The 9th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2024
3–7 June 2024, Lisboa, Portugal

TRAFFIC FORECASTING WITH UNCERTAINTY: A CASE FOR CONFORMALIZED QUANTILE REGRESSION

Gašper Petelin[1,2], Jože M. Rožanec[2,3,4] and Gregor Papa[1,2]

[1]Computer Systems Department, Jožef Stefan Institute, Ljubljana, 1000, Slovenia
Email: gasper.petelin@ijs.si

[2]Jožef Stefan International Postgraduate School, Ljubljana, 1000, Slovenia

[3]Laboratory of Artificial Intelligence, Jožef Stefan Institute, Ljubljana, 1000, Slovenia

[4]Qlector d.o.o., Rovšnikova 7, 1000 Ljubljana, Slovenia

Key words: Traffic Flow Forecasting, Prediction Intervals, Conformalized Quantile Regression Models

Summary. Accurate and reliable traffic flow forecasting is of primary importance for traffic planning and management. While there is a growing interest in real-time traffic forecasting models, accurate predictions remain a challenge due to the dynamic nature of traffic systems and the multiple factors that affect the traffic flow. Point forecasts do not provide insights regarding uncertainties associated with forecasts. Furthermore, many traffic flow models fail to produce prediction intervals that accurately capture the uncertainty of the forecasts. Therefore, we investigate the use of quantile regression models for traffic flow forecasting and highlight their tendency to generate prediction intervals that are too narrow and poorly calibrated. To address this issue, we propose using conformal predictions, which allow us to achieve well-calibrated prediction intervals leading to more accurate, reliable and therefore trustworthy predictions.

1 INTRODUCTION

Accurate traffic flow forecasting is crucial for improving traffic planning and management, reducing congestion and increasing the overall efficiency of transport systems. In recent years, there has been increasing interest in developing models and algorithms that can accurately forecast traffic conditions in real-time. However, accurate traffic forecasting remains a challenging task due to the complex and dynamic nature of transport systems and the numerous factors that can affect traffic flows. Despite their popularity, point forecasts do not have an uncertainty associated with the forecast. This can reduce their usability and trustworthiness in the context of decision-making. On the other hand, knowing the range of possible outcomes, including worst-case and best-case scenarios, can help traffic managers use their resources more effectively, plan detours and alternate routes, and implement safety measures when needed. It is, therefore, crucial to generate prediction intervals with coverage levels when forecasting future traffic flow, as this can contribute significantly to the usability of the models.

Our contribution: The paper explores the use of quantile regression models for traffic flow forecasts, highlighting their advantages but also their tendency to generate prediction intervals that are too narrow and therefore poorly calibrated. Secondly, it demonstrates how the use of conformal prediction can mitigate this problem by providing well-calibrated prediction intervals, ensuring the model's predictions remain closer to the desired probability coverage, leading to more trustworthy forecasts. We published the dataset and code, to make them publicly available [1].

The paper is structured as follows: Section 2 provides a literature review on the topic. Section 3 offers a detailed insight into the forecasting system we used, the results of which are presented in Section 4. Finally, in Section 5 we conclude our work.

2 RELATED WORK

In this section, we briefly describe related work regarding traffic forecasting (Section 2.1) and forecasting methods related to probabilistic forecasting and forecasts with prediction intervals (Section 2.2).

2.1 Short-term Traffic Forecasting

Short-term traffic forecasting has been an active area of research for many years [21]. Various techniques have been proposed for accurate traffic flow prediction, including statistical methods (e.g., ARIMA and exponential smoothing), machine learning algorithms, and hybrid approaches. While statistical methods accurately capture seasonality and trends in traffic flow data, they may fail to model complex and nonlinear systems. Furthermore, they require assumptions about the underlying data distribution (which may not always hold in practice) and are ill-posed to address issues such as missing values in data [18]. In addition, statistical methods usually only model the temporal component of traffic flow and neglect the spatial dimension of the traffic flow problem. On the other hand, machine learning algorithms have shown promising results for short-term traffic forecasting [12]. Machine learning methods can often incorporate external factors (such as weather conditions, holidays, or special events, among others) in a more natural way, work with missing data, or even take into account the topology of road networks. Examples of recent techniques that can model both spatial and temporal components include can be found in [11] but mainly include using graph neural networks. For a more detailed overview of methods for traffic forecasting, we refer the reader to [28] and [4].

