Journal of Physics:
Conference Series

Multi-region Probabilistic Load Forecasting with Graph Bayesian Transformer Network

Multi-region Probabilistic Load Forecasting with Graph Bayesian Transformer Network

Shiqi Zhou[1], Junfan Wang[1], Zhekang Dong[1*], Xiaoyue Ji[2], Yujie Yuan[3] and Chun Sing Lai [4]

[1] School of Electronics and Information, Hangzhou Dianzi University, Hangzhou, China
[2] Department of Precision Instrument, Tsinghua University, Beijing, China
[3] Department of Air Traffic Management, Civil Aviation University of China, Tianjin, China
[4] Department of Electronic and Electrical Engineering, Brunel University of London, UK

*E-mail: englishp@hdu.edu.cn

Abstract. Accurate probabilistic load forecasting is essential for efficient energy management and the safety operation of power system. Existing load forecasting methods suffer from two limitations: 1) Inadequate utilization of feature; 2) insufficient modelling capability for fine-grained dependencies. To end these problems, a multi-region probabilistic load forecasting method based on graph Bayesian Transformer network is proposed. Specifically, the proposed forecasting framework consists of graph neural network and hybrid Bayesian Transformer connected in cascaded configuration. The former one is used to develop multi-graph spatial-temporal features, which can enhance the feature learning ability and share the graph structure information to realize the joint forecasting of multi-region. The latter one is used to capture multi-scale information, which can improve the adaptability of model to complex dynamic data and forecasting accuracy. For validation, a series of compared experiments and ablation analysis are conducted under New England dataset. The experimental results demonstrate that the proposed method has good performance in foresting accuracy, and adaptability. In particular, compared to other comparative methods, the Continuous Ranked Probability Score (CRPS) is reduced 32.7%.

1. Introduction

As industrialization continues, electricity demand in numerous countries has risen significantly. As a result of the imbalance between energy supply and demand, the world is experiencing an unprecedented energy crisis [1-3]. Consequently, efficient and rational utilization of power resources has become a focal point for nations worldwide [4,5]. Probabilistic load forecasting, by addressing the inherent uncertainties in load distribution, enables power system operators to make risk-informed decisions in areas such as economic dispatch and steady-state estimation of transmission networks [6-9]. Accurate probabilistic load forecasting is essential for optimizing grid dispatch and promoting the rational use of power resources [10,11].

C. Li *et al.* [12] proposed an interpretable Long Short-Term Memory (LSTM) network for probabilistic residential load forecasting within a memristor-based neuromorphic computing architecture. A. Faustine *et al.* [13] introduced parameterized depth quantile regression to achieve short-term probabilistic load forecasting. Álvarez *et al.* [14] proposed an adaptive probabilistic load forecasting methodology based on Hidden Markov model to capture load demand uncertainty and dynamics. Bayesian neural networks are extensively applied for probabilistic forecasting, as they capture both epistemic and aleatoric uncertainties [15-18]. Alessandro *et al.* [19] proposed an approach framed on Bayesian Mixture Density Networks, which enhance neural network mapping by integrating predictive distributions that address both uncertainty types. C. Wang *et al.* [20] proposed a multi-region Bayesian neural network, Bayesian Multiple-Decoder Transformer (BMDeT), to achieve probabilistic forecasting of multi-energy loads, accounting for their complex interdependencies and uncertainties. However, there still some difficulties and challenges:

1) **Inadequate utilization of feature:** Electricity load typically exhibits spatial dependency and load transfer effects. Utilizing features from a single region fails to fully leverage the spatio-temporal information across regions, resulting in inadequate utilization of features.

2) **Insufficient modelling capability for fine-grained dependencies:** Due to the complex spatial heterogeneity, the insufficient modeling of fine-grained dependencies leads to local properties that are then difficult to capture adequately on individual regions.

Based on this, a multi-dimensional Bayesian Autoformer network for short-term load probabilistic forecasting is proposed. The contributions are summarized below:

- Graph neural network (GNN) is proposed to develop spatial-temporal features, which can enhance the feature learning ability and share the graph structure information to realize the joint forecasting of multi-regions.
- Hybrid Bayesian attention is proposed to capture multi-scale information, which can more acutely identify short-term, neighborhood, and fine-grained spatiotemporal patterns and quantify electricity load uncertainty more accurately.
- The proposed model was subjected to a series of experimental analysis (compared analysis, ablation analysis, adaptive analysis) on real load datasets from multiple regions. The experimental results demonstrate that the proposed method has good performance in foresting accuracy, and adaptability.

