Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.083723

ARTICLE

A Space-Air-Ground Integrated Network Traffic Estimation Algorithm Based on Time-Varying Higher-Order Moments

Xiaoxiong Yang[1,2], Yi Zhang[1], Dingde Jiang[1,*] and Shuqing He[3]

[1]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
[3]School of Information Science and Engineering, Linyi University, Linyi, China

*Corresponding Author: Dingde Jiang. Email: jiangdd@uestc.edu.cn

Received: 09 April 2026; Accepted: 12 May 2026; Published: 15 June 2026

ABSTRACT: With the proliferation of network users, traffic engineering has become increasingly important for the management and optimization of networks. As a crucial component of traffic engineering, the traffic matrix can assist network managers in making informed decisions to optimize resource utilization. However, in the current complex and heterogeneous space-ground integrated network, the cost of direct real-time measurement of traffic matrix is high and the delay is high. To address this challenge, we propose a network traffic estimation algorithm based on time-varying higher-order moments and deep learning, which leverages the time-varying higher-order moments property of traffic to improve the understanding of non-stationary traffic. First, we introduce an extended generalized autoregressive conditional heteroskedasticity model (THM-GARCH) that incorporates higher-order moment information to predict traffic volatility. Then, the THM-GARCH model is integrated with a long short-term memory network, and a dynamic feature update mechanism is developed to address the issue. The experimental results indicate that the proposed algorithm achieves the highest qualitative accuracy among all traffic estimation experiments, with a 17.78% reduction in root mean square error and a 14.69% reduction in mean square error.

KEYWORDS: Traffic estimation; higher-order moments; non-stationary traffic; deep learning

1 Introduction

Advancements in communication technology have resulted in a growing number of network devices and communication links, leading to an enormous increase in network traffic [1]. As the space-air-ground integrated edge network promotes a wider range of Internet services, the operation and maintenance of the network is facing major challenges [2]. For instance, the number of network users continues to rise, and it is expected that global mobile data traffic will reach 303 EB per month by 2030 [3]. To effectively manage networks, traffic engineering (TE) has been employed to assist in decision-making. TE enhances network resource utilization by optimizing data traffic flow within the network. Techniques such as network traffic analysis, route planning, and load balancing are essential components of traffic engineering [4,5]. In this context, network traffic analysis employs monitoring techniques to understand and plan the behavior of network traffic. The traffic matrix of a network comprises multiple time slices of original destination (OD) flows, which help service providers allocate network resources more efficiently [6,7]. Typically, OD flows can be obtained through methods such as NetFlow, sFlow, network control plane data, and network log analysis [8]. However, due to the explosive growth of traffic and the expansion of physical networks, the real-time acquisition of flow matrices demands substantial computational resources. Consequently, network operators are considering the use of a limited amount of observed traffic to predict future traffic matrices. This approach not only reduces monitoring costs but also facilitates efficient edge network management.

With the advancement of computational power, artificial neural networks, particularly convolutional neural networks, are becoming increasingly popular. This popularity stems from the ability of neural networks to model nonlinear relationships and demonstrate significant predictive capabilities [9]. Consequently, traffic estimation methods are gradually shifting from traditional statistical approaches to machine learning techniques, and ultimately to deep reinforcement learning strategies. However, existing algorithms face challenges in capturing the dependencies between different moments of traffic. While network traffic exhibits periodic fluctuations, it is also influenced by various factors. For example, the periodic overflights of low-Earth orbit (LEO) satellite nodes can cause significant non-stationary fluctuations in network traffic. These complexities render network traffic multiscale and nonlinear [10].

Higher-order moments provide additional statistical information that can enhance the characterization of flow dynamics. The fluctuation of network traffic is influenced not only by historical variance but also by dynamic skewness and kurtosis [11]. For example, bursty traffic may result in a right-skewed distribution, while Distributed Denial of Service attacks can significantly increase kurtosis. Furthermore, the time-varying characteristics of traffic sequences require further characterization. In light of this, we propose a network traffic matrix estimation algorithm (TYHM) that takes into account the time-varying higher-order moment characteristics. This algorithm leverages the benefits of time-varying higher-order moment fluctuations and long short-term memory (LSTM) neural networks to enhance the accuracy of network flow matrix estimation. Specifically, the contributions of this paper are as follows:

- A generalized autoregressive conditional heteroskedasticity model based on time-varying higher-order moments (THM-GARCH) is designed to provide a multidimensional statistical characterization by utilizing higher-order moment information to capture the non-normal characteristics of flow distribution. Specifically, in the GARCH model, skewness and kurtosis are incorporated to account for the asymmetry and tail characteristics of the flow, while the volatility of the flow is estimated through dynamic recursion.
- The design of a neural network that integrates THM-GARCH statistical features with LSTM significantly enhances the modeling capabilities of non-stationary flow series. First, THM-GARCH is employed to update the higher-order moment information and volatility of the flow. Subsequently, the flow values and higher-order moment information are transformed into feature matrices for input into the network. The incorporation of time-varying higher-order moments increases the model's robustness in handling outliers and extreme cases, thereby improving the algorithm's adaptability to complex flow fluctuations and reducing estimation errors in flow predictions.
- Through rigorous experiments, we validate the performance of the proposed algorithm for network traffic estimation. Specifically, in the presence of complex traffic fluctuations, the TYHM algorithm significantly decreases the overall estimation error when compared to other algorithms.

The rest of the paper is organized as follows. Section 2 describes related work on traffic prediction and higher-order moments. Section 3 presents a model of the network traffic estimation problem along with evaluation metrics. Section 4 describes the proposed algorithm. Section 5 shows the test environment and results of the algorithm and analyzes them. Finally, Section 6 gives conclusions.

2 Related Works

This section provides a systematic review of existing research in traffic estimation and higher-order moment analysis. Following the reviewers' suggestions, we categorize the related work into three dimensions to better position the TYHM framework.

2.1 Traditional Statistical Estimation Methods

Traditional statistical methods rely on mathematical frameworks to model traffic time and space characteristics. Literature [12] established a network traffic prediction algorithm that employs a simulated annealing technique to optimize a differential integrated moving average autoregressive model. To address limited link load observations, some studies proposed flow count estimation methods that operate without historical data, utilizing correlation analysis to infer missing values [13]. Additionally, moment-matching techniques have been introduced, such as the moment-generating function matching method used to estimate OD flow rates under Poisson model conditions [14]. While these models offer high interpretability, their fixed parametric assumptions often struggle to adapt to the highly complex and non-stationary fluctuations inherent in modern space-air-ground integrated networks.

2.2 Deep Learning-Based Traffic Prediction Models

Deep learning has demonstrated significant advantages in processing high-dimensional non-linear features [15]. Existing frameworks frequently employ dynamic graph convolutional networks (GCN) combined with recurrent neural networks (RNN) to capture temporal and topological dependencies [16]. To enhance adaptability, graph structure learning has been applied to improve the performance of GCNs in traffic prediction [17]. For time-varying network traffic, the authors of literature [18] introduced a traffic prediction algorithm that combines deep Q-learning with generative adversarial networks. However, most current deep learning models treat traffic sequences as Gaussian distributed, neglecting the statistical information contained in skewness and kurtosis. This limitation restricts their robustness when facing extreme traffic bursts or anomalous spikes.

