

Parametric Versus Non-Parametric Time Series Forecasting Methods: A Review

Anjali Gautam^{1,2,*} and Vrijendra Singh¹

¹Department of Information Technology, Indian Institute of Information Technology, Deoghat Jhahwa, Allahabad-211015, Uttar Pradesh, India

²Department of Computer Engineering and Applications, GLA University Mathura-281406, Uttar Pradesh, India.

Received 19 March 2020; Accepted 22 May 2020

Abstract

The non-parametric methods have been proposed in the research literature as an alternative to parametric methods for time series forecasting. However, scarce evidence is available about the relative performance and computational ability of both parametric and non-parametric methods. This paper reviews the comparative studies conducted for evaluating the accuracy of parametric and non-parametric methods, especially machine learning methods. For this, we briefly review widely applicable the parametric and non-parametric methods. Moreover, an empirical study has been carried out on real time series datasets to evaluate the comparative performance of parametric methods over machine learning methods. Additionally, the limitations of the machine learning methods are highlighted which leads to the selection of parametric methods over non-parametric methods by the researchers in recent years. Further, some recommendations for future research are presented.

Keywords: Parametric methods; non-parametric methods; machine learning.

1. Introduction

Forecasting of future events such as the next year rainfall measurement [1] or a stock market volatility [2] with achieving higher accuracy constitutes a challenging problem for the field of applied and theoretical time series research. Time series analysis is applicable in many areas like statistics, econometrics, signal processing, pattern recognition, mathematical finance, weather forecasting, ECG, control engineering and astronomy etc. The methods and techniques for time series analysis can be categorized as parametric and non-parametric methods. The parametric methods assume that the basic stochastic stationary process has a certain structural formation which may be described by utilizing a small number of parameters (for e.g., applying an autoregressive moving average (ARMA) model [3] or an autoregressive integrated moving average (ARIMA) model [4]). The main task of parametric approaches is to evaluate the parameters of the time series model that outlines a stochastic process. [5] In contrast, non-parametric methods explicitly evaluate the spectrum or the covariance of the stochastic process without assuming that the stationary process may have any specific structure [6] [7]. Furthermore, only parameters are required from the historic data of the model while using parametric models, to forecast the future values of time series. On contrary, the non-parametric models require current state of data with number of parameters to predict the future values of time series data [8]. Moreover, non-parametric models are computationally slower than parametric models, but make lesser assumptions about the time series data. Further, parametric models significantly

varies from non-parametric models due to some assumptions required by parametric models regarding the temporal and spatial covariance structure and the marginal probability distributions of the time series data, whereas non-parametric models only maintain the empirical structure of the observed time series data.

The parametric methods and non-parametric methods both can be opted for time series analysis, however, it is important to know beforehand which type is best suited for particular time series problem. Therefore, a comprehensive review of comparative studies of parametric methods and non-parametric methods is conducted in this paper to address the problem of method selection. Moreover, an empirical study has been carried out on real time series datasets to evaluate the comparative performance of parametric methods over machine learning methods. Additionally, the limitations of the machine learning methods are highlighted which leads to the selection of parametric methods over non-parametric methods by the researchers in recent years. It is important to note the limitations of machine learning as these methods claims superiority [9], however several recent studies chose parametric methods over machine learning methods for addressing different forecasting problems [10-14].

The remaining part of the paper is organized as follows. Brief description of widely applicable parametric and non-parametric time series forecasting methods is presented in Section 2. Thereafter, comparative studies of parametric methods and non-parametric methods are reviewed in Section 3. Further, an empirical study of popular parametric and non-parametric methods is presented in the Section 4. In Section 5, the reasons for applicability of parametric methods in machine learning era are highlighted. Some recommendations for future studies are discussed in Section 6. Finally, conclusions are presented in Section 7.

*E-mail address: rs176@iitit.ac.in

ISSN: 1791-2377 © 2020 School of Science, IITU. All rights reserved.

doi:10.25103/jestr.133.18

2. Widely used parametric and non-parametric time series forecasting models

This section describes popular parametric and non-parametric time series forecasting models. The discussion begins with exponential smoothing and stochastic models (ARMA and ARIMA). Thereafter, the non-parametric which includes regression models, neural networks, Gaussian process and long-short term memory (LSTM) model are briefly described.