The remainder of this work is organized as follows: Section 2 introduces the proposed model and the specific design. A series of comparison experiments and joint analysis (ablation analysis and adaptive analysis) are conducted in Section 3. Finally, Section 4 draws the conclusion.

2. Proposed Methodology

In this paper, we propose a multi-region probabilistic forecasting model based on Graph Bayesian Transformer network. The model framework is shown in Figure 1, it includes graph structure learning, graph Bayesian hybrid Transformer and performance optimize. The following is a further description of the model.

2.1 Graph structure learning

In electricity load forecasting, each node represents the electrical load of a region, while the edges of each node represent the connections between regions (proximity relationships). Based on this, the directed graph structure G can be represented as

$$G=(V,E) \tag{1}$$

where V denotes the set of nodes (regions). E denotes the set of edges, i.e., the lines connecting the regions. Adjacency matrix $A\in\mathbf{R}^{N\times N}$ describes the connectivity relationship between each region.

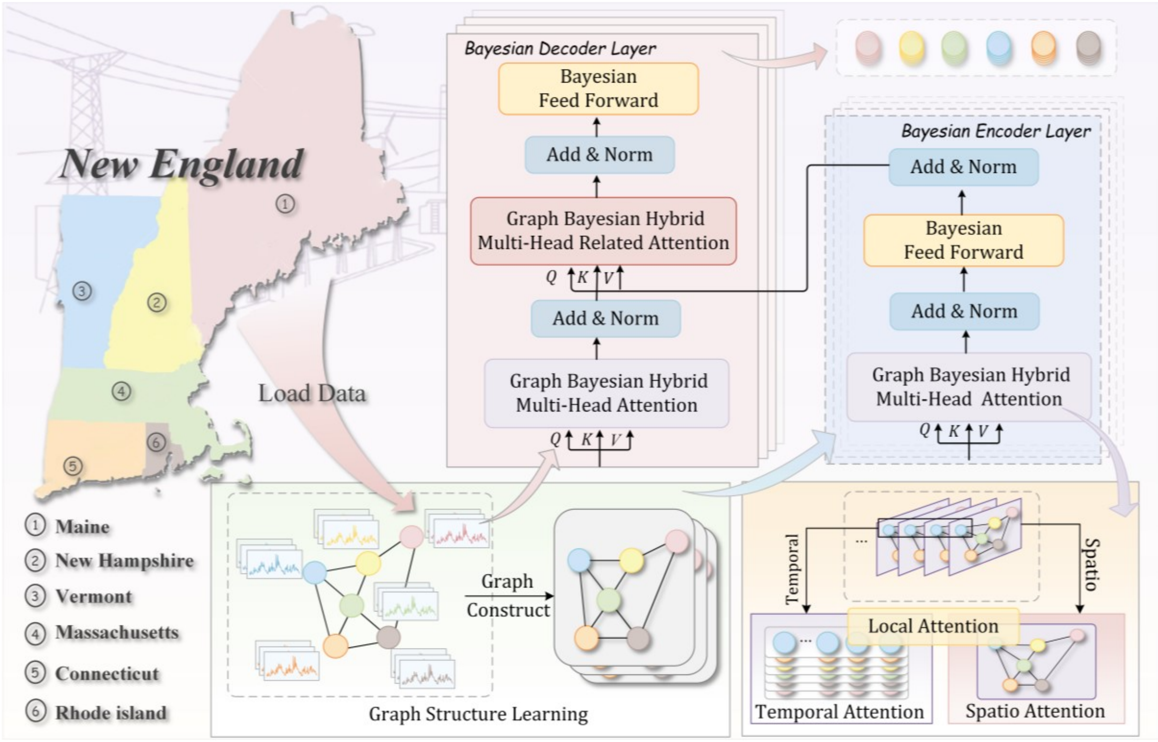


Figure 1. The framework of the proposed multi-region probabilistic forecasting model

The feature matrix $X \in \mathbf{R}^{N \times F}$ contains F-dimensional historical load data of each node. Then, the update formula is as follows:

$$H^{(l+1)} = \sigma\left(AH^{(l)}W^{(l)}\right) \tag{2}$$

where $H^{(l)} \in \mathbf{R}^{N \times D_l}$ is the node identity matrix of the l_{th} layer and $H^{(0)}=X$ is the initial input feature. $\hat{A}$ is the normalized matrix of the adjacency matrix A after adding the identity matrix I. $W^{(l)} \in \mathbf{R}^{D_l \times D_{l+1}}$ is the learnable weight matrix of the l_{th} layer. σ (·) is a nonlinear activation function.

