A NEURAL NETWORK BASED MODEL FOR FORECASTING THE POWER OUTPUT OF A COMMERCIAL SCALE PHOTOVOLTAIC POWER PLANT

M.A.A.P. Manchanayaka

(178253F)

Degree of Master of Engineering

Department of Mechanical Engineering

University of Moratuwa

Sri Lanka

April 2022

A NEURAL NETWORK BASED MODEL FOR FORECASTING THE POWER OUTPUT OF A COMMERCIAL SCALE PHOTOVOLTAIC POWER PLANT

M.A.A.P. Manchanayaka

(178253F)

Dissertation submitted in partial fulfillment of the requirements for the degree Master of Engineering

Department of Mechanical Engineering

University of Moratuwa

Sri Lanka

April 2022

Declaration

I declare that this is my own work, and this dissertation does not incorporate without acknowledgement, any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works such as articles or books.

Signature: UOM Verified Signature Date: 21/04/2022

M.A.A.P. Manchanayaka

The above candidate has carried out research for the Masters Dissertation under my supervision.

Signature: UOM Verified Signature Date: 21/04/2022

Dr. Indrajith D. Nissanka

Acknowledgement

The research was carried out under the supervision of Dr. Indrajith D. Nissanka, Senior lecturer Department of Mechanical Engineering, University of Moratuwa. I am very much grateful to Dr. Nissanka for giving me his immense support and invaluable guidance throughout this study. I am very much grateful to Dr. Himan Punchihewa, Head of the Department of Mechanical Engineering, University of Moratuwa, and the former course coordinator of MEng /PG Diploma of Energy Technology, for his extended guidance and encouragement from the first academic day of the programme to the very end. I would also like to thank Mr. Pubudu Pathirana, Senior Engineer, Panasian Power PLC for supporting me on collecting the data and other required information of Beliatta solar PV plant. Finally, I would like to acknowledge all the lecturers and supporting staff for their utmost support during the study of entire MEng programme.

Abstract

Solar photovoltaic (PV) is penetrating electrical grids with a substantial growth of new additions, as a result of renewable energy policies and plans implemented locally and globally. However, the intermittent nature of the availability of solar energy brings an uncertainty into electrical power systems making it complex for power management and integrating into existing electricity infrastructure. This has been a key issue in promoting renewable energy in developing countries. Accurate solar power forecasts in different time horizons can play a vital role to bring down the uncertainty by a significant margin. In this work, a neural network (NN) model was coupled with a decomposition and transposition (D&T) model to forecast day(s) ahead hourly PV output of a grid connected 1 MW solar PV plant located in Hambantota, Sri Lanka. Historical weather and solar radiation data for last 14 years were collected from two APIs (Application Programming interfaces) for the location of PV plant and variation of global horizontal irradiation (GHI) with percentage cloud cover, rain, temperature, relative humidity, and wind speed were analysed. The selected parameters from the analysis together with day and hour numbers were fed in to the NN model through a scaling layer and trained it using Levenberg–Marquardt backpropagation algorithm. Optimum NN model was selected by changing the hidden layer sizes and calculating the mean squared error. The forecasted GHI values of the optimized NN model were decomposed to diffuse horizontal irradiance (DHI) and direct normal irradiance (DNI) using Erbs correlation, as the first step of D & T model. Then, DHI and DNI components were converted to global tilted irradiance (GTI) using HDKR correlation, in order to calculate solar PV output, including possible plant specific losses. The correlation coefficient (R) between GHI output and target values of the trained NN model for an unseen testing data set was observed to be 0.86. For final model, mean percentage forecasting accuracy was observed to be 86% with 12% standard deviation. The model could be adopted to any commercial or utility scale solar PV plant which is in a tropical climate region.

Key words: Machine Learning Model, Neural Network, Solar Power Forecast, Solar PV, Utility scale power plant

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
ANN	Artificial Neural Network
API	Application Programming Interface
ARIMA	Autoregressive Integrated Moving Average
BFGS	Broyden–Fletcher–Goldfarb–Shanno formula
CCGT	Combined Cycle Gas Turbine
CEB	Ceylon Electricity Board
DFP	Davidon–Fletcher–Powell algorithm
DHI	Diffuse Horizontal Irradiance
EBH	Direct Horizontal Irradiance
FFNN	Feed Forward Neural Network
GA-ANN	Genetic Algorithm-based Artificial Neural Network
GHI	Global Horizontal Irradiance
GTI	Global Tilted Irradiance
HDKR	Hay-Davies-Klucher-Reindl model
ICE	Internal Combustion Engines
IEA	International Energy Agency
IEC	International Electrotechnical Commission
IPCC	Intergovernmental Panel on Climate Change
IRENA	International Renewable Energy Agency
LM	Levenberg-Marquardt algorithm
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
MIT	Massachusetts Institute of Technology
MLP	Multilayer Perceptron
MLR	Multiple linear regression
MSE	Mean Squared Error
NCRE	Non-Conventional Renewable Energy

NOAA	National Oceanic and Atmospheric Administration
NOCT	Nominal Operating Cell Temperature conditions
OCGT	Open Cycle Gas Turbine
PV	Photovoltaic
RH	Relative Humidity
RMSE	Root Mean Squared Error
RNN	Recurrent neural network
SD	Standard Deviation
SE	Squared Error
STC	Standard Test Condition
STL	Seasonal-Trend decomposition procedure based on Loess
UTC	Universal coordinated Time
VRE	Variable Renewable Energies
WRC	World Radiation Center

CHAPTER 1

INTRODUCTION

This chapter includes a detailed background study of evolution of world's primary energy sources from the past to present status and the future inclinations towards renewable energy. The discussion is focused on the rising of fossil fuels as the primary source, global issues developed with the increase of usage of fossil fuels, such as pollution, climate change and sustainability issues. Further, the discussion explores the way solar power generation emerge as the fastest growing renewable energy technology in recent times, globally and locally. The final quarter of the chapter describes the challenges associated with high solar power share in electricity generation and the importance of accurate solar power forecasting which is the main goal of this study.

1.1 Brief History of World's Energy

Energy is the prime factor required to drive the world forward. Earth receives energy from sun which is the very first energy source. With the evolutionary process of humans, the fire had been discovered and used for the purposes such as cooking, lighting, and generate heat to keep their body warm in cold environments. For this they have burnt wood which was the easily available energy source around them. Thousands of years ago people have realized that they can use the energy of wind for sailing across oceans to reach different destinations and to discover the world. From 2000 years ago, the Greeks have used waterwheels for grinding grains. Some research reveal that the windmills have been used in Persia from ninth century for the same purpose. Harvesting energy from animals for food production and transport was an essential option, almost everywhere in the world [1].

It can be understood that the energy sources used by human being for their activities were purely renewable in ancient times. But with the exponential growth of the population, the energy demand has also increased and low efficiency and less effectiveness of using the above-mentioned energy sources for some human activities became a barrier to move forward. Industrial revolution in US and Europe which was begun at the tail end of the 18th century brought new inventions, manufacturing processes and industries. As a result of this, the global energy demand surged and the transition from use of renewable based energy sources to Fossil fuels such as coal was begun. At the end of 18th century, the main energy source used was biomass, however compared to biomass, coal has high specific energy capacity and most importantly coal was available in areas where biomass was not readily available. Also, coal was best suited as an energy source for most of the outcomes of industrial revolution. With this the growth of the demand of coal was increased rapidly and peaked at the end of the 19th century where crude oil was invented. Even though history of natural gas runs to as early as 500 BC in China, the production and use of natural gas began to increase significantly with the evolvement of crude oil extraction technologies in the latter part of the 19th century. Since then, coal, oil and gas (fossil fuel or hydrocarbon-based energy sources) dominated in world energy share during the entire 20th century [1]. Figure 1.1 shows a history from 1750 BC, how different energy sources competing each other to fulfill the growth of world energy demand.

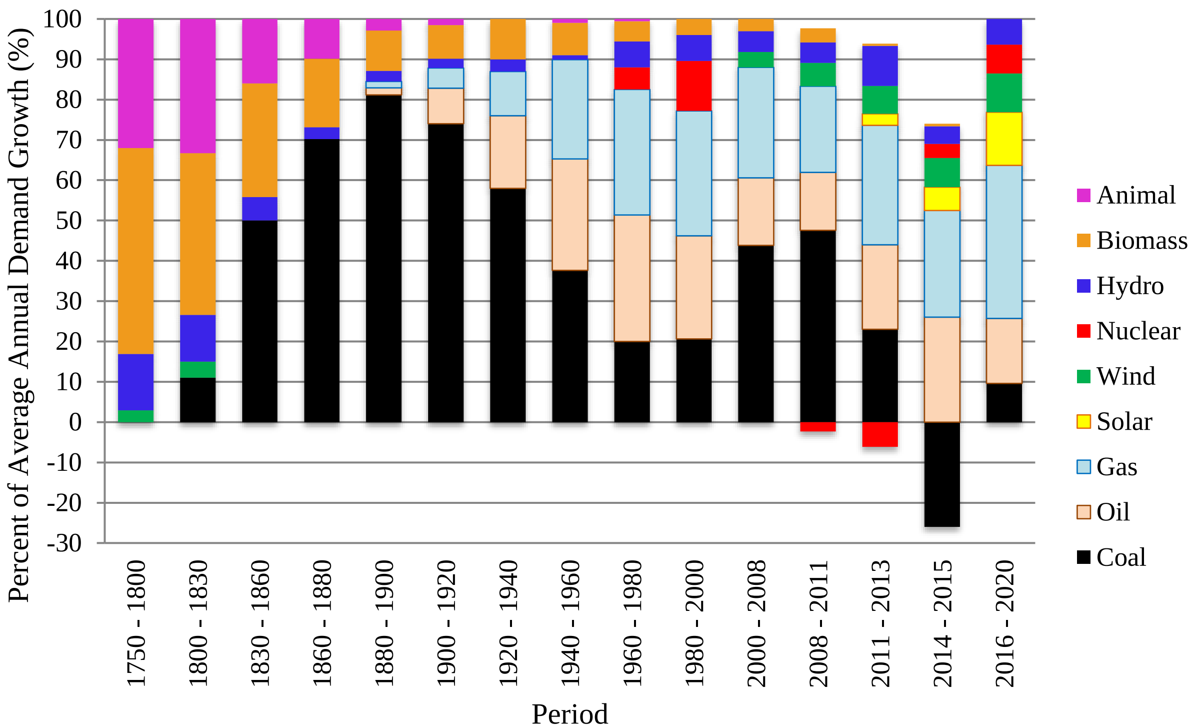


Figure 1-1: Percentage growth of global energy supplied by source [1]

1.2 Dominance of Fossil Fuels as the Primary Energy Source

World total energy consumption was estimated to less than 500 Mtoe at the beginning of the 19th century and it was taken about a century to double the consumption [2]. The 19th century is very significant in the history of world energy because share of the fossil fuel energy sources rapidly increased from 2% to 49% mainly due to growth of coal production. This share was further increased to 77% at the end of the 20^{th} century, which means 8110 Mtoe of energy from 10496 Mtoe total energy consumption was supplied from fossil fuel-based energy sources. The contribution from renewable sources was 1770 Mtoe at this point and seems almost impossible to catch-up the contribution given by fossil fuel-based energy sources in any means [2]-[4]. This can be better viewed from the Figure 1.2 which gives the world energy consumption for fossil fuel and renewable energy sources, separately from year 1800 to 2019. Nuclear energy also came into the picture from 1965, however concerns over the radioactive waste and associated risk factors resulted to almost stuck this sector for more than two decades from year 2000, throwing the future of nuclear energy into a serious doubt so far. Therefore, this discussion is focused on the fossil fuels and renewables from here onwards.

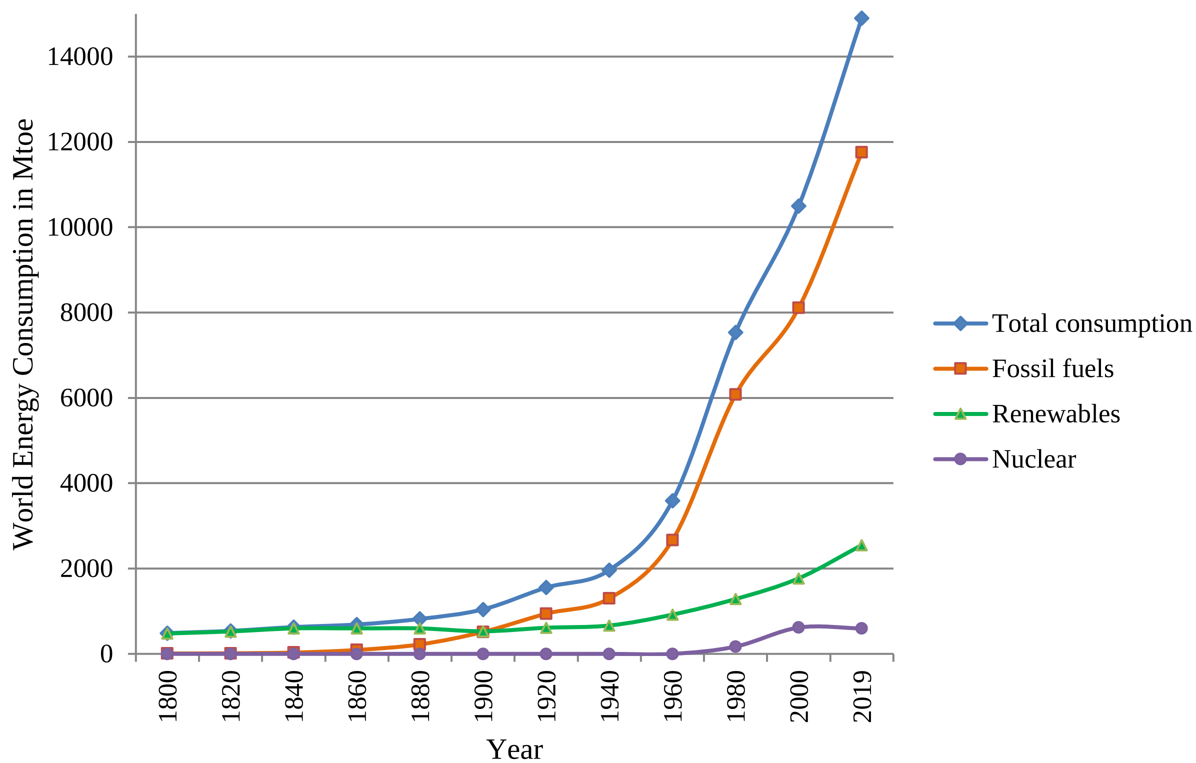


Figure 1-2: World Energy Consumption - Total, based on fossil fuels, renewables and nuclear from 1800 to 2019 [2]–[4].

There are several possible reasons for the dominance of fossil fuels over the other energy sources. The main reason is that all of them are incredibly 'energy dense' or acting as 'high dense energy reservoirs'.

Typical energy density values for wood: 16 MJ/kg, coal: 23.9 - 25 MJ/kg, crude oil: 42 - 47 MJ/kg, natural gas: 42 - 55 MJ/kg [5]. Thus, as an energy source, fossil fuels are more efficient compared to other sources. Secondly, the fossil fuels have been created through an entirely natural process for millions of years and can be easily harvested using the modern-day technologies. Because of this reason, the fossil fuel energy sources can be considered as 'ready-made' fuels. However, comparing with the other energy sources such as wind, geothermal and solar energy, the effectiveness of these sources depends on how efficiently they are collected, transformed, and stored for a useful work.

