Forecasting of Solar Electricity Generation and Performance Evaluation of Forecasting models using Time Series data

MSc Research Project
Data Analytics

Siddharth Chaudhary
x16137001

School of Computing
National College of Ireland

Supervisor: Dr.Paul Stynes

National College of Ireland
Project Submission Sheet – 2017/2018
School of Computing

Student Name:	Siddharth Chaudhary
Student ID:	x16137001
Programme:	Data Analytics
Year:	2016
Module:	MSc Research Project
Lecturer:	Dr.Paul Stynes
Submission Due Date:	11/12/2017
Project Title:	Forecasting of Solar Electricity Generation and Performance Evaluation of Forecasting models using Time Series data
Word Count:	7092

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	24th December 2017

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Forecasting of Solar Electricity Generation and Performance Evaluation of Forecasting models using Time Series data

Siddharth Chaudhary
x16137001
MSc Research Project in Data Analytics

24th December 2017

Abstract

The present study applies four time series models named TBATS, ARIMA, Simple Exponential Smoothing and Holt method to forecast the solar power generation in two Indian cities, Delhi and Jodhpur. Since solar power generation is dependent on solar radiation hence the later one is forecasted with the help of time series models and former one with the help of forecasted solar radiation value. The ARIMA method outperforms the TBATS, Holt method and Simple exponential by 11 percent, 31 percent and 32 percent respectively. Finally the forecasted values of ARIMA were used to calculate the total electricity production can be made in two sites, Delhi and Jodhpur. The difference in production came out to be 7.7 MWH, 36 MWH, 111 MWH and 65 MWH for November, December, January and February months with Jodhpur being on higher side for all four months.

1 Introduction

World is looking to use renewable sources of energy like wind and solar. One cannot deny the fact that efficiency of these sources rely on weather which varies according to time and location. With solar power system, (Vignola et al.; 2012) the time window has shrunk to day time when solar radiations are available and with what angle they are transmitted to the location will also create difference in power generation. So one of the important factor , which will impact the system performance, is location. A location with 50 more days of (Wilcox; 2007) sunlight will create a major impact in the performance of solar power system. This difference has been clearly shown by comparing power generation capacity of a similar solar system in two cities, Delhi and Jodhpur, with different climate. The first step of solar power forecasting is to forecast solar irradiance and other related weather aspects (Mathiesen and Kleissl; 2011). In this study not only solar radiation has been used for forecasting. Apart from location, this study compares the different time series models like Auto-regressive Integrated Moving averages(ARIMA), HOLT, TBATS and Simple Exponential for forecasting solar radiation data. ARIMA have been chosen because it is among the best models which are used for solar radiation forecast (Voyant et al.; 2017). While this is the first research on using TBATS model for forecasting of solar radiation which discussed in literature review.

Section 2 is presenting the literature review of subject. Section 3 is explaining the methodology of the study. Section 4 focuses on design and implementation. Section 5 explains the Evaluation and section 6 comparison of result generated from various models and comparison of forecasted electricity generation.

This study has been done to answer various questions like which model is best to for time series modelling of solar radiation data, which location is suitable for implementing solar power generation.

Research Objective 1:To Forecast the solar electricity generation of Jodhpur and Delhi.

Research Objective 2:To evaluate the performance of forecasting models.

Research Objective 3:Comparison between both the cities which one can produce more solar electricity on same amount of investment over setting up solar power plant.

2 Related Work

This section presents the work done by various researchers regarding solar power forecasting using different models.

Johansson and Burnham (1993) conducted a research on electricity and renewable fuels for the growing world energy demand. As per Author, Energy demand have never been decreased even though many strenuous efforts were made in using energy in efficient ways. Efforts has been made to use renewable sources of energy.Johansson investigated that by 2050 renewable sources of energy would account three-fifths of the worlds total electricity generation. Their findings have also made it necessary to forecast solar electricity generation which is the base of this research. Herzog et al. (2001) in his report on Renewable energy sources states that there is proliferation in power generation using renewable sources, generation of electricity using solar photovoltaic energy is experiencing rapid growth while declining the cost of electricity generation. The above two finding gives the intuition that future will be of Renewable energy sources which again gives the concrete base to this research of predicting solar electricity generation. As (richter; 2009) European Solar Thermal Electricity Association (ESTLA) report states that concentrated Solar Power could meet one quarter of worlds total energy demand by 2050. Research on forecasting of solar irradiance became an essential field due to its demand as per REN21 [1] solar photovoltaic capacity production was lifted to 227 gigawatt(GW) in 2015 which was 177 GW in 2014. It is almost 10 times the worlds solar photovoltaic(PV) generation of a decade earlier. Building on the above two finding gives the concrete base to this research of predicting solar energy production for the future. Forecasting of solar electricity is highly dependent on solar radiation, solar radiation needs to be forecasted first and it can be done using historical time series data.

Time series is explained by (Granger; 1981) on his investigation on properties of time series data. Author states that, Data which is auto correlated over time is categorized as time series data. Most of time series study is to extract statistical important information from the time-dependent data of single series (univariate time series) and from time dependent data of multiple series (multivariate time series data). Solar radiation (weather data) comes under the category of univariate time series data. Here one observation depends on previous observations and the order matters. A time series data is consisting of various components as trend, seasonality, cyclic component and Noise. Trend is defined