2.2 Graph Bayesian Hybrid Transformer Network

Hybrid Bayesian Transformer Network consists of an encoder module, a decoder module. The main task of the encoder is to extract and compress the multi-scale features of the input data. It efficiently captures the temporal trends and interdependencies between regions. The task of the decoder is to generate probabilistic forecasts based on the embedding of the encoder outputs.

Hybrid Bayesian Encoder: The Bayesian encoder consists of four encoder layers, including Graph Bayesian Hybrid Multi-Head Attention Layer, Bayesian Feed Forward Layer and Normalization Layer. Graph Bayesian Hybrid Multi-Head Attention Layer includes hybrid attention, it can be represented by

$$H_{hybrid} = \alpha H_t + \beta H_s + \gamma H_l \tag{3}$$

$$H_t = Attention_{temproal}\left(Q_t W_i^{Q_t}, K_t W_i^{K_t}, V_t W_i^{V_t}\right) = \operatorname{soft}\max\left(\frac{Q_t W_i^{Q_t} K_t^T W_i^{K_t}}{\sqrt{d_k}}\right) V_t W_i^{V_t} \tag{4}$$

$$H_s = Attention_{spatia}\left(Q_s W_i^{Q_s}, K_s W_i^{K_s}, V_s W_i^{V_s}\right) = \operatorname{soft}\max\left(\frac{Q_s W_i^{Q_s} K_s^T W_i^{K_s}}{\sqrt{d_k}} + A\right) V_s W_i^{V_s} \tag{5}$$

$$H_l = Attention_{local}\left(Q_c W_i^{Q_l}, K_c W_i^{K_c}, V_c W_i^{V_c}\right) = \mathrm{soft}\max\left(\frac{Q_c W_i^{Q_c} K_c^T W_{i-\omega:i+\omega}^{K_c}}{\sqrt{d_k}}\right) V_c W_{i-\omega:i+\omega}^{V_c} \tag{6}$$

where α and β denote the weight of learnable. W_i^{Qt}, W_i^{Kt}, and W_i^{Vt} represent the weights of Q_t, K_t, and V_t, respectively. Similarly, W_i^{Qs}, W_i^{Ks}, W_i^{Vs}, W_c^{Qc}, W_c^{Kc}, and W_c^{Vc} denote the weights of Q_s, K_s, V_s, Q_c, K_c, and V_c. d_k denotes scaling coefficient, and A denotes the adjacency matrix. ω denotes the size of sliding windows. Then, we utilize multiple hybrid attention mechanisms to parallelize the computational:

$$MultiHead(Q,K,V) = W_o \cdot Concat(Head_1, \ldots, Head_2) \tag{7}$$

$$Head_i = H_{hybrid}^i \tag{8}$$

where W_o is the weight of *Concat* output. Bayesian neural networks evaluate uncertainty by sampling from a probability distribution to obtain weights and biases, such as Gaussian distribution. Weights, including W_i^{Qt}, W_i^{Kt}, W_i^{Vt}, W_i^{Qs}, W_i^{Ks}, W_i^{Vs}, W_c^{Qc}, W_c^{Kc}, W_c^{Vc} and W_o, are taken by sampling from the Gaussian distribution. Similarly, the parameters of Bayesian Feed Forward Layer are sampled from the Gaussian distributions.

Hybrid Bayesian Decoder: The Bayesian encoder includes Graph Bayesian Hybrid Multi-Head Attention Layer, Graph Bayesian Hybrid Multi-Head Related Attention Layer, Bayesian Feed Forward Layer and Normalization Layer. Among them, the input Q and K of the Related Attention Layer are outputs from the encoder, while V is the output from the Attention Layer in decoder. Bayesian decoder combines known information about future time steps with spatio-temporal features of the encoder outputs and generates a probability distribution of the load forecast.

