Applied Energy 213 (2018) 499–509

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

# Power load probability density forecasting using Gaussian process quantile regression

Yandong Yang[a,b], Shufang Li[a,b,*], Wenqi Li[c], Meijun Qu[a,b]

[a] Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Beijing Laboratory of Advanced Information Network, University of Posts and Telecommunications, Beijing 100876, China
[c] State Grid Henan Electric Power Company, Jinshui District, Zhengzhou 450052, China

## HIGHLIGHTS

- Propose a Gaussian process quantile regression (GPQR) model.
- The proposed method can provide power load probability density forecasting.
- PICP, PINAW and CWC are adopted to assess the GPQR model.
- The quality of PIs can be significantly improved by the GPQR model.
- The power load forecast accuracy are evaluated by two cases of PJM.

## ARTICLE INFO

*Keywords:*
Power load forecasting
Gaussian process
Quantile regression
Probability density forecasting

## ABSTRACT

Accurately predicting the power load in certain areas is of great importance for grid management and power dispatching. A great deal of research has been conducted within the smart grid system community in developing an assortment of different algorithms that seek to increase the accuracy of these predictions. However, these predictions suffer from various sources of error, such as the variations in weather conditions, calendar effects, economic indicators, and many other sources, which are caused by the inherent stochastic and nonlinear characteristics of power demand. In order to quantify the uncertainty in load forecasting effectively, this paper proposes a comprehensive probability density forecasting method employing Gaussian process quantile regression (GPQR). GPQR is a type of Bayesian non-parametric method which can handle the uncertainties in power load data in a principled manner. Consequently, the probabilistic distribution of power load data can be statistically formulated. The effectiveness of the proposed method for short-term load forecasting has been assessed adopting the real dataset provided by American PJM electric power company. Numerical results demonstrate that the uncertainties in power load data can be effectively acquired based on the proposed method. Meanwhile, the competitive predictive performance could be yielded with respect to the conventional adopted methods.

## 1. Introduction

The success of smart grid applications relies on the quality of grid information. This is especially true for state grid intelligent control system, where reliable and accurate grid information is highly desired for system operators. One of the critical needs for smart grid is to forecast the future power load. As is known, electricity cannot be stored in energy storage devices efficiently in large quantities, therefore system operators need to ensure that the amount generated during a certain period is sufficient to satisfy the load while not exceeding this demand significantly. Accurately predicting the power load not only can provide insightful information for system operators to reduce the maintenance costs [1], but also can ensure reliable power systems planning and operations [2].

Due to the significance of Short-Term Load Forecasting (STLF), many efforts have been devoted to developing varieties of STLF techniques such as statistical methods, machine learning methods and hybrid models. For an overview of the related works, the interested readers can refer to the recent book by [3]. The classical statistical models for STLF are various types of ARIMA which express the forecast as a function of historical load and possible exogenous variables [4–6]. Meanwhile, machine learning-based models such as Support Vector

* Corresponding author at: Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China.
*E-mail address:* bupt_paper@126.com (S. Li).

https://doi.org/10.1016/j.apenergy.2017.11.035
Received 28 August 2017; Received in revised form 2 November 2017; Accepted 4 November 2017
Available online 10 November 2017

Regression (SVR) [7,8] and Gaussian Process Regression (GPR) [9,10] have achieved great success in the last decade. Neural Network (NN) models have gained popularity for their excellent ability to model complex nonlinear relationship as well [11]. In addition, hybrid models such as [12–15], taking advantages of existing models, have attracted considerable attention of researchers. Selection of any of these techniques for STLF depends on problem domain, computation burden, dataset size, and analysis purpose.

Generally speaking, most existing literatures on electric load forecasting can be divided into two categories which are point forecasts and interval predictions [16,17] according to forecasting outputs. Point forecasts are the most traditional techniques, which provide an estimate of the future load for each step throughout the forecasting horizons as precise as possible. Rather than providing single-valued load forecast information, interval prediction methods attempt to construct well-calibrated lower and upper bounds of the future prediction associated with a prescribed probability called the confidence level. Different from these two types of forecasts, probability density forecasting can quantify the uncertainty by constructing probability density function of forecasting results. What is more, it can provide full probability distribution description of the future demand, which is especially desirable for power system management. Although calculating the probability for each possible prediction requires extra efforts, the additional information is highly useful to facilitate the full understanding of the service reliability.