1.3 Global Issues of Using Fossil Fuels

Use of fossil fuels has major drawbacks for the long-term sustainability of the world. It is the major source for environmental pollution and climate change, on the other hand the fossil fuels are non-renewable.

1.3.1 Pollution and Climate Change

Pollution based drawbacks are associated with fuel extraction, transportation, processing and finally with the end use which is the combustion of the fuel. For example, coal excavation demolishes lands and for several decades these lands will be out of use. In the case of oil and gas, spills and leakages occur during extraction, transportation and storage processes causing air and water pollution. Oil and gas refining also involves environmental pollution, however the major impact of the fossil fuels occurs during the end use. Generation of electricity, heating, transportation or whatever the purpose, the end use of all types of fossil fuels is combustion. As a result of this, various unwanted combusted products and chemical compounds such as SO_x, NO_x, CO_x, CH, soot and ash, tar droplets are released to the environment causing air pollution. Air pollution has many adverse effects on the health of human being, animals, and crops. CO_2 and greenhouse gasses such as methane, nitrogen oxide and chlorofluorocarbons which are produced by combustion of fossil fuels, absorb infrared solar radiation reflected from earth's surface causing global warming. Global warming is believed to be the main reason for melting of glaciers and polar ice caps, sea level rise, climate changes and extreme weather conditions. Further, the above-mentioned primary pollutants may undergo chemical reactions with water and other atmospheric compounds, resulting the formation of secondary pollutants such as ozone, aerosols and various acids [6].

CO_2 is emitting to the atmosphere in different ways such as from solid waste, trees, and other biological materials, burning of wood and from certain chemical reactions in manufacturing processes. However, all these emissions are negligible compared to the emission from burning of fossil fuels. Global CO_2 emission from fossil fuel combustion was 0.03 billion metric tons in the year 1800. It has been increased to 1.95 billion metric tons in the year 1900, after another 100 years it has been skyrocketed to 25.12 billion metric tons in the year 2000 [7]. According to Figure 1.3 the global CO_2 emission is almost entirely a function of the global fossil fuel consumption.

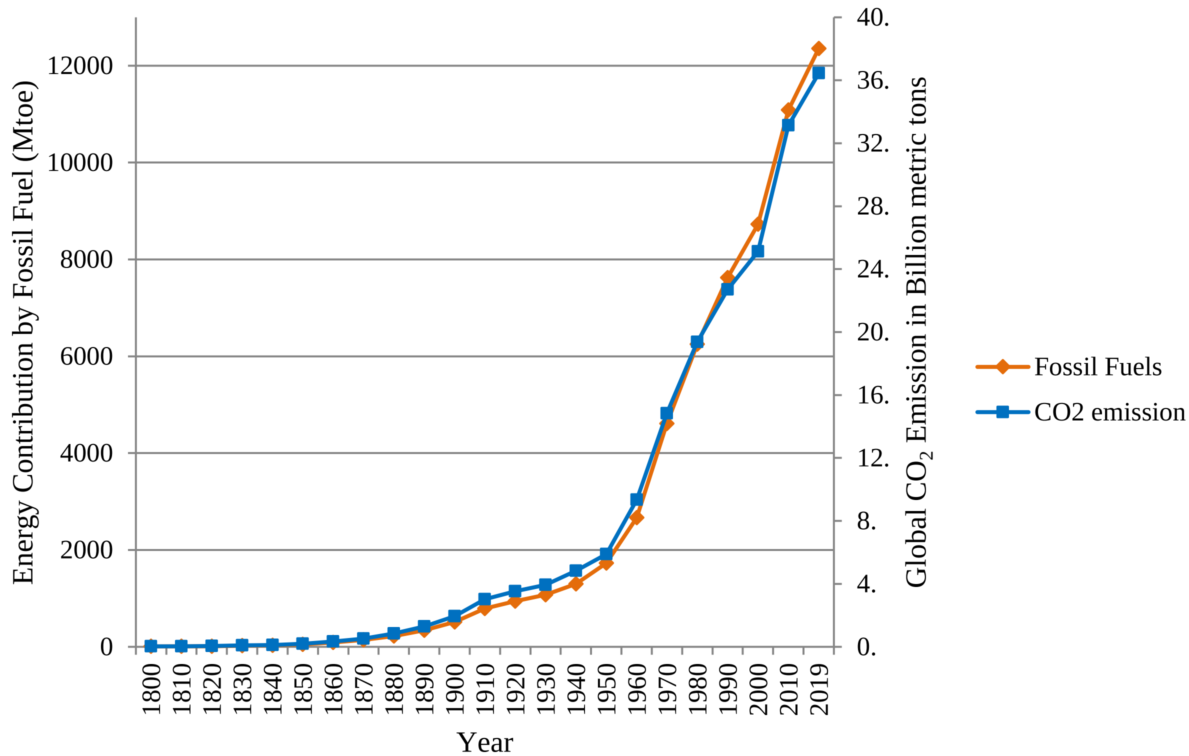


Figure 1-3: Comparison of global CO_2 emission with world energy contribution by burning fossil fuel [2]-[4], [7].

Since CO_2 is the primary greenhouse gas which causes global warming, the impact of fossil fuel consumption on the global temperature change is clearly interconnected. The global warming for last 140 years can be better visualized in the mean temperature anomalies plot shown in Figure 1.4. For this graph the base line (mean) temperature has been calculated for the period from 1901 to 2000 and temperature anomaly is the difference of the particular year's temperature from the base line temperature. Positive anomaly indicates that the observed temperature is warmer than the base line and negative anomaly indicate that the observed temperature is cooler than the base line temperature. The mean temperature anomaly plot suggests that the global temperature has been increased exponentially for last 140 years and this increase is becoming a huge threat for the sustainability of the earth system.

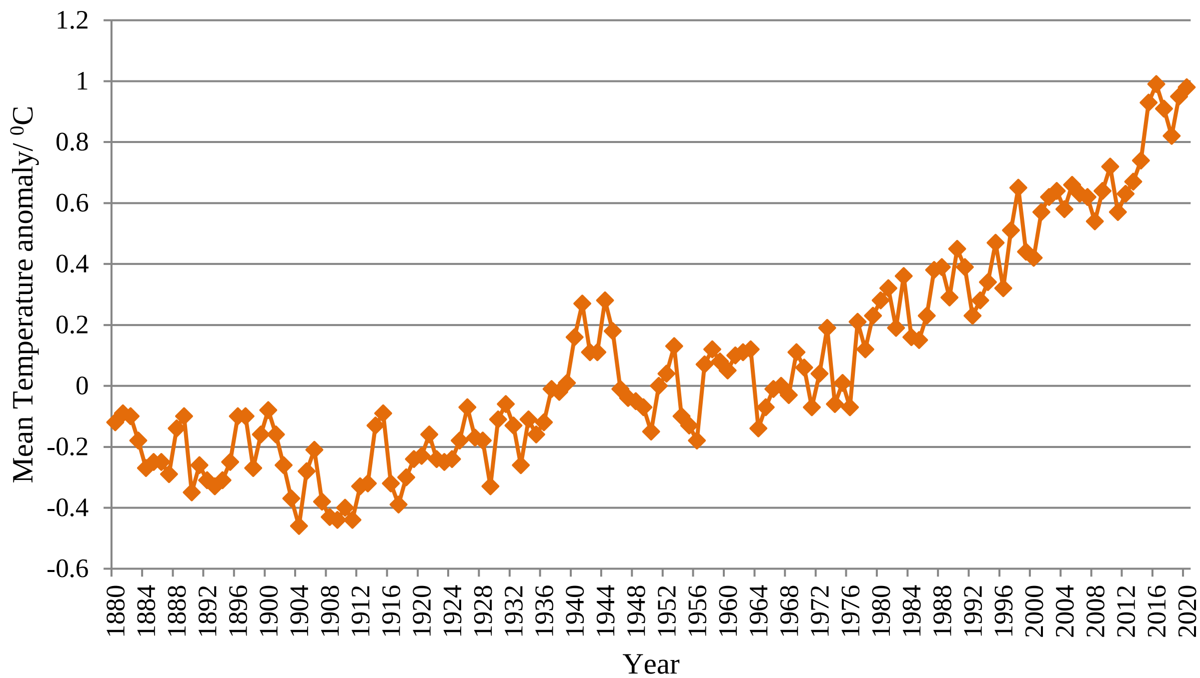


Figure 1-4: Annually mean temperature anomalies relative to the base period 1901 – 2000 [8], [9]

The global temperature change was a discussion among international community since in 1970's, but global warming became a significant debate among the public and political arena worldwide in 1988. IPCC (Intergovernmental Panel on Climate Change) was established and met for the first time, to examine the effect of greenhouse gasses on global warming and climate change. Since then, numerous efforts have been taken and policies have been put in place globally to control and minimize CO_2 and other greenhouse gas emission. The Kyoto Protocol in 1997 which targeted to bring down CO_2 and greenhouse gases emission in main industrialized countries by at least 5% below 1990 levels for five years commitment period starting from 2008, The Paris agreement in 2015 which aimed to limit the global temperature increase well below 2 ^{0}C in this century compared to level of pre-industrial era, are some of the highlights of those efforts [10], [11]. The importance of keeping global temperature rise below 1.5 ^{0}C was again foregrounded in the COP 26 UN Climate Change Conference, took place in November 2021 in Glasgow [12].

1.3.2 Sustainability of Using of Fossil Fuels

Forming of fossil fuels is a completely natural process and it is believed that decomposing plants and animals underground through millions of years resulted the formation of coal, oil and natural gas. However, compared to the time duration they take to originate, the consumption rate is fast and the reserves are depleting rapidly. Some researchers predict the world fossil fuel reserves will end within this century with the continues increase of use of fossil fuels. According to BP's Statistical review of world energy 2020, remaining oil and gas reserves last for next 50 years, and coal reserves last for next 132 years [13].

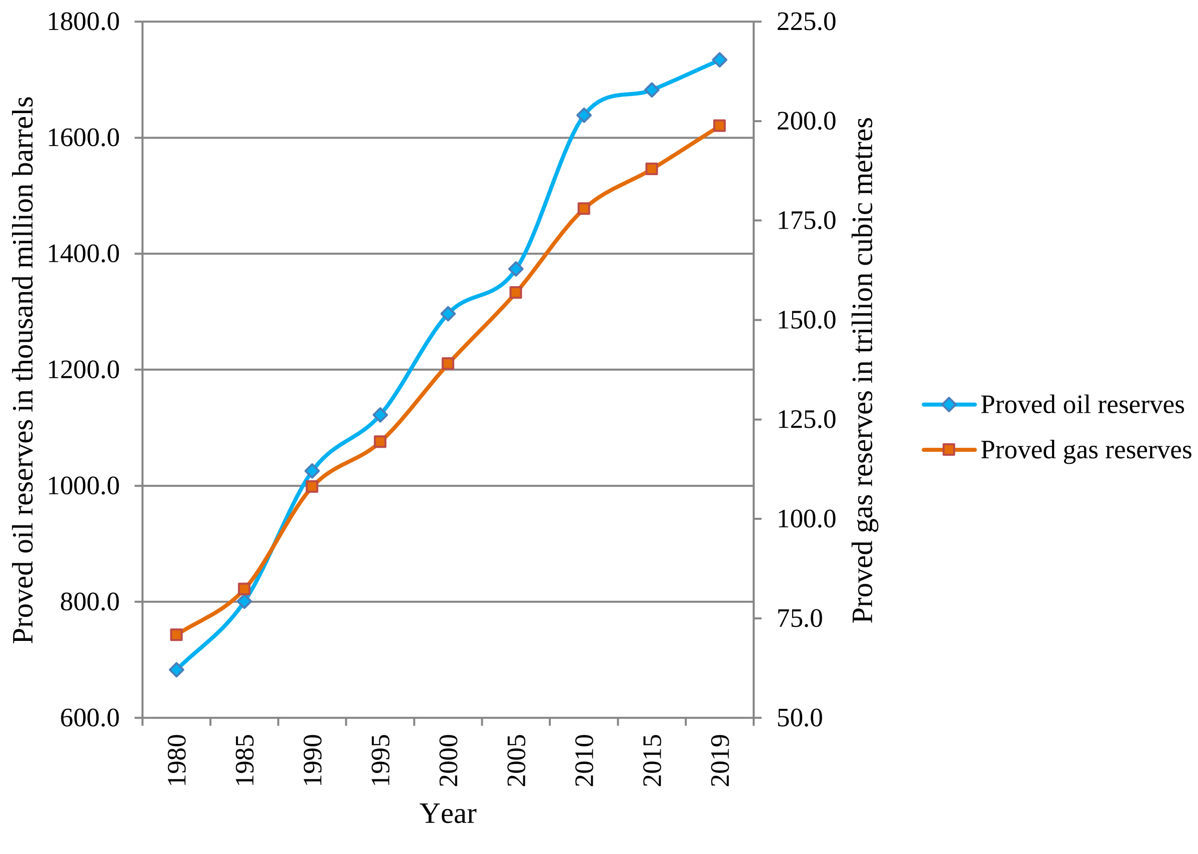


Figure 1-5: History of world's proved oil & gas reserves for 1980 - 2019 period [13]

However, world proven fossil fuel energy reserves keep increasing for years due to the advancing technologies which allow to discover reserves which are previously unknown or inaccessible.

Specially the offshore discoveries have been increased significantly in the recent history and overtaken the energy potential of those found on land in the same period. The increase of oil and gas reserves for the last three decades are shown in Figure 1.5. Because of these discoveries of new deposits, it is very difficult to predict the time frame which the fossil fuel reserves going to be last, however for sure a finite volume of deposits is available in the earth.

1.4 Energy Transition from Fossil Fuels to Renewables

As mentioned previously the earth's fossil fuel deposits are a result of concentrating solar and biomass energies for million years. Since the invention of these fuels, the consumption has been increased exponentially with the population growth and the improving of living standards, resulting the continuous depletion of the fossil fuel reserves and initiation of climate change.

The concerns over the consumption of fossil fuels as the primary energy source in the world popped up time to time from about 50 years back. In 1972, a project group from MIT (Massachusetts Institute of Technology) led by Dr. Dennis L. Meadows published the report *"The Limits to Growth*" highlighting the concerns on both depletion of fossil fuels and climate change [14]. This report was an eye opener for most of the industrialized countries who consume fossil fuels heavily. Establishment of the International Energy Agency (IEA) in 1974, aiming to work globally on energy policies and to ensure a sustainable future, is considered as an outcome of the Meadows report. The policy framework set by IEA initially, triggered the use of renewable energies directly or indirectly, also lots of countries started to invest heavily on improving the available renewable energy extracting technologies and inventing new technologies. The renewable energy sector was further boosted by Kyoto protocol in 1997 and Paris agreement in 2015 [15].

Outcomes of different global policy frameworks and protocols do not reflect immediately form world energy consumption patterns which have been illustrated in Figure 1.6. Considering the history of renewables, traditional biomass was dominating the global renewable energy share for centuries and still continuing on. Contribution of hydropower has been increased with the invention of the turbine technology in the middle of the 19th century which was well before the global awareness on greenhouse gas emissions and depletion of fossil fuel deposits. In 1974 when IEA was establishing, 18 percent of global energy consumption was from renewable sources. 99% of this 18 percent was supplied by traditional biomass (72%) and hydropower (27%). After 26 years, in 2000 the share of renewables was dropped to 17% due to the exponential growth of the world energy demand, even though the renewable energy consumption was increased from 13733 TW h to 20590 TW h within this period. However, no significant contribution from the renewables other than traditional biomass and hydropower was observed during this period since the renewable share still dominated by traditional biomass (61%) and hydropower (36%) [3].