2.3 Performance optimize

During the training process, model performance is quantified and optimized in multiple dimensions such as accuracy of predicted values, coverage of interquartile range, and matching of prediction distributions. This multi-loss setting helps the model to generate more reliable predictions that not only give single-point predictions, but also quantify uncertainty through probability distributions, making the predictions richer and more robust. The overall loss function for the training model is obtained by:

$$Loss = a_1 Loss1 + a_2 Loss2 + a_3 Loss3 \tag{9}$$

$$Loss1 = \mathrm{RMSE}(\hat{y}, y) \tag{10}$$

$$Loss2 = pinball(\hat{y}, y) \tag{11}$$

$$Loss3 = CRPS(\hat{y}, y) \tag{12}$$

where parameters a_1, a_2, a_3 represent the weight assignments for *Loss1*, *Loss2*, *Loss3*.

To optimize the model and achieve probabilistic forecasting, variational inference is employed to enable effective Bayesian inference [21]. Specially, inference task is addressed by minimizing the Kullback–Leibler (KL) divergence from the latent posterior, formulated as:

$$\begin{aligned} \mathrm{KL}\left[q(\omega|\theta) \,||\, P(\omega|D)\right] &= \int q(\omega|\theta) \log\left(\frac{q(\omega|\theta)}{P(\omega)P(D|\omega)}\right) d\omega \\ &= -\mathrm{E}_q\left[\log\frac{q(\omega|\theta)}{P(\omega)} - \log P(D|\omega)\right] + \log P(D) \end{aligned} \tag{13}$$

where $q(\omega|\theta)$ represents the variational distribution of parameter θ, $P(\omega|D)$ is the posterior distribution approximated by $q(\omega|\theta)$. In equation (13), the second term is constant with respect

Journal of Physics: Conference Series **3001** (2025) 012013 doi:10.1088/1742-6596/3001/1/012013

to the network parameters. Therefore, the primary component requiring optimization through backpropagation is the Evidence Lower Bound (ELBO):

$$\mathrm{ELBO}=\mathrm{E}_q\left[\log\left(q(\omega|\theta)-P(\omega)\right)-\log P\left(D|\omega\right)\right]=KL\left(q(\omega|\theta)||P(\omega)\right)-\mathrm{E}_q\left[\log P\left(D|\omega\right)\right] \quad (14)$$

where $\mathrm{E}_q[\log P(D|\omega)]$ is reconstruction loss, we better adapt $\mathcal{L}$ for the probabilistic forecasting model, we further customize it by jointly minimizing the multi-loss:

$$\mathcal{L}=-\mathrm{E}[Loss]+KL\left(q(\theta)||p(\theta)\right) \quad (15)$$

3. Experiment and analysis

3.1 Dataset

The dataset chosen for the experiment was collected by New England Independent System Operator and is publicly available on the web https://www.iso-ne.com/. It is a real electricity load dataset that contains electric load data for the New England region of the northeastern U.S., including historical load profiles for six states (e.g. Miane, New Hampshire, Vermont, Massachusetts, Connecticut, and Rhode Island). It records the demand for electricity over time for each region at an hourly frequency.

3.2 Evaluation Indicators

In this paper, the three evaluation indicators are used to evaluate the performance of the forecasting model: Root Mean Square Error (RMSE), pinball loss, Continuous Ranked Probability Score (CRPS), the special formula is as follows [22-23]:

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\mathrm{y}_t-\hat{\mathrm{y}}_t)^2} \quad (16)$$

$$pinball\_loss(q,y_t)=\begin{cases} q\left(y_t-c_t^{(q)}\right) & y_t\geq\hat{y}_t^{(q)} \\ 1-q\left(\hat{y}_t^{(q)}-y_t\right) & y_t<\hat{y}_t^{(q)} \end{cases} \quad (17)$$

$$CRPS(F_t,\mathrm{y}_t)=\sum_{i=1}^{N}\left(F_t\left(\tau_i\right)-\varepsilon\left\{\tau_i\geq\mathrm{y}_t^{(i)}\right\}\right)^2 \quad (18)$$

where N is the number of prediction data, y_t is the actual value and $\hat{y}_t$ is the predicted value. q is quartile, $\hat{y}_t^q$ is the estimate when the quartile is q at time t.

