PrEF: Probabilistic Electricity Forecasting via Copula-Augmented State Space Model

Zhiyuan Wang[1]*, Xovee Xu[1]*, Goce Trajcevski[2], Kunpeng Zhang[3], Ting Zhong[1], Fan Zhou[1]†

[1] University of Electronic Science and Technology of China
[2] Iowa State University
[3] University of Maryland, College park
zhy.wangcs@gmail.com, xovee@ieee.org, gocet25@iastate.edu, kpzhang@umd.edu, {zhongting, fan.zhou}@uestc.edu.cn

Abstract

Electricity forecasting has important implications for the key decisions in modern electricity systems, ranging from power generation, transmission, distribution and so on. In the literature, traditional statistic approaches, machine-learning methods and deep learning (e.g., recurrent neural network) based models are utilized to model the trends and patterns in electricity time-series data. However, they are restricted either by their deterministic forms or by independence in probabilistic assumptions – thereby neglecting the uncertainty or significant correlations between distributions of electricity data. Ignoring these, in turn, may yield error accumulation, especially when relying on historical data and aiming at multi-step prediction. To overcome these, we propose a novel method named **Pr**obabilistic **E**lectricity **F**orecasting (PrEF) by proposing a non-linear neural state space model (SSM) and incorporating copula-augmented mechanism into that, which can learn uncertainty-dependencies knowledge and understand interactive relationships between various factors from large-scale electricity time-series data. Our method distinguishes itself from existing models by its traceable inference procedure and its capability of providing high-quality probabilistic distribution predictions. Extensive experiments on two real-world electricity datasets demonstrate that our method consistently outperforms the alternatives.

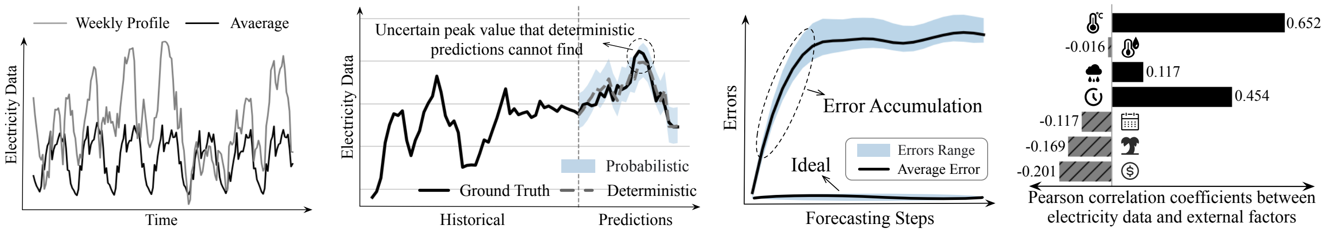
1 Introduction

Analyzing electricity time-series data (e.g., power generation and electricity demand) and making accurate forecasting has been a topic of interest starting in (and going on for most of) the 20^{th} century (Taylor 1975; Moghram and Rahman 1989), and continuing through contemporary research (Khan et al. 2016; Kuster, Rezgui, and Mourshed 2017; Aslam et al. 2021). Stable and efficient managements in modern electricity power systems – including power generation scheduling, transmission, distribution, and sales decision-making – are of extreme socio-economic importance, and heavily depend on the high-quality prediction of various electricity data at short-term. One of the issues is that the growing renewable energies (e.g., wind energy or hydropower) which reduce the carbon emissions, are unstable due to the rapidly changed climate conditions. A complementary issue is that the electricity cannot be stored in large quantities or for a significant duration, to be subsequently used when needed. This brings great uncertainty and increases difficulties in effective management of operations, as well as making reliable predictions for electric systems. In this paper, we investigate how to effectively integrate the use of historical observations and various external factors to model electricity time-series data and make high-quality forecasting for operations of electricity systems.

Electricity time-series data was analyzed by knowledgeable experts who would typically design mathematical/statistical models and tune suitable parameters from specific data type to simulate the dynamical procedure (Da Silva, Do Coutto Filho, and De Queiroz 1983; Moghram and Rahman 1989). In addition to relying on domain knowledge, such models generalize poorly on future data. More recently, researchers have proposed learning-based approaches, e.g, based on recurrent neural network (RNN) and variational-autoencoders (VAE) to conduct quantile regressions, variational inference, or Gaussian noise building for electricity time-series data (He et al. 2016; Dalal et al. 2020; Xia et al. 2021; Qiu et al. 2021).

However, the existing methods suffer from the following limitations (illustrated in Figure 1): **(1)**: They are unable to capture the uncertainty-dependencies of electricity time-series. Uncertainties widely exist in modern electricity systems due to the instabilities of power generations, especially for the growing renewable resources. Most of the existing models generate single-valued predictions and thus fail to consider the future uncertainty, or only utilize independent probabilistic techniques such as simple Gaussian process and Monte Carlo simulation (Ameli 2004; Nguyen and Quanz 2021), ignoring the dependencies between probabilistic distributions in a traceable way. **(2)**: They are prone to error accumulation. Existing models solely based on RNNs or autoregressive methods are ineffective in making the multi-step ahead predictions for electricity time-series. **(3)**: They cannot establish the interactive relationships between electricity data and various external factors. Electricity data are closely related with natural (e.g., weather or seasonal changes) and social (e.g., economics or policy) factors which greatly affect the performance of electricity forecast-

*Equal contribution.
†Corresponding author.



(a) Data (b) Uncertainty (c) Error accumulation (d) External factors

Figure 1: Illustration of electricity time-series data and three research questions.

ing. However, researchers found that classical methods cannot capture the complex dependencies among multivariate time-series and the non-linear interactions between different causal factors for their own limitations (Gallicchio, Micheli, and Pedrelli 2019).

