Predict Beam Normal Irradiation and Global Horizontal Irradiation using Deep learning and Time series Algorithms

MSc Research Project
Data Analytics

Karan Sachdeva
Student ID: x18185916

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Karan Sachdeva
Student ID:	x18185916
Programme:	Msc Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	28/09/2020
Project Title:	Predict Beam Normal Irradiation and Global Horizontal Irradiation using Deep Learning and Time Series Algorithms
Word Count:	7343
Page Count:	25

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 required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

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):	

Predict Beam Normal Irradiation and Global Horizontal Irradiation using Deep Learning and Time Series Algorithms

Karan Sachdeva
x18185916

Abstract

Solar forecasting is one of important use case in field of data analytics that has grown exponentially in past couple of decades. Advent of neural network with improvement in computation systems has radically improved solar forecasts and enabled more accurate prediction. In recent years, a lot of emphasize is given to not only predict solar forecast but also improve existing results by applying various models. This research focus on hybrid approach of combining time series and neural network to improve solar forecasts results and take up existing challenges in the area of solar energy. Hybrid model forecast produced results with decent evaluation metrics, i.e. RMSE of 38.34 W/m^2, MAE of 27.771 W/m^2 for GHI while RMSE of 97.7 W/m^2 and MAE of 78.46 W/m^2 for BNI respectively. Also, Time series and deep neural networks are implemented to compare metrics with hybrid model metrics and comparison is done between metrics in current literature review and those obtained from all model implementation.

1 Introduction

The exponential expansion of human population has made energy a critical need for sustainable existence. In current circumstances of growing energy demand along with ever depleting fossil fuel supplies has highlighted the importance of renewable source of energy. International bodies such as International Energy Agency(IEA) emphasized on exploring renewable resources and measured that, in 2014, the world's total global primary energy production (TPES) is 159342 TWh of which renewable sources accounted a meagre 13.8% i.e. 22,027 TWh (Ghimire et al.; 2019). To enforce development of these sources of energy, mandatory renewable energy targets(MRET) are set by government agencies all over the world. For instance, European Union has set MRET of 20% of EU energy production and plan to further increase it by 7% by year 2030 (Ghimire et al.; 2019). Consequently, there is huge investments, and hence, research is carried out for commercial exploration of renewable resources.

1.1 Motivation and Domain Overview

There is abundance of renewable energy and many forms are harnessed to sustain human needs. Solar energy is widely exploited source used for commercial production of energy. Solar energy is extensively used in photo-voltaic(PV) and concentrated solar plants(CSP) with approximate energy of 100GWp and 100MWp respectively as per data in 2017 (Benali et al.; 2019). China, too has set steep goals and targets to raise 15% by the year 2020. It accounted for 45% of the global investments on renewable energy exploration (Chen et al.; 2019). As part of this research, components of solar energy, namely, Beam Normal Irradiation(BNI) and Global Horizontal Irradiation(GHI) forecast, principal source of energy generation from CSP and PV plants respectively, is performed.

1.2 Research Question

The research question investigates potential scope of application of different models in domain and attempts on exploring new machine learning techniques for obtaining better forecast.
RQ- "Can prediction of BNI and GHI using hybrid technique of Deep Learning and Time Series models, enhance traditional forecasting models of Neural network and Time Series(LSTM, MLP and ARIMA)?"
Different models are implemented, evaluated and results are compared for various forecast error metrics, for instance RMSE, to evaluate and compare forecast results.

1.3 Project Objectives

To address the research question, project objectives are clearly laid out in Table 1. These objectives are comprehensive critique of existing literature review, and development of different modelling techniques to forecast components of solar irradiation. The research question is investigated by objectives and sub-objectives listed in the table below.

Obj.	**Project Objectives**	**Modelling Technique Used**	**Evaluation Metrics**
1	Critical literature review of existing research articles of state of art techniques in the area of Solar Irradiation.		
2	Input dataset into Postgres and perform data cleaning and roll up data from hour to daily grain. Also, perform exploratory data analysis on time series data.		Correlation Plots, Test for Stationarity, Seasonality, Trend.
3	Implementation and evaluation of time series models for GHI and BNI forecast		
(3.1)	Implement and Evaluate RNN model(LSTM)	LSTM	Root Mean Squared Error(RMSE),
(3.2)	Implement and Evaluate Hybrid ARIMA ANN model	Hybrid ARIMA ANN	Mean Absolute Error(MAE),
(3.3)	Implement and Evaluate MLP model architecture	MLP	Mean Absolute Percentage Error(MAPE)
(3.4)	Implement and Evaluate ARIMA model architecture	ARIMA	

4	Compare results obtained from each model application		
5	Compare evaluation metrics obtained from each model with those listed in literature review		
6	Interpret how hybrid model of time series and neural network performs in comparison to neural network and time series models applied independently.		

Table 1: Table of Project Objectives

The major contribution is development of novel approach of combining neural network and time series models for prediction. The model would contribute towards body of knowledge by enhancing existing modeling technique and introducing hybrid model for prediction.

Rest of the report is structured as follows. **Chapter 2** describes peer reviewed literature in accordance with project objectives. **Chapter 3** describes implementation methodology and proposed design to be used. **Chapter 4** describes implementation steps for each model while **Chapter 5** describes evaluation, result and discussion of models implemented. Finally, **chapter 6** concludes whether implementation is successful in answering research question and identifies area of future work.

2 Literature Review for Solar Irradiation(2015-2019)

Solar energy is widely explored source of renewable energy and due to huge commercial aspect and potential for energy generation, it is subject of prime importance for researchers. The setup of both CSP and PV plants requires huge financial investment, forecast of solar irradiation subject to great degree of accuracy is essential for exploration of energy. The forecasting models used are broadly classified into empirical or weather based models, time series models and machine learning models. Literature review explores these model techniques, and identify gaps in existing models and types of models in area of solar forecasting in subsection 2.2, subsection 2.3 and subsection 2.4.

2.1 Empirical models

Empirical models are based upon results obtained from weather or metallurgical parameters, namely, air temperature, sunshine hours, among others inputed into system of mathematics based on assumption of normality of data. Models are calibrated using various functions, namely, linear function, logarithmic function, among others with metallurgical factors as inputs for forecast. Hassan et al. (2016) worked on development of model based on temperature inputs, namely, air temperature(T) and temperature gradient to predict daily solar radiation for different locations of Egypt. 17 different models are applied in the research with best model obtained at daily level with metrics with MAPE of 2.8% and RMSE of 0.7308 MJ/m^2. The study obtained decent performance metrics, however, the approach or results are not validated as study doesn't extensively performs statistical tests for forecasts obtained. Ozoegwu (2016) devised a forecast methodology of incorporating temperature gradient(T_{max}-T_{min}) in Hargreaves-Samani model to improve the existing model for predictions. The modified equation incorporated temperature gradient and other exogenous parameters i.e. relative humidity, extraterrestrial radiation for solar forecast. The model performed with lower forecasting errors with RMSE 0.02MJ/m^2 and R^2 of 0.96. Moreover, research validated results obtained with hypothesis testing such as ANOVA, to prove statistical significance and further strengthen the postulated model.

2.2 Cloud and Weather Based models

Cloud and weather based models have proved effective to forecast BNI, in particular for forecast short term horizon. Pereira et al. (2019) proposes improvement in solar irradiance forecast results for both GHI and BNI by employing novel offline coupling procedures(OCP). The study proposed the usage of weather research and forecast method(WRF) along with coupling methods that are used in forecast and prediction. Research is carried out in two stages, namely, derive WRF simulations using numerical weather prediction model that understands cloud formation to predict GHI and other meteorological factors. These are then used as inputs to OCP. These set routines are responsible in deriving additional features, namely, terrain effect that is not captured in WRF simulations. In addition, it employs external data, that of aerosol and gases distribution considered as constituents of atmosphere at given instance of space and time. The idea of research is prediction of Solar Irradiance under clear sky conditions by decomposing Solar Irradiance into GHI, BNI and Diffuse Irradiation. The forecast results obtained are decent GHI forecast with rRMSE of 2.5% approximately. But forecast errors i.e. MAPE for BNI and diffuse irradiation prediction that are approximately 55% and 51%.