2.2 Probabilistic Forecasting

Probabilistic forecasts differ from point forecasts in that they provide a range of probabilities rather than a single outcome. Because some systems are inherently uncertain, probabilistic forecasts can greatly facilitate decision-making. Many solutions to probabilistic forecasting have been proposed (e.g., Bayesian models, Quantile regression, and Deep Learning-based methods). While forecasting a probability range is useful to decision-making, some models make assumptions about the underlying distribution of data and therefore constraining their application. For a more detailed introduction to probabilistic forecasting methods, we refer the reader to [27].

The technique we focus on in this paper is Quantile Regression (QR) [15]. QR estimates the conditional quantiles (groups produced by dividing a frequency distribution into equal groups) of a response variable rather than just the conditional mean. It provides a more complete picture

[1] Code: `https://gitlab.com/ijs-e7/traffic-qcr`

of the distribution of the response variable and can be particularly useful when the distribution is non-normal or heteroscedastic. Unfortunately, the disadvantage of QR is that the prediction intervals are often poorly calibrated (too small or too large), which means that the true value may fall outside the prediction interval (e.g., for a model that predicts intervals from the 5% quantile to the 95% quantile, the interval may or may not contain 90% of the data). To overcome this issue, Conformal Quantile Regression (CQR) [23] has been proposed. Conformal inference is a paradigm for creating statistically rigorous uncertainty intervals for predictive models without any distribution assumptions, where the intervals are guaranteed to contain the ground truth with a specified probability [1]. In regression settings, this is achieved by extending or shortening prediction intervals at both ends by a factor based on a calibration set. While the intervals are shortened or extended by the same margin, the entire process is considered adaptive given the quantile regression produces prediction intervals of multiple sizes. CQR is a method that combines QR with conformal prediction to provide prediction intervals with guaranteed coverage probabilities (under the i.i.d. sample assumption). CQR has been successfully used in various domains [10, 29]. For a detailed description of the conformal prediction framework, we refer the reader to [1, 9, 6].

3 METHODOLOGY

This section provides an overview of the modeling process, including the dataset used, the feature engineering steps, the machine learning models used, and the error metrics used to quantify model performance.

3.1 Dataset

We use the MOL-TF dataset[2] in our experiments, which consists of traffic flow data at 15-minute intervals from 2012 to 2020 using inductive loop counters installed under high-traffic roads. Despite having altogether 190 sensors, only 30-40% of them were operational at any given time, as were taken out of service or were malfunctioning. The dataset has unique properties, such as irregularly sized gaps of missing data ranging from hours to years, and a lack of exact sensor locations for older sensors which were later discontinued (i.e., removed years ago but are still in the dataset). Therefore, when building a model, it is essential to consider these limitations and engineer suitable features to address the challenges posed by the abovementioned peculiarities.

3.2 Feature Engineering and Exogenous Variables

For the purpose of traffic forecasting, we created a total of 275 features. We describe the feature types in Table 1. Among them we find frequently used features to describe time series in a tabular format (e.g., lag_n or $moving\_average_{n,w}$), and other time-based features (e.g., $public\_holiday$ or $date\_features$ (e.g., year, whether it is a weekend or not, the day of the week, the hour, and the minute of the day)). Among domain-specific features we may find the $quantile\_lag_{n,q}$, designed to capture the information from other measuring stations by incorporating traffic flow from them. Due to the large number of measuring station and their high unavailability, a raw traffic count from each station would result in a large number of features

[2]Data available at: `https://repo.ijs.si/cs/datasets/mol-traffic-data`

with missing values. For this reason, traffic flow at all measuring stations is aggregated with quantiles, i.e., new covariates are created to represent the q-th quantile of traffic flow across all stations. In addition, these covariates are converted into features by taking their lags. For example, the feature $quantile\_lag_{3,0.95}$ would indicate the 95-th quantile of total traffic flow 45 minutes ago.

Table 1: Table of features used for forecasting together with a short description.

Feature type	Description
lag_n	Lag feature of the time-series being forecasted for different lags n. [3]
$moving\_average_{n,w}$	Moving average with different window lengths w and lags n. [4]
$quantile\_lag_{n,q}$	Traffic flow divided into different quantiles q and different lags n. [5]
$public\_holiday$	Boolean variable describing if a given day is a holiday or not.
$date\_features$	Features constructed from a date for which forecasts are being made.

