artif. Intell.* **33**, 5668–5675 (2019).
- Liu, L. et al. Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction. *IEEE Trans. Intell. Transport. Syst.* **23**, 3377–3391 (2020).
- Stover, V. W. & McCormack, E. D. The impact of weather on bus ridership in Pierce County, Washington. *J. Public Transport.* **15**, 95–110 (2012).
- Kashfi, S. A., Lee, B., & Bunker, J. Impact of rain on daily bus ridership: a Brisbane case study. In *Australasian Transport Research Forum 2013 Proceedings* (pp. 1–18). (Australasian Transport Research Forum, 2013).
- Chen, A., Yang, H., Lo, H. K. & Tang, W. H. Capacity reliability of a road network: an assessment methodology and numerical results. *Transport. Res. Part B Methodol.* **36**, 225–252 (2002).
- Mo, Y. & Qiao, X. A study on the reliability evaluation of urban transit system. In *2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS)* Vol. 2, pp. 299–302 (IEEE, 2009).
- Aparicio, J. T., Arsenio, E. & Henriques, R. Assessing robustness in multimodal transportation systems: a case study in Lisbon. *Eur. Transp. Res. Rev.* **14**, 28 (2022).
- Sklar, M. Fonctions de répartition à  $n$  dimensions et leurs marges. *Annales de l'ISUP* **8**, 229–231 (1959).
- Bedford, T. & Cooke, R. M. Probability density decomposition for conditionally dependent random variables modeled by vines. *Ann. Math. Artif. Intell.* **32**, 245–268 (2001).
- Aas, K., Czado, C., Frigessi, A. & Bakken, H. Pair-copula constructions of multiple dependence. *Insurance Math. Econ.* **44**, 182–198 (2009).
- Brechmann, E., Czado, C. & Paterlini, S. Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *J. Bank. Financ.* **40**, 271–285 (2014).
- Gräler, B. et al. Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.* **17**, 1281–1296 (2013).
- Low, R. K. Y., Alcock, J., Faff, R., & Brailsford, T. Canonical vine copulas in the context of modern portfolio management: are they worth it? Asymmetric Dependence in Finance: Diversification, Correlation and Portfolio Management in Market Downturns, 263–289 (2018).
- Meloni, I., Eluru, N., Spissu, E., Portoghese, A. & Bhat, C. R. A copula-based joint model of commute mode choice and number of