[1] www.ren21.net/wp-content/uploads/2016/06/GSR$_2$016$_{F}ull_{R}eport.pdf$

as the long term movement in a time series without calendar related and irregular effects, and it is a reflection of the underlying level. All the components associated with time are known as seasonal component. So, in time series data seasonality refers to periodic fluctuations that occur regularly in particular time-frame. Seasonality is always of a fixed and known period. A cyclic pattern exists when data exhibit rises and falls that are not of fixed period. In discrete time, white noise is a discrete signal whose samples are regarded as a sequence of serially uncorrelated random variables with zero mean and finite variance. The analysis of above findings is also supported by (Mahalakshmi et al.; 2016). Mahalakshmi investigation on forecasting of different types of time series data. Author in his study also analyzed the use of time series and have explained the time series component and forecasting models like Auto-regressive Integrated Moving Averages(ARIMA), Simple Exponential Smoothing (SES).Time series analysis is an essential area in forecasting that focuses on forecasting of time series data to study the data and extract meaningful information and statistics from it. From the above research findings,univariate time series data has been considered for this research. Ren et al. (2015). Conducted a research on using ensemble methods for solar and wind power forecasting using time series data. Author mentioned the fact that solar irradiance forecasting categorization is based on two approaches physical and statistical methods. Former takes meteorological data as input and models used to analyse the historical data using Artificial Neural Networks(ANN), ARIMA. Building on the above research findings, forecasting models like ARIMA and historical data have been used.The above three finding helps in forecasting the data which is eventually the main motivation of this research

Four forecasting models have been used to analysis univariate solar radiation time series data for this Research and they are as follow: Simple Exponential Smoothing (SES), Holts method, TBATS and ARIMA.

TBATS model have not been used earlier for forecasting of solar radiation as best of my knowledge. This is the first research using this model. As no literature paper have been found on solar radiation forecasting. However, it has been used in forecasting Gold price as per (Hassani et al.; 2015) and outperform the Bayesian auto-regression and Bayesian vector auto-regressive models. TBATS model was also used in forecasting household electricity demand as per (Veit et al.; 2014).

According to De Livera et al. (2011) and Rob.J.Hyndman[2],TBATS model is evaluated as T stands for Trignometric term, B for Box-Cox, A for ARMA errors, T for Trend and S for seasonality. It is the model that combines other models component like Trigonometric term for seasonality this is similar to fourier term in harmonic regression. Fourier terms are used to handle periodic seasonality, Box-Cox transformation for heterogeneity, It has ARMA error like dynamic regression, level and trend term similar to automated exponential smoothing.

TBATS {W,{P,Q} $\phi, \{M, K\}$}

$Where$,

W= Box-Cox transformation for heterogeneity

P,Q= ARMA errors

$\phi = levelandtrend(dampingparameter)$

M,K= seasonality period and Fourier term

Some of the key advantages of the TBATS modeling

1.Handles long seasonality. As the solar radiation data dataset used in this research has long seasonality

[2]https://www.otexts.org/fpp

2.It is fully automated

In this data, seasonality and non-linear features are present this is the reason to use this model

Unlike Naive method, that uses the recent observations to forecast the future periods and as Mean method that uses average of all the observation to predict future values Simple Exponential Smoothing(SES) is the model which forecasts the value based on all the observation but recent observation are heavily weighted than earlier observations i.e forecasting depends more on last observations than earlier observations. This model assumes that the fluctuation is around the mean of heavily weight value it gives the point forecast values as the average of all the predicted values. Point forecast values are all the forecast values which lies within the 95 percent confidence interval of models.SES is the method that deals with data having no trend and no seasonality. While the Holt Method is extended SES model just added an another feature as trend. holt method deals better with Time series data which has trend factor in it as per (Kalekar; 2004) and Rob.J.Hyndman[3] studies on time series forecasting Using exponential smoothing.

According Box et al. (2015) ARIMA is one of the most famous and efficient algorithm for modeling additive time series data. Although it also has a prerequisite that time series should be linear and stationary. ARIMA implements three algorithms, differencing to make data stationary, and moving average and auto regression for overall forecasting. ARIMA is suitable for univariate forecasting. ARIMA(p,d,q) p is the last number of observations that are used as predictor in regression equation of the model, d is the value that states number of times the data needed to be differencing to make it stationary, q is the value of past lagged error used in regression equation. (Box and Cox; 1964) and (Sakia; 1992)in their research on box-cox transformation states that the non-linear data can often significantly be transformed to fit in the model. It is the transformation for variance stabilization. The Box-Cox transformation provides a convenient way to make the data normally distributed and to find a suitable transformation. In this research box cox transformation have been used.

(Das et al.; 2018)conducted a research on forecasting models optimization and forecasting of photovoltaic power generation using time series data. Das states researchers classifies that the photovoltaic power forecasting depends upon time horizon, solar irradiance and other meteorological data and forecasting models furthermore describing three categories of time horizon short-term forecasting, medium-term forecasting, long-term forecasting. As per author, Long-term forecasting is worthy of planning of electricity generation and forecasting models strength and limitations have been discussed and finding shows that Root mean standard error(RMSE) was used more frequently to evaluate performance of forecasting models.Furtermore (Lonij et al.; 2013) found that forecasting accuracy changes with the change of forecast horizon with identical parameters in the same model. Building on the above findings Long-term forecasting have been used in this research for solar electricity generation and RMSE, MAPE have been used for finding accuracy of models. (Jiang and Dong; 2017)conducted a case study on Tibet area in china using Kernel Support Vector Machine(KSVM) with Regularized Estimation under Structural Hierarchy(GRESH) hybrid model which is optimization based, for forecasting of hourly global horizontal irradiance jiang used one year of data from four different sites. First 23 days of each month have been used as training set and last seven days as test set. Author investigated that on criteria of Mean Absolute Percentage Error(MAPE) and RMSE performance of KSVM-GRESH is better than KSVM models. However, lim-

[3]https://www.otexts.org/fpp

itation of this case study is that no cross-validation have been done. As well as test data should have been chosen randomly rather than the last seven days of each month. Concluding from the above two research, it shows that models are usually compared on basis of RMSE, MAPE. Therefore, for comparison of the model RMSE, MAPE and Mean Absolute Error(MAE) have been used in this research.

