Data Mining 1 (2020) 1-6 Submitted 14/12; Published 21/12

Wind Speed Forecasting Using Artificial Intelligence And Box-Jenkins Methods

Abdellah Djaadi ABDELLAH.DJAADI@STUD.MIF.VU.LT
Data Science
Modelling and Data Analysis
Faculty of Mathematics and Informatics

Advisor: Jurgita Markevičiūtė

1. Abstract

The threat of global climate change is increasing every day; this fact has led the world to search for an alternative energy resources, which is renewable energy (RE). Moreover, energy generation has been continuously increasing during recent years. Wind and solar have had the most significant growths among all renewable resources. These resources have the potential to help alleviate the global climate change threat. The ARIMA and artificial neural networks methods have showen a vital role in predicting the wind power which has a significant effect on the production. The purpose of this paper is compare the predictive performance of these two methods.

Keywords: Prediction, ARIMA, ANN.

2. Introduction

The wind power is one of the most attractive renewable energy technologies in the world because of its high efficiency and low pollution (Chang, 2013). The advantages of RE are that it is low in pollution and sustainable (Tsai et al., 2017). Moreover, The existence of air, sunlight and other resources on earth motivate the government to use use in an appropriate way for human. Because, the climate change is rising everyday. Furthermore, the world economy is strongly related on the effective ways of electrical power generation (Johansson et al., 1993). The wind energy represent the most interesting solution to resolve the world energy challenges. In the past decade, many countries planned to shift from traditional energy to renewable energy(RE)(**??**). According to REN21's 2017 report, Renewable power generating capacity saw its largest annual increase ever in 2016, with an estimated 161 gigawatts (GW) of capacity added Total global capacity was up nearly 9% compared to 2015, to almost 2,017 GW at year's end (Pillot et al., 2019). This shows that it has developed very rapidly in last few years. The wind speed often is influenced by natural conditions therefore , it exhibits strong seasonal variations (Tsai et al., 2017). Thus, the wind speed forecasting is very important. Also, it is recommended to select the best location for the wind turbines, and performance prediction. Wind speed can be forecasted by performing the traditional autoregression models like ARIMA (Rehman and Halawani, 1994) or the

ANN (Mohandes et al., 1998).

3. Literature review

Nowdays, many scientific articles have been written which have same proposition as our paper. On this literature review will cover several articles about renewable energy in general and particularly, the wind power.
In (Mohandes et al., 2004) article, the researchers used Mean daily wind speed data from Madina city in Saudi Arabia. They introduce the support vector machines (SVM) and multilayer perceptron (MLP) methods. Few preprocessing techniques were used on their data set, such as the cross validation technique of machine learnig were performed to obtain reasonable parameters for both methods. The team tried to compare the prediction perfomrance of these two methods.
The wind farm of Kaberten in Algeria became a recommendation of the public power consumption.One of the wind energy potential studies in this region were performed by (Mohamed et al., 2015). The author talked about the distribution of the wind turbines, where the orientation of these turbines in Kaberten is based on hourly measures of speed and wind.
Besides, in the same wind frame, the team of Gasmi et al. (2018) performed the neural networks ANN and the autoregressive integrated moving average ARIMA for forcasting wind power production using time series analysis. The purpose of their work was to compare the prediction performance of these models. Results show that the ANN perform better than ARIMA.
(Palomares-Salas et al., 2009) performed the ARIMA model to predict time-series involving wind speed measurements. The paper covers the validation model process, and a regression analysis. The ARIMA model show better results back propagation neural network, and particularly, in prediction short time-intervals.
Besides, (Erdem and Shi, 2011) the aim of paper was to predict wind speed and direction, they introduced four approaches based on ARMA method. The component model is better at forecasting the wind direction than the traditional-linked ARMA model as the results show. Also, (Chang, 2013) presented ANN method based on back propagation neural network and described a wind power prediction methodologies. Results showed a good accuracy for short-term wind forecasting.
Besides,(Li and Shi, 2010) presented three types of typical ANN, which are back propagation, adaptive linear element, and radial basis function, to predict the wind speed. The proposed methodologies were performed in the same wind dataset and the results show that no single ANN model outperforms others according to the evaluation metrics.
In (Zeng and Qiao, 2011) paper, the researches team described a SVM-based method for wind power prediction. They used real wind speed and wind power data which recorded by the National Renewable Energy Laboratory. Results illustrate that the SVM method has good results.
Also, (Zeng and Qiao, 2011) introduced a new short-term forecasting method based on the application of evolutionary optimization algorithms for the automated specification of neu-