2.3 Robustness-Oriented Learning Strategies

Research has explored various data recovery and robustness enhancement strategies to mitigate the impact of abnormal data environments [19]. In recent years, neural networks have proven to be well-suited for forecasting time series data. Literature [20] proposes a hybrid forecasting method that combines the Savitzky–Golay filter with an LSTM architecture, effectively improving the accuracy of traffic flow forecasts. Literature [21] employed higher-order moment normalization to model the time-varying characteristics of network traffic. Subsequently, a moment crossing predictor was proposed to enhance traffic prediction accuracy effectively. The proposed TYHM framework differs from these models by deeply coupling THM-GARCH with LSTM. This design converts time-varying statistical features directly into input matrices, enabling the neural network to perceive structural changes in traffic distributions.

3 Problem Statement

In this section, we first provide a formal description of the network traffic estimation problem. Then, the evaluation metrics used in this paper to validate the performance of the algorithm are presented.

3.1 Modeling of the Traffic Estimation Problem

In this paper, the network system is abstractly modeled as $G = \{V, L, R\}$, where $V = \{v_1, v_2, \ldots, v_{|V|}\}$ denotes the set of source and destination nodes in the network, and v_i represents the i-th node. The variable g denotes the set of links, including links between two neighboring nodes and links from a node to the node itself, and l_i denotes the i-th link. The routing matrix of the network, denoted as R, can be derived from the network profile. Specifically, R is a binary matrix, as illustrated in the following equation.

$$R = \begin{bmatrix} r_1^{1,1} \cdots r_1^{|N|,|N|} \\ \vdots \quad \ddots \quad \vdots \\ r_{|L|}^{1,1} \cdots r_{|L|}^{|N|,|N|} \end{bmatrix}. \quad (1)$$

If the OD flow between nodes v_i and v_j passes through link l_k, then $r_k^{i,j} = 1$. Otherwise, $r_k^{i,j} = 0$. The essence of the flow matrix is a time series composed of OD flows, as shown in Eq. (2).

$$X = \begin{bmatrix} x_1^1\, x_1^2 \cdots x_1^{|N|^2} \\ x_2^1\, x_2^2 \cdots x_2^{|N|^2} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ x_T^1 x_T^2 \cdots x_T^{|N|^2} \end{bmatrix}, \quad (2)$$

where x_t^i denotes the flow rate of the i-th OD pair at time t. As a result, the traffic estimation problem can be defined as Eq. (3).

$$x_{t+1}^y = f(x_t^i, x_{t-1}^i, \cdots, x_2^i, x_1^i). \quad (3)$$

The primary objective of this paper is to estimate future traffic using a set of known historical data. Single-step predictions require more computational resources. Therefore, we employ a multi-step prediction strategy. The problem must adhere to the following constraints.

$$\begin{cases} Y_t = RX_t \\ X_t^i \geq 0, \quad i \in 1, 2, \cdots, |N|^2 \\ Y_t^j \geq 0, \quad j \in 1, 2, \cdots, L \end{cases}. \quad (4)$$

$Y = [Y_1, Y_2, \cdots, Y_t, \cdots Y_T]$ is a link load matrix of dimension $L \times T$. $y_t^j \in Y_t, (j = 1, 2, \ldots L)$ denotes the traffic load on the j-th link at time t.

3.2 Algorithmic Assessment Indicators

In this work, we use the following metrics to evaluate the performance of the proposed time-varying higher-order moments-based traffic estimation algorithm.

(1) Spatial Error: Spatial error encompasses both spatial absolute error and spatial relative error. This type of error allows for the visualization of the estimation deviation of the algorithm across different OD flows and assesses its performance in addressing the spatial correlation of these flows.

$$ASE\,(i) = \left(\sum_{t=1}^{T} \left| \hat{x}_t^i - x_t^i \right| \right) \Big/ T. \quad (5)$$

Eq. (5) is used to calculate the spatial absolute error of the n-th OD flow, where $\hat{x}_t^n$ denotes the traffic estimate of the n-th OD flow at time t.

(2) Temporal Error: The temporal error evaluates the estimation bias of the algorithm at various moments in time. Eq. (6) is used to calculate the time absolute error at the t-th time point.

$$ARE\left(t\right) = \left(\sum_{i=1}^{N} \left|\hat{x}_t^i - x_t^i\right|\right) \Big/ N. \tag{6}$$

(3) Root Mean Square Error (RMSE): The RMSE is calculated based on the square of the error and is more sensitive to larger errors, as shown below.

$$RMSE = \sqrt{\left(\sum_{i=1}^{N} \left(\hat{x}_t^i - x_t^i\right)^2\right) \Big/ N}. \tag{7}$$

(4) Mean Absolute Error (MAE): The MAE quantifies the average deviation between predicted values and actual values, effectively minimizing the impact of outliers on the overall error assessment. The definition is shown below.

$$MAE = \left(\sum_{i=1}^{N} \left|\hat{x}_t^i - x_t^i\right|\right) \Big/ N. \tag{8}$$

Using these metrics, it is possible to analyze the error profile of traffic estimates in multiple dimensions.

4 Traffic Estimation Algorithm Design

In this section, we implement a time-varying higher-order moments-based algorithm for network traffic estimation. The algorithm integrates volatility prediction with Long Short-Term Memory LSTM networks and is divided into two main stages. First, we design a volatility prediction model, THM-GARCH, based on the GARCH model, which accounts for time-varying higher-order moments. Second, we propose a network traffic estimation algorithm that utilizes features derived from time-varying higher-order moments in conjunction with the LSTM network.

4.1 Prediction of Flow Volatility Based on THM-GARCH

Network traffic typically exhibits a volatility clustering effect, characterized by periods of high volatility and autocorrelation. Abnormal large fluctuations in network traffic are often accompanied by subsequent significant fluctuations. During stable periods, the fluctuations remain relatively small. The Generalized autoregressive conditional heteroskedasticity (GARCH) model is specifically designed for volatility prediction in time series data. Traditional GARCH models primarily consider basic statistical features of the data and rely on variance to capture the volatility of the time series. However, traffic series frequently display abnormal volatility and asymmetry. Consequently, we have developed a THM-GARCH model based on time-varying higher-order moments to capture the asymmetry and tail characteristics of the flow, utilizing skewness and kurtosis.