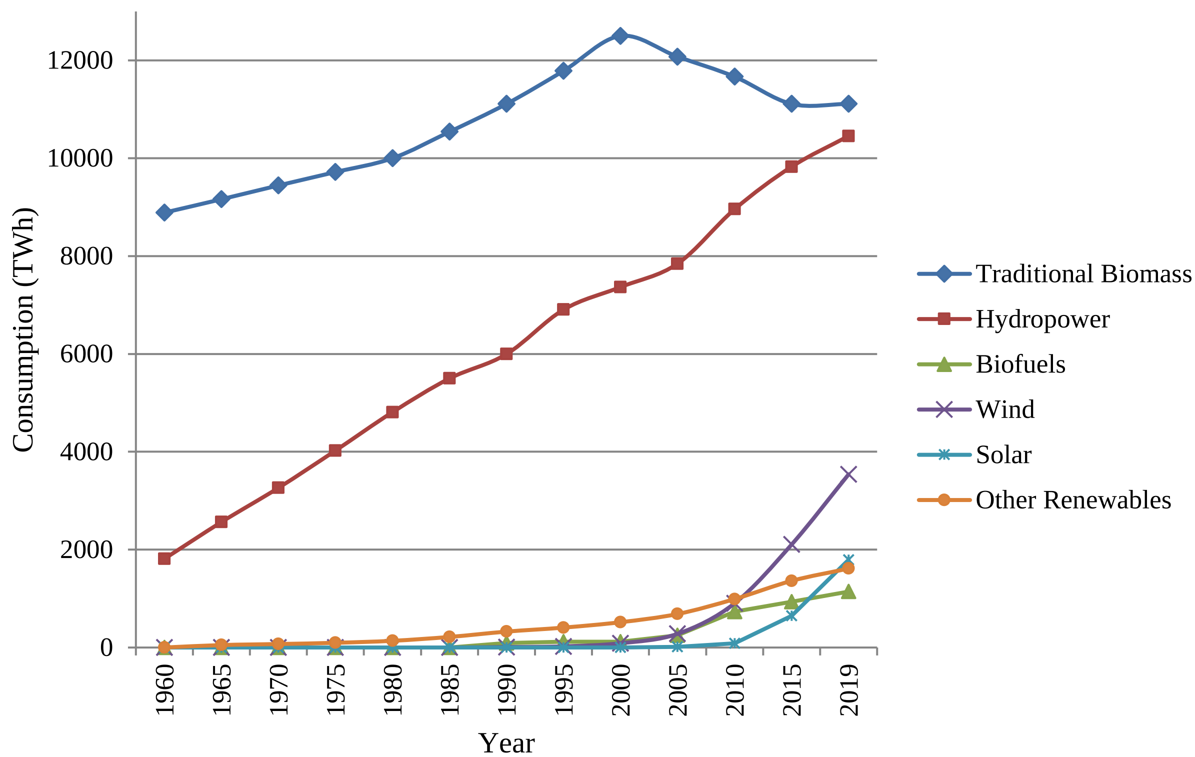


Figure 1-6: World share of renewable energy consumption 1960 - 2019 [3]

However, during this period, lot of background research and technological advancements have been carried out on different renewable energy sources and existing technologies around the world. A noticeable boost on the contribution can be observed from wind energy since year 2000 and from solar energy since year 2010. Rate of consumption growth of solar and wind is higher than any of the other renewables, contribution to the world energy consumption by both dominated the sources other than traditional biomass and hydropower at the end of year 2019. Increasing trend of wind and solar energy capacity is an incredibly good sign specially for the future of global electricity generation. In the year 2000, the global electrical energy supply was around 26 000 TW h, from which 7 500 TW h came from renewables. The renewable share was still dominated by hydropower with 4 400 TW h, and the contributions of wind and solar were 1600 TW h and 850 TW h respectively, which are higher than the contribution of other sources [3].

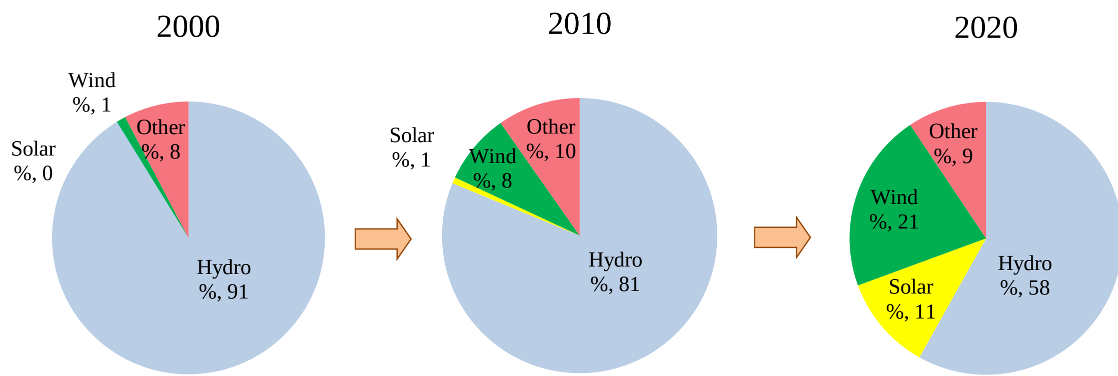


Figure 1-7: Evolving of Solar and Wind share in Global Electricity Production by renewables for last two decades [3]

1.5 Solar PV Energy - The Fastest Growing Renewable Energy Technology

Solar energy is the primary energy source of the earth. Direct use of solar energy for heating and lighting has a long history. Well-designed solar energy absorbing systems such as concentrated solar power systems are being used in the modern world for some heating and power generation applications. However, the technologies of converting solar energy into a useful energy form such as electricity was remained rather undeveloped or expensive for a long period of time.

The first silicon - single crystal photovoltaic (PV) cell to convert solar energy into electricity was built at Bell Labs in United State in 1954. The conversion efficiency of this cell was 6%. The efficiency value hits 40% or more in modern solar cells, however it drops to the range 15%-20 % when it comes to the panel efficiency. The cost associated in solar PV sector was very high in the early stages. It was at high as 77 dollars/W even after 20 years of first solar cell production in 1954. However, as a combined effect of enforced global environmental goals, technology advancements and increasing manufacturing volumes, the cost has reduced to as low as 0.3 USD/ W nowadays depending on the type of panel used [16], [17].

Combination of low cost, increased efficiency, and durability of solar photovoltaic attracted more and more attention of the people in recent years. This has resulted to lift solar PV to the fastest growing energy source by the global annual capacity additions for the last five years among the leading renewable energy sources which contribute to the world electrical energy supply. Annual capacity additions by different renewable sources for last 7 years are shown in Figure 1.8, as given in Renewables 2021 Global Status Report published by REN 21, only solar PV is having the annual capacity additions exceeding 90 GW for the last consecutive four years.

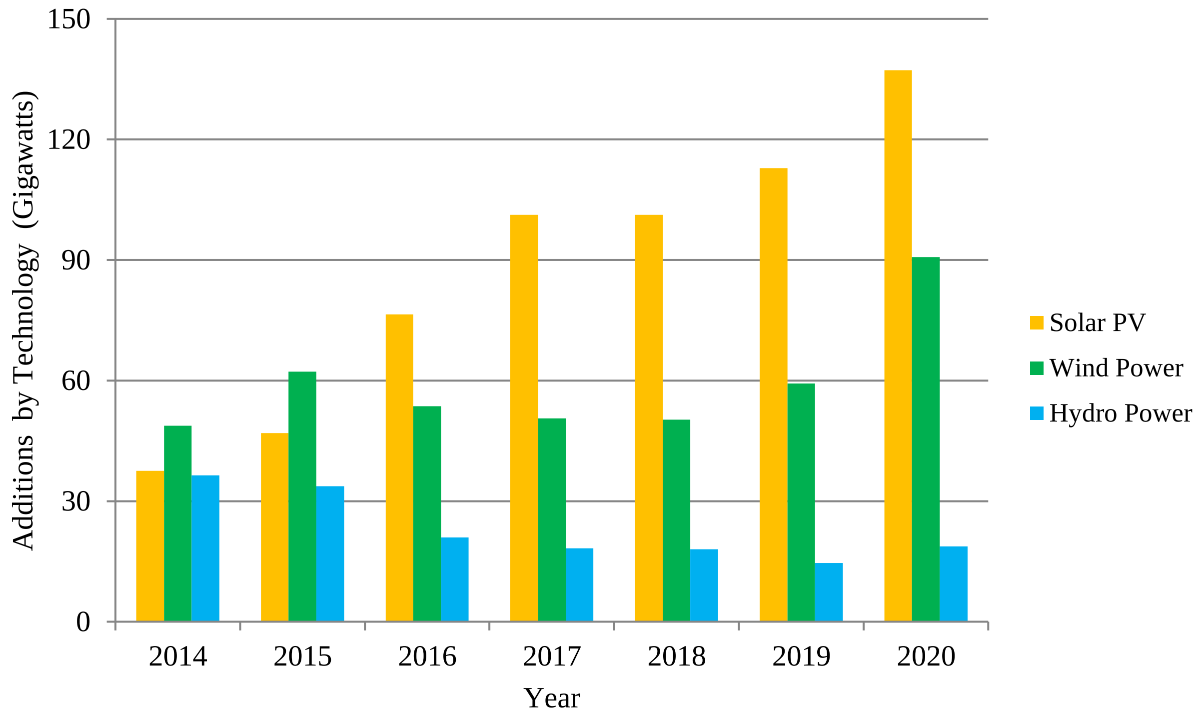


Figure 1-8: Annual capacity additions of solar, wind and hydro, 2014 - 2020 [18]

The global installed solar PV capacity has been increased from 72 GW to 707.5 GW during the 2011 - 2020 decade, with average annual percentage growth of 29% [19]. This growing pattern will essentially be continued for the next coming decades as well. According to IRENA's (International Renewable Energy Agency) REmap analysis, the global installed capacity could climb up to 2840 GW by 2030 and up to 8519 GW by 2050. Reaching to these numbers means, solar PV contribution for the reduction of global CO_2 emission would go up to 4.9 Gt in 2050, representing 21% of energy sector related emissions reductions from which the goals set by the Paris agreement [20].

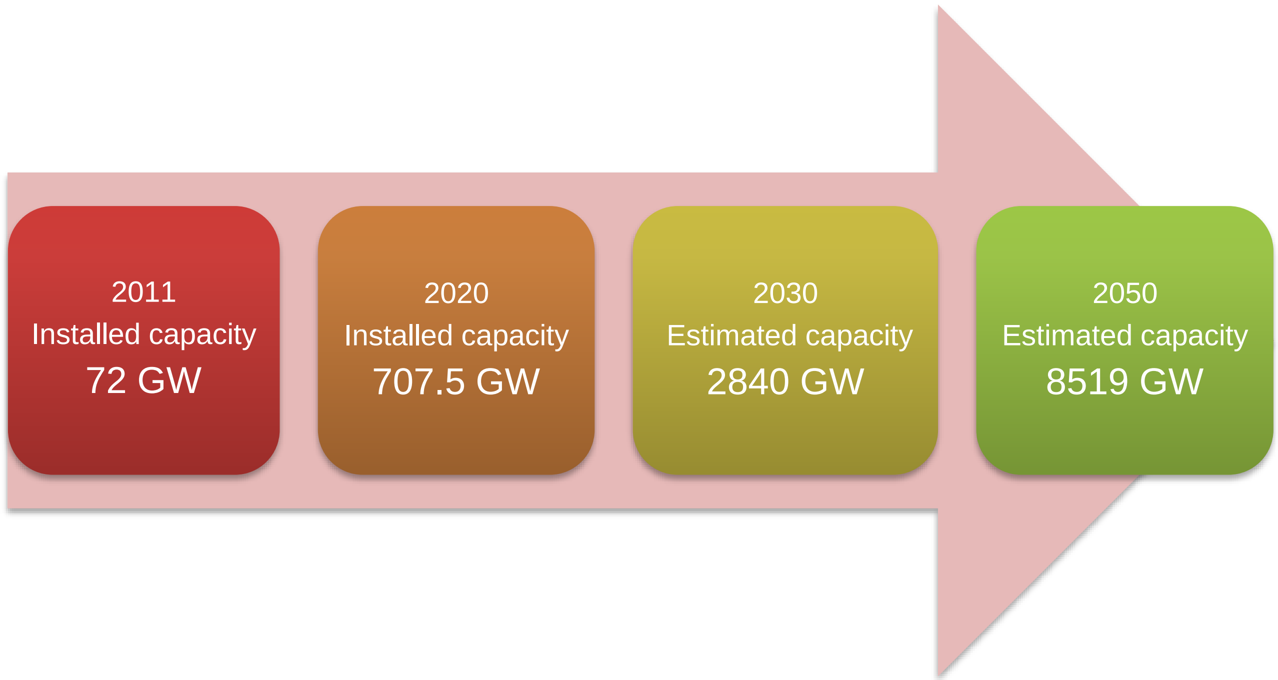


Figure 1-9: Global Solar PV capacities from past to future [20]

1.6 The Future of Utility Scale Solar PV Plants

Generally, the solar PV systems can be classified as standalone systems and grid connected systems. Standalone systems are applicable for the areas where the electricity grid is not available or not easily accessible. Due to the unavailability of solar radiation in nighttime, the stand-alone systems have to be used with an energy storage device. The grid connected systems can be further classified as distributed solar PV systems and utility scale solar PV systems. The distributed systems are small scale typically in the range of 1 kW - 1000 kW, connected to the grid in the proximity to the end user of the electricity through a low voltage transformer. Installation of distributed PV systems on the rooftops of buildings and houses became very popular in the recent decades and it will remain same for the next few years, as a result of ongoing energy policies and relief measures offered by many countries around the world and also due to the continuous cost reduction associated with the solar PV systems.

Utility scale solar PV systems are different to distributed systems in few ways. They are centralised, which means the generated power is supplied directly to the transmission or distribution network through high voltage transformers under a purchasing agreement between the power distributing authority and the owner of the PV plant. The capacity of the utility scale PV systems or plants are high, typically starting from 1 MW.

The 1 MW solar PV system constructed in 1982 at Hesperia, California, USA by Arco Solar can be considered as the world first utility scale PV plant. After the first installation, a very little growth in the utility scale PV sector can be observed for next 20 years. The large utility scale plants were not cost effective with expensive PV panels in 1980 - 2000 period. Hence, the main focus in the sector was on relatively small residential and commercial distributed PV systems and off grid applications. As the cost started to drop, the attention re-focused towards the development of utility scale PV power plants. A huge growth in terms of annual capacity additions of utility scale PV plants can be observed since 2010 as a result of cost reductions and newly introduced tariff systems in some countries. Even though the trend of utility scale plants initialized in Europe, particularly in Germany and Spain, China is the leading force of the sector [21]–[23].