ral networks and the nearest neighbour search. The test results showed that the wind power prediction error can be reduced by using the proposed automated specification method.

4. Data

4.1 Data set

The dataset is about Wind Turbines's scada system that is working and generating power in Turkey. The data were measured at every 10 minutes intervals for about 12 months starting from January 2018 to December 2018. It contains 50530 observations with 5 variables. the attributes are wind speed, wind direction, generated power, etc. The majority of attributes will not be relevant to this paper, so we will focus on 2 of them(wind speed will be noted as wind s and generated power as generated p). The wind speed is the most important variable that affect a wind turbine's output as we will see later in this paper.

Table 1: interpretation of all attributes.

Features	**Description**	**Type**
Date/Time	the observations were recorded every 10 minutes	numeric
LV ActivePower (KW)	The power generated by the turbine at that moment	numeric
Wind Speed (m/s)	The wind speed at the hub height of the turbine	numeric
TheoreticalPowerCurve (KWh)	The theoritical power values at that moment	numeric
Wind Direction (°)	The direction of the wind power	numeric

4.2 Exploratory analysis and data pre-processing

The pre-processing step is essential in analysing the data. It helps to get better interpretation from the data set and, it improves the accuracy of the model. In this paper, we will keep only two variables; wind speed (Wind Speed (m/s)) and generated power (LV ActivePower (KW)). Since the other attributes do not fit the purpose of this paper. Also, our data set has 50530 observations and we will generate a new data set that has 2500 features only to get better results. Furthermore, we tried to detect the missing values and spot the outliers then, we removed them.

Table 2: Descriptive statistics of the attributes

D_Stats	Wind Speed	Power Generated
Min	0.000	-2.741
1st Qu.	4.201	50.678
median	7.105	825.838
Mean	7.558	1307.684
3rd Qu.	10.300	2482.508
Max	25.06	3618.733

This data set has no missing values but some outliers are existed as the Box plot shows in **Figure 1**. The outliers of the wind speed are the cases when the wind speed is less than the min 4.201 m/s, because there will be now generated power as we will see later in Figure3. Also, when the wind speed is greater then 18.4 (m/s) since this case might cause damages to the turbines. However, the generated power, it is strange the power has negative values.

