Confidence Interval for Traffic Demand Prediction with Coverage Guarantee

Chao Yang [1,2], Xiannan Huang[1], Yan Cheng[1]

[1]Key Laboratory of Road and Traffic Engineering, Ministry of Education at Tongji University
[2]Urban Mobility Institute, Tongji University

Abstract

Accurate short-term traffic demand prediction is critical for the operation of traffic system. Besides point estimation, confidence interval of the prediction is also important because many models about traffic operation, such as shared bike rebalancing and taxi dispatching, take the uncertainty of future demand into account and require confidence interval as the input. However, existing methods require strict assumptions such as unchanging traffic pattern and correct model specification ensure that enough coverage. Therefore, the confidence intervals provided could be invalid, especially in a changing traffic environment. To fill this gap, we propose a simple but efficient method, **CONTINA** (**Con**formal **T**raffic **In**tervals with **A**daptation). The main idea of this method is collecting errors of interval during deployment, and the interval will be widened in the next step if the errors are larger, and shorted otherwise. Besides, we theoretically prove that the coverage of the confidence intervals provided by our method can converge to the target coverage level. Experiments across four real-world datasets and prediction models demonstrate that our method can provide valid confidence interval with shorter length. Our method is able to help traffic management personnel develop more reasonable operation plan in practice. And we release the code, model and dataset in xiannanhuang/CONTINA

Keywords: confidence interval, traffic prediction, conformal prediction

Introduction

Short-term traffic demand prediction refers to forecasting the traffic demand, such as taxi or bike-sharing demand, across different areas of a city for the next half hour or several hours. This prediction is crucial as accurate forecasts can help traffic management authorities to rebalance shared bikes or dispatch taxis efficiently [1]. Such efforts contribute to solving resources, alleviating traffic congestion, and building an environmentally friendly society. Therefore, in recent years, substantial work has emerged in this field and most of them focus on providing more accurate point prediction of future traffic demand.

However, point predictions alone are insufficient. It is also important to consider the confidence intervals of these predictions. This is because traffic demand inherently involves uncertainty, making it nearly impossible to build a perfectly accurate point prediction model. Therefore, many studies on bike rebalancing or taxi dispatching do not rely solely on point predictions. Instead, they consider confidence intervals as inputs for their models [2], [3], [4], [5], [6], [8]. These methods often assume that future bike usage in each area falls into a certain interval, and return the rebalancing plan that keeps efficient as long as the future demand actually falls into those intervals [8], [9], [10].

From this perspective, it can be concluded that the key requirements for predicting confidence intervals are twofold. 1) **Valid**. It means that the model's confidence interval must have a high probability to cover the actual demand in the future. If coverage cannot be guaranteed, it means that future traffic demand may exceed what the rebalancing plan can handle, reducing its efficiency. 2) **Efficient.** It means that when ensuring coverage, the confidence interval should be as short as possible. If the confidence interval is too long, the rebalancing plan would need to account for highly unlikely extreme scenarios, making the plan overly conservative. As a result, **we aim to propose a method to provide valid and effective confidence interval for traffic demand prediction with coverage guarantee** in this paper.

Recently, some researches have focused on predicting confidence intervals for traffic demand prediction [1], [11]. Typically, these methods decompose the uncertainty of model predictions into two parts: model uncertainty and data uncertainty [12], [13], [14]. Model uncertainty is often addressed through model ensemble [12] or MC Dropout [11], [15], while data uncertainty is tackled by methods such as output variance or quantiles [16]. Further details can be found in the literature review section.

While these methods are insightful, they require some rigorous assumptions to ensure valid. For example, the functional form of the model, as well as the distributional assumptions about the errors, are correctly specified. And the data distribution during deployment is consistent with that during training. However, these assumptions may not hold in traffic prediction. For instance, traffic patterns always change over time [17] and it could violate the identical distribution assumption. Therefore, confidence intervals derived using these methods may not maintain adequate coverage rate, and typically, coverage rate deteriorates as deployment. And this will also be demonstrated in our experiments.

To address the issue of confidence intervals deteriorating over time during deployment, we propose a dynamic calibration method for confidence intervals. The idea is intuitive: as the model is deployed, its prediction errors on real data become observable. These observed errors can then be used to estimate the model's future errors. For example, if the model has been deployed for 10 days and the observed errors are 1, 2, 3, ..., 10, we can infer that the error on day 11 is likely to fall within this range. If the model predicts a traffic demand y for day 11, then adding and subtracting 9 to y will result in a 90% confidence interval. This approach, which uses additional data to calibrate the confidence intervals, is known as

conformal prediction [18].

It should be noticed that some other studies have applied conformal prediction to traffic forecasting [17], [19], [20]. However, these works either focus solely on point predictions or use overly simplistic method, which fail to ensure valid. As a result, we reference conformal prediction mothed and propose a valid and efficient confidence interval modeling method in this paper.

In summary, the contributions of our paper are:

1) **Algorithmic Contributions.** We propose **CONTINA** (**Con**formal **T**raffic **In**tervals with **A**daptation), a novel method to provide valid and efficient confidence intervals for traffic demand prediction which combines adaptive conformal prediction, quantile prediction and dynamic learning rate mechanism. Experiments across 4 datasets demonstrate that our approach provides shorter confidence intervals while maintaining sufficient coverage rates.
2) **Theoretical Contributions**. We prove that the average coverage of the confidence intervals provided by our approach will converge to the target coverage rate (Theorem 4.1). Additionally, it is also proved that even for the region with the worst coverage, the coverage of it will still converge to the target coverage rate (Theorem 4.2).

Literature review and preliminary

Modeling Uncertainty in Traffic Demand Prediction

In traffic demand prediction, the modeling of confidence intervals typically separates the prediction uncertainty into model uncertainty and data uncertainty. Model uncertainty refers to the mismatch between the patterns captured by the model and the true underlying patterns, which leads to prediction errors. This can be mitigated by increasing the training data or developing more appropriate models. Data uncertainty reflects the inherent uncertainty in the problem itself. A more rigorous definition, according to Schweighofer [21], is as follows:

Given a training dataset D, model parameters w, and a test sample x, the predictive distribution of the model for y should be:

$$p(y|x, D) = \int p(y|x, w)p(w|D)dw$$

Here, p(w|D) represents the probability of different models learned from the training data. p(y|x,w) represents the predicted distribution of y given a specific model. The prediction uncertainty is defined as the entropy H(p(y|x,D)) of the predictive distribution p(y|x), which can be decomposed as follows:

$$\mathrm{H}\big(\mathrm{p}(\mathrm{y}|\mathrm{x},\mathrm{D})\big) = \int \mathrm{H}[\mathrm{p}(\mathrm{y}|\mathrm{x},\mathrm{w}')]\mathrm{p}(\mathrm{w}'|\mathrm{D})\mathrm{dw}' + \int \mathrm{D}_{\mathrm{kl}}[p(y|x,w')||p(y|x,D)]p(w'|D)dw'$$
$$= E_{w\sim p(w'|D)}H\big(\mathrm{p}(\mathrm{y}|\mathrm{x},\mathrm{w}')\big) + E_{w\sim p(w'|D)}\mathrm{D}_{\mathrm{kl}}[p(y|x,w')||p(y|x,D)]$$

The first term represents the average uncertainty of the predicted distribution for each model, which corresponds to data uncertainty. The second term measures the distance between each model's predicted distribution ($p(y|x,w')$) and the optimal predicted distribution ($p(y|x,D)$), representing model uncertainty.

The primary challenge in obtaining model uncertainty lies in estimating p(w' | D), which means the probability over models given the training dataset. A relatively simplified approach is to use ensemble learning. This involves generating multiple models through different parameter initializations or hyperparameter configurations, and using the variance of predictions across these models to estimate model uncertainty [12], [22], [23]. This method assumes that training each model is equivalent to sampling from p(w'|D). Another approach is to use Bayesian neural networks (BNNs), which treat network parameters not as fixed values but as distributions. The posterior distribution of the parameters is inferred from the training data. Since the posterior distribution is often intractable, methods like variational inference [24] or Markov Chain Monte Carlo (MCMC) [13], [14], [25] are employed to perform sampling.

Methods for modeling data uncertainty include directly modeling the predicted distribution. For example, in [26], the true value is assumed to follow a Gaussian distribution, and the neural network outputs the mean and variance of the distribution. Alternatively, some models assume a negative binomial distribution for the true value and output the corresponding parameters [27]; Outputting confidence interval is another method to obtain data uncertainty [28];

Most studies focusing on confidence interval in traffic prediction combine these two types of uncertainty to derive the final confidence interval. However, this approach has certain limitations. For instance, it is practically impossible to obtain the true posterior distribution of models conditioned on the training data, whether using model ensembles or Bayesian inference. The validation of these algorithms is hard to guarantee and often relies on strong assumptions [21], such as large sample size (assume an infinite amount of training data) and model correctness (the model accurately captures the relationships in the training data). Only under these ideal conditions can these methods provide true model uncertainty. However, such assumptions are nearly impossible in real-world scenarios. In terms of modeling data uncertainty, some approaches assume specific distributions, such as normality, for the data. It is also unverifiable.

Another critical flaw of these methods is their reliance on the assumption that the training and test data come from the same distribution. If this assumption is violated, which is common in traffic prediction [17], these methods will fail. This raises doubts about their applicability and effectiveness in traffic prediction tasks.

In addition, some studies use inherently probabilistic neural networks to model uncertainty, such as Gaussian Process Regression [1], [29] or diffusion models [30], [31]. However, these approaches also typically assume that the training and test data share the same distribution, thereby overlooking the dynamic nature of traffic patterns and leading to invalid confidence interval.

Conformal prediction

Conformal prediction includes full conformal prediction and split conformal prediction. Here, we mainly focus on split conformal prediction, whose core idea is to infer the error on the test set using the error on the validation set, thus obtaining the confidence interval for test samples. Specifically, given a validation set $\{(x_i, y_i)\}_{i=1}^{n}$ and a prediction model f, the procedure works as follows: compute prediction error for each data in validation set and gather these errors together, resulting in a set $E = \{|f(x_i) - y_i| : i = 1,2, \dots, n\}$. Then for a test data x_{n+1}, the prediction interval $C_{1-\alpha}(x_{n+1})$ can be constructed as:

$$C_{1-\alpha}(x_{n+1}) = [f(x_{n+1}) - Q_{1-\alpha}(E), f(x_{n+1}) + Q_{1-\alpha}(E)]$$

Where the $Q_{1-\alpha}(E)$ is the $1-\alpha$ quantile of E. In detail, $Q_{1-\alpha}(E)$ is the $(1-\alpha)n$-th smallest value in E. And there is a theorem vindicating this procedure.

Theorem [32]: If all the n+1 samples are exchangeable, then:

$$1 - \alpha \leq P\big(y_{n+1} \in C_{1-\alpha}(x_{n+1})\big) \leq 1 - \alpha + \frac{1}{n}$$

The definition of exchangeability is: if the π is a permutation of $\{1,2, \dots, n+1\}$ then:

$P((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), (x_{n+1}, y_{n+1})) =$
$P((x_{\pi_1}, y_{\pi_1}), (x_{\pi_2}, y_{\pi_2}), \dots, (x_{\pi_n}, y_{\pi_n}), (x_{\pi_{n+1}}, y_{\pi_{n+1}}))$

The assumption of exchangeability is weaker than i.i.d, which is often used in many models.

In practice, the assumption of exchangeability often does not hold. For example, the traffic pattern in future is not the same as the pattern in the history. As a result, standard conformal prediction methods cannot guarantee the required coverage for confidence intervals in such cases. In the context of dynamic forecasting for time series, methods like online conformal prediction have been proposed.

The earliest work on online conformal prediction introduced a method to adjust the width of confidence intervals based on their performance during deployment [33]. For instance, if a confidence interval fails to cover the true value at a given time step, it is widened for the next step; if it succeeds, it is narrowed [34]. The rates of widening and narrowing are predefined. Subsequent research extended this idea by removing the need for fixed adjustment rates, proposing adaptive approaches using methods like aggregating experts [35], [36] to determine these rates. Some studies framed this as an online convex optimization (OCO) problem [37] and used some OCO algorithms to improve interval width adjustments [38], [39],

[40]. Beyond adjusting interval widths, other studies focused on dynamically updating the calibration set [41], [42]. New data observed during deployment is added to the calibration set, while the oldest data is removed, ensuring the set is updated at each time step. Under certain conditions, this method can also guarantee coverage [41]. Additionally, research has shown that weighting data by similarity to prioritize relevant patterns can also improve coverage accuracy [43].

Apart from online conformal prediction, other related works focus on constructing shorter prediction intervals. Traditional methods often yield intervals of uniform length, which can be suboptimal. Some studies have shown that variance differs across data, suggesting that confidence intervals should be longer in high-variance data and shorter in low-variance data [32]. Accounting for variance can produce data-specific intervals. Others proposed constructing intervals for each test sample using errors from its nearest neighbors in the calibration set [45], [46]. Other approaches include partitioning data by features to assign feature-specific interval lengths [46], [47]. Replacing point prediction models with conditional distribution prediction models [48], [49] to derive intervals has also been attempted. When data distributions are asymmetric, directly adding or subtracting the same value to the predicted mean is inappropriate. Conformal prediction based on quantile regression [51] has been proposed to address this issue.

Recent works have applied conformal prediction methods to model confidence intervals for traffic demand forecasting. However, these approaches are often simplistic, relying on the original conformal prediction framework, which fails to guarantee coverage under dynamic conditions [17]. Additionally, some methods focus solely on pointwise predictions, which is inadequate [19].

Summary

In summary, most existing studies on confidence interval modeling for traffic demand forecasting assume that traffic patterns remain unchanged, which is inconsistent with real-world scenarios. Conformal prediction, particularly its extensions, offers an effective way to account for changing traffic patterns when constructing confidence intervals.

Our work aims to reform some conformal prediction methods from machine learning area and develop a conformal prediction framework specifically tailored to traffic demand forecasting problem, with theoretical guarantees. Furthermore, since traffic demand forecasting involves multiple regions, our method seeks to ensure both global and region-specific coverage.

Method

Problem definition

Suppose that there are n regions and for each region, we need to estimate the confidence interval for inflow and outflow. We use $y_{t,i,1}$ to express the actual inflow in time t for region i, and $y_{t,i,2}$ to the actual outflow. The estimated confidence bound is $[low_{t,i,1}, up_{t,i,1}]$ and $[low_{t,i,2}, up_{t,i,2}]$. And we have a model f to predict the traffic demand in different regions and the confidence level we want to control in $1-\alpha$.

The task is to provide confidence intervals with enough coverage. First, we focus on the average coverage. Suppose we deploy our method for T time steps, the average coverage is defined as:

$$cov = \frac{1}{2nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{2} I\left(y_{t,i,j} \in [low_{t,i,j}, up_{t,i,j}]\right)$$

Additionally, since we provide prediction intervals for each region, we also aim to ensure good coverage for each individual region. This means that even for the region with the lowest coverage, we can still obtain a reasonably high coverage rate. As a result, the following metric is also our focus:

$$minRC = \min_{i}\frac{1}{2T}\sum_{t}^{T}\sum_{j=1}^{2} I\left(y_{t,i,j} \in [low_{t,i,j}, up_{t,i,j}]\right)$$

In conclusion, the target is to ensure:

$$cov = 1-\alpha$$
$$minRC = 1-\alpha$$

Method overview

As the aim of our research is to reform original conformal prediction to a method more suitable for traffic demand prediction task, we need to summarize some challenges about using original conformal prediction in our task.

1) **Asymmetric distribution.** It means that the distribution of traffic demand is not symmetric and the confidence interval constructed by adding and subtracting a same value to the prediction might be inappropriate. For example, if the prediction is 7 and the estimated error is 10, then we will get a prediction interval of [-3,17]. The -3 is pretty unreasonable since the traffic demand is at least 0. As a result, proposing a method for the asymmetric traffic demand distribution is important. 2) **Dynamic traffic pattern**. It means that the traffic pattern is changing over time and data in different time points cannot be considered as exchangeable. This fact will make the confidence interval provided by original conformal prediction invalid. 3) **Multiple regions**. The traffic demand prediction task is a multivariate task and we need to

predict traffic demand in each region. However, the traffic laws might change at different rates in different regions. For example, the traffic law in the region where a new railway station opens might change drastically, but the laws in other regions may not change significantly.

To address the asymmetry in traffic demand, we reference quantile conformal prediction[17], [51], which uses quantile regression to predict different quantiles of prediction and tackling asymmetry in these quantiles.

To overcome the second challenge and obtain confidence intervals with coverage guarantee when traffic pattern changes, we draw inspiration from adaptive conformal prediction [33] and propose to update confidence interval dynamically. For example, if the true demand is not in the confidence interval, it will be elongated in the next time step. Despite its simplicity, this method is efficient and has rigorous theoretical coverage guarantee.

To address the third challenge, when refreshing the lengths of confidence intervals for different regions, the rate of updating should vary according to regions. In consequence, we propose a method that deciding refreshing rate adaptively for each region.

Quantile conformal prediction

As we analyzed before, the original conformal prediction returns the prediction interval like $[\hat{y}-\delta, \hat{y}+\delta]$ and this is a symmetric interval. However, the distribution of traffic demand is **asymmetric** as the traffic demand is always above 0. Therefore, quantile conformal prediction has been proposed to address this challenge. We refer to this idea and reform the traditional point prediction model to quantile prediction model first.

Specifically, we want to transfer point prediction model to predict the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the traffic demand. This transformation is easy to accomplish since most deep-learning-based traffic prediction model consist of a spatial-temporal net to excavate features and a prediction head to get the prediction. And the only change is adjusting the output dimension of the prediction head from 1 to 2 (from predicting only mean value to predict the $\frac{\alpha}{2}$,$1-\frac{\alpha}{2}$ quantiles). Then when training, quantile loss function is used to calculate loss:

$$l_{\frac{\alpha}{2}}\left(y, y_{\frac{\alpha}{2}}\right)=\left(1-\frac{\alpha}{2}\right)\left(y_{\frac{\alpha}{2}}-y\right) I\left(y \leq y_{\frac{\alpha}{2}}\right)+\frac{\alpha}{2}\left(y-y_{\frac{\alpha}{2}}\right) I\left(y>y_{\frac{\alpha}{2}}\right)$$

$$l_{1-\frac{\alpha}{2}}\left(y, y_{\frac{\alpha}{2}}\right)=\frac{\alpha}{2}\left(y_{1-\frac{\alpha}{2}}-y\right) I\left(y \leq y_{1-\frac{\alpha}{2}}\right)+\left(1-\frac{\alpha}{2}\right)\left(y-y_{1-\frac{\alpha}{2}}\right) I\left(y>y_{\frac{\alpha}{2}}\right)$$

Where y is the true traffic value and $y_{\frac{\alpha}{2}}, y_{1-\frac{\alpha}{2}}$ are the predicted $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of traffic demand, $l_{\frac{\alpha}{2}}$ and $l_{1-\frac{\alpha}{2}}$ are the loss function for the two quantiles. And $I(*)$ is indicator

function which equals 1 if the event in the bracket happens and 0 otherwise. Therefore, the total loss is the summation of these 2 losses as:

$$l = l_{\frac{\alpha}{2}} + l_{1-\frac{\alpha}{2}}$$

Then we can calculate gradients for the parameters in the model and use some optimization algorithms to refresh them.

