Applied Energy 224 (2018) 13–33

Contents lists available at ScienceDirect

Applied Energy

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

# Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model

Guo-Feng Fan[a], Li-Ling Peng[a], Wei-Chiang Hong[b,*]

[a] School of Mathematics and Statistics, Ping Ding Shan University, Ping Ding Shan 467000, Henan, China
[b] School of Education Intelligent Technology, Jiangsu Normal University/101, Shanghai Rd., Tongshan District, Xuzhou 221116, Jiangsu, China

## HIGHLIGHTS

- A novel hybrid STLF model based on PSR algorithm and BSK regression model is proposed.
- The data series is sufficiently reconstructed by PSR algorithm to demonstrate its evolution mechanism.
- The spatial structure among regression and neighbor points are reasonably illustrated by BSK regression model.
- The proposed PSR-BSK model significantly receives higher forecasting accuracy and shorter running time than other models.

## ARTICLE INFO

*Keywords:*
Electricity load forecasting
Phase space reconstruction (PSR) algorithm
Spatial geographical weighted
Bi-square kernel (BSK) regression

## ABSTRACT

Short term load forecasting (STLF) is an important issue for an electricity power system, to enhance its management efficiency and reduce its operational costs. However, STLF is affected by lots of exogenous factors, it demonstrates complicate characteristics, particularly, the multi-dimensional nonlinearity. Therefore, it is desired to extract some valuable features embedded in the time series, to demonstrate the relationships of the nonlinearity, eventually, to improve the forecasting accuracy. Due to the superiorities of phase space reconstruction (PSR) algorithm in reconstructing the phase space of time series, and of bi-square kernel (BSK) regression model in simultaneously considering original spectral signature and spatial information, this paper proposes a novel electricity load forecasting model by hybridizing PSR algorithm with BSK regression model, namely PSR-BSK model. The electricity load data can be sufficiently reconstructed by PSR algorithm to extract the evolutionary trends of the electricity power system and the embedded valuable features information to improve the reliability of the forecasting performances. The BSK model reasonably illustrates the spatial structures among regression points and their neighbor points to receive the rules of rotation rules and disturbance in each dimension. Eventually, the proposed PSR-BSK model including multi-dimensional regression is successfully established. The short term load data from the New South Wales (NSW, Australia) market and the New York Independent System Operator (NYISO, USA) are employed to illustrate the forecasting performances with different alternative forecasting models. The results demonstrate that, in these two employed numerical examples, the proposed PSR-BSK models all significantly receive the smallest forecasting errors in terms of MAPE (less than 2.20%), RMSE (less than 30.0), and MAE (less than 2.30), and the shortest running time (less than 400 s) than other forecasting models.

## 1. Introduction

Accurate short-term load forecasting (STLF) plays an important role in operation management of an electricity power system where it is used to make critical decisions such as schedule of electricity production and purchase, reliability and security analysis, and maintenance plans of systems in the competitive energy markets [1]. Therefore, improving the accurate level of STLF could not only increase the management efficiency in terms of schedule planning, but also significantly reduce operational costs [2]. STLF is affected by lots of exogenous factors, such as weather, social economic activities, holidays, and so on, thus, it demonstrates several complicate characteristics, such as time varying, uncertainty, and multi-dimensional nonlinearity [3–5]. The common STLF models often focus on one or more of those factors, however, the forecasting results almost illustrate that those exogenous influences on the loads could not be completely

* Corresponding author.
*E-mail addresses:* samuelsonhong@gmail.com, hongwc@jsnu.edu.cn (W.-C. Hong).

https://doi.org/10.1016/j.apenergy.2018.04.075
Received 13 February 2018; Received in revised form 6 April 2018; Accepted 26 April 2018

**Nomenclature**

| | |
|---|---|
| $x_i(i = 1,2,...,N)$ | a time series data |
| $N$ | the number of sampling points |
| $X_t(t = 1,2,...,N-(m-1)\tau)$ | the delay sequence of $x_i$ |
| $\tau$ | the delay time |
| $m$ | the embedded dimension |
| $C(m,N,r,\tau)$ | the correlation integrals |
| $r$ | the spatial distance |
| $S(m,N,r,\tau)$ | the serial correlation of a nonlinear time series |
| $S(m,r,\tau)$ | the serial correlation of a nonlinear time series when $N \to \infty$ |
| $\tau_{opt}$ | the local optimal delay time |
| $\Delta S(m,\tau)$ | the quantity of different spatial distances |
| $s$ | the standard deviation of the data set |
| $\bar{S}(\tau)$ | the mean of the quantity, $S(m,r,\tau)$ |
| $\Delta\bar{S}(\tau)$ | the mean of the quantity, $\Delta S(m,\tau)$ |
| $S_{cor}(\tau)$ | the sum of the absolute value of $\bar{S}(\tau)$ and $\Delta\bar{S}(\tau)$ |
| $\tau_w$ | the global optimal delay time |
| $\mathbf{R}$ | the recurrence matrix |
| $\theta(x)$ | Heaviside function |
| $\varepsilon$ | the distance threshold |
| $w_{ij}$ | the associate weight of $d_{ij}$ |
| $d_{ij}$ | the distance between the data point ($j$) and regression point ($i$) |
| $b$ | the bandwidth between $w_{ij}$ and $d_{ij}$ |
| $n$ | the number of observations |
| $\mathbf{S}$ | the regression coefficient matrix |
| $tr(\mathbf{S})$ | the trace of $\mathbf{S}$ |
| $\hat{\sigma}$ | the maximum likelihood estimator of the variance |
| RSS | the residual sum of squares |
| $\overrightarrow{k}^{i}_{mm}$ | the $k$ point with minimum distance in the $m$th dimension |
| $\hat{f}_i$ | the forecasting load of $x_i$ |
| $a_i$ | the actual load of $x_i$ |
| $q$ | the number of compared models |
| $Rank_j$ | the average rank sum received from each forecasting value for each model |