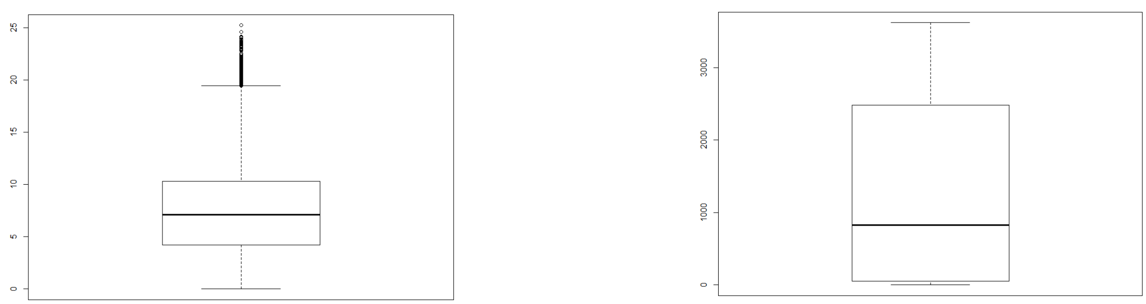

Figure 1: Box plot of Wind_S and Power_G

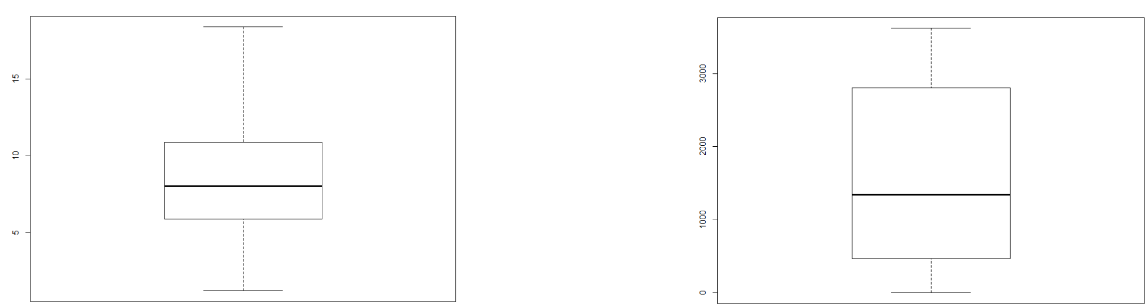

Figure 2: New box plot of Wind_S and Power_G

Furthermore, we applied the correlation test to find the association between the variables. The result was approximately **0.83**. This value means that there is high positive correlation between the variables. Another way to phrase this statement would be to say that wind with high speed tends to have higher energy production as this graph illustrate:

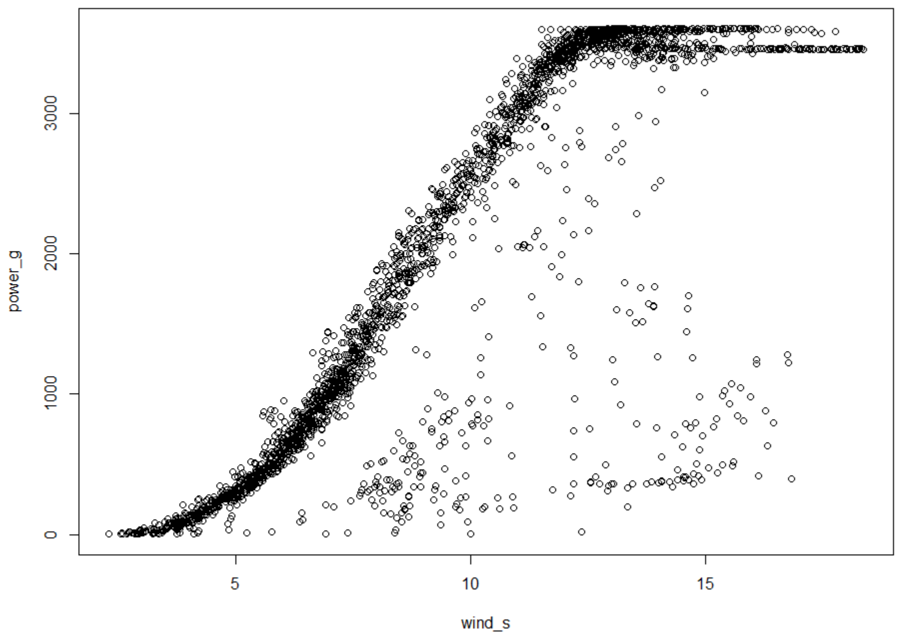

Figure 3: The correlation between the Wind_S and Power_G

5. Methdology

At this stage, we will introduce the methods which are used in this paper and, show the results of each method.

5.1 ARIMA

The auto regressive integrated moving average (*ARIMA*). It is one of the statistical method and it can be used in engineering, economics and natural sciences to solve the problems that have a great deal of data where the observations. are interdependent (Chang et al., 2014).