Table 1. Parameter setting

Model	Parameters	Model	Parameters
QLSTM [12]	LSTM unit: 16 FC unit: 16 layer: 4	BNN [19]	FC-1 unit: 512 layer:1 FC-2 unit: 256 layer:4
APLF [14]	Forgetting factors 1: 0.2 Forgetting factors 2: 0.7	BMDeT [20]	Encoder layer: 4 Decoder layer: 4 Decoder multi-head: 8 model dimension: 24
GNN [26]	Hidden Dimension:24 Convolution layer: 4 Edge dimension: 7 Node dimension: 48	**Proposed Model**	Encoder layer: 4 Decoder layer: 4 Decoder multi-head: 8 model dimension: 24

Journal of Physics: Conference Series **3001** (2025) 012013 doi:10.1088/1742-6596/3001/1/012013

3.3 Experimental Environment and Parameter Settings

The proposed model is trained and tested on a server equipped with NVIDIA GeForce RTX 4090 GPUs and Programming language Python 3.10. Suitable parameters facilitate good performance of the model. Parameters of experienced models are determined by grid search method [24, 25], and the parameter settings for all models are demonstrated in Table 1.

3.4. Comprised Experiment

The performance of the proposed model is compared with other maintains load forecasting methods, including QLSTM [12], APLF [14], GNN [26], BNN [19], BMDeT [20]. Then, the special comprised results as shown in Table 2. From Table 2, the proposed model performs the best in overall performance, the average RMSE, Pinball Loss and CRPS are 0.0559, 0.0234, and 0.0409, respectively. Specifically, compared with signal-region forecasting model (QLSTM, APLF), the performance of proposed model is more accurate and stable. Compared with GNN, which also builds graph structure features, the accuracy of the proposed model is improved by 14.2%, and GNN only enables point forecasting. It is shown that the proposed model is able to better learn graph structure features and realize multi-region joint forecasting. Compared with multi-regions probabilistic forecasting models (BNN, BsDeT), the proposed model has significant advantages, with average RMSE, Pinball loss, and CRPS improving by at least 10.8%, 12%, and 32.7%, respectively. It indicates that the proposed model is able to efficiently capture the coupling relationship between the proximity region and the local, realizing the accurate probabilistic forecasting of the electricity loads in multiple regions.

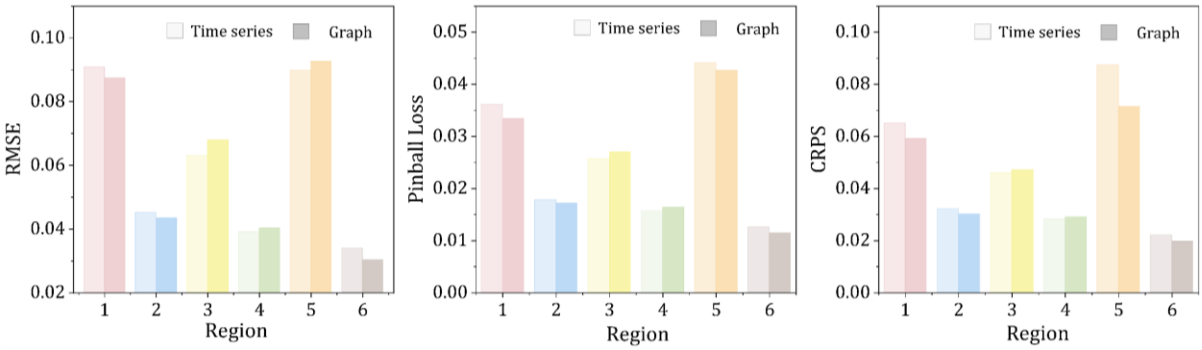
Table 2 Comprised experiment with different load forecasting methods