In THM-GARCH, we first dynamically adjust the mean through a first-order autoregressive process, as shown in Eq. (9).

$$\mu_{i,t} = \phi\mu_{i,t-1} + (1-\phi)x_{t-1}^i,\ \ (0 < \phi < 1), \tag{9}$$

where u_t^i denotes the conditional mean value of the i-th OD flow at time t, which is used to capture the flow trend. ϕ serves as an autoregressive parameter that regulates the influence of historical flow on the current value. Conditional variance was calculated dynamically using the following equation.

$$\begin{cases} \sigma_{i,t}^2 = \omega + \alpha\left(x_{t-1}^i - \mu_{t-1}^i\right)^2 + \beta\sigma_{i,t-1}^2 + \eta_1 S_{i,t-1} + \eta_2 K_{i,t-1}, \\ \qquad s.t.\ \omega > 0, \alpha \geq 0, \beta \geq 0\ , \alpha + \beta < 1, \end{cases} \tag{10}$$

where $\sigma_{i,t}^2$ represents the conditional variance at time t used to characterize flow volatility, while ω is a constant term that indicates the base level of variance. The parameters α, β, η_1, and η_2 are employed to condition the influence of historical values on current values. Including these terms in the estimation equation represents a moment-based correction strategy. This approach is based on the principle of statistical bias correction, which utilizes higher-order moments to compensate for the bias caused by skewness and the variance driven by kurtosis that cannot be captured by the standard mean-variance model. Specifically, these terms act as linear correction factors that calibrate the original forecast based on the asymmetry and tail heaviness of the flow distribution. In addition, $\alpha + \beta < 1$ is used to maintain smoothness.

Skewness is a statistical measure used to quantify the asymmetry of a probability distribution. It is calculated as follows:

$$\mathrm{S}_{i,t} = \gamma_0 + \gamma_1 \mathrm{S}_{i,t-1} + \gamma_2 x_{i,t-1}^3,\ \ |\gamma_1| < 1, \tag{11}$$

where $\mathrm{S}_{i,t}$ denotes the conditional skewness at time t, γ_0 denotes the baseline skewness, and γ_1 is used to measure the effect of historical skewness. $\gamma_2 > 0$ denotes the asymmetric contribution, where large positive residuals boost future skewness.

Kurtosis is used to characterize the tails of the distribution, as demonstrated in Eq. (12).

$$\mathrm{K}_{i,t} = \delta_0 + \delta_1 \mathrm{K}_{i,t-1} + \delta_2 x_{i,t-1}^4,\ \ |\delta_1| < 1, \tag{12}$$

where $\mathrm{K}_{i,t}$ denotes the conditional kurtosis of the traffic sequence at moment t and δ_0 denotes the baseline kurtosis. $\delta_2 > 0$ denotes that extreme fluctuations increase the kurtosis.

We fit the residuals using a dynamic skewed t-distribution, which effectively captures the anomalous fluctuations and asymmetric characteristics of the traffic residuals. By introducing the skewness parameter λ, the non-normal characteristics of the network traffic are modeled. The probability density function is presented below.

$$f\left(z_{i,t} | \nu_{i,t}, \lambda_{i,t}\right) = \frac{c_{i,t} \cdot \Gamma\left(\dfrac{\nu_{i,t}+1}{2}\right)\left(1 + \dfrac{adj_z^2}{\nu_{i,t}-2}\right)^{-\dfrac{\nu_{i,t}+1}{2}}}{\sqrt{\pi\left(\nu_{i,t}-2\right)}\Gamma\left(\dfrac{\nu_{i,t}}{2}\right)}. \tag{13}$$

$$adj_z = \begin{cases} z_{i,t} \cdot \lambda_{i,t} \text{ if } z_{i,t} \geq 0 \\ z_{i,t}/\lambda_{i,t} \text{ if } z_{i,t} < 0 \end{cases}. \tag{14}$$

$$c_{i,t} = \frac{2}{\lambda_{i,t} + 1/\lambda_{i,t}}, \tag{15}$$

where $\Gamma(\cdot)$ denotes the Gamma function and adj_z denotes the adjusted residual, which introduces asymmetry. Additionally, $c_{i,t}$ serves as the normalization constant. The variable $z_{i,t}$ is the normalized residual, calculated as shown in Eq. (16).

$$z_{i,t} = \frac{x_t^i - \mu_{i,t}}{\sqrt{\sigma_{i,t}^2}}. \tag{16}$$

The variable $\lambda_{i,t}$ denotes skewness at time t and is used to control the asymmetry of the distribution, as illustrated below.

$$\lambda_{i,t} = \max\left(\lambda_0 + \lambda_1 S_{i,t}, 0.1\right). \tag{17}$$

The variable $\nu_{i,t}$ represents the degrees of freedom at moment t, which is utilized to regulate the thickness of the distribution's tails. It is calculated as demonstrated below.

$$\nu_{i,t} = \max(\nu_0 + \nu_1 K_{i,t}, 2.1). \tag{18}$$

The transition from statistical modeling to deep learning is facilitated by a unified feature construction process. The YHM framework uses the THM-GARCH model to map the extracted time-varying higher-order moments into a six-dimensional feature matrix, enabling the LSTM network to accurately predict non-stationary network traffic by recognizing the morphological evolution patterns of traffic distribution. The initialization of the THM-GARCH model uses a parameter vector that encompasses the mean process, variance dynamics, higher-order moment evolution, and distribution characteristics. These initial values are randomly initialized within the range [0, 1], providing a starting point for maximum likelihood estimation. Next, the parameters were optimized using the maximum likelihood estimation method via a sequential quadratic programming algorithm.

4.2 Network Traffic Estimation Algorithms

After obtaining the volatility, we apply an LSTM model during the network traffic estimation phase. This model is responsible for capturing long-term dependencies and nonlinear characteristics in network traffic. The selection of data features is crucial for neural network-based algorithms to achieve optimal results. We trained the algorithm by extracting a feature matrix from the network traffic to serve as input for the neural network.

We extracted six key attributes from the traffic time series: traffic value, mean, variance, skewness, kurtosis, and predicted volatility. The mean and variance reflect the overall trend and volatility of network traffic and can be utilized directly as features to enhance the model's understanding of global patterns. Skewness and kurtosis are particularly effective in capturing abnormal fluctuations in network traffic. We calculate the higher-order moments at the current point in time based on the traffic data within the designated time window. The skewness is calculated as follows:

$$skew_{i,t} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^3}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2\right)^{\frac{3}{2}}} = \frac{\mu_3}{\sigma^3}, \tag{19}$$

where $skew_{i,t}$ denotes the skewness of the i-th OD flow at moment t, x is the sequence of observation data points within the sliding window corresponding to the ith OD flow at time step t, $\bar{x}$ denotes the sample mean, μ_3 denotes the third-order central moment, and σ denotes the standard deviation. The formula for kurtosis is shown below.

$$kurt_{i,t} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2\right)^2} - 3 = \frac{\mu_4}{\sigma^4} - 3, \tag{20}$$

where μ_4 denotes the fourth-order central moment, and 3 is subtracted so that the kurtosis of the normal distribution is 0. The predicted values of volatility are calculated based on the predicted values of variance, as shown below.

$$vola_{i,t} = \sqrt{\sigma^2}. \tag{21}$$

where σ^2 denotes the variance of the traffic samples within that time window. The volatility rate measures the fluctuations in network traffic over a specified period and is used to characterize the uncertainty or trend of changes in traffic. Higher volatility indicates significant variations in traffic rates within a given time window, while lower volatility suggests that the traffic remains relatively stable. Following the aforementioned processing, we can obtain the feature matrix of the traffic, which has a dimensionality of 6N.

$$M = \begin{bmatrix} X_1 & \mu_1 & \sigma_1^2 & skew_1 & kurt_1 & vola_1 \\ X_2 & \mu_2 & \sigma_2^2 & skew_2 & kurt_2 & vola_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_t & \mu_t & \sigma_t^2 & skew_t & kurt_t & vola_t \end{bmatrix}. \tag{22}$$

This matrix integrates the time-varying higher-order moment statistics of the flow with the original flow data, creating a multidimensional input for the LSTM network. This integration enhances the LSTM's capacity to model flow patterns effectively. Fig. 1 illustrates the processing of flow data within the algorithm.

