SHORT-PAPER

Probabilistic Forecast of EV Charging Demand using Quantile Regression and LSTM with Attention Mechanism

SILVANA MATRONE, Politecnico di Milano, Milan, MI, Italy

AMIRHOSSEIN HEYDARIAN ARDAKANI, University of Twente, Enschede, Overijssel, Netherlands

EMANUELE OGLIARI, Politecnico di Milano, Milan, MI, Italy

ELHAM SHIRAZI, University of Twente, Enschede, Overijssel, Netherlands

SONIA LEVA, Politecnico di Milano, Milan, MI, Italy

Published: 16 June 2025

Citation in BibTeX format

E-Energy '25: The 16th ACM International Conference on Future and Sustainable Energy Systems
June 17 - 20, 2025
Rotterdam, Netherlands

Conference Sponsors:
SIGENERGY

E-Energy '25: Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems (June 2025)
https://doi.org/10.1145/3679240.3734687
ISBN: 9798400711251

Probabilistic Forecast of EV Charging Demand using Quantile Regression and LSTM with Attention Mechanism

Silvana Matrone
Politecnico di Milano
milano, Italy
silvana.matrone@polimi.it

Amirhossein Heydarian Ardakani
University of Twente
Enschede, Netherlands
a.heydarian@utwente.nl

Emanuele Ogliari
Politecnico di Milano
milano, Italy
emanuelegiovanni.ogliari@polimi.it

Elham Shirazi
University of Twente
Enschede, Netherlands
e.shirazi@utwente.nl

Sonia Leva
Politecnico di Milano
milano, Italy
sonia.leva@polimi.it

Abstract

The electrification of transportation has significantly increased electric vehicle (EV) charging demand on energy systems. Accurately capturing the uncertainty of future EV loads is essential for flexible system operation. This study proposes an LSTM-attention model for probabilistic EV load forecasting, featuring an encoder-decoder architecture with an intermediate attention layer. Using Quantile Regression (QR), the model predicts upper, median, and lower load quantiles. Evaluation is performed on parking station data from the SmoothEMS Met GridShield project in the Netherlands.

CCS Concepts

• **Computing methodologies → Supervised learning by regression**.

Keywords

Electric Vehicles, EV Load Forecast, Long Short-Term Memory, Probabilistic Forecast, Quantile Regression

ACM Reference Format:
Silvana Matrone, Amirhossein Heydarian Ardakani, Emanuele Ogliari, Elham Shirazi, and Sonia Leva. 2025. Probabilistic Forecast of EV Charging Demand using Quantile Regression and LSTM with Attention Mechanism. In *The 16th ACM International Conference on Future and Sustainable Energy Systems (E-ENERGY '25), June 17–20, 2025, Rotterdam, Netherlands.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3679240.3734687

1 Introduction

Electric load forecasting has become increasingly important due to the rapid growth of the global electric vehicle (EV) fleet. As power demand from charging stations rises, accurate EV charging load forecasting is essential for balancing electricity demand and power generation, ensuring grid stability. Many forecast models have been proposed in the literature [6]. In previous work, we introduced an LSTM with an attention layer model [4].

A key challenge in managing distributed energy systems like microgrids is handling uncertainty. In this context, point forecasting alone may be inadequate. Only a few studies have explored probabilistic forecasting for EV charging demand, which offers insights into uncertainties via distribution estimates [2], quantile estimations [1], or Bayesian techniques [5]. This study applies a Quantile Regression (QR) framework to the EV demand forecasting method from [4], using the Dutch ASR dataset.

2 Methodology

The point forecast of EV charging is performed by an LSTM encoder–decoder with attention. To incorporate probabilistic forecasting, QR is applied.

LSTM-Attention model. The Encoder is composed of stacked LSTM layers, it captures its essential features into a lower dimensional representation, namely the last hidden state of the last LSTM layer. This becomes the Context Vector that is fed to the Decoder to be interpreted. The Decoder then generates predictions for each required time step. The Attention mechanism dynamically refines the context vector by computing a weighted sum of all hidden states. The alignment scores determine these weights by assessing the relevance between the previous output and the current hidden state. The strucuture of the model is shown in Figure 1.