Region		①	②	③	④	⑤	⑥	**Average**
QLSTM	RMSE	0.6187	0.1173	0.4618	0.2178	0.2617	0.0890	0.2943
	Pinbal	0.2921	0.0509	0.2207	0.1029	0.1272	0.0361	0.1383
	CRPS	0.5694	0.0969	0.4295	0.2003	0.2503	0.0684	0.2691
APLF	RMSE	0.8307	0.2210	0.3844	0.2848	0.1577	0.0809	0.3265
	Pinbal	0.5554	0.0994	0.1112	0.1557	0.0528	0.0159	0.1650
	CRPS	0.6023	0.1729	0.2963	0.2128	0.1099	0.0431	0.2395
GNN*	RMSE	0.0988	0.0647	0.0779_2	0.0522	0.0937	0.0429	0.0717
	Pinbal	-	-	-	-	-	-	-
	CRPS	-	-	-	-	-	-	-
BNN*	RMSE	0.1908	0.1598	0.1690	0.1954	0.1755	0.1412	0.1719
	Pinbal	0.1135	0.1092	0.1167	0.1576	0.1158	0.0695	0.1137
	CRPS	0.1728	0.1325	0.1528	0.1778	0.1590	0.1144	0.1516
BMDeT*	RMSE	0.0924_2	0.0491_2	0.0819	0.0391_2	0.0849_2	0.0311_2	0.0690_2
	Pinbal	0.0374_2	0.0195_2	0.0352_2	0.0158_2	0.0403_2	0.0119_2	0.0266_2
	CRPS	0.0833_2	0.0708_2	0.0579_2	0.0595_2	$\mathbf{0.0611}_1$	0.0488_2	0.0636_2
Our	RMSE	$\mathbf{0.0800}_1$	$\mathbf{0.0409}_1$	$\mathbf{0.0627}_1$	$\mathbf{0.0380}_1$	$\mathbf{0.0846}_1$	$\mathbf{0.0292}_1$	$\mathbf{0.0615}_1$
	Pinbal	$\mathbf{0.0316}_1$	$\mathbf{0.0162}_1$	$\mathbf{0.0255}_1$	$\mathbf{0.0155}_1$	$\mathbf{0.0403}_1$	$\mathbf{0.0108}_1$	$\mathbf{0.0234}_1$
	CRPS	$\mathbf{0.0566}_1$	$\mathbf{0.0288}_1$	$\mathbf{0.0451}_1$	$\mathbf{0.0278}_1$	0.0683_2	$\mathbf{0.0188}_1$	$\mathbf{0.0428}_1$

Note: * represents multi-region forecasting model; Superscripts 1 and 2 denote the first and second rankings, respectively.

3.5. Ablation studies

Graph structure based GNN: The experiment result with different input feature as shown in Table 3. The input with the graph structure demonstrates superior overall performance than time series, indicating its effectiveness to enhance the feature learning ability and share the graph structure information to realize the joint forecasting of multi-regions.



(a) (b) (c)

Figure 2. Ablation experiment with graph structure based on GNN, (a) RMSE, (b) pinball loss, (c)CRPS

Graph Bayesian hybrid attention: As shown in Table 3, the proposed Graph Bayesian Hybrid Transformer Network demonstrates superior overall performance, indicating its effectiveness in capturing the spatiotemporal relationships of regional electricity load. For certain specific regions (e.g., Regions 4 and 5), the Spatio-temporal attention performs less effectively than the local-global attention model. This discrepancy may be attributed to significant differences in load patterns between these regions and others. The proposed Graph Bayesian Hybrid attention can alleviate the problem of regional variations and improve the accuracy of forecasting.

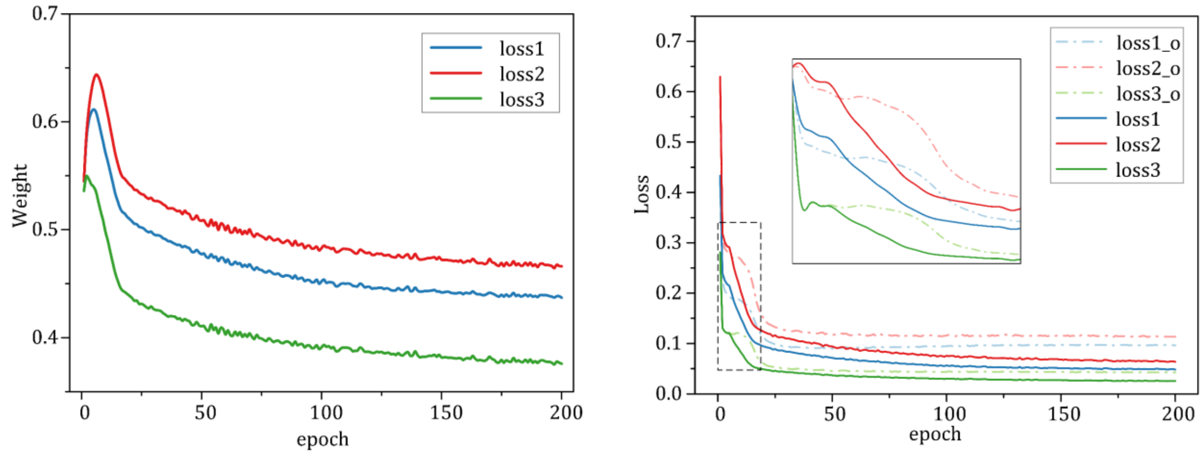
Table 3 Ablation experiment with Bayesian hybrid attention