This next four research findings supports to use ARIMA and TBATS in forecasting (Paoli et al.; 2010) conducted a research using an optimized MLP, a most used form of ANN in renewable energy domain and in time series forecasting. For this research, global daily radiation data have been taken from meteorological station of Ajaccio to forecast for the next day. Poali identified that the MLP model outperforms the ARMA model. Paoli also identified that ARIMA, Bayesian inferences, Markov chain and K-nearest Neighbor are most popular forecasting models on reviewing previous related work of forecasting of time series he further finds that optimized MLPs prediction is similar to ARIMA. (Voyant et al.; 2011) conducted a research in forecasting of daily global radiation using optimized ANN(MLP) induced with multivariate. voyant took Nine years of data(January 1988 t0 December2007) from Corsica meteorological department which contain global solar irradiance as endogenous and temperature, wind speed, pressure, sunshine duration as exogenous input. Endogenous Input is taken on principles of ARIMA methodology. On studying the result, use of exogenous input improves the forecasting quality only about 0.5 percent.The drawback of three next studies is that non-stationary time series have been used and converted to stationary time series before being used in ARMA model. Instead of ARMA implementation, ARIMA would have been used as it is advanced version of ARMA developed for non-stationary data as per (Wan et al.; 2015). Another approach for forecasting of solar radiation was proposed by (Ji and Chee; 2011) by using an ARMA/Time Delay Neural Networks for hourly solar prediction the time series data used for this research is non-stationary before implementing ARMA this data have been made stationary using detrend models Chee investigate that hybrid model of ARMA/TDNN presents better result than standalone ARMA. (Sun et al.; 2015) have done in-depth investigation of six different ARMA-GARCH models, applied on 2 datasets(1983-2012) of monthly mean total daily solar radiation from two different climate station of china. The Autoregressive and moving averages are determined using Akaike information Criterion. Authors finding shows that ARMA (3,3)-GARCH-M outperforms the other model. Sun concluded that ARMA-GARCH model are better than ARMA Box-Jenkins method as the former one maintains the variability of solar radiation. The drawback of the above two studies is that non-stationary time series have been used and converted to stationary time series before being used in ARMA model. Instead of ARMA implementation, ARIMA, TBATS, SES would have been used as it is advanced version of ARMA developed for non-stationary data.

Reikard (2009) conducted a comparison of time series forecasting models to predict global horizontal radiation using six time series data sets which runs from January 1, 1987 to December 31,1990 at resolutions of 5, 15, 30, and 60 min. Data contributes nonlinear variability, due to variations in cloud cover and weather. Nevertheless, the dominance daily data makes it straightforward to build predictive models. Forecasting models like regression in logs, Autoregressive Integrated Moving Average (ARIMA), and Unobserved Components models(UCM) are compared. Reikard concluded UCM performed better than Regression model but it is outperformed by ARIMA. Building upon his finding it support to use ARIMA. In the field of forecasting solar radiation (Hussain and Al Alili; 2016)carried out a research on day head forecast. Hussian used one month of data 1 to

30 march 2016 of Abu Dhabi and forecasted the global solar irradiance of day ahead and on basis of ACF plot ARIMA model have been used. Further he used ARIMA(2,1,3), ARIMA(2,1,2),ARIMA(2,1,4).Hussian investigated that ARIMA(2,1,4) model performed best. The drawback of the above research is we need to find the perfert model manually this is time consuming as building on Rob.J.Hyndman[4] findings appropriate ARIMA model can be automatically implemented using auto.arima() function and that has been used in this research.

Wan et al. (2015) conducted a research on forecasting methodology and application of solar energy forecasting in smart Grids energy management. In his finding of forecasting methodology, Wan proposed upgraded variations of ARMA like ARIMA and ARMAX which have been proved better than itself. Because ARMA assumes a linear relationship between series which is always not the case. WAN concluded that ARMAX has a capability to include the external factors like temperature, wind and humidity that could increase the forecasting accuracy. On his finding it suggest to be the future work of this research.

Kumar and Sudhakar (2015) carried out the research on performance evaluation of solar PV power plant commissioned at Ramagundam a place located In India.in his research he carried out the process flow how the solar radiation is converted in electricity using photovoltaic cell, evaluated performance of various component like inverter, solar panel and also presented the methodology for calculation of solar electricity generation which is highly valuable for this research.

3 Methodology

This research uses the 'Cross Industry Standard Process for Data Mining' (CRISP-DM) methodology. CRISP-DM is hierarchical process which comprises of six level breakdowns as per (Chapman et al.; 2000) which is followed by this research. The Crisp-DM is modified according to this researchs need as shown in figure.1 the arrow direction between two figure indicates the modified form. First is Data extraction, data cleaning and data aggregating. Second Time series data Modelling, Third is Implementation of forecasting models. Fourth is the forecasting model evaluation have been done. Furthermore, Fifth is the predicted value of forecasting models which have been used to achieve the research objective that is the sixth phase

Data from Central Pollution Control Board(CPCB)[5] Ministry of environment and forests (Govt. of India) is collected for this research. CPCB is an organization comprises of many meteorological stations which stores and maintain historical weather and pollution data. Two datasets (January 1, 2016 to October 31, 2017) of daily measured solar radiation from two cities Delhi (latitude 28.6502 N, longitude 77.3027 E) and Jodhpur (latitude 26.2389 N, longitude 73.0243 E) of Indias CPCB meteorological station equipped with pyranometers which measures the solar radiation have been taken. The dataset contains two attributes date and daily measured solar radiation.