To address the aforementioned limitations, we propose a novel copula-augmented state space model for **Pr**obabilistic **E**lectricity **F**orecasting (**PrEF**). It consists of two main modules: (1) *probabilistic temporal inference*; and (2) *interactive covariance establishment*. The first module proposes neural non-linear state space model (SSM) which transforms observations and external factors into SSM's parameter space for probabilistic time-series inference and learns the dependencies between electricity distributions. SSM maintains a latent state obeying Gaussian distribution at each timestamp and makes inference based on previous state, which can effectively ignore short-term interference from fluctuations and cumulative errors. The second module integrates a copula mechanism with SSM to learn the interactive relationships between electricity data and various external factors, enabling better approximations of the true distributions of electricity data. To summarize, our contributions are threefold:

- We propose PrEF, a probabilistic electricity forecasting framework that learns cross-dependencies between electricity distributions and eliminates the cumulative errors during the inference by a non-linear neural SSM.
- We propose a copula-augmented mechanism for incorporating the impact of complex external factors. It establishes the covariance matrix between electricity data and various external factors, which explicitly accounts for the interactive relationships of multivariate input, while giving probabilistic distribution predictions.
- We empirically validate our model on two large-scale real-world electricity forecasting datasets: one for renewable power generation and another for electricity demand. The experimental results demonstrate the superior performance of PrEF compared to baselines.

2 Problem Definition and Model Overview

We now formalize the problem and present the basic architecture components of PrEF.

2.1 Problem Definition

Let $\mathbf{X}_{1:T} = \{\mathbf{X}_1, \cdots, \mathbf{X}_t, \cdots, \mathbf{X}_T\}$ denote a set of multivariate time series observations at a sequence of time instants. Each $\mathbf{X}_t = \{\mathbf{X}_t^e, \mathbf{X}_t^a\} \in \mathbb{R}^{d_x}$ consists of two componente: (1) $\mathbf{X}_t^e$ denoting the value of electricity data; and (2) $\mathbf{X}_t^a$ denoting other auxiliary observations (e.g., temperature and humidity) at time t, where d_x denotes the feature dimensions of $\mathbf{X}_t$.

Let $\mathbf{E}_{1:T} = \{\mathbf{E}_1, \cdots, \mathbf{E}_t, \cdots, \mathbf{E}_T\}$ denote the set of associated external factors. Each $\mathbf{E}_t$ is composed of certain time dependent variables (e.g., the hourOfDay and dayOfWeek), as well as certain external variables which, although may occur at particular time instant, they need not have an explicitly known dependence on it (e.g., electricity price and capacity of electric power plants). Given past time series observations $\mathbf{X}_{1:T}$ and external factors $\mathbf{E}_{1:T+\tau}$, the goal of *electricity forecasting* problem is to produce a set of probabilistic forecastings, i.e., we are interested in the conditional distribution of future electricity values $\mathbf{X}^e_{T+1:T+\tau}$:

$$p(\mathbf{X}^e_{T+1:T+\tau} | \mathbf{X}_{1:T}, \mathbf{E}_{1:T+\tau}; \Theta) , \quad (1)$$

where τ is the future time-step for forecasting and Θ denotes learnable parameters of the model. In this paper, we study two forecasting problems: electricity demand forecasting (EDF) and electricity generation forecasting (EGF).

2.2 Model Overview

As a time-series inference neural network, PrEF provides a principled framework for modeling the uncertainty patterns of electricity time-series data. In addition, it establishes the correlations between variables and learns non-linearities from copula-augmented neural network.

The overall framework of PrEF is shown in Figure 2 and we note that, from a high-level perspective, there are two main modules of our proposed PrEFmodel: *probabilistic temporal inference* and *interactive covariance establishment*. The first one includes SSM into RNN-based network to transform the observations and external factors into SSM's parameter space for probabilistic time-series inference. The second one builds the covariance matrix to capture the cross-sectional and temporal-locality dependencies.

3 Methodology

This section details our methodology for probabilistic electricity time-series modeling and forecasting. We begin by explaining the novel probabilistic temporal inference modules where we create a non-linear neural SSM, followed by the establishing of interactive covariance, which explicitly accounts for external factors aggregation, and conclude with the details of optimization and prediction approach of PrEF.

3.1 Probabilistic Temporal Inference

Recurrent neural networks and their variants (e.g., LSTM and GRU) are widely used for modeling sequential dependencies in electricity forecasting (Kong et al. 2017; Dudek, Pełka, and Smyl 2021). As a simple but effective method, RNNs achieve satisfactory performance in traditional tasks. However, when applied to large-scale multivariate time-series, RNNs cannot effectively model the long-term correlations between external features (Gallicchio, Micheli, and Pedrelli 2019). Besides, direct sequential forecasting and temporal label-forced alignment lead to error accumulation in multi-step ahead forecasting, causing inaccuracies in long-term electricity predictions. More importantly, the deterministic structure of RNNs cannot support learning uncertainties from time-series data. This, in turn, impacts the effectiveness of the interval predictions, which is of a great significance for electric power industry (Hong and Fan 2016).