addressed. Therefore, the critical issue of STLF is well dealing with the above mentioned characteristics of loads to increase the forecasting accurate level.

In the past decades, many STLF models have been continuously proposed to improve forecasting accuracy. These STLF models are often classified into two categories, traditional statistical models and artificial intelligent models. The statistical models use historical data to find out the relationships between exogenous variables and electric loads, or the relationships among the time periods of data itself, in which the relationships are theoretically defined as linearly. There are many famous statistical models, including ARIMA models [2,6,7], regression models [8–10], exponential smoothing models [11,12], Kalman filtering models [13,14], and Bayesian estimation models [15,16]. For example, in Ref. [7], authors apply auto-regressive (AR) and moving-average (MA) methods to model separately non-seasonal and seasonal cycles of the hourly electric load without any additional factors (such as weather data) from the PJM interconnection power network. The forecasting performance of the proposed model receives 0.86% in terms of MAPE, and demonstrates 21% reduction compared with Dudek's forecaster proposed in 2016. In Ref. [10], authors also demonstrate a new STLF technique for forecasting 1-week-ahead daily load by using two separate forecast processes: seasonal and trend items, for the short term load data from Victoria (VIC) grid in Australia. After eliminating the seasonal item in the original load demand, the regression model is employed to forecast the trend item. The forecasting results indicate that the proposed model can significantly improve the accuracy in terms of MAPE.

These statistical models are easily to implement, however, the embedded drawback is that they are theoretically based on the linear definition. They could hardly deal with the complicate nonlinear characteristics of electric load series, and almost could not receive satisfied forecasting performances [17]. Recently, to overcome the limitation of linear definition, Ref. [14] uses hourly electric load from the TEPCO (Tokyo Electric Power Company) to propose a novel framework for forecasting electric loads by combining ensemble Kalman filter technique with multiple regression model. The proposed ensemble Kalman filter consists of a linear observation model with Gaussian noise and a linear or nonlinear system model with any type of noise distribution to receive much greater nonlinear flexibility than does the original Kalman filter. The forecasting accuracy of the proposed model is significantly better than that of the current state-of-the-art models.

Due to superior nonlinear computing capability, the artificial intelligent models have been widely explored in STLF to receive higher and satisfied forecasting accuracy, such as artificial neural network (ANN) [17–20], knowledge based (expert) system models [1,21], fuzzy theory models [22–24], and support vector regression (SVR) models [25–28]. For example, in Ref. [20], authors use the hourly temperatures and electric loads from the New England Pool region to present the boosted neural networks for short-term load forecasting by minimizing the error between the estimated output from the previous iteration and the target output. The forecasting results confirm that the proposed model outperforms other existing techniques. In Ref. [1], authors use the temperature and load data from Iran's national power network to investigate a knowledge-based STLF model by proposing a novel priority index for selection of similar days, which temperature similarity and date proximity are simultaneously considered. The proposed model demonstrates the superiorities compared with Bayesian neural network and linear neuro-fuzzy methods in aspects of forecasting accuracy and computation time. In Ref. [24], authors apply fuzzy logical weights to compensate the presence of bias to improve the STLF accuracy in Malaysian electricity market. In Ref. [25], authors propose a new SVR-based STLF model with the ambient temperature of two hours as input variables and electric loads from four typical office buildings in China. The empirical results demonstrate that the proposed SVR model offers a higher degree of prediction accuracy and stability. Ref. [27] is authors' previously proposed SVR-based STLF model by hybridizing with the differential empirical mode decomposition (DEMD) method and auto regression (AR). The numerical results, by using NSW and NYISO electric load data sets, illustrate the validity of the idea that the proposed model can simultaneously provide forecasting with good accuracy and interpretability.

Moreover, since the change tendency of the short term electric load often demonstrates fluctuation and non-stationary, some analysis tools have been applied into these artificial intelligent models to deal with the specific problems [24,29,30]. Therefore, the research direction of STLF in the recent years is concentrated on proposing hybrid or combined models:

(1) hybridizing or combining these artificial intelligent models with each other [31–33]. In Ref. [31], authors combine self-organizing map technique and $k$-means algorithm with a multilayer perceptron model to conduct STLF by employing the electric load data from Spanish utility Iberdrola. In Ref. [32], authors propose two hybrid STLF models, by hybridizing wavelet transform with an ANN and with adaptive neural fuzzy inference system, respectively, to forecast Iran's electric load. This kind of hybrid or combined model is empirically superior to other alternatives. In Ref. [33], authors combine the self-organizing map, the SVR model, and the fuzzy