After training, it is needed to adjust the quantile prediction in validation set. We follow the procedure proposed in [19]. For any given region i and inflow outflow index j, suppose there is a data $(x_{t,i,j}, y_{t,i,j})$ in validation set and the predicted low and up quantile are $y_{t,i,j,\frac{\alpha}{2}}, y_{t,i,j,1-\frac{\alpha}{2}}$. The conformal score can be calculated as:

$$\mathrm{e_{t,i,j}} = \max\left\{y_{t,i,j} - y_{t,i,j,1-\frac{\alpha}{2}}, y_{t,i,j,\frac{\alpha}{2}} - y_{t,i,j}\right\} \quad (1)$$

This means if $y_{t,i,j} \leq y_{t,i,j,\frac{\alpha}{2}}$ happens,$e_t = y_{t,i,j,\frac{\alpha}{2}} - y_{t,i,j}$ and if $y_{t,i,j} \geq y_{t,i,j,1-\frac{\alpha}{2}}$, $e_{t,i,j}$ will be $y_{t,i,j} - y_{t,i,j,1-\frac{\alpha}{2}}$. Then all $e_{t,i,j}$ can be collected together as $E_{i,j}$. Then the $1-\alpha$ quantile of E_i can be expressed as $Q_{1-\alpha}(E_i)$. When there is an observation $x_{i,j}$ in test set, the final prediction interval of $y_{i,j}$ can be expressed as:

$$C_{1-\alpha}(x_{i,j}) = [y_{i,j,\frac{\alpha}{2}} - Q_{1-\alpha}(E_{i,j}), y_{i,j,1-\frac{\alpha}{2}} + Q_{1-\alpha}(E_{i,j})] \quad (2)$$

where $y_{i,j,\frac{\alpha}{2}}, y_{i,j,1-\frac{\alpha}{2}}$ are the predicted lower and upper quantile. [19] have proved that if y is the actual value and each data in validation set and test set is exchangeable then: $P(y_i \in C_{1-\alpha}(x_i)) \geq 1-\alpha$. This conclusion means the coverage of quantile conformal prediction can be guaranteed.

Dynamic updating of confidence intervals

Although previous studies have shown that quantile conformal prediction can provide coverage guarantees under the assumption of exchangeable data, our problem involves data that are not exchangeable according the challenge **Dynamic traffic pattern**. Therefore, the aforementioned method of using only the validation set to adjust quantile prediction cannot ensure coverage. To address this issue, we draw inspiration from online conformal prediction and dynamically adjusted our confidence intervals during deployment based on the actual coverage achieved by the past intervals. We will elaborate it in the following part.

For a specific region i, and time step t and index j, the prediction interval is

$$C_{1-\alpha}(x_{t,i,j}) = [y_{t,i,j,\frac{\alpha}{2}} - Q_{1-\alpha_{t,i}}(E_{t,i,j}), y_{t,i,j,1-\frac{\alpha}{2}} + Q_{1-\alpha_{t,i}}(E_{t,i,j})] (4)$$

The above equation means replacing the $1-\alpha$ quantile of $E_{t,i,j}$ with the $1-\alpha_{t,i}$ quantile of $E_{t,i,j}$. We will first explain how we decided the $\alpha_{t,i}$ for each time step, and then we will elaborate how $E_{t,i,j}$ is defined.

$\alpha_{t,i}$ is decided in a iterate style. In time t, we can calculate the coverage error as:

$$err_{t,i} = 1 - \frac{\left(I\left(y_{t,i,1} \in C_{1-\alpha}(x_{t,i,1})\right) + I\left(y_{t,i,2} \in C_{1-\alpha}(x_{t,i,2})\right)\right)}{2} \quad (5)$$

This error means the proportion of true values being out of the confidence interval. Then we use the following equation to obtain $\alpha_{t+1,i}$

$$\alpha_{t+1,i} = \alpha_{t,i} + \gamma_{t,i}(\alpha - err_{t,i}) \quad (6)$$

Where $\gamma_{t,i}$ can be regarded as learning rate and we will explain how to decide it in the following section. Intuitively, this formula compares the actual coverage error at time t with the target coverage error we aim to control. If the actual error exceeds the desired level, $\alpha_{t+1,i}$ decreases. For example, if $\alpha_{t,i}$ is 0.1 at time t, it might reduce to 0.09 at time $t+1$. This means that at the next time step, the 91st percentile will be used in (4) instead of the 90th percentile, resulting in a wider prediction interval, as the 91st percentile is greater than the 90th percentile.

As for $E_{i,t,j}$, we add the most recent conformal score $e_{i,t,j}$ into $E_{i,t}$, and delete the earliest one in each time step, as suggested by [41]. The process is shown in Figure 1 (For simplicity, we omit region index i and inflow outflow index j in this figure).

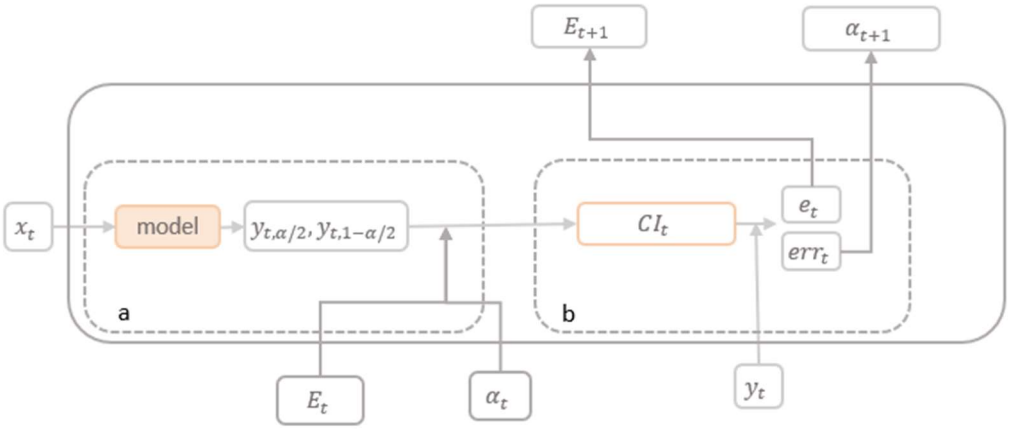

Figure 1 The generation procedure of confidence interval

Adaptive learning rate determination

In Eq6, the learning rate $\gamma_{t,i}$ should be determined. In the earliest adaptive conformal prediction research [33], the learning rate was set as a constant. And some later researches [35], [36] pointed out that a constant learning rate could be suboptimal and the situation in our problem could be even more complicated because the rate of traffic pattern change in different regions might be different. As a result, the learning rate for each region should be distinct.

Besides, setting distinct learning rate for each parameter using second order momentum is common in optimization algorithms for deep learning, such as Adam [52] and SGDM [53]. Therefore, we draw inspiration from these methods and propose to use second order

momentum to adjust the learning rate for each region and result in an adaptive learning rate scheme.

Suppose the initial learning rate is γ_1 and initial moment is $v_{1,i}$ for region i. Then in time step t, we have $v_{t-1,i}$ form the past step and obtain $err_{t,i}$ in this step, then:

$$v_{t,i} = \beta v_{t-1,i} + (1-\beta)\left(err_{t,i} - \alpha\right)^2 \quad (7)$$

Then:

$$\alpha_{t+1,i} = \alpha_{t,i} - \frac{\gamma_1}{\sqrt{v_{t,i}+\epsilon}}\left(err_{t,i} - \alpha\right) \quad (8)$$

Therefore, the learning rate at each time step t for each region i is:

$$\gamma_{t,i} = \frac{\gamma_1}{\sqrt{v_{t,i}+\epsilon}}$$

Where ϵ is a small number used to prevent dividing zero.

[54] discussed the advantages of using second order momentum to adjust learning and proposed that this method could keep the learning more stable and accelerate convergent rate.

We summarize our algorithm as follows:

Algorithm 1 Conformal Traffic Intervals with Adaptation

Require: Training dataset D_1, validation dataset D_2, confidence level α.
Ensure: Obtain test data $(x_{t,i,j}, y_{t,i,j})$ one by one in future T steps.

Train quantile prediction model M using D_1.
for $i \in [1, n]$ **do**
  for $j \in \{1, 2\}$ **do**
    Obtain error set $E_{1,i,j}$ in D_2 using Eq.1.
  end for
  Initialize $\alpha_{1,i} = \alpha$, $v_{1,i} = 0$.
end for
for $t \in [1, T]$ **do**
  for $i \in [1, n]$ **do**
    for $j \in \{1, 2\}$ **do**
      Observe $x_{t,i,j}$.
      Obtain predicted quantiles $y_{t,i,j,\alpha/2}$, $y_{t,i,j,1-\alpha/2}$ using model M.
      Output predicted interval:

$$C_{1-\alpha}(x_{t,i,j}) = \left[y_{t,i,j,\alpha/2} - Q_{1-\alpha_{t,i}}(E_{t,i,j}),\ y_{t,i,j,1-\alpha/2} + Q_{1-\alpha_{t,i}}(E_{t,i,j})\right].$$

      Observe $y_{t,i,j}$.
      Calculate $e_{t,i,j}$ using Eq.1.
      Obtain $E_{t+1,i,j}$ by adding $e_{t,i,j}$ to $E_{t,i,j}$ and deleting the oldest element of it.
    end for
    Calculate error $err_{t,i}$ using Eq.5.
    Obtain $\alpha_{t+1,i}$, $v_{t+1,i}$ using Eq.7 and Eq.8.
  end for
end for

Theoretical results

Theorem 4.1: Average guarantee

If the error defined by Equation (5) is bounded, then:

$$|cov - (1-\alpha)| \leq \frac{c}{T}$$

Where c is a constant

This theorem states that, overall, the coverage rate achieved by our method will converge to the target coverage rate as the deployment time increases. Moreover, the convergence rate is inversely proportional to the deployment time. It should be highlighted that the bounded error and it is not too strict. We do not need any unrealistic assumptions, such as i.i.d. data, larger sample size or correct model specification, to ensure coverage.

Additionally, we aim to establish a result indicating that, even in the region with the lowest coverage, our method can still maintain a reasonable level of coverage or converge to the desired target coverage rate. To achieve this result, we need to introduce an additional assumption: the coverage error defined by Equation (5) at a given time step only depends on the data from the K-most recent time steps and is independent of data from earlier time steps. This assumption is not so strict because it is reasonable to assume the traffic demand in pretty early time is independent of the traffic demand in the future, and this will result in the independence of errors.

Theorem 4.2: coverage guarantee for the worst region:
For any region i *and time step* t*, If:*

$$P\left(err_{t,i}\middle|err_{t-1,i}, err_{t-2,i}, \dots, err_{1,i}\right) = P\left(err_{t,i}\middle|err_{t-1,i}, err_{t-2,i}, \dots, err_{t-K,i}\right)$$

Then:

$$minRC \geq 1 - \alpha - \frac{c_1}{T} - \sqrt{\frac{c_2 K log n}{T}}$$

Where c_1, c_2 are two constants.

This theorem demonstrates that even for the region with the lowest coverage, our method can maintain a relatively high coverage level. Furthermore, as the deployment time increases, the coverage rate will converge to the desired target coverage rate. And it is needed to emphasized that the number of regions n just causes error proportion to $\sqrt{\log n}$, which means that even for a city with a larger number of regions, the worst regional coverage will not deteriorate very much.

Experiments

Datasets

In the experiment, we used 4 datasets. Each dataset spans 16 months, for January to the April of the next year.

NYCbike: This dataset covers shared bike usage record in New York city from *January 2022*

to April 2023. Each entry represents a bike pickup and drop-off event. Following previous methods in [55], we divided New York into 200 grids and calculated the bike pickup and drop-off quantities for each grid every hour.

NYCtaxi: This dataset covers taxi usage data from *January 2018 to April 2019* in New York city, with hourly usage for each region, including both pick-up and drop-off dimensions. The division of regions is based on the scheme provided by the official website.

CHIbike: The dataset contains shared bike usage record in Chicago from *January 2022 to April 2023.* Chicago is divided into 200 grids and hourly bike pickup and drop-off quantities for each grid are collected.

CHItaxi: This dataset contains hourly taxi pick-up and drop-off values of census regions in Chicago from *January 2016 to April 2017.*

The descriptions of datasets are summarized in Table 1 and we plot the regions or grids of each dataset in Figure 2.

Table1 Descriptions of datasets

Dataset	No. regions	Mena length	Mena wide	Mean area	Mean usage	Std of usage
NYCBike	200	1.02	2.89	2.95	17.8	51.0
NYCTaxi	263	7.59	7.69	32.45	41.8	111.2
ChiBike	200	0.92	0.82	0.75	2.03	5.20
ChiTaxi	171	0.97	1.06	0.89	10.3	41.0

Figure 2 The girds or regions for datasets. a) the grids of NYCbike dataset, b) the regions of NYCtaxi dataset, c) the grids of CHIbike dataset, d) the regions of CHItaxi datasets.

Setup

The task is predicting bike/taxi usage number in the following hour using usage number in the preceding 6 hours. We used data from January to November for training and data in December for validation, then deployed the trained model on data from January to April in next year. And the grids or regions with average bike or taxi usage blow 2 are deleted in the experiments.

To validate that our method can be applied to a wide range of models, we select 4 classic spatial temporal prediction models, STGCN [56], DCRNN [57], MTGNN [58] and GWNET [59], as our prediction model. $\alpha \; is \; set \; as \; 0.1$, in another word, target coverage rate is 90%. In adaptive learning rate algorithm, β is set as 0.99, γ_1 is 0.005, and ϵ is e^-8.

Baselines and evaluation metrics

We choose baseline methods form three perspectives:

1) Traditional confidence interval prediction methods: quantile regression (QR), Bootstrap[12], MC dropout [60], directly minimizing Mean interval score (MIS) [61].

2) Methods for confidence interval modeling in traffic prediction task: DESQRUQ[12], ProbGNN[14], UATGCN [13] and QUANTARFFIC[17].

3) Conformal prediction and its online versions: CP (traditional conformal prediction) [18], ACI (adaptive conformal prediction) [33], DtACI [36], QCP (quantile conformal prediction) [51],

We used 3 metrics to evaluate the quality of confidence intervals:
Coverage (Cov): The proportion of true values included within the confidence interval. Defined as:

$$cov = \frac{1}{2nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{2} I\left(y_{tij} \in \left[low_{tij}, up_{tij}\right]\right)$$

Minimum Regional Coverage (minRC): The coverage of the confidence interval in the region with the lowest coverage. Defined as:

$$minRC = \min_{i} \frac{1}{2T}\sum_{t}^{T}\sum_{j=1}^{2} I\left(y_{tij} \in \left[low_{tij}, up_{tij}\right]\right)$$

Length: The average length of the confidence interval. Define as:

$$Length = \frac{1}{2nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{2} up_{tij} - low_{tij}$$

Additionally, to evaluate how the quality of confidence intervals changes over time, we calculated the four above-mentioned metrics for each month from January to April.

Results

The results of our experiments are summarized in Table2. The result with coverage greater than 88% and minimum regional coverage greater than 85% is considered as valid and we color the cells of valid results in blue. Among these valid results, the minimum length is expressed in red text and the second minimum length is expressed in green text. Besides, the results in Table2 are the average values among all 4 prediction models, and we leave the full results in Appendix C.

Table 2 results of experiments

Dataset	time	metric	QR	MC-dropout	boostrap	MIS	DESQRUQ	UATGCN	ProbGNN	QuanTraffic	CP	ACI	QCP	DtACI	CONTINA
NYCbike	1month	cov	89.6%	54.8%	30.8%	88.7%	91.7%	91.7%	93.1%	91.3%	89.2%	89.8%	90.0%	89.6%	89.6%
		length	0.265	0.218	0.084	0.278	0.284	0.284	0.306	0.288	0.285	0.300	0.266	0.291	0.276
		minRC	81.1%	36.4%	20.4%	80.1%	86.6%	85.0%	87.8%	86.4%	84.3%	88.8%	85.3%	87.3%	88.7%
	2month	cov	90.0%	55.8%	30.7%	89.0%	91.9%	92.3%	93.4%	92.0%	89.3%	90.1%	90.5%	90.0%	89.8%
		length	0.270	0.227	0.087	0.283	0.283	0.291	0.315	0.285	0.289	0.304	0.271	0.298	0.275
		minRC	80.9%	37.6%	18.6%	79.6%	88.0%	86.6%	87.9%	86.4%	84.8%	88.9%	84.1%	85.9%	89.3%
	3month	cov	89.3%	56.2%	30.6%	88.6%	91.2%	91.9%	93.0%	91.4%	87.6%	89.7%	89.9%	89.2%	89.8%
		length	0.289	0.250	0.094	0.301	0.296	0.315	0.340	0.306	0.295	0.322	0.290	0.321	0.298
		minRC	80.5%	38.1%	18.1%	80.9%	84.5%	85.0%	86.5%	84.8%	82.3%	88.6%	82.4%	86.8%	89.2%
	4month	cov	87.7%	57.1%	30.5%	87.6%	89.8%	90.3%	91.3%	89.8%	82.9%	90.0%	88.4%	89.3%	89.7%
		length	0.343	0.323	0.117	0.355	0.356	0.386	0.419	0.360	0.312	0.411	0.345	0.400	0.363
		minRC	73.1%	33.4%	19.9%	72.8%	78.8%	78.4%	80.8%	76.8%	69.1%	88.8%	74.1%	88.4%	89.1%
	AVG	cov	89.2%	56.0%	30.6%	88.5%	91.2%	91.5%	92.7%	91.1%	87.3%	89.9%	89.7%	89.5%	89.7%
		length	0.292	0.254	0.095	0.304	0.305	0.319	0.345	0.310	0.295	0.334	0.293	0.327	0.303
		minRC	78.9%	36.4%	19.2%	78.3%	84.5%	83.8%	85.8%	83.6%	80.1%	88.8%	81.5%	87.1%	89.1%
NYCtaxi	1month	cov	87.9%	64.5%	47.4%	88.9%	92.4%	89.6%	94.3%	82.6%	91.0%	89.9%	89.5%	90.2%	89.7%
		length	0.241	0.148	0.096	0.243	0.262	0.264	0.300	0.294	0.283	0.270	0.258	0.271	0.237
		minRC	77.3%	31.8%	34.2%	77.9%	82.9%	75.1%	83.0%	59.8%	78.4%	88.6%	80.4%	87.1%	89.0%
	2month	cov	87.3%	64.9%	46.5%	88.7%	92.0%	89.0%	93.8%	82.3%	89.7%	89.9%	89.0%	89.7%	89.9%
		length	0.248	0.141	0.099	0.252	0.270	0.275	0.312	0.301	0.281	0.284	0.265	0.280	0.248
		minRC	75.4%	31.9%	34.1%	76.5%	63.4%	73.6%	80.8%	59.7%	76.7%	88.6%	78.9%	87.4%	89.3%
	3month	cov	86.6%	65.2%	46.4%	88.7%	91.8%	88.9%	93.7%	81.8%	89.6%	90.1%	88.5%	89.9%	89.9%
		length	0.251	0.138	0.098	0.252	0.272	0.275	0.311	0.304	0.280	0.286	0.268	0.280	0.248
		minRC	72.6%	26.3%	33.9%	77.0%	77.1%	71.2%	78.0%	56.9%	73.9%	88.5%	67.0%	85.9%	89.2%
	4month	cov	83.1%	65.6%	46.3%	88.9%	92.0%	89.0%	93.7%	80.9%	89.8%	90.0%	87.5%	90.3%	89.9%
		length	0.259	0.214	0.098	0.250	0.274	0.271	0.312	0.312	0.281	0.290	0.275	0.285	0.252
		minRC	74.2%	32.0%	32.9%	79.5%	80.3%	73.8%	77.7%	55.8%	71.9%	88.9%	78.5%	86.7%	89.4%
	AVG	cov	86.2%	65.1%	46.7%	88.8%	92.1%	89.1%	93.9%	81.9%	90.0%	90.0%	88.6%	90.0%	89.8%
		length	0.250	0.160	0.098	0.249	0.270	0.271	0.309	0.303	0.281	0.283	0.267	0.279	0.246
		minRC	74.9%	30.5%	33.8%	77.7%	75.9%	73.4%	79.9%	58.1%	75.2%	88.6%	76.2%	86.8%	89.2%
CHIbike	1month	cov	89.5%	29.8%	22.9%	89.8%	93.5%	92.1%	93.6%	87.0%	90.0%	90.2%	89.9%	90.2%	89.7%
		length	0.514	0.188	0.107	0.513	0.531	0.553	0.593	0.527	0.623	0.638	0.514	0.624	0.521
		minRC	82.4%	19.0%	9.3%	83.3%	90.4%	86.9%	90.3%	74.9%	86.2%	88.3%	83.2%	86.4%	89.0%
	2month	cov	89.1%	31.4%	24.0%	89.4%	93.1%	92.1%	93.6%	86.8%	88.1%	89.5%	89.5%	89.1%	89.8%
		length	0.553	0.215	0.120	0.554	0.572	0.593	0.637	0.566	0.624	0.696	0.553	0.655	0.563
		minRC	81.4%	20.1%	10.7%	84.0%	89.8%	87.3%	90.2%	75.9%	83.9%	87.9%	82.8%	86.5%	89.2%
	3month	cov	88.7%	32.3%	23.8%	89.1%	92.3%	91.4%	93.1%	86.8%	86.6%	90.2%	88.9%	90.0%	89.8%
		length	0.613	0.251	0.134	0.617	0.635	0.657	0.706	0.625	0.655	0.789	0.613	0.765	0.629
		minRC	81.7%	21.0%	10.8%	83.8%	89.0%	87.4%	90.4%	77.4%	82.2%	89.4%	82.9%	87.8%	89.4%
	4month	cov	87.5%	36.0%	22.8%	88.1%	90.9%	90.5%	92.4%	86.4%	79.8%	90.0%	87.7%	89.4%	89.9%
		length	0.835	0.405	0.188	0.844	0.889	0.906	0.964	0.843	0.713	1.124	0.835	1.059	0.883
		minRC	83.0%	24.2%	12.6%	79.5%	87.4%	86.3%	89.5%	78.6%	71.0%	88.6%	83.0%	86.7%	89.1%