2.1. Parametric models

For analysis of time series data, a parametric model is a good choice as it is easy to use due to transparency and less computing power requirement. The theory behind parametric time series models is rich and their applications can be found in many fields, including the augmentative area of studies which involve financial and environmental data. In a parametric modeling, we estimate the parameters of the probability distribution assumed for the time series data. In next sections, few popular parametric methods are described briefly.

2.1.1. Exponential smoothing (ES)

Exponential smoothing model is originated with the work of Holt [2004], Brown [2004] and Winters [1960] by generalizing exponentially weighted moving average (EWMA) to include trend component in producing forecasts and it also allowed effects of seasonality in time series forecasting. Mathematically, simple ES model can be represented as

$$Z_t = wY_t + (1 - w)Z_{t-1} \tag{1}$$

where Z_t is the smoothed level of the series computed after Y_t is observed, and Y_t is observed value of time series in period t . The ES methods are considered as ad-hoc since these methods are formulated without any well-specified statistical model. In [18], provided a statistical basis for exponential smoothing in terms of Wold decomposition for a random walk with noise. Further, research to provide a statistical framework was done by [19] and [20] which shows that the forecasts produced by some linear ES methods arose as a special case of ARIMA methods. Later, several researchers provide various exponential methods to include damped trend [21], state space model [22], kernel regression [23] and double smoothing [24] etc. to produce forecasts.

2.1.2. Autoregressive Moving Averages (ARMA) model

ARMA is an arrangement of two different types of model, namely the autoregressive (AR) model, and the moving average (MA) model. These models are developed with the

work of [25] which describes the notion of stochasticity in time series. This idea led to the development of AR and MA time series models [26] [27]. An AR (p) model is a serially dependent forecasting model which predicts the future time series values based on past values of the observed series. This is done by calculating a coefficient or set of coefficients that refer to consecutive values of the series from particular and time-lagged values. Mainly MA(q) is used for identification of patterns in time series and smoothing of time series data but this is also modeled for predicting future values of time series [26]. Although, this method smoothens out the irregular patterns in time series, however, these smoothing abilities of MA model is useful in decomposing a time series. This property of MA model is well utilized in ARMA and ARIMA model for identifying the shock effects in white noise terms. Moreover, an ARMA method describes the conditional mean of Y_t as a function of both previous observations, Y_{t-1}, \dots, Y_{t-p} and previous innovations, $\epsilon_{t-1}, \dots, \epsilon_{t-q}$. The number of previously observed values that, Y_t depends on, p , is the order of AR. The number of previous innovations that Y_t depends on, q , is the order of MA [5]. In general, these models are represented by ARMA (p, q). A stationary ARMA model of orders p and q for a time series Y_t is defined as

$$Z_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \epsilon_t + \sum_{i=1}^q \beta_i \epsilon_{t-i} \tag{2}$$

where Z_t is the smoothed level of the series computed after Y_t , is observed, Y_t , is observed value of time series in period t , ϵ is the white noise and α, β are parameters [3]. Moreover, it is possible that an effect of an AR term may be canceled out by an MA term in ARMA model or vice versa. Although both the terms are important in modeling the time series using ARMA and ARIMA model.

2.1.3. Autoregressive Integrated Moving Averages (ARIMA) model

Basically, the ARMA models were developed for stationary time series, however a new class of models was introduced by integrating a phase for removal of non-stationarity in a time series. These models known as ARIMA models and developed by Box and Jenkins [5] which is a stochastic process. ARIMA is described as three stage iterative model which includes time series identification, estimation, and verification. The Figure 1 shows the Box-Jenkins approach for ARIMA modeling of time series data. The Box-Jenkins approach mainly use three filters namely integration filter, AR filter and MA filter. The integration filter produces a filtered (differenced) series from observed data. The AR filter generates an intermediate series which is further processed by MA filter which results in random white noise [5].



Fig. 1. Process of ARIMA Modeling

Usually, ARIMA is represented as ARIMA (p, d, q) where, p denotes the number of Autoregressive terms, d denotes the degree of differencing, and q denotes the number of lagged forecast error. All these parameters are computed with the help of autocorrelation function (ACF) and partial autocorrelation function (PACF) [4]. The ARIMA models are linear in nature, but it is not satisfactorily possible to model a real time series in linear fashion obtained through real-life problems. Further, it is difficult to apply these models to non-stationary time series, as it is only applicable to the stationary or weakly stationary time series. Assuming that the non-stationarity can be removed with differencing which may distort the actual behaviour of time series. It is noted that the residuals of MA model grows exponentially when the estimated parameters are not in the appropriate range and the computed residuals for later observed values of the series could be very large or could overflow. The orders of ARMA and ARIMA models are identified through the ACF and PACF plots, which is difficult and requires much experience even for the simplest models. ARMA models are a conditional expectation and non-conditional variance models, thus it is not really recommended for heteroscedastic time series.