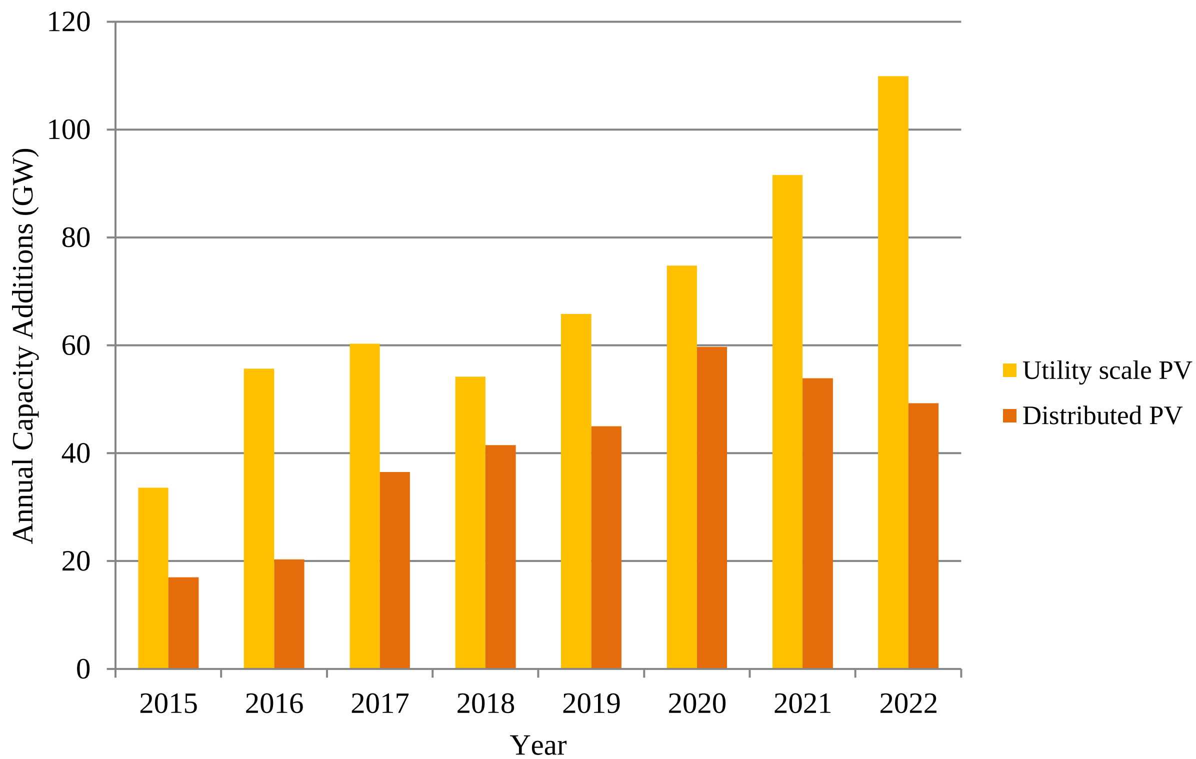


Figure 1-10: Global annual solar PV capacity additions by application segment, 2015-2022 [22]

Dominance of global annual addition of utility scale solar PV systems over distributed PV systems for the period of 2015 -2020 has been illustrated in Figure 1.10. The figure clearly shows that the capacity addition of utility scale power plant has continuously exceeds the distribution PVs for 2015-2022. The forecasted additions for the period of 2021 - 2022 based on planned projects show the same continuation. According to the report “Renewables 2020 - Analysis and Forecast to 2025” published by IEA, the generation costs of utility scale solar power are estimated to reduce up to 36% from the present values. As a result, utility scale solar PV systems would become the cheapest option for adding new electricity capacities in most countries. Furthermore, as a long-term forecast, utility scale PV systems would contribute 60% from total solar PV capacity by 2050 [20], [24].

1.7 Local Energy Scenario and Context of Solar Energy

Sri Lanka's primary energy supply in 2020 was estimated to 330, 000 TJ excluding biomass use of domestic applications such as cooking. The supply is mainly driven by oil in transportation [25].

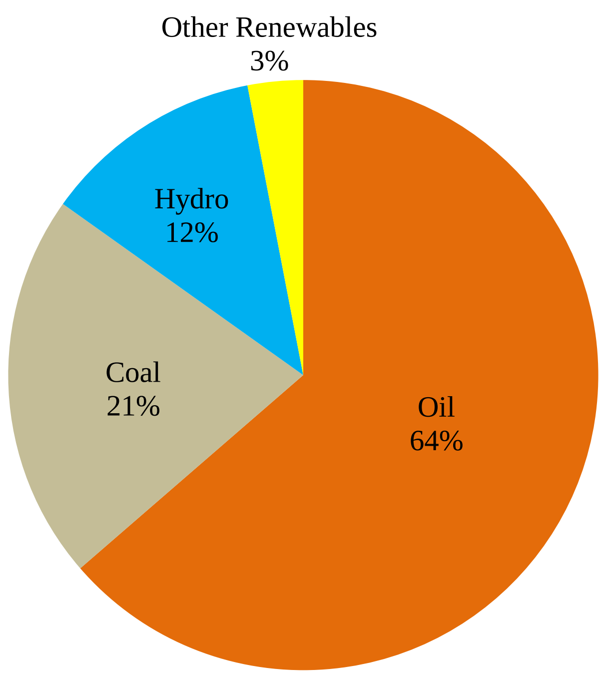


Figure 1-11: Sri Lanka's primary energy supply by source in 2020 [25]

Contribution from renewables for the total energy supply is around 50000 TJ, representing 15% from overall. The share of renewables other than hydropower has been started to increase gradually during the past few years. Specially in the sector of electricity generation, the contribution from non-conventional renewable energy sources has been noticeably improved as in the global case described before. For example, the total electricity generation in the country in 2018 was 15.3 TW h from which 2% wind (326 GW h) and 1% solar (141 GW h). In 2020, the total generation was 16.9 TW h, the contribution of wind and solar has been increased to 3.6% and 2.4%, respectively [26].

Since Sri Lanka is a country located near the equator, the solar energy potential is high almost everywhere in the island throughout the year. However, the exploitation of solar energy is very small compared to the available potential. Putting the available potential together with solar PV technology improvements, cost reduction and global environmental goals makes solar energy, one of the main contenders in Sri Lanka's future energy plans. This has been identified by the authorities since 10 - 12 year back, and different plans have been set in place. In 2010, Ministry of power and energy published a 15 points agenda planning various programmes through Ceylon Electricity Board (CEB) in order to improve the power generation sector. The programme 'Lanka Jananee' was suggested focusing on the development of new renewable energy sources such as wind power, Biomass, solar, Dendro and Wave power [27]. One year later, the first ever grid connected solar PV plant with capacity 1.24 MW was added to the national grid in 2011 [28]. In 2016, Sri Lanka launched the programme 'Soorya Bala Sangramaya' (Battle for Solar Energy) aiming to boost solar power generation. Addition of 200 MW by 2020 and 1000 MW by 2028, specially through the rooftop PV systems has been expected through the programme [29].

According to the long-term expansion plan for the period 2020 - 2039 published by CEB in 2018, it is expected to add 3425 MW power from non-conventional renewable energy sources (NCRE) to the national grid by the end of year 2039. In December 2018 when submitting the report, only 585 MW of grid connected NCRE capacity was available, with 8.7% (51.4 MW) contribution from solar. However, this has been planned to increase up to 56% (1900 MW) by 2039 which would be the biggest capacity addition among all NCRE sources.

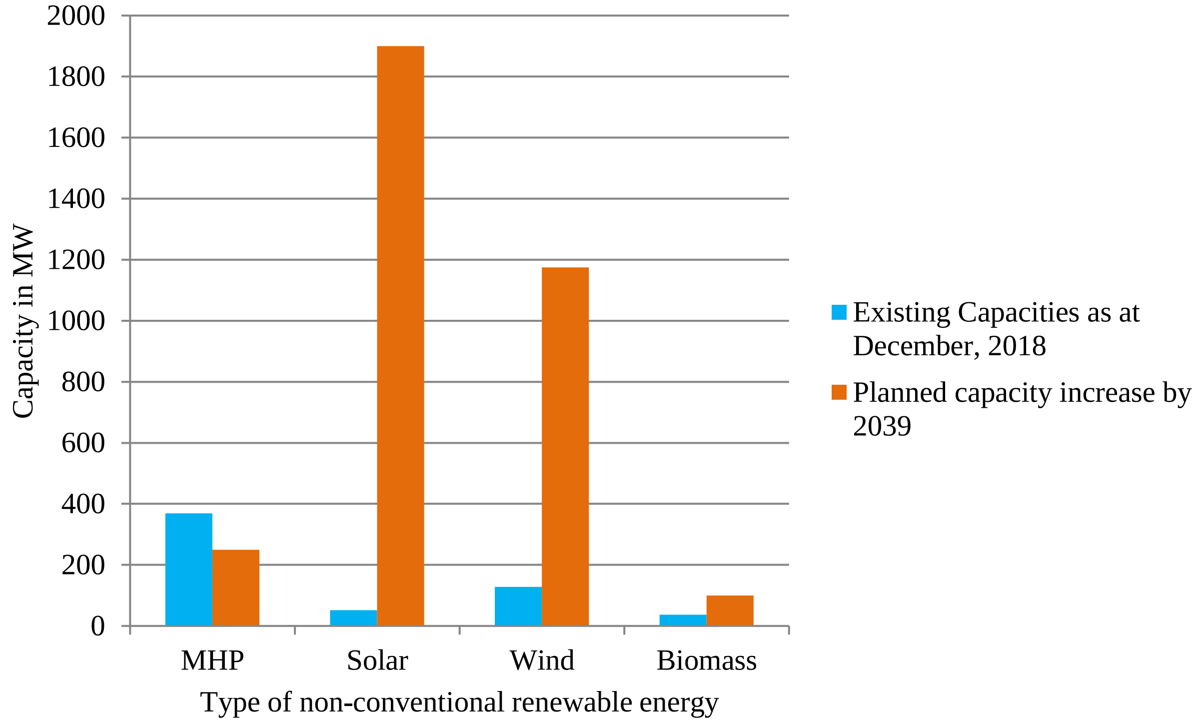


Figure 1-12: Expecting NCRE capacity increase in power generation by 2039 [30]

1.8 Challenges in Grid Operation and Power Management with High Solar PV Share

The primary energy sources for global electricity generation are still fossil fuels, accounting 63% from total in 2019. However, it is obvious that the solar PV share in electricity generation would be increased substantially in coming years according to the current global and local trends. Renewables such as solar PV and wind are called variable renewable energies (VRE) because of their intermittent nature of the availability [31]. As shown in Figure 1.13, first the intensity of incident solar radiation changes in a roughly bell-shaped pattern as the position of the sun changes throughout the daytime causing changes in solar PV output in a similar pattern. Second, weather changes affect significantly on incident solar radiation, bringing more uncertainties in to the solar PV output. Wind energy does not vary with the sun's position, but it changes randomly. In addition, both solar and wind have seasonal patterns as well. As a result, power inputs to the grid through wind and solar are not steady. On the other hand, generated power output using conventional energy sources such as fossil fuels and hydro are much stable, since reserved quantities of energy sources are used for the purpose. Handling an electrical grid which has stable power feeds is not that complicated.

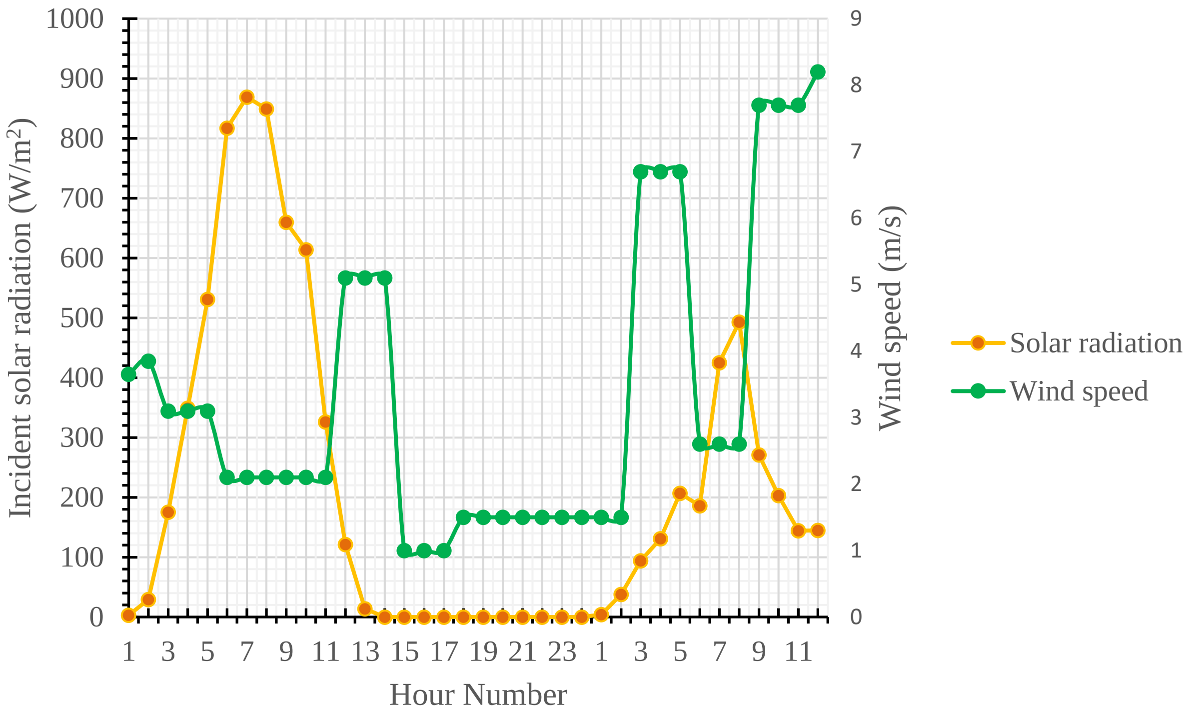


Figure 1-13: Intermittent nature of VRE energy sources [32], [33]

Modern electrical grid systems are having many advanced features. They have the capacity to supply the power 'on demand' basis for a span starting from small household application to large commercial and industrial applications. In terms of voltage tolerance limits and the frequency, the reliability and the quality of most of the systems are robust. It can be understood that the current dominance of conventional energy sources in electricity generation is one of the reasons behind a reliable and good quality electrical grid. However, uncertainty in demand side is always present in any grid and operation strategies have been developed to tackle these challenges. Hence, up to a certain level of capacity additions from VRE sources, the intermittent nature associated with them does not affect on grid operation strategies. However, with the VRE share increases, uncertainty level in supply side also increases creating more challenges in grid operation and power management. For example, power input to the grid from a large solar PV plant can drop suddenly due to an unexpected weather condition while the power demand remains unchanged, and the situation has to be dealt with. On the other hand, VRE sources are 'must dispatch when available' type of sources since there is no way to store incident solar radiation or wind energy as it is and re-use when required. With a significant share of VRE capacity in a particular grid system, generation may exceed the demand in some occasions and this also must be correctly handle to manage the grid [34].

1.9 Role of Solar PV Power Forecasting in Future Electricity Sector

The basic requirement in an electrical grid operation is maintaining a balance between supply and demand. Imbalance between the supply and demand may lead to a total crash or blackout of the grid. The equilibrium between supply and demand is monitored using alternative current frequency and, most of the countries have set the frequency value to 50 Hz with a small tolerance. An increase of demand beyond the supply results in dropping the frequency below 50 Hz, whereas a decrease of demand results in increase in the frequency above 50 Hz. In order to maintain the frequency within the tolerance level, the grid operators have to either ramp up or ramp down the power generation. Ramping capacity or possible ramping rate depends on the type of the power source connected to the grid. Power sources such as internal combustion engines (ICE) and hydro are having relatively high ramping rates among the conventional power sources. Because of this reason, they provide more flexibility to the electrical grids. On the other hand, oil and coal plants do not have good ramping capacity and their down time and up time are high [35].