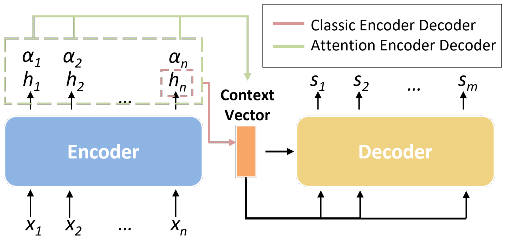


Figure 1: Encoder Decoder with Attention stucuture of the forecasting model.

Quantile Regression. QR is a technique used to estimate conditional quantiles of a target variable. To apply this approach, we trained three different LSTM-Attention models: one for the median (50th quantile), one for the 20th quantile, and one for the 80th quantile. The model predicts specific quantiles by minimizing a quantile-dependent loss function known as the pinball loss function, defined as follows:

E-ENERGY '25, Rotterdam, Netherlands

ACM ISBN 979-8-4007-1125-1/25/06
https://doi.org/10.1145/3679240.3734687

$$L_\tau(y, \hat{y}) = \begin{cases} \tau(y - \hat{y}) & \text{if } y \geq \hat{y} \\ (1-\tau)(\hat{y} - y) & \text{if } y < \hat{y} \end{cases} \tag{1}$$

where y is the measured value at a given timestep and $\hat{y}$ is the corresponding prediction. This function is asymmetric, meaning that when the quantile value τ is between 0 and 0.5, underforecasting is penalized less (to predict a lower quantile), whereas for values between 0.5 and 1, overforecasting receives a lower penalty (to predict a higher quantile). If τ is 0, the pinball loss function simplifies to the absolute error.

Evaluation metrics. The classic metrics used for point forecast models are not effective when dealing with probabilistic ones. The Continuous Ranked Probability Score (CRPS) is a proper scoring rule used to evaluate the accuracy of probabilistic forecasts. It measures the difference between the predicted cumulative distribution function (CDF) and the actual observation. The CRPS formula is shown in (2) for a given forecast $\hat{F}(y)$ and true observation y.

$$\text{CRPS}(\hat{F}, y) = \sum_{i=1}^{N} \left(\hat{F}(y_i) - \mathbf{1}\{y_i \geq y\} \right)^2 \Delta y_i, \tag{2}$$

where $\hat{F}(y_i)$ is the cumulative probability of the forecast up to the value y_i, y_i is the set of possible forecast values, y is the observed value, $\mathbf{1}\{y_i \geq y\}$ is the Heaviside step function, which is 1 if $y_i \geq y$ and 0 otherwise. Prediction Interval Coverage Probability (PICP) is used for coverage measurement. It is the percentage of true values that fall within the predicted uncertainty interval.

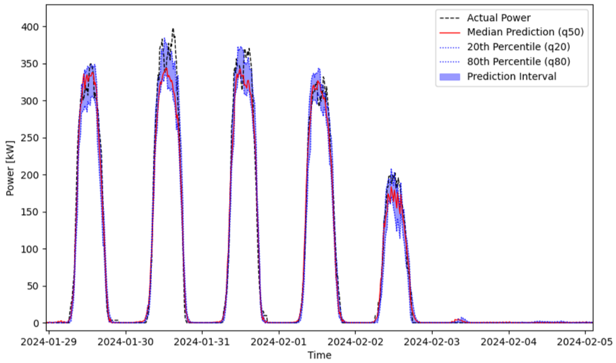
3 Case Study

The dataset utilized in this study is sourced from the SmoothEMS Met GridShield project in the Netherlands, collected from the ASR office parking lot in Utrecht, while weather data are from KNMI [3]. The features extracted to provide as input are month, day, hour, minute, temperature, global radiation, number of active sessions. The dataset has 15-minute discretization, the model is trained on 80% of the data and tested on the remaining 20% using 96 input steps (previous day) to predict the next day with no lag. The forecast is performed with a rolling horizon once every hour and provides a 24-hour forecast horizon.