Inference with Nonlinear State Space Model. To address the issues mentioned above, we introduce a novel learning-based state space model (SSM). Its main benefit is that it has the ability to provide traceable multi-step ahead forecast distributions while accounting for data uncertainties. It can be described by the following process:

$$\begin{cases} \mathbf{L}_t = \mathbf{A}_t \mathbf{L}_{t-1} + \mathbf{B}_t + \mathbf{Q}_t \varepsilon, \\ \mathbf{X}_t = \mathbf{C}_t \mathbf{L}_t + \mathbf{D}_t + \mathbf{R}_t \varepsilon, \end{cases} \quad \varepsilon \sim \mathcal{N}(0,1). \tag{2}$$

At time t, $\mathbf{L}_t \in \mathbb{R}^{d_s}$ is the latent state in SSM which includes temporal patterns with respect to trend and seasonality, and d_s is the dimension of latent states. $\mathbf{A}_t \in \mathbb{R}^{d_s \times d_s}$ denotes the translation matrices evolving the latent states of the previous time to the current time, and $\mathbf{C}_t \in \mathbb{R}^{d_x \times d_s}$ denotes the parameters which transform latent states $\mathbf{L}_t$ into observations $\mathbf{X}_t$. $\mathbf{B}_t \in \mathbb{R}^{d_s}$ and $\mathbf{D}_t \in \mathbb{R}^{d_x}$ are system bias. Here we assume the vectors in $\mathbf{L}_t$ are independent, then positive $\mathbf{Q}_t \in \mathbb{R}^{d_s}$ and $\mathbf{R}_t \in \mathbb{R}^{d_f}$ denote the respective variances of latent states and observations. This iterative structure simulates the dynamics of complex states in time series. However, simple linear SSM ignores the non-linear relationships, which have impact on fitting ability. Thus we further employ the activation function to model the nonlinearity and the conditional probabilistic distribution of each time t, which can be described as follows:

$$\begin{cases} p(\mathbf{L}_0) &= \mathcal{N}(\mathbf{L}_0 | \mathbf{0},\ \mathbf{I}), \\ p(\mathbf{L}_t) &= \mathcal{N}(\mathbf{L}_t | \delta(\mathbf{A}_t \mathbf{L}_{t-1} + \mathbf{B}_t),\ \mathtt{diag}(\mathbf{Q}_t^2)), \\ p(\mathbf{X}_t) &= \mathcal{N}(\mathbf{X}_t | \delta(\mathbf{C}_t \mathbf{L}_t + \mathbf{D}_t),\ \mathtt{diag}(\mathbf{R}_t^2)), \end{cases} \tag{3}$$

where $\mathbf{L}_0$ is the initial latent state, $\delta(\cdot)$ denotes Tanh activation function and $\mathtt{diag}(\cdot)$ the diagonal function. Notably, after the introduction of nonlinearity in Eq. (3), the key distribution in SSM – i.e., filter distribution – is still Gaussian and equals to the distribution of canonical SSM. Based on

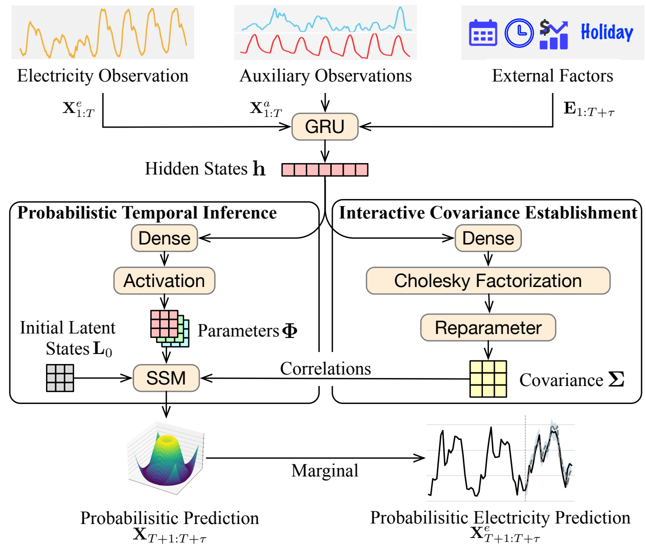


Figure 2: An illustration of PrEF. It takes electricity multivariate as input and outputs probabilistic distribution. There are two main modules: (1) probabilistic temporal inference; and (2) interactive covariance establishment.

this design, we can obtain the probabilistic results of time-series observations $\mathbf{X}_t$ at time t from the distributions in Eq. (3).

Autoregressive Parameter Learning. Traditionally, knowledgeable experts set translation matrices and variance of the SSM in a rule-based manner, relying on their experience, which can also be compute-intensive. However, in modern forecasting task with larger-scale dataset, it becomes more difficult to find suitable parameters to describe the temporal processes and to learn shared patterns of time-series data. Furthermore, in their classical setting, parameters are considered time-invariant in SSM, which is not flexible enough to model the dynamics of all time-steps.

To address these issues, we first let parameters be different at each time-step. The temporal evolution dynamics at time t are driven by parameters $\Phi_t = \{\mathbf{A}_t, \mathbf{B}_t, \mathbf{Q}_t, \mathbf{C}_t, \mathbf{D}_t, \mathbf{R}_t\}$. A general way to estimate the parameters Φ_t is to maximize the following likelihood:

$$\mathrm{argmax}_{\Phi_t} p(\mathbf{X}_{1:t} | \Phi_{1:t}). \tag{4}$$

We employ autoregressive models GRU (Cho et al. 2014) to estimate these parameters. Specifically, at each time step t, the GRU cells take the previous hidden states $\mathbf{h}_{t-1}$, observations $\mathbf{X}_{t-\tau}$, and the current external factors $\mathbf{E}_t$ as input. Notably, employing $\mathbf{X}_{t-\tau}$ aims to let the model learn the contemporaneous situation since electricity forecasting has a certain periodicity. For example, in hourly data, when $\tau = 24$, the model can utilize the electricity time-series at the same time yesterday to make the prediction. This design ensures GRU understand both trends and seasonality from external factors $\mathbf{E}_{1:t}$ and temporal patterns from observations $\mathbf{X}_{1:t}$. Thus, we can obtain the hidden states $\mathbf{h}_t$ from the GRU cell at time t, i.e.,

$$\mathbf{h}_t = \mathtt{GRU\text{-}CELL}(\mathbf{h}_{t-1}, \mathbf{X}_{t-\tau}, \mathbf{E}_t; \theta_g), \tag{5}$$