2.2. Non-parametric models

Non-parametric methods have attained prominence in the last decade. Its success lies on the ability to learn by trial-error method and also by improving the accuracy over iterations. Different models such as neural networks (NN), support vector regression (SVR) [28] [29], regression trees [30] [31], Gaussian process (GP) [32], long short-term memory (LSTM) [7] and others that are collectively known as machine learning algorithms which have been applied to time series forecasting.

2.2.1. Regression models

Regression models depicts the relationship between an output (response) variable, and one or more input (predictor) variables. Most popular machine learning regression methods include linear regression, CART regression trees (CART) [31] and support vector regression models (SVR) [28]. The CART is recursive tree like method for partitioning the input variable and based on the idea of decision trees. Additionally, the regression trees have been widely used in diverse regression problems such as environment [31], electricity load [30], and neuroscience [33]. Further, some different variations of regression trees are also available in literature such as random forest regression [33], gradient boost regression tree [31], and bagging regression trees [30].

Moreover, the SVR is developed from support vector machines which identifies the hyperplane that maximizes the margin between two different classes. Further, SVR have been utilized for forecasting problems in different fields, such as financial time series [34], software management [35], traffic management [29], electric load [28] and many more. Empirically, SVR reported the poor forecasting accuracy, due to lack of knowledge regarding the selection of parameters in SVR model. Although, structural methods for ensuring optimal parameters selection are scant in literature. Moreover, few searching algorithms such as simulated annealing algorithm and genetic algorithm were also utilized to test the suitability of parameter selection for SVR model [36]. Further, a chaotic version of genetic algorithm was applied to parameter selection for electric load forecasting [28]. However, these searching algorithms are time consuming for searching the optimal parameters of SVR model [29].

2.2.2. Neural network model

A neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes [37]. A simplest neural network is known as multi-layer perceptron (MLP) [38]. Moreover, neural networks are extremely popular for load forecasting, the review of different architectures can be found in [6] and [39]. The main reason for the preference of neural network models over other non-linear methods is the neural network models emerged as universal approximators. Furthermore, these methods can achieve high degree of prediction accuracy by approximating a large class of functions due to its ability to process the information parallelly [40]. There are also some promising features such as robustness, high learning capabilities and fault tolerance. However, neural network models are also have some shortcomings, such as selection of architecture according to complexity of problem, disruptive and unstable training, over-fitting, hyper-parameter selection and weak extrapolation capacity. The most demanding shortcoming of neural network is to find optimal architecture which include relevant input variables, hidden layer size, learning rate, momentum term.

2.2.3. Gaussian process model (GP)

A GP is a generalization of a multivariate Gaussian distribution to infinitely many variables. GP can also be considered as special case of Bayesian inference in which all the variables can be assumed to be Gaussian. Moreover, GP assumes that all the data is utilized to make a prediction and all the attributes can be represented by one joint distribution [32]. GPs are relatively recent when compared to other machine learning methods which are applied to non-linear modeling [41]. A GP is a generalization of a multivariate Gaussian distribution to infinitely many variables. GP can also be considered as special case of Bayesian inference in which all the variables can be assumed to be Gaussian. Moreover, GP assumes that all the data is utilized to make a prediction and all the attributes can be represented by one joint distribution. Further, covariance function of GP requires fewer hyper parameters in comparison to other advanced machine learning approaches such as neural network and SVR. GP is also suitable for the development of control strategies based on model as GP describes the distribution of new forecasts. These advantages leads to the implementation of GP for different areas such as load forecasting, [32] solar radiation [42] and remote sensing [43].

2.2.4. Long-short term memory model (LSTM)

LSTM architectures [44] is a deep neural architecture with long-term memory which considered as an improved architecture over recurrent neural networks (RNN) architecture. RNNs are able to process the input information as well as previously obtained information, due to its recurrent connections in the architecture. These re-current connections allows the direct processing of temporal dependencies for time series modeling. However, RNNs cannot capture the long temporal dependencies for long term time series prediction. This leads to the exploration of different LSTM architectures for the problem of time series forecasting [7] [45]. There is also one limitation with the LSTMs that these cannot perform well, when forecasts are dependent on recent past observations [45].