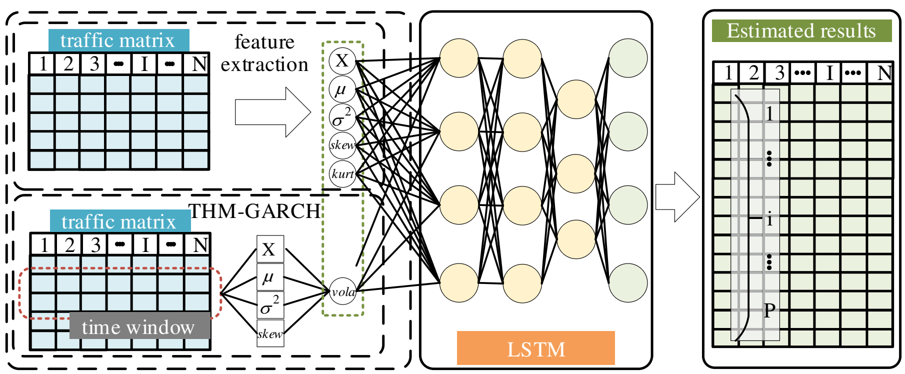

Figure 1: Illustration of the model of the network traffic estimation algorithm TYHM.

5 Experimental Setup and Analysis of Results

In this section, we conduct comprehensive simulation experiments to evaluate the effectiveness of the proposed traffic estimation algorithm. The experimental results are thoroughly analyzed based on the evaluation metrics to validate the efficacy of the proposed algorithm fully.

5.1 Simulation Environment Construction

Our experiments are conducted based on the real IoT topology and traffic dataset Abilene [22]. Specifically, the network topology used contains 12 nodes and 54 links. All the linked physical nodes form a total of 144 OD pairs. Network traffic is sampled every five minutes. Traffic data was collected over three weeks, encompassing a total of 6048 time points. Among them, the sampled data of the first week is used for the training process and the data of the second and third weeks is used for network testing. The detailed parameters of the simulation environment are shown in Table 1. In addition, we selected TomoG [23], SOFIA [24], and SRSVD [25] as the comparison algorithms. The TomoG algorithm has the advantage of fast speed by utilizing link load traffic and routing matrix to compute the OD flow. SOFIA utilizes tensor factorization and temporal pattern detection to achieve the estimation of network traffic. SRSVD implements a compressed sensing technique to address the traffic matrix problem, demonstrating effectiveness even in the presence of a significant number of missing values.

Table 1: Configuration of the experimental environment.

Parameters	Value
Number of nodes	12
Number of links $\lvert L \rvert$	54
Training Data Set	2016
Test Data Set	4032
Time Window	144
Learning Rate	0.7
Dropout probability	0.3

A sliding window mechanism is employed for both training and inference phases. For each time step t, higher-order moments are calculated using a window size of $W = 144$, which corresponds to a 12-h observation period in the 5-min sampled dataset. To ensure numerical stability and gradient efficiency in the LSTM network, the input feature matrix M undergoes global scaling by a factor of 10^4 instead of simple min-max normalization, which preserves the relative magnitude of bursty traffic spikes. In addition, the multi-step prediction is executed using a segment-based recursive strategy. During inference, the LSTM model generates predictions for a continuous time segment. The predicted outputs are then used to re-calculate the time-varying higher-order moments for the subsequent segment, ensuring that the model maintains awareness of the evolving statistical distribution throughout the evaluation horizon. The proposed TYHM framework is implemented using MATLAB R2023b on a workstation equipped with an Intel Core i9-13900K CPU and an NVIDIA RTX 4060 GPU. The LSTM component consists of a single hidden layer with 200 units to maintain a balance between representation capacity and computational efficiency. We employ the Adam optimizer with an initial learning rate of 0.01, which is reduced by a factor of 0.1 every 50 epochs. The network is trained for 100 epochs with a mini-batch size of 32 to ensure stable gradient convergence.

5.2 Simulation Results and Analysis

5.2.1 Algorithm Effectiveness Verification

To verify the effectiveness of the proposed algorithm in the volatility prediction phase, we randomly selected four OD flows to observe their traffic values and volatility trends. Fig. 2 illustrates the trends of predicted volatility values alongside actual traffic values for the second and third weeks. The fluctuation patterns of the two curves in each subfigure are similar. As the fluctuations in the real flow increase, the predicted volatility also rises, and vice versa. This indicates that our algorithm can effectively learn the intensity of network traffic changes from historical data and accurately track its fluctuation characteristics.