As more and more solar PV plants are penetrated a grid, more flexibility and efficient ramping of power sources in the grid would be required. Accurate solar power forecasts would play a critical role to overcome this challenge, because they provide options for taking proactive actions in scheduling of power dispatch, management, and planning in different time horizons. Probability of power imbalances and risks of grid blackouts could be reduced through precise power ramping and management, improving the reliability and quality of the power supply through the grid. Also, it can contribute to reduce energy costs by optimizing the operation of conventional and reserved power sources [36], [37]. From the viewpoint of solar PV plant owners, accurate solar power forecasts allow them to go for higher bids in electricity market because the risks of dropping the reliability in the bid would reduce substantially and quality of the supply is guaranteed with accurate forecasts [38].

1.10 Importance of one day ahead hourly forecasts

Table 1.1 includes details of different flexibility measures for commonly used conventional thermal power generation technologies in electrical grids. As described in section 1.9, ICE and hydropower plants are having high ramping rates and they can dispatch the required power on demand within very short time. On the other hand, operating the other power sources in grids with high solar PV share will be a challenge without an accurate solar PV forecasting system. For example, Open Cycle Gas Turbine (OCGT) plants are having a moderate ramping capacity of 8 – 12 % from nominal load (P_{Nom}) per minute [39]. This means, for an OCGT plant operating at minimum load of 40 % P_{Nom}, it will take around 5 to 8 minutes ramp up time to dispatch full load. Therefore, operating OCGT plants with a short term (10 – 15 min) solar PV forecasting system will be much easier for grid operators.

Table 1-1: Comparison of commonly used power plants for each generation technology with regards to flexibility [35], [39], [40]

Generation Technology	Minimum load (% P_{Nom})	Average ramp rate (% P_{Nom} per min)	Hot start-up time (min) or (h)	Cold start-up time (min) or (h)
OCGT	40 - 50%	8 – 12 %	5 – 11 min	5 – 11 min
CCGT	40 - 50%	2 – 4%	60 - 90 min	3 – 4 h
Hard coal-fired power plant	25 - 40%	1.5 – 4%	2.5 – 3 h	5 – 10 h
Lignite-fired power plant	50 - 60%	1 – 2%	4 – 6 h	8 – 10 h

However, Combined Cycle Gas Turbine plants (CCGT), Hard coal-fired power plants and Lignite-fired power plants are having very low ramping rates (1 – 4 % from P_{Nom}) and they will take 15 – 60 minutes to dispatch full load from a condition where operating at minimum load [39]. Therefore, a PV power forecasting system with hourly granularity will be most suited for these generation technologies. Further, start-up time (the time required to attain stable operation when starting up from standstill) for the same technologies has a span of 1 – 10 hours [39]. Here start-up times are longer for cold start-up (plant shut for more than 48 hours) than for hot start-up (plant shut for less than 8 hours) [40]. Hence, minimum of 10 hours ahead solar PV forecasting system make it easier for decision making on plant maintenance and management, and it helps to minimize curtailment of solar power from the grid. However, considering that the non-availability of solar radiation in the nighttime, one day ahead forecasting system will be useful for the purpose.

1.11 Problem Statement

Various solar PV power forecasting models have been developed over the years through different approaches, some of them are currently in practice in different levels. For example, 11 centers equipped with solar PV forecasting tools have been established in India to provide information to grid operators [38]. Similar kind of movements can be found in Germany and California, USA [38]. Nevertheless, most of the countries including Sri Lanka are not practicing such a methodology since the percentage of solar PV penetrations into the grids is not still threatening to their power management systems. However, the requirement of Solar power forecasting would be mandatory in future with the increase of solar PV electricity generation share, as the global energy and environmental goals.

The complexity, and accuracy of existing solar PV power forecasting models are in different levels. Adopting an existing forecasting model to different solar plants or power systems might be a challenge because of various reasons such as different model input parameters, different forecasting time horizons, changes in climate conditions etc. Therefore, solar power forecasting can be recognized as a research area which needs improvements and more attention in coming years and localized studies are vital in this regard. Some countries have already invested heavily on VRE power forecasting related projects. For example, Australian Renewable Energy Agency have awarded funding of USD 5.6 million for 11 VRE power forecasting projects in 2019. Sun4cast project in USA, EWeLiNE, ORKA, ORKA2 and Gridcast projects in Germany are some other examples [38]. Countries such as Sri Lanka also need to step up and focus on the development of solar power forecasting technologies, since huge growth of solar PV share in electricity has been planned for the future. Individual researchers also can play a critical role on this, by exploring the opportunities to develop reliable solar power forecasting models using the available resources within their reach. Hence, this research has been conducted for finding a way to develop a reliable forecasting model which could be easily implemented in any solar PV plant built specially in tropical climate regions.

1.12 Aim and Objectives of the Study

Aim of the research is to formulate a one day ahead energy output forecasting mechanism for a commercial or utility scale solar PV plant using readily available data sources. To achieve the above aim of the project, following set of objectives were established.

- ❑ To identify key factors affecting the energy generation of a solar PV plant.
- ❑ To establish the suitable modeling approaches for designing of a power forecasting model for solar PV plant
- ❑ To develop a one day ahead forecasting model for PV plant and validate the model
- ❑ To evaluate the model performance for a selected location

1.12.1 Scope of the Study

Historical solar radiation and weather data for 14 years have been used for the analysis. Analysis is limited to hourly averaged values of both solar radiation and weather data. Validation of the entire forecasting model is conducted in two stages. In stage 1, the global horizontal irradiance (GHI) values predicted through an artificial neural network are validated presenting an unseen GHI data set. In stage 2, hourly solar power output values are calculated through a decomposition and transposition model and validated against the daily power plant generation records.

1.12.2 Outline of the Thesis

Chapter 1 of the thesis describes the history of the global energy and current status of the solar PV in electricity generation, both globally and locally resulting the solar PV power forecasting a mandatory requirement in future electricity generation sector. The Chapter 2 is mainly focused on the literature review where first part is focused on basic solar radiation theories and PV energy calculation method. Then it explores various available forecasting methods which have been used in different sectors, and the forecasting methods used by engineers and scientists in solar radiation or energy forecasting sector in different time horizons. Research method used in the study is described in detail in Chapter 3. In Chapter 4, the results obtained from the study has been discussed in a range of aspects. Final chapter gives a conclusion of the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Basics of Solar Radiation

This section of the chapter describes the basic solar radiation theories and concepts such as solar angles, solar constant and solar radiation models which might be used in the estimation of hourly solar radiation on a tilted surface.

2.1.1 Solar Constant and Basic Solar Angles

Intensity of solar radiation received on the earth, outside the atmosphere depends on geometrical parameters such as diameter of sun, diameter of earth and sun-earth distance. The earth's orbit around the sun is elliptical and sun-earth distance varies slightly through one revolution. The rate of radiation energy received from the sun on a unit area of surface perpendicular to the direction of propagation of the radiation at mean earth-sun distance outside the atmosphere is defined as solar constant G_{sc}.

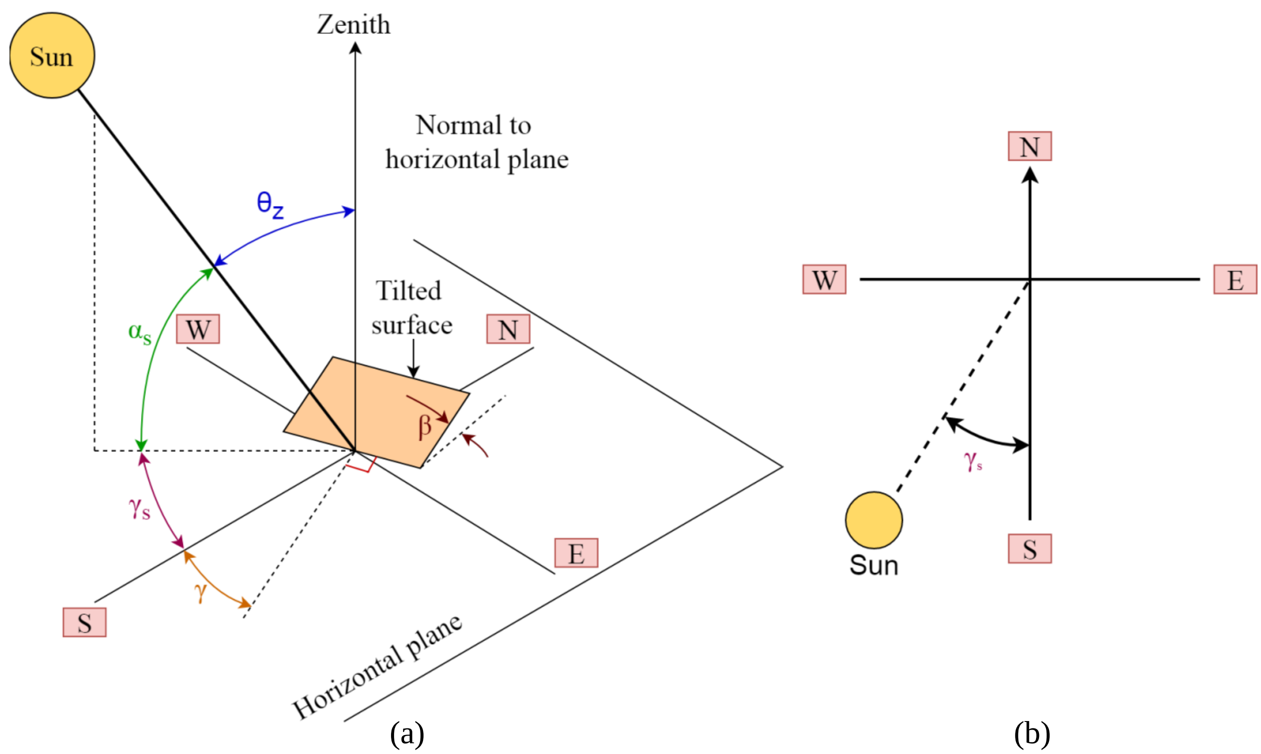


Figure 2-1: Zenith angle, slope, surface azimuth angle, and solar azimuth angle for a tilted surface (a). Plan view showing solar azimuth angle (b) [41]

Different values ranging from 1322 W/m^2 to 1395 W/m^2 have been estimated for G_{sc} by various scientists over the years [54]. A value of 1367 W/m^2 which is the recommended value by The World Radiation Center (WRC) is adopted for the calculations in this study [41].

As depicted in Figure 2.1, incoming solar radiation on to a surface of any orientation relative to the earth depends on various factors, including angles described below. Applicable sign convention is also given in the description.

1. Latitude angle (ϕ)

   The angular location north or south of the equator, where angles in north are positive; $-90^0 \leq \phi \leq 90^0$.

2. Declination angle (δ)

   The angular position of the sun, when the sun is on the local meridian with respect to the plane of the equator, angles in the north are positive; $-23.45^0 \leq \delta \leq 23.45^0$. Declination angle be calculated approximately using the equation 2.1 (Cooper's equation).

$$\delta = 23.45 \sin\left[\frac{2\pi(284+n)}{365}\right] \qquad (2.1)$$

where n is the day of the year starting from the first of January (i.e., n = 1 for January 1, n = 32 for February 1, etc.)

3. Solar hour angle (ω)

   The solar hour angle is the displacement of the sun due to the rotation of earth at the rate of 15^0 per hour, measured with respect to the position of sun at solar noon (when the sun crosses local meridian). Solar angle has negative values for the morning and positive values for afternoon.

4. Zenith angle (θ_z)

   Angle between vertical and sun's rays.

5. Solar altitude angle (α_s)

   Angle between horizontal and sun's rays. α_s is the complementary angle of θ_z

6. Solar azimuth angle (γ_s)

   Displacement of the projection of line to the sun on the horizontal plane measured from south. Displacements from south to east are negative and south to west are positive.

7. Surface azimuth angle (γ)

   Displacement of the projection of the normal to a tilted surface on the horizontal plane measured from south. Displacements from south to east are negative and south to west are positive. $-180^0 \leq \gamma \leq 180^0$

8. Slope (β)

   β is the angle between the tilted surface and the horizontal plane. $\beta \leq 180^0$ ($\beta > 90^0$ means, the tilted surface has a downward facing component)

2.1.2 Estimating Hourly Solar Radiation on a Tilted Surface

Extraterrestrial radiation on a horizontal surface for an hour period (I_0) can be given as;

$$I_0 = \frac{12 \times 3600}{\pi} G_{sc} \left[1 + 0.033 \cos\left(\frac{360n}{365}\right)\right]$$

$$\times \left[\cos\phi \cos\delta \left(\sin\omega_2 - \sin\omega_1\right)\right] + \pi \left(\frac{\omega_2 - \omega_1}{180}\right) \sin\phi \sin\delta \qquad (2.2)$$

where $\omega_1$and ω_2 are the hour angles at the beginning and ending of the hour.

However, equation 2.2 cannot be applied to estimate the solar radiation on a horizontal surface on the earth, because a portion of solar radiation is scattered by the earth's atmosphere before the radiation reaches the surface. The portion of solar radiation received without having been scattered is called beam radiation (I_b) or direct horizontal irradiance (EBH). As explained, the balance portion of solar radiation is reached on a horizontal surface after having been scattered by the atmosphere and it is called diffuse radiation (I_d) or diffuse horizontal irradiance (DHI). The total radiation on a horizontal surface (I) which is also called global horizontal irradiance (GHI), is the sum of beam and diffuse components of radiation.

2.1.2.1 Isotropic Transposition Model

Isotropic sky and anisotropic sky models have been developed to estimate the total solar radiation on a tilted surface (I_T) or global tilted irradiance (GTI). These models can be commonly called transposition models as they estimate I_T from I [41].