3.3 Machine Learning Model

LightGBM [13] is a popular implementation of Gradient Boosted Decision Trees. It uses a gradient-boosting technique to iteratively improve the accuracy of the model by adding new decision trees to the ensemble. The main advantages of the LightGBM are that it can work with missing data, supports different data types (categorical, numerical), and usually does not require scaling of the data to achieve good results. In addition, it is a commonly used model when dealing with data in tabular form [25] where it often outperforms other models [2]. While there are many different machine learning models for traffic flow forecasting, LightGBM is relatively robust, provides good performance without tuning hyperparameters, and is easy to train even in the absence of data. We use one LightGBM model for each of the forecasting horizons separately. Note that the calibration approach is model-independent and does not prescribe a particular machine-learning model.

3.4 Metrics

Quantifying the performance of models is crucial for selecting which models to use. For point forecasts, there is an extensive literature on various metrics [16] for quantifying the quality of forecasts. For quantile forecasts, on the other hand, the selection of metrics is a more complex problem. In quantile regression, the metrics must take into account how often the true value falls within the predicted interval and how wide the interval is. Ideally, one would like to have the smallest possible interval sizes while ensuring that the true values fall within the predicted range in most cases. For this reason, we use two metrics that capture both of the described properties. The mean prediction interval width (MPIW) [24, 14] describes the length of the predicted interval, while the prediction interval coverage probability (PICP) [17, 14] quantifies how often the true value falls outside the predicted range. Thus, one would like to minimize the

[3] lag_n: $\{n \in N \mid (1 \leq n < 100) \vee (n = 100 + 4k \wedge 0 \leq k < 25) \vee (n = 200 + 16k \wedge 0 \leq k < 32)\}$

[4] $moving\_average_{n,w}$: $w \in \{3, 6, 12, 24, 48, 96\}, n \in \{1, 2, 3, 6, 12, 24, 48, 96\}$

[5] $quantile\_lag_{n,q}$: $q \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}, n \in \{1, 2, 3, 6, 12, 24, 48, 96\}$

MPIW while keeping the PICP close to the desired coverage probability. Both the MPIW and PICP metrics are described by the following equations:

$$MPIW_h = \frac{1}{n}\sum_{i=0}^{n}\left|y_{i,h}^{u} - y_{i,h}^{l}\right| \tag{1}$$

$$PICP_h = \frac{1}{n}\sum_{i=0}^{n} I_{i,h}; \quad I_{i,h} = \begin{cases} 0, & y_{i,h} \in [y_{i,h}^{u}, y_{i,h}^{l}] \\ 1, & y_{i,h} \notin [y_{i,h}^{u}, y_{i,h}^{l}] \end{cases} \tag{2}$$

Here, $y_{i,h}$ is the observed number of vehicles while $y_{i,h}^{l}$ and $y_{i,h}^{u}$ are the predicted lower and upper bound for instance i and forecasting horizon h.

3.5 Forecasting Horizon

Many studies [3, 19], use prediction horizons that range from a few minutes to a few hours in advance. In this study, we performed forecasting at multiple time horizons. In particular, we considered forecasting up to eight hours ahead, considering 15-minute intervals. By doing so, we consider we could generate accurate forecasts and eventually support effective decision-making.

4 RESULTS

4.1 Experimental Design

Our dataset was split into two distinct sets - the training set and the test set - based on chronological order. The training set consists of the oldest data from 2012 to 2019, while the test set contains the most recent examples from 2019 to 2022. To generate conformal predictions, we also divided the training and testing data into two sets - one for training the machine learning algorithm and one for calibration. In this case, the calibration set was limited to 2018-2019, while the models were trained with data between 2012-2018. When creating a prediction interval, one must select the desired coverage. In our scenario, we chose a coverage probability of 90%. This means that when future forecasts are made, the actual number of vehicles should be within the prediction interval of approximately 90% of instances (i.e., $PICP \approx 90\%$). As described in [8], the training process is performed in a global setting, i.e., using a single model trained on all time series data for a given time horizon. This approach eliminates the need for separate models for each measuring station since the same model is trained for the entire dataset.

It is a common practice in forecasting to train a separate model for each forecast horizon. In our case, this means training 32 separate models, each of which focuses on forecasting a particular time horizon in the future. The models are calibrated considering the conformal framework for the forecasting horizon of interest.

To obtain reliable performance estimates, the algorithms were trained ten times, and the resulting tables report the mean values of MPIW and PICP. Given that transforming the forecasting problem into a tabular format generates approximately 43 million instances, the training process can be extremely slow. Therefore, to mitigate this issue, we used only 5% of the data. For this study, we used the default LightGBM parameters without optimizing the hyperparameters.