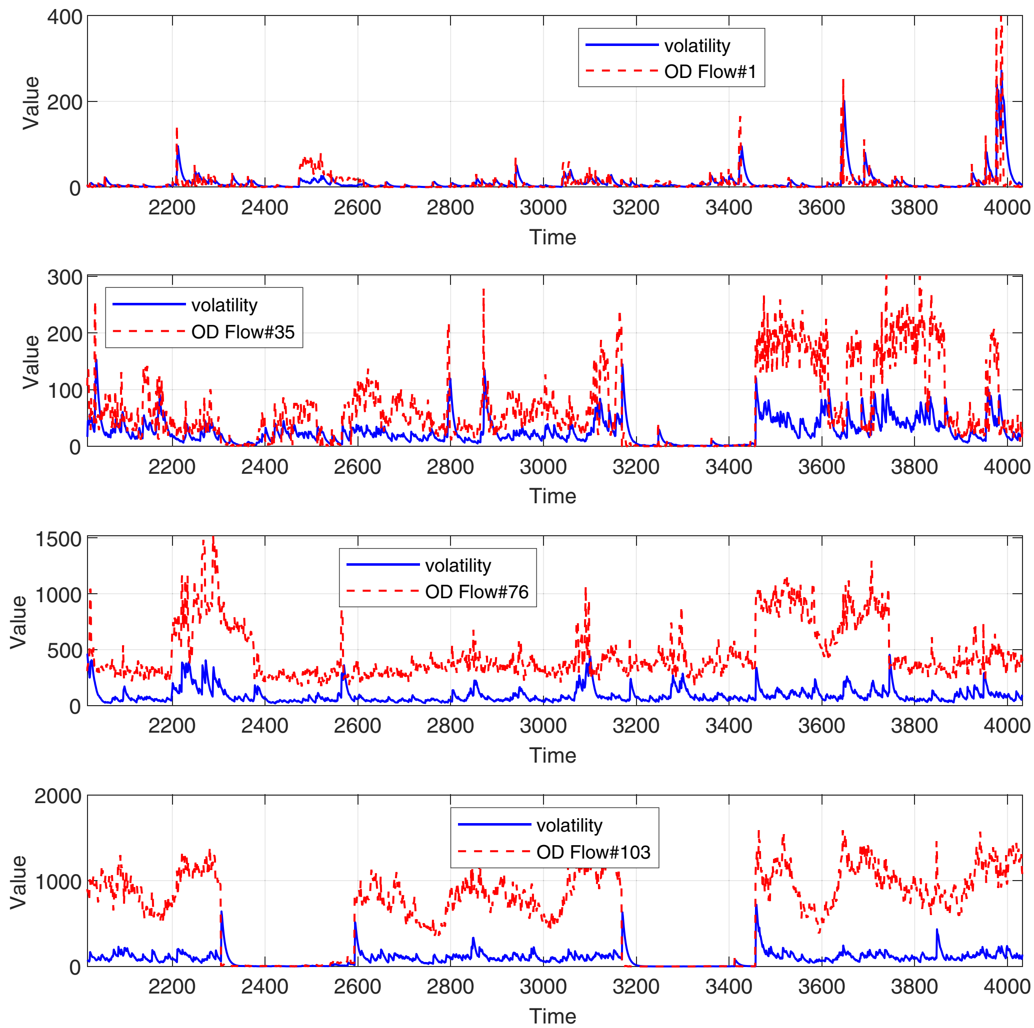

Figure 2: Schematic illustration of the trend in the predicted value of volatility and the real value of flow.

Fig. 3 illustrates the results of comparing the real and estimated values generated by various algorithms for different OD flows. Overall, the TYHM algorithm accurately predicts the fluctuations in flow rate, and its overall trend closely aligns with the true values. Notably, our algorithms demonstrate greater accuracy in predicting the traffic surge, aligning more closely with the true values. In contrast, the SOFIA and SRSVD algorithms tend to overestimate the traffic surge, resulting in significant errors. Conversely, the flow value for OD flow 126 shows a sharp decline near the 3750 time point. In this instance, the TYHM algorithm continues to demonstrate optimal performance. This indicates that the algorithm proposed in this paper effectively reduces prediction errors and adapts well to various complex network scenarios.

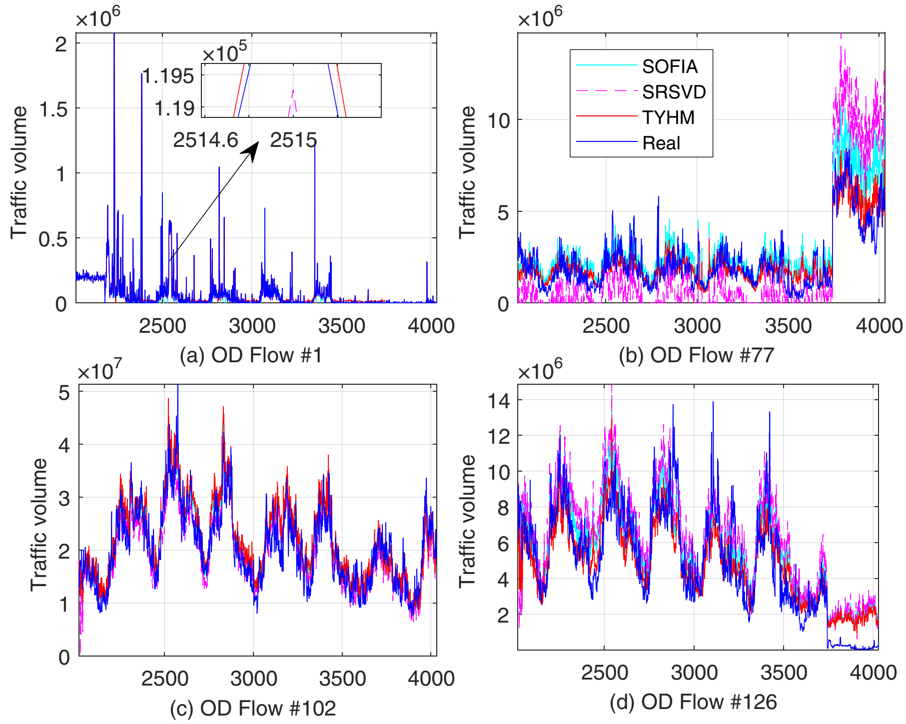

Figure 3: Comparison of real values with those predicted by different algorithms.

To further evaluate the traffic estimation capability of the proposed algorithm, we compare its results with those of the SOFIA algorithm through scatterplot analysis. Specifically, Fig. 4a–d illustrates the estimation results and error ranges for both algorithms. For OD Flow 1 and OD Flow 102, the scatter distributions of the SOFIA and TYHM algorithms are relatively more centralized, but there are still some points outside the 30% error range. In the case of OD Flow 102, most estimates from both algorithms exceed the true values. Conversely, for OD Flow 77 and OD Flow 126, the scatter distribution is more dispersed, suggesting a larger prediction error. In summary, the TYHM algorithm effectively estimates future flows, with the distribution of estimation errors generally maintained within 30% for most flows. By utilizing higher-order moment information from the flow data, our algorithm demonstrates excellent estimation accuracy, even in the presence of sudden changes in flow magnitude.

5.2.2 Impact of Time Windows on Performance

Since the size of the time window $tsize$ is used to compute higher-order moment information, it significantly impacts estimation performance. Therefore, we compare and summarize the estimation errors at different t in Fig. 5. The figure illustrates that the RMSE of the estimation results initially decreases and then increases as the time window expands. The error is largest at a window size of 12. This occurs because when t is too small, the algorithmic model struggles to capture comprehensive higher-order moment information, leading to poor performance. Additionally, the algorithm requires more time to run due to the frequent calculations necessitated by the small time window. When the time window size is 144, the model effectively acquires features and higher-order moment information from the captured data, resulting in traffic estimates that align more closely with the true values. However, when t is excessively large, historical fluctuations in the flow rate can adversely affect the current estimation, leading to a slight decrease in performance.