	AVG	cov	88.7%	32.4%	23.4%	89.1%	92.5%	91.5%	93.2%	86.8%	86.1%	90.0%	89.0%	89.7%	89.8%
		length	0.629	0.265	0.137	0.632	0.657	0.677	0.725	0.640	0.654	0.812	0.629	0.776	0.649
		minRC	82.1%	21.1%	10.8%	82.6%	89.2%	87.0%	90.1%	76.7%	80.8%	88.6%	83.0%	86.9%	89.2%
CHItaxi	1month	cov	90.3%	46.3%	44.4%	91.3%	93.2%	92.8%	94.4%	89.2%	92.0%	90.0%	90.9%	90.2%	89.6%
		length	0.208	0.269	0.081	0.220	0.222	0.279	0.307	0.227	0.297	0.273	0.210	0.265	0.219
		minRC	83.7%	15.0%	27.3%	85.6%	87.2%	87.8%	91.0%	86.9%	89.6%	87.5%	83.4%	87.5%	89.0%
	2month	cov	90.6%	47.9%	44.2%	91.6%	93.4%	93.1%	94.5%	89.5%	90.8%	90.0%	91.1%	90.0%	89.8%
		length	0.221	0.297	0.085	0.234	0.235	0.300	0.330	0.240	0.283	0.282	0.222	0.264	0.228
		minRC	84.7%	15.5%	27.1%	86.2%	88.4%	88.4%	91.4%	86.6%	88.3%	87.6%	84.3%	87.5%	89.4%
	3month	cov	89.8%	49.0%	43.3%	90.8%	92.6%	92.4%	93.9%	89.6%	88.6%	89.7%	90.4%	89.7%	89.8%
		length	0.241	0.333	0.096	0.257	0.258	0.331	0.365	0.262	0.283	0.326	0.243	0.294	0.256
		minRC	83.3%	16.9%	25.7%	84.7%	86.9%	86.5%	89.3%	86.5%	84.0%	86.9%	82.9%	86.5%	89.4%
	4month	cov	90.1%	48.7%	42.8%	91.1%	92.8%	92.5%	94.3%	89.3%	89.4%	90.1%	90.6%	90.2%	89.9%
		length	0.237	0.319	0.091	0.252	0.252	0.319	0.351	0.257	0.288	0.320	0.238	0.302	0.251
		minRC	82.1%	20.4%	28.3%	84.7%	84.7%	86.3%	90.3%	84.2%	82.3%	88.2%	82.0%	87.0%	89.4%
	AVG	cov	90.2%	48.0%	43.7%	91.2%	93.0%	92.7%	94.3%	89.4%	90.2%	89.9%	90.8%	90.0%	89.8%
		length	0.227	0.305	0.088	0.241	0.242	0.307	0.338	0.247	0.288	0.300	0.228	0.281	0.238
		minRC	83.4%	17.0%	27.1%	85.3%	86.8%	87.3%	90.5%	86.0%	86.1%	87.5%	83.1%	87.1%	89.3%
1st			0	0	0	0	0	0	0	0	0	0	1	0	19

First, from an overall perspective, our method achieved the best performance in 19 out of 20 cases and the second-best result in the remaining one case. This validates the effectiveness of our proposed approach, which is capable of achieving shorter confidence intervals while maintaining coverage.

Next, we report the experimental results for each dataset in detail. For the NYCbike dataset, several methods (DESQRUQ, UATGCN, ProbGNN) are able to maintain validity during the first month; however, by the last month, only three conformal prediction-based methods (ACI, DtACI, CONYINA) could produce valid confidence intervals. This indicates that the effectiveness of traditional confidence interval construction methods may gradually diminish over time. Additionally, for many confidence interval construction methods (QR, MIS), although they sometimes achieve a coverage rate of near 90% on average, their performance can be poor in the worst-case region, even below 75%.

The situation for the NYCtaxi dataset is similar to that of the NYCbike dataset. While some methods (DESQRUQ, UATGCN, ProbGNN) could achieve the target coverage rate on average, their coverage in the worst-performing region are significantly lower than required. Among all valid methods, our approach consistently produced the shortest confidence intervals.

For the two datasets from Chicago, our method also achieved the shortest confidence intervals while ensuring coverage guarantees. Compared to the datasets from New York, the baseline methods show slightly more competitive performance. Notably, many baseline methods are able to maintain validity, particularly in the worst-performing region, where their coverage rates do not drop significantly below the overall coverage. This suggests that the

heterogeneity of traffic patterns across different regions in the Chicago Bike and Chicago Taxi datasets may not be as pronounced as in the New York datasets. Furthermore, the decline in coverage rates over time for the baseline methods is less significant in the Chicago datasets, indicating that the changes in traffic patterns in the Chicago datasets may not be as substantial as those in the New York datasets.

Finally, the average coverage and regional minimum coverage obtained by our approach is always greater 89% and 88%, respectively, which cannot be achieved by the other method. This demonstrates the effectiveness of our approach to maintain coverage. And if we adjust the threshold for regional minimum coverage from 85% to 86%, our method can provide the best result in 20 out of 20 cases.

Method analysis

Sensitive Analysis of initial learning rate

We conducted an additional sensitivity analysis for the initial learning rate, γ_1. Specifically, we adjusted γ_1 to four different values: 0.001, 0.002, 0.005, and 0.01, and repeated our experiments. The results of NYCbike dataset are presented in the following Figure 3.

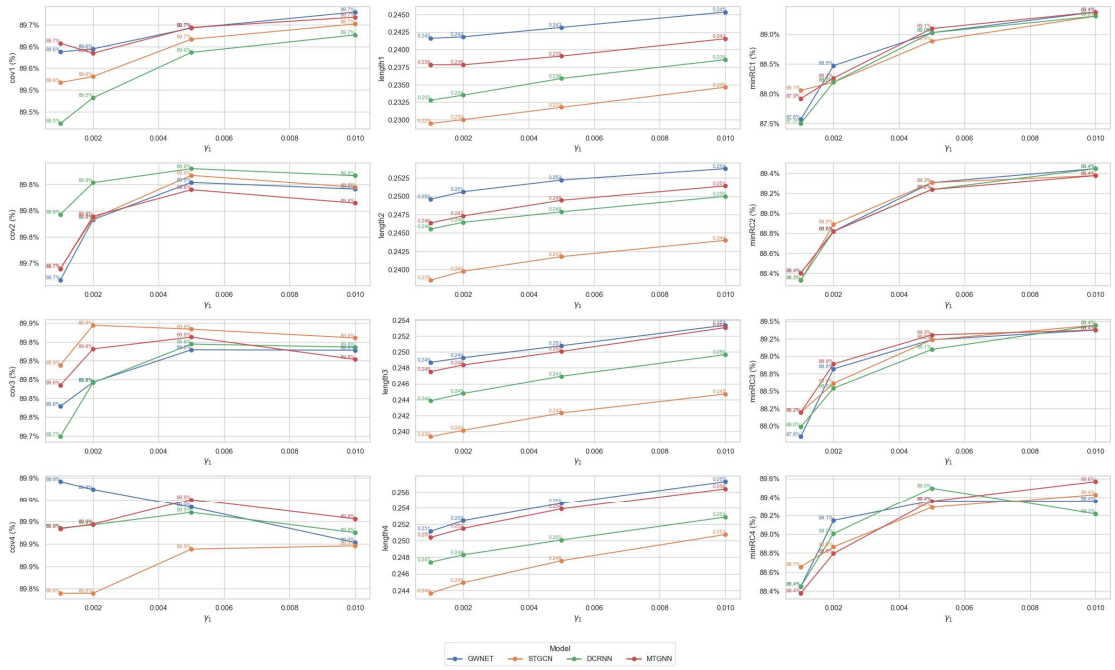

Figure 3 results of using different initial rates in NYCbike dataset

It can be observed that our method is relatively robust to different initial learning rates. When the initial learning rate is adjusted, the average coverage rate obtained by our method generally remains above 89%, and the coverage rate in the worst-case regions stays above 87.5%. The length of the returned intervals tends to become longer as the learning rate

increases, but the magnitude of this change is very limited.

The benefit of using adaptive learning rate

To validate whether using an adaptive learning rate for different regions improves the results of the confidence interval, we conducted additional experiments by replacing the adaptive learning rates with a fixed learning rate. We summarized the coverage of the confidence interval for each region and each day, and plotted the results in Figure 4 to Figure 7. The solid lines represent the mean coverage across all regions in each day. We also calculate the standard deviation of coverage across regions in each day and plot it in shadow part. (These Figures 4-7 show the situations where GWNET was used as prediction model and the results for other prediction models are in Appendix.)

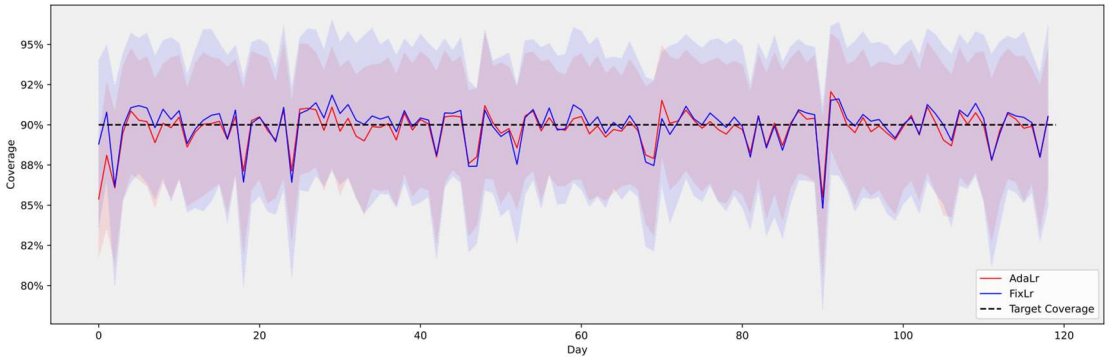

Figure 4 Daily regional coverage for NYCbike dataset

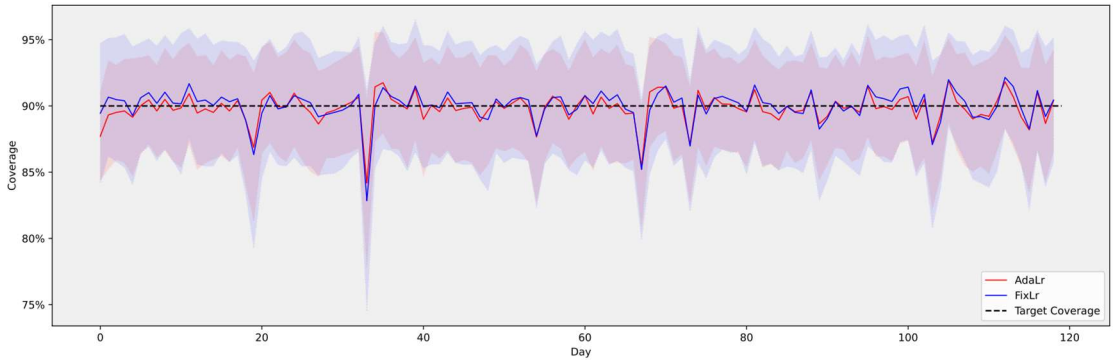

Figure 5 Daily regional coverage for NYCtaxi dataset

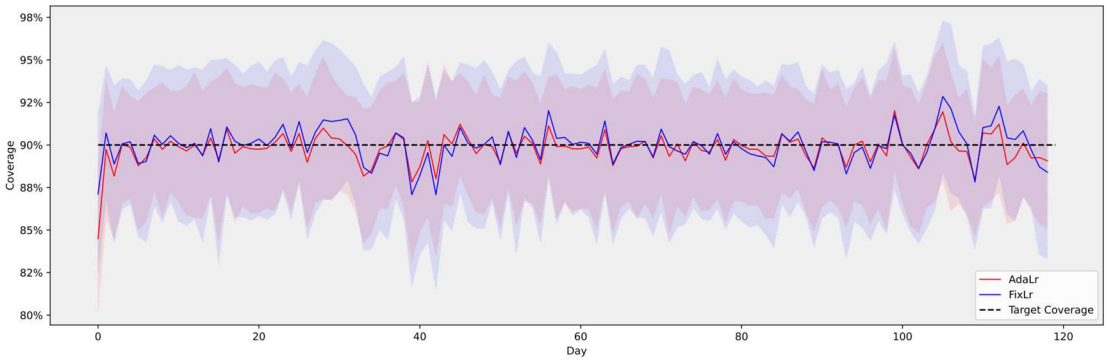

Figure 6 Daily regional coverage for CHIbike dataset

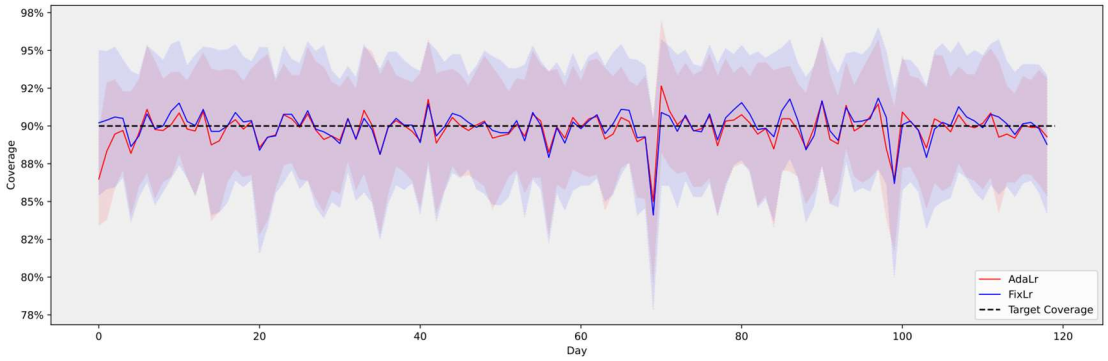

Figure 7 Daily regional coverage for CHItaxi dataset

The results suggest that when using an adaptive learning rate, the coverage for all regions is more concentrated around 90%. However, when using a fixed learning rate, the coverages across regions are more spread out. This indicates that while the overall coverage may still be around 90%, some regions may have a coverage rate greater than 95%, while others may have a coverage rate below 85%. This disparity suggests that fixed learning rate cannot handle the varying transportation patterns in different regions as well as adaptive learning rate. For some regions, the learning rate could be too fast, while for others, it could be too slow. In conclusion, using an adaptive learning rate results in more stable performance, with coverage more consistently aligned with the target rate.

Conclusion

In this paper, a method for constructing confidence intervals for traffic demand forecasting is proposed. Unlike traditional approaches, the method accounts for potential changes in traffic patterns during deployment. To overcome changes in traffic pattern, an adaptive quantile conformal prediction method is introduced. And an adaptive learning rate scheme is used to manage the heterogeneity of traffic changes across different regions.

Theoretical guarantees for the proposed method are also provided. First, our method ensures that the desired coverage rate is achieved at the citywide level. Second, even for the regions with the poorest coverage, our method delivers satisfactory performance. Furthermore, both overall and regional coverage rates converge to the desired levels as the deployment period increases. Experiments were conducted using real-world data from bike-sharing and taxi systems to validate the effectiveness of the proposed method. Notably, for regions with the poorest coverage, the coverage rates provided by our method significantly exceeded those achieved by existing approaches.

The proposed approach has practical applications in traffic operation, such as shared-bike rebalancing or taxi dispatching. Additionally, the method allows for real-time monitoring of

the prediction interval widths. When the intervals become excessively wide, practitioners are alerted that the current model no longer reflects traffic patterns adequately and need some updates.

Reference

[1] M. Xu, Y. Di, H. Yang, X. Chen, and Z. Zhu, 'Multi-task supply-demand prediction and reliability analysis for docked bike-sharing systems via transformer-encoder-based neural processes', *Transportation Research Part C: Emerging Technologies*, vol. 147, p. 104015, Feb. 2023, doi: 10.1016/j.trc.2023.104015.

[2] L. Yu, Y. Xu, H. Shi, and Z. Zhang, 'Robust Optimization Model Based on a Static Rebalancing Design for Bike- Sharing Systems Affected by Demand Uncertainty', *IEEE Access*, vol. 12, pp. 73936–73949, 2024, doi: 10.1109/ACCESS.2024.3405643.

[3] Z.-Y. Zhang and X. Zhang, 'Shared Bikes Scheduling Under Users' Travel Uncertainty', *IEEE Access*, vol. 8, pp. 3123–3143, 2020, doi: 10.1109/ACCESS.2019.2961628.

[4] F. Miao *et al.*, 'Data-Driven Robust Taxi Dispatch Under Demand Uncertainties', *Ieee T Contr Syst T*, vol. 27, no. 1, pp. 175–191, Jan. 2019, doi: 10.1109/TCST.2017.2766042.

[5] M. Hua, X. Chen, J. Chen, and Y. Jiang, 'Minimizing fleet size and improving vehicle allocation of shared mobility under future uncertainty: A case study of bike sharing', *Journal of Cleaner Production*, vol. 370, p. 133434, Oct. 2022, doi: 10.1016/j.jclepro.2022.133434.

[6] Q. Chen, S. Ma, H. Li, N. Zhu, and Q.-C. He, 'Optimizing bike rebalancing strategies in free-floating bike-sharing systems: An enhanced distributionally robust approach', *Transportation Research Part E: Logistics and Transportation Review*, vol. 184, p. 103477, Apr. 2024, doi: 10.1016/j.tre.2024.103477.

[7] S. Huang, K. Liu, and Z.-H. Zhang, 'Column-and-constraint-generation-based approach to a robust reverse logistic network design for bike sharing', *Transportation Research Part B: Methodological*, vol. 173, pp. 90–118, Jul. 2023, doi: 10.1016/j.trb.2023.04.010.

[8] R. Zhao, Z. Tian, L. Tian, W. Liu, and D. Z. W. Wang, 'Research on rebalancing of large-scale bike-sharing system driven by zonal heterogeneity and demand uncertainty', *Transportation Research Part C: Emerging Technologies*, vol. 170, p. 104933, Jan. 2025, doi: 10.1016/j.trc.2024.104933.

[9] Q. Chen, C. Fu, N. Zhu, S. Ma, and Q.-C. He, 'A target-based optimization model for bike-sharing systems: From the perspective of service efficiency and equity', *Transportation Research Part B: Methodological*, vol. 167, pp. 235–260, Jan. 2023, doi: 10.1016/j.trb.2022.12.002.