ARIMA models were popularized by Box and Jenkins in 1970 (Box et al., 2011). It have been used several times in time series forecasting problems, because they are simple as well as easy to understand and implement. The disadvantage of this method is that the prediction error increases as the prediction time increases.(Chang et al., 2014).In *ARIMA* model, the next value of a variable is assumed to be a linear function of several previous observations and random errors. Denoted as *ARIMA(p; d; q)* or expressed as:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j}$$

$$\phi(B)\left(1 - B^d\right) y_t = \theta(B) e_t$$

where yt is the actual value , e_t is a white noise process with mean zero and variance σ^2, B is the backshift operator (Brockwell and Davis, 2013).

5.2 Neural network model

Recently, the science has seen a great development in artificial intelligence (AI), several new AI methods for wind speed and power prediction have been developed. .
Neural networks are widely use to deal with the non-linear modelling approach that provides an approximation to any function. For time series prediction, it has this following form;

$$y_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i f\left(\sum_{j=1}^{n} \beta_{ij} y_{t-j} + \beta_{0j}\right) + \varepsilon_t$$

where n is the number of input nodes, m is the number of hidden nodes, f is a sigmoid transfer function: $f(x) = \frac{1}{(1+e^{-x})}$, $(\alpha_i)_{1\leq i\leq m}$ weights from the hidden to output nodes and $(\beta_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ are weights from the input to hidden nodes. $\alpha_0 = \beta_{i0} = 1$ from the bias terms and ϵ_t is the random error.

5.3 Performing ARIMA and ANN

Our dataset is time series data. Thus to perform *ARIMA* method we first must verify that our time series is stationary by applying the Dickey-Fuller test. The result show that P-Value is 0.01 for wind speed attribute. Therefore, our time series is stationary as it is shown in **Figure 4**. Moreover, we can apply directly the AUTO ARIMA model which gives directly the reasonable parameters for our model.

[H] Table 1: Best ARIMA model

ARIMA (p,d,q)	*AIC*
$ARIMA(2,0,0)$ with non-zero mean	854.3203
$ARIMA(0,0,0)$ with non-zero mean	1446.793
$ARIMA(1,0,0)$ with non-zero mean	874.5046
$ARIMA(0,0,0)$ with zero mean	2182.676
ARIMA(3,0,0) with non-zero mean	**851.523**
$ARIMA(4,0,0)$ with non-zero mean	854.1045
$ARIMA(3,0,0)$ with zero mean	*Inf*