Several software tools were used in our study, such as the *numpy* [7], *darts* [5], *sklearn* [22], *MAPIE* [26], and *pandas* [20].

4.2 Probabilistic Forecasts With Different Horizons

Here we compare the effectiveness of QR and CQR in generating prediction intervals for different forecast horizons. Table 2 illustrates the comparative performance of QR and CQR for horizons between fifteen minutes and eight hours into the future (aggregated across all measuring stations). It can be seen that QR often produces prediction intervals that are too narrow and do not reach the desired 90% coverage threshold on the test dataset. For example, when predicting values for 8 hours ahead, the ordinary intervals from QR only encompass the true value 83.9% of the time, raising concerns about the effectiveness of the model even when trained to produce intervals with a 90% coverage rate. In contrast, CQR produces prediction intervals closer to the 90% requirement and is therefore better calibrated. The discrepancy between QR and CQR is more pronounced at horizons greater than three hours, as uncalibrated prediction intervals are far from the required 90% coverage. The importance of using calibrated models that can produce accurately calibrated prediction intervals is emphasized, as many QR models overestimate their confidence level and can produce narrower prediction intervals. On the other hand, CQR can adjust prediction intervals to be much closer to the desired coverage even for previously unseen data.

4.3 Forecasting Accuracy for different stations

It is important not only to compare QR and CQR for different horizons, but also to evaluate the accuracy of their predictions for different measuring stations. Figure 1 illustrates a comparison of prediction interval calibration for different measuring stations. It can be observed that QR produces prediction intervals that are often too narrow for some stations. In extreme cases, the resulting intervals only achieve coverage of about 40% instead of the intended 90%. Calibrated models generate prediction intervals that are much closer to the specified 90% compared to uncalibrated models. This highlights the main advantage of the conformal prediction framework, where prediction intervals can be adjusted to produce better prediction intervals. On the other hand, the drawbacks of a conformalized prediction framework are also evident from the above figure. Since CQR uses only a single global constant value to adjust the initial prediction interval, the intervals can often be adjusted even if it is not necessary. This can be seen in Figure 1, where some QR intervals for individual stations are closer to 90% compared to CQR prediction intervals. In summary, although the prediction intervals obtained with CQR for different time horizons achieve the desired 90% conditional coverage, individual measurement stations may exceed or fall below the 90% desired coverage threshold.

4.4 Example of Forecasts With Prediction Intervals

For clarity, we give examples of forecasts and their prediction intervals. Figure 2 shows three examples of prediction intervals obtained with regular QR and CQR. While conclusions from a few hand examples can be misleading, the selected examples show that both methods produce similar prediction intervals, with CQR generally having slightly larger prediction intervals compared to QR. In addition, CQR by default can produce prediction intervals that are negative. Since traffic flow cannot be negative, such values can simply be set to zero. In our case, we intentionally do not perform such a correction, even though this would narrow the prediction intervals of the CQR in some cases.

Table 2: Performance comparison of the MPIW and PICP metrics between QR and CQR over different forecasting horizons. Note that PICP should be close to the 90% specified coverage.

Horizon min	QR-PICP [%]	CQR-PICP [%]	QR-MPIW	CQR-MPIW
15	0.842214	0.898796	28.955776	31.622821
30	0.841815	0.897364	31.433076	33.867398
45	0.841924	0.898461	33.511282	35.367419
60	0.839802	0.899384	34.796558	36.555611
75	0.840788	0.897814	36.009832	37.244738
90	0.840183	0.898972	37.118646	38.200107
105	0.842494	0.898627	37.156039	38.544909
120	0.842322	0.899264	37.578008	38.580316
135	0.842541	0.900107	36.834235	38.046392
150	0.840340	0.899556	38.856038	39.960259
165	0.839329	0.898564	39.835607	41.228992
180	0.841134	0.898303	40.948913	42.052492
195	0.838409	0.899326	41.967217	42.754069
210	0.839587	0.898832	43.130416	43.604264
225	0.839139	0.899784	43.405047	44.511488
240	0.839413	0.899970	43.847630	44.854851
255	0.838934	0.900270	43.710950	44.887580
270	0.839090	0.899911	44.157199	45.124140
285	0.838359	0.899595	43.715106	44.945387
300	0.836904	0.900029	43.214179	44.875615
315	0.835988	0.900483	42.233077	44.183448
330	0.839534	0.901006	42.033808	44.011103
345	0.839875	0.901768	40.968346	43.547498
360	0.838058	0.900614	40.497423	42.387177
375	0.840325	0.900578	38.615834	40.822243
390	0.839126	0.900392	40.283268	42.346649
405	0.838463	0.900279	40.962669	43.716520
420	0.841016	0.900265	41.647127	44.431421
435	0.840530	0.900136	42.330318	44.767469
450	0.839703	0.900363	43.539172	45.840313
465	0.837991	0.899892	44.291754	46.519851
480	0.839280	0.900267	44.413916	46.800569