Qualitative checks have been done for any absurd or corrupt value. All this validation and cleaning of this research was done in Excel and R (programming language) (Shumway and Stoffer; 2006). On examining the data set it was determined that data uphold lot of missing values. The dataset has been investigated thoroughly and the missing values

[4]https://www.otexts.org/fpp
[5]http://www.cpcb.gov.in/CAAQM/frmUserAvgReportCriteria.aspx

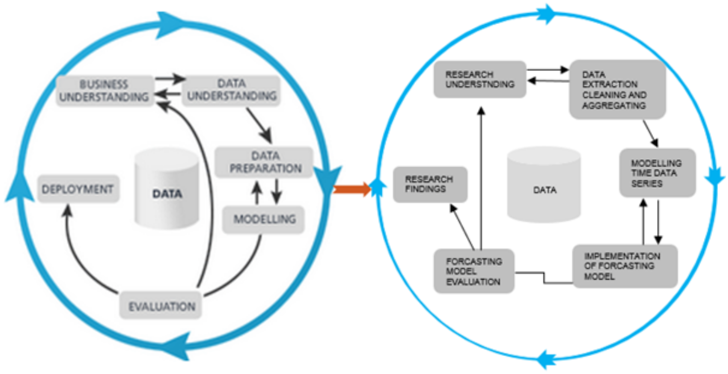

Figure 1: Modified Crisp-DM

have been filled on analyzing the previous data values using excel and outliers have been removed using R as dicussed by (Brockwell and Davis; 2016). The time series data is prepared using time series R package (zoo, tseries) as per (Shumway and Stoffer; 2006) book. This time series data is univariate as it contains single attribute that be used in forecasting . After cleaning up the data, it will be ready to parse into the forecasting models which forecast the univariate data.

The time series data of both the cities have been decomposed and seen in form of plots to examine the trend, seasonality and noise. Data have been analysed for linearity and non-stationary using Box test, qqnorm plot, qqline plot and Auto-Correlation function (ACF) (Theiler et al.; 1992).

After analysing the data, Different forecasting have been built in R Studio (Integrated Development Environment) (Racine; 2012) using Rs Forecast package. Different forecasting models like Simple Exponential Smoothing(SES), HOLTs, TBATS and ARIMA model have been applied. As per (Brockwell and Davis; 2016)the forecast package of R consist of all the forecasting models used for this research which is capable of processing the solar radiation time series data. To call a forecast model form the package their specific function names is needed to pass as discussed next. For simple exponential smoothing function ses() has been used, for Holts model function holt() has been used, for TBATS model function tbats() has been used and for ARIMA model function auto.arima() has been used. ARIMA model have lot of variation as discussed in literature review. The auto.arima() choose the best fit model itself for the data. After implementing the forecasting models, accuracy of the models have been calculated using error parameters like Root Mean Standard Error(RMSE), Mean Absolute Errors(MAE), Mean Absolute Percentage Error(MAPE). On basis of these error parameters, models having the least error in predicting the next 120 observations is regarded as the best model. The predicted solar radiation value of the best model is used for calculation of solar electricity generation.

The forecasted Solar radiation of four months (November 1,2017 to February 28,2018) of both the cities have bern extracted from the best performing forecasting models. As the daily measured datas unit is in w/m2 therefore, the forecast datas unit is also in w/m2.For solar electricity calculation the data needs to be in kwh/m2 as per global formula to calculate the photovoltaic system generated electricity as per (Šúri et al.; 2007) and. To convert the w/m2 unit of data into kwh/m2 (Šúri et al.; 2007) gave the formulae that kwh/m2 = w/m2*24/1000[6][7][8]. To calculate the electricity generation of a month daily data have been summed up and is used in Standard formulae for solar electricity generation by (Šúri et al.; 2007) E=A*R*H*PR[9].

E = Electricity produced (kWh)

A = Solar Panel Area (m2)

r = solar panel efficiency or yield(percent),r is taken as 15.6 as per standard test condition

H = Solar radiation on panels

PR = Performance ratio, coefficient for losses (default value = 0.75)

The forecasted electricity for both the cities will be compared using table, visualization graph and the performance of forecasting models will be evaluated in results section.

4 Implementation

The overall architecture of this study can be explained below in figure 2. The programming language used for this research is R.

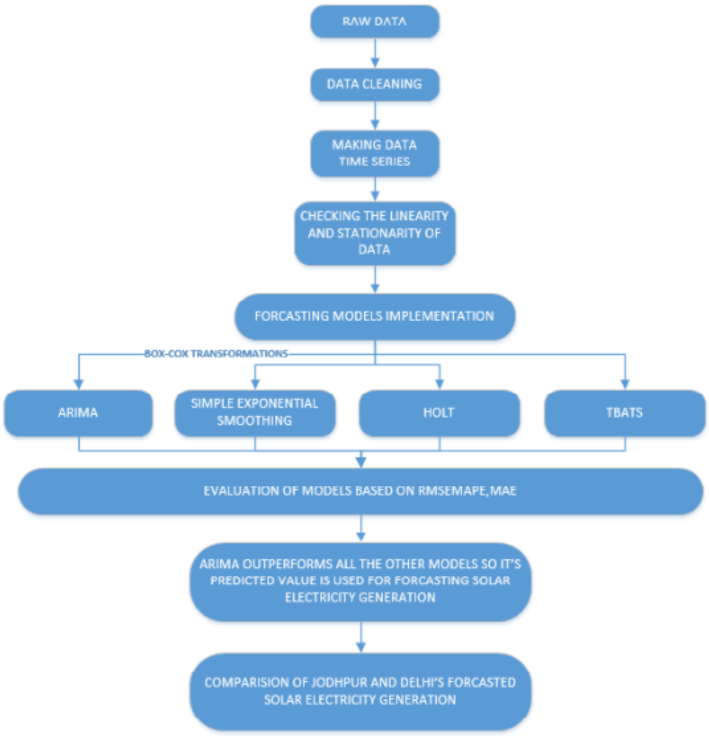

Figure 2: Data Flow of the Research

Data is collected from the Central Pollution Control Board (CPCB) Ministry of environment and forests(Govt. of India). Data for the city Jodhpur is downloaded from Jodhpur Station and for Delhi it is downloaded from the Anand vihar station. Both