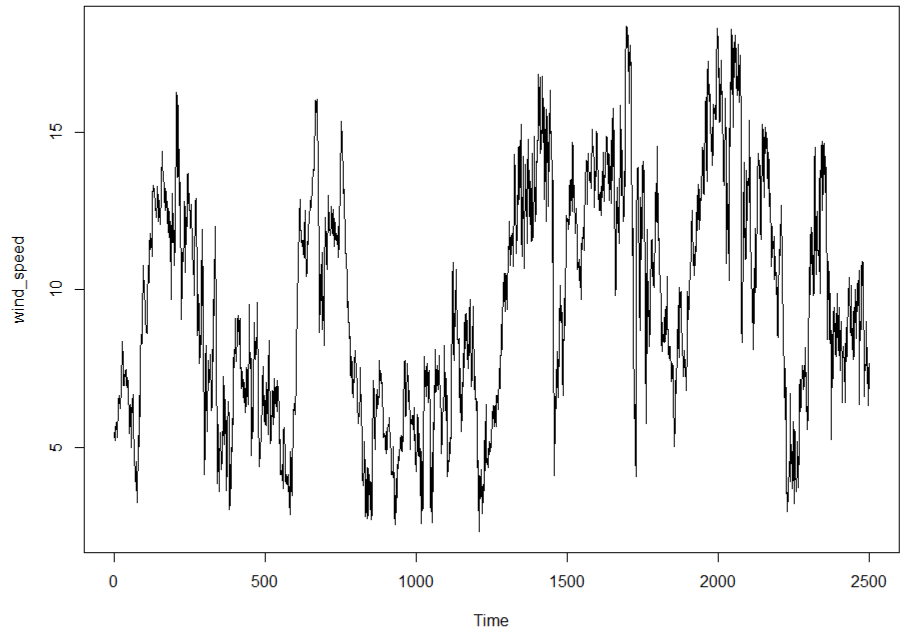

Figure 4: Wind speed data

As we can see in **Table 1** that the best *ARIMA(3,0,0)* is the best model that can be used to predict the wind speed in this data. However, *NNAR(3,2)* is the recommended model for the artificial intelligence method.

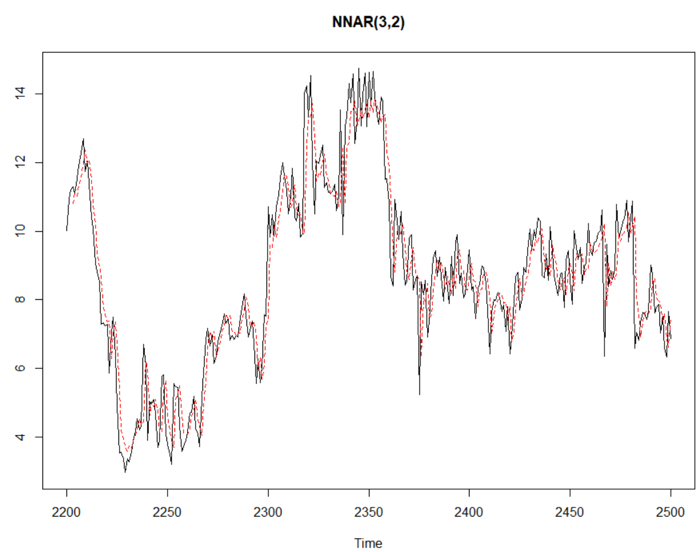

Figure 5: Wind speed forecasting by ANN

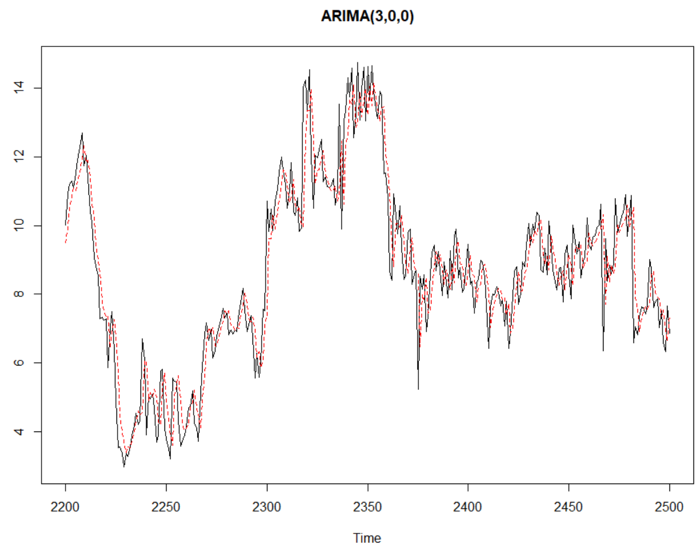

Figure 6: Wind speed forecasting by ARIMA

In the **Figure 5** and **Figure 6**, the predicted graph is in red colour. It can be seen that the forecasting of the both models is acceptable. However, it is impossible to decide which of the methods is more accurate. so we applied the estimation of differentiation criteria between the estimated models, the mean absolute percentage error (MASE), the mean absolute deviation(MAD) and the mean absolute scaled error(MASE). The results can be seen in **Table3**

- The Mean absolute percentage error (MAPE):