In isotropic sky model, the sum of diffuse radiation from the sky and ground reflected radiation on a tilted surface is assumed to be same regardless of the orientation of the surface. This concept was suggested by Hottel and Woertz (1942), and subsequently an improved model was derived by Liu and Jordan (1963) [41]. The model includes three segments to estimate beam, sky diffuse, and ground reflected components as shown in the equation 2.3 [41].

$$I_T = I_b R_b + I_d \left(\frac{1 + \cos\beta}{2}\right) + I\rho_g \left(\frac{1 - \cos\beta}{2}\right) \tag{2.3}$$

where $\rho_g$is ground reflectance or albedo and R_b is the ratio of beam radiation on a tilted plane to that on a horizontal surface. R_b can be calculated using the equation 2.4 for a south facing tilted surface (γ = 0) in northern hemisphere [41].

$$R_b = \frac{\cos(\phi - \beta)\cos\delta\cos\omega + \sin(\phi - \beta)\sin\delta}{\cos\phi\cos\delta\cos\omega + \sin\phi\sin\delta} \tag{2.4}$$

where, ω is the hour angle at the midpoint of the hour for which the R_b is to be calculated. If the sunrise or sunset occurs somewhere in the middle of the hour, calculated value of R_b using the instantaneous equation 2.4 can be very large. The best approach for such situations is to switch from the instantaneous equation to one integrated over a time period from ω_1 to ω_2 excluding sunset or sunrise. An average value of R_b ($R_{b,ave}$) can be calculates as given in the equation 2.5.

$$R_{b,ave} = \frac{a}{b} \tag{2.5}$$

where, parameters a, b are calculated using below equations.

$$a = (\sin\delta\sin\phi\cos\beta - \sin\delta\cos\phi\sin\beta\cos\gamma) \times \frac{1}{180}(\omega_2 - \omega_1)\pi$$

$$+(\cos\delta\cos\phi\cos\beta + \cos\delta\sin\phi\sin\beta\cos\gamma) \times (\sin\omega_2 - \sin\omega_1)$$

$$-(\cos\delta\sin\beta\sin\gamma) \times (\cos\omega_2 - \cos\omega_1)$$

and

$$b = (\cos\phi\cos\delta) \times (\sin\omega_2 - \sin\omega_1) + (\sin\phi\sin\delta) \times \frac{1}{180}(\omega_2 - \omega_1)\pi$$

To avoid the effect of sunset or sunrise, sunset hour angle (ω_s) needs to be determined using the equation 2.6, before calculating hourly R_b for a particular day. The sunrise hour angle is the negative value of sunset hour angle.

$$\cos\omega_s = -\frac{\sin\phi\sin\delta}{\cos\phi\cos\phi} = -\tan\phi\tan\delta \qquad (2.6)$$

As discussed above, total radiation on a horizontal surface (I) is the sum of the beam horizontal radiation (I_b) and the diffuse horizontal radiation (I_d). If I, I_b and I_d are known, the total radiation on a tilted surface (I_T) can be calculated using the Liu and Jordan isotropic sky model given by equation 2.3. However, a decomposition model such as Erbs correlation (given in the equation 2.7) should be used for estimating I_d.

$$\frac{I_d}{I} = \begin{cases} 1.0 - 0.09k_T & ; for \;\; k_T \leq 0.22 \\ 0.9511 - 0.1604k_T + 4.388k_T^2 - 16.638k_T^3 + 12.336k_T^4 & ; for \; 0.22 < k_T \leq 0.80 \\ 0.165 & ; for \; k_T > 0. \end{cases} \qquad (2.7)$$

Here k_T is called the hourly cleanness index and it can be calculated using the equation 2.8.

$$k_T = \frac{I}{I_0} \qquad (2.8)$$

2.1.2.2 Anisotropic Transposition Model

Anisotropic sky model is an improved version of isotropic model, where circumsolar diffuse and horizon brightening components have also been taken into account. This has been illustrated in the Figure 2.2. First improvement was done by Hay and Davies (1980) by introducing a modification for the circumsolar diffuse radiation assuming the direction as same as the beam radiation. Reindl et al. (1990) further modified the model by including a correction for the diffuse from horizon which was first suggested by Klucher (1979) [41].

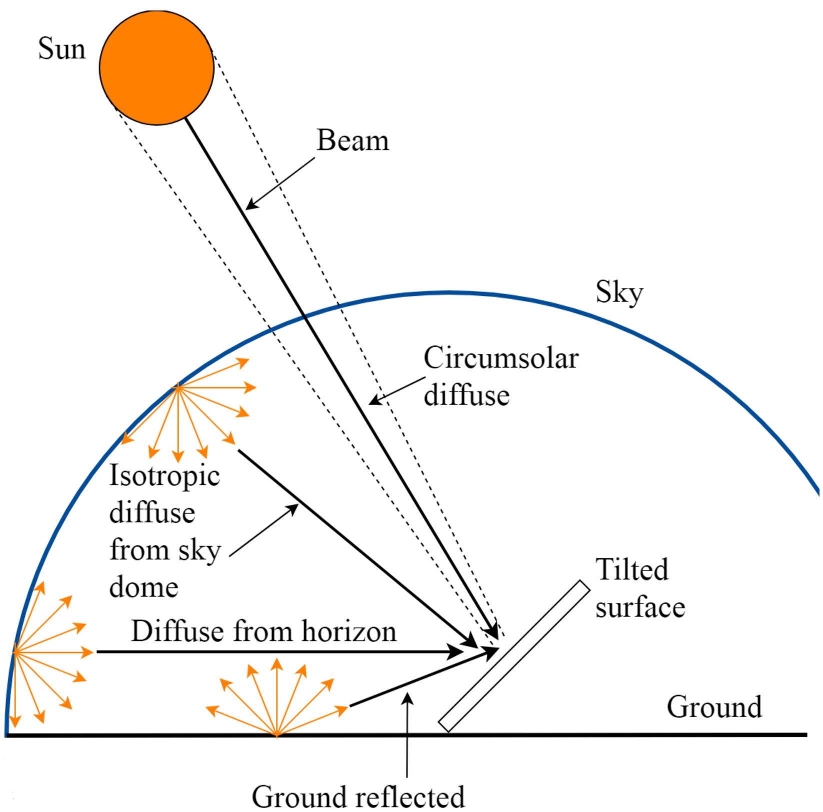


Figure 2-2: Components of solar radiation accounted for anisotropic sky model [41]

This improved anisotropic model is called HDKR (Hay-Davies-Klucher-Reindl) model, and it is given by equation 2.9.

$$I_T = (I_b + I_d A_i)R_b + I_d(1 - A_i)\left(\frac{1 + \cos\beta}{2}\right)\left[1 + f\sin^3\left(\frac{\beta}{2}\right)\right] + I\rho_g\left(\frac{1 - \cos\beta}{2}\right) \qquad (2.9)$$

where anisotropic index A_i and modulating factor f can be estimated from the equations 2.10 and 2.11, respectively.

$$A_i = \frac{I_b}{I_0} \quad (2.10)$$

$$f = \sqrt{\frac{I_b}{I}} \quad (2.11)$$

2.2 Estimating Energy Generation of PV Systems

The energy generation of a PV system in kWh per hour (E_{PV}) can be estimated using the total radiation on the PV panels (I_T) as given in the equation 2.12 [42].

$$E_{PV} = 10^{-3} A_{PV} I_T \eta_{con} \quad (2.12)$$

where, A_{PV} is the area of the PV panels or arrays (m^2) and η_{con} is the conversion efficiency of the system. η_{con} depends on many factors associated with entire solar PV system. If an individual PV module is considered, the conversion efficiency changes with the change of operating conditions of the module, especially with the change of PV cell temperature. Equation 2.13 includes a temperature adjustment to estimate conversion efficiency (η_{con}) of a PV system, accounting for this operating condition change to a certain extent.

$$\eta_{con} = \eta_{STC}[1 - \beta_T(T_C - T_{STC})]\eta_{sys} \quad (2.13)$$

where, η_{STC}, β_T, T_C and T_{STC} are module efficiency at Standard Test Conditions (STC), temperature coefficient at the maximum power of the module ($^0C^{-1}$), operating temperature of a PV cell and standard test condition temperature of a PV cell, respectively. η_{STC} and β_T should be provided by the manufacturer of the solar PV modules. Standard Test Condition (STC) and Nominal Operating Cell Temperature conditions (NOCT) are two of the test conditions recognized by IEC (International Electrotechnical Commission) for standardization of solar PV modules of different manufacturers [43]. The accepted values of the parameters for these two test conditions are given in the Table 2.1

Table 2-1: STC and NOCT test conditions for a solar PV cell [43]

Test Condition	**Incident solar irradiance SI (W/m^2)**	**PV cell temperature (^{0}C)**	**Ambient temperature (^{0}C)**	**Wind speed (m/s)**	**Air mass**
STC	1000	25	N/A	N/A	1.5
NOCT	800	N/A	20	1	1.5

According to the Table 2.1, T_{STC} is 25 ^{0}C. T_C can be calculated using the equation 2.14 below.

$$T_C = T_a + \frac{SI}{SI_{NOCT}}\left(T_{NOCT} - T_{a,NOCT}\right) \qquad (2.14)$$

where T_a is the ambient temperature, SI is the incident solar irradiance on the PV plane in W/m^2 and SI_{NOCT} is the incident solar irradiance on the PV plane at nominal operating cell temperature which is 800 W/m^2 as given in the Table 2.1. Nominal operating cell temperature (T_{NOCT}) should be provided by the manufacturer and it is usually around (43 ± 3) ^{0}C. $T_{a,NOCT}$ is the ambient temperature for the nominal operating cell temperature which is 20 ^{0}C as given in Table 2.1. η_{sys} in the equation 2.13 represents the efficiency of an entire solar PV system or plant considering all possible aspects of losses associated, other than the module efficiency η_{STC}. Typical range of some of the losses have been listed in Table 2.2.

Here, array over sizing loss (inverter clipping loss) occurs when a solar array has a higher peak capacity than the inverter capacity. This is mostly common in present PV plants because designers tend to suggest attaching more modules to a single array than the inverter rating since the PV modules produce their rated power very rarely. Array over sizing losses are depending on DC/AC ratio which is the ratio between the installed DC capacity and inverter AC power rating of a solar system. For example, DC/AC ratio of 1.5 results a clipping loss of 2% - 5% [44].

Table 2-2: Common types of losses represented in η_{sys}

	Types of losses	**Range of percentage loss**	**Reference**
1.	Inverter loss	2.0 - 7.0 %	[45], [46]
2.	Reflection loss	2.0 - 4.0 %	[47], [48]
3.	Soiling loss	1.5 - 35 %	[49], [50]
4.	Array mismatch loss	1.0 - 7.0 %	[46], [51]
5.	Loss due to low irradiance level	1.0 - 4.0 %	[52], [53]
6.	Loss in the Transformer	1.0 - 1.5 %	[44]
7.	Loss in DC cables	0.5 - 1.5 %	[44], [54]
8.	Loss in AC cables & Transmission	0.2 - 0.5 %	[44], [46]
9.	Auxiliary power loss	0.1 - 1.0 %	[44]
10.	Array over sizing loss	0.0 - 10 %	[46], [55]
11.	Spectral loss	0.0 - 3.0 %	[46], [56]

Mismatch losses can occur due to various reasons such as manufacturer tolerance, light - induced power degradation, uneven surface soiling, discoloration, cracking etc. If the properties of two interconnected PV modules in series are not identical, because of those reasons, mismatch losses occur [51]. Also, mismatch losses can occur when adjacent modules are experiencing different conditions such as one shaded and the other one exposed to solar radiation directly. Furthermore, the conversion efficiency decreases for lower irradiance levels due to the voltage losses. Generally, the solar radiation spectrum and incident angles in normal operations are not same to that of standard test conditions. Because of this, reflection losses and spectral losses (or gains) can be expected in actual running conditions. Auxiliary power losses represent various loads which have been connected to the PV system and they absorb power from the system for running throughout the day.

Soiling loss occurs as dust and dirt deposit on the modules with the time. Loss factor depends heavily on the environmental conditions around the solar PV system, and it might increase up to a high as 35% - 40 % within several months in a highly urbanized area as shown in the Figure 2.3 [57].

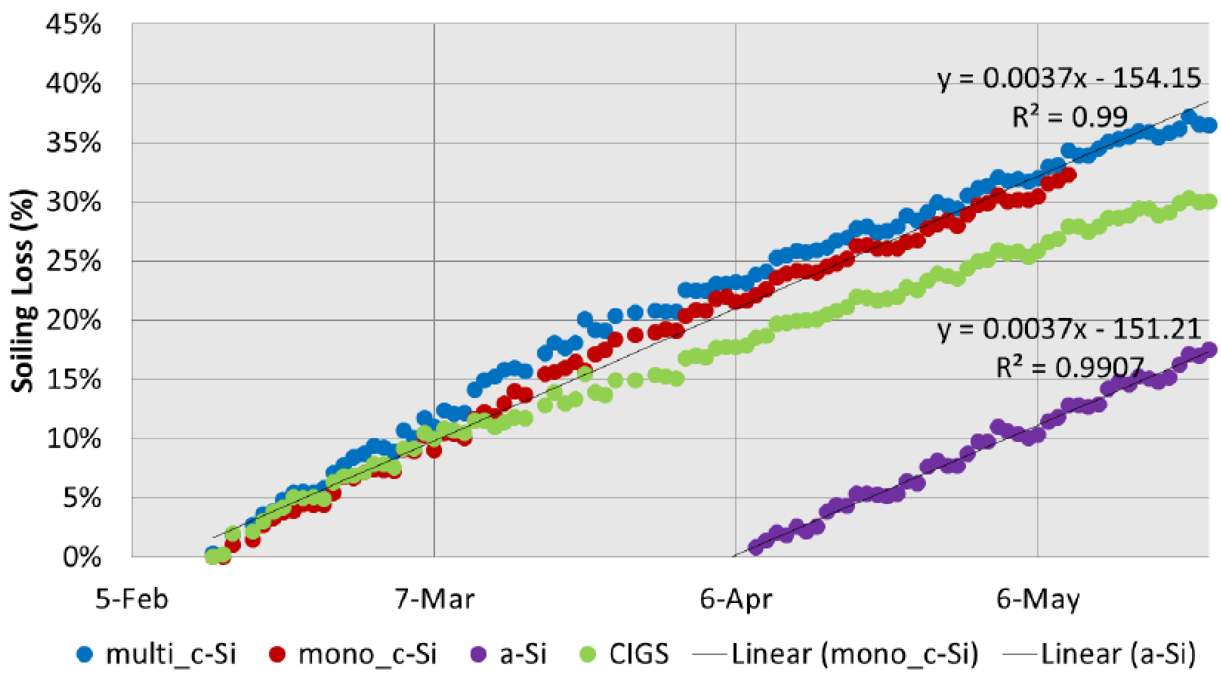


Figure 2-3: Variation of the soiling loss with the time in a solar system installed in IIT Bombay, Mumbai [57]

Figure 2.3 also shows that the soiling loss for different types of PV modules such as multi-crystalline silicon (multi_c_Si), mono-crystalline silicon (mono_c_Si), amorphous silicon (a-Si) and Copper indium gallium selenide (CIGS) thin film are slightly different to each other. Soiling effect can be reduced by regular cleaning of modules. Rain helps to wash out the deposits on tilted PV modules for certain extent, however deposits such as bird's dung are hardly removed without proper cleaning or washing.

None of the tabulated losses are stable and they are varying with the factors such as weather, load conditions and design parameters. Other than the losses listed in the Table 2.2, about 2.0 - 3.0 % loss in power can be observed due to the light induced degradation occurred in PV modules around the first year or so. However, this drop in efficiency increases to a range of 0.5 - 0.7 % per year afterwards [44], [58]. Various kinds of shading losses also causing significant generation losses in a PV plant, however this needs to be addressed separately since they are more site specific. For an example, accurate results for shading losses could be obtained through a good 3-D simulation of the PV system. If the manufacturer tolerance for PV modules have been given as a percentage range around zero (for example ±3%), this will also be added a loss into the system and the generation becomes less. This is generally called 'Module quality loss' and it can also be added to the list of loss factors [44], [59].

2.3 Classification of Forecasting Methods

Forecasting or prediction techniques can be primarily classified as qualitative and quantitative methods as given in Figure 2.4. Qualitative methods would be implemented for situations which have a little data or no data available. In general, qualitative methods are less structured compared to the quantitative methods. For example, experts can make predictions for future operations of a company using existing knowledge of past operations. These predictions fall under the category of qualitative methods. Sometimes qualitative predictions are more beneficial because they give a flexibility to use information other than numerical data to predict future trends. However, the qualitative forecasting can be subjective and stubborn.