Region		①	②	③	④	⑤	⑥	Average
Full	RMSE	0.1278	0.0745	0.0706	0.0648	0.0778^2	0.0490	0.0774
	Pinball	0.0530	0.0322	0.0285	0.0277	0.0364^2	0.0203	0.0331
	CRPS	0.0970	0.0608	0.0504	0.0516	0.0695^2	0.0369	0.0610
Auto	RMSE	0.1124	0.0540	0.0951	0.0389	0.0901	0.0333	0.0717
	Pinball	0.0472	0.0224	0.0412	0.0157	0.0431	0.0131	0.0305
	CRPS	0.0879	0.0423	0.0774	0.0285	0.0838	0.0233	0.0572
Mix	RMSE	0.0998	0.0539	0.0609^2	0.0429	**0.0713^1**	0.0308	0.0607
	Pinball	0.0410	0.0227	0.0233^2	0.0176	**0.0329^1**	0.0118	0.0249
	CRPS	0.0743	0.0420	0.0415^2	0.0322	**0.0632^1**	0.0205	0.0456
Spatio-temporal	RMSE	0.0888^2	0.0427^2	0.0676	0.0380^2	0.0856	0.0326^2	0.0599^2
	Pinball	0.0348^2	0.0166^2	0.0262	0.0170^2	0.0410	0.0126^2	0.0246^2
	CRPS	0.0634^2	0.0294^2	0.0473	0.0265^2	0.0694	0.0221^2	0.0446^2
Our	RMSE	**0.0800^1**	**0.0409^1**	**0.0579^1**	**0.0371^1**	0.0846	**0.0302^1**	**0.0574^1**
	Pinball	**0.0316^1**	**0.0159^1**	**0.0229^1**	**0.0147^1**	0.0403	**0.0112^1**	**0.0231^1**
	CRPS	**0.0566^1**	**0.0276^1**	**0.0397^1**	**0.0259^1**	0.0683	**0.0187^1**	**0.0417^1**

Note: Superscripts 1 and 2 denote the first and second rankings, respectively.

3.5. multi-loss balance method

To verify the effectiveness of the proposed multi-loss balance method, the experimental results as shown in Figure 3. Figure 3(a) illustrates the weight of loss during model training. It can be observed that each task begins with the same initial weight, the weights adaptively adjust as training progresses. Figure 3(b) shows Loss curves for models trained with and without the multi-loss balancing method. It can be seen that the proposed multi-loss balance method can effectively accelerate the training of multi-task models. The results indicate that the proposed method significantly accelerates model training by dynamically adjusting the weight ratios of each loss, enhancing the training efficiency.



(a) (b)

Figure 3 The effectiveness of the proposed multi-loss balance method. (a) the weight of loss during model training; (b) loss curves for models trained with and without the multi-loss balancing method.

4. Conclusion

In this work, we mainly focus on the investigation of multi-region probabilistic load forecasting. Specially, GNN is proposed to enhance the feature learning ability and share the graph structure information with multi-region. Then, hybrid Bayesian attention is proposed to achieve more acutely identify short-term, neighbourhood, and fine-grained spatiotemporal patterns and quantify electricity load uncertainty more accurately. Based on these, the spatial-temporal information is inadequate utilized, and the capability of fine-grained dependencies is capture. The experimental results and joint analysis demonstrate that the proposed model outperforms the existing mainstream methods in terms of accuracy and adaptability. In the future, datasets need to analyse for better understand pattern differences and focus on different forecasting dimensions.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grants (62206062), Yangtze River Delta Science and Technology Innovation Community Jointly Tackled Key Project (2023CSJGG1300), and Fundamental Research Funds for the Provincial University of Zhejiang under Grant GK229909299001-06.