Arbizu-Barrena et al. (2017) performed short horizon forecast for GHI and BNI by advection and diffusion of Cloud index using methodology based on WRF models. Maps based on cloud index maps are generated using cloud index transformed into GHI and BNI. Hence, the maps obtained from the cloud index maps are fed into WRF models to predict GHI and BNI upto a forecst of 6h horizon. The results generated from cloud index advection and diffusion(CIADCast) is compared with those obtained from various empirical models, namely, smart persistence model, Cloud motion vector model(CMV) model, among others that are used in area solar radiation forecast. Model performed exceptionally for short term forecast of BNI of 1-4h horizon. But, WRF-Solar model proved better at long period interval forecast. The study concluded that CIADCast model is more precise in determining cloud cover, and hence, achieve accurate BNI forecast. Wang et al. (2019) performed similar study with a novel idea that is based on calculation of cloud physical properties instead of cloud index and solar radiation. The study introduced a new model for intra day forecast of GHI and BNI on Spinning Enhanced Visible and Infrared Imager (SEVIRI) mounted on geostationary satellite MSG. The cloud physical properties are input into CPP-SICCS (Surface Insolation under Clear and Cloudy skies derived from SEVIRI imagery) model to impute solar irradiation on surface. SICC model distinguishes between cloud and cloud free image pixels and calculate CMVs based on cloud optical thickness (COT), cloud effective radius, and cloud top height derived from SEVIRI. The model gives rRMSE for GHI ranged from approximately 21% to 43% for 0–2h ahead forecasts for various locations of Netherlands and approximately 50% at 3h interval. The rRMSE for direct normal irradiation ranged from between 43.8% and 100% for 0–2h time horizon. Sirch et al. (2017) applied the similar technique for forecast. The model derived acceptable results for forecast upto a time period of 2h but for long period forecast, forecast errors increased substantially. Pecenak et al. (2016) used cloud optical depth and solar zenith angle to derive clear sky index that is considered as yet another important factor in determining prediction for BNI. Hence, approach to derive CMV employing cloud imagery and cloud index, and clear sky index are applicable for forecast of short term horizon i.e 0-2h ahead and the approach gave good forecast metrics for Beam Normal and Global Horizontal Irradiation. But, model gives poor results for intraday forecast for both BNI and GHI due to frequent changes in cloud cover.

2.3 Time Series models

Empirical and Cloud models are effective in some capacity but doesn't consider stochastic characteristics and time series of solar irradiation data (Inman et al.; 2013). These shortcomings are instrumental for the development of time series models, based upon stochastic regression and ANN (Ozoegwu; 2019). Reikard et al. (2017) proposed the application of time series model i.e. ARIMA coupled with exogenous input as Clearness index to forecast solar irradiation. The study compared the forecast results with other empirical models, namely, smart persistence, WRF-Solar and Cloud Advection, among other models. The evaluation methods used for the implementation are RMSE and MAE, measured by units of W/m^2. The results highlighted that ARIMA as most suitable model for short time duration forecast of 15m-45m horizon, followed by DICast. The MAE obtained for ARIMA is approximately 40W/m^2 while RMSE is approximately 72.4W/m^2 for data collated and calibrated from all sites. However, for long term forecast WRF based models proved more suitable with less prediction errors. Also, research highlighted that interpolation of missing values, not required for meteorological based model, is mandatory for application of ARIMA model. Alanazi et al. (2017) employed the technique of two-stage hybrid model for hourly GHI prediction. Time series data is used as input into Nonlinear Autoregressive Neural Network (NARNN) and output data is further used as exogenous input into ARMAX Model. Only daytime data is considered for model for expected sunshine hours, thus filtering and reducing data fed into model for computation and reducing simulation time for the model. Also, data is made stationary by differencing and normalized to scale to eliminate variations introduced by changes in the clear sky irradiance.Data obtained from second stage is in normalized form and comprises of GHI excluding nighttime hours. So, in order to evaluate the model, original data is normalized and nighttime hours are removed. Normalized RMSE or nRMSE is thus calculated and evaluated for prediction or forecast errors. The model exhibited nRMSE of 0.08 for cloudy weather condition, 0.12 for partly cloudy conditions and 0.04 for sunny weather condition at a average intraday level. The model also generated result for non stationary data and one stage NARNN, and forecast errors are thus in comparable range to stationary hybrid model.

Li et al. (2016) researched on the impact of cloud velocity on BNI forecast for short horizon. The research decodes the impact of real-time cloud transmit on BNI forecast accuracy. Cloud velocity is used as input for Multi layer Perceptron model(MLP), and model with real time and invariant transmittance of velocity of cloud is used for comparison to obtain performance metrics. Evaluation criteria set for analysis are MAE and RMSE. Forecast with real time transmittance showed improvement in MAE of 5% approximately over invariant transmittance for 15min ahead horizon forecast. Similarly, for 5min and 10min ahead forecast, model showed considerable change of 2.23% and 2.95% respectively. It is ascertained from study that cloud velocity is an important parameter for short horizon forecast.

Reikard and Hansen (2019) proposed solar irradiance forecast and clear sky index forecast that are important parameters for BNI prediction, by application of frequency and time models. The research compared results of ARIMA model, smart persistence model , frequency domain model and regression models for forecast in horizon of 15min-3hours. The forecasts are generated in order and data from previous iteration is not used in further iteration. For instance,forecast for 3h forward horizon, errors obtained for t+2h are discarded and only errors for t+3h time period are considered for evaluation metrics. Evaluation metrics used in study are MAE and RMSE and metrics from all models are compared. For short period estimation, i.e. 15mins forecast, persistence models performed considerably better in comparison to ARIMA and frequency models respectively. For 1h ahead forecast ARIMA gave better forecast results as compared to other models. For 2h and 3h ahead forecast, frequency domain models gave better forecast metrics as compared to other model metrics for GHI and clear sky index forecast. None of the model considerably outperformed one another for short horizon of forecast for GHI and clear sky index. Mukaram and Yusof (2017) proposed model for univariate time series by application of seasonal ARIMA(SARIMA) and artificial neural network (ANN) to predict daily average solar radiation. SARIMA is a special version of model ARIMA that has the ability to handle seasonal data and is also based on box-jenkins approach or method. In research, seasonal data is differentiated by seasonal period lag (Lag S) until seasonality is removed from said data. Model parameters p,q is estimated based upon maximum likelihood and parameter d is determined based upon number of times time series is differentiated. Research is carried out using different modeling techniques, namely, SARIMA, hybrid model and ANN model. In hybrid model, time series data is input into SARIMA and residuals obtained from model is used as input for ANN model to obtain forecasting results. Evaluation methods used in research are RMSE, MAPE and MAE. Hybrid model gave better forecast results followed by ANN and SARIMA respectively. RMSE , MAE and MAPE for hybrid model is 1.5, 1.1 and 0.065% respectively. The research also suggests ANN models perform better forecast in comparison to forecast obtained from time series models due to non- linearity in data . Also, researchers postulated that neural network are expected to give better forecast results when exposed to exogenous factors in modeling. (Reikard and Hansen; 2019).

2.4 Machine Learning models

ANN is successfully applied in many areas in the field of machine learning. It is typically used in forecasting, image and object detection problems, recommend systems, among others. It has performed better than typical regression and time series models used in forecasts, especially in area of solar forecasting. There is extensive work done in area of solar forecasting and a lot of literature is availabe in said area. Renno et al. (2016) used neural networks in research for hourly BNI and daily GHI forecast. Study employed exogenous factors, namely, meteorological factors such as mean temperature, humidity, wind speed and precipitation, and astronomical and geographical factors such as sunshine hours, latitude, longitude and daylight hours for GHI prediction. For BNI prediction clearness index, time series data of BNI, hour angle is used for forecast. Modeling architecture includes Multilayer perceptron (MLP) neural network with error back propagation(BP) along with optimization using levenberg-marquadt(LM) for both area of prediction. MAE, MAPE and RMSE are employed to evaluate forecast results. Various combination of model layers, input exogenous variables and transfer function are used for prediction. The forecast result for daily GHI with MAPE and nRMSE of 4.45% and 0.04 respectively and hourly forecast for BNI with MAPE and nRMSE of 5.3% and 0.05 respectively is obtained. The study highlights ANN generate good forecast results for solar irradiation. Also, it emphasizes on importance of exogenous factors considered in research. But, research lacks mention of data pre processing techniques and methodology employed for implementation, an important aspect to reproduce results highlighted in study. Also, it merely mentions that normalization is undertaken for exogenous variables but does not mention details on model employed. Premalatha and Arasu (2016) compared forecast metrics obtained from different training model of back propagation, namely, gradient descent(GD), Lavenberg Marquadt(LM), resilient back propagation (RP) and scaled conjugate gradient (SCG) for forecast employing various exogenous factors. Data is normalized and ranged between (-1,1) using equation mentioned below, duly suggested in the research conducted by Solmaz and Ozgoren (2012).

$$X_N = 0.8\left(\frac{\mathrm{X}_R\text{-}X_{max}}{\mathrm{X}_{max}\text{-}X_{min}}\right)+0.1$$

where,
X_N: normalized value;
X_r: original value;
X_{max}: maximum value of variable;
X_{min}: minimum value of variable.