[10] C. Fu, N. Zhu, S. Ma, and R. Liu, 'A two-stage robust approach to integrated station location and rebalancing vehicle service design in bike-sharing systems', *European Journal of Operational Research*, vol. 298, no. 3, pp. 915–938, May 2022, doi: 10.1016/j.ejor.2021.06.014.

[11] A. Sengupta, S. Mondal, A. Das, and S. I. Guler, 'A Bayesian approach to quantifying uncertainties and improving generalizability in traffic prediction models', *Transportation*

Research Part C: Emerging Technologies, vol. 162, p. 104585, May 2024, doi: 10.1016/j.trc.2024.104585.

[12] T. Mallick, J. Macfarlane, and P. Balaprakash, 'Uncertainty Quantification for Traffic Forecasting Using Deep-Ensemble-Based Spatiotemporal Graph Neural Networks', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 9141–9152, Aug. 2024, doi: 10.1109/TITS.2024.3381099.

[13] W. Qian, T. D. Nielsen, Y. Zhao, K. G. Larsen, and J. J. Yu, 'Uncertainty-Aware Temporal Graph Convolutional Network for Traffic Speed Forecasting', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 8578–8590, Aug. 2024, doi: 10.1109/TITS.2024.3365721.

[14] Q. Wang, S. Wang, D. Zhuang, H. Koutsopoulos, and J. Zhao, 'Uncertainty Quantification of Spatiotemporal Travel Demand With Probabilistic Graph Neural Networks', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 8770–8781, Aug. 2024, doi: 10.1109/TITS.2024.3367779.

[15] C. Li, L. Bai, W. Liu, L. Yao, and S. T. Waller, 'Graph Neural Network for Robust Public Transit Demand Prediction', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4086–4098, May 2022, doi: 10.1109/TITS.2020.3041234.

[16] Y. Li, S. Chai, G. Wang, X. Zhang, and J. Qiu, 'Quantifying the Uncertainty in Long-Term Traffic Prediction Based on PI-ConvLSTM Network', *IEEE Trans. Intell. Transport. Syst.*, vol. 23, no. 11, pp. 20429–20441, Nov. 2022, doi: 10.1109/TITS.2022.3193184.

[17] Y. Wu, Y. Ye, A. Zeb, J. J. Yu, and Z. Wang, 'Adaptive Modeling of Uncertainties for Traffic Forecasting', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 4427–4442, May 2024, doi: 10.1109/TITS.2023.3327100.

[18] G. Shafer and V. Vovk, 'A Tutorial on Conformal Prediction', *Journal of Machine Learning Research*, 2008.

[19] I. Laña, I. Olabarrieta, and J. D. Ser, 'Measuring the Confidence of Single-Point Traffic Forecasting Models: Techniques, Experimental Comparison, and Guidelines Toward Their Actionability', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 11180–11199, Sep. 2024, doi: 10.1109/TITS.2024.3375936.

[20] P. Dong, X.-L. Wang, I. Bose, K. K. H. Ng, X. Zhang, and X. Zhang, 'Causally-Aware Spatio-Temporal Multi-Graph Convolution Network for Accurate and Reliable Traffic Prediction', arXiv.org. Accessed: Oct. 23, 2024. [Online]. Available: https://arxiv.org/abs/2408.13293v1

[21] K. Schweighofer, L. Aichberger, M. Ielanskyi, G. Klambauer, and S. Hochreiter, 'Quantification of Uncertainty with Adversarial Models', in *Neural Information Processing Systems*, arXiv, 2023. doi: 10.48550/ARXIV.2307.03217.

[22] B. Lakshminarayanan, A. Pritzel, and C. Blundell, 'Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles', Nov. 03, 2017, *arXiv:* arXiv:1612.01474. Accessed: Oct. 22, 2022. [Online]. Available: http://arxiv.org/abs/1612.01474

[23] F. Wenzel, J. Snoek, D. Tran, and R. Jenatton, 'Hyperparameter Ensembles for Robustness and Uncertainty Quantification', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 6514–6527. Accessed: Nov. 05, 2023. [Online]. Available: https://papers.nips.cc/paper/2020/hash/481fbfa59da2581098e841b7afc122f1-Abstract.html

[24] C. Louizos and M. Welling, 'Multiplicative normalizing flows for variational Bayesian neural

networks', in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, in ICML'17. Sydney, NSW, Australia: JMLR.org, Aug. 2017, pp. 2218–2227.

[25] T. Salimans, D. P. Kingma, and M. Welling, 'Markov Chain Monte Carlo and variational inference: bridging the gap', in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, in ICML'15. Lille, France: JMLR.org, Jul. 2015, pp. 1218–1226.

[26] A. Kendall and Y. Gal, 'What uncertainties do we need in Bayesian deep learning for computer vision?', in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 5580–5590.

[27] X. Jiang, D. Zhuang, X. Zhang, H. Chen, J. Luo, and X. Gao, 'Uncertainty Quantification via Spatial-Temporal Tweedie Model for Zero-inflated and Long-tail Travel Demand Prediction', in *arXiv.org*, Jun. 2023. Accessed: Oct. 11, 2023. [Online]. Available: https://arxiv.org/abs/2306.09882v1

[28] T. Pearce, A. Brintrup, M. Zaki, and A. Neely, 'High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach', in *Proceedings of the 35th International Conference on Machine Learning*, PMLR, Jul. 2018, pp. 4075–4084. Accessed: Jan. 10, 2023. [Online]. Available: https://proceedings.mlr.press/v80/pearce18a.html

[29] Y. Jiang, J. Fan, Y. Liu, and X. Zhang, 'Deep Graph Gaussian Processes for Short-Term Traffic Flow Forecasting From Spatiotemporal Data', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 20177–20186, Nov. 2022, doi: 10.1109/TITS.2022.3178136.

[30] H. Wen *et al.*, 'DiffSTG: Probabilistic Spatio-Temporal Graph Forecasting with Denoising Diffusion Models', Jun. 27, 2023, *arXiv*: arXiv:2301.13629. doi: 10.48550/arXiv.2301.13629.

[31] L. Lin, D. Shi, A. Han, and J. Gao, 'SpecSTG: A Fast Spectral Diffusion Framework for Probabilistic Spatio-Temporal Traffic Forecasting', Aug. 06, 2024, *arXiv*: arXiv:2401.08119. Accessed: Oct. 23, 2024. [Online]. Available: http://arxiv.org/abs/2401.08119

[32] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, 'Distribution-Free Predictive Inference for Regression', *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1094–1111, Jul. 2018, doi: 10.1080/01621459.2017.1307116.

[33] I. Gibbs and E. Candes, 'Adaptive Conformal Inference Under Distribution Shift', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 1660–1672. Accessed: Oct. 17, 2024. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html

[34] Z. Lin, S. Trivedi, and J. Sun, 'Conformal Prediction with Temporal Quantile Adjustments', May 23, 2022, *arXiv*: arXiv:2205.09940. Accessed: Oct. 30, 2024. [Online]. Available: http://arxiv.org/abs/2205.09940

[35] M. Zaffran, O. Feron, Y. Goude, J. Josse, and A. Dieuleveut, 'Adaptive Conformal Predictions for Time Series', in *Proceedings of the 39th International Conference on Machine Learning*, PMLR, Jun. 2022, pp. 25834–25866. Accessed: Oct. 27, 2024. [Online]. Available: https://proceedings.mlr.press/v162/zaffran22a.html

[36] I. Gibbs and E. J. Candès, 'Conformal Inference for Online Prediction with Arbitrary Distribution Shifts', *Journal of Machine Learning Research*, vol. 25, no. 162, pp. 1–36, 2024.

[37] E. Hazan, 'Introduction to Online Convex Optimization', Aug. 06, 2023, *arXiv*: arXiv:1909.05207. doi: 10.48550/arXiv.1909.05207.

[38] A. Bhatnagar, H. Wang, C. Xiong, and Y. Bai, 'Improved Online Conformal Prediction via Strongly Adaptive Online Learning', presented at the the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA, 2023, pp. 2337–2363.

[39] Z. Zhang, Z. Lu, and H. Yang, 'The Benefit of Being Bayesian in Online Conformal Prediction', Oct. 03, 2024, *arXiv*: arXiv:2410.02561. Accessed: Oct. 30, 2024. [Online]. Available: http://arxiv.org/abs/2410.02561

[40] Z. Zhang, D. Bombara, and H. Yang, 'Discounted Adaptive Online Learning: Towards Better Regularization', presented at the Forty-first International Conference on Machine Learning, Jun. 2024. Accessed: Nov. 29, 2024. [Online]. Available: https://openreview.net/forum?id=ZoTIdyExx6

[41] C. Xu and Y. Xie, 'Conformal prediction interval for dynamic time-series', in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Jul. 2021, pp. 11559–11569. Accessed: Oct. 24, 2024. [Online]. Available: https://proceedings.mlr.press/v139/xu21h.html

[42] C. Xu and Y. Xie, 'Conformal Prediction for Time Series', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11575–11587, Oct. 2023, doi: 10.1109/TPAMI.2023.3272339.

[43] J. Jonkers, G. V. Wallendael, L. Duchateau, and S. V. Hoecke, 'Conformal Predictive Systems Under Covariate Shift', Sep. 16, 2024, *arXiv*: arXiv:2404.15018. Accessed: Oct. 31, 2024. [Online]. Available: http://arxiv.org/abs/2404.15018

[44] J. Lee, C. Xu, and Y. Xie, 'Kernel-based optimally weighted conformal prediction intervals', May 27, 2024, *arXiv*: arXiv:2405.16828. Accessed: Oct. 27, 2024. [Online]. Available: http://arxiv.org/abs/2405.16828

[45] J. Lei and L. Wasserman, 'Distribution-free Prediction Bands for Non-parametric Regression', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 76, no. 1, pp. 71–96, Jan. 2014, doi: 10.1111/rssb.12021.

[46] S. Kiyani, G. Pappas, and H. Hassani, 'Conformal Prediction with Learned Features', Apr. 26, 2024, *arXiv*: arXiv:2404.17487. Accessed: Oct. 31, 2024. [Online]. Available: http://arxiv.org/abs/2404.17487

[47] Anonymous Authors, 'Enhanced Conformal Prediction for Deep Learning Based Time Series Forecasting', presented at the Forty-second International Conference on Machine Learning, Jan. 2025. Accessed: Feb. 15, 2025. [Online]. Available: https://openreview.net/forum?id=tiBln4U0dB&referrer=%5BAuthor%20Console%5D(%2Fgroup%3Fid%3DICML.cc%2F2025%2FConference%2FAuthors%23your-submissions)

[48] V. Chernozhukov, K. Wüthrich, and Y. Zhu, 'Distributional conformal prediction', *Proceedings of the National Academy of Sciences*, vol. 118, no. 48, p. e2107794118, Nov. 2021, doi: 10.1073/pnas.2107794118.

[49] M. Sesia and Y. Romano, 'Conformal Prediction using Conditional Histograms', presented at the Advances in Neural Information Processing Systems, Nov. 2021. Accessed: Nov. 27, 2024. [Online]. Available: https://openreview.net/forum?id=EvhsTX6GMyM

[50] C. Xu and Y. Xie, 'Sequential Predictive Conformal Inference for Time Series', arXiv, 2022. doi: 10.48550/ARXIV.2212.03463.

[51] Y. Romano, E. Patterson, and E. Candes, 'Conformalized Quantile Regression', in *Advances*

in Neural Information Processing Systems, Curran Associates, Inc., 2019. Accessed: Nov. 25, 2024. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html

[52] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', Jan. 30, 2017, *arXiv*: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.

[53] Y. Nesterov, 'A method for solving the convex programming problem with convergence rate O(1/k^2)', *Proceedings of the USSR Academy of Sciences*, 1983, Accessed: Dec. 17, 2024. [Online]. Available: https://www.semanticscholar.org/paper/A-method-for-solving-the-convex-programming-problem-Nesterov/8d3a318b62d2e970122da35b2a2e70a5d12cc16f

[54] J. Duchi, E. Hazan, and Y. Singer, 'Adaptive Subgradient Methods for Online Learning and Stochastic Optimization'.

[55] J. Wang, J. Jiang, W. Jiang, C. Han, and W. X. Zhao, 'Towards Efficient and Comprehensive Urban Spatial-Temporal Prediction: A Unified Library and Performance Benchmark', arXiv.org. Accessed: May 06, 2023. [Online]. Available: https://arxiv.org/abs/2304.14343v2

[56] B. Yu, H. Yin, and Z. Zhu, 'Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting', in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden: International Joint Conferences on Artificial Intelligence Organization, Jul. 2018, pp. 3634–3640. doi: 10.24963/ijcai.2018/505.

[57] Y. Li, R. Yu, C. Shahabi, and Y. Liu, 'Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting', *arXiv: Learning*, Jul. 2017, Accessed: Jun. 26, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Diffusion-Convolutional-Recurrent-Neural-Network%3A-Li-Yu/9ba0186ed40656329c421f55ada7313293e13f17

[58] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, 'Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks', May 24, 2020, *arXiv*: arXiv:2005.11650. Accessed: May 06, 2023. [Online]. Available: http://arxiv.org/abs/2005.11650

[59] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, 'Graph WaveNet for Deep Spatial-Temporal Graph Modeling', in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao, China: International Joint Conferences on Artificial Intelligence Organization, Aug. 2019, pp. 1907–1913. doi: 10.24963/ijcai.2019/264.

[60] Y. Gal and Z. Ghahramani, 'Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning', presented at the International Conference on Machine Learning, Jun. 2015. Accessed: Nov. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Dropout-as-a-Bayesian-Approximation:-Representing-Gal-Ghahramani/f35de4f9b1a7c4d3fa96a0d2ab1bf8937671f6b6

[61] D. Wu *et al.*, 'Quantifying Uncertainty in Deep Spatiotemporal Forecasting', in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event Singapore: ACM, Aug. 2021, pp. 1841–1851. doi: 10.1145/3447548.3467325.

[62] R. Vershynin, 'High-Dimensional Probability'.

Appendix

A. Implementation details

A1. Base prediction models

STGCN [56] (Spatial Temporal Graph Convolutional Network) consists of multiple spatial-temporal convolution blocks which integrate graph convolutions to extract spatial features and gated temporal convolutions to capture temporal dynamics, allowing it to effectively process spatial temporal data.

DCRNN [57] (Diffusion Convolutional Recurrent Neural Network) models traffic flow as a diffusion process over a directed graph and captures spatial and temporal dependencies through diffusion convolutions on graphs and an encoder-decoder architecture. This model leverages bidirectional random walks on graphs to capture spatial correlations and uses recurrent neural networks like GRUs to model temporal sequences.

MTGNN [58] (Multivariate Time Series Forecasting with Graph Neural Networks) automatically extracts relationships between regions (or grids) via a graph learning module. With its mix-hop propagation layers and dilated inception layers, MTGNN captures complex spatial and temporal dependencies while addressing challenges like unknown graph structures and joint optimization of graph structure and network parameters.

GWNET [59] (Graph WaveNet) extends the concept of Wavenet to graph-structured data by incorporating adaptive adjacency matrices learned during training, thus overcoming limitations posed by predefined graph structures. And it employs dilated causal convolutions along with graph convolutions to efficiently capture long-range dependencies.

The implementation of these 4 models is based on [55], and the hyper-parameters, such as hidden dimension of DCRNN, number of layers in STGCN, are the default value in [55]. For all models, we transfer it from point prediction version to quantiles prediction version by adding an additional prediction head and training it with quantile loss. Before training starts, the dataset is normalized using z-score standardization. These models are trained with Adam for 100 epochs with learning rate 0.005. And the training will be stopped if the loss in validation set dose not decrease for 10 epochs continuously.

A2. Details of baseline methods

Bootstrap: It is a technique used to estimate statistics on a population by sampling a dataset

with replacement. In our experiments, we train 20 models with randomly selected training samples and the variance of outputs of difference models (σ^2) is considered as the variance of prediction. Then if the mean of predictions of all models is y, the confidence interval is $[y - 1.645\sigma, y + 1.645\sigma]$.

MC Dropout. It is a Bayesian approximation technique that leverages dropout to estimate model uncertainty. In our experiments, we set dropout rate as 0.3, the same as [17]. And during deployment, we use models to predict traffic value 20 times and obtain the variance (σ^2) and means of these predictions (y). Then the confidence interval is $[y - 1.645\sigma, y + 1.645\sigma]$.

Directly Minimizing Mean Interval Score (MIS) [61] is training models with the following loss:

$$MIS = \frac{1}{2nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{2}(up_{tij} - low_{tij}) + \frac{2}{\alpha}(low_{tij} - y_{t,i,j})I(y_{t,i,j} < low_{tij}) + \frac{2}{\alpha}(y_{t,i,j} - up_{t,i,j})I(y_{t,i,j} > up_{tij})$$

DESQRUQ[12]: The main idea of it is training multiple quantile prediction models with different hyper parameters and ensemble these results. Bayesian optimization and Gaussian copula are used to find better hype-parameter. The search space of hype-parameters in our experiments is: learning rate [0.0001, 0.02], batch size [8, 16, . . . , 256], number of layers for the encoder [1, 2, 3, 4, 5], number of training epoch [20,21,. . . ,100].

ProbGNN[14]: This method regards the traffic demand as a distribution to consider data uncertainty and use ensembles to account for model uncertainty. In our experiments, Gaussian distribution is considered as data distribution, the same as [14]. Besides, we train 20 models with different initialization and five top models by validation set loss are selected to create an ensemble model, which is also the same as [14].

UATGCN [13]: This method uses Monte Carlo dropout and predictive variances to estimate model and data uncertainty. We set dropout rate as 0.2, the same as [13].

QUANTARFFIC[17]: Training a quantile repression model and using validation set to adjusts the quantile prediction of each region independently. The settings in our experiments are the same in original paper.

B. Proofs

B.1 proofs of Theorem 4.1

To proof the theorem of average coverage, we need some lemmas first.

Lemma1: If errors are bound by B and $Q_1(E_{i,t}) = B$, then, for any region $i, and\ any\ time\ t$:

$$-\frac{\gamma_1}{\alpha} \leq \alpha_{i,t} \leq 1 + \frac{\gamma_1}{\alpha}$$

Proof:

1 for $t = 1$ $\alpha_{i1} = 1 - \alpha$, this is in $[-\frac{\gamma_1}{\alpha}, 1 + \frac{\gamma_1}{\alpha}]$

2 suppose for t

$$-\frac{\gamma_1}{\alpha} \leq \alpha_{i,t} \leq 1 + \frac{\gamma_1}{\alpha}$$

Then we proof for $-\frac{\gamma_1}{\alpha} \leq \alpha_{i,(t+1)} \leq 1 + \frac{\gamma_1}{\alpha}$

1)If $-\frac{\gamma_1}{\alpha} \leq \alpha_{it} \leq 0$, then $C_{it} = [0, B]$ and $P(y_{it} \in C_{it}) = 1$ then

$$\alpha_{i,(t+1)} = \gamma_{it}\left(\alpha - \frac{1}{2}\sum_{j=1}^{2} I\left(y_{t,i,j} \notin [low_{i,t,j}, up_{i,t,j}]\right)\right) + \alpha_{i,t} = \alpha\gamma_{i,t} + \alpha_{i,t}$$

Because $v_{i,t+1} = \beta v_{i,t} + (1-\beta)g_t^2$ and $g_t^2 = \left(\alpha - \frac{1}{2}\sum_{j=1}^{2} I\left(y_{tij} \notin [low_{itj}, up_{itj}]\right)\right)^2 > \alpha^2$ (when $\alpha \leq 0.25$). Note that if $\alpha > 0.25$, g_t^2 can also be bounded and the result of our main theorem will not change.

