Applied Energy 218 (2018) 159–172

Contents lists available at ScienceDirect

Applied Energy

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

Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data

Mahmoud Shepero*, Dennis van der Meer, Joakim Munkhammar, Joakim Widén

Department of Engineering Sciences, Uppsala University, P.O. Box 534, SE-751 21 Uppsala, Sweden

HIGHLIGHTS

- Probabilistic residential load forecasting using Gaussian and log-normal processes.
- Deterministic and probabilistic error metrics evaluated the proposed processes.
- Our results produced sharper forecasts compared with previous models.
- The log-normal process outperformed the Gaussian process in the forecast sharpness.
- The log-normal, unlike the Gaussian, process produced a varying forecast sharpness.

ARTICLE INFO

Keywords:
Gaussian process
Probabilistic load forecasting
Residential load

ABSTRACT

Probabilistic load forecasting (PLF) is of important value to grid operators, retail companies, demand response aggregators, customers, and electricity market bidders. Gaussian processes (GPs) appear to be one of the promising methods for providing probabilistic forecasts. In this paper, the log-normal process (LP) is newly introduced and compared to the conventional GP. The LP is especially designed for positive data like residential load forecasting—little regard was taken to address this issue previously. In this work, probabilisitic and deterministic error metrics were evaluated for the two methods. In addition, several kernels were compared. Each kernel encodes a different relationship between inputs. The results showed that the LP produced sharper forecasts compared with the conventional GP. Both methods produced comparable results to existing PLF methods in the literature. The LP could achieve as good mean absolute error (MAE), root mean square error (RMSE), prediction interval normalized average width (PINAW) and prediction interval coverage probability (PICP) as 2.4%, 4.5%, 13%, 82%, respectively evaluated on the normalized load data.

1. Introduction

Electric load forecasting is vital for several businesses that are dealing with the operation, trading, and planning of energy, for example, banks, electric utilities, and insurance companies [1]. Probabilistic load forecasting (PLF) has become increasingly important in the last decade, as PLF provides a probability density function (pdf) rather than a point forecast, which might be more valuable to the stakeholders compared with the point forecast [1]. It was estimated that a one percent improvement in the mean absolute percentage error (MAPE) can save hundreds of thousands of dollars for utilities [2].

In the literature there is a plethora of PLF models, for example [3–6]. In the recent years, hybrid models which combine two forecasting methods have been gaining popularity [7]. For example, models using support vector quantile regression (SVQR) along with copula [8], genetic algorithm along with artificial neural networks (GA-ANN) [9], and other hybrid methods in [10–13] have recently been developed. However, there is an underrepresentation of non-parametric methods in the research literature [14]. One of the relatively under used non-parametric methods in the field of PLF is the Gaussian process (GP) [15]. This introduction focuses on reviewing previous implementations of the GP in PLF; for more information on other PLF methods the reader is directed to two recent literature reviews [1,15].

A GP, as defined in [16], is a collection of random variables, any finite number of which have a joint Gaussian distribution. A GP can be defined by the covariance function, or kernel, which is used to describe relationships or nearness between inputs [16]. GP models have been used previously to forecast wind power [17,18], solar power [19],

* Corresponding author.
E-mail addresses: mahmoud.shepero@angstrom.uu.se (M. Shepero), dennis.vandermeer@angstrom.uu.se (D. van der Meer), joakim.munkhammar@angstron.uu.se (J. Munkhammar), joakim.widen@angstron.uu.se (J. Widén).

https://doi.org/10.1016/j.apenergy.2018.02.165
Received 21 September 2017; Received in revised form 22 January 2018; Accepted 25 February 2018
0306-2619/

electricity price [20,21], and temperature for electric load forecast [22]. It has also been used for clustering of electricity consumption profiles [23], and modeling of households' response to demand response (DR) signal from an aggregator [24]. Rasmussen and Williams [16] provided a detailed mathematical description of the GP models and their implementation.

Forecasts can be categorized based on the resolution of the forecast. On the monthly resolution, Ploysuwan et al. [25] utilised the GP to forecast the monthly peak load in Thailand. The predictors, or explanatory variables, contained the load of the previous month, the gross domestic product (GDP) and gross national product (GNP). However, when performing forecasting, the authors set the GDP and GNP to zero. Though the authors did not explain how doing so resulted in an accurate forecast, it is likely that the GP model did not put high predictive weight on the GDP and the GNP. Atsawathawichok et al. [26] also forecasted the monthly peak load in Thailand. The differences between [25,26] are that several kernels were incorporated in the latter model while the former one only used the Gaussian kernel, and that the latter model only used the previous lags in the time series as predictors while the former model used the GDP and the GNP—at least in training. Direct comparison of the performance between the two models is unfortunately not possible since they forecasted different years, and no error metrics for model evaluation were provided in [25]. Alamaniotis et al. [27] compared four different kernels to forecast the monthly electricity consumption using the monthly data from the four previous years as predictors, i.e., (12×4) values. The Matérn kernel demonstrated superior performance compared with the other kernels. Yan et al. [28] used a mixture of kernels to represent the monthly load trend in their GP forecasting model.

