Conference Paper No: PF-06

A hybrid model for wind speed prediction in Anuradhapura, Sri Lanka

K. M. Lenora[1*], S. P. Abeysundara[2] and K. Perera[3]

[1]Postgraduate Institute of Science, University of Peradeniya, Sri Lanka
[2]Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka
[3]Department of Engineering Mathematics, Faculty of Engineering, University of Peradeniya, Sri Lanka
kushanilenora092@gmail.com*

Abstract

Wind energy plays a major role in a sustainable future as a useful, environmentally friendly energy alternative. Wind speed is the most important parameter in the design and implementation of wind energy. This paper aims to define a methodology capable of providing accurate monthly average wind speed predictions in the Anuradhapura region, Sri Lanka. Hybrid forecasting of time series is considered to be a potentially effective alternative compared with the conventional stand-alone forecasting modeling approaches like seasonal autoregressive integrated moving average (SARIMA) and artificial neural network (ANN). In this study, at first, SARIMA and ANN models are used to separately recognize and forecast the linear and nonlinear components of time series, respectively. Then, the study suggests a hybrid approach combining SARIMA and ANN for forecasting wind speed and its forecasting results are compared with the single SARIMA and ANN models. The mean absolute error (MAE), root mean square error (RMSE), and paired sample *t*-test are used as performance measures. Results obtained by a case study show that the SARIMA-ANN hybrid approach is the most suitable for wind speed forecasting. This approach demonstrates the potential to be applied to wind speed forecasting in other regions of the country.

Keywords: Artificial neural networks (ANN), Hybrid approach, Seasonal autoregressive integrated moving average (SARIMA), Wind speed

Introduction

The wind speed affects decisions related to agriculture, aviation and maritime, constructions, urban air pollution management, and many other important domains. Recently, as **a** renewable energy source, wind energy source has received considerable attention worldwide. Moreover, wind energy plays a major role in a sustainable future as a green energy alternative. Accurate forecasting and estimation of wind speed are more important if wind energy is to reach its full potential**.** However, accurate and reliable wind speed forecasts become a challenging task due to its stochastic, nonlinear**,** random, indeterminant, discontinuous, and fluctuating nature (Xu et al., 2017;Duan & Liu, 2019; Aasim et al., 2019).

In general, statistical methods such as auto regressive moving average (ARMA), auto regressive integrated moving average (ARIMA) and seasonal auto regressive integrated moving average (SARIMA) have been extensively utilized for wind speed prediction (Haddad et al., 2020; Grigonytė, 2021;Torres et al., 2005). However, the major limitation of such models is their ability to capture only the linear form of time series data under the normality, linearity and stationary assumptions.

Artificial neural networks (ANNs) overcome the limitation of ARIMA. The major advantages of neural networks are their nonlinear modeling capability and the fewer number of assumptions on data prior to the model building process. In wind speed literature, the ANNs have been used for forecasting for many years with impressive results, which have proven to be the better approach for wind speed forecasting. For instance, (Li & Shi, 2010) investigated three types of typical ANN, namely, adaptive linear element, back propagation, and radial basis function, for wind speed forecasting. According to the compared results of three types of ANN show that even for the same wind dataset, no single ANN model outperforms others universally in terms of all evaluation metrics.

In order to overcome the limitations of each of these models, a hybrid methodology that includes both linear and nonlinear modelling capabilities can be one of the best approaches. Once the pipeline of the model is properly trained these types of models show optimal forecasting performance (Shi et al., 2012). Two hybrid models; ARIMA with ANN and a Kalman filter (KF) were proposed by (Liu et al., 2012). The results are showed that the hybrid models have superior outcomes than the classical ARIMA model. Moreover, (Shi et al., 2012) proposed two hybrid models, namely, ARIMA-ANN and ARIMA-SVM, for wind speed and power forecasting. The study investigated the applicability of the proposed hybrid models based on wind speed and wind power generation, respectively. The results indicated that the hybrid approaches are viable options for forecasting both wind speed and wind power generation time series.

The major objective of the present study is to propose an accurate and efficient forecast model to predict the future behaviors of wind speed. This is achieved by developing SARIMA models, ANNs and hybrid methods to forecast the wind speed. Finally, based on the most accurate and efficient forecast model will be determined using the model validation and the selection criteria.