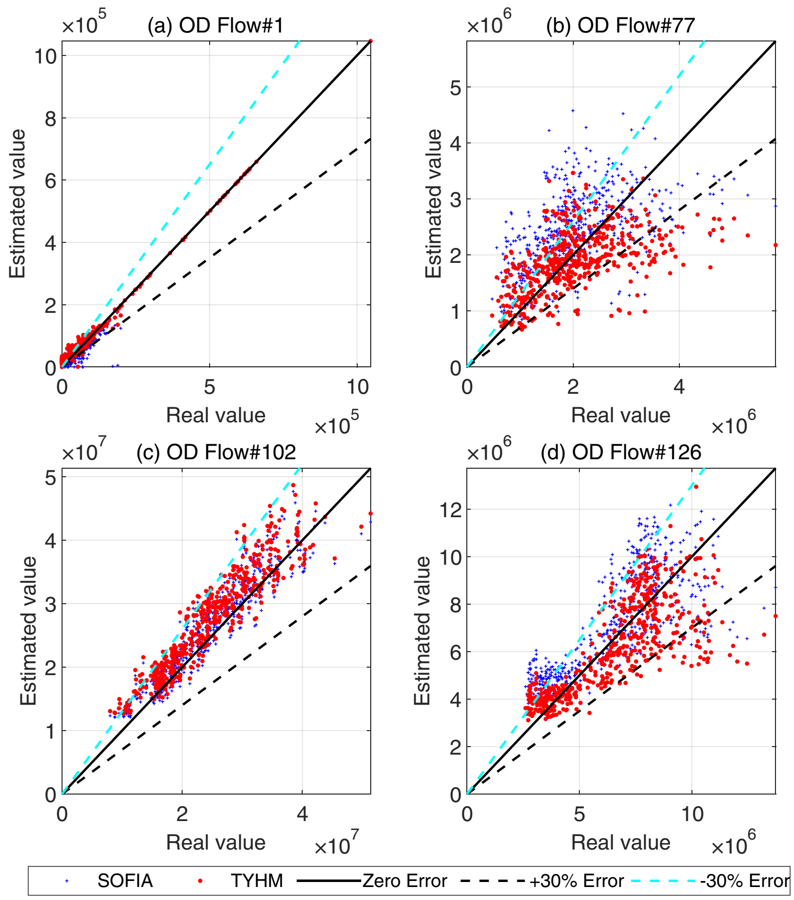

Figure 4: Scatter error distribution of predicted and real values of different algorithms.

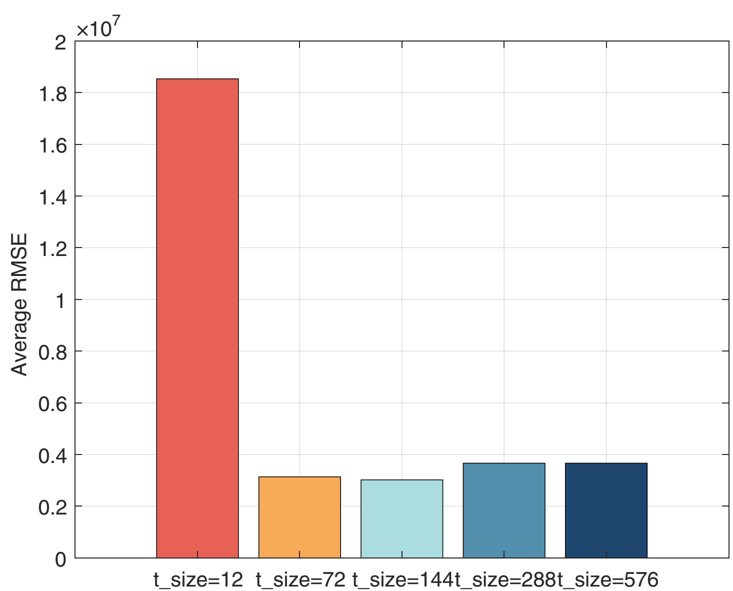

Figure 5: The relationship between predictive performance and time window.

5.2.3 Error Analysis of Traffic Estimates

Time absolute error and time relative error are used to assess the estimation bias of algorithms at various time intervals. Fig. 6 illustrates the average temporal absolute error of five different algorithms using a cloud and rain plot. As depicted in the figure, the TYHM algorithm exhibits a more concentrated error distribution, while the SRSVD algorithm displays the most dispersed error distribution. These outliers in temporal error may arise during peak traffic hours. Our algorithm enhances the model's robustness by incorporating time-varying higher-order moment features, while maintaining high accuracy despite extreme fluctuations in flow. Compared to the LSTM, SOFIA, TomoG, and SRSVD algorithms, the TYHM algorithm reduces median errors by 14.3%, 31.2%, 36.46%, and 44.95%, respectively.

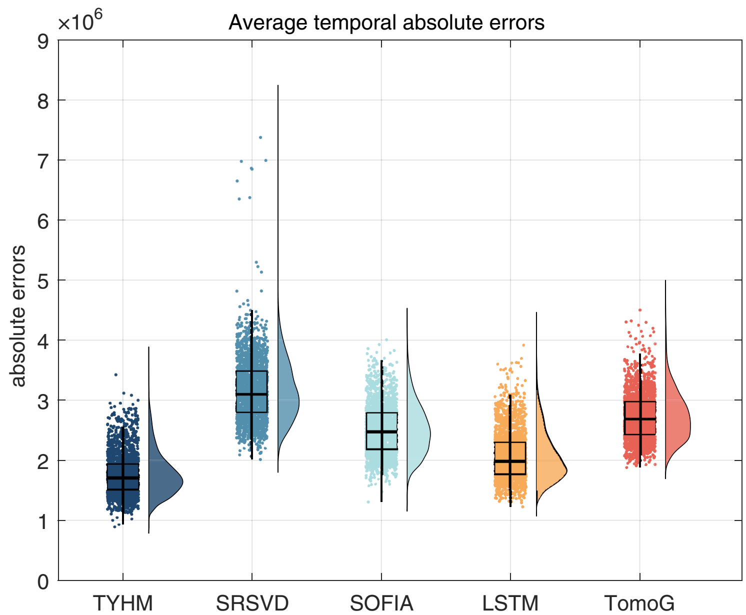

Figure 6: Comparison of the average time absolute error of different algorithms.

The spatial relative error and spatial absolute error can be used to assess the bias in the algorithms' estimations of different OD flows. As illustrated in Fig. 7, the temporal absolute error of the TYHM algorithm is primarily concentrated in lower locations, while the other four algorithms exhibit some outliers with higher errors. The median spatial absolute error of TYHM is reduced by 18.26%, 20.12%, 31.48%, and 43.15% when compared to the LSTM, SOFIA, TomoG, and SRSVD algorithms, respectively. This low median error indicates that TYHM performs effectively in spatial modeling and is capable of addressing the spatial correlation between OD flows.

Fig. 8 illustrates the RMSE of the estimation results obtained using various algorithms. The RMSE of the TYHM algorithm is reduced by 17.78%, 28.28%, 33.31%, and 43.3% compared to the LSTM, SOFIA, TomoG, and SRSVD algorithms, respectively. The TYHM algorithm demonstrates superior performance in terms of average RMSE, highlighting its effectiveness in overall prediction accuracy. This advantage is primarily attributed to the THM-GARCH model's capability to capture the non-Gaussian characteristics of traffic distribution through skewness and kurtosis, as well as the LSTM model's proficiency in modeling long-term dependencies in time series data. In contrast, the other comparison algorithms may struggle to adapt to sudden changes in traffic, resulting in higher RMSE values.