[6] https://answers.yahoo.com/question/index?qid=20091027141143AARtZmS

[7] http://www.energylens.com/articles/kw-and-kwh

[8] https://www.researchgate.net/post/How$_{c}an_{I}convert_{W}h_{m}2_{t}o_{W}{}_{m}2$

[9] http://photovoltaic-software.com/PV-solar-energy-calculation.php

dataset is having the daily frequency, as forecasting of solar radiation, evaluation of models and comparison of solar electricity generation have been done using these datasets. Downloaded datasets were having lot of missing values which have been filled using Excel and outliers have been removed using R. Cleaned datasets are then Loaded in r using read.csv(). All the implementation part is done in RStudio(IDE) using R language. For processing the data, packages like zoo,tseries,forecast,ggplot2,fpp2 have been installed using install.packages() function. Then these packages are loaded in memory using library() function. The data is then converted into time series format using zoo() function and ts() function as per the forecasting models need. Non-linearity of data have been checked using qqnorm() and qqline() ploting. Then the Box.test() function, Auto-Correlation function(ACF) is applied to check whether the date is a white noise or not. As the time series data used for forecasting models should not be of random distribution or white noise they should show some correlation with the previous as well as with next variable.Then the graphical representation of the time series data have been done using plot() function. The solar radiation time series data is then decomposed in its seasonal, trend and noise component. The Box-cox transformation have been done for variance stabilization of the data.

Now it comes the implementation of the forecasting models ARIMA have been implemented using auto.arima() and passing the box-cox transformation value as 0.21 which is close to cube root transformation. Coding for backend process of all the models were already done in the forecast package. The auto.arima() function have been used which will eventually call the best model for the data from the "forecast" package consist of various version of ARIMA. Forecast() function is then applied to forecast the next 120 days value of solar radiation. Forecasted value is then extracted in form of excel into the local memory from Rstudio using write.table() function. Simple Exponential Smoothing model is applied on the solar radiation time series data using ses() function. The SES models code is also present in the forecast package which is called by ses() function. For forecasting and implementation of SES model ses() function have been used. Therefore, next four months of value have been forecasted by ses(). Then plotting of forecasted value have been done using autoplot(). Forecasted value is then extracted in form of excel into the local memory from Rstudio using write.table() function.

Holts model is applied on the solar radiation time series data using holt() function. . The holt models code is also present in the forecast package which is called by holt() function. For forecasting and implementation of HOLT model holt() function have been used. Therefore, next four months of value have been forecasted by holt(). Then plotting of forecasted value have been done using autoplot(). Forecasted value is then extracted in form of excel into the local memory from Rstudio using write.table() function.

TBATS model is applied on the solar radiation time series data using tbats() function. TBATS models code is also present in the forecast package which is called by tbats() function. For forecasting and implementation of HOLT model it is needed to pass tbats() function. Therefore, next four months of value have been forecasted by forecast() function. Then plotting of forecasted value have been done using autoplot(). Forecasted value is then extracted in form of excel into the local memory from Rstudio using write.table() function.

RMSE, MAPE, MAE have been used to compare each model using the 20 percent of test and 80 percent training data to check the performance of the model which is discussed in evaluation and result section. Focus will be on TBATS model how it perform on solar radiation time series data.

The extracted data of best performing model has been used to calculate the solar electricity generation.As the daily measured datas unit is in w/m2 therefore, the forecasted datas unit is also in w/m2.Each forecasted value is then converted to kwh using kwh/m2 = w/m2*24/1000 formulae. Electricity generation of a month is calculated using Standard formulae for solar electricity generation by E=A*R*H*PR.

For this research, Area for solar panel is assumed as of 1000 m2 for both the cities, r is taken as 15.6 as per standard test condition. As per (Dubey et al.; 2013) air temperature also affect the solar electricity generation but the average temperature of both the cities of each month are quite similar just having the max difference of 0.7 c. the temperature is below 25 for four months of both the cities [10][11]. So that does not affect the electricity generation calculation of the cities. The generated electricity for both the cities have been compared using table and visualization graph.Comparison of forcasted solar electricity generation and the performance of forecasting models is evaluated in results section. Focus will also be on TBATS model performance as this is the first research on solar radiation forecasting using this model.

5 Evaluation

A comparative time series plot of data for two cities is displayed in figure 3. We can see a seasonality of 12 months in both the series. There is not much trend. There is a fluctuation in Delhi data during the month of July. As July, August is monsoon season, so there is less solar radiation during monsoon season.

Figure 3: Comparative time series plotting of two cities

To check linearity dataset is examined using qqnorm() and qqline plot.by the plot below figure 4 it shows the data is somehow non-linear. To make variance stabilization box-cox transformation have been done for ARIMA while TBATS have in-built Box-Cox Transformation method.

Further, To determine the trend and seasonality of the two solar radiation time series data, acf and pacf of both the cities have been plotted as follows. Fig 5 is showing the auto correlation and partial auto correlation of solar irradiance data of Delhi and Jodhpur. As it is quite clear from acf that Delhi and Jodhpur both have a seasonality of 365 days. PACF explains the relationship of current observation with past observation.