Evaluation criteria used for implementation shows that LM algorithm have the best forecast metrics. The RMSE, MAE, MAPE and R reported for the study are 0.82, 1.08, 4.2%, 0.9 respectively that asserts Lavenberg Marquadt(LM) works well in conjunction with MLP for solar forecast. The study performs iterations for different number of neurons to obtain optimal solution for given ANN model. The emphasize of the study is to determine the best optimization model for solar forecast and concentrates its effort on result metrics from different optimization techniques rather than overall methodology and implementation of research . Ozoegwu (2019) carried out study of solar energy availability using hybrid Non-linear autoregressive neural network(NAR) and ANN model. The MLP with LM training model is used with structural input month number. Input for NAR is time series solar irradiation data. The output data forecast of solar radiation obtained from NAR is used as input into ANN model combined with month number as exogenous variable. The evaluation metrics of the proposed hybrid model is compared against NAR and NARX model with month number as the structural input used. Hybrid model is known to give superior forecast result as compared to both NAR and NARX. While NAR predicted the solar energy availability for a location with metric R^2=0.7, RMSE=0.7 MJ/m^2 and R=0.89, hybrid ANN predicted with metrics R^2=0.97, RMSE=0.3 MJ/m^2 and R=0.9. MAPE obtained for 16-month ahead NAR forecast is 8.4% while that obtained for hybrid ANN forecast is 5.7%. The optimization is performed based upon Lavenberg Marquadt equation(LM) model for each implementation

Literature review has highlighted the importance of factors, namely, astronomical and climatic factors such as cloud cover, sunshine hours, clear sky index, hour angle as important predictors for BNI and GHI forecast.Also, it highlights the usability of machine learning model for obtaining accurate forecast results. Time series models, namely, ARIMA and SARIMA are effective in forecasts with hybrid models of time series and ANN proving more accurate in forecast metrics. There is lot of research performed in area GHI forecast, while research in BNI forecast are performed with moderate level of success and concentrated on short horizon forecast owing to intermittent weather patterns and high dependence of BNI on climatic factors, namely, cloud cover and sky clearness index that changes intermittently. Hence, literature review fulfils objective 1 of project objectives mentioned in table1.

2.5 Literature Review Summary

Table 2 below summarizes different modeling techniques for GHI and BNI forecast with metrics listed for relevant researches.

Sno.	Authors	Modeling Techniques	Results
1	(Benali et al.; 2019)	Application of ANN and Random forest to predict BNI, GHI and Beam diffuse irradiation for short term hour ahead forecast	GHI for Average RMSE=130Wh/m^2, Average MAE=100Wh/m^2, BNI for Average RMSE=270 Wh/m^2 and Average MAE=200Wh/m^2
2	(Chiteka and Enweremadu; 2016)	Prediction of GHI using ANN for day ahead forecast.	Forecast Day ahead MAPE= 8%
3	(Reikard and Hansen; 2019)	Forecast GHI and clearness index for short term forecast using ARIMA Time series models.	Average RMSE for minutes ahead forecast=100.3W/m^2 approximately and Average MAE for minutes ahead forecast=63.8W/m^2 approximately. Average RMSE for hour ahead forecast = 110 W/m^2 and Average MAE=74.6 W/m^2
4	(Ozoegwu; 2019)	Forecast Monthly Solar Energy based on Month Number and Time series using Hybrid NAR ANN model and ANN models	Hybrid model Average MAPE=5.67% and for NAR Average MAPE=9%

5	(Mukaram and Yusof; 2017)	Forecast monthly Solar Radiation using SARIMA, ANN and Hybrid SARIMA ANN model.	SARIMA Average RMSE=1.82 MJ/m^2, Average MAE=1.48MJ/m^2 and MAPE=8.3%. ANN Average RMSE=1.98MJ/m^2, Average MAE=1.5MJ/m^2 and MAPE=8.79%. Hybrid model Average RMSE =1.49MJ/m^2, Average MAE=1.0921MJ/m^2, MAPE=6.48%
6	(Renno et al.; 2016)	Forecast Daily GHI and Hourly BNI using MLP with Back Propagation(BP) and LM algorithm	Daily GHI MAPE=4.46%, nRMSE= .043 and Hourly BNI MAPE=5.4%, nRMSE=.05

Table 2: Table of Literature Review Summary

With documentation of literature review and critical analysis of studies in field of research, **chapter 1 objective 1** in **table of objectives** 1 is attained.

3 Global Horizontal Irradiation and Beam Normal Irradiation Methodology and Design

3.1 Introduction

The project implementation is based on Knowledge Discovery Database(KDD) as proposed by (Fayyad et al.; 1996). Subsequent sections outlines the methodology approach used in implementation and a three tier architecture used for design and implementation, detailing steps followed and technologies used in the implementation. Also, insights into dataset with exploratory data analysis(EDA) is detailed in sections below.

3.2 Global Horizontal Irradiation and Beam Normal Irradiation Methodology and Design

The research methodology used in project is shown in figure 1 below. Various steps used in the methodology are explained in subsections below.

3.2.1 Data Extraction and Ingestion

Data for the research is downloaded from public domain site data world [1] with weather and geographical predictor variables, released by department of energy for government of China. These factors include atmospheric pressure, sky cover metrics, temperature metrics, among other factors along with GHI and BNI time series data. Grain of data is at hourly level. Data is extracted into CSV file for various locations of China.It is imported into Postgres database for cleaning and processing.

[1]https://data.world/doe/solar-hourly-solar-dni-ghi

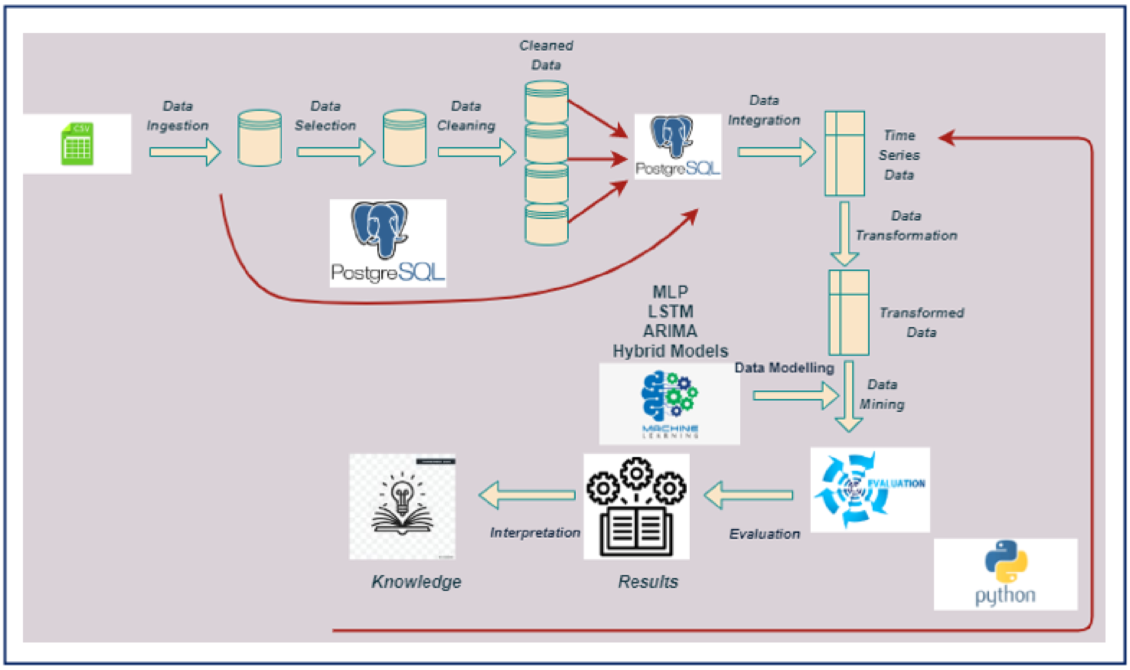


Figure 1: GHI and BNI Prediction Methodology

3.2.2 Data Selection and Cleaning in Postgres

Data is ingested into Postgres table LND_BUGT_DATA for Bugt location in China. The grain of data is at hourly level and required fields is selected for analysis. Data rolled up to day level data and night hour data where radiation is zero is removed from analysis, and finally, mean is taken to roll up. Daily Data is placed into BUGT_DATA_DAILY.

3.2.3 Data Integration

Once, cleaning and roll up is done, data is integrated into Python using Postgres Connector to fetch data from database. Checks for missing data in the attributes and exploratory data analysis is performed on the dataset. **SQL Alchemy** and **Psycopg** package of python is used to connect Postgres with Python into a pandas dataframe.

3.2.4 Data Transformation Techniques Used

Time series data is converted into scaled numerical data to provide as input to Neural Networks. Time Series Data is transformed using MinMax scaler function of scikit learn package in python and data is then converted into supervised machine learning dataset using pandas shift function. Text data is converted into three dimensional data using Numpy library to be used as input for LSTM and other neural network implementation as it require three dimensional input for processing.

3.2.5 Data Modeling

Models used in research are Long Short Term Memory(LSTM), Multi-Layer Perceptron(MLP), ARIMA and hybrid ARIMA ANN model. Models are described as below.