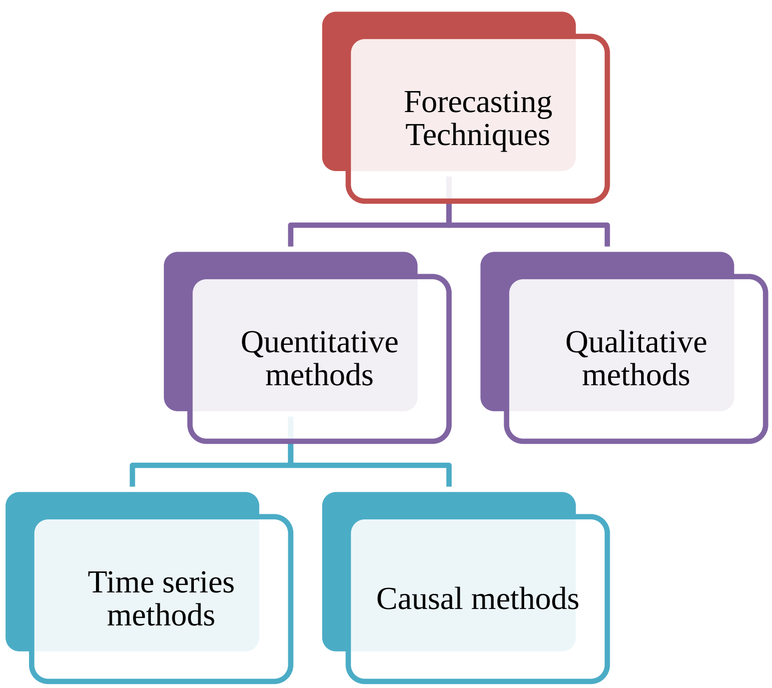


Figure 2-4: Classification of forecasting methods [60]

On the other hand, quantitative methods rely on numerical data and calculations to produce the forecasts. Therefore, more historical & associative data should be available to achieve high accuracy in forecasts. The quantitative forecasting techniques can be further categorized as Time Series Methods and Causal Methods. In time series methods, the future patterns are predicted based on the patterns observed in the historical data. In causal methods, all possible factors which can impact on the dependent variable are taken into consideration for the forecasts [60]. Some of the Time Series and Causal Methods are explained below.

2.3.1 Time Series Methods

Moving Average, Exponential Smoothening, Decomposition and Box- Jenkins forecasting can be considered as some of the most commonly used time series forecasting methods [61].

2.3.1.1 Moving Average Forecasting

Generally moving average forecasting is a good choice for time series which are fairly steady over the time in the absence of seasonal variations. Let the forecasting is to be produced based on the following time series.

$$Y_t, Y_{t-1}, Y_{t-2}, \dots \dots \dots, Y_{t-n+1}, Y_{t-n}$$

where Y_t is the observation at time t, Y_{t-1} is the observation at time t-1 etc., n is the number time steps to be averaged.

Suppose the forecast value at time t+1 is $\hat{Y}_{t+1}$; this value is obtained by calculating the moving average at time t (M_t),

$$\hat{Y}_{t+1} = M_t = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-n+1}}{n} \tag{2.15}$$

Similarly;

$$\hat{Y}_t = M_{t-1} = \frac{Y_{t-1} + Y_{t-2} + \cdots + Y_{t-n}}{n} \tag{2.16}$$

Also, $\hat{Y}_{t+1}$ can be calculated using;

$$\hat{Y}_{t+1} = \hat{Y}_t + \frac{Y_t - Y_{t-n}}{n} \tag{2.17}$$

The forecasting error can be calculated as,

$$e_t = Y_t - \hat{Y}_t \tag{2.18}$$

$$e_{t+1} = Y_{t+1} - \hat{Y}_{t+1} \tag{2.19}$$

The moving average forecasting fails to capture maximum and minimum values in the series. It gives over-predictions when the series is having continuously downward trend and under-predictions when the series is having continuously upward trend. Forecasting error can be reduced introducing the weights to the moving average calculation.

$$\hat{Y}_t = w_1 Y_{t-1} + w_2 Y_{t-2} + \cdots + w_i Y_{t-i} \ldots + w_n Y_{t-n} \tag{2.20}$$

where w_i is the weight to be given to the actual value in the time series at the time t – i, where w_i satisfy,

$$\sum_{i=1}^{n} w_i = 1 \tag{2.21}$$

The weights are basically chosen by trial-and-error methods and using the experience. However, a higher weighing can be set for the most recent actual value in the time series because it is the best indicator of the future trend of the series.

2.3.1.2 Exponential Smoothening Forecasting

Exponential smoothening forecasting models have been developed assuming the significance of most recent data is high, and it diminishes with the time goes on. Therefore, a high weight is given for most recent data, and it is decreased by (1-α) for data belongs to every previous time step. Here α is called the smoothening constant or desired response rate ($0 < \alpha < 1$).

The forecast value for period t+1 made at time t by exponential smoothening method can be given by;

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \tag{2.22}$$

where Y_t is actual value at time t; $\hat{Y}_t$ is the forecast value for period t made at time t -1

By substituting the previous forecasting values back to the starting point, the data series can be expanded as follows.

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1} Y_1 + (1-\alpha)^t Y_0 \tag{2.23}$$

The equation can be written in a compact form as follows.

$$\hat{Y}_{t+1} = \alpha \sum_{k=0}^{t-1} (1-\alpha)^k Y_{t-k} + (1-\alpha)^t Y_0 \tag{2.24}$$

Here Y_0 is the initial value which can be taken as the average value of first few observations of the series, or the first observation itself. Best value of α can be decided by minimizing the mean square error (MSE) or root mean square error (RMSE) which is calculated based on samples.

The exponential smoothening method is preferred for the situations where few observations are available or specially for the series moving around a constant mean. If a certain series is having a trend or a seasonal pattern, some modifications for the method are required. Holt's model [61] and Winter's model [61] are examples for such efforts where Holt's model introduces a linear-trend component and Winter's model introduces a seasonality factor.

2.3.1.3 Decomposition Forecasting

In the decomposition forecasting methods, the main time series is broken down into different patterns. The forecasting for these patterns is made separately and then combined them to get the forecasting of the original series. Therefore, for decomposition forecasting any time series data can be assumed as;

$$\text{Data} = \text{Pattern(s)} + \text{error} \tag{2.25}$$

The above approach can be mathematically represented as;

$$Y_t = f(S_t, T_t, E_t) \tag{2.26}$$

where;

Y_t is the time series value (actual value) at time t, S_t is the seasonal component (or index) at time t, T_t is the trend component at time t, E_t is the error or random component at time t.

The exact mathematical form of the Equation 2.26 depends on the decomposition model use. Generally, the models are found in additive form or multiplicative form as below.

$$Y_t = S_t + T_t + E_t \text{ (Additive form)} \tag{2.27}$$

$$Y_t = S_t \times T_t \times E_t \ \ (Multiplicative) \tag{2.28}$$

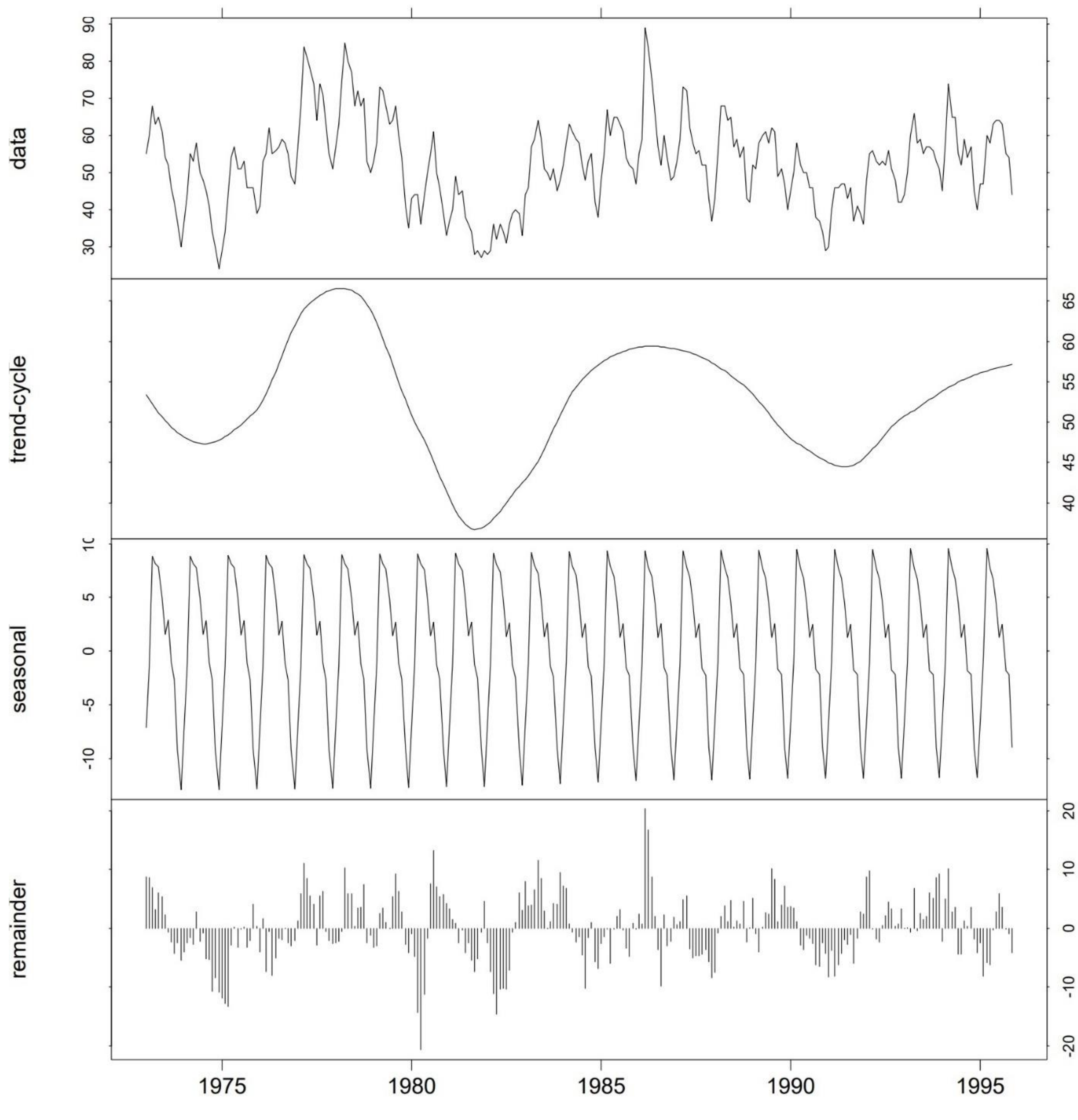


Figure 2-5: Different components of a time series after decomposition [61]

Additive models have been recommended for time series where the fluctuations of seasonal patterns are not significant. Otherwise, the multiplicative decomposition models are preferred [61].

Various decomposition models based on additive and/or multiplicative form(s) have been developed over the years. STL decomposition model (Seasonal-Trend decomposition procedure based on Loess) proposed by Cleveland et al (1990), X-12-ARIMA model (Autoregressive integrated moving average) developed by both the U.S. Census Bureau and Statistics Canada (1997), are examples for some of the most sophisticated decomposition models which were developed in recent history. X-13ARIMA - SEATS, a successor to X-12-ARIMA is the current version available from U.S. Census Bureau [61], [62].

Even though well-developed decomposition models are available, projecting individual decomposed components into the future and recombining to get the forecast of the time series is always challenging. However, decomposition gives a good graphical outlook for the behavior of a time series, which helps to recognize the structure and causes of variations, leading to a better understanding of the problem. However, obtaining adequate forecasts of the components is difficult.

2.3.1.4 Box - Jenkins Forecasting

Box - Jenkins method tries to identify best fitted ARIMA model for the available data set and uses the fitted model for the forecasting. The method consists of well-ordered five stage process as shown in the Figure 2.6, starting from data preparation and then model selection, parameter estimation, model checking and finally the forecasting stage [61].

Data preparation includes the transformation and/or differencing of original data. Transformation is generally required for time series where the variance is highly fluctuating along the series. The variance can be stabilized in such series by transformation methods such as taking square roots or converting to logarithm scale. Differencing is required to remove the visible patterns such as seasonal variations in the original time series and to make the series stationary. Differencing is carried out by taking the difference between two adjacent data values or between observations a month or a year apart. In general, differenced stationary data series can be effectively modeled than the original data series [61].

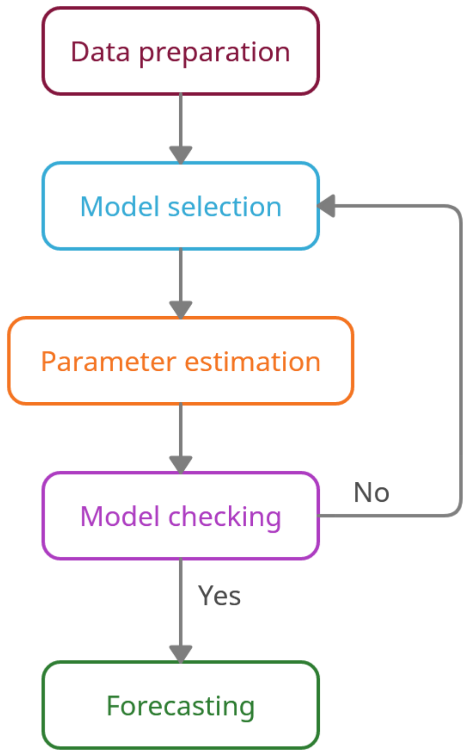


Figure 2-6 : Schematic representation of the Box-Jenkins methodology for time series modeling [61]

The best fitted ARIMA model is selected by evaluating various graphs which have been plotted using the transformed and/or differenced data series. ‘White noise’ and ‘random walk’ are examples for some of the basic ARIMA models to select among various available models [61]. A model selection tool such as Akaike’s Information Criterion could also be used to select the model which best suit for the task. However, a considerable experience is required for interpreting data samples and to select the best suited model [63]. Under parameter estimation, the best fit values for the model coefficients are calculated. After this step the model should be tested for any area of inadequacy. If an inadequacy is found, it is required to go back to model selection step again and select a suitable model. After the successful completion of these steps, the model can be considered ready for forecasting [61].

2.3.2 Causal Methods

Causal methods use a set of dependent variables possibly also including time series components to establish a statistical relationship with the forecasted variable [60]. Two causal methods, regression analysis and artificial neural networks are discussed below.

2.3.2.1 Regression Analysis Method for Forecasting

As classified in the Figure 2.4, regression analysis is a causal forecasting method which is developed by considering all possible factors affecting to the dependent variable. The linear regression model which is in practice could be written as follows [61].

$$Y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \cdots + b_kX_{k,i} + e_i \tag{2.29}$$

where Y_i , $X_{1,i}$, $X_{2,i}$,.... $X_{k,i}$ represent the i^{th} observation of each of the variables Y, X_1, X_2,.... X_k, respectively. b_0, b_1, b_2,.... b_k are fixed but unknown parameters which are called regression coefficients, and e_i is the estimated error or residual (forecast error) for the i^{th} observation which is a random variable.