On the weekly resolution, Leith et al. [29] employed the GP to forecast the weekly electricity load. Two other models were compared, namely basic structural models (BSM), and seasonal auto-regressive integral (SARI)—described in detail in [30]. The authors concluded that the GP performed better than the other methods provided that the correct kernel is chosen. Moreover, the authors noted that the GP forecast decays to zero outside the space of the training data. The GP forecast decayed to the mean function in the cases of absence of knowledge learned from the training data [16]. This phenomenon can be avoided if (1) one of the predictors is the sequential index or time and the kernel have a periodic component, or (2) the prediction is not far outside the training data. Otherwise, the GP forecast decays to the mean function, which is commonly set to zero for simplicity [16,24].

GPs have also been used on daily resolutions. For example, Mori and Ohmi [31] showed that the GP outperformed multi-layer perceptron (MLP), support vector machine (SVR) and radial basis function network (RBFN) models in forecasting the daily maximum load. In their GP model, the predicted temperature of the forecasted day was the most important predictor for the forecast using the GP. In [32] historical load data from 1 month, and 1–3 years behind were used to predict the load of 30 days ahead. The authors used the Gaussian kernel. They also compared the GP, relevance vector regression (RVR) and autoregressive moving average (ARMA) models. The GP in this paper was the least accurate, i.e., worst performing, method. This bad performance might be attributed to the fact that GPs are designed for predicting single output rather than multiple output, i.e., one-step ahead rather than multi-step ahead prediction [33,34].

An analysis of hourly autocorrelation coefficients was done in [35] in order to select the most important time lags, as predictors, for the GP model. They developed the model for three different distribution feeders each representing a consumption profile: residential sector, non-residential sector, and service sector profiles. Lauret et al. [36] compared GP to ANN and Bayesian NN models in forecasting of the hourly load. The authors used the Gaussian kernel, and the GP model outperformed the other two methods on the test data set. In [37], the GP model suffered from overfitting when forecasting the hourly load using the Gaussian kernel. The GP could not achieve an MAPE of less than 15.9%.

Alamaniotis et al. [38] developed an ensemble GP using kernel machines for hourly forecasting. The authors employed a set of kernels to capture the different features of the demand patterns. A multi-objective optimization was used to find the linear combination between the Matérn, the NN and Rational Quadratic kernels that reduced the error metrics. One of the error metrics they used was the Theil inequality coefficient (Theil). The adoption of the Theil is sometimes accompanied by confusion since there are two versions of the Theil, and each has its own interpretation and parameters [39].

Yang et al. [40] used a hybrid GP quantile regression (GPQR) model to capture the relations between inputs and outputs of their probabilisitic forecast. The authors used the current and previous temperature measurements as exogenous variables in their model. The model was able to achieve a prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW) of 99.4% and 23.8% on a hourly data. Such good results were achieved on a similar sinusoidal dataset using another hybrid forecasting method in [13].

Five minute forecasts were developed in [41] for providing a PLF for the energy intensive enterprises. The authors used the automatic relevance determination (ARD) version of the Gaussian kernel. They also used probabilistic error metrics, however, they did not normalize them. They could achieve a prediction interval with a width of around 100 MW in a 1150 MW peak load. On the minute resolution, Alamaniotis and Tsoukalas [42] compared GPs, with three different kernels, to the ARMA model. The Matérn kernel was ranked first as it performed best in 14 out of 24 trials.

The limitation of the GP to the forecsting of one-step ahead led to the development of the twin Gaussian processes (TGP) in [14,34,33] and other inference methods in [28,43]. Yan et al. [14] adopted the TGP for the PLF on a time series with hour resolution. They used the 24th and the 168th time lag besides six other predictors to perform multi-step ahead forecast of daily and hourly load. The authors showed that the percent error is higher in case of the hourly prediction compared with the daily prediction. On the other hand, Yan et al. [28] compared the recursive and direct multi-step ahead forecasting strategies using conventional GPs. In the recursive approach a single step ahead GP model was used to forecast multiple one-step ahead recursively. The direct approach trains several GP models each for forecasting one specific step. The direct approach outperformed the recursive approach. The uncertainty of the multi-step ahead prediction is supposed to increase with the index of the forecasted steps [43], i.e., in the multi-step ahead forecast the variance of the first step is supposed to be smaller than the variance of the 10th step. To model this phenomena, Girard et al. [43] used the recursive approach, however, they viewed the lagged values as random variables. As a result, the uncertainty propagated from the forecasted first-step to the last-step ahead.

Learning the GP is based upon selecting the appropriate kernel for the problem at hand and learning the optimal hyperparameters [28]. The choice of kernel should be made by the modeller [16,24,29,38], and it is essential for ensuring the correct performance of the forecast [28]. In the reviewed papers, the Gaussian kernel was used in [25,27,31,32,35–37,40,42]. The ARD version of the Gaussian kernel was used in [41]. This kernel is smooth and might not be suitable for representing many physical processes, which is why the Matérn kernel is a better alternative [16]. The Matérn kernel was used in [27,42]. The NN kernel was employed in [27,42], linear kernel in [27]. Summation of several kernels was done in [29,38]. In [26,28] a similar kernel mixture was used. This mixture was used previously in [16] for a similar problem.

Alamaniotis et al. [38] used deterministic error metrics, like the mean square error (MSE), the root mean square error (RMSE), the mean absolute error (MAE), the MAPE, the maximum absolute percentage error (MAP) and the Theil, for the optimization of the hyperparameters of the kernel. The use of deterministic error metrics for learning the GP