- **LSTM** - It is a type of Recurrent Neural Network(RNN) used vigorously in field of time series forecast and other machine learning application. Unlike standard feed-forward network, it has feedback loop and predicts catering historical predictions.
- **MLP** - It is a class of feedforward artificial neural network and consists of three layers, namely, input layer, hidden layer and output layer. Each node comprises of neuron, with exception to input layer, activated by non linear activation function. MLP utilizes supervised learning models of back propagation for training the neural network.
- **ARIMA** - It is statistical model that captures different temporal structures in a time series data for prediction and forecast. It uses previous or historical time series data to predict the data for ahead time forecast range.
- **Hybrid ARIMA ANN**- It uses combination of both ARIMA and ANN model for prediction. Predictions are made from ARIMA model and residuals obtained are fed to artificial neural network which uses these residuals along with training data to forecast and predict.

3.2.6 Evaluation Metrics Used

Root mean square error(RMSE), normalized Root Mean Square Error(nRMSE), Mean Squared error(MSE), Mean absolute Error(MAE), Mean Absolute Percentage Error(MAPE) and Mean forecast Error(Bias) are used as evaluation metrics to determine model performance. The equation for metrics are mentioned below, where,

y_i = Predicted value of variable
x_i = Actual value of variable
n = Number of observations

RMSE $= \sqrt{\sum_{i=1}^{n} \left(\frac{\mathrm{y}_i \text{ - } x_i}{\mathrm{n}}\right)^2}$

MAE $= \frac{1}{\mathrm{n}} \sum_{i=1}^{n} |y_i - x_i|$

MAPE $= \frac{100}{\mathrm{n}} \sum_{i=1}^{n} |\frac{\mathrm{y}_i \text{ - } x_i}{\mathrm{x}_i}|$

3.3 Design Architecture and Technologies Used

The Research is carried out in three tier architecture. Figure 2 below details the architecture with technologies used for implementation. Three layers of design are, namely, Client tier, Logical Layer and Data Persistent layer. Client tier are potential data capture points for the research where data is fetched from public domain website into Csv files. It is then fed to logical layer comprising of Database system i.e. Postgres, used for cleaning and merging of data. Time Series data is fed into Python for machine learning and time series model application, namely, LSTM, ARIMA, MLP and Hybrid ARIMA ANN, using Tensorflow and Keras libraries constituting backflow to Logical layer. Evaluation metrics are then obtained,analyzed and discussed to complete the back flow to client tier.

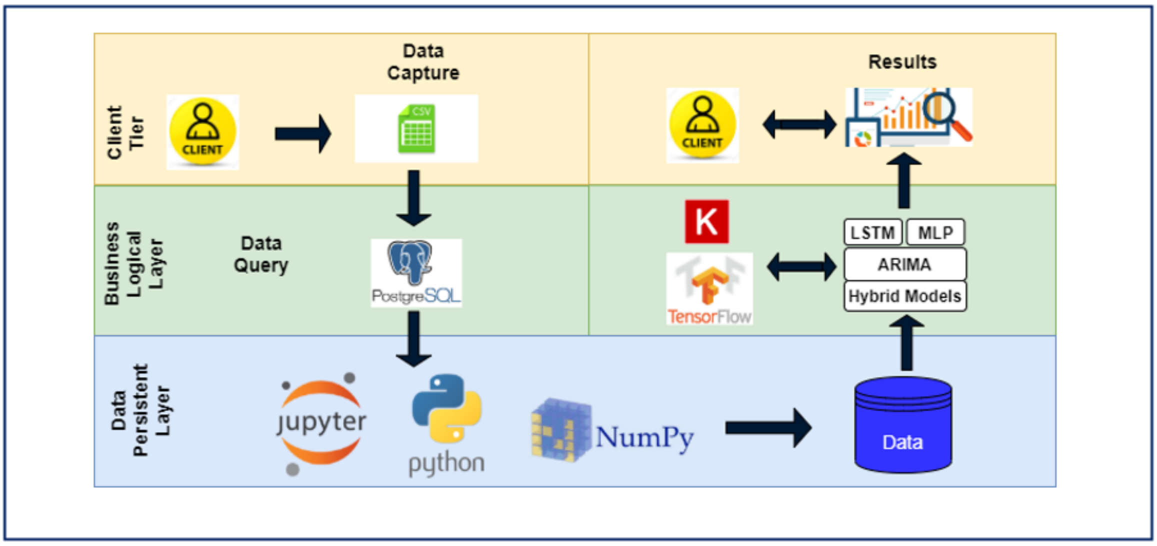


Figure 2: Data Flow Design Architecture

3.4 Exploratory Data Analysis

The section details important analysis required for time series data analysis and exploring relation among predictor variables, and understanding correlation among them. This analysis fulfils objective 2 of project objectives listed in table 1.

3.4.1 Null and Duplicate Value checks

The figure 3 below shows there is no missing value in the dataset. Also, there is no duplicate value in the dataset as well.

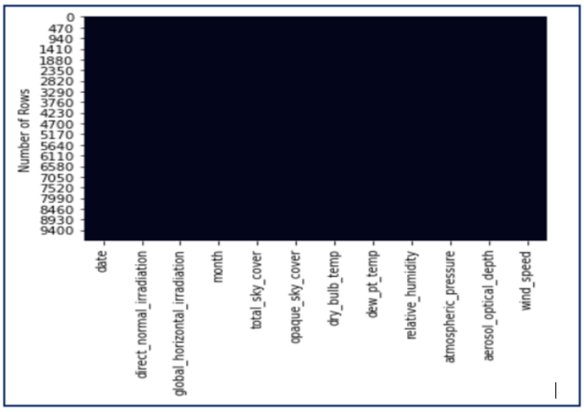


Figure 3: Null Value Check

3.4.2 Correlation Plot

Correlation plot shows strength of linear relationship among exogenous variables. The correlation plot in figure 4 below clearly shows strong relationship between Beam Normal Irradiation and predictor variable such as total sky cover and opaque sky cover. Similarly, for global horizontal irradiation, there is strong relationship with dry bulb temperature and dew point temperature.

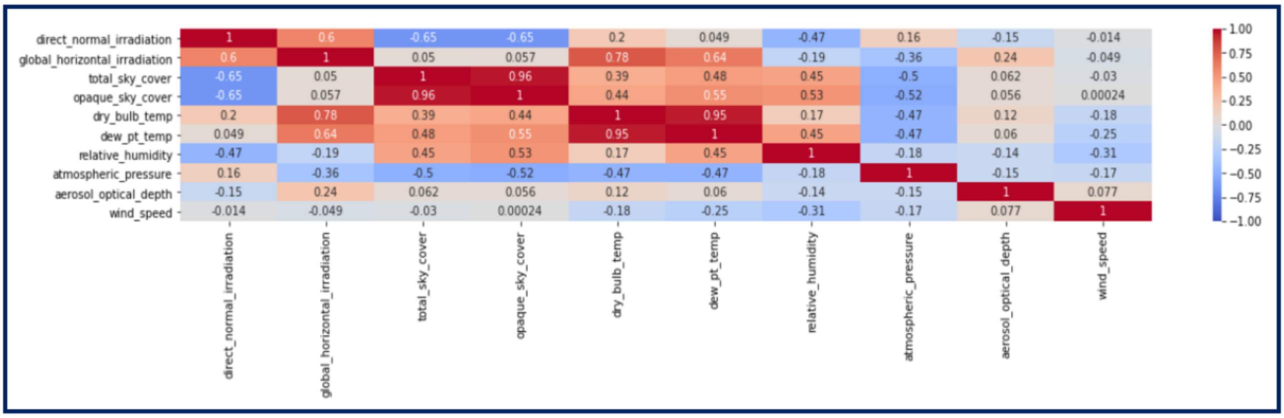


Figure 4: Correlation Plot

3.4.3 Feature Importance

Another important aspect to understand important exogenous factors in study is feature importance plots. In this research random forest regressor is employed to understand non-linear relationship among said factors and rank predictors in order of importance for prediction. The figure 5 and figure 6 below shows feature importance plots for both GHI and BNI prediction respectively. Month and opaque sky cover are important predictors for BNI, while for GHI prediction time series GHI data and opaque sky cover are important features.