This paper is organized as follows. Materials and methods section describe the data set and the theoretical background of the time series forecast models. Results and discussion section discuss the significant results of the data analysis. Finally, conclusions section presents the conclusion of the study.

Methodology and methods

In this study, seasonal autoregressive integrated moving average (SARIMA), artificial neural networks (ANNs) and hybrid method have been used in forecasting monthly wind speed. The methodology for analyzing data consists of the following three steps. which are:

Step 1: Modelling of SARIMA using Box-Jenkins procedure.

Step 2: Modelling of ANNs with three types of input variables, which are based on seasonal and non-seasonal lag orders of SARIMA model.

a. The inputs based on the order of the SARIMA model
b. The inputs based on the seasonal lag.
c. The inputs based on the lag 1 and seasonal lags ± 1

Step 3: Modelling of SARIMA-ANN hybrid models

Finally, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and paired sample t-test are used as performance measures to find the suitable model for forecasting wind speed.

SARIMA model

SARIMA was proposed by box and Jenkins in 1976 as an extension of the well-established ARIMA model (Janacek, 2009). It is a most widely used linear time series forecasting method for univariate time series data that contains trends and seasonality. The $SARIMA(p,d,q)(P,D,Q)_S$ model can be written in the general form as follows:

$$\phi_p(B)\Phi_P B^s(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)e_t \quad (1)$$

where B is denoted as the backward shift operator, d and D are denoted as the non-seasonal and seasonal orders of difference respectively, ϕ is the non-seasonal AR polynomial of order p, ϕ is the seasonal AR polynomial of order P, ϕ is the regular MA polynomial of order q, Θ is the seasonal MA polynomial of order Q, Y_t is the observed value at time t and e_t is the residual value at time t.

Based on the Box-Jenkins methodology, the SARIMA approach provides the main four steps of identification, estimation, and validation and prediction (Society, 2016).

ANN model

ANN is a nonlinear model which maps a set of input variables through several layers of processing elements or neurons into a set of output variables (Faraway & Chatfield, 1998). Generally, in a time series modelling ANN input consist of previous value of observations while the output is the observation at time t. The following equation illustrate the relationship between the input variables and output variable in a neural network.

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j\, g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij}\, y_{t-i}\right) \quad (2)$$

where y_t is the output variable, α_0 is the bias of the hidden layer, α_j is the weights of the hidden layer weight, $g(x)$ is the activation function, β_{0j} input bias, β_{ij} is the input weight and y_{t-i} is the lag input variables.

Hybrid (SARIMA-ANN) model

The SARIMA and ANNs both have their own strength to analyze their own pattern data, where SARIMA is suitable for linear pattern while ANNs are suitable for nonlinear pattern. (Zhang, 2003) has developed a hybrid approach that is described under the two phases based on their linear and non-linear behaviors. According to Zhang's hybrid methodology, time series Y_t can be defined as,

$$Y_t = L_t + N_t \quad (3)$$

where Y_t is the time series observation at time t, $L_t$is the linear component of the time series, N_t is the non-linear component of the time series. The hybrid approach consists of two steps. First, the SARIMA approach is mainly used to analyze the linear component of the time series. Then we assume that the residual from the linear model will contain only the non-linear behavior of the time series. Next, the ANN-based approach was applied to capture the nonlinear component of the series. The forecasted value of the hybrid model is the sum of the forecasted value of the SARIMA and the ANN models.

Data

The data employed in this study were collected from the Department of Meteorology, Colombo, and represent the monthly average wind speed in Anuradhapura region from January 1995 through December 2019.

Results and discussion

The time series plot of the original observations for the period 1995-2019 is shown in Figure 1. As an initial step of the study, Box-Jenkins procedure is applied to find the suitable SARIMA model for the time series. The results suggested that, $SARIMA\ (1,0,2) \times (1,1,2)_{12}$, with AIC value of 808.319 is the most suitable model for forecasting time series