After verifying the data using Box test, it have been analyzed that the solar radiation time series data used in this research is not a white noise as p-value is below 0.05 which

[10]https://www.yr.no/place/India/rajasthan/jodhpur/statistics.html
[11]https://www.yr.no/place/India/Delhi/New$_Delhi/statistics.html$

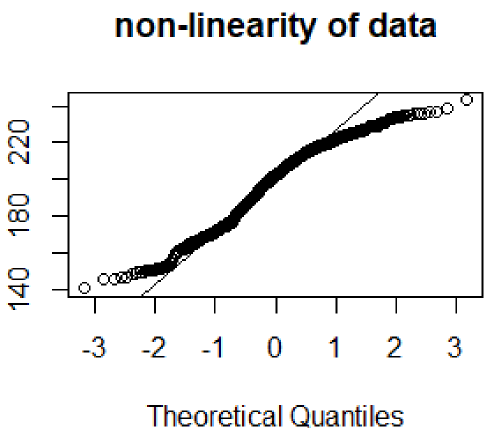

Figure 4: Checking Linearity

shows that the data can be be used for forecasting and Auto-Correlation function(ACF) plot of Solar radiation data shows that the data is non-stationary as ACF of stationary time series will drop to zero relatively quickly while ACF of the solar radiation data decreasing slowly as mentioned by Rob.J.Hyndman[12],hence it is non-stationary data.

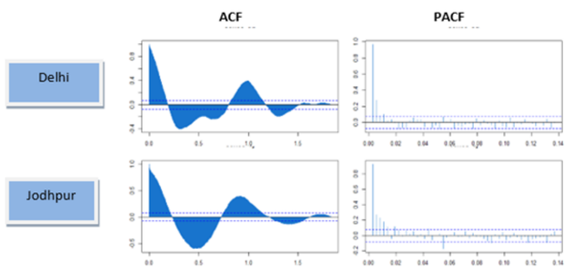

Figure 5: ACF and PACF for Delhi and Jodhpur

Time series decomposition figure 6 and 7 is explained below. This graph clearly evaluate the different components of solar radiation data of Delhi and Jodhpur. The overall variation of data is same in two cities. But Delhi is more fluctuating in comparison to jodhpur. In the seasonal component, both are displaying seasonality of 365 days in the overall duration of two years. There is not much trend as each city has observed almost same amount of sunlight in the duration of 2 years. But some trend might have been visible if this study was done for 5 years or more. Third is Noise data which have been used for analysing the underlying effect. After taking out the seasonality and trend, left over series is stationary hence can be used for forecasting.

Different methods like TBATS, ARIMA, Simple Exponential Smoothing, Holt model are applied in this research for forecasting the solar radiation data. The evaluation of each forecasting model is based on figures given below and is explained thoroughly in Result section. The blue line in each figure is the forecasted value of each model and the dark blue area is the models 80 percent confidence that all the value will come within

[12]https://www.otexts.org/fpp

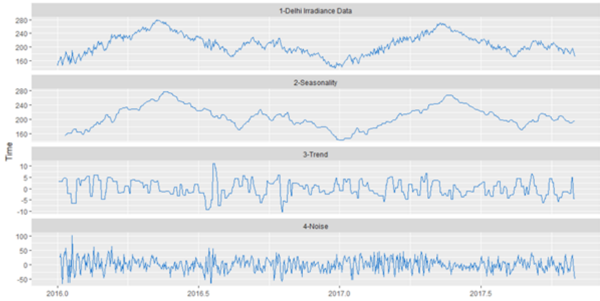

Figure 6: Component of Delhi data

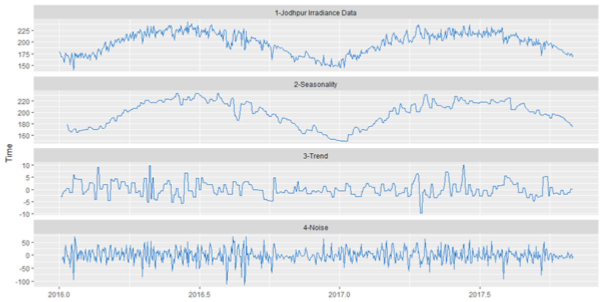

Figure 7: Component of Jodhpur data

the dark blue patch while the model is 95 percent confidence that all the value will come within the light blue patch.

Figure 8 below is forecasting result of TBATS model. TBATS graph shows that forecasted values are following the seasonal variation but very smooth. So this method is been checked is it suitable for this data. TBATS (1,1,3,-,,365,1) is being evaluated as the first variable is 1 that tells us substantial box cox transformation have been made. The next two variable is similar to p,q of ARMA. The 1,3 values here states that 1 last observation is used as predictor in regression equation and last 3 past lagged error are used in regression equation. Next - value signifies on damping parameter have been used as data doesnt have any trend, next variable 365,1 states that seasonality is found at 365 values and is handled by one fourier term.

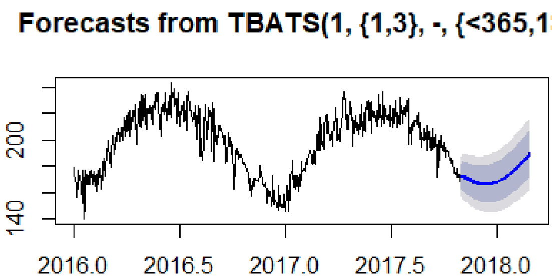

Figure 8: TBATS forecasting plot

ARIMA models plot Figure 9 below shows the closest pattern of forecasted value to actual value. ARIMA(3,1,2)(0,1,0)[365] is evaluated as the models first variable(3,1,2) states that 3 last observation is used as predictor in regression equation. 1 time differencing have been done for making data stationary and 2 last lagged error have been used in regression equation.(0,1,0) states that 1 seasonal differencing have been done and [365] states that it is on daily frequency.

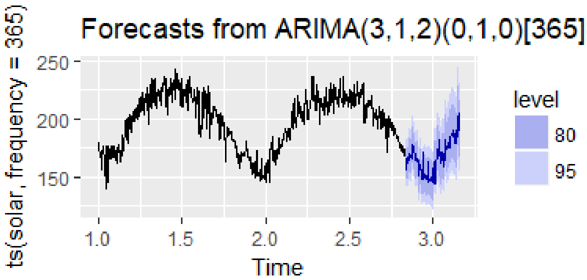

Figure 9: ARIMA forecasting plot

Simple exponential Smoothing model figure 10 is evaluated here the straight line seen is the point forecast as simple exponential smoothing gives the mean of all the predicted value. Therefore, it is evaluated that this model wont provide the value for each day. Therefore, it wont to help predicting the research objective of calculating how much electricity can be produced as for predicting electricity generation this model is not an effective one.