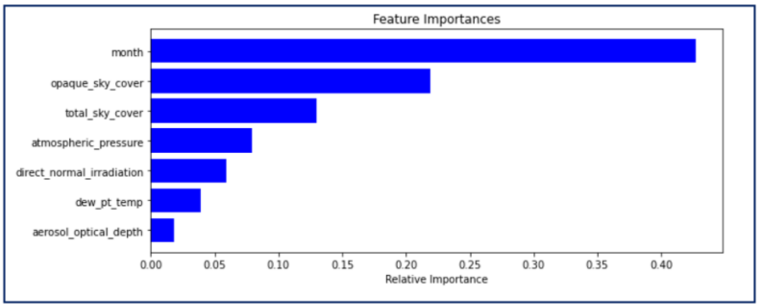


Figure 5: Feature Importance Plot BNI

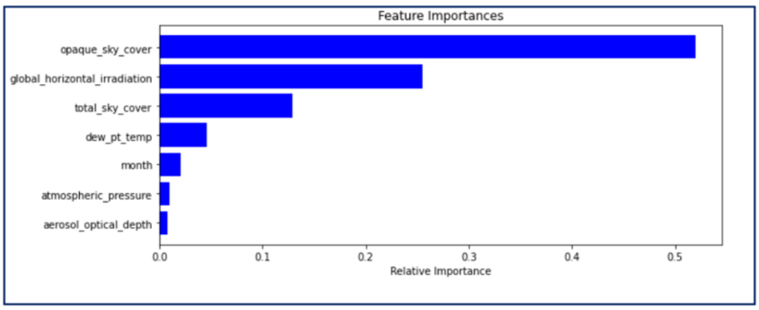
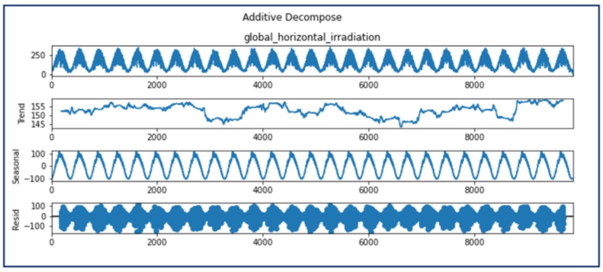


Figure 6: Feature Importance Plot GHI

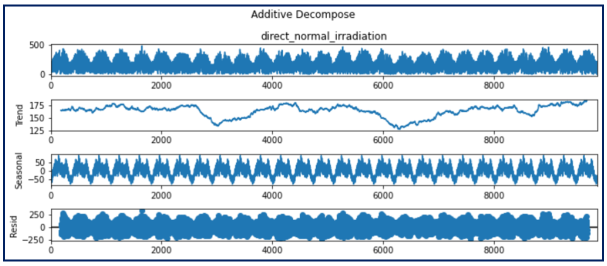
3.4.4 Stationarity, Seasonality and Trend in Time Series Data

There is seasonality in data for GHI and no trend is present in the data. It is evident from figure 7 below for additive decompose. Similarly, for BNI, there is seasonality in data but no trend present as evident from figure 8 below.



	seas	trend	resid	actual_values
0	-104.768751	NaN	NaN	45.13
1	-105.512382	NaN	NaN	62.17
2	-103.015545	NaN	NaN	50.75
3	-102.989702	NaN	NaN	45.71
4	-102.381771	NaN	NaN	57.21

Figure 7: Seasonal Decompose for GHI



	seas	trend	resid	actual_values
0	-64.060191	NaN	NaN	35.26
1	-63.027305	NaN	NaN	208.42
2	-53.440800	NaN	NaN	101.58
3	-60.123166	NaN	NaN	52.67
4	-58.127155	NaN	NaN	157.08

Figure 8: Seasonal Decompose for BNI

Data for GHI and BNI is stationary as conformed by Dicker Fuller test for Stationarity. The null hypothesis states that time series data is not stationary. Test was carried out at 5% confidence interval and null hypothesis is rejected.

4 Implementation for Global Horizontal Irradiation and Beam Normal Irradiation Prediction

This section details about the implementation steps carried out in the research in accordance to plan laid out in methodology and design. The following sections are mainly divided into data pre-processing, feature selection and modeling architecture for each implementation.

4.1 Data Pre-processing in Postgres

Data Pre-processing is performed in two stages. Data cleaning is done in Postgres and data transformation is done in python framework for model application. So, data processing in Postgres includes below steps.

- Data ingestion into Postgres and removing fields not required for research. Only exogenous factors, and GHI, BNI data are kept in tables for implementation.
- Combine fields to form date timestamp column, namely, hour, day, month, year.

- Convert hourly data for GHI and BNI into daily grain data by taking mean.

With this implementation, **chapter 1 objective 2** in **table of objectives** 1 is attained. Both Exploratory data analysis and data pre-processing is successfully completed.

4.2 Feature Selection for Prediction

An important aspect for multivariate time series forecast is estimation of important exogenous features. This is undertaken in section 3.4, namely, correlation analysis to check for linear relationship and regression analysis using random forest regressor to determine non-linear relationship at play between variables. Using both analysis, it is determined that opaque sky cover and total sky cover have strong relationship with forecast variable, i.e. GHI and BNI. Also, time series data of GHI is important predictor for GHI forecast while Month number is important predictor for BNI forecast.

4.3 Data Transformation and Normalization

Data transformation techniques employed in the study are introduced in section 3.2.4. MinMax Scaler function is used to scale values between 0 and 1 for exogenous factors and predicted. Also, seasonality from GHI and BNI is removed by differencing seasonality obtained from seasonal decomposition of time series data as depicted in section 3.4.4. Once, these steps are undertaken, time series dataset is converted into supervised machine learning dataset using pandas shift function. The idea is to forecast at time interval t using value at instance (t-1). Another important transformation is conversion of data into three dimensional data as LSTM model needs 3-D input. These inputs represents data or sample, time steps and features. Time step is set as 1 as single lag is taken and 9 features are taken for prediction.

4.4 Model Implementation

This section details about various modeling techniques employed for prediction. As mentioned above in section 3.2.5, LSTM, MLP, ARIMA and ARIMA ANN hybrid models are used for prediction in multivariate and univariate versions. Data is split into 80%, 5%, 15% for training, validation and test set for ANN models.Data split is 85% and 15% for training and test set for ARIMA model implementation.For each model execution loss function used on validation set is MAE. Keras callback functionality is used to identify the best model used for final prediction.

4.4.1 LSTM Implementation for Global Horizontal Irradiation Prediction

LSTM architecture is a feedforward neural network with feedback loop to cater to historical data. In this research, LSTM layer is coupled with two dense layers. Baseline model is built on LSTM layer and two dense layers are added to improve computation and evaluation metrics. Neurons used are 50,64,1 for LSTM layer and two dense layers respectively. Iterations are carried out to arrive at optimal neurons for each layer. Relu is used as Activation function in two dense layers. Model is executed for 100 epochs, 72 batch size with learning rate set to 0.001. Figure 9 below shows model layers, and plot loss on validation and training data. Model has converged and there is no over-fitting.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 50)                12000
_________________________________________________________________
dense (Dense)                (None, 64)                3264
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
=================================================================
Total params: 15,329
Trainable params: 15,329
Non-trainable params: 0
_________________________________________________________________
```

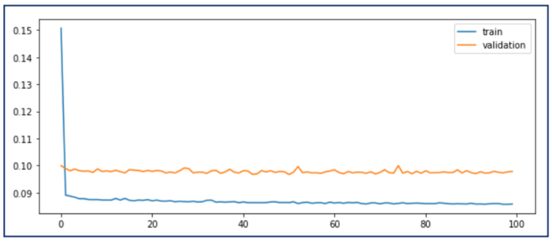


Figure 9: Model loss plot and Model layers LSTM

4.4.2 MLP Implementation for Global Horizontal Irradiation Prediction

MLP is feedforward neural network with input layer, hidden layer and output layer. In this implementation, flatten layer is used as input layer to convert three dimension data input into two dimension. Hidden layer comprise of 64 neurons while output layer comprises of 50 neurons. Iterations are done by varying number of neurons to find optimal solution. Activation function used is Relu and model is executed at 100 epochs and 72 batch size, and learning rate set to 0.001. Figure 10 shows model layers, and training, validation error plots. Model has converged and no signs of over-fitting is observed.