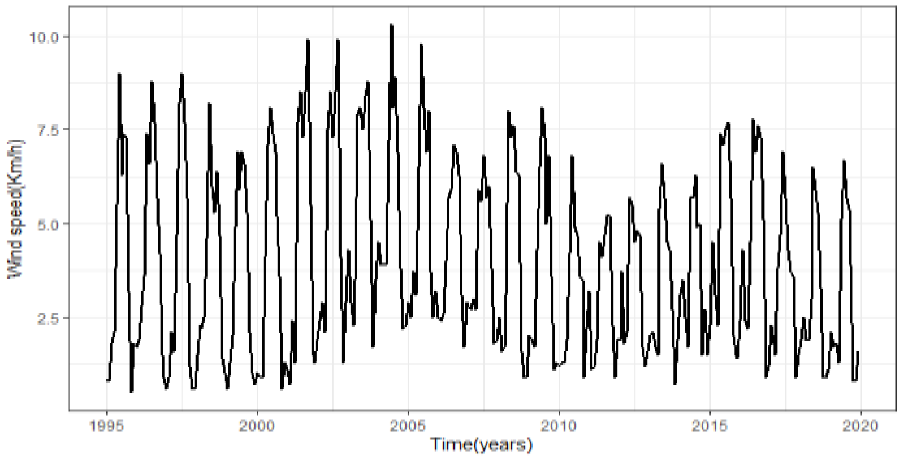

Figure 1. *Time Series plot for Monthly wind speed figures from period 1995-2019*

After selecting the best model, diagnostic checking was carried out. Jarque-Bera test was used to check normality, Ljung-Box test was used to check the autocorrelation and heteroskedasticity ARCH test was used to check for ARCH effect. The Jarque-Bera test results indicated that, the residuals are normally distributed (p-value = 0.124). Ljung–Box test for autocorrelation up to the maximum lag order of 48 was 29.1 at the p-value of 0.36. Test results concluded that the residual has no autocorrelation. Furthermore, the Heteroskedasticity ARCH test results indicated that the model hasn't an ARCH effect (p-value = 0.408)

Next, an ANN with three types of input nodes based on SARIMA and single hidden layer with up to 10 hidden nodes is fitted.

***Table 1.** Model accuracy measures for wind speed data*

Models	RMSE	MAE
$SARIMA\ (1,0,2) \times (1,1,2)_{12}$	0.5894	0.4808
ANN		
All Lags	0.7855	0.5179
Seasonal Lags 12,24	0.5575	0.4630
Lag 1 with seasonal Lag ± 1	1.4170	1.0733
Hybrid		
All Lags	0.5102	0.4271
Seasonal Lags 12,24	0.4680	0.3481
Lag 1 with seasonal Lag ± 1	0.5595	0.4900

The fitted model is used to forecast the wind speed using hybrid methodology and the corresponding results are summarized in Table I. According to the error analysis results, the proposed SARIMA-ANN hybrid model is the most suitable model with the lowest RMSE error. Moreover, MAE that the hybrid model is more significant than stand-alone SARIMA and ANN models for the speed forecasting.

In terms of RMSE, the percentage decrements of the hybrid model over the SARIMA and ANN are 20.59% and 16.05%, respectively. In terms of MAE, the percentage decrements of the hybrid model over the SARIMA and ANN are 27.60% and 24.82%, respectively. Thus, the hybrid model is better in wind speed forecasting.

The 12-month forecasted values for the test data of the selected SARIMA, ANN, and SARIMA-ANN hybrid model is represented in Figure 2. The mean absolute percentage error (MAPE) of the SARIMA, ANN and SARIMA-ANN hybrid model are 27.93% ,19.46% and 17.66%, respectively. The paired sample t-test (at 5% level of significance) was carried out to validate any significant differences between the actual values and predicted values of each of the models. Results revealed that there is no significant difference between the forecasted and actual values of wind speed for SARIMA-ANN hybrid model (p-value = 0.8625).