where θ_g denotes the learned parameters of GRU. Then the dense layers are applied to accomplish the affine transformation from $\mathbf{h}_t$ to Φ_t:

$$\Phi_t = \delta(\texttt{Dense}(h_t; \theta_d)), \tag{6}$$

where θ_d are parameters of the dense layers and $\delta(\cdot)$ denotes the activation function. For transformation matrices $\mathbf{A}_t, \mathbf{C}_t$ and bias $\mathbf{B}_t, \mathbf{D}_t$, we use *Sigmoid* function for training stability. For variance $\mathbf{Q}_t, \mathbf{R}_t$, we use *SoftPlus* function to ensure positive results. The probabilistic likelihood of the time-series is distributed according to:

$$p(\mathbf{X}_{1:t} | \mathbf{h}_{1:t-1}, \mathbf{X}_{t-\tau}, \mathbf{E}_{1:t}, \theta_g, \theta_d) = p(\mathbf{X}_{1:t} | \Phi_{1:t}), \tag{7}$$

which is equal to Eq. (4). Hence we can optimize the parameters $\Phi_{1:t}$ in temporal inference and parameters θ_g, θ_d in network by maximizing this likelihood with SSM, which is elaborated in Section 3.3.

3.2 Interactive Covariance Establishment

The accuracy of the electricity forecasting is typically influenced by various external factors, such as economy, weather, policy, historical data and real-time electricity price (Pardo, Meneu, and Valor 2002; Alamaniotis, Gatsis, and Tsoukalas 2018). However, in most probabilistic forecasting models, especially those based on neural networks, random variables are treated as independent (mostly for efficiency), which cannot reflect the significant correlations between electricity time-series and various external factors. To address this issue, we introduce copula mechanism to establish the interactive covariance and approximate the true distribution of time-series. In our PrEF, we naturally select Gaussian copula to establish the covariance since the assumption that variables follow multivariate Gaussian distribution.

According to Sklar's Theorem (Sklar 1997), given a series of random variables $\{k_i\}, i \in \{1, \cdots, n\}$ and their marginal cumulative distribution functions (CDFs) $F_i(k_i)$, the joint CDF is decomposed by a unique copula function $C(\cdot)$:

$$F(k_1, \cdots, k_n) = C(F_1(k_1), \cdots, F_n(k_n)). \tag{8}$$

Via this formula, we can easily construct the joint distribution from the known marginal Gaussian distribution $F_i(k_i) \sim \mathcal{N}(0, \sigma_i)$ and obtain the joint probability density function (PDF) by taking the derivative with respect to random variables:

$$\begin{aligned} p(k_1, \cdots, k_n) &= \frac{\partial F(k_1, \cdots, k_n)}{\partial k_1 \cdots k_n} \\ &= \frac{\partial C(F_1(k_1), \cdots, F_n(k_n))}{\partial F_1(k_1) \cdots F_n(k_n)} \prod_i \frac{\partial F_i(k_i)}{\partial k_i} \\ &= c(F_1(k_1), \cdots, F_k(k_n)) \prod_i p(k_i), \end{aligned} \tag{9}$$

where $c(\cdot)$ denotes the copula density. Then we can employ neural networks to build the covariance and to estimate the copula densities. On the one hand, we build $c^{\text{cross}}(\cdot)$ with cross-sectional correlations, i.e., covariance between electricity $\mathbf{X}_t^e$ and auxiliary observations $\mathbf{X}_t^a$. On the other hand, we construct $c^{\text{time}}(\cdot)$ for modeling the temporal locality dependencies of adjacent latent states, i.e., covariance between $\mathbf{L}_{t-1}$ and $\mathbf{L}_t$. Notably, in the following we will use superscript "$*$" to represent all variables during the establishment of $c_t^{\text{cross}}(\cdot)$ and $c_t^{\text{time}}(\cdot)$, since the only difference between them is their dimension.

To estimate the copula density at time t, we first build a covariance matrix through dense layers:

$$\alpha_t^* = \text{Tanh}(\texttt{Dense}(\mathbf{h_t}; \theta_\alpha)), \tag{10}$$

$$\mathbf{\Sigma}_t^* = \sigma_t^* \mathbf{I} + \alpha_t^* \alpha_t^{*\mathrm{T}}, \tag{11}$$

$$\text{where} \quad \sigma_t^{\text{cross}} = \mathbf{R}_t,\ \sigma_t^{\text{time}} = \mathbf{Q}_t, \tag{12}$$

where θ_α denotes the learned parameters and Eq. (12) guarantees $\mathbf{\Sigma}_t^*$ to be positive definite. Then we take the Cholesky factorization (Chen et al. 2008) with respect to $\mathbf{\Sigma}_t^*$ and get the factor $\mathbf{\Omega}_t^*$, from which we can utilize the reparameterization approach (Kingma and Welling 2014) for differentiable copula samples $\mathbf{\Gamma}_t^* \sim C_t^*(\cdot)$ and subsequent parameters optimization:

$$\mathbf{\Gamma}_t^* = \mathbf{\Omega}_t^* \mathbf{\Upsilon}, \quad \mathbf{\Upsilon} \sim \mathcal{N}(0, \mathbf{I}). \tag{13}$$