3. Comparison of parametric models to non-parametric models

This section reviews the some studies published in the recent years which are conduct for the comparison of parametric models and non-parametric models. The non-parametric models have been proposed as an alternative to parametric models in the academic literature for time series analysis, however, there is scant evidence available about relative performance of parametric and non-parametric models in terms of computational ability and accuracy. Many a times, it is reported that parametric models are more efficient than non-parametric models and vice versa. Although, this difference in performance is normally not that much of a problem, because it depends on the type of problem being addressed. Some researches demonstrated that non-parametric model for forecasting can outperform in chaotic stochastic dynamical systems [9] [46]. These studies highlights two main reasons for the failure of parametric models. The first one is the instability of parametric inference in chaotic systems [47] and the other one is the “true myth model” which is presented in [48] and showed the necessary gap of real problem with its simplified model representation. However, Hartig and Dormann [2013] argued to the previous claims regarding non-parametric models and presented a study which demonstrated parametric modeling also preferred for forecasting by applying theta-logistic example described in [48]. Moreover, Hartig and Dormann [2013] showcased the advantages of theoretical understanding and transferability offered by parametric modeling which lacks in non-parametric modeling. Furthermore, Jabot [2015] provide claims that parametric modeling favors forecasting by focusing on two comparative advantages over non-parametric modeling. First one is the diagnosis can be done using simple modeling procedures such as Bayesian approach for the likelihood of parametric modeling failure. Second one is the reliability of forecasts produced by parametric forecasting.

Recently, literally hundreds of different machine learning algorithms have been proposed, that claim the accuracy improvements and methodological advances for different time series forecasting problems [6] [7] [39]. However, machine learning models’ superiority proclamations are only based on single or few time series data which raises questions about the generality of produced outputs. These proposed machine learning models are tested for one-step ahead or short-term forecasts, without considering medium or long-term forecasts. In spite of that, scant evidences are available that certify performance claims of machine learning models as a forecasting tool [51-53]. In Adya and Collopy [1998] and Ahmed et.al [2010] the authors summarized numerous empirical studies and concluded that the outcome of all compared studies in terms of accuracy are somewhat mixed. Several similar empirical competitions are conducted for exploring machine learning methods such as NN3 [54] NN5 [55] and ESTSP [56] under different conditions, which results in debates among the scientific community for the area of time series forecasting [57]. Later, in Crone et.al [2011] the authors published the outcome of NN competitions and stated that no machine learning methods is able to outperform the Theta method [58] which is a statistical decomposition method (parametric method). Thus, the results obtained from objectively evaluated machine learning methods cannot be generalized and also not enough to persuade researchers and academicians to utilize machine learning methods over parametric methods for time series forecasting. Further, it has been shown that machine learning methods does not

automatically guarantees the improvement in accuracy of the produced forecasts [59]. Recently, Makridakis et.al [2018] also presented an empirical study which provides insights about the inability of non-parametric models for time series forecasting using M3 competition monthly time series dataset. This study reveals that the traditional parametric forecasting outperforms the non-parametric machine learning models. It also suggested that non-parametric model require to be more accurate and should be of less black box. To examine the computational ability and performance of parametric and non-parametric methods, a simulation study is presented in the next section.

4. Comparative analysis via simulation

A simulation study has been carried out to compare the relevant parametric models and machine learning models. To conduct the empirical comparison, 8 para-metric methods are selected namely: ES, AR, ARMA, ARIMA, seasonal ARIMA (SARIMA), SARIMA with exogenous variable (SARIMAX), ARMA with exogenous variable (ARMAX) and ARIMA with exogenous variable (ARIMAX). Further, 6 popular machine learning methods for prediction are chosen which are MLP, SVR, GP, LSTM, linear regression and regression trees. To test the aforementioned methods, 5 real time series datasets of different fields are selected from Kaggle website. The description of 5 real time series datasets is presented in Table 1.