5 CONCLUSIONS

This research shows how CQR can be used to produce traffic forecasts with an associated uncertainty range. Furthermore, we demonstrate how the conformal prediction framework allows for creating well-calibrated forecasts. Accurate prediction intervals are particularly important in decision-making processes, where the consequences of incorrect predictions can be significant.

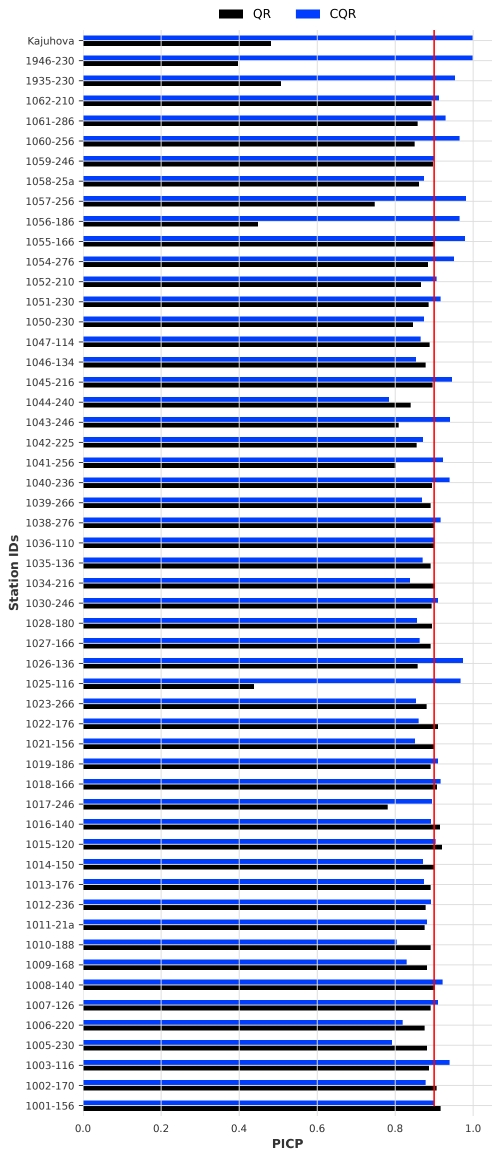

Figure 1: PICP for different traffic measuring stations obtained with regular QR and CQR. Only stations that appear in the test set are shown.

Our future work will involve testing a variety of machine learning models that can more effectively use the structure of the road network and capture time dependence without having to convert the data into a tabular format. We also plan to conduct more extensive feature engineering to improve the accuracy and reliability of our predictions. This includes exploring different feature selection methods, incorporating additional domain-specific features, and experimenting with feature engineering techniques that can capture the spatio-temporal dynamics of traffic patterns.