We compute the gradient for optimization with respect to copula samples as $\mathbf{\Gamma}_t^* \sim \mathcal{N}(0, \mathbf{\Sigma}_t^*)$, and then the logarithmic copula density can be calculated as:

$$\begin{aligned} \log c_t^*(\cdot) = \ & \frac{1}{2} \log |\mathbf{\Sigma}_t^*| - \sum_i \log \sigma_t^{*(i)} \\ & - \frac{1}{2} \mathbf{\Gamma}_t^{*\mathrm{T}} (\texttt{diag}(\mathbf{\Sigma}_t^{*-1}) - \mathbf{\Sigma}_t^{*-1}) \mathbf{\Gamma}_t^*, \end{aligned} \tag{14}$$

where $\sigma_t^{*(i)}$ is the i-th value of σ_t^*. Now, the distributions of latent states $\mathbf{L}_t$ and the observations $\mathbf{X}_t$ at time t are:

$$\begin{cases} p(\mathbf{L}_t) = \mathcal{N}(\mathbf{L}_t | \delta(\mathbf{A}_t \mathbf{L}_{t-1} + \mathbf{B}_t), \mathbf{\Sigma}_t^{\text{time}}), \\ p(\mathbf{X}_t) = \mathcal{N}(\mathbf{X}_t | \delta(\mathbf{C}_t \mathbf{L}_t + \mathbf{D}_t), \mathbf{\Sigma}_t^{\text{cross}}), \end{cases} \tag{15}$$

where both the correlations between various observations and temporal locality dependencies are fully taken into account.

3.3 Optimization and Prediction

To get the probabilistic distribution predictions, a straightforward way is to maximize Eq. (1). However, the probabilistic temporal inference in PrEF is an autoregressive process from the initial latent states $\mathbf{L}_0$. Ignoring the dynamics in time $t \in \{1, \cdots, T\}$ and treating them as "black box" will not only lead to training instability but also weaken the generalization capability of our model. Similarly, when only considering the electricity value $\mathbf{X}^e$, the model is prone to overfitting issue on training data. Thus, we set the optimization goal as maximizing the following loss:

$$\begin{aligned} \mathcal{L}(\Theta) &= \log p(\mathbf{X}_{1:T+\tau} | \mathbf{X}_{1:T}, \mathbf{E}_{1:T+\tau}; \Theta) \\ &= \sum_{t=1}^{T+\tau} \log p(\mathbf{X}_t | \mathbf{X}_{1:t-1}, \mathbf{E}_{1:t}; \Theta) \\ &= \sum_{t=1}^{T+\tau} \log c_t^{\text{cross}} + \sum_{i=1}^{d_x} \log p(\mathbf{X}_t^{(i)} | \mathbf{\Sigma}_{1:t}^*, \Phi_{1:t}), \end{aligned} \tag{16}$$

where $\mathbf{X}_t^{(i)}$ is the value of i-th observation of $\mathbf{X}_t$. In this loss, the first term optimizes the covariance, and we estimate it via Eq. (14). The second term is a likelihood driven by SSM parameters $\Phi_{1:t}$, and we employ Kalman Filter (Kalman 1960) to finish the deduction. This loss provides the sequential inference with structural assumptions and keeps the evolution traceable, significantly reducing the error accumulation impact – and we can use general optimization methods such as Adam and RMSprop to calculate the gradients and update parameters.

When the optimization finishes, we iteratively employ Eq. (15) at each time steps $t \in \{T+1, \cdots, T+\tau\}$ to obtain the probabilistic distribution of $\mathbf{X}_t$. To obtain the electricity value, we can sample values from the distribution of $\mathbf{X}_t$ and calculate the average as output. We use the joint distribution of $\mathbf{X}_t$ to build the probabilistic electricity distribution of $\mathbf{X}_t^e$.

4 Experiments

We now discuss the experiments for evaluating PrEF by comparing its performance with state-of-the-art electricity forecasting baselines and time-series prediction methods on two large-scale real-world electricity datasets.

Research Questions: We aim to address the following three research questions (RQ):

- **RQ1**: How does PrEF perform on electricity forecasting when compared to existing time-series learning baselines especially those based on probabilistic inference?
- **RQ2**: Can PrEF address the error accumulation issue in multi-step electricity forecasting?
- **RQ3**: Can PrEF approximate the true predictive distributions by building the interactive relationships of multivariate input?

Datasets and Processing: We use large-scale real-world electricity forecasting datasets:

(1) *Sichuan* dataset is collected from a hydroelectric station located in Sichuan, China and we use it for renewable power generation forecasting. Notably, hydropower generation is greatly influenced by external factors (e.g., water flow volume and rate) and thus makes the forecasting more challenging.

(2) *Panama* dataset contains the total electricity consumption in Panama in four years and we use it for electricity demand forecasting.

Baselines. We select nine baselines covering three classical methods, three RNN-based methods, and three probabilistic methods for comparison.

- **Historical Average (HA)**: uses the average value of previous $p = 7$ periods as prediction.
- **ARIMA** (Lee and Ko 2011): is a generalization of autoregressive moving average (ARMA) model which conducts statistical analysis on series data.
- **SVR**: is the regression type of support vector machine (SVM) which has been widely used for electricity data forecasting (Hong et al. 2013; Fan et al. 2016).
- **LSTM and GRU**: are two popular variants of RNN, used for modeling electricity time-series data (Kong et al. 2017; Dudek, Pełka, and Smyl 2021).
- **CNN-RNN**: combines the CNN with RNN to consider both rich features and dependencies in electricity time-series (Guo et al. 2020; Qu, Qian, and Pei 2021).
- **GRU-VAE**: employs GRU as encoder and decoder for prediction. It builds the distributions in latent space and reconstructs the observation via VAE (Qiu et al. 2021).
- **DSSM** (Rangapuram et al. 2018): bridges the gap between state space model and neural network, which can be used to make probabilistic predictions.
- **DeepAR** (Salinas et al. 2020): is an RNN-based probabilistic method and makes forecastings in the form of Monte Carlo samples. Besides, it can learn seasonal behaviors and dependencies on given features.

The deterministic models (LSTM/GRU) are used to parameterize a normal distribution instead of directly making predictions.