```
Model: "sequential_7"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
flatten_1 (Flatten)          multiple                  0
_________________________________________________________________
dense_13 (Dense)             multiple                  500
_________________________________________________________________
dense_14 (Dense)             multiple                  3264
_________________________________________________________________
dense_15 (Dense)             multiple                  65
=================================================================
Total params: 3,829
Trainable params: 3,829
Non-trainable params: 0
```

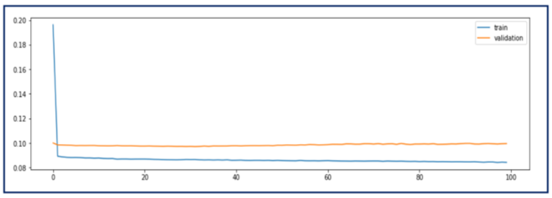


Figure 10: Model loss plot and Model layers MLP

4.4.3 ARIMA Model Implementation for Global Horizontal Irradiation Prediction

For implementation of ARIMA, estimation of p i.e. number of lag observations, d i.e. number of non seasonal differencing and q i.e. number of lag forecast error is determined using Auto Arima function that estimates the value based upon minimum AIC and BIC value to obtain optimal integral values. Also, partial auto correlation(PACF) and auto correlation plots(ACF) are analyzed to determine lag. For GHI, p, d, q is estimated as 10, 1, 0 respectively. Figure 11 below shows ARIMA model residual statistics. that clearly signifies residuals can be further modelled.

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                 8382
Model:                ARIMA(10, 1, 0)   Log Likelihood              -41831.198
Method:                       css-mle   S.D. of innovations             35.573
Date:                Wed, 12 Aug 2020   AIC                          83686.396
Time:                        21:45:59   BIC                          83770.802
Sample:                             1   HQIC                         83715.220

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       7.641e-06      0.083   9.23e-05      1.000      -0.162       0.162
ar.L1.D.y     -0.6526      0.011    -59.971      0.000      -0.674      -0.631
ar.L2.D.y     -0.5765      0.013    -44.668      0.000      -0.602      -0.551
ar.L3.D.y     -0.5193      0.014    -36.660      0.000      -0.547      -0.491
ar.L4.D.y     -0.4763      0.015    -31.897      0.000      -0.506      -0.447
ar.L5.D.y     -0.3912      0.015    -25.465      0.000      -0.421      -0.361
ar.L6.D.y     -0.3423      0.015    -22.283      0.000      -0.372      -0.312
ar.L7.D.y     -0.2864      0.015    -19.182      0.000      -0.316      -0.257
ar.L8.D.y     -0.2168      0.014    -15.310      0.000      -0.245      -0.189
ar.L9.D.y     -0.1486      0.013    -11.517      0.000      -0.174      -0.123
ar.L10.D.y    -0.0862      0.011     -7.930      0.000      -0.108      -0.065
                                    Roots
```

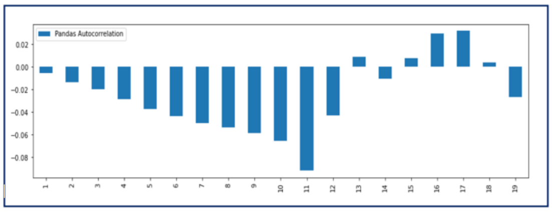


Figure 11: ARIMA Model Residual Statistics

4.4.4 Hybrid ARIMA ANN Model Implementation for Global Horizontal Irradiation Prediction

The approach of prediction is based on analyzing non-linear relationships in ARIMA residuals. As mentioned above in figure 11, lags is present in PACF plot of residuals that can be further modeled. Figure 12 below shows flowchart of hybrid model implementation and training, validation error loss. Residuals obtained from ARIMA forecast are fed into ANN model architecture for prediction, which in this research is MLP. Input layer comprises of 50 neurons while hidden and output layer layer comprises of 64 and 1 neuron respectively and activation function used is Relu. Predictions of ARIMA and residual forecast obtained from ANN model are added to form final forecast value.

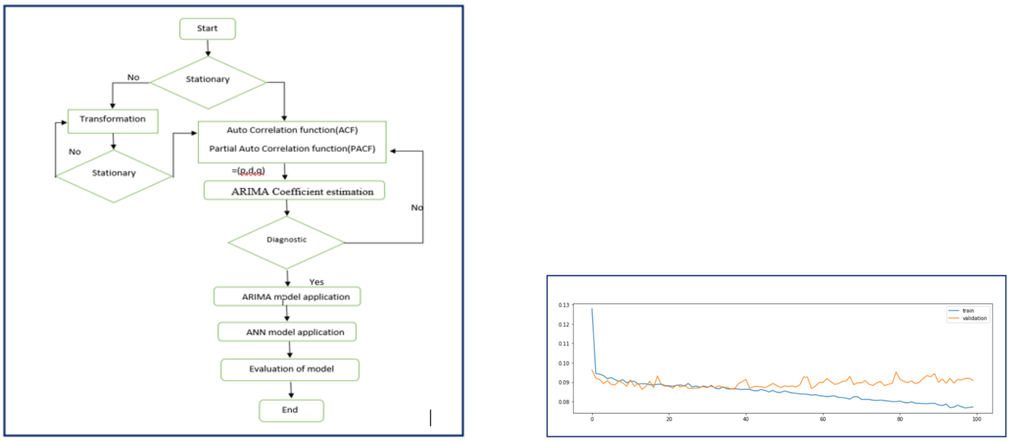


Figure 12: Hybrid Model flowchart and Model Loss plots

As evident, model converged well for intial epochs and then model tend to overfit on training dataset. 14 epoch is identified as best fit for implementation. Hence, it is used for prediction.

4.4.5 LSTM Model Implementation for Beam Normal Irradiation Prediction

For prediction of BNI, LSTM layer is combined with two dense layers to improve on computation, and thus, improve forecast results. Neurons used in implementation are 50,64,1 for LSTM and two dense layers respectively. Relu is used as activation function in two dense layers and learning rate is set to 0.001 with ADAM used as optimizer. Figure

13 below shows validation and training loss plot along with layer architecture of model. Clearly, model has converged and there is no sign of over-fitting.

```
Model: "sequential_23"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_8 (LSTM)                (None, 50)                11000
_________________________________________________________________
dense_34 (Dense)             (None, 64)                3264
_________________________________________________________________
dense_35 (Dense)             (None, 1)                 65
=================================================================
Total params: 14,329
Trainable params: 14,329
Non-trainable params: 0
_________________________________________________________________
```

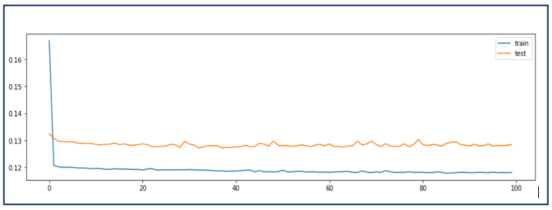


Figure 13: Model loss plot and Model layers LSTM

Model is executed for 100 epochs, 72 batch size with Keras callback. Best model is identified among 100 epochs and used for prediction.

4.4.6 MLP Model Implementation for Beam Normal Irradiation Prediction

Flatten layer is used as input layer to convert three dimension data input into two dimension. Hidden layer comprises of 64 neurons while output layer comprises of 50 neurons. Iterations are carried out by varying number of neurons is done to find optimal solution. Activation function used is Relu and model is executed at 100 epochs and 72 batch size with learning rate set to 0.001. ADAM optimizer is used in compilation of model.Figure 14 shows model layers, and training, validation plots. Loss plot clearly shows model has converged and there is no signs of over-fitting.

```
Model: "sequential_24"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
flatten_4 (Flatten)          multiple                  0
_________________________________________________________________
dense_36 (Dense)             multiple                  250
_________________________________________________________________
dense_37 (Dense)             multiple                  3264
_________________________________________________________________
dense_38 (Dense)             multiple                  65
=================================================================
Total params: 3,579
Trainable params: 3,579
Non-trainable params: 0
```

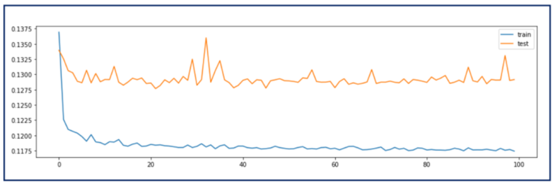


Figure 14: Model loss plot and Model layers MLP

4.4.7 ARIMA Model Implementation for Beam Normal Irradiation Prediction

This implementation follows the same procedure as mentioned in section 4.4.3 for ARIMA implementation in GHI prediction. Model with p,d,q of 10,1,0 is used for prediction. These parameters are decided analyzing PACF and ACF plots, and Auto ARIMA function estimates.Figure 15 shows residual statistics for ARIMA implementation and PACF plot for residual obtained. It clearly shows residual for BNI could be modeled further as there is still lag of 12-13 present.