So, for each t $v_{i,t} \geq \alpha^2$. Then

$$\gamma_{i,t} = \frac{\gamma_1}{\sqrt{v_{i,t}} + \epsilon} \leq \frac{\gamma_1}{\alpha}$$

So $\alpha\gamma_{it} + \alpha_{ir}$ is in $\left[-\frac{\gamma_1}{\alpha}, 1 + \frac{\gamma_1}{\alpha}\right]$,

2) if $1 \leq \alpha_{it} \leq 1 + \frac{\gamma_1}{\alpha}$, then $C_{it} = \emptyset$ and $P(y_{it} \in C_{it}) = 0$ then

$$\alpha_{i,(t+1)} = \gamma_{it}\left(\alpha - \frac{1}{2}\sum_{j=1}^{2} I\left(y_{t,ij} \notin [low_{t,i,j}, up_{t,i,j}]\right)\right) + \alpha_{i,t} = (\alpha - 1)\gamma_{i,t} + \alpha_{i,t}$$

is in $[-\frac{\gamma_1}{\alpha}, 1 + \frac{\gamma_1}{\alpha}]$

3) if $0 < \alpha_{i,t} < 1$

$$\alpha_{i,(t+1)} = \gamma_{i,t}\left(\alpha - \frac{1}{2}\sum_{j=1}^{2} I\left(y_{t,i,j} \notin \left[llow_{t,i,j}, up_{t,i,j}\right]\right)\right) + \alpha_{i,t}$$

because $\gamma_{it}\left(\alpha - \frac{1}{2}\sum_{j=1}^{2} I\left(y_{tij} \notin \left[low_{t,i,j}, up_{t,i,j}\right]\right)\right)$ is in $[-\frac{\gamma_1}{\alpha}, \frac{\gamma_1}{\alpha}]$, $-\frac{\gamma_1}{\alpha} \le \alpha_{i,(t+1)} \le 1 + \frac{\gamma_1}{\alpha}$

combing these three cases, we have: $-\gamma_1 \le \alpha_{i,(t+1)} \le 1 + \gamma_1$

Then for any t,

$$-\frac{\gamma_1}{\alpha} \le \alpha_{i,t} \le 1 + \frac{\gamma_1}{\alpha}$$

Lemma2: for any region i

$$\left|\frac{1}{2T}\sum_{t=1}^{T}\sum_{j=1}^{2} I\left(y_{t,i,j} \in \left[low_{t,i,j}, up_{t,i,j}\right]\right) - (1-\alpha)\right| \le (2-\beta)\frac{1+2\gamma_1}{T\alpha}$$

Proof:

Define Δ_i as $\Delta_{i1} = \frac{1}{\gamma_{i1}}, \Delta_{i2} = \frac{1}{\gamma_{i2}} - \frac{1}{\gamma_{i1}}, \Delta_{i3} = \frac{1}{\gamma_{i3}} - \frac{1}{\gamma_{i2}}, \ldots$ and $\left|\left|\Delta_i\right|\right|_1$ is the $l1$ norm of Δ_i, which is defined as $\sum_j |\Delta_{ij}|$

$$\left|\frac{1}{2T}\sum_{t=1}^{T}\sum_{j=1}^{2} I\left(y_{t,i,j} \in \left[low_{t,i,j}, up_{t,i,j}\right]\right) - (1-\alpha)\right| = \left|\frac{1}{2T}\sum_{t=1}^{T}\sum_{j=1}^{2} I\left(y_{t,i,j} \notin \left[low_{t,i,j}, up_{t,i,j}\right]\right) - \alpha\right|$$

$$= \left|\frac{1}{2T}\sum_{t=1}^{T}\left(\sum_{r=1}^{t}\Delta_{ir}\right)\gamma_{it}\sum_{j=1}^{2}\left(y_{t,i,j} \notin \left[low_{t,i,j}, up_{t,i,j}\right]\right) - \alpha\right|$$

$$= \left|\frac{1}{2T}\sum_{t=1}^{T}\Delta_{ir}\sum_{t=r}^{T}\sum_{j=1}^{2}\gamma_{i,t} I\left(y_{t,i,j} \notin \left[low_{t,i,j}, up_{t,i,j}\right]\right) - \alpha\right|$$

$$= \left|\frac{1}{T}\sum_{t=1}^{T}\Delta_{ir}\left(\sum_{t=r}^{T}\frac{1}{2}\sum_{j=1}^{2}\gamma_{it} I\left(y_{t,i,j} \notin \left[low_{t,i,j}, up_{t,i,j}\right]\right) - \alpha\right)\right|$$

$$= \left|\frac{1}{T}\sum_{t=1}^{T}\Delta_{ir}(\alpha_{iT} - \alpha_{ir})\right| \quad (1)$$

$$\le \frac{1}{T}\max_r |\alpha_{iT} - \alpha_{ir}| \sum_{t=1}^{T} |\Delta_{ir}|$$

(1) Is because $\alpha_{i,(r+1)} = \gamma_{ir}\left(\alpha - \frac{1}{2}\sum_{j=1}^{2} I\left(y_{tij} \notin \left[low_{t,i,j}, up_{t,i,j}\right]\right)\right) + \alpha_{ir}$

According to lemma 1, we have $-\gamma_1 \le \alpha_{ir} \le 1 + \gamma_1$ thus $\max_r |\alpha_{iT} - \alpha_{ir}| \le 1 + 2\gamma_1$

$$\frac{1}{2T}\max_r |\alpha_{iT} - \alpha_{ir}| \sum_{t=1}^{T} |\Delta_{ir}| \le \frac{1 + 2\gamma_1}{T\alpha} \left|\left|\Delta_{\mathrm{i}}\right|\right|_1 (2)$$

Then we bound $\left|\left|\Delta_{\mathrm{i}}\right|\right|_1$

Because:

$$\gamma_{it} = \frac{\gamma_1}{\sqrt{v_{i,t}} + \epsilon}$$

Then

$$\frac{1}{\gamma_{i,t+1}} - \frac{1}{\gamma_{i,t}} = \sqrt{v_{i,t}} - \sqrt{v_{i,t+1}}$$

Therefore:

$$\left|\left|\Delta_{\mathrm{i}}\right|\right|_1 = \sqrt{v_{i,T}} - \sqrt{v_{i,1}}$$

Because:

$v_{i,t+1} = \beta v_{i,t} + (1-\beta) g_t^2$, so, for any t

$$v_{i,t+1} = \sum_{k=1}^{t} (1-\beta)\beta^k g_{t-k}^2 + (1-\beta) g_t^2 \le (1-\beta)\sum_{k=1}^{\infty} \beta^k + (1-\beta) \le 2 - \beta$$

Therefore:

$$\left|\left|\Delta_{\mathrm{i}}\right|\right|_1 = \sqrt{v_{i,T}} - \sqrt{v_{i,1}} \le \sqrt{v_{i,T}} \le 2 - \beta (3)$$

Combing (2) and (3) can get this lemma.

Theorem 4.1:

$$\left| \frac{1}{2T} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{2} I\left(y_{tij} \in \left[low_{itj}, up_{itj}\right]\right) - (1-\alpha) \right| \le \frac{c}{T}$$

Proof:

Using lemma2 for each i and combining them together can easily proof this theorem.

Where $c = (2-\beta)\frac{1+2\gamma_1}{\alpha}$

B.2 proofs of Theorem 4.2

We need some lemmas first.

Lemma3 Hoeffding inequality[62].

If $x_1, x_2, \dots, x_n$ are independent random variables and $a_i \le x_i \le b_i$, then for any $\epsilon > 0$:

$$P\left(\left|\sum_i x_i - E\sum_i x_i\right| \ge \epsilon\right) \le 2\exp\left(-\frac{2\epsilon^2}{\sum(a_i - b_i)^2}\right)$$

Lemma 4 propositions of sub-guassion random variable [62]:

For a random variable x with zero mean, the following three inequalities are equal:

for any $t > 0$

$$P(|x| \ge t) \le 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$$

For any $\lambda \in R$

$$Eexp(\lambda x) \le \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

For any $p > 1$:

$$\left|\left|x\right|\right|_{L^p} = (E|x|^p)^{\frac{1}{p}} \le K\sigma$$

Where K is a constant

Lemma 5

For random variables $x_1, x_2, \dots, x_n$ if we have:

$$P(|x_i| \ge t) \le 2\exp\left(-\frac{t^2}{2\sigma_i^2}\right)$$

Then:

$$P\left(\left|\frac{1}{n}\sum_i x_i\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{2\max_i \sigma_i^2}\right)$$

Proof:

$$||\frac{1}{n}\sum_i x_i||_{L^p} \le \sum\frac{1}{n}|x_i|_{L^p} = \sum\frac{1}{n}K\sigma_i \le K\max_i \sigma_i$$

Using lemma 4, this theorem can be proofed.

Theorem 4.2:

if the error in each region is independent, and the data in one time step only depends on the data in the preceding K time slices, then:

$$\min_i \frac{1}{2T}\sum_{t}^{T}\sum_{j=1}^{2} I\left(y_{t,i,j} \in \left[low_{t,i,j}, up_{t,i,j}\right]\right) \geq 1-\alpha-\frac{c}{T}-\sqrt{\frac{Klogn}{2T}}$$

Proof:

Define M_{it} as $\frac{1}{2}\sum_{j=1}^{2}\left(I\left(y_{t,i,j} \in \left[low_{t,i,j}, up_{t,i,j}\right]\right)-(1-\alpha)\right)$,each $M_{i,t}$ is bounded in [0,1].

Define $S_k = \{i | i = nK + k, n = 1,2 \ldots and\ i \leq T\}$. Then for a given k each element in $\{M_{it} | t \in S_k\}$ is independent, then according to lemma3, we have:

$$P\left(\left|\sum_{t\in S_k} M_{it} - E\sum_{t\in S_k} M_{it}\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2\epsilon^2}{|S_k|}\right)$$

Where $|S_k|$ is the number of elements in S_k. Without loss of generality, assume T is an integral multiple of K, then, $|S_k| = \frac{T}{K}$

Therefore:

$$P\left(\frac{1}{T}\left|\sum_{t<T} M_{it} - E\sum_{t<T} M_{it}\right| \geq \epsilon\right)$$

$$= P\left(\frac{1}{K}\sum_{k}\frac{1}{|S_k|}\left|\sum_{t\in S_k} M_{it} - E\sum_{t\in S_k} M_{it}\right| \geq \epsilon\right)\ by\ lemma\ 5$$

$$\leq 2\exp\left(-\frac{2\epsilon^2 T}{K}\right)$$

In the worst-case $E\frac{1}{T}\sum_{t<T} M_{it} = 1-\alpha-\frac{c}{T}$ for every i. Therefore:

$$P\left(\left(1-\alpha-\frac{c}{T}-\frac{1}{T}\sum_{t<T} M_{it}\right) \geq \epsilon\right) \leq \exp\left(-\frac{2\epsilon^2 T}{K}\right)$$

Then:

$$E\left(1-\alpha-\frac{c}{T}-\min_i \frac{1}{T}\sum_{t<T} M_{it}\right) = E\left(\max_i\left(1-\alpha-\frac{c}{T}-\frac{1}{T}\sum_{t<T} M_{it}\right)\right)$$

To be clearer, let u_i be $1-\alpha-\frac{c}{T}-\frac{1}{T}\sum_{t<T} M_{it}$ and

$$P(u_i \geq \epsilon) \leq \exp\left(-\frac{2\epsilon^2 T}{K}\right)$$

Then:

$$\exp\left(\lambda E\left(\max_i u_i\right)\right) \le E\exp\left(\lambda \max_i u_i\right) \le E\sum_i \exp(\lambda u_i) \le n\exp\left(\frac{\lambda^2 K}{8T}\right) = \exp\left(\log n + \frac{K\lambda^2}{8T}\right)$$

The first $\le$ is because of Jensen's inequality, the third inequity is because of lemma 4

Therefore, if we let $\lambda = \sqrt{\frac{8T\log}{K}}$

$$E\left(\max_i u_i\right) \le \frac{\log n}{\lambda} + \frac{K\lambda}{8T} \le \sqrt{\frac{K}{2T}\log n}$$

Then recall the definition of u_i and $M_{i,t}$, we can finish this proof.

C. Full result

C1 Full result of main experiments

We report the results of all prediction models in the following 4 tables, each one represents the results for one dataset.

Table A1 full results of NYCbike dataset

Time	Base model	metric	QR	MCD	boostrop	MIS	DESQRUQ	UATGCN	ProbGNN	QuanTraffic	CP	ACI	QCP	DtACI	CONTINA
1 month	STGCN	cov	88.0%	55.4%	32.4%	87.8%	91.8%	90.8%	91.9%	91.1%	89.1%	89.8%	89.5%	89.9%	89.6%
		length	0.241	0.177	0.089	0.239	0.270	0.269	0.327	0.278	0.284	0.284	0.244	0.299	0.257
		minRC	70.2%	36.6%	25.8%	81.5%	86.3%	81.7%	89.2%	86.5%	81.7%	88.9%	85.2%	88.8%	88.5%
	DCRNN	cov	89.8%	48.1%	36.3%	88.4%	91.7%	91.4%	92.6%	91.2%	89.3%	90.0%	89.8%	89.5%	89.6%
		length	0.264	0.203	0.100	0.310	0.286	0.286	0.268	0.289	0.285	0.295	0.264	0.286	0.272
		minRC	85.7%	30.8%	24.5%	74.4%	87.6%	87.6%	88.2%	86.4%	84.7%	88.8%	85.7%	86.5%	88.7%
	MTGNN	cov	90.5%	55.3%	26.4%	88.8%	91.9%	92.3%	93.1%	91.8%	89.1%	89.8%	90.8%	89.6%	89.8%
		length	0.275	0.229	0.067	0.278	0.290	0.298	0.302	0.289	0.283	0.306	0.276	0.287	0.286
		minRC	82.6%	33.5%	14.7%	81.2%	85.7%	85.1%	85.9%	85.8%	85.9%	88.9%	83.3%	87.3%	88.9%
	GWNET	cov	90.0%	60.3%	28.2%	89.8%	91.6%	92.2%	94.9%	91.3%	89.2%	89.7%	90.0%	89.4%	89.6%
		length	0.280	0.262	0.078	0.284	0.289	0.284	0.329	0.294	0.290	0.315	0.280	0.293	0.287
		minRC	86.0%	44.8%	16.5%	83.4%	86.7%	85.8%	87.8%	87.0%	84.9%	88.8%	87.0%	86.5%	88.8%
	AVG	cov	89.6%	54.8%	30.8%	88.7%	91.7%	91.7%	93.1%	91.3%	89.2%	89.8%	90.0%	89.6%	89.6%
		length	0.265	0.218	0.084	0.278	0.284	0.284	0.306	0.288	0.285	0.300	0.266	0.291	0.276
		minRC	81.1%	36.4%	20.4%	80.1%	86.6%	85.0%	87.8%	86.4%	84.3%	88.8%	85.3%	87.3%	88.7%
2 month	STGCN	cov	88.0%	56.5%	30.1%	87.5%	91.5%	91.7%	91.4%	92.0%	89.4%	89.9%	89.3%	90.0%	89.8%
		length	0.246	0.185	0.094	0.240	0.276	0.272	0.335	0.262	0.289	0.289	0.246	0.304	0.256
		minRC	71.5%	39.7%	17.6%	81.1%	88.1%	83.6%	87.9%	86.9%	83.9%	88.7%	83.4%	88.8%	89.2%
	DCRNN	cov	90.3%	49.2%	36.8%	88.9%	92.3%	91.8%	93.1%	91.7%	89.3%	90.4%	90.3%	90.0%	89.8%
		length	0.270	0.211	0.104	0.319	0.291	0.298	0.275	0.285	0.287	0.302	0.270	0.294	0.272
		minRC	83.5%	30.6%	24.6%	74.5%	88.8%	89.0%	89.5%	85.9%	85.2%	89.0%	83.8%	84.8%	89.3%
	MTGNN	cov	91.2%	56.7%	27.0%	89.0%	91.4%	92.9%	93.8%	92.4%	89.3%	90.0%	91.5%	89.9%	89.8%

		length	0.278	0.240	0.070	0.282	0.291	0.306	0.312	0.292	0.286	0.308	0.279	0.293	0.285
		minRC	83.1%	35.0%	14.7%	80.9%	86.9%	86.4%	85.8%	85.3%	85.1%	89.0%	83.6%	84.9%	89.3%
		cov	90.6%	60.8%	28.9%	90.5%	92.2%	92.7%	95.3%	91.9%	89.3%	90.0%	90.6%	89.9%	89.8%
	GWNET	length	0.288	0.270	0.082	0.292	0.275	0.289	0.339	0.302	0.293	0.316	0.287	0.300	0.288
		minRC	85.5%	45.1%	17.4%	81.7%	88.1%	87.6%	88.5%	87.6%	85.1%	89.1%	85.6%	85.3%	89.2%
		cov	90.0%	55.8%	30.7%	89.0%	91.9%	92.3%	93.4%	92.0%	89.3%	90.1%	90.5%	90.0%	89.8%
	AVG	length	0.270	0.227	0.087	0.283	0.283	0.291	0.315	0.285	0.289	0.304	0.271	0.298	0.275
		minRC	80.9%	37.6%	18.6%	79.6%	88.0%	86.6%	87.9%	86.4%	84.8%	88.9%	84.1%	85.9%	89.3%
		cov	87.0%	56.5%	29.7%	86.6%	90.7%	91.2%	91.1%	91.4%	88.0%	89.1%	88.7%	89.6%	89.8%
	STGCN	length	0.263	0.206	0.101	0.258	0.299	0.292	0.360	0.286	0.297	0.297	0.266	0.329	0.280
		minRC	74.0%	39.0%	14.8%	82.1%	84.6%	83.8%	88.0%	85.1%	83.8%	88.5%	81.5%	88.6%	89.0%
		cov	89.7%	49.9%	36.7%	89.0%	91.6%	91.2%	92.5%	91.0%	87.4%	89.7%	89.8%	89.1%	89.8%
	DCRNN	length	0.288	0.234	0.116	0.341	0.279	0.324	0.298	0.303	0.292	0.322	0.288	0.310	0.293
		minRC	81.1%	32.4%	24.1%	77.7%	83.4%	85.1%	85.6%	83.2%	81.7%	88.1%	81.1%	86.3%	89.2%
		cov	90.6%	57.4%	26.8%	88.8%	90.9%	92.4%	93.4%	91.9%	87.6%	89.9%	90.9%	89.1%	89.8%
3 month	MTGNN	length	0.300	0.269	0.073	0.295	0.311	0.329	0.341	0.314	0.291	0.332	0.301	0.329	0.309
		minRC	82.8%	35.2%	15.8%	80.9%	85.8%	85.3%	85.4%	85.1%	81.0%	88.7%	83.1%	85.9%	89.3%
		cov	90.0%	61.2%	29.2%	89.8%	91.6%	92.6%	95.0%	91.3%	87.6%	90.0%	90.0%	89.1%	89.8%
	GWNET	length	0.307	0.292	0.086	0.312	0.294	0.315	0.361	0.321	0.299	0.338	0.307	0.316	0.311
		minRC	84.0%	46.0%	17.6%	82.8%	84.2%	85.8%	86.9%	86.0%	82.8%	89.0%	83.9%	86.5%	89.2%
		cov	89.3%	56.2%	30.6%	88.6%	91.2%	91.9%	93.0%	91.4%	87.6%	89.7%	89.9%	89.2%	89.8%
	AVG	length	0.289	0.250	0.094	0.301	0.296	0.315	0.340	0.306	0.295	0.322	0.290	0.321	0.298
		minRC	80.5%	38.1%	18.1%	80.9%	84.5%	85.0%	86.5%	84.8%	82.3%	88.6%	82.4%	86.8%	89.2%
		cov	85.1%	57.3%	28.3%	85.0%	89.2%	89.7%	88.2%	89.7%	84.8%	89.2%	87.3%	90.3%	89.7%
	STGCN	length	0.313	0.279	0.105	0.311	0.369	0.350	0.445	0.339	0.317	0.377	0.322	0.428	0.345
		minRC	69.9%	31.0%	19.4%	75.2%	77.9%	74.9%	80.0%	78.5%	74.8%	88.5%	73.5%	88.3%	89.1%
		cov	88.5%	51.7%	37.1%	88.7%	90.2%	89.6%	91.2%	89.7%	82.3%	90.6%	88.7%	89.1%	89.8%
	DCRNN	length	0.339	0.300	0.163	0.400	0.330	0.427	0.383	0.354	0.308	0.416	0.340	0.388	0.355
		minRC	74.0%	31.2%	25.4%	65.6%	78.6%	80.2%	81.2%	76.4%	67.5%	89.5%	74.3%	88.6%	89.1%
		cov	88.7%	58.5%	27.2%	88.3%	89.8%	90.8%	91.9%	90.0%	82.5%	90.1%	89.0%	89.1%	89.7%
4 month	MTGNN	length	0.353	0.355	0.092	0.334	0.376	0.392	0.409	0.367	0.307	0.424	0.354	0.390	0.373
		minRC	74.6%	33.1%	17.2%	75.6%	83.0%	77.7%	79.3%	76.8%	65.7%	88.7%	75.0%	88.6%	88.9%
		cov	88.7%	60.8%	29.2%	88.6%	90.1%	91.2%	93.9%	89.9%	82.2%	90.1%	88.7%	88.8%	89.7%
	GWNET	length	0.365	0.360	0.107	0.374	0.351	0.375	0.438	0.380	0.315	0.429	0.365	0.393	0.377
		minRC	73.7%	38.4%	17.6%	74.7%	75.5%	80.7%	82.8%	75.4%	68.6%	88.4%	73.6%	88.2%	89.1%
		cov	87.7%	57.1%	30.5%	87.6%	89.8%	90.3%	91.3%	89.8%	82.9%	90.0%	88.4%	89.3%	89.7%
	AVG	length	0.343	0.323	0.117	0.355	0.356	0.386	0.419	0.360	0.312	0.411	0.345	0.400	0.363
		minRC	73.1%	33.4%	19.9%	72.8%	78.8%	78.4%	80.8%	76.8%	69.1%	88.8%	74.1%	88.4%	89.1%
1st			3	0	0	0	0	0	0	0	1	0	4	0	15