Table 1. Summary of selected datasets

Name	Time Period	Type	Description
Oil	1960-2017	Monthly	Oil prices in USD from World banks global economic monitor.
AP	1949-1960	Monthly	US air passengers data
HEC	2004-2018	Hourly	Hourly electricity consumption data from American electrical power
S&P	2013-2018	Daily	Stock prices data of American airlines group
GT	1751-2015	Monthly	Global land average temperatures data

To test the forecasting performance, we obtained the 5-steps ahead forecasts for each selected datasets. Moreover, symmetric mean absolute percentage error (sMAPE) measure has been used for evaluating the prediction performance of time series forecasting methods. Mathematically, sMAPE can be defined as

$$sMAPE = \frac{2}{r} \sum_{t=1}^r \frac{|Y_t - \bar{Y}_t|}{|Y_t| + |\bar{Y}_t|} \times 100 \quad (3)$$

where r is the forecasting horizon, Y_t are the original observations and \bar{Y}_t are the produced forecasts by the time series model at time period t . The methods are implemented using MATLAB and Python. The comparisons are carried out on a workstation with 3.4 GHz processor and 32 GB RAM memory. Furthermore, the comparative results are presented

in Table 2. The 8 parametric methods: ES, AR, ARMA, ARIMA, SARIMA, ARMAX, ARIMAX and SARIMAX are represented as P1-P8 respectively in Table 2. Moreover, 6

machine learning methods SVR, linear regression, GP, Regression trees, MLP and LSTM are represented as M1-M6 respectively in Table 2.

Table 2. Forecasting performance (sMAPE) of widely used parametric and machine learning methods

Data	P1	P2	P3	P4	P5	P6	P7	P8	M1	M2	M3	M4	M5	M6
Oil	6.1	6.5	6.7	6.1	6.1	7.6	7.6	7.0	4.2	33.0	7.0	6.9	8.5	14.4
AP	18.1	28.1	27.2	28.4	17.6	26.2	2.9	26.3	13.6	12.7	24.3	15.6	9.5	33.8
HEC	2.2	1.3	0.9	5.0	5.0	5.0	1.5	6.7	36.0	33.3	3.5	6.5	23.2	5.4
S&P	3.4	5.0	5.1	5.44	5.3	5.6	5.3	5.6	2.2	2.4	4.8	3.7	2.3	11.0
GT	36.9	31.8	32.3	6.0	6.0	41.6	30.2	41.6	33.4	34.0	33.8	33.0	76.4	10.9

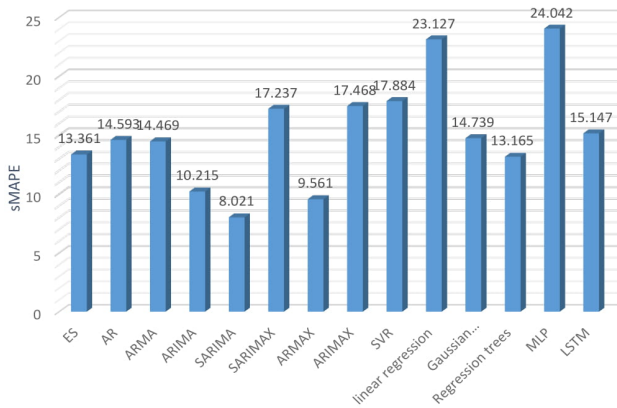


Fig. 2. Comparison of forecasting performance of parametric and machine learning methods.

The forecasting performance (in terms of sMAPE error) of each aforementioned 14 methods (both parametric and non-parametric) for predicting 5-steps ahead are displayed in Table 2 for five selected datasets. The minimum error achieved for each dataset is highlighted with a boldface. The results indicate that the parametric methods outperforms machine learning methods for 3 datasets (AP, HEC and GT). Further, for 2 datasets (Oil and S&P) parametric methods are competitive with machine learning methods. Thus, it is difficult to infer whether parametric metric is better than non-parametric methods or not. Furthermore, the averaged error for each 5 datasets for 14 methods is presented in Figure 2. The Figure 2 shows that the parametric methods have lower approximated error for all datasets in comparison to widely used machine learning methods. Lastly, it is important to note that parametric methods which are much simpler and computationally in-expensive than machine learning methods are highly competitive with compared machine learning methods. Therefore, the limitations of machine learning methods are highlighted in subsequent Section, which leads to the applicability of parametric methods for different forecasting problems [10-14] during the outburst of machine learning methods.