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                    D.y   No. Observations:                 8875
Model:                ARIMA(10, 1, 0)   Log Likelihood              -52261.464
Method:                       css-mle   S.D. of innovations             87.323
Date:                Thu, 13 Aug 2020   AIC                         104546.927
Time:                        02:24:51   BIC                         104632.019
Sample:                             1   HQIC                        104575.902

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0058      0.194      0.030      0.976      -0.375       0.386
ar.L1.D.y     -0.6661      0.011    -62.956      0.000      -0.687      -0.645
ar.L2.D.y     -0.6051      0.013    -47.850      0.000      -0.630      -0.580
ar.L3.D.y     -0.5415      0.014    -38.701      0.000      -0.569      -0.514
ar.L4.D.y     -0.4801      0.015    -32.406      0.000      -0.509      -0.451
ar.L5.D.y     -0.4122      0.015    -27.064      0.000      -0.442      -0.382
ar.L6.D.y     -0.3458      0.015    -22.707      0.000      -0.376      -0.316
ar.L7.D.y     -0.2885      0.015    -19.478      0.000      -0.318      -0.259
ar.L8.D.y     -0.2190      0.014    -15.644      0.000      -0.246      -0.192
ar.L9.D.y     -0.1349      0.013    -10.662      0.000      -0.160      -0.110
ar.L10.D.y    -0.0830      0.011     -7.840      0.000      -0.104      -0.062
```

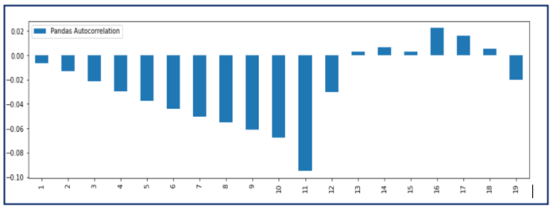


Figure 15: ARIMA Model Residual Statistics

4.4.8 Hybrid ARIMA ANN Model Implementation for Beam Normal Irradiation Prediction

Implementation follows similar strategy as discussed in section 4.4.4. As highlighted in section 4.4.7, residual could be further modeled to obtain forecasts. Hence, they are modeled using ANN architecture. MLP architecture is used for implementation. Input layer comprises of 50 neurons while hidden and output layer layer comprises of 64 and 1 neuron respectively and activation function used is Relu. Predictions of ARIMA and residual forecast obtained from ANN model are added to form final forecast value. Figure 16 below shows loss plots for training and validation set. As shown in figure above, model

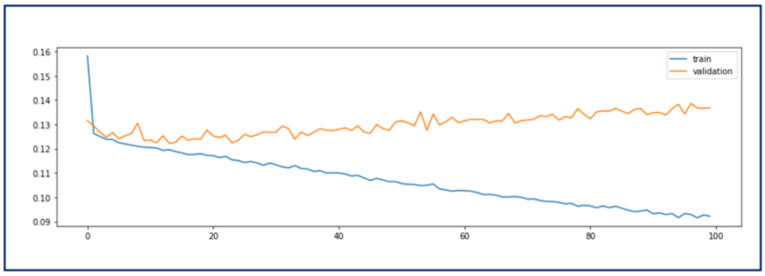


Figure 16: Model loss plot for hybrid model BNI

converges for initial epochs and then over-fits on training dataset as number of epochs increases. Model of epoch 14 is used for prediction.

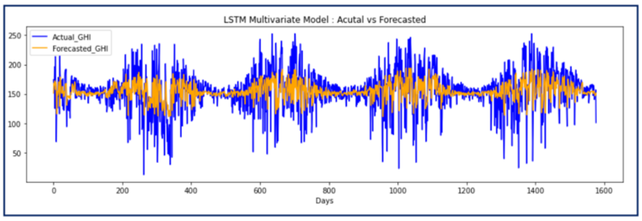
With implementation of all models, **chapter1 objective 3.1, 3.2, 3.3, 3.4** in table of objectives 1 is completed for implementation part.

5 Evaluation, Results and Discussion

Evaluation metrics for the models used are RMSE, MAE and MAPE. Also, line plots for predicted value versus actual value would be analyzed. Based upon these metrics each model is compared with each other. Also, results obtained for models would be compared in detail with model metrics mentioned in table 2 in literature review section.

5.1 Evaluation Metrics LSTM for Global Horizontal Irradiation Prediction

Evaluation metrics for LSTM model implementation for GHI and line plot for actual versus predicted value is shown in figure 17 below.



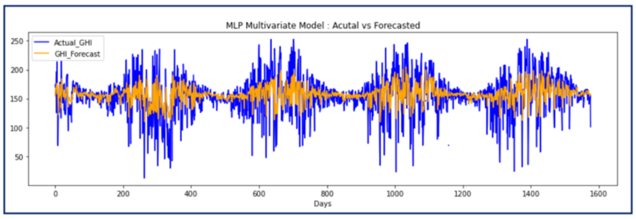
Evaluation Metrics	Values
Root Mean Square Error(RMSE)	32.475
Mean Absolute Error(MAE)	23.338
Mean Absolute Percentage Error(MAPE)	19.94

Figure 17: Evaluation Metrics LSTM for GHI

Yellow line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction is 32.475 Wh/m^2, 23.338 Wh/m^2, 19.94% respectively.

5.2 Evaluation Metrics MLP for Global Horizontal Irradiation Prediction

Evaluation metrics for MLP model implementation for GHI and line plot for actual versus predicted value is shown in figure 18 below.



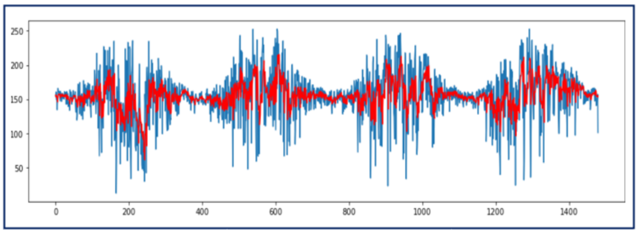
Evaluation Metrics	Values
Root Mean Square Error(RMSE)	32.468
Mean Absolute Error(MAE)	23.02
Mean Absolute Percentage Error(MAPE)	19.94

Figure 18: Evaluation Metrics MLP for GHI Prediction

Yellow line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction is 32.468 Wh/m^2, 23.020 Wh/m^2, 19.94%.

5.3 Evaluation Metrics ARIMA for Global Horizontal Irradiation Prediction

Evaluation metrics for ARIMA model implementation for GHI prediction along with line plot for actual versus predicted value is shown in figure 19 below.Red line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction is 34.838 Wh/m^2, 24.737 Wh/m^2, 21.13% respectively.

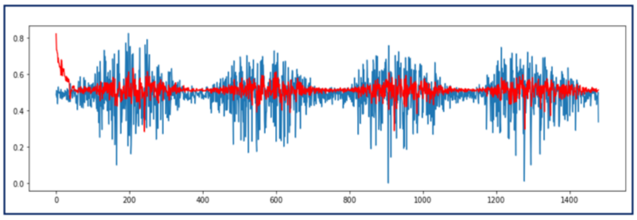


Evaluation Metrics	Values
Root Mean Square Error(RMSE)	34.838
Mean Absolute Error(MAE)	24.737
Mean Absolute Percentage Error(MAPE)	21.13

Figure 19: Evaluation Metrics ARIMA for GHI Prediction

5.4 Evaluation Hybrid ARIMA for Global Horizontal Irradiation Prediction

Evaluation metrics for hybrid ARIMA ANN model implementation for GHI and line plot for actual versus predicted value of residuals is shown in figure 20 below.



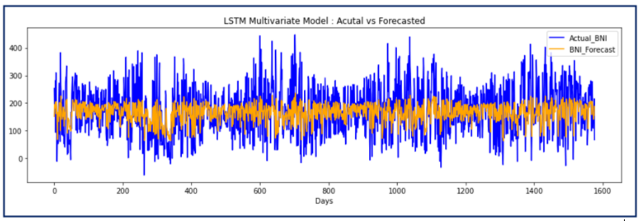
Evaluation Metrics	Values
Root Mean Square Error(RMSE)	38.344
Mean Absolute Error(MAE)	27.771
Mean Absolute Percentage Error(MAPE)	23.711

Figure 20: Evaluation Metrics Hybrid ARIMA for GHI Prediction

Red line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction are 38.344 Wh/m^2, 27.771 Wh/m^2, 23.71%.

5.5 Evaluation Metrics LSTM for Beam Normal Irradiation Prediction

Evaluation metrics for LSTM model implementation for BNI and line plot for actual versus predicted value is shown in figure 21 below.



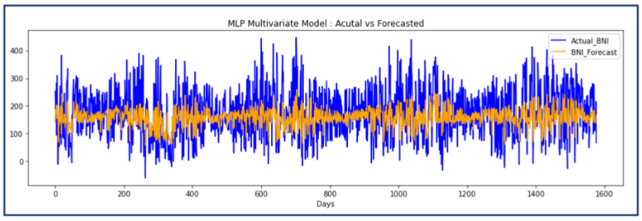
Evaluation Metrics	Values
Root Mean Square Error(RMSE)	85.838
Mean Absolute Error(MAE)	69.846
Mean Absolute Percentage Error(MAPE)	42.88

Figure 21: Evaluation Metrics LSTM for BNI

Yellow line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction is 85.838 Wh/m^2, 69.846 Wh/m^2, 42.88% respectively.

5.6 Evaluation Metrics MLP for Beam Normal Irradiation Prediction

Evaluation metrics for MLP model implementation for BNI and line plot for actual versus predicted value is shown in figure 22 below.



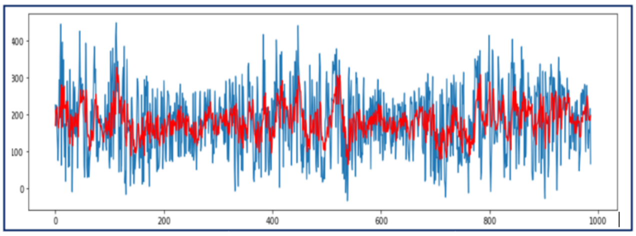
Evaluation Metrics	Values
Root Mean Square Error(RMSE)	86.427
Mean Absolute Error(MAE)	70.208
Mean Absolute Percentage Error(MAPE)	41.23

Figure 22: Evaluation Metrics MLP for BNI Prediction

Yellow line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction is 86.477 Wh/m^2, 70.208 Wh/m^2, 41.23%.

5.7 Evaluation Metrics ARIMA for Beam Normal Irradiation Prediction

Evaluation metrics for ARIMA model implementation for BNI prediction along with line plot for actual versus predicted value is shown in figure 23 below.Red line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction is 92.358 Wh/m^2, 75.490 Wh/m^2, 73.46%.

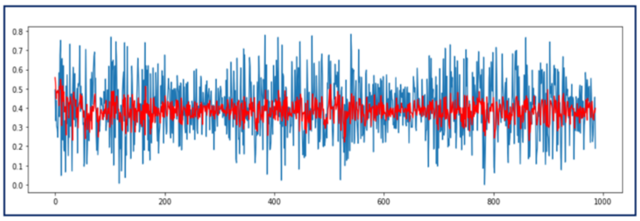


Evaluation Metrics	Values
Root Mean Square Error(RMSE)	92.455
Mean Absolute Error(MAE)	75.503
Mean Absolute Percentage Error(MAPE)	73.46

Figure 23: Evaluation Metrics ARIMA for BNI Prediction

5.8 Evaluation Hybrid ARIMA for Beam Normal Irradiation Prediction

Evaluation metrics for hybrid ARIMA ANN model implementation for GHI and line plot for actual versus predicted value of residuals is shown in figure 24 below.



Evaluation Metrics	Values
Root Mean Square Error(RMSE)	97.206
Mean Absolute Error(MAE)	78.46
Mean Absolute Percentage Error(MAPE)	71.16

Figure 24: Evaluation Metrics Hybrid ARIMA for BNI Prediction