Table A2 Full results of NYCtaxi dataset

Time	Base model	metric	QR	MCD	boostrop	MIS	DESQRUQ	UATGCN	ProbGNN	QuanTraffic	CP	ACI	QCP	DtACI	CONTINA
1month	STGCN	cov	88.2%	60.0%	33.9%	88.0%	92.5%	90.2%	92.6%	90.7%	90.5%	90.0%	90.1%	90.1%	89.7%

		length	0.223	0.168	0.072	0.233	0.262	0.285	0.272	0.241	0.316	0.303	0.226	0.305	0.232
		minRC	79.2%	42.2%	20.7%	74.3%	81.4%	72.0%	80.1%	81.3%	50.3%	88.7%	80.9%	87.5%	88.9%
	DCRNN	cov	83.3%	60.7%	60.9%	89.5%	91.3%	86.9%	95.4%	78.8%	91.2%	89.9%	87.2%	90.2%	89.6%
		length	0.263	0.116	0.116	0.237	0.271	0.218	0.291	0.463	0.268	0.254	0.325	0.256	0.236
		minRC	75.1%	17.2%	47.5%	81.6%	81.3%	78.2%	89.7%	41.8%	88.1%	88.7%	84.0%	87.0%	89.0%
	MTGNN	cov	89.8%	70.3%	43.6%	89.1%	93.4%	90.8%	93.7%	81.0%	91.0%	89.9%	90.3%	90.2%	89.7%
		length	0.238	0.192	0.085	0.254	0.259	0.280	0.303	0.236	0.277	0.266	0.239	0.265	0.239
		minRC	78.7%	35.4%	29.0%	81.3%	87.2%	76.7%	78.4%	59.6%	87.8%	88.6%	79.9%	87.3%	89.1%
	GWNET	cov	90.2%	66.9%	51.3%	89.0%	92.5%	90.6%	95.5%	80.0%	91.2%	89.9%	90.3%	90.3%	89.7%
		length	0.242	0.116	0.113	0.247	0.257	0.272	0.336	0.237	0.272	0.257	0.242	0.259	0.243
		minRC	76.3%	32.4%	39.6%	74.4%	81.6%	73.5%	83.8%	56.5%	87.3%	88.3%	76.7%	86.7%	89.0%
	AVG	cov	87.9%	64.5%	47.4%	88.9%	92.4%	89.6%	94.3%	82.6%	91.0%	89.9%	89.5%	90.2%	89.7%
		length	0.241	0.148	0.096	0.243	0.262	0.264	0.300	0.294	0.283	0.270	0.258	0.271	0.237
		minRC	77.3%	31.8%	34.2%	77.9%	82.9%	75.1%	83.0%	59.8%	78.4%	88.6%	80.4%	87.1%	89.0%
	STGCN	cov	87.8%	59.6%	32.9%	87.7%	91.8%	89.5%	92.3%	90.2%	89.9%	90.0%	89.7%	90.1%	89.9%
		length	0.231	0.170	0.073	0.242	0.271	0.298	0.283	0.249	0.316	0.317	0.234	0.320	0.242
		minRC	77.7%	40.4%	22.2%	75.8%	80.1%	67.9%	78.3%	81.6%	49.2%	88.9%	80.6%	88.3%	89.3%
	DCRNN	cov	82.3%	61.9%	60.0%	89.1%	91.2%	86.3%	94.8%	78.0%	89.7%	89.9%	86.5%	89.5%	89.9%
		length	0.265	0.141	0.120	0.245	0.274	0.226	0.301	0.466	0.265	0.269	0.327	0.262	0.248
		minRC	74.2%	20.6%	45.7%	77.2%	10.0%	79.1%	87.6%	41.3%	86.2%	88.4%	84.0%	86.5%	89.2%
2month	MTGNN	cov	89.3%	70.7%	42.6%	89.1%	93.0%	90.2%	93.0%	80.9%	89.6%	89.8%	89.9%	89.6%	89.8%
		length	0.246	0.118	0.088	0.263	0.269	0.293	0.315	0.244	0.275	0.280	0.248	0.273	0.249
		minRC	76.3%	36.0%	29.7%	79.1%	84.7%	75.1%	75.6%	60.6%	85.3%	88.5%	77.4%	87.5%	89.2%
	GWNET	cov	89.9%	67.6%	50.6%	88.9%	92.2%	89.9%	95.1%	80.2%	89.6%	89.8%	90.0%	89.5%	89.9%
		length	0.251	0.135	0.117	0.257	0.266	0.283	0.350	0.246	0.269	0.272	0.251	0.265	0.252
		minRC	73.3%	30.6%	38.8%	73.7%	78.9%	72.5%	81.7%	55.5%	86.0%	88.7%	73.8%	87.2%	89.3%
	AVG	cov	87.3%	64.9%	46.5%	88.7%	92.0%	89.0%	93.8%	82.3%	89.7%	89.9%	89.0%	89.7%	89.9%
		length	0.248	0.141	0.099	0.252	0.270	0.275	0.312	0.301	0.281	0.284	0.265	0.280	0.248
		minRC	75.4%	31.9%	34.1%	76.5%	63.4%	73.6%	80.8%	59.7%	76.7%	88.6%	78.9%	87.4%	89.3%
	STGCN	cov	87.8%	59.7%	32.9%	87.7%	91.6%	89.5%	92.1%	90.2%	90.1%	90.1%	89.8%	90.1%	89.9%
		length	0.231	0.169	0.073	0.243	0.271	0.299	0.282	0.249	0.315	0.315	0.234	0.316	0.242
		minRC	68.3%	42.6%	21.3%	72.4%	70.8%	66.7%	70.8%	73.5%	51.2%	88.0%	75.3%	86.2%	89.2%
	DCRNN	cov	79.6%	62.3%	60.1%	89.0%	90.4%	85.7%	94.7%	75.9%	89.4%	90.2%	84.3%	89.8%	89.8%
		length	0.277	0.137	0.118	0.245	0.283	0.225	0.298	0.477	0.264	0.273	0.339	0.264	0.247
		minRC	71.0%	22.4%	45.9%	80.6%	74.2%	69.7%	81.0%	36.7%	83.1%	88.8%	40.1%	86.2%	89.1%
3month	MTGNN	cov	89.2%	71.0%	42.3%	89.3%	92.9%	90.0%	93.0%	80.8%	89.4%	90.1%	89.7%	89.8%	89.9%
		length	0.246	0.113	0.087	0.263	0.268	0.292	0.316	0.244	0.274	0.283	0.248	0.274	0.250
		minRC	75.3%	19.3%	29.0%	80.1%	81.4%	74.1%	76.0%	59.9%	79.2%	88.8%	76.7%	85.2%	89.3%
	GWNET	cov	89.9%	68.0%	50.5%	88.9%	92.3%	90.2%	95.1%	80.3%	89.5%	90.1%	90.0%	90.0%	89.8%
		length	0.251	0.132	0.116	0.257	0.266	0.286	0.349	0.246	0.268	0.274	0.251	0.267	0.251
		minRC	75.8%	20.7%	39.7%	74.9%	81.9%	74.3%	84.2%	57.3%	82.0%	88.5%	75.8%	86.0%	89.2%
	AVG	cov	86.6%	65.2%	46.4%	88.7%	91.8%	88.9%	93.7%	81.8%	89.6%	90.1%	88.5%	89.9%	89.9%
		length	0.251	0.138	0.098	0.252	0.272	0.275	0.311	0.304	0.280	0.286	0.268	0.280	0.248

		minRC	72.6%	26.3%	33.9%	77.0%	77.1%	71.2%	78.0%	56.9%	73.9%	88.5%	67.0%	85.9%	89.2%
4month	STGCN	cov	88.1%	59.4%	32.6%	88.0%	91.8%	89.7%	91.2%	90.4%	89.6%	89.9%	90.0%	89.9%	89.9%
		length	0.231	0.170	0.073	0.241	0.269	0.295	0.294	0.248	0.316	0.324	0.233	0.324	0.248
		minRC	71.8%	42.6%	20.6%	74.3%	74.9%	68.9%	70.9%	77.1%	51.5%	88.7%	78.3%	86.9%	89.3%
	DCRNN	cov	64.8%	62.6%	59.9%	88.9%	90.6%	85.6%	94.7%	71.9%	89.8%	90.1%	79.6%	90.5%	89.9%
		length	0.309	0.239	0.118	0.243	0.295	0.222	0.296	0.511	0.265	0.274	0.371	0.269	0.250
		minRC	66.7%	21.1%	43.0%	81.7%	80.4%	71.1%	81.6%	29.8%	79.9%	89.0%	74.7%	87.0%	89.5%
	MTGNN	cov	89.6%	71.8%	42.0%	89.7%	93.1%	90.5%	93.4%	81.1%	89.9%	90.1%	90.1%	90.6%	89.9%
		length	0.247	0.215	0.085	0.262	0.268	0.287	0.311	0.244	0.274	0.284	0.248	0.277	0.254
		minRC	76.3%	32.0%	28.4%	81.0%	80.2%	76.5%	76.8%	60.4%	78.1%	88.9%	78.4%	86.4%	89.4%
	GWNET	cov	90.0%	68.5%	50.8%	89.1%	92.4%	90.3%	95.4%	80.3%	89.9%	90.1%	90.2%	90.4%	89.9%
		length	0.249	0.232	0.115	0.255	0.265	0.281	0.346	0.245	0.268	0.277	0.249	0.270	0.255
		minRC	82.1%	32.1%	39.5%	81.2%	85.9%	78.8%	81.3%	56.0%	78.1%	88.8%	82.4%	86.6%	89.4%
	AVG	cov	83.1%	65.6%	46.3%	88.9%	92.0%	89.0%	93.7%	80.9%	89.8%	90.0%	87.5%	90.3%	89.9%
		length	0.259	0.214	0.098	0.250	0.274	0.271	0.312	0.312	0.281	0.290	0.275	0.285	0.252
		minRC	74.2%	32.0%	32.9%	79.5%	80.3%	73.8%	77.7%	55.8%	71.9%	88.9%	78.5%	86.7%	89.4%
1st			0	0	0	0	0	0	0	0	0	0	0	0	20

Table A3 Full results of CHIbike dataset

Time	Base model	metric	QR	MCD	boostrop	MIS	DESQRUQ	UATGCN	ProbGNN	QuanTraffic	CP	ACI	QCP	DtACI	CONTINA
1month	STGCN	cov	88.0%	32.0%	22.5%	86.2%	93.3%	90.8%	92.3%	87.7%	90.0%	90.2%	90.0%	90.2%	89.7%
		length	0.487	0.164	0.088	0.481	0.499	0.507	0.534	0.504	0.609	0.627	0.488	0.611	0.494
		minRC	78.5%	22.1%	11.9%	81.3%	90.6%	84.2%	88.3%	79.5%	86.0%	88.3%	84.5%	86.2%	89.2%
	DCRNN	cov	89.7%	27.4%	23.7%	89.3%	93.1%	92.2%	94.9%	88.0%	90.1%	90.3%	90.3%	90.3%	89.6%
		length	0.555	0.162	0.144	0.558	0.570	0.588	0.674	0.572	0.659	0.670	0.556	0.659	0.565
		minRC	83.3%	15.0%	0.9%	84.6%	89.0%	84.7%	92.0%	71.9%	87.0%	88.8%	83.6%	86.8%	88.8%
	MTGNN	cov	89.3%	27.4%	22.9%	91.8%	93.9%	92.4%	93.6%	86.4%	90.0%	90.2%	89.4%	90.2%	89.9%
		length	0.513	0.177	0.093	0.511	0.535	0.547	0.581	0.522	0.611	0.627	0.513	0.613	0.522
		minRC	80.2%	16.1%	11.9%	82.3%	91.2%	88.7%	91.0%	75.8%	86.4%	87.4%	80.3%	86.7%	89.4%
	GWNET	cov	91.1%	32.4%	22.7%	91.9%	93.6%	93.1%	93.5%	85.7%	90.0%	90.2%	89.9%	90.0%	89.7%
		length	0.500	0.251	0.105	0.503	0.521	0.570	0.582	0.509	0.614	0.627	0.500	0.613	0.503
		minRC	87.6%	22.7%	12.3%	85.1%	90.9%	89.8%	90.1%	72.4%	85.5%	88.9%	84.4%	86.0%	88.8%
	AVG	cov	89.5%	29.8%	22.9%	89.8%	93.5%	92.1%	93.6%	87.0%	90.0%	90.2%	89.9%	90.2%	89.7%
		length	0.514	0.188	0.107	0.513	0.531	0.553	0.593	0.527	0.623	0.638	0.514	0.624	0.521
		minRC	82.4%	19.0%	9.3%	83.3%	90.4%	86.9%	90.3%	74.9%	86.2%	88.3%	83.2%	86.4%	89.0%
2month	STGCN	cov	87.8%	33.8%	23.2%	86.0%	93.0%	90.7%	92.3%	87.6%	88.1%	89.5%	89.7%	89.1%	89.9%
		length	0.519	0.189	0.096	0.520	0.537	0.542	0.571	0.536	0.610	0.678	0.520	0.642	0.530
		minRC	75.3%	23.0%	12.1%	83.1%	90.1%	85.1%	88.7%	80.4%	83.8%	87.7%	83.6%	86.8%	89.5%
	DCRNN	cov	89.1%	29.2%	24.6%	88.9%	92.7%	92.0%	94.8%	87.7%	88.1%	89.4%	89.6%	89.1%	89.8%
		length	0.591	0.188	0.159	0.594	0.608	0.627	0.720	0.609	0.657	0.741	0.592	0.693	0.606
		minRC	83.0%	14.7%	2.9%	83.8%	88.2%	86.0%	92.1%	73.7%	84.9%	87.9%	83.2%	86.7%	89.2%
	MTGNN	cov	88.6%	29.2%	24.2%	91.0%	93.4%	92.5%	93.7%	86.1%	88.1%	89.5%	88.7%	89.1%	89.7%

		length	0.558	0.205	0.106	0.554	0.578	0.592	0.628	0.566	0.613	0.682	0.558	0.645	0.569
		minRC	79.1%	18.8%	14.0%	81.3%	90.5%	89.0%	89.6%	75.3%	83.3%	88.1%	79.3%	86.4%	89.0%
	GWNET	cov	90.9%	33.3%	23.7%	91.7%	93.4%	93.0%	93.5%	85.9%	88.2%	89.4%	89.8%	89.0%	89.8%
		length	0.542	0.278	0.118	0.548	0.566	0.612	0.627	0.552	0.616	0.681	0.542	0.640	0.547
		minRC	88.1%	23.8%	14.0%	87.6%	90.4%	89.3%	90.3%	74.2%	83.5%	88.0%	85.2%	86.3%	89.0%
	AVG	cov	89.1%	31.4%	24.0%	89.4%	93.1%	92.1%	93.6%	86.8%	88.1%	89.5%	89.5%	89.1%	89.8%
		length	0.553	0.215	0.120	0.554	0.572	0.593	0.637	0.566	0.624	0.696	0.553	0.655	0.563
		minRC	81.4%	20.1%	10.7%	84.0%	89.8%	87.3%	90.2%	75.9%	83.9%	87.9%	82.8%	86.5%	89.2%
3month	STGCN	cov	87.7%	34.3%	22.4%	86.2%	92.3%	90.4%	92.0%	87.6%	86.6%	90.2%	89.3%	90.0%	89.8%
		length	0.582	0.227	0.109	0.586	0.602	0.609	0.641	0.599	0.640	0.769	0.583	0.748	0.597
		minRC	78.3%	23.1%	12.6%	82.6%	89.3%	85.8%	89.4%	80.9%	81.6%	89.4%	83.7%	87.4%	89.4%
	DCRNN	cov	88.3%	29.5%	25.5%	88.4%	91.7%	90.7%	94.2%	87.3%	86.3%	90.2%	88.7%	89.9%	89.8%
		length	0.648	0.220	0.177	0.652	0.669	0.683	0.789	0.666	0.692	0.842	0.649	0.813	0.675
		minRC	81.5%	15.1%	2.6%	82.8%	87.4%	84.9%	91.9%	77.7%	82.3%	89.4%	81.9%	87.6%	89.2%
	MTGNN	cov	88.2%	30.7%	23.7%	90.6%	92.7%	92.1%	93.2%	86.2%	86.7%	90.2%	88.3%	90.2%	89.9%
		length	0.620	0.244	0.118	0.616	0.639	0.660	0.695	0.623	0.642	0.769	0.620	0.752	0.637
		minRC	80.6%	20.1%	13.3%	81.8%	89.5%	89.7%	90.3%	77.4%	82.9%	89.3%	80.9%	88.3%	89.4%
	GWNET	cov	90.4%	34.8%	23.6%	91.1%	92.6%	92.4%	93.2%	86.0%	86.6%	90.2%	89.3%	89.8%	89.8%
		length	0.601	0.310	0.133	0.612	0.629	0.676	0.697	0.611	0.646	0.777	0.601	0.747	0.608
		minRC	86.5%	25.5%	14.7%	87.9%	89.9%	89.4%	89.9%	73.6%	81.9%	89.5%	85.2%	88.1%	89.4%
	AVG	cov	88.7%	32.3%	23.8%	89.1%	92.3%	91.4%	93.1%	86.8%	86.6%	90.2%	88.9%	90.0%	89.8%
		length	0.613	0.251	0.134	0.617	0.635	0.657	0.706	0.625	0.655	0.789	0.613	0.765	0.629
		minRC	81.7%	21.0%	10.8%	83.8%	89.0%	87.4%	90.4%	77.4%	82.2%	89.4%	82.9%	87.8%	89.4%
4month	STGCN	cov	86.8%	37.8%	20.8%	86.1%	90.9%	89.7%	91.5%	87.0%	80.0%	90.0%	88.1%	89.6%	89.8%
		length	0.804	0.399	0.154	0.824	0.882	0.858	0.898	0.821	0.697	1.093	0.804	1.035	0.848
		minRC	82.4%	27.4%	11.6%	71.7%	87.8%	85.4%	87.9%	81.9%	70.4%	88.6%	83.4%	87.1%	89.4%
	DCRNN	cov	86.2%	33.4%	26.8%	86.7%	89.9%	88.9%	93.2%	86.1%	79.4%	90.1%	86.5%	89.3%	89.9%
		length	0.839	0.378	0.248	0.854	0.914	0.912	1.050	0.857	0.755	1.213	0.840	1.143	0.930
		minRC	79.8%	18.6%	8.6%	79.3%	84.8%	82.4%	91.0%	77.7%	70.8%	88.5%	79.9%	87.0%	88.3%
	MTGNN	cov	87.4%	35.5%	21.1%	89.6%	91.3%	91.6%	92.4%	86.3%	79.8%	90.0%	87.4%	89.1%	89.8%
		length	0.858	0.409	0.159	0.850	0.896	0.926	0.949	0.846	0.698	1.093	0.859	1.018	0.893
		minRC	82.9%	22.5%	14.2%	80.3%	88.8%	89.1%	88.9%	78.6%	71.4%	88.6%	83.1%	86.1%	89.4%
	GWNET	cov	89.4%	37.4%	22.5%	90.0%	91.5%	91.7%	92.6%	86.3%	80.0%	90.1%	88.6%	89.5%	89.9%
		length	0.839	0.434	0.189	0.849	0.866	0.926	0.960	0.848	0.703	1.096	0.839	1.040	0.860
		minRC	87.0%	28.4%	15.8%	86.7%	88.2%	88.4%	90.3%	76.0%	71.3%	88.9%	85.5%	86.5%	89.5%
	AVG	cov	87.5%	36.0%	22.8%	88.1%	90.9%	90.5%	92.4%	86.4%	79.8%	90.0%	87.7%	89.4%	89.9%
		length	0.835	0.405	0.188	0.844	0.889	0.906	0.964	0.843	0.713	1.124	0.835	1.059	0.883
		minRC	83.0%	24.2%	12.6%	79.5%	87.4%	86.3%	89.5%	78.6%	71.0%	88.6%	83.0%	86.7%	89.1%
1st			4	0	0	0	1	0	0	0	0	0	2	0	16