5. Reasons for applicability of parametric methods in machine learning era

In recent years, non-parametric models gained popularity in the field of time series forecasting due to its ability to capture subtle patterns in time series data [6] [7] [39]. However, it is evident from the literature that the parametric models still are widely applicable for time series data obtained from real world problems such as electricity load forecasting [10] [11]

[61] [62], demand of different quantities such as coal and gold [12] [63] [64], healthcare [65] and environmental problems [13] [14] [66] [67]. The researchers and academicians choose parametric models over machine learning models for these time series forecasting problems in during the outburst of machine learning methods is due to the following reasons:

- (1) The machine learning models are of black-box type, thus it has been difficult to understand how the forecasts are generated using machine learning models. Moreover, obtaining numbers from a black box is not acceptable to practitioners who need to know how forecasts arise and how they can be influenced or adjusted to arrive at workable predictions [68].
- (2) The machine learning methods have high computing requirement than parametric forecasting methods.
- (3) The machine learning methods require specialist knowledge for its implementation as compared to statistical time series forecasting methods.
- (4) Due to randomness of time series data, machine learning methods should be able to differentiate patterns from noise in the time series data, so over-fitting can be avoided. In contrast, parametric methods can directly control over-fitting by applying some information criteria such as AIC or BIC [60].
- (5) Mostly machine learning methods require preprocessing of the input data which specifically needed attention for the proper transformation selection. However, there is no best transformation available in the literature that can be utilized to automate the learning process and probably improve the forecasting accuracy [40].
- (6) Machine learning methods are unable to address the uncertainty in forecasts or defining the confidence intervals for them. To overcome this issue, many researchers propose simulating the intervals by iteratively generating multiple future sample paths. Yet, even in that case, the forecast distribution of the methods is empirically and not analytically derived, raising many doubts about its quality [60].
- (7) Too much effort is required to select optimal architectures of machine learning models. Moreover, these methods also need to ensure proper hyper-parameter selection for particular time series forecasting problems [36] [69].

Usually, machine learning methods are expected to perform better but these sophisticated methods are not able to surpass much simpler parametric methods (see Table 2).

However, it is little disappointing from scientific view as machine learning methods require specialist knowledge and ample time for implementation.

6. Recommendations for future research

There is no critical evidence about the clear superiority of parametric over non-parametric time series forecasting models and vice versa [9] [46] [49] [50] [60]. Therefore, some researchers have tried to incorporate the strengths of both parametric and non-parametric models to time series forecasting [70-72]. These studies have reported the success of hybrid methods. Thus, we encourage the researchers to further explore the evolving machine learning models with the parametric models for forecasting and to assess the reliability of generated forecasts. Moreover, a recent empirical study [60] suggested the need for unbiased and specific ways to evaluate the performance of time series forecasting models which can be accomplished by comparing open competitions. Finally, we recommend designing of the machine learning methods which particularly include the time dependency and volatility of time series data.

7. Conclusion

Predicting a time series has been a very challenging problem owing to the fact that the higher uncertainty exists in larger forecasting horizon. Since 1960, researchers put their significant effort in developing sophisticated methods, frameworks, and accuracy or error measures for time series forecasting. Mainly, the time series methods are categorized as parametric and non-parametric methods. The non-parametric methods have been proposed in the research

literature as an alternative to para-metric methods for time series forecasting. However, scarce evidence is available about the relative performance and computational ability of both parametric and non-parametric methods. This paper reviews the comparative studies conducted for evaluating the accuracy of parametric and non-parametric methods, especially machine learning methods. Furthermore, a simulation study has been conducted on real time series data to compare the performance of popular parametric and machine learning methods for different forecasting problems. The empirical study has indicated that the simple parametric methods are highly competitive with the machine learning methods. The simulation study supports our intuition that the parametric methods are highly competitive with the machine learning methods which require more expertise and demand more computational power. Although, recently proposed hybrid methods and semi-parametric methods which uses machine learning concepts provided very impressive results for M4 competition [73]. The authors are not denying the potential of machine learning methods, however more research is needed to develop the efficient machine learning methods for time series data. As structural instabilities, noise and uncertainty exist in time series data and therefore, should be considered before modeling time series data with machine learning methods that might interfere in obtaining optimal weights. Additionally, the limitations of the machine learning methods are highlighted which leads to the selection of parametric methods over non-parametric methods by the researchers during the outburst of machine learning methods.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License