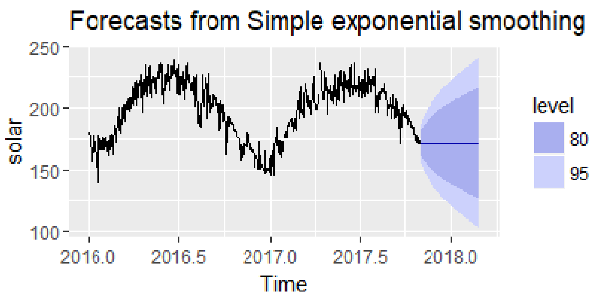

Figure 10: Simple Exponential Smoothing forecasting plot

The is the forecasting evaluation of HOLTs method in figure 11. Holt method is performing double exponential smoothing which is good with trend. Since this data does not show much trend variation hence forecasting result is a straight line or in other words HOLT is also not suitable for this dataset. Therefore, the SES and Holt model is not good for forecasting solar electricity generation as it needs the actual forecasted data not the Point forecast.

Below table in Result section presents the numeric comparison of four models. Here we have presented the RMSE, MAE and MAPE for all four models. Root mean square error or root mean square deviation (RMSE) is used to measure the difference between actual value and predicted value. Mean Absolute error (MAE) measures the average of all errors of prediction. So, it is the mean of absolute difference of actual value and

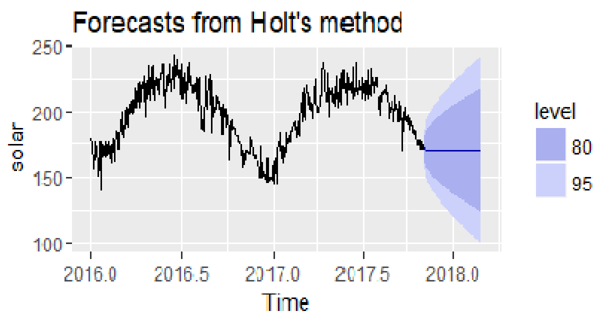

Figure 11: Holt Trend method forecasting plot

predicted value. Mean absolute percentage error (MAPE) measures the accuracy of a method for constructing fitted time series values in statistics.

The predicted value of the best fit model on basis of the above error parameter will be used for forecasting solar electricity generation. Next is the result section.

6 Results

The aim of the project of forecasting solar radiation,forecasting of solar electricity generation and performance of models have been achieved. Models have been evaluated using error parameters RMSE,MAE,MAPE,MASE.On the basis of the error parameter ARIMA outperformed all the other model and TBATS performed better than SES and Holt model.As ARIMA outperformed all the models its predicted value will be used for forecasting solar electricity generation of Jodhpur and Delhi for the next four month.

Model Name	RMSE(Root mean square error)	MAE(mean absolute error)	MAPE(Mean absolute error percentage)	MASE(mean absolute scaled error)
Simple Exponential Smoothing	8.136	6.19	3.141	0.583
TBATS	7.64	5.945	2.16	0.38
Holt's method	8.13	6.202	3.14	0.58
ARIMA	6.17	2.90	1.44	0.27

Figure 12: Performance of forecasting models

It is clear from the table that mean error is maximum for holt method as it simply averaging the entire series. It can be analyzed from RMSE, MAE and MAPE values that they are least for ARIMA as it ignores the sign while calculating the error, hence a better parameter to rely and compare. MASE compares the result to one obtained from naive method and if value is greator than 1 then performance is very poor. MASE is least for ARIMA, hence it is the best algorithm for modelling this data.

Electricity comparison of both cities

This is the final outcome of the research project where electricity generation has been calculated and compared. First of all the solar radiation for next four months has been calculated for two cities with the help of solar power generation formula, total electricity of each month has been calculated. The figure.13 and graph figure.14 below shows that jodhpur solar electricity production is higher than that of Delhi.The table shows that jodhpurs electricity generation on 1000m2 of solar panel is significantly high.1000 kwh is equal to 1 mgh.

Month\City	Jodhpur	Delhi
November-2017	1399086	1391364
December-2017	1323153	1287000
January-2018	1454427	1342692
February-2018	1465659	1400022

Figure 13: Comparing Jodhpur's and Delhi's forecasted solar electricity generation

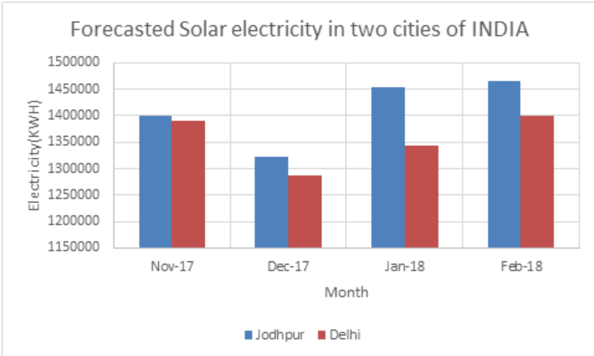

Figure 14: Graphical representation of Jodhpur's and Delhi's forecasted solar electricity generation

7 Conclusion and Future Work

The research objective of forecasting solar electricity generation of both the cities,comparison between them and performance evaluation of forecasting models have been achieved. The research states that the solar electricity generation of jodhpur is higher than that of Delhi for four months from November 2017 to February 2018. Forecasting of solar electricity significantly depend upon forecasting of solar radiation which is done by the forecasting models. Talking about model performance Arima model outperforms all the other three model. This is the first research on solar radiation forecasting using TBATS model. TBATS doesnt perform well in predicting the solar radiation value. Therefore, it is concluded that TBATS model is not a good option for forecasting solar radiation. Simple Exponential Smoothing and Holt model gives the point forecast which again is not a good option for forecasting the solar radiation.