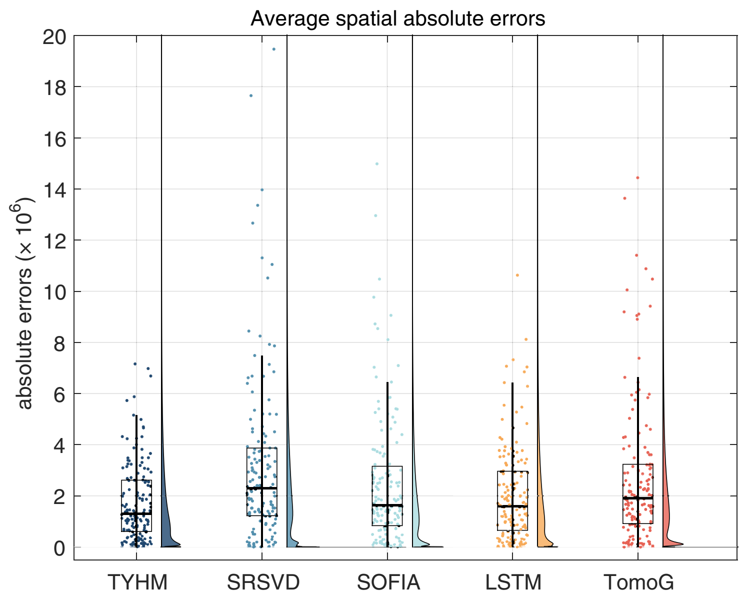

Figure 7: Comparison of the average spatial absolute errors of different algorithms.

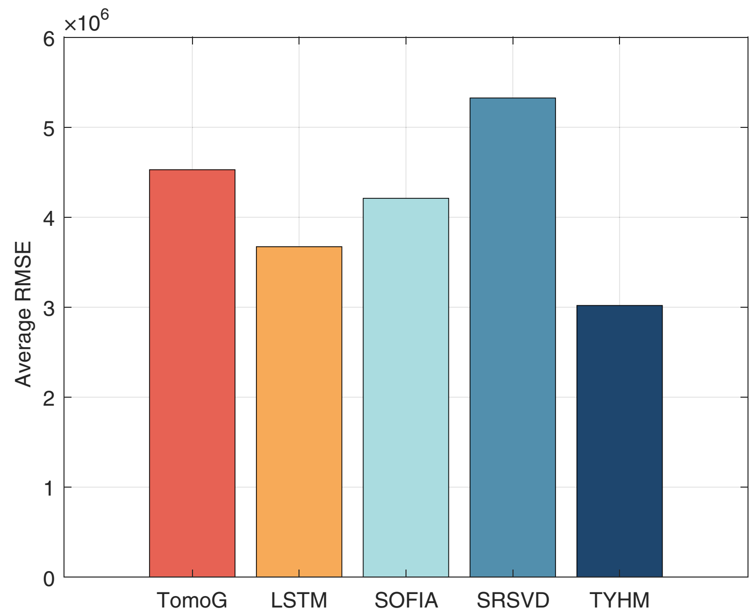

Figure 8: Comparison of average RMSE for different algorithms.

The MAE balances the positive and negative offsets that occur when errors accumulate over time, providing a more accurate reflection of prediction deviation. To illustrate the long-term estimation capabilities of the algorithms, Fig. 9 compares the trends in the MAE of various algorithms for predicting the next seven days. On day seven, the MAE of the TYHM algorithm decreases by 14.69%, 29.69%, 35.28%, and 44.29% compared to the LSTM, SOFIA, TomoG, and SRSVD algorithms, respectively. The increase in MAE on day seven may result from cumulative errors or periodic fluctuations in traffic patterns that the model may not fully capture. Nevertheless, TYHM's consistently low MAE over an extended period indicates its superior ability to model the characteristics of network traffic accurately. The superior performance of TYHM under sudden traffic fluctuations stems from its ability to detect changes in the distribution through higher-order moments. Unlike other baselines, the integration of time-varying skewness and kurtosis enables the model to quantify the asymmetry and heavy-tailed characteristics of traffic bursts. These statistical metrics serve as early warning signals of changes in traffic distribution structure, allowing the LSTM network to adjust its predictions before significant changes in the mean occur. Consequently, incorporating these higher-order features provides a more robust representation of non-stationary characteristics.

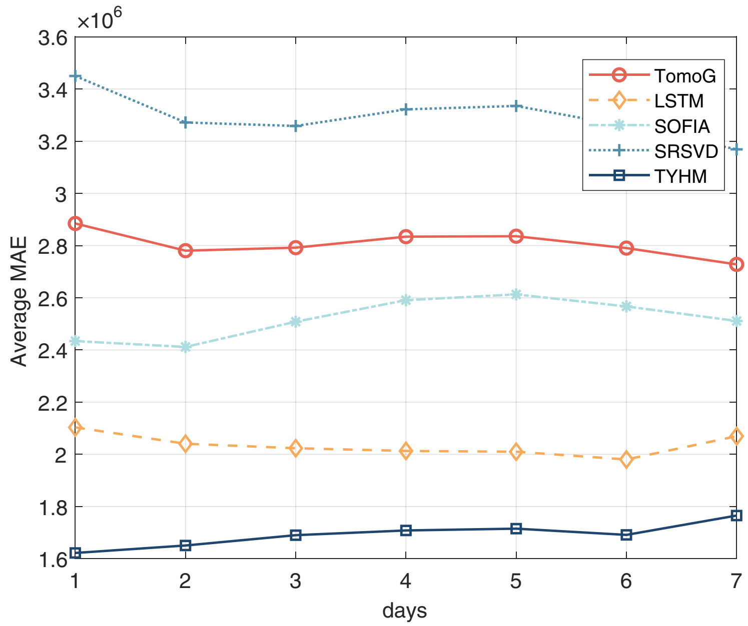

Figure 9: Comparison of average MAE of different algorithms over a week.

Overall, the TYHM framework achieves significant performance improvements compared to other baseline algorithms. First, while some traditional algorithms also focus on the temporal evolution of mean and variance, our framework extends this scope to include time-varying skewness and kurtosis. This enables the model to quantify the asymmetry and heavy-tailed nature of traffic residuals through a dynamically skewed t-distribution, aspects that are often overlooked in standard second-order volatility models. Second, unlike traditional moment-based models that typically rely on static window statistics, TYHM introduces a recursive evolution mechanism for skewness and kurtosis. This approach captures the temporal persistence of structural changes in traffic distributions, enabling the model to distinguish between transient noise and persistent anomalous fluctuations. Third, we propose a dedicated feature matrix that facilitates synergy between statistical methods and neural networks. This integration enables the LSTM network to perceive the underlying distribution patterns of traffic, significantly enhancing the network's inference robustness during periods of sudden traffic surges.