References

1. P. Momani and P. Naill, American Journal of Environmental Sciences 599 (2009).
2. J. Yu, Applied Financial Economics 12, pp.193-202 (2002).
3. X. Zhongjie, Case studies in time series analysis, volume 3 (World Scientific, 1993).
4. B. J. West, Physiology, promiscuity, and prophecy at the millennium: A tale of tails (World Scientific, 1999).
5. G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time series analysis: forecasting and control (John Wiley & Sons, 2015).
6. B. Yildiz, J. I. Bilbao and A. B. Sproul, Renewable and Sustainable Energy Reviews 73, pp. 1104-1122 (2017).
7. L. Deng, APSIPA Transactions on Signal and Information Processing 3 (2014).
8. J. Fan and Q. Yao, Nonlinear time series: nonparametric and parametric methods (Springer Science & Business Media, 2008).
9. C. T. Perretti, G. Sugihara and S. B. Munch, Ecology 94, pp. 794-800 (2013).
10. J. W. Taylor and M. B. Roberts, IEEE Transactions on Power Systems 31, pp. 1925-1932 (2016).
11. E. M. de Oliveira and F. L. C. Oliveira, Energy 144, pp. 776-788 (2018).
12. S. A. David, J. Machado, L. R. Trevisan, C. Inacio Jr and A. M. Lopes, Fundamenta Informaticae 151, pp. 389-408 (2017).
13. A. Ahmad, T. Anderson and T. Lie, Solar Energy 122, pp. 1398-1408 (2015).
14. A. Pohoata and E. Lungu, REVISTA DE CHIMIE 68, pp. 818-823 (2017).
15. C. C. Holt, International journal of forecasting 20, pp. 5-10 (2004).
16. R. G. Brown, Smoothing, forecasting and prediction of discrete time series (Courier Corporation, 2004).
17. P. R. Winters, Management science 6, pp. 324-342 (1960).
18. J. F. Muth, Journal of the american statistical association 55, pp. 299-306 (1960).
19. G. E. Box and G. M. Jenkins, Time series analysis: forecasting and control, revised ed (Holden-Day, 1976).
20. B. Abraham and J. Ledolter, International Statistical Review/Revue Internationale de Statistique pp. 51-66 (1986).
21. E. S. Gardner Jr and E. McKenzie, Management Science 31, pp. 1237-1246 (1985).
22. R. Snyder, Journal of the Royal Statistical Society. Series B (Methodological) pp. 272-276 (1985).
23. I. Gijbels, A. Pope and M. P. Wand, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61, pp. 39-50 (1999).
24. J. W. Taylor, Journal of the Operational Research Society 54, pp. 799-805 (2003).
25. G. U. Yule et al., Phil. Trans. R. Soc. Lond. A 226, pp. 267-298 (1927).
26. E. Slutsky, Econometrica: Journal of the Econometric Society pp. 105-146 (1937).
27. A. M. Walker, Biometrika 49, pp. 117-131 (1962).
28. W.-C. Hong, Y. Dong, W. Y. Zhang, L.-Y. Chen and B. Panigrahi, International Journal of Electrical Power & Energy Systems 44, pp. 604-614 (2013).
29. W.-C. Hong, Y. Dong, F. Zheng and C.-Y. Lai, Applied Mathematical Modelling 35, pp. 1282-1291 (2011).
30. C. Gonzalez, J. Mira-McWilliams and I. Juarez, IET Generation, Transmission & Distribution 9, pp. 1120-1128 (2015).
31. C. Persson, P. Bacher, T. Shiga and H. Madsen, Solar Energy 150, pp. 423-436 (2017).
32. J. R. Lloyd, International Journal of Forecasting 30, pp. 369-374 (2014).