With the penetration of renewable energies such as wind and solar power, the level of uncertainty in power systems has significantly increased. It is imperative to quantify potential uncertainties associated with forecasts. The applications of NN-based prediction intervals (PIs) for quantifying uncertainties associated with forecasted loads were investigated in [16,18], and the numerical results showed that prediction intervals provide more information about uncertainties existed in the process of load forecasting. Nevertheless, construction of (PIs) using these NN-based methods are computationally expensive. The interval type-2 FLS (IT2 FLS) was adopted to STLF for handling uncertainties [19], whereas the output of the IT2 FLS is an interval rather than a prediction interval. Liu et al. [20] utilized the quantile regression averaging (QRA) methodology to a set of sister point forecasts and generated PIs of future electric loads. Besides, a boosting additive quantile regression model for a set of quantiles of the future distribution was proposed by Taieb et al. [21]. Unlike interval prediction, probability density forecasting can provide a new perspective to solve this problem of uncertainties evaluation in power systems. There have been only a limited number of studies investigating the underlying probabilistic information. A semi-parametric additive model for forecasting the probability density functions (PDF) of the half-hourly electricity demand for power system was proposed in [22]. Recently, quantile regression neural network (QRNN) [23] has been proposed to draw complete conditional probability density curve of future load, but the most frequently occurring value of probability density curve could not achieve the most accurate prediction value. This phenomenon reveals that the shallow architecture of QRNN lacks enough capability to model the complex temporal characteristics of load series [24]. Later, a kernel-based support vector quantile regression [25] was put forward to improve the forecasting accuracy and quantify the uncertainty. In the smart grid era, the electricity demand is more active and less predictive than before, probabilistic load forecasting should be capable of quantifying the inherent uncertainties in load series and helpful to assess the risk of relying on the forecasts. Hence, a reliable and efficient probabilistic load forecasting technique is urgent to be designed.

As to time series modeling, Gaussian processes (GPs) are one of the most popular and advanced choice in the current state of the art for regression, for they are naturally able to handle complex relationships contained in time series. In addition, the uncertainty about the variance of the series at each point could be maintained, which is important for STLF since the nonstationarity and variability of load series can lead to high uncertainty. Kernel-based GPR [9] has been developed to fully make use of the relationships between multiple profiles. Meanwhile, the performance of three combinations of kernels have been compared. It is demonstrated that kernels with a multiplicative structure yield superior predictive performance than the widely adopted additive models. However, these kernel-based GPR models only provided point predictions instead of probabilistic predictions of the future demand.

Compared with GPs regression, Quantile regression (QR) is a type of regression method which aims at estimating quantiles of the response variable given certain values of the predictor variables [26,27]. Relative to the classical least squares estimation, quantile regression estimates are more robust against to outliers in the response measurements. What's more, different measures of central tendency and statistical dispersion can be useful to obtain a more comprehensive analysis of the relationship between variables. Therefore, quantile regression has been applied in the field of energy extensively, where probabilistic analysis and reliability assessment of power load are very significative [28,29]. The traditional regression models assume that any uncertainty in the learned model results from incomplete knowledge of underlying deterministic function. Quantile regression is the most relevant when the response is likely to be subject to variability or intrinsic randomness [30].

Due to the stochastic characteristics of power demand and various external impacts such as calendar effects, seasonal factors and weather conditions, the load signals would exhibit volatile temporal characteristics. Thus, a more meaningful prediction scheme should provide the most probable distribution of power load rather than one crisp value. As is known, there is no certain in forecasting. In this paper, we solve the problem of short-term probabilistic load forecasting by the means of Gaussian process quantile regression (GPQR), which incorporates Gaussian processes into the quantile regression to construct a more powerful nonlinear quantile regression model. GPQR can be thought of as a Bayesian alternative to the kernel methods [31]. It is the first application of method to handle the uncertainties in load forecasting to the best of our knowledge, which will enrich the literatures on probabilistic forecasting for electricity load data. A key advantage of this approach is non-parametric, which means our method can model arbitrarily complex systems given enough data.

The contributions could be summarized as follows:

- This paper proposes a comprehensive probabilistic load forecasting method based on Gaussian process quantile regression which can generate complete probability distribution of the future demand. We compare three different kernel functions and choose the optimal kernel function for the proposed model.
- Three Prediction Intervals (PIs) assessment criteria are employed to evaluate the performance of the proposed probabilistic load forecasting method, i.e., PI coverage probability (PICP), PI normalized average (PINAW) and coverage width-based criterion (CWC).
- Based on the open datasets obtained from PJM electric power company, we demonstrate the advantages of the proposed approach compared with the common used methods under various time-scales.
- The obtained results from two case studies show that the quality of PIs has been significantly improved compared with BPQR and SVQR models.

The reminder of this paper is organized as follows: An overview of existing literatures on short-term load forecasting is provided in the first section, and then, the mathematical background of the GPQR model is described in Section 2. In Section 3, we introduce several point forecasting metrics and PI evaluation indices for the model assessment. Section 4 validates the effectiveness and superiority of the proposed algorithm by comprehensive case studies adopting real-world PJM electricity datasets. Finally, the conclusions and guidelines for future work are outlined in Section 5.