Experimental settings. Each dataset is split into two subsets **P1** (Jan 2017 - Dec 2018) and **P2** (Jan 2019 - Dec 2020), both spanning two years. Each subset is split into training (50%), validation (25%), and test (25%) sets. For time-series sequence, we let T be 48 and prediction steps τ be 24, i.e., we employ two days of observations as well as external factors to make future one-day hourly electricity forecasting. The dimensions of latent states in SSM and GRU are 128 and 256, respectively. Same with our PrEF, we employ standard GRU cell for DeepAR and DSSM and let dimension of latent state of GRU be 256. In DSSM, we also set consistent dimensions of latent state of state space model as 128. All models are tuned to the best performance with early stopping when validation loss has not declined for 50 consecutive epochs.

Evaluation Protocols. We use four common metrics to evaluate the forecasting performance. RMSE and MAE are aiming to measure the accuracy of predictions. Logarithmic density $\log p$ and continuous ranked probability score (CRPS) evaluate the quality of probabilistic distribution. We employ truncated normality for the computation of CRPS. Specifically, we only consider the range where the forecasting target locates. For the two tasks in our paper, we keep the positive value. Among them, larger $\log p$ is better, while for others the smaller the value – the better.

4.1 RQ1: Modeling Uncertainty-dependencies Improves Electricity Forecasting

We now consider the benefits of PrEF on improving the electricity forecasting by modeling the uncertainty-dependencies. Table 1 summarizes the performance of PrEF and the baselines in probabilistic electricity forecasting on both datasets. We can observe that PrEF achieves the best results in terms of all four metrics. In addition, we have following conclusions: (1) Classical approaches HA and statistic model ARIMA perform poorly since they can only make static predictions and cannot capture the uncertainties existing in electricity time-series; (2) All other baselines are

Model	Sichuan P1				Sichuan P2				Panama P1				Panama P2			
	RMSE	MAE	$\log p$	CRPS	RMSE	MAE	$\log p$	CRPS	RMSE	MAE	$\log p$	CRPS	RMSE	MAE	$\log p$	CRPS
HA	584.9	442.7	n/a	n/a	597.1	454.9	n/a	n/a	132.2	96.79	n/a	n/a	123.7	86.10	n/a	n/a
ARIMA	446.2	346.7	n/a	n/a	453.5	357.9	n/a	n/a	101.5	86.05	n/a	n/a	100.6	80.43	n/a	n/a
SVR	397.3	307.4	1.184	0.036	401.0	310.8	0.940	0.037	92.76	70.77	2.453	0.017	78.30	58.48	2.667	0.016
LSTM	366.9	282.0	1.300	0.033	356.7	267.4	1.237	0.032	79.24	56.68	2.649	0.015	73.60	52.04	2.694	0.015
GRU	365.0	280.6	1.349	0.034	355.3	265.7	1.239	0.032	80.24	57.49	2.677	0.015	74.34	51.84	2.680	0.016
CNN-RNN	361.5	277.1	1.351	0.034	352.6	261.6	1.240	0.032	72.43	51.60	2.734	0.012	70.85	51.74	2.784	0.014
GRU-VAE	346.2	264.5	1.390	0.032	339.4	254.6	1.363	0.030	76.04	60.24	2.727	0.013	73.74	52.18	2.675	0.015
DSSM	324.4	247.0	1.546	0.029	324.0	247.3	1.469	0.027	64.14	43.66	2.801	0.010	71.21	51.87	2.793	0.013
DeepAR	326.2	246.8	1.539	0.029	325.0	248.1	1.463	0.029	66.85	45.49	2.764	0.011	67.87	46.03	2.753	0.015
PrEF$^-$	315.0	242.2	1.597	0.028	320.4	244.9	1.506	0.027	62.29	42.60	2.901	0.009	64.35	43.11	2.823	0.012
PrEF	**308.6**	**239.5**	**1.626**	**0.026**	**318.5**	**242.7**	**1.522**	**0.026**	**60.00**	**40.84**	**3.118**	**0.008**	**63.94**	**42.01**	**2.851**	**0.012**

Table 1: Performance comparison on renewable power generation and demand forecastings on two real-world datasets. Best performance is in bold font and the best baselines result is underlined. PrEF$^-$ demotes removing the copula mechanism. “n/a” denotes not applicable since HA and ARIMA cannot make probabilistic predictions.

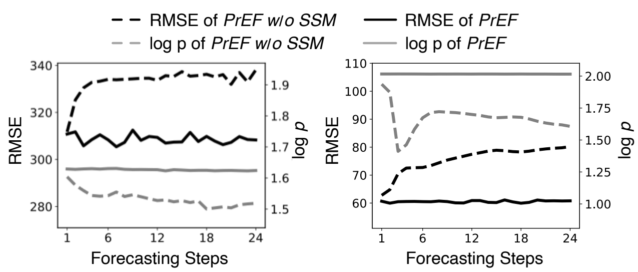


Figure 3: The effect of probabilistic temporal inference at multi-step electricity forecasting on Sichuan-P1 (left) and Panama-P1 (right) datasets.

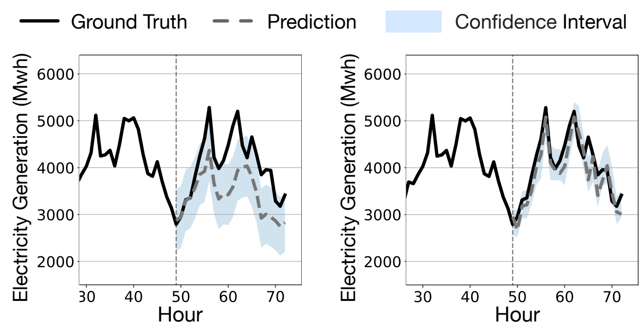


Figure 4: Comparison of *PrEF w/o SSM* (left) and PrEF (right) on a multi-step forecasting case.