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

- The Mean Absolute Deviation(MAD):

$$MAD_k = \frac{1}{N}\sum_{t=1}^{N}|e_{t+k|t}|$$

- The Mean Absolute Scaled Error(MASE):

$$\frac{\frac{1}{J}\sum_j |e_j|}{\frac{1}{T_1}\sum_{t=2}^{T}|Y_t - Y_{t-1}|}$$

MODEL	*MASE*	*MAPE*	*MAD*
$ARIMA(3,0,0)$	**0.9821206**	**0.09889631**	**0.7861313**
$NNAR(3,2)$	0.992431	0.09988658	0.8335178

Table 3: Assessment criteria for differentiation between the two models.

we find that the performance of the box and Jenkins model is better than the neural network model in terms of all evaluation metrics, with the mean absolute percentage error (MAPE) of 0.098, the mean absolute deviation (MAD) of 0.786 , and the mean absolute scaled error(MASE) of 0.982.

6. Conclusion

The objective of this paper is to compare the forecasting performance of the autoregressive integrated moving average *ARIMA* and artificial neural networks *ANN*. The methods were applied in data set which was recorded by Wind Turbines's scada system in Turkey. The obtained results show that *ARIMA* model performed better than *ANN* model in predicting the wind speed. The criteria of the MASE, MAD and MAPE. Thus, the ARIMA is the recommended model and can be used for the prediction of wind speed confirmed the results.

7. The R code

```
library(dplyr)
library(tidyr)
library(forecast)
library("stringr")
library(e1071)
library(lattice)
library(caret)
######### Reading the data ########

mydata <- read.csv("wind1.csv")
summary(mydata)
###### Preprocessing ######
boxplot(mydata$Wind.Speed..m.s.)
boxplot(mydata$LV.ActivePower..kW.)
######### REMOVING THE OUTLIERS #######"
summary(mydata)

df_new <- mydata %>%
dplyr::select (LV.ActivePower..kW.,Wind.Speed..m.s.) %>% #getting mentioned var
filter (LV.ActivePower..kW. > 0 & Wind.Speed..m.s. < 18.4 ) %>%
drop_na() #dropping NA values.
df_new <- df_new %>% drop_na()
summary(df_new)
```

```
boxplot(df_new$Wind.Speed..m.s.)
boxplot(df_new$LV.ActivePower..kW.)

############## setting the variables #########
observ = 2500
library(caret)

wind_s <- df_new$Wind.Speed..m.s.[ 1 : observ]
power_g <- df_new$LV.ActivePower..kW.[1:observ]
summary(wind_s)

cor(wind_s,power_g)
plot(wind_s,power_g)
###################

########## TIME SERIES #################"

library(tseries)
yt <- ts(df_new)
plot(yt)
wind_speed <- ts(wind_s[1:observ])
plot(wind_speed)

train_set <- window(wind_speed, start = 2200, end = 2500)
plot(train_set)
str(train_set)

###################  STATIONARITY TEST ###########

adf.test(train_set, alternative="stationary", k=0)

################## AUTO-ARRIMA ################
Bmodel<-auto.arima(train_set,max.p=5, max.q=0,d=0,trace=T,test="adf"
                                              ,stationary=F)
Bmodel
fit1 <- forecast(Bmodel)

plot(train_set, type = 'l',main="ARIMA(3,0,0)",ylab="", xlim = c(2200,2500))
lines(fitted(fit1), col = 2, lty = 2)
summary(fit1)
########### ANN ############
library(nnet)
library(NeuralNetTools)

fittt <- nnetar(train_set)
```

```
fit2 <- forecast(fittt)
plot(fit2)
summary(fit2)
plot(train_set, type = 'l',main="NNAR(3,2)",ylab="", xlim = c(2200, 2500))
lines(fitted(fit2), col = 2, lty = 2)

############## ERRORS TESTING ############"
require(Metrics)
AR_pre <- fitted(fit1)
NR_pre <- fitted(fit2)

a <- AR_pre[282:301]
n <- NR_pre[282:301]
s <- test_set[282:301]

##### the estimation of differentiation criteria for ARIMA #####
mad(s,a)
mape(s,a)
mase(s, a,step_size = 1)

##### the estimation of differentiation criteria for ANN #####
mad(s,n)
mape(s,n)
mase(s, n,step_size = 1)
############ THE END ###########
```

References

Joseph Aidan and Issa Abdourahamane Ide. The use of autoregressive moving average and artificial neural network as short term wind speed forecasting tools for lagos, nigeria.