- non-work stops during the commute. *Int. J. Transp. Econ.* **38**, 337–364 (2011).
30. Ermagun, A., Hossein Rashidi, T. & Samimi, A. A joint model for mode choice and escort decisions of school trips. *Transportmetrica A Transp. Sci.* **11**, 270–289 (2015).
  31. Sener, I. N. & Bhat, C. R. A copula-based sample selection model of telecommuting choice and frequency. *Environ. Plan. A* **43**, 126–145 (2011).
  32. Seyedabrishami, S. & Izadi, A. R. A copula-based joint model to capture the interaction between mode and departure time choices in urban trips. *Transport. Res. Procedia* **41**, 722–730 (2019).
  33. Rasaizadi, A. & Seyedabrishami, S. Analysis of the interaction between destination and departure time choices. *Sci. Iran.* **28**, 2471–2478 (2021).
  34. Jafari Shahdani, F., Rasaizadi, A. & Seyedabrishami, S. The interaction between activity choice and duration: application of copula-based and nested-logit models. *Sci. Iran.* **28**, 2037–2052 (2021).
  35. Pourabdollahi, Z., Karimi, B. & Mohammadian, A. Joint model of freight mode and shipment size choice. *Transport. Res. Rec.* **2378**, 84–91 (2013).
  36. Keya, N., Anowar, S. & Eluru, N. Joint model of freight mode choice and shipment size: a copula-based random regret minimization framework. *Transport. Res. Part E Logist. Transport. Rev.* **125**, 97–115 (2019).
  37. Xiaoliang, Z. & Limin, J. Analysis of bus line operation reliability based on copula function. *Sustainability* **13**, 8419 (2021).
  38. Di, Y., Xu, M., Zhu, Z. & Yang, H. A copula-based approach for multi-modal demand dependence modeling: temporal correlation between demand of subway and bike-sharing. *Travel Behav. Soc.* **38**, 100908 (2025).
  39. Asgari, L., Shahangian, R. S. & Mamdoohi, A. R. A joint mode change and mode choice decision under transportation demand management policies (case of a copula approach). *Int. J. Transport. Eng.* **11**, 1515–1531 (2024).
  40. Su, Z., Lee, E., Lo, H.K., & Chow, J.Y.J. Resilient bus services design under correlated stochastic metro system disruption. In: *International Conference on Advanced Systems in Public Transport (CASPT)*, Kyoto, Japan, 2025 (2025).
  41. Bedford, T. & Cooke, R. M. Vines—a new graphical model for dependent random variables. *Ann. Stat.* **30**, 1031–1068 (2002).
  42. Joe, H., Li, H. & Nikoloulopoulos, A. K. Tail dependence functions and vine copulas. *J. Multivar. Anal.* **101**, 252–270 (2010).
  43. Nations, U. UNISDR terminology on disaster risk reduction. *United Nations Office for Disaster Risk Reduction, Report*, 1–13 (2009).
  44. Berdica, K. An introduction to road vulnerability: what has been done, is done and should be done. *Transp. Policy* **9**, 117–127 (2002).
  45. Hirai, C., Kunimatsu, T., Tomii, N., Kondou, S. & Takaba, M. A train stop deployment planning algorithm using a petri-net-based modelling approach. *Q. Rep. RTRI* **50**, 8–13 (2009).
  46. Diab, E. & Shalaby, A. Metro transit system resilience: understanding the impacts of outdoor tracks and weather conditions on metro system interruptions. *Int. J. Sustain. Transport.* **14**, 657–670 (2020).
  47. Adjetey-Bahun, K., Birregah, B., Châtelet, E., Planchet, J. L., & Laurens-Fonseca, E. (2014, May). A simulation-based approach to quantifying resilience indicators in a mass transportation system. In *ISCRAM*.
  48. Zhu, Y., Xie, K., Ozbay, K., Zuo, F. & Yang, H. Data-driven spatial modeling for quantifying networkwide resilience in the aftermath of hurricanes Irene and Sandy. *Transport. Res. Rec.* **2604**, 9–18 (2017).
  49. Metropolitan Transportation Authority. (2023). Subway and bus ridership for 2023. <https://www.mta.info/agency/new-york-city-transit/subway-bus-ridership-2023>
  50. Wang, H. & Noland, R. B. Bikeshare and subway ridership changes during the COVID-19 pandemic in New York City. *Transp. Policy* **106**, 262–270 (2021).
  51. Yu, L., Chen, Q., & Chen, K. Deviation of peak hours for urban rail transit stations: a case study in Xi'an, China. *Sustainability*. **2733** (2019).
  52. Gu, L. & Ye, X. Research on peak time of passenger flow entering and leaving stations in Osaka rail transit stations. *Compr. Transport.* **2**, 57–61 (2014).
  53. Zippenfenig, P. Open-Meteo.com Weather API [Computer software]. *Zenodo*. <https://doi.org/10.5281/ZENODO.7970649>
  54. Dissmann, J., Brechmann, E. C., Czado, C. & Kurowicka, D. Selecting and estimating regular vine copulae and application to financial returns. *Computat. Stat. Data Anal.* **59**, 52–69 (2013).
  55. Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333 (1989).

### Acknowledgements

B.H. acknowledges support for the research from the Climate Resilience through Multidisciplinary Big Data Learning, Prediction & Building Response Systems (CLIMBS) project (Award #2344533) funded by the National Science Foundation.

### Author contributions

The authors confirm contribution to the paper as follows: study conception and design: B.H. and J.C.; data collection: Y.G. and B.H.; analysis and interpretation of results: Y.G., B.H., J.C., Z.S., and O.W.; draft manuscript preparation: Y.G. and B.H. All authors reviewed the results and approved the final version of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Brian Yueshuai He.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026