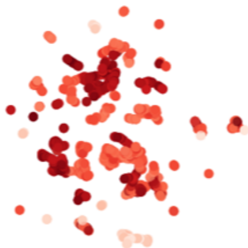
learning-based methods thus they can parameterize a distribution (mostly Gaussian) to account the uncertainties and output a probabilistic prediction; (3) Unlike PrEF, learning-based methods consider electricity distributions whether in observation or inference, which cannot fully estimate the uncertainty-dependencies. For example, GRU-VAE focuses on Gaussian noise of latent states, while others only consider the noise of predictions; (4) Although DSSM introduced SSM to model time-series for uncertainty-pattern perception, it ignores the contemporaneous observations and infers the latent states linearly. In PrEF, we consider the interactive relationships between various factors while DSSM treats them independently; Overall, our proposed PrEF handles the uncertainty that widely exists in electricity time-series in a traceable way and captures the uncertainty-dependencies between electricity distributions, enabling significant improvements in electricity forecasting.

4.2 RQ2: Probabilistic Temporal Inference Mitigates Error Accumulation

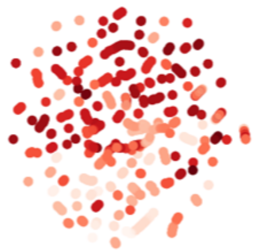
In this part, we investigate the effect of probabilistic temporal inference on the error accumulation. Towards that, we replace the evolution of SSM with fully connected layers to generate the probabilistic distributions of latent states $\mathbf{L}_t$ and electricity forecasting $\mathbf{X}_t^e$ at each time step. We keep other settings unchanged and use *PrEF w/o SSM* to denote it. To highlight the capability of our model, we report the performance changes at each forecasting step from $T+1$ to $T+\tau$, and the results are shown in Figure 3.

We can observe that our probabilistic temporal inference module outperforms the counterpart, especially when prediction steps increase. Like most existing models, *PrEF w/o SSM* generates the distributions at each time step independently, while our non-linear SSM provides a traceable way to consider the cross-dependencies between electricity distributions. Specifically, each distribution of latent state in inference is deduced at the previous step, making the distributions closely related. In addition, we use a Kalman filter method (Kalman 1960) to optimize the evolution procedure during inference and hence the performance degradation caused by error accumulation can be alleviated to a large extent.

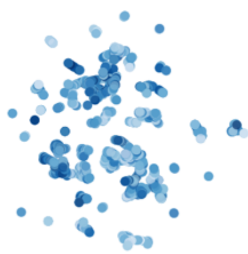
We further validate the quality of multi-step electricity predictions, comparing PrEF with *PrEF w/o SSM* in Figure 4. We can see that for *PrEF w/o SSM*, the gaps be-



(a) Temperature

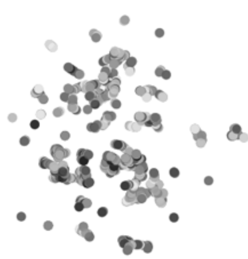


(b) Temperature



(c) Precipitation

(d) Precipitation



(e) HourOfDay

(f) HourOfDay

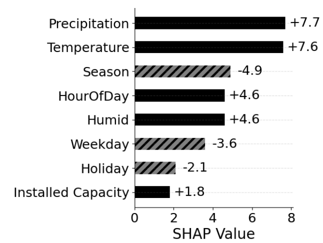
Figure 5: Visualization of the latent states in the 2-D space, where each point denotes a sample and the distances between points can reflect their original Euclidean distance. In each subfigure, we color points by the values of a specific external factor. The latent states are learned by baseline CNN-RNN (a,c,e) and PrEF (b,d,f), separately.

tween ground truth and predictions are getting larger as time increases. Conversely, PrEF approximates the ground truth better, demonstrating that probabilistic temporal inference helps in mitigating the impact of error accumulation.

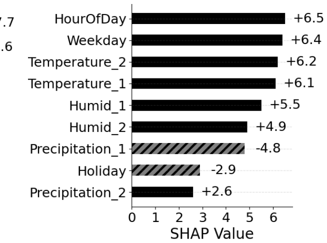
4.3 RQ3: Interactive Covariance Establishment Improves External Factor Aggregation

We now analyze the impact of incorporating external factors on electricity forecasting via interactive covariance establishment module. We first remove the this module and denote such model as PrEF^{-}. As shown in Table 1, our proposed copula mechanism brings additional performance improvements on both datasets and helps us better approximate the true distribution of electricity data.

We demonstrate our conclusion in a more intuitive way by visualizing the latent states. We project the latent states $\mathbf{L}_t$ to a 2D space using *UMAP* algorithm (McInnes, Healy, and Melville 2020) and color them by the values of external factors including temperature, precipitation, and hourOfDay, the results are shown in Figure 5, where we compare the latent states of CNN-RNN (a,c,e) with PrEF (b,d,f). We can observe that the positions of latent states in PrEF obey the Gaussian distribution and many of them with the similar color are clustered together. This demonstrates that our model can learn certain patterns between electricity time-series and external features. In contrast, the latent states in CNN-RNN cannot be disentangled within the color. These results suggest that our proposed PrEF not only improves the performance of electricity forecasting but also enables better explanations of prediction.



(a) Sichuan-P1



(b) Panama-P1

Figure 6: Influence of different factors estimated by SHAP values. The bars with hatch "/" denote negative correlations.

Lastly, we study the influence of each external factor on electricity forecasting for better understanding the behaviors of factors and providing prediction interpretability. We employ Shapley additive explanations (SHAP) (Lundberg and Lee 2017) to show the *importance values* of factors in Figure 6. We can see that climatic factors such as temperature and precipitation, play an important role in electricity forecasting since there exist certain intrinsic correlations; periodicity can also be understood well with time-dependent variables such as hourOfDay, weekday, etc. Additionally, holidays are negatively correlated with electricity demand on Panama dataset and precipitation has strong positive correlation with hydropower generation on Sichuan dataset. These indicate the benefits PrEF for improving the interpretability for the electricity forecasting, confirming that operations of electricity systems require overall coordination and optimization – and our model can provide an enabling perspective for probabilistic electricity forecasting.