6 Conclusion

In this paper, we propose a time-varying higher-order moments based network traffic estimation algorithm designed to effectively address the challenges posed by high-dimensional, non-stationary, and non-normal traffic data. The THM-GARCH model enhances the LSTM network's ability to model complex time series patterns by incorporating higher-order moments, which provide comprehensive statistical features. The algorithm initially incorporates higher-order moment adjustment factors through a sliding window to preserve distributional properties. Subsequently, the THM-GARCH model estimates the parameters and predicts statistical features, while the LSTM network facilitates long-term estimation of network traffic. Experimental results validate the superiority of this method in terms of prediction accuracy, demonstrating its ability to reduce flow estimation errors and improve adaptability in complex flow scenarios.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Dingde Jiang and Yi Zhang; methodology, software, Xiaoxiong Yang; validation, Dingde Jiang; formal analysis, Xiaoxiong Yang; resources, Shuqing He; data curation, Dingde Jiang; writing—original draft preparation, Xiaoxiong Yang; writing—review and editing, Yi Zhang; visualization, Yi Zhang; supervision, Dingde Jiang; project administration, Shuqing He; funding acquisition, Shuqing He. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data openly available in a public repository.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Azab A, Khasawneh M, Alrabaee S, Choo KKR, Sarsour M. Network traffic classification: techniques, datasets, and challenges. Digit Commun Netw. 2024;10(3):676–92.
2. Zhang Y, Zhang P, Jiang C, Wang S, Zhang H, Rong C. QoS aware virtual network embedding in space-air-ground-ocean integrated network. IEEE Trans Serv Comput. 2024;17(4):1712–23. doi:10.1109/tsc.2024.3357707.
3. Ericsson. Ericsson mobility report. 2024 [cited 2026 Jan 1]. Available from: https://www.ericsson.com/zh-cn/about-us/company-facts/ericsson-worldwide/china/mobility-report.
4. Liu H, Wang H. Real-time anomaly detection of network traffic based on CNN. Symmetry. 2023;15(6):1205. doi:10.3850/978-981-18-8813-7_p129-cd.
5. Kumar R, Venkanna U, Tiwari V. Optimized traffic engineering in software defined wireless network based IoT (SDWN-IoT): state-of-the-art, research opportunities and challenges. Comput Sci Rev. 2023;49(6):100572. doi:10.1016/j.cosrev.2023.100572.
6. Wang F, Xin X, Lei Z, Zhang Q, Yao H, Wang X, et al. Transformer-based spatio-temporal traffic prediction for access and metro networks. J Light Technol. 2024;42(15):5204–13. doi:10.1109/jlt.2024.3393709.
7. Zhang Y, Jiang D, Zhang P, Guizani M, Kostromitin KI. Adaptive resource management in edge networks: task-aware virtual network embedding. IEEE Trans Veh Technol. 2025;75(5):8429–38. doi:10.1109/tvt.2025.3628126.
8. Kumar A, Vidyapu S, Saradhi VV, Tamarapalli V. A multi-view subspace learning approach to internet traffic matrix estimation. IEEE Trans Netw Serv Manag. 2020;17(2):1282–93. doi:10.1109/tnsm.2020.2983329.
9. Li Z, Fu Y, Tian M, Li C, Yu FR, Cheng N. FedSTDN: a federated learning-enabled spatial-temporal prediction model for wireless traffic prediction. IEEE Trans Mob Comput. 2025;24(9):8945–58.
10. Zhou J, Han T, Xiao F, Gui G, Adebisi B, Gacanin H, et al. Multiscale network traffic prediction method based on deep echo-state network for internet of things. IEEE Internet Things J. 2022;9(21):21862–74. doi:10.1109/jiot.2022.3181807.
11. Abbasi M, Shahraki A, Taherkordi A. Deep learning for network traffic monitoring and analysis (NTMA): a survey. Comput Commun. 2021;170(3):19–41. doi:10.1016/j.comcom.2021.01.021.
12. Yang H, Li X, Qiang W, Zhao Y, Zhang W, Tang C. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. Comput Netw. 2021;193(3):108102. doi:10.1016/j.comnet.2021.108102.

13. Dey S, Winter S, Tomko M, Ganguly N. Traffic count estimation at basis links without path flow and historic data. IEEE Trans Intell Transp Syst. 2023;24(10):11410–23. doi:10.1109/tits.2023.3279279.
14. Ephraim Y, Coblenz J, Mark BL, Lev-Ari H. Traffic rate network tomography via moment generating function matching. In: Proceedings of the 2022 56th Annual Conference on Information Sciences and Systems (CISS); 2022 Mar 9–11; Princeton, NJ, USA. p. 148–53.
15. Yu X, Wang H, Wang J, Wang X. A common feature-driven prediction model for multivariate time series data. Inf Sci. 2024;677(3):120967. doi:10.1016/j.ins.2024.120967.
16. Weng W, Fan J, Wu H, Hu Y, Tian H, Zhu F, et al. A decomposition dynamic graph convolutional recurrent network for traffic forecasting. Pattern Recognit. 2023;142(2):109670. doi:10.1016/j.patcog.2023.109670.
17. Ta X, Liu Z, Hu X, Yu L, Sun L, Du B. Adaptive spatio-temporal graph neural network for traffic forecasting. Knowl Based Syst. 2022;242:108199. doi:10.1016/j.knosys.2022.108199.
18. Nie L, Wang X, Zhao Q, Shang Z, Feng L, Li G. Digital twin for transportation big data: a reinforcement learning-based network traffic prediction approach. IEEE Trans Intell Transp Syst. 2024;25(1):896–906. doi:10.1109/TITS.2022.3232518.
19. Dong S, Zhang H. Research on network traffic prediction and management based on logarithmic barrier method. In: Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC); 2020 Dec 11–13; Chongqing, China. p. 627–30.
20. Bi J, Zhang X, Yuan H, Zhang J, Zhou M. A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM. IEEE Trans Autom Sci Eng. 2022;19(3):1869–79. doi:10.1109/tase.2021.3077537.
21. Ge Y, Zhang Y, Shi K, Li H. A moment cross predictor for non-stationary mobile traffic forecasting. In: Proceedings of the 2024 IEEE/CIC International Conference on Communications in China (ICCC); 2024 Aug 7–9; Hangzhou, China. p. 2059–64.
22. Traffic datsets: abilene. 2004 [cited 2026 Jan 1]. Available from: https://dx.doi.org/10.21227/7x3c-5p06.
23. Singhal H, Michailidis G. Structural models for dual modality data with application to network tomography. IEEE Trans Inf Theory. 2011;57(8):5054–71. doi:10.1109/tit.2011.2158474.
24. Lee D, Shin K. Robust factorization of real-world tensor streams with patterns, missing values, and outliers. In: Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE); 2021 Apr 19–22; Chania, Greece. p. 840–51.
25. Roughan M, Zhang Y, Willinger W, Qiu L. Spatio-temporal compressive sensing and internet traffic matrices (extended version). IEEE/ACM Trans Netw. 2012;20(3):662–76. doi:10.1109/tnet.2011.2169424.