Red line represents the predicted value and blue line represents the actual value. RMSE, MAE and MAPE for the prediction are 97.206 Wh/m^2, 78.46 Wh/m^2, 71.16%.

With evaluation metrics, **chapter 1 objective 3.1, 3.2, 3.3, 3.4** in **table of objectives** 1 is completed.

5.9 Comparison of Developed Prediction Models and Discussion

It is clearly evident from metrics, neural network performed better than time series and hybrid model. Time Series models gave slightly better metrics as compared to hybrid models for GHI and BNI forecast. In neural network implementation, MLP performed marginally better in comparison to LSTM for both GHI and BNI prediction. Upon comparison with recent literature review, it can be ascertained that implementation for ARIMA model gave better forecasts in comparison to existing published study, i.e. (Reikard et al.; 2017). The model predicted GHI with RMSE of 34.838 W/m^2 while RMSE published in research is 100 W/m^2. Similarly, MLP and LSTM model in this study performed considerably better than those obtained in (Benali et al.; 2019). However, research carried out by (Chiteka and Enweremadu; 2016), (Renno et al.; 2016) and (Mukaram and Yusof; 2017) have better forecast metrics, i.e. MAPE between 5-9% for GHI prediction while this research predicted with MAPE 19.9%. This is due to application optimization technique i.e. Lavenberg-Marquadt(LM) used in these published research is known to works well with MLP model for solar forecasts. Hence, **chapter 1 objective 4 and objective 5 in project objectives table** 1 is completed.

6 Conclusion and Future Work

The research extensively explored area of solar forecast and various modeling techniques, namely, time series and machine learning is applied and compared in study. A modeling technique of hybrid ARIMA ANN is introduced and applied in the study. Upon evaluation and comparing results with those of time series and machine learning models, it is gathered that both time series and machine learning models have slightly better evaluation metrics than those of hybrid model. As stated in research question, it can be concluded that time series and machine learning model perform better than hybrid modeling approach introduced in the research and hybrid model does not enhance the performance of traditional time series and machine learning models. With this interpretation, **chapter 1 objective 6** of **project objectives** table 1 is complete. Also, all objectives of **project objectives** table 1 is complete based on which research has concluded and answered the research question effectively.

However, model has potential application in forecasting problems. It can further be taken up for research and optimization techniques such as Lavenberg-Marquadt(LM) equation could be employed to improve on forecast metrics. It is clear from literature review that models with optimization based on LM equation improves the forecast results extensively. Also, different architecture of feedforward ANN, for instance LSTM network and different architecture of Hybrid model approach could be explored as well for better results.

References

Alanazi, M., Mahoor, M. and Khodaei, A. (2017). Two-stage hybrid day-ahead solar forecasting, *North American Power Symposium* **112**(1): 474–485.

Arbizu-Barrena, C., Ruiz-Arias, J. A., Rodríguez-Benítez, F. J., Pozo-Vázquez, D. and Tovar-Pescador, J. (2017). Short-term solar radiation forecasting by advecting and diffusing msg cloud index, *Solar Energy* **155**(1): 1092–1103.

Benali, L., Notton, G., Fouilloy, A., Voyant, C. and Dizene, R. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components, *Renewable Energy* **132**: 871–884.

Chen, C., Xue, B., Cai, G., Thomas, H. and Stückrad, S. (2019). Comparing the energy transitions in germany and china: Synergies and recommendations, energy reports, *Energy Reports* **5**(1): 1249–1260.

Chiteka, K. and Enweremadu, C. (2016). Prediction of global horizontal solar irradiance in zimbabwe using artificial neural networks, *Journal of Cleaner Production* **135**(1): 701–711.

Fayyad, U. M., Piatetsky-Shapiro, G. and Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework., **96**: 82–88.

Ghimire, S., Deo, R. C., Downs, N. J. and Raj, N. (2019). Global solar radiation prediction by ann integrated with european centre for medium range weather forecast fields in solar rich cities of queensland australia, *Journal of Cleaner Production* **216**(5): 288–310.

Hassan, G. E., Youssef, M. E., Mohamed, Z. E., Ali, M. A. and Hanafy, A. A. (2016). New temperature-based models for predicting global solar radiation, *Applied Energy* **179**(1): 437–450.

Inman, R. H., Pedro, H. T. and Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration, *Progress in Energy and Combustion Science* **39**(6): 535–576.

Li, M., Chu, Y., Pedro, H. T. and Coimbra, C. F. (2016). Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term dni forecasts, *Renewable Energy* **86**(1): 1362–1371.

Mukaram, M. Z. and Yusof, F. (2017). Solar radiation forecast using hybrid sarima and ann model: a case study at several locations in peninsular malaysia, *Malays J Fundam Appl Sci Spec Issue Some Adv Ind Appl Math* **1**(1): 346–350.

Ozoegwu, C. G. (2016). New temperature-based models for reliable prediction of monthly mean daily global solar radiation, *Renewable and Sustainable Energy* **10**(2): 023706.

Ozoegwu, C. G. (2019). Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number, *Journal of Cleaner Production* **216**(1): 1–12.

Pecenak, Z., Mejia, F., Kurtz, B., Evan, A. and Kleissl, J. (2016). Simulating irradiance enhancement dependence on cloud optical depth and solar zenith angle, *Solar Energy* **136**(1): 675–681.

Pereira, R., Santos, C. S. and Rocha, A. (2019). Solar irradiance modelling using an offline coupling procedure for the weather research and forecasting (wrf) model, *Solar Energy* **188**(1): 339–352.

Premalatha, N. and Arasu, A. V. (2016). Prediction of solar radiation for solar systems by using ann models with different back propagation algorithms, *Journal of Applied Research and Technology* **14**(3): 206–214.

Reikard, G. and Hansen, C. (2019). Forecasting solar irradiance at short horizons: Frequency and time domain models, *Renewable Energy* **135**(1): 1270–1290.

Reikard, G., Haupt, S. E. and Jensen, T. (2017). Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models, *Renewable Energy* **112**(1): 474–485.

Renno, C., Petito, F. and Gatto, A. (2016). Ann model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building, *Journal of Cleaner Production* **135**(1): 1298–1316.

Sirch, T., Bugliaro, L., Zinner, T., Möhrlein, M. and Vazquez-Navarro, M. (2017). Cloud and dni nowcasting with msg/seviri for the optimized operation of concentrating solar power plants, *Atmospheric Measurement Techniques* **10**(2): 409—-429.

Solmaz, O. and Ozgoren, M. (2012). Prediction of hourly solar radiation in six provinces in turkey by artificial neural networks, *Journal of Energy Engineering* **138**(1): 194–204.

Wang, P., van Westrhenen, R., Meirink, J. F., van der Veen, S. and Knap, W. (2019). Surface solar radiation forecasts by advecting cloud physical properties derived from meteosat second generation observations, *Solar Energy* **177**(1): 47–58.