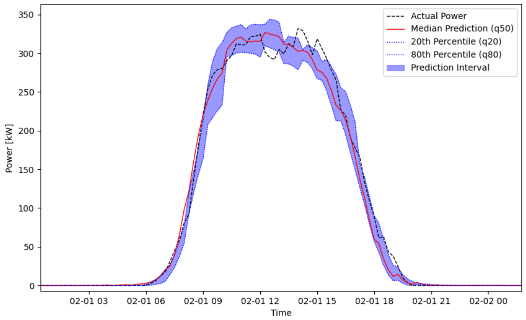
4 Results and Discussion

The proposed model is evaluated with different hyperparameter configurations, including different numbers of LSTM layers in both Encoder and Decoder (1 and 2 layers) and different number of units in each layer (24, 48, and 64). As shown in Table 1, there is a trade-off between model complexity and forecasting performance. Models with 48 and 64 units generally outperform those with 24 units, especially for the point forecast. Model 48-2 achieves the best overall performance, with the lowest CRPS (6.111), competitive pinball losses across the two quantiles and highest PICP.

To further assess the model performance, we consider a benchmark based on naive persistence and a fixed-width quantile prediction. The benchmark used to calculate the median prediction is the weekly persistence model, which assumes that the forecast for each time step is the same as the corresponding value from the previous week. It reflects a simplistic assumption that power consumption patterns repeat weekly. On top of that, the benchmark uses fixed-width (equal to 10% of the persistence prediction) intervals around it for the forecast of the 20th and 80th quantiles, providing a constant prediction range. In comparison to the benchmark, the 48-2 model outperformed it with significant improvements: Pinball Loss (q20) decreased by 81.39%, Pinball Loss (q50) improved by 62.09%, Pinball Loss (q80) showed a 68.04% reduction, the CRPS (mean) improved by 67.70% and PICP increased of 26.82% points. These results indicate that the 48-2 model provides a substantial enhancement over the benchmark in forecasting accuracy. Future work could explore the inclusion of additional input lags and experiment with extending the forecasting horizon to assess the model's performance over longer timeframes.



(a) EV load forecasts for a sample week.



(b) EV load forecasts for a sample day.

Figure 2: Probabilistic forecast with LSTM-attention model.

Table 1: Results from simulation on different hyperparameters configurations

Model	Pinball loss(50)	Pinball loss(20)	Pinball loss(80)	CRPS	PICP
24-1	1.730	3.870	4.028	6.754	61.00%
24-2	1.851	4.210	3.487	6.874	62.05%
48-1	1.494	4.095	2.800	6.197	72.12%
48-2	**1.486**	**4.020**	**2.847**	**6.111**	**78.60%**
64-1	1.715	3.821	3.523	6.532	73.58%
64-2	1.570	3.906	2.812	6.217	66.02%
benchmark	10.606	7.986	8.885	18.918	51.78%

References

[1] Julian Huber, David Dann, and Christof Weinhardt. 2020. Probabilistic forecasts of time and energy flexibility in battery electric vehicle charging. *Applied Energy* 262 (March 2020).

[2] Md Shariful Islam, Nadarajah Mithulananthan, and Duong Quoc Hung. 2018. A Day-Ahead Forecasting Model for Probabilistic EV Charging Loads at Business Premises. *IEEE Transactions on Sustainable Energy* 9 (2018).

[3] Royal Netherlands Meteorological Institute (KNMI). [n. d.]. Daggegevens van het weer in Nederland. https://www.knmi.nl/nederland-nu/klimatologie/daggegevens Accessed: 2025-04-07.

[4] Silvana Matrone, Emanuele Ogliari, Alfredo Nespoli, and Sonia Leva. 2024. Electric vehicle supply equipment day-ahead power forecast based on deep learning and the attention mechanism. *IEEE Transactions on Intelligent Transportation Systems* (2024).

[5] Dan Zhou, Zhonghao Guo, Yuzhe Xie, Yuheng Hu, Da Jiang, Yibin Feng, and Dong Liu. 2022. Using Bayesian Deep Learning for Electric Vehicle Charging Station Load Forecasting. *Energies* 15, 17 (2022).

[6] Juncheng Zhu, Zhile Yang, Monjur Mourshed, Yuanjun Guo, Yimin Zhou, Yan Chang, Yanjie Wei, and Shengzhong Feng. 2019. Electric Vehicle Charging Load Forecasting: A Comparative Study of Deep Learning Approaches. *Energies* 12, 14 (2019).