George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons, 2011.

PJ Brockwell and RA Davis. 11.7 estimating the cross spectrum, estimation of the phase spectrum. *Time series: theory and methods: Springer Science & Business Media*, pages 448–450, 2013.

Wen-Yeau Chang. Short-term wind power forecasting using the enhanced particle swarm optimization based hybrid method. *Energies*, 6(9):4879–4896, 2013.

Wen-Yeau Chang et al. A literature review of wind forecasting methods. *Journal of Power and Energy Engineering*, 2(04):161, 2014.

Ergin Erdem and Jing Shi. Arma based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, 88(4):1405–1414, 2011.

Laid Gasmi, Zouaoui Chikr Elmezouar, and Mohammed Kadi Attouch. Wind power forecasting using neural network and arima models (field of" kabertene", in southern algeria). *INTERNATIONAL JOURNAL OF ECOLOGICAL ECONOMICS & STATISTICS*, 39 (2):71–79, 2018.

Mahmoud Ghofrani and A Suherli. Time series and renewable energy forecasting. *Time series analysis and applications*, pages 77–92, 2017.

Sui-Lau Ho, Min Xie, and Thong Ngee Goh. A comparative study of neural network and box-jenkins arima modeling in time series prediction. *Computers & Industrial Engineering*, 42(2-4):371–375, 2002.

Thomas B Johansson, Henry Kelly, Amulya KN Reddy, and Robert H Williams. Renewable fuels and electricity for a growing world economy: defining and achieving the potential. *Renewable Energy. Sources for fuels and electricity. Island Press, Washington DC*, 1993.

Gong Li and Jing Shi. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, 87(7):2313–2320, 2010.

Benmbarek Mohamed, Benzergua Fadela, and Khiat Mounir. Optimization of the wind turbines location in kaberten wind farm in algeria. *Energy Procedia*, 74:122–129, 2015.

Mohamed A Mohandes, Shafiqur Rehman, and Talal O Halawani. A neural networks approach for wind speed prediction. *Renewable Energy*, 13(3):345–354, 1998.

Mohammad A Mohandes, Talal O Halawani, Shafiqur Rehman, and Ahmed A Hussain. Support vector machines for wind speed prediction. *Renewable energy*, 29(6):939–947, 2004.

JC Palomares-Salas, JJG De La Rosa, JG Ramiro, J Melgar, A Aguera, and A Moreno. Arima vs. neural networks for wind speed forecasting. In *2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, pages 129–133. IEEE, 2009.

Benjamin Pillot, Marc Muselli, Philippe Poggi, and João Batista Dias. Historical trends in global energy policy and renewable power system issues in sub-saharan africa: The case of solar pv. *Energy policy*, 127:113–124, 2019.

Shafiqur Rehman and Talal Omar Halawani. Statistical characteristics of wind in saudi arabia. *Renewable Energy*, 4(8):949–956, 1994.

Gordon Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342–349, 2009.

Sang-Bing Tsai, Youzhi Xue, Jianyu Zhang, Quan Chen, Yubin Liu, Jie Zhou, and Weiwei Dong. Models for forecasting growth trends in renewable energy. *Renewable and Sustainable Energy Reviews*, 77:1169–1178, 2017.

Huaizhi Wang, Zhenxing Lei, Xian Zhang, Bin Zhou, and Jianchun Peng. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, 198: 111799, 2019.

Jianwu Zeng and Wei Qiao. Support vector machine-based short-term wind power forecasting. In *2011 IEEE/PES Power Systems Conference and Exposition*, pages 1–8. IEEE, 2011.