5 Conclusions

We presented PrEF, a novel framework for probabilistic electricity forecasting, integrating traceable uncertainty-dependencies and interactive relationships. It provides high-quality probabilistic forecastings in short-term with stable performance, enabling significant improvements for operations of modern electricity systems. Our extensive experiments demonstrate that PrEF is superior to existing methods and provides a satisfactory interpretability. In our future work, we plan to extend PrEF to capture spatial-temporal correlations among both producers and consumers of electricity.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No. 62072077 and No. 62176043), and NSF SWIFT (Grant No. 2030249).

References

Alamaniotis, M.; Gatsis, N.; and Tsoukalas, L. H. 2018. Virtual Budget: Integration of electricity load and price anticipation for load morphing in price-directed energy utilization. *Electric Power Systems Research*, 158: 284–296.

Ameli, M. 2004. *On Monte Carlo Simulation and Analysis of Electricity Markets*. Ph.D. thesis, Royal Institute of Technology, Stockholm.

Aslam, S.; Herodotou, H.; Mohsin, S. M.; Javaid, N.; Ashraf, N.; and Aslam, S. 2021. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renew. Sust. Energ. Rev.*, 144: 110992.

Chen, Y.; Davis, T. A.; Hager, W. W.; and Rajamanickam, S. 2008. Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate. *ACM Transactions on Mathematical Software*, 35(3): 1–14.

Cho, K.; van Merrienboer, B.; Gülçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *EMNLP*, 1724–1734.

Da Silva, A. L.; Do Coutto Filho, M.; and De Queiroz, J. 1983. State forecasting in electric power systems. In *IEE Proceedings C (Generation, Transmission and Distribution)*, volume 130, 237–244. IET.

Dalal, N.; Mølnå, M.; Herrem, M.; Røen, M.; and Gundersen, O. E. 2020. Day-Ahead Forecasting of Losses in the Distribution Network. In *AAAI*, volume 34, 13148–13155.

Dudek, G.; Pełka, P.; and Smyl, S. 2021. A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting. *TNNLS*, 1–13.

Fan, G.-F.; Peng, L.-L.; Hong, W.-C.; and Sun, F. 2016. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing*, 173: 958–970.

Gallicchio, C.; Micheli, A.; and Pedrelli, L. 2019. Comparison between DeepESNs and gated RNNs on multivariate time-series prediction. arXiv:1812.11527.

Guo, X.; Zhao, Q.; Zheng, D.; Ning, Y.; and Gao, Y. 2020. A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. *Energy Reports*, 6: 1046–1053.

He, Y.; Xu, Q.; Wan, J.; and Yang, S. 2016. Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function. *Energy*, 114: 498–512.

Hong, T.; and Fan, S. 2016. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecasting*.

Hong, W.-C.; Dong, Y.; Zhang, W. Y.; Chen, L.-Y.; and Panigrahi, B. 2013. Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. *International Journal of Electrical Power & Energy Systems*, 44(1): 604–614.

Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems. *J. basic eng.*, 82(1): 35–45.

Khan, A. R.; Mahmood, A.; Safdar, A.; Khan, Z. A.; and Khan, N. A. 2016. Load forecasting, dynamic pricing and DSM in smart grid: A review. *Renew. Sust. Energ. Rev.*

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *ICLR*.

Kong, W.; Dong, Z. Y.; Jia, Y.; Hill, D. J.; Xu, Y.; and Zhang, Y. 2017. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1): 841–851.

Kuster, C.; Rezgui, Y.; and Mourshed, M. 2017. Electrical load forecasting models: A critical systematic review. *Sustainable cities and society*, 35: 257–270.

Lee, C.-M.; and Ko, C.-N. 2011. Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, 38(5): 5902–5911.

Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. In *NIPS*, 4768–4777.

McInnes, L.; Healy, J.; and Melville, J. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426.

Moghram, I.; and Rahman, S. 1989. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on power systems*, 4(4): 1484–1491.

Nguyen, N.; and Quanz, B. 2021. Temporal Latent Auto-Encoder: A Method for Probabilistic Multivariate Time Series Forecasting. In *AAAI*, volume 35, 9117–9125.

Pardo, A.; Meneu, V.; and Valor, E. 2002. Temperature and seasonality influences on Spanish electricity load. *Energy Economics*, 24(1): 55–70.

Qiu, Y.; Sun, Y.; Liu, C.; Li, B.; Wang, S.; and Peng, T. 2021. Aggregate Model for Power Load Forecasting Based on Conditional Autoencoder. In *International Conference on Intelligent Computing*, 406–416. Springer.

Qu, J.; Qian, Z.; and Pei, Y. 2021. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy*, 232: 120996.

Rangapuram, S. S.; Seeger, M. W.; Gasthaus, J.; Stella, L.; Wang, Y.; and Januschowski, T. 2018. Deep state space models for time series forecasting. *NeurIPS*, 31: 7785–7794.

Salinas, D.; Flunkert, V.; Gasthaus, J.; and Januschowski, T. 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecasting*.

Sklar, A. 1997. Random variables, distribution functions, and copulas – a personal look backward and forward. In *Distributions With Fixed Marginals & Related Topics*. JSTOR.

Taylor, L. D. 1975. The Demand for Electricity: a Survey. *The Bell Journal of Economics*, 6(1).

Xia, M.; Shao, H.; Ma, X.; and de Silva, C. W. 2021. A stacked GRU-RNN-based approach for predicting renewable energy and electricity load for smart grid operation. *IEEE Trans. Ind. Informat.*, 17(10): 7050–7059.