There is lot of scope in this study for future work. First of all this work can be improved by including weather factors to improve the solar power forecasting. As temperature, humidity and wind are also responsible for the variation in power generation. This can be done with the help of ARIMAX, ARIMA-ANN or more complex models. Another important factor is to include more cities so that it can be determined that which locations will be more profitable in implanting a solar power generation system. Data of 10 years or more should be needed to forecast for a complete year with more accuracy There are lots of rural areas in India which are not connected to grid. So these cities/rural areas should be included in order to fulfill the requirement demand and fill the current demand supply gap.

References

Box, G. E. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 211–252.

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). *Time series analysis: forecasting and control*, John Wiley & Sons.

Brockwell, P. J. and Davis, R. A. (2016). *Introduction to time series and forecasting*, springer.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). Crisp-dm 1.0 step-by-step data mining guide.

Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., Van Deventer, W., Horan, B. and Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review, *Renewable and Sustainable Energy Reviews* **81**: 912–928.

De Livera, A. M., Hyndman, R. J. and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association* **106**(496): 1513–1527.

Dubey, S., Sarvaiya, J. N. and Seshadri, B. (2013). Temperature dependent photovoltaic (pv) efficiency and its effect on pv production in the world–a review, *Energy Procedia* **33**: 311–321.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification, *Journal of econometrics* **16**(1): 121–130.

Hassani, H., Silva, E. S., Gupta, R. and Segnon, M. K. (2015). Forecasting the price of gold, *Applied Economics* **47**(39): 4141–4152.

Herzog, A. V., Lipman, T. E. and Kammen, D. M. (2001). Renewable energy sources, *Encyclopedia of Life Support Systems (EOLSS). Forerunner Volume-Perspectives and Overview of Life Support Systems and Sustainable Development* .

Hussain, S. and Al Alili, A. (2016). Day ahead hourly forecast of solar irradiance for abu dhabi, uae, *Smart Energy Grid Engineering (SEGE), 2016 IEEE*, IEEE, pp. 68–71.

Ji, W. and Chee, K. C. (2011). Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn, *Solar Energy* **85**(5): 808–817.

Jiang, H. and Dong, Y. (2017). Forecast of hourly global horizontal irradiance based on structured kernel support vector machine: A case study of tibet area in china, *Energy Conversion and Management* **142**: 307–321.

Johansson, T. B. and Burnham, L. (1993). *Renewable energy: sources for fuels and electricity*, Island press.

Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing, *Kanwal Rekhi School of Information Technology* **4329008**: 1–13.

Kumar, B. S. and Sudhakar, K. (2015). Performance evaluation of 10 mw grid connected solar photovoltaic power plant in india, *Energy Reports* **1**: 184–192.

Lonij, V. P., Brooks, A. E., Cronin, A. D., Leuthold, M. and Koch, K. (2013). Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors, *Solar Energy* **97**: 58–66.

Mahalakshmi, G., Sridevi, S. and Rajaram, S. (2016). A survey on forecasting of time series data, *Computing Technologies and Intelligent Data Engineering (ICCTIDE), International Conference on*, IEEE, pp. 1–8.

Mathiesen, P. and Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states, *Solar Energy* **85**(5): 967–977.

Paoli, C., Voyant, C., Muselli, M. and Nivet, M.-L. (2010). Forecasting of preprocessed daily solar radiation time series using neural networks, *Solar Energy* **84**(12): 2146–2160.

Racine, J. S. (2012). Rstudio: A platform-independent ide for r and sweave, *Journal of Applied Econometrics* **27**(1): 167–172.

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

Ren, Y., Suganthan, P. and Srikanth, N. (2015). Ensemble methods for wind and solar power forecastinga state-of-the-art review, *Renewable and Sustainable Energy Reviews* **50**: 82–91.

richter, W. R. E. (2009). Concentrating solar power global outlook 09.

Sakia, R. (1992). The box-cox transformation technique: a review, *The statistician* pp. 169–178.

Shumway, R. H. and Stoffer, D. S. (2006). *Time series analysis and its applications: with R examples*, Springer Science & Business Media.

Sun, H., Yan, D., Zhao, N. and Zhou, J. (2015). Empirical investigation on modeling solar radiation series with arma–garch models, *Energy Conversion and Management* **92**: 385–395.

Šúri, M., Huld, T. A., Dunlop, E. D. and Ossenbrink, H. A. (2007). Potential of solar electricity generation in the european union member states and candidate countries, *Solar energy* **81**(10): 1295–1305.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data, *Physica D: Nonlinear Phenomena* **58**(1-4): 77–94.

Veit, A., Goebel, C., Tidke, R., Doblander, C. and Jacobsen, H.-A. (2014). Household electricity demand forecasting: benchmarking state-of-the-art methods, *Proceedings of the 5th international conference on Future energy systems*, ACM, pp. 233–234.

Vignola, F., Michalsky, J. and Stoffel, T. (2012). *Solar and infrared radiation measurements*, CRC press.

Voyant, C., Muselli, M., Paoli, C. and Nivet, M.-L. (2011). Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation, *Energy* **36**(1): 348–359.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F. and Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review, *Renewable Energy* **105**: 569–582.

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J. and Hu, Z. (2015). Photovoltaic and solar power forecasting for smart grid energy management, *CSEE Journal of Power and Energy Systems* **1**(4): 38–46.

Wilcox, S. (2007). National solar radiation database 1991-2005 update: User's manual, *Technical report*, National Renewable Energy Laboratory (NREL), Golden, CO.