33. P. F. Smith, S. Ganesh and P. Liu, *Journal of neuroscience methods* 220, pp. 85-91 (2013).
34. W.-M. Hung and W.-C. Hong, *Control & Cybernetics* 38 (2009).
35. W.-C. Hong and P.-F. Pai, *The International Journal of Advanced Manufacturing Technology* 28, pp. 154-161 (2006).
36. P.-F. Pai, W.-C. Hong and Y.-S. Lee, *International Journal of Operations Research* 2, pp. 1-7 (2005).
37. J. E. Menke and T. R. Martinez, *Intelligent Data Analysis* 13, pp. 135-149 (2009).
38. R. L. Welch, S. M. Ru ng and G. K. Venayagamoorthy, *Comparison of feedforward and feedback neural network architectures for short term wind speed prediction*, in 2009 International Joint Conference on Neural Networks, pp. 3335-3340 (IEEE, 2009).
39. M. Q. Raza and A. Khosravi, *Renewable and Sustainable Energy Reviews* 50, pp. 1352-1372 (2015).
40. G. P. Zhang and M. Qi, *European journal of operational research* 160, pp. 501-514 (2005).
41. C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, volume 2 (MIT press Cambridge, MA, 2006).
42. C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte and A. Fouilloy, *Renewable Energy* 105, pp. 569-582 (2017).
43. L. Pasolli, F. Melgani and E. Blanzieri, *IEEE Geoscience and Remote Sensing Letters* 7, pp. 464-468 (2010).
44. J. Wang, L. Zhang, Y. Chen and Z. Yi, *International journal of neural systems* 28, 1750061 (2018).
45. D. Hsu, arXiv preprint arXiv:1707.00666 (2017).
46. C. T. Perretti, S. B. Munch and G. Sugihara, *Proceedings of the National Academy of Sciences*, 201216076 (2013).
47. S. N. Wood, *Nature* 466, 1102 (2010).
48. C. T. Perretti, S. B. Munch and G. Sugihara, *Proceedings of the National Academy of Sciences* 110, E3976-E3977 (2013).
49. F. Hartig and C. F. Dormann, *Proceedings of the National Academy of Sciences* 110, E3975- E3975 (2013).
50. F. Jabot, *Journal of theoretical biology* 372, pp. 205-210 (2015).
51. M. Adya and F. Collopy, *Journal of forecasting* 17, pp. 481-495 (1998).
52. N. K. Ahmed, A. F. Atiya, N. E. Gayar and H. El-Shishiny, *Econometric Reviews* 29, pp. 594-621 (2010).
53. S. F. Crone, M. Hibon and K. Nikolopoulos, *International Journal of forecasting* 27, pp. 635-660 (2011).
54. S. Crone and K. Nikolopoulos, *Nn3 artificial neural networks & computational intelligence forecasting competition-motivation*, (2007).
55. S. Crone, *Nn5 forecasting competition for arti cial neural networks & computational intelligence*, (2008).
56. A. Lendasse, *Multiprint Oy/Otamedia* (2007).
57. S. F. Crone, *International Journal of Forecasting* 25, pp. 456-460 (2009).
58. V. Assimakopoulos and K. Nikolopoulos, *International journal of forecasting* 16, pp. 521-530 (2000).
59. G. P. Zhang, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, pp. 3-16 (2007).
60. S. Makridakis, E. Spiliotis and V. Assimakopoulos, *PLoS one* 13 (2018).
61. Z. Yang, L. Ce and L. Lian, *Applied Energy* 190, (2017).
62. P. M. Macaira, R. C. Sousa and F. L. C. Oliveira, *IEEE Latin America Transactions* 14, pp. 1252-1258 (2016).
63. S. Jiang, C. Yang, J. Guo and Z. Ding, *Energy Sources, Part B: Economics, Planning, and Policy* 13, pp. 190-195 (2018).
64. H. Hassani, E. S. Silva, R. Gupta and M. K. Segnon, *Applied Economics* 47, pp. 4141-4152 (2015).
65. D. Lawrence and J. M. Williams, *Nicotine & Tobacco Research* 18, pp. 1463-1470 (2015).
66. K. Yunus, P. Chen and T. Thiringer, *IET Renewable Power Generation* 11 (2016) 132.
67. J.-Y. Zhang, D.-X. Li and T.-L. Huang, *Heavy metal pollution prediction based on hybrid arima and grey model*, in *Materials, Manufacturing Technology, Electronics and Information Science (MMTEI2015)* pp. 19-29 (World Scientific, 2016).
68. J. Friedman, T. Hastie and R. Tibshirani, *The elements of statistical learning*, volume 1 (Springer series in statistics New York, 2001).
69. M. Khashei, *Master of Science Thesis*, Isfahan University of Technology (2005).
70. M. Ghofrani, A. Arabali and M. Ghayekhloo, *Electric Power Systems Research* 117, pp. 134-142 (2014).
71. R. T. Fleifel, S. S. Soliman, W. Hamouda and A. Badawi, *Lte primary user modeling using a hybrid arima/narx neural network model in cr*, in 2017 IEEE Wireless Communications and Networking Conference (WCNC) pp. 1-6 (2017).
72. S. Voronin and J. Partanen, *International Journal of Energy Research* 38, pp. 626-637 (2014).
73. S. Makridakis, E. Spiliotis and V. Assimakopoulos, *International Journal of Forecasting* 34, pp. 802-808 (2018).