References

[1] Wang, Zhi, et al. "Study on the multi-time scale rolling optimization operation of a near-zero energy building energy supply system." *Energy Conversion and Management*, 270 (2022): 116255.
[2] C. S. Lai et al., "Multi-View Neural Network Ensemble for Short and Mid-Term Load Forecasting," IEEE Transactions on Power Systems, 2021,36(4): 2992-3003.
[3] Dong Z. et al. "Periodic Segmentation Transformer-Based Internal Short Circuit Detection Method for Battery Packs." *IEEE Transactions on Transportation Electrification*, 2024.
[4] Zhou Y, Wang J, Li Z, et al. "Short-term photovoltaic power forecasting based on signal decomposition and machine learning optimization." Energy Conversion and Management, 2022, 267: 115944.
[5] Cao Z, Wang J, Xia Y. "Combined electricity load-forecasting system based on weighted fuzzy time series and deep neural networks." Engineering Applications of Artificial Intelligence, 2024, 132: 108375.
[6] Bracale, Antonio, et al. "Multivariate quantile regression for short-term probabilistic load forecasting." *IEEE Transactions on Power Systems*, 2019, 35(1): 628-638.

[7] Li, Binghui, and Jie Zhang. "A review on the integration of probabilistic solar forecasting in power systems." *Solar Energy*, 210 (2020): 68-86.
[8] Zhao, Hang, et al. "A generation-storage coordination dispatch strategy for power system based on causal reinforcement learning." *Sustainable Energy, Grids and Networks*, (2024): 101427.
[9] Dong Z, Ji X, Zhou G, et al. Multimodal neuromorphic sensory-processing system with memristor circuits for smart home applications[J]. IEEE Transactions on Industry Applications, 2022, 59(1): 47-58.
[10] Lai C S, Jia Y, et al. A review of technical standards for smart cities. Clean Technologies, 2020, 2(3): 290-310.
[11] Wang, Jianguo, et al. "Electrical load forecasting based on variable T-distribution and dual attention mechanism." *Energy*, 283 (2023): 128569.
[12] Li, Chaojie, et al. "Interpretable memristive LSTM network design for probabilistic residential load forecasting." *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69.6 (2022): 2297-2310.
[13] A. Faustine, and P. Lucas. "FPSeq2Q: Fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates." *IEEE Transactions on Smart Grid*, 13.3 (2022): 2440-2451.
[14] V. Álvarez, S. Mazuelas, and J. A. Lozano. "Probabilistic load forecasting based on adaptive online learning." *IEEE Transactions on Power Systems*, 36.4 (2021): 3668-3680.
[15] Zhou Y, Ding Z, Wen Q, et al. "Robust load forecasting towards adversarial attacks via Bayesian learning." *IEEE Transactions on Power Systems*, 2022, 38(2): 1445-1459.
[16] Li B, Mo Y, Gao F, et al. "Short-term probabilistic load forecasting method based on uncertainty estimation and deep learning model considering meteorological factors." *Electric Power Systems Research*, 2023, 225: 109804.
[17] Xu, Lei, Maomao Hu, and Cheng Fan. "Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks." *Journal of Building Engineering*,46 (2022): 103853.
[18] Zhou S et al. Multi-view Adaptive Probabilistic Load Forecasting Combing Bayesian Autoformer Network[J]. Journal of Electronics & Information Technology. 47(2024):1-9.
[19] Brusaferri A, Matteucci M, Spinelli S, et al. "Probabilistic electric load forecasting through Bayesian mixture density networks." *Applied Energy,* 2022, 309: 118341.
[20] Wang C, Wang Y, Ding Z, et al. "Probabilistic multi-energy load forecasting for integrated energy system based on Bayesian transformer network." *IEEE Transactions on Smart Grid*, 2023.
[21] Dong Z. et al. "PFFN: A Parallel Feature Fusion Network for Remaining Useful Life Early Prediction of Lithium-ion Battery." *IEEE Transactions on Transportation Electrification*, 2024.
[22] Ji X, Chen Y, Wang J, et al. "Time-Frequency Hybrid Neuromorphic Computing Architecture Development for Battery State-of-Health Estimation." IEEE Internet of Things Journal, 2024.
[23] Dong Z, Ji X, Wang J, et al. ICNCS: internal cascaded neuromorphic computing system for fast electric vehicle state of charge estimation[J]. IEEE Transactions on Consumer Electronics, 2023.
[24] Bergstra, James, et al. "Algorithms for hyper-parameter optimization." *Advances in neural information processing systems*, 24 (2011).
[25] Dong Z, Hu C, Zhou S, et al. "DECNet: A Non-Contacting Dual-Modality Emotion Classification Network for Driver Health Monitoring". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.
[26] Lv, Yunlong, et al. "Multi-area short-term load forecasting based on spatiotemporal graph neural network." *Engineering Applications of Artificial Intelligence*, 138 (2024): 109398.