Table A4 Full results of CHItaxi dataset

Time	Base model	metric	QR	MCD	boostrop	MIS	DESQRUQ	UATGCN	ProbGNN	QuanTraffic	CP	ACI	QCP	DtACI	CONTINA
1month	STGCN	cov	91.6%	45.6%	44.6%	91.5%	93.0%	93.1%	94.0%	89.8%	92.1%	89.9%	91.3%	90.4%	89.6%

		length	0.206	0.282	0.076	0.216	0.217	0.287	0.300	0.228	0.297	0.273	0.206	0.266	0.215
		minRC	83.9%	14.2%	27.4%	84.7%	86.6%	87.8%	90.1%	87.4%	89.8%	87.3%	83.9%	87.8%	89.0%
	DCRNN	cov	87.8%	45.8%	44.2%	89.8%	92.7%	92.0%	94.9%	88.9%	92.0%	90.0%	91.3%	90.3%	89.6%
		length	0.198	0.235	0.098	0.211	0.217	0.320	0.370	0.227	0.297	0.270	0.204	0.264	0.217
		minRC	80.3%	15.8%	27.6%	84.7%	86.9%	86.7%	91.5%	86.3%	89.4%	88.0%	81.0%	87.6%	89.0%
	MTGNN	cov	91.1%	46.9%	49.3%	90.8%	93.1%	92.9%	94.4%	89.3%	91.9%	89.9%	90.7%	90.2%	89.7%
		length	0.213	0.276	0.075	0.223	0.226	0.258	0.283	0.228	0.299	0.275	0.213	0.268	0.221
		minRC	85.8%	15.9%	30.5%	85.3%	88.3%	88.7%	91.0%	86.4%	89.6%	87.2%	85.8%	87.0%	89.2%
	GWNET	cov	90.6%	46.9%	39.7%	93.2%	93.9%	93.3%	94.3%	88.9%	91.9%	90.0%	90.2%	90.0%	89.6%
		length	0.216	0.282	0.073	0.230	0.227	0.249	0.275	0.227	0.294	0.275	0.216	0.263	0.222
		minRC	84.8%	14.2%	23.8%	87.6%	86.9%	88.1%	91.3%	87.3%	89.8%	87.6%	82.8%	87.5%	88.8%
	AVG	cov	90.3%	46.3%	44.4%	91.3%	93.2%	92.8%	94.4%	89.2%	92.0%	90.0%	90.9%	90.2%	89.6%
		length	0.208	0.269	0.081	0.220	0.222	0.279	0.307	0.227	0.297	0.273	0.210	0.265	0.219
		minRC	83.7%	15.0%	27.3%	85.6%	87.2%	87.8%	91.0%	86.9%	89.6%	87.5%	83.4%	87.5%	89.0%
2month	STGCN	cov	91.7%	46.8%	44.0%	91.9%	93.2%	93.4%	94.1%	90.3%	90.9%	90.0%	91.4%	89.9%	89.8%
		length	0.219	0.311	0.080	0.231	0.231	0.308	0.322	0.241	0.282	0.281	0.219	0.260	0.225
		minRC	85.5%	14.2%	27.4%	85.5%	88.9%	89.6%	90.6%	87.7%	88.2%	87.2%	85.5%	87.4%	89.4%
	DCRNN	cov	88.2%	47.7%	44.4%	90.1%	92.8%	92.3%	95.0%	89.4%	90.8%	90.0%	91.6%	90.0%	89.8%
		length	0.209	0.261	0.105	0.224	0.229	0.350	0.405	0.240	0.282	0.279	0.215	0.262	0.223
		minRC	80.6%	17.6%	27.1%	84.8%	87.2%	86.9%	92.1%	86.5%	88.6%	87.9%	81.0%	87.6%	89.4%
	MTGNN	cov	91.6%	48.6%	48.7%	91.0%	93.4%	93.2%	94.6%	89.0%	90.8%	90.0%	91.2%	90.1%	89.8%
		length	0.227	0.308	0.078	0.237	0.240	0.276	0.302	0.241	0.286	0.285	0.226	0.267	0.230
		minRC	87.2%	16.2%	31.4%	85.9%	88.9%	89.9%	91.3%	86.5%	87.8%	87.5%	87.2%	87.3%	89.4%
	GWNET	cov	90.8%	48.5%	39.5%	93.4%	94.1%	93.4%	94.4%	89.4%	90.7%	90.0%	90.4%	89.9%	89.8%
		length	0.228	0.309	0.078	0.245	0.241	0.264	0.292	0.240	0.282	0.284	0.228	0.266	0.232
		minRC	85.4%	14.0%	22.7%	88.5%	88.4%	87.3%	91.8%	85.5%	88.5%	87.6%	83.8%	87.6%	89.4%
	AVG	cov	90.6%	47.9%	44.2%	91.6%	93.4%	93.1%	94.5%	89.5%	90.8%	90.0%	91.1%	90.0%	89.8%
		length	0.221	0.297	0.085	0.234	0.235	0.300	0.330	0.240	0.283	0.282	0.222	0.264	0.228
		minRC	84.7%	15.5%	27.1%	86.2%	88.4%	88.4%	91.4%	86.6%	88.3%	87.6%	84.3%	87.5%	89.4%
3month	STGCN	cov	91.0%	48.1%	42.9%	91.0%	92.4%	92.7%	93.5%	90.2%	88.7%	89.6%	90.7%	89.7%	89.8%
		length	0.240	0.350	0.090	0.253	0.259	0.339	0.355	0.262	0.282	0.324	0.240	0.290	0.253
		minRC	82.9%	16.0%	27.5%	83.2%	85.6%	84.7%	86.0%	86.1%	84.2%	86.5%	82.4%	86.7%	89.4%
	DCRNN	cov	87.2%	48.5%	43.9%	89.4%	92.0%	91.6%	94.5%	89.5%	88.6%	89.7%	90.3%	89.7%	89.8%
		length	0.226	0.294	0.120	0.244	0.249	0.391	0.449	0.262	0.282	0.325	0.232	0.294	0.255
		minRC	78.7%	19.4%	25.3%	83.3%	85.8%	85.1%	91.0%	87.0%	83.8%	87.1%	79.3%	86.3%	89.2%
	MTGNN	cov	91.0%	50.0%	47.2%	90.1%	92.7%	92.5%	94.0%	89.9%	88.6%	89.6%	90.6%	89.7%	89.8%
		length	0.249	0.344	0.086	0.260	0.262	0.305	0.334	0.262	0.286	0.332	0.248	0.298	0.256
		minRC	86.3%	17.2%	27.4%	85.0%	87.9%	88.3%	90.8%	86.7%	84.5%	87.2%	86.3%	86.8%	89.4%
	GWNET	cov	90.2%	49.3%	39.0%	92.8%	93.4%	92.8%	93.8%	88.5%	88.6%	89.7%	89.9%	89.8%	89.8%
		length	0.250	0.343	0.088	0.270	0.263	0.290	0.321	0.262	0.283	0.326	0.250	0.296	0.258
		minRC	85.2%	15.1%	22.6%	87.2%	88.3%	88.1%	89.4%	86.1%	83.5%	86.7%	83.4%	86.2%	89.5%
	AVG	cov	89.8%	49.0%	43.3%	90.8%	92.6%	92.4%	93.9%	89.6%	88.6%	89.7%	90.4%	89.7%	89.8%
		length	0.241	0.333	0.096	0.257	0.258	0.331	0.365	0.262	0.283	0.326	0.243	0.294	0.256

		minRC	83.3%	16.9%	25.7%	84.7%	86.9%	86.5%	89.3%	86.5%	84.0%	86.9%	82.9%	86.5%	89.4%
4month	STGCN	cov	91.1%	48.0%	42.4%	91.3%	92.6%	93.0%	93.8%	90.6%	89.4%	90.1%	90.8%	90.1%	89.9%
		length	0.236	0.336	0.086	0.249	0.247	0.330	0.343	0.258	0.287	0.318	0.236	0.300	0.250
		minRC	82.7%	18.9%	27.8%	84.5%	82.6%	86.5%	88.2%	84.9%	82.6%	88.3%	82.8%	86.8%	89.3%
	DCRNN	cov	87.6%	47.9%	43.5%	89.7%	92.3%	91.1%	94.9%	89.7%	89.4%	90.1%	90.7%	90.2%	90.0%
		length	0.222	0.283	0.113	0.240	0.245	0.366	0.426	0.257	0.287	0.320	0.229	0.302	0.250
		minRC	79.1%	20.0%	29.8%	82.7%	83.2%	83.5%	91.7%	83.0%	82.6%	88.4%	79.4%	87.0%	89.3%
	MTGNN	cov	91.2%	49.5%	46.8%	90.4%	93.0%	92.9%	94.3%	88.3%	89.4%	90.1%	90.9%	90.2%	89.9%
		length	0.244	0.327	0.082	0.256	0.257	0.294	0.323	0.258	0.291	0.322	0.244	0.305	0.252
		minRC	83.5%	21.4%	31.4%	84.7%	87.1%	88.7%	91.3%	85.8%	81.4%	87.9%	82.5%	86.8%	89.4%
	GWNET	cov	90.4%	49.5%	38.4%	93.0%	93.4%	93.0%	94.1%	88.7%	89.3%	90.1%	90.0%	90.1%	89.9%
		length	0.245	0.331	0.084	0.264	0.257	0.284	0.311	0.257	0.288	0.321	0.245	0.300	0.254
		minRC	83.2%	21.4%	24.2%	86.8%	85.9%	86.6%	90.1%	83.1%	82.7%	88.2%	83.1%	87.2%	89.4%
	AVG	cov	90.1%	48.7%	42.8%	91.1%	92.8%	92.5%	94.3%	89.3%	89.4%	90.1%	90.6%	90.2%	89.9%
		length	0.237	0.319	0.091	0.252	0.252	0.319	0.351	0.257	0.288	0.320	0.238	0.302	0.251
		minRC	82.1%	20.4%	28.3%	84.7%	84.7%	86.3%	90.3%	84.2%	82.3%	88.2%	82.0%	87.0%	89.4%
1st			5	0	0	0	1	0	0	0	0	0	3	0	13

Our method achieves the best prediction results 15, 16, 20, and 13 times across four datasets, respectively. In cases where our method fails to achieve the best prediction result, it typically obtained the second-best prediction result. This undoubtedly demonstrates the competitiveness of our approach.

C2 Full results of sensitive analysis

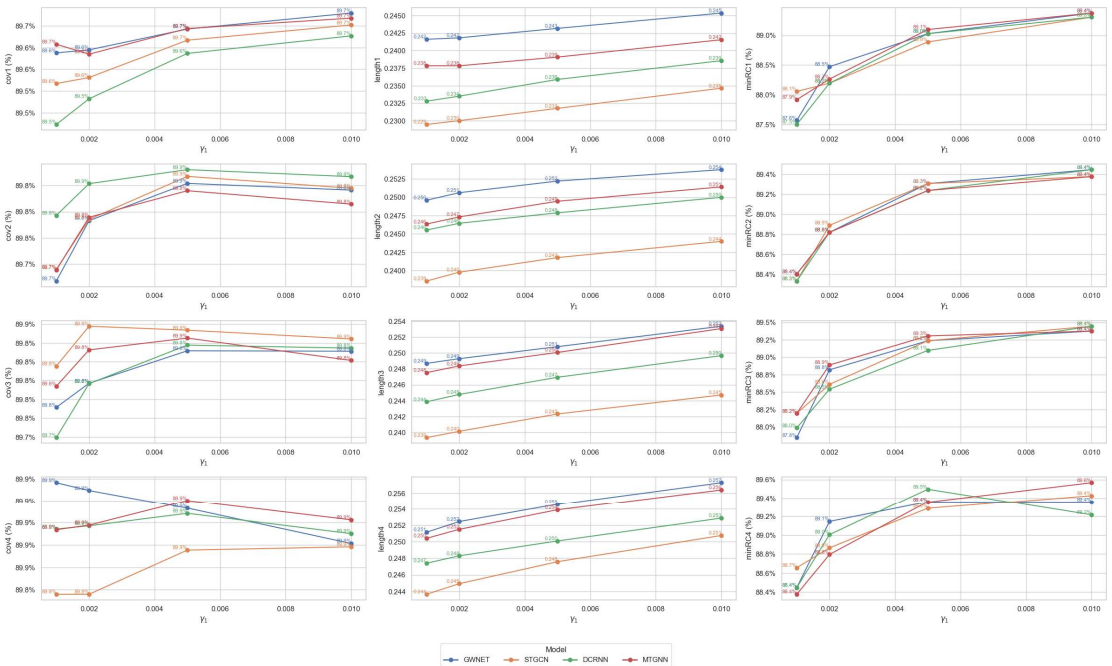

Figure A1. Results of sensitive analysis in NYCTaxi dataset

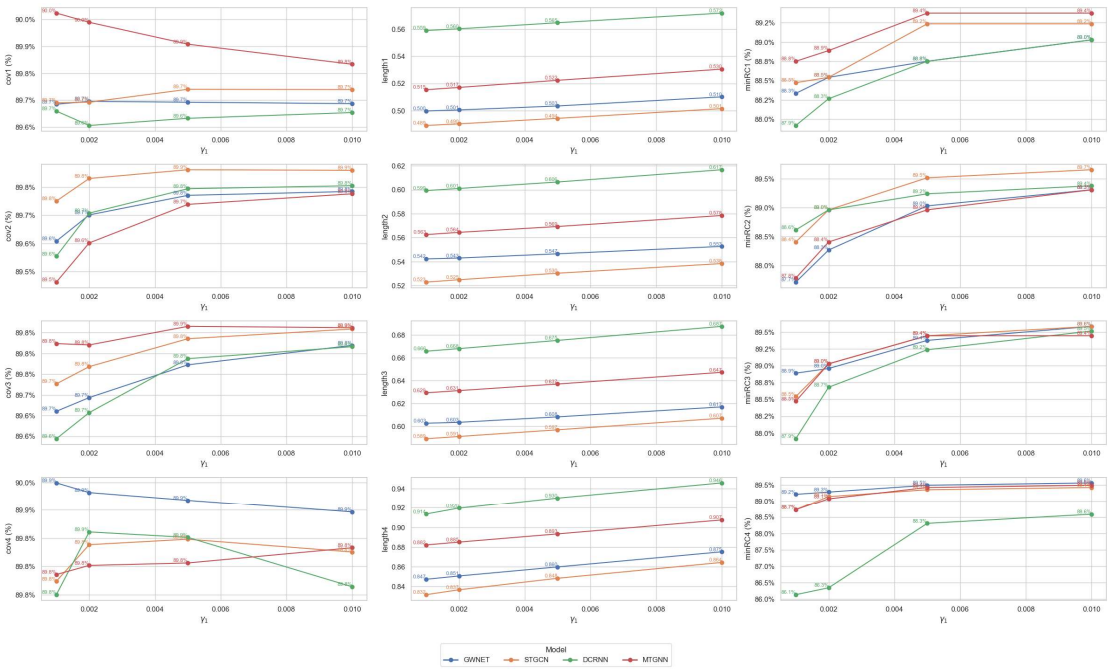

Figure A2. Results of sensitive analysis in CHIbike dataset

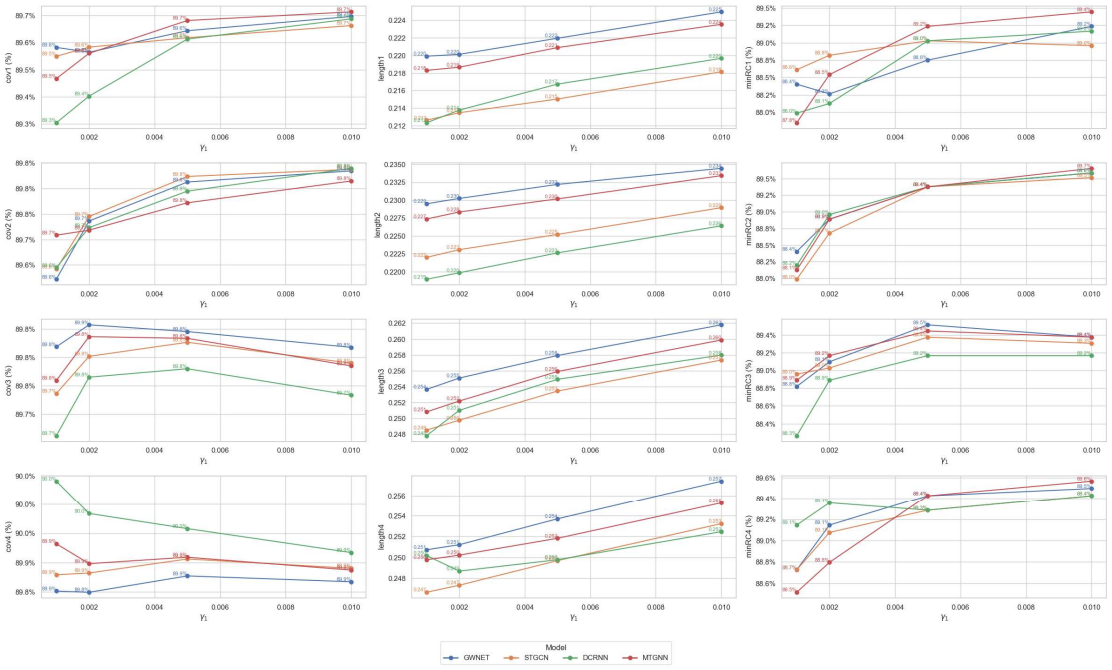

Figure A3. Results of sensitive analysis in CHITaxi dataset

The results of sensitive analysis in NYCtaxi, CHIbike and CHItaxi datasets are similar to the results in NYCBike datasets.