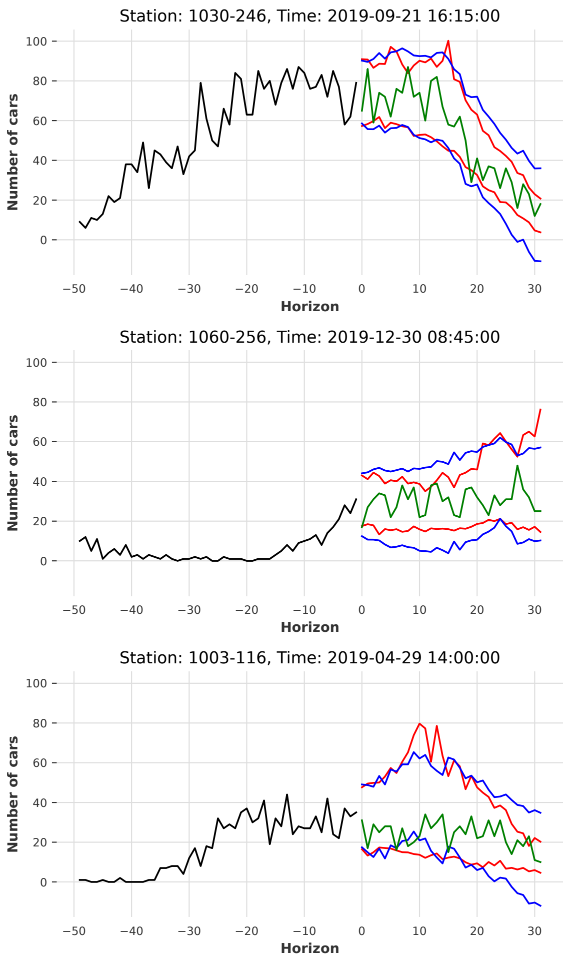

Figure 2: Example of prediction intervals for QR (red) and CQR (blue) for three different stations together with the actual traffic flows (green).

ACKNOWLEDGMENT

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P2-0098 and young researcher grants). The work is also part of a project that is funded by the European Union from Horizon Europe under grant agreement No 101077049 (CONDUCTOR).

REFERENCES

[1] Anastasios N Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

[2] Candice Bentéjac, Anna Csörgő, and Gonzalo Martínez-Muñoz. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54:1937–1967, 2021.

[3] Giovanni Buroni, Gianluca Bontempi, and Karl Determe. A tutorial on network-wide multi-horizon traffic forecasting with deep learning. In *EDBT/ICDT Workshops*, 2021.

[4] Alireza Ermagun and David Levinson. Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews*, 38(6):786–814, 2018.

[5] Julien Herzen etc. all. Darts: User-friendly modern machine learning for time series. *Journal of Machine Learning Research*, 23(124):1–6, 2022.

[6] Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

[7] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[8] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Global models for time series forecasting: A simulation study. *Pattern Recognition*, 124:108441, 2022.

[9] Jianming Hu, Qingxi Luo, Jingwei Tang, Jiani Heng, and Yuwen Deng. Conformalized temporal convolutional quantile regression networks for wind power interval forecasting. *Energy*, 248:123497, 2022.

[10] Vilde Jensen, Filippo Maria Bianchi, and Stian Normann Anfinsen. Ensemble conformalized quantile regression for probabilistic time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[11] Weiwei Jiang and Jiayun Luo. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, page 117921, 2022.

[12] Matthew G Karlaftis and Eleni I Vlahogianni. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3):387–399, 2011.

[13] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[14] Abbas Khosravi, Saeid Nahavandi, Doug Creighton, and Amir F Atiya. Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE transactions on neural networks*, 22(3):337–346, 2010.

[15] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[16] Diamantis Koutsandreas, Evangelos Spiliotis, Fotios Petropoulos, and Vassilios Assimakopoulos. On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, 73(5):937–954, 2022.

[17] Chaoshun Li, Geng Tang, Xiaoming Xue, Adnan Saeed, and Xin Hu. Short-term wind speed interval prediction based on ensemble gru model. *IEEE transactions on sustainable energy*, 11(3):1370–1380, 2019.

[18] Rumana Majid and Shakeel Ahmad Mir. Advances in statistical forecasting methods: An overview. *Economic Affairs*, 63(4):815–831, 2018.

[19] Osama Mohammed and Jalil Kianfar. A machine learning approach to short-term traffic flow prediction: A case study of interstate 64 in missouri. In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–7. IEEE, 2018.

[20] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[21] Dmitry Pavlyuk. Short-term traffic forecasting using multivariate autoregressive models. *Procedia Engineering*, 178:57–66, 2017.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.

[24] Durga L Shrestha and Dimitri P Solomatine. Machine learning approaches for estimation of prediction interval for the model output. *Neural networks*, 19(2):225–235, 2006.

[25] Ravid Shwartz-Ziv and Amitai Armon. Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90, 2022.

[26] Vianney Taquet, Vincent Blot, Thomas Morzadec, Louis Lacombe, and Nicolas Brunel. Mapie: an open-source library for distribution-free uncertainty quantification. *arXiv preprint arXiv:2207.12274*, 2022.

[27] Hristos Tyralis and Georgia Papacharalampous. A review of probabilistic forecasting and prediction with machine learning. *arXiv preprint arXiv:2209.08307*, 2022.

[28] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.

[29] Ying Wu, Yongchao Ye, Adnan Zeb, James Jianqiao Yu, and Zheng Wang. Adaptive modeling of uncertainties for traffic forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 2023.