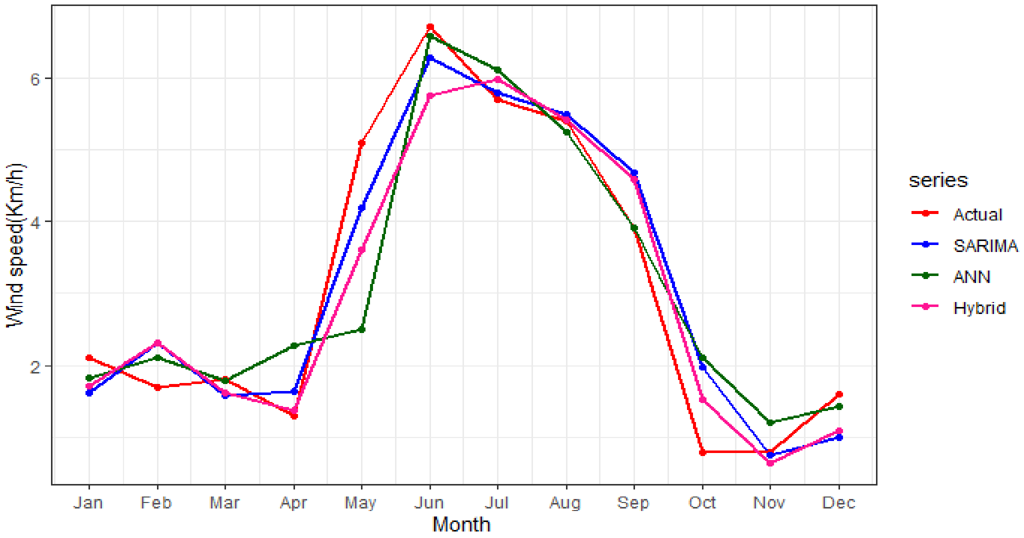

Figure 2. *Actual vs. predicted values for selected SARIMA, ANN, and hybrid model*

Conclusions

The study considered forecasting the wind speed based on three models such as ANN, SARIMA and Hybrid model. A hybrid model comprising ANN and ARIMA is proposed for wind speed forecasting. The mean absolute error (MAE), root mean square error (RMSE), and paired sample *t*-tests are used to compare the performance of all the developed models. Results showed that the SARIMA-ANN hybrid approach is the most suitable for forecasting the wind speed.

References

Aasim, Singh, S. N., & Mohapatra, A. (2019). Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renewable Energy*, *136*, 758–768. https://doi.org/10.1016/j.renene.2019.01.031

Duan, Z., & Liu, H. (2019). An evolution-dependent multi-objective ensemble model of vanishing moment with adversarial auto-encoder for short-term wind speed forecasting in Xinjiang wind farm, China. *Energy Conversion and Management*, *198*(July), 111914. https://doi.org/10.1016/j.enconman.2019.111914

Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: A comparative study using the airline data. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *47*(2), 231–250. https://doi.org/10.1111/1467-9876.00109

Grigonytė, E. (2021). *Short-term wind speed forecasting using ARIMA model. July 2016*. https://doi.org/10.6001/energetika.v62i1-2.3313

Haddad, M., Nicod, J., Mainassara, Y. B., Masry, Z. Al, Péra, M., Haddad, M., Nicod, J., Mainassara, Y. B., Rabehasaina, L., & Masry, Z. Al. (2020). *Wind and Solar Forecasting for Renewable Energy System using SARIMA-based Model To cite this version : HAL Id : hal-02867736 Wind and Solar Forecasting for Renewable Energy System using SARIMA-based Model*.

Janacek, G. (2009). Time Series Analysis Forecasting and Control. *Journal of Time Series Analysis*, *2009*(December 2009), 9892. https://doi.org/10.1111/j.1467-

9892.2009.00643.x

Li, G., & Shi, J. (2010). On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, *87*(7), 2313–2320. https://doi.org/10.1016/j.apenergy.2009.12.013

Liu, H., Tian, H., & Li, Y. (2012). Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Applied Energy*, *98*, 415–424. https://doi.org/10.1016/j.apenergy.2012.04.001

Shi, J., Guo, J., & Zheng, S. (2012). Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renewable and Sustainable Energy Reviews*, *16*(5), 3471–3480. https://doi.org/10.1016/j.rser.2012.02.044

Society, A. (2016). *American Society for Quality Review Author (s): Beat Kleiner Review by : Beat Kleiner Published by : Taylor & Francis , Ltd . on behalf of American Statistical Association and American Society for Quality Stable URL : http://www.jstor.org/stable/1267710 Accessed : 27-06-2016 04 : 44 UTC Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use , available at*. *19*(3), 343–344.

Torres, J. L., Blas, M. De, Francisco, A. De, & Garcı, A. (2005). *Forecast of hourly average wind speed with ARMA models in Navarre (Spain)*. *79*, 65–77. https://doi.org/10.1016/j.solener.2004.09.013

Xu, Y., Ren, Q., Zheng, Z. J., & He, Y. L. (2017). Evaluation and optimization of melting performance for a latent heat thermal energy storage unit partially filled with porous media. *Applied Energy*, *193*, 84–95. https://doi.org/10.1016/j.apenergy.2017.02.019

Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0