C2 Full results of adaptive learning rate

STGCN:

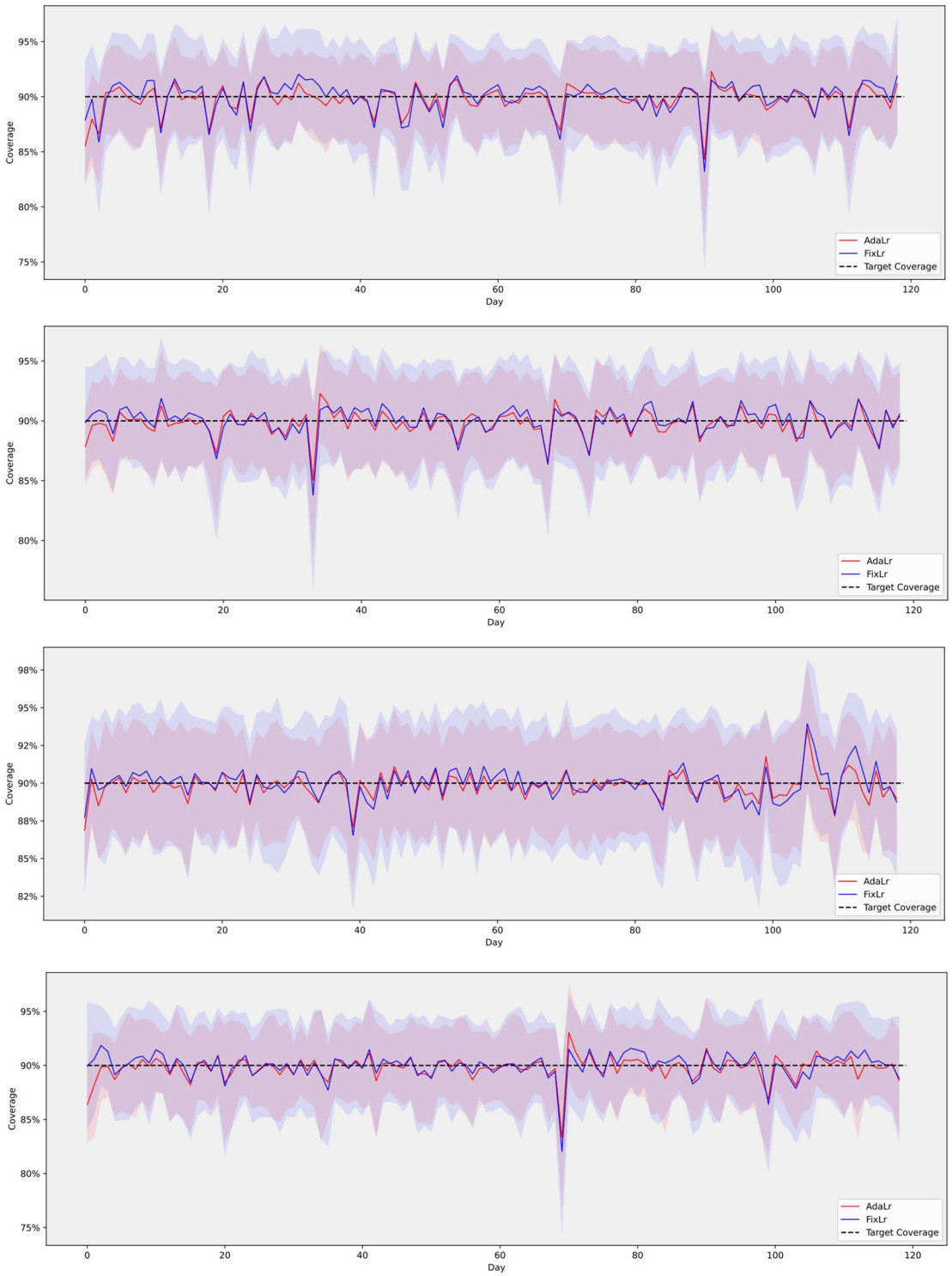

Figure A4-7. Regionl coverage of NYCbike, NYCtaxi, CHIbike and CHITaxi dataset using STGCN

DCRNN

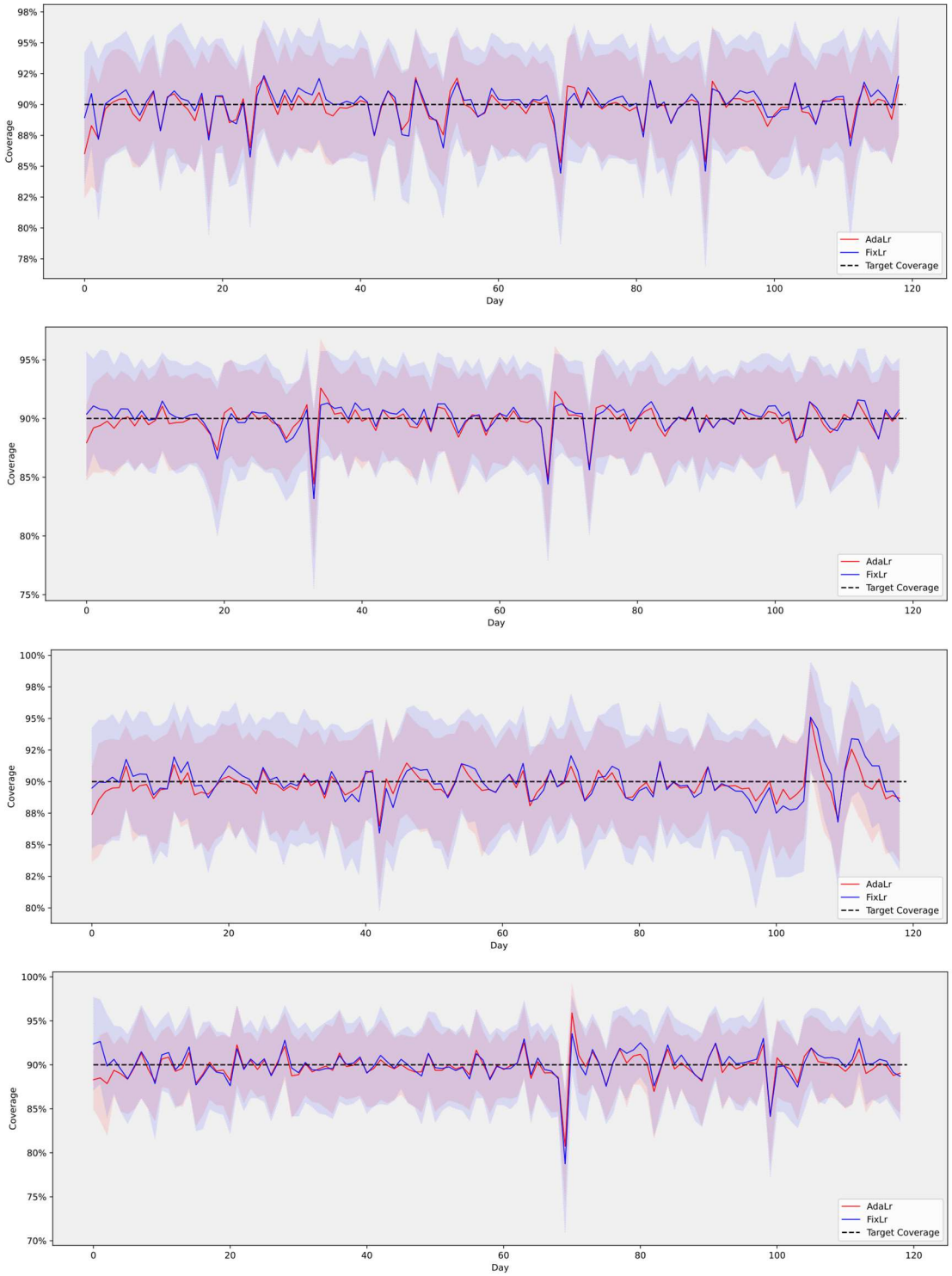

Figure A8-11. Regionl coverage of NYCbike, NYCtaxi, CHIbike and CHITaxi dataset using DCRNN

MTGNN

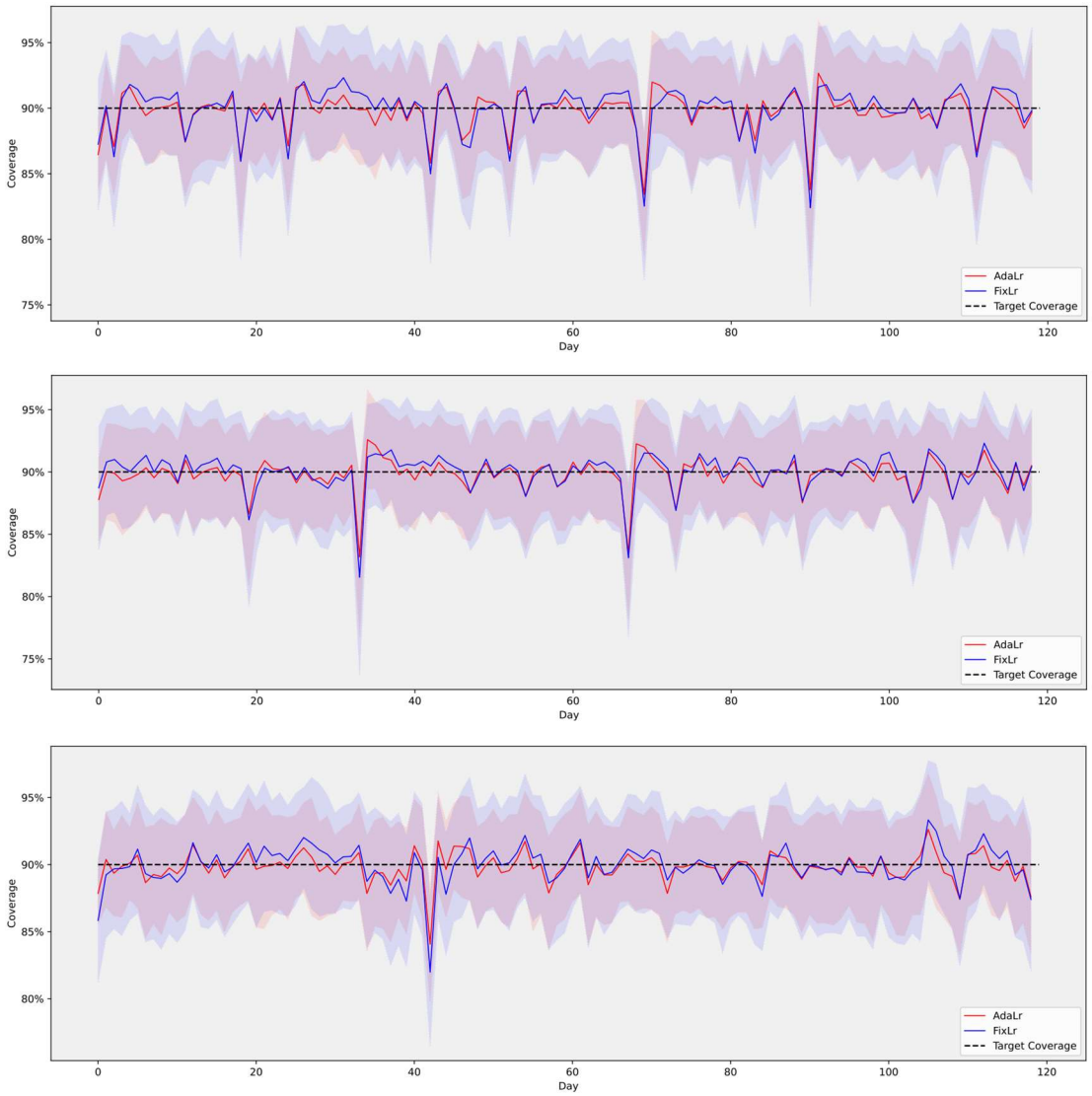

Figure A12-15. Regionl coverage of NYCbike, NYCtaxi, CHIbike and CHITaxi dataset using MTGNN

It could be found that the coverages of regions when using adaptive learning rate are more concentrated on 90% than using fixed learning rate.

E Visualization of confidence intervals

We plot the confidence based on STGCN model in NYCbike, NYCTaxi and CHIbike datasets in the following figures.

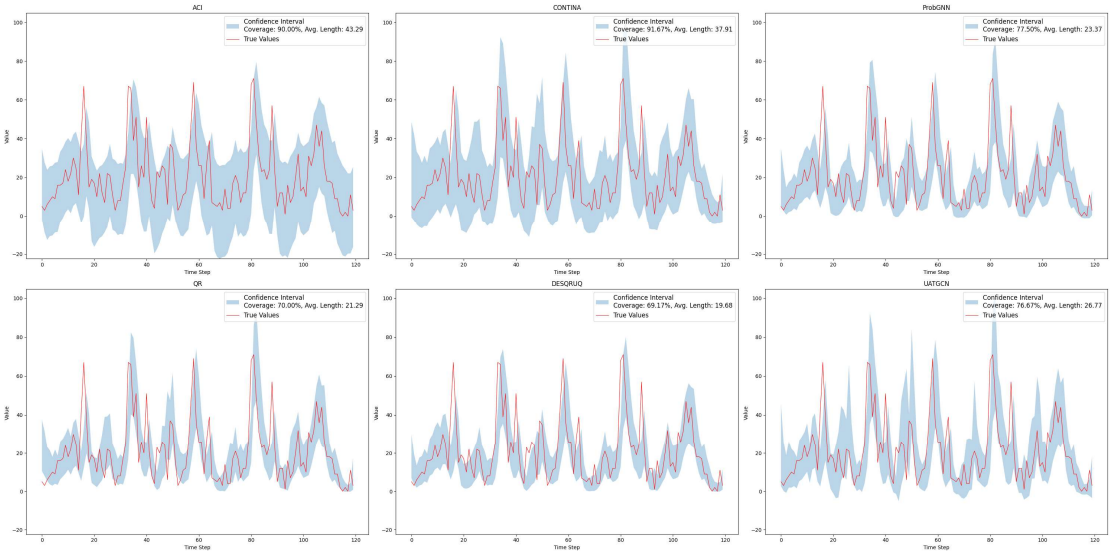

Figure A16. Confidence intervals in NYCBike dataset

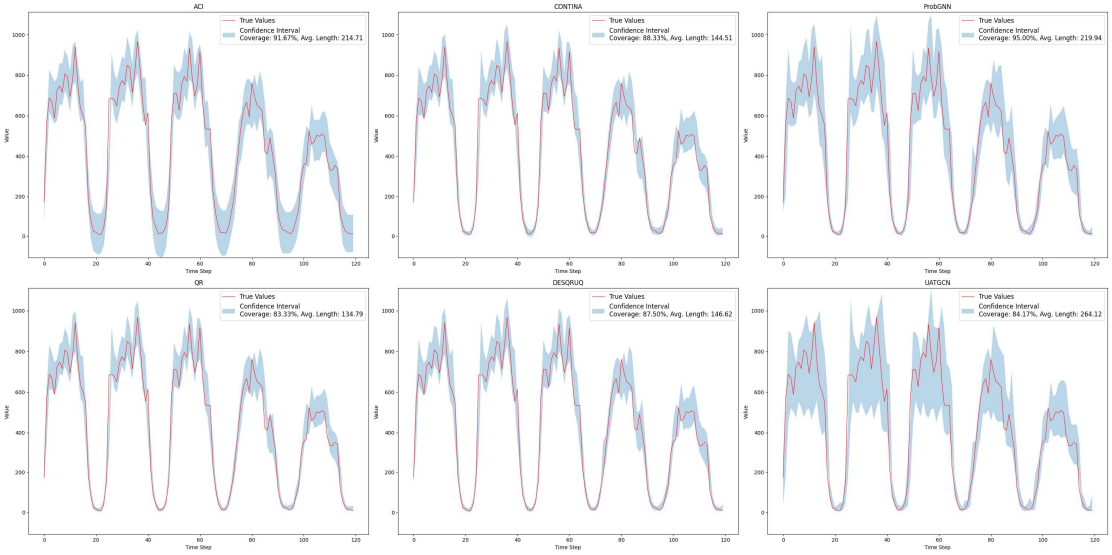

Figure A17. Confidence intervals in NYCTaxi dataset

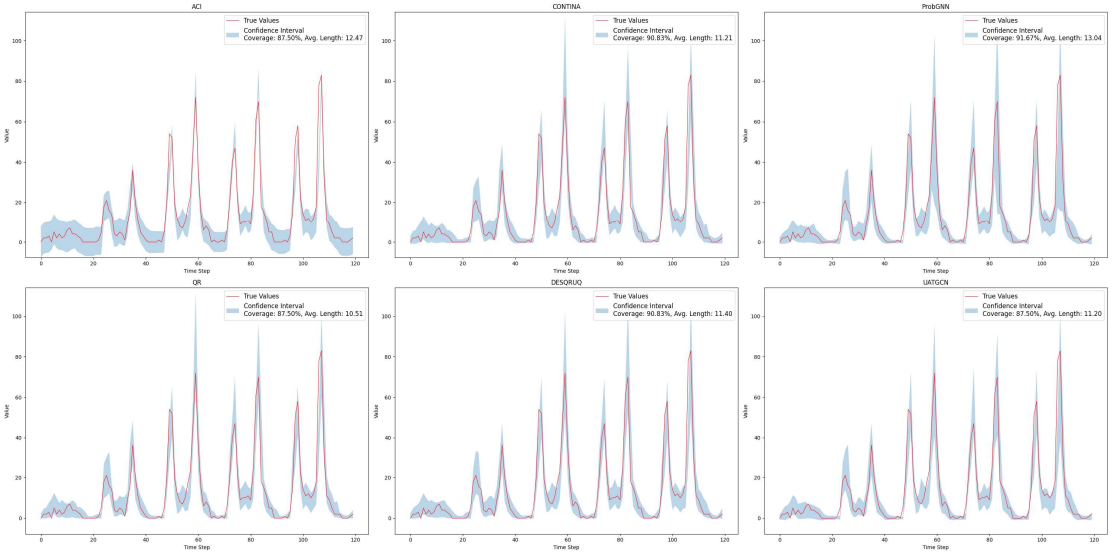

Figure A18. Confidence intervals in CHibike dataset

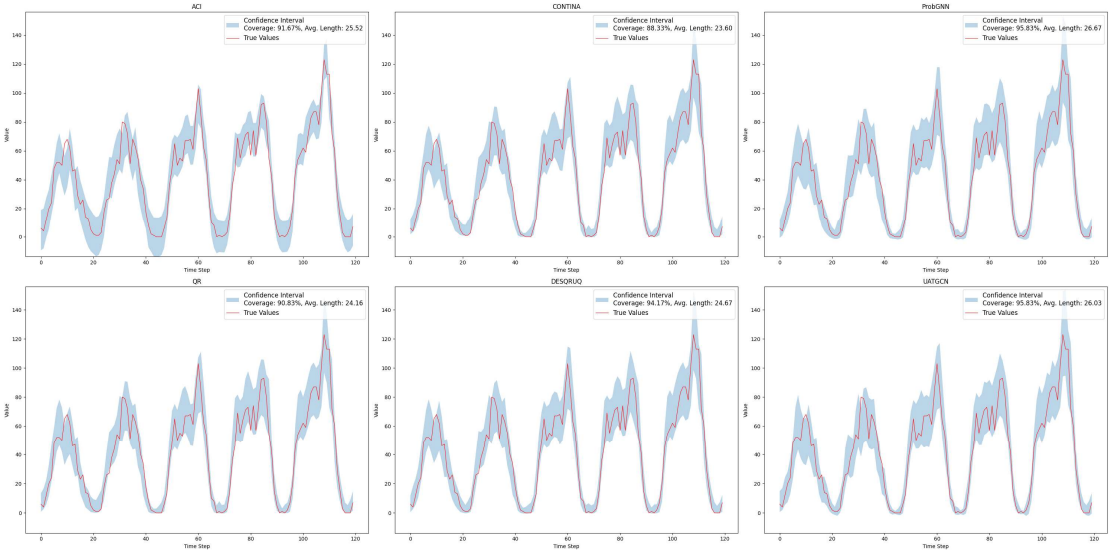

Figure A18. Confidence intervals in CHitaxi datase