The forecasted dependent variable $\hat{Y}_i$ for the i^{th} observation of each of the variables can be given as;

$$\hat{Y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \cdots + b_kX_{k,i} \tag{2.30}$$

So, the forecast error could be calculated using the equation 2.18 discussed in section 2.3.1.1. In order to determine the regression coefficients b_0, b_1, b_2,.... b_k, sum of the square error terms (S) is minimized [61].

$$S = \sum_{i=1}^{n} e_i^2 \tag{2.31}$$

$$SS = \sum_{i=1}^{n} \left(Y_i - b_0 - b_1X_{1,i} - b_2X_{2,i} - \cdots - b_kX_{k,i}\right)^2 \tag{2.32}$$

For the minimum square error, partial derivatives of S with respect to each regression coefficient should be equal to zero.

$$\frac{\partial S}{\partial b_0} = 0; \frac{\partial S}{\partial b_1} = 0; \frac{\partial S}{\partial b_2} = 0 \ldots\ldots \frac{\partial S}{\partial b_k} = 0 \qquad (2.33)$$

Developing a regression model for forecasting with a good accuracy is always challenging. Preparation of a long list of variables which are potentially having an impact on the dependent variable is a good point to start with. Long list is to be reduced to a short list by analyzing how significant each explanatory variable X_i for deciding the value of dependent variable Y_i. The methods such as plotting Y_i against each explanatory variable X_i, observing the inter correlation among the explanatory variables etc. have been used in identifying the sensitivity to dependent parameters. Once the short list is finalized, the regression model can be developed and fitted to the existing data set.

2.3.2.2 Forecasting Using Artificial Neural Networks

Artificial neural network (ANN) models have been developed using the concept of how human brain works. Generally neural networks require much larger data set for accurate forecasting, compared to the methods discussed above. One of the major plus points of ANN models is that they can be fitted to nonlinear and complicated forecasting problems [64].

The basic processing unit of an ANN is called a neuron or perceptron which can be considered as a small computing engine where it receives inputs, processes them, and transmits an output. The importance of receiving input signals into the neuron might not be equal from one input to the other. Therefore, to give different priorities to different input signals, they are weighted with different values which can be adjusted as required. A basic neuron model which receives four inputs x_0, x_1, x_2 and x_3 through the weights w_0, w_1, w_2 and w_3 is shown in Figure 2.7.

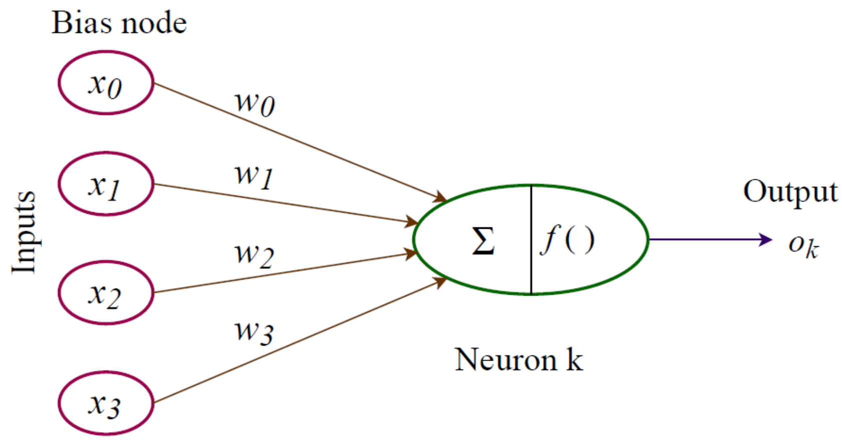


Figure 2-7: Neuron model with weighted inputs and embedded transfer function f [64]

Weighted sum of the inputs or net input (net_k) which neuron receives can be calculated as [64];

$$net_k = x_0 w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \tag{2.34}$$

where k is the number given to the neuron.

Input x_0 is considered as the bias, which is an extra input added, usually set the value $x_0 = 1$. Adding a bias allows to shift the activation function left or right providing an opportunity to better fit the data with neural network model. Therefore, equation 2.34 can be re-written as;

$$net_k = w_0 + x_1 w_1 + x_2 w_2 + x_3 w_3 \tag{2.35}$$

In general form equation 2.35 can be expresses as a summation.

$$net_k = \sum_{i=0}^{n} w_i x_i + w_0 \tag{2.36}$$

The output from the neuron k is computed through a transfer function (activation function) embedded in the unit. Most used transfer function is Sigmoid function (Logistic function) which brings a nonlinearity into the neural networks, making it capable of learning and performing more complex tasks. The output o_k through the sigmoid activation function is given as [64];

$$o_k = f(net_k) = \frac{1}{1+e^{-net_k}} \tag{2.37}$$

Combining number of neurons into layers is resulted in forming of an artificial neural network which could be used successfully for complex forecasting problems. Generally, ANNs can be classified as feed-forward neural networks (FFNN) and recurrent neural networks (RNN) depending on the nature of their operation [64]. In FFNNs, information flows only in one direction which is from input layer to output layer. The weights would not be further adjusted once the training of the network is completed. On the other hand, in RNNs, the results are obtained through the information flow from input layer to output layer and the error in results is fed back to the previous layers in order to re-adjust the weights [64].

Figure 2.8 shows a multilayer FFNN (Multilayer perceptron or MLP) which includes input layer with five inputs, two hidden layers and output layer with two outputs. Hidden layers are used in NNs to capture the complexity of the problem. However, adding more hidden layers to the NN architecture would lead to over-fitting the problem or reduction of the accuracy. Generally, a neural network with a single hidden layer can produce very good results for majority of problems. There is no exact method of deciding the number of hidden layers and number of nodes (hidden layer sizes) for a particular neural network to produce a forecast with minimum error. Usually this is done through a trial-and-error process.

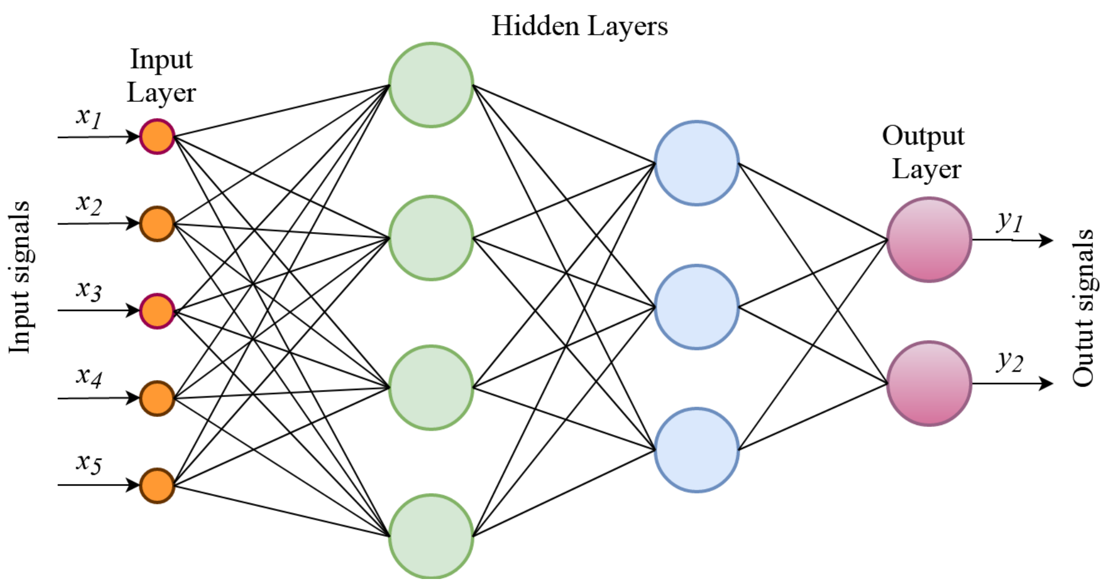


Figure 2-8: Architecture of a multilayer feed-forward neural network [64]

In the discussion from here, the following notations are used for the parameters associated with j^{th} neuron in the l^{th} layer (Figure 2.9) and for the entire ANN. The input layer of a network is considered as 0^{th} layer.

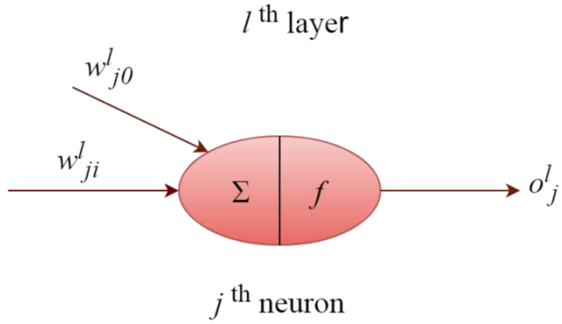


Figure 2-9: Notations for the j^{th} neuron in l^{th} layer of an ANN [64]

w_{ji}^{l} - weight linking i^{th} neuron in the l - 1^{st} layer to the j^{th} neuron in the l^{th} layer.

w_{j0}^{l} - bias term written as a weight to a unitary input linking j^{th} neuron in the l^{th} layer

o_{j}^{l} - output of j^{th} neuron in l^{th} layer

o_{pj}^{l} - output of j^{th} neuron in l^{th} layer when pattern p is presented

net_{j}^{l} - weighted sum of the inputs which j^{th} neuron in l^{th} layer receives

N^{l} - number of neurons in the l^{th} layer

L - number of layers for the network excluding the input layer

P - number of patterns in the training set

p - input pattern index (1, . . . , P)

δ_{ij}^{l} - delta term containing the error needed to update the connection linking the i^{th} neuron in the $l-1^{st}$ layer to the j^{th} neuron in the l^{th} layer

f_{j}^{l} - activation function for the j^{th} neuron in ^{lth} layer

$\dot{f}_{j}^{l}$ - derivative of the activation function

Using the above notations, equations for net_{j}^{l} and o_{j}^{l} with Sigmoid activation function can be re-written as;

$$net_{j}^{l} = \sum_{i=1}^{N^{l-1}} w_{ji}^{l}\, o_{i}^{l-1} + w_{j0}^{l} \tag{2.38}$$

$$o_j^l = f_j^l(net_j^l) = \frac{1}{1 + e^{-net_j^l}} \tag{2.39}$$

Main highlights on the forecasting methods discussed in the section 2.3 are given in the table 2.3. It is understood that the causal method based on Artificial Neural Network can handle large data sets with nonlinear and complex behavior, to produce accurate forecasting results. Therefore, an ANN forecasting method might be better suited for this study especially considering the size of the data set, number of variables and the complexity associated with the problem. The suitability of an ANN model for the study is further examined in detail in the forthcoming sections 2.4 and 2.5.

Table 2-3: Main highlights of different forecasting techniques

Category	Techniques	Highlights
Quantitative - Time series	Moving averages	• Good choice for time series which are steady over the time and in the absence of seasonal variations. • Give over-predictions when the series is having continuously downward trend and under-predictions when the series is having continuously upward trend.
	Exponential smoothing	• Preferred for situations where few observations are available or specially for the series moving around a constant mean.
	Decomposition approach	• Gives a good graphical outlook for the behavior of a time series, which helps to recognize the structure and causes of variations. However, obtaining adequate forecasts of the components is difficult.
	Box–Jenkins method	• Identifies best fitted ARIMA model for the available data set and uses the fitted model for the forecasting. • Considerable experience is required in interpreting sample correlation functions to identify best fitted model.
Quantitative - Causal	Regression analysis	• All possible factors for the dependent variable are considered. Forecasting accuracy depends largely on the accuracy of estimates for the independent variables.
	Artificial neural network	• Generally, a much larger data set is required for accurate forecasting. They can be fitted to non-linear and complicated forecasting problems.

2.4 Training of an Artificial Neural Network

Once an appropriate ANN architecture is selected for a particular forecasting problem, next challenge is to decide suitable values for the weights of the network which gives optimum results. This is achieved by initially assuming the weights and adjusting them until the ANN reaches to its optimum performance for the available data set. This process is called 'training' and the network is said to be learning through the process. Learning methods of ANNs can be categorized as supervised learning and unsupervised learning. In supervised learning, both input data and output data are provided to the model for the training process, while only the input data are presented for the model in unsupervised learning [64].

Whole learning process of a feed-forward neural network can be divided into few stages. Once input parameters are fed into the model, the information propagates to the hidden neurons of each layer and finally the model produces output. This step is called forward propagation. The output obtained through forward propagation is different to the actual or target output. This difference can be considered as the error of the model and an error function is usually defined mostly using squared error (SE), mean error (ME), mean squared error (MSE) or mean absolute error (MAE) [64].

For example, if an input pattern is presented to the network, the error function using squared error can be written as [64];

$$E_p = \frac{1}{2}\sum_{i=1}^{N^L}\left(t_{pi} - o_{pi}\right)^2 \tag{2.40}$$

where t_{pi} is the actual or target output, o_{pi} is the output obtained through the model and E_p is the squared error of the model after the first forward-propagation when pattern p is presented.

The error function for the total error for all patterns presented can be written as [64];

$$E_{total} = \sum_{p=1}^{P} E_p = \frac{1}{2}\sum_{p=1}^{P}\sum_{i=1}^{N^L}\left(t_{pi} - o_{pi}\right)^2 \tag{2.41}$$

The error function is sometimes called cost function or lost function. The cost function is a measure of how predicted values have been deviated from the actual values and the performance of the model can be optimized by minimizing it. Various optimization algorithms have been developed over the years to achieve this goal. Some of the commonly used optimization algorithms for training are shown in the Figure 2.10.

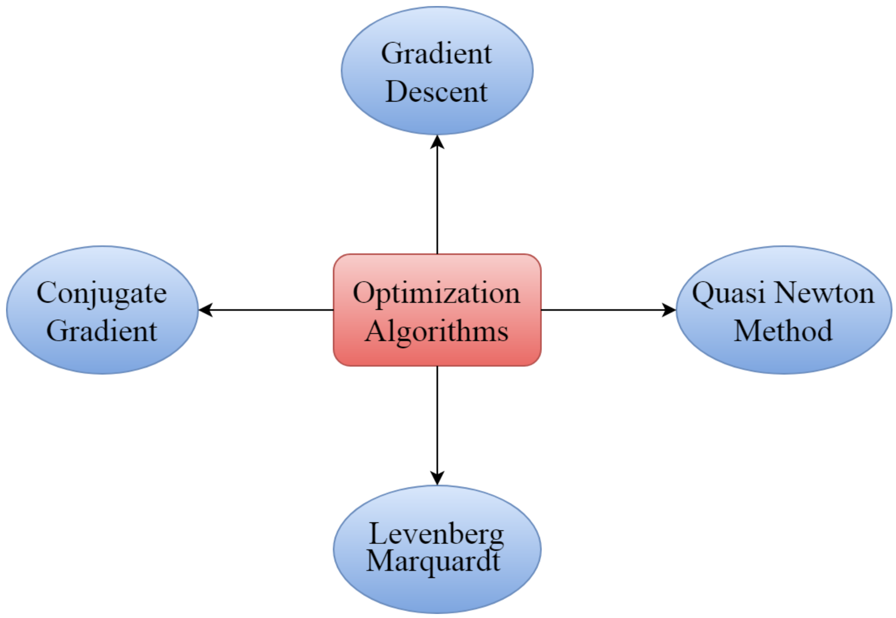


Figure 2-10: Examples of Optimization algorithms used for training of an ANN model [65]

2.4.1 Backpropagation and Gradient Descent Method

Gradient descent algorithm uses backpropagation, to compute the gradients of the cost function with respect to parameters of the NN model. The word backpropagation is used because, information from the cost function is flowed backwards through the network to compute the gradients.

To minimize the error, partial derivative of the cost function with respect to all the weights in the network are forced to zero [36].

$$\frac{\partial E_{total}}{\partial w_{ji}^{l}} = 0 \tag{2.42}$$

The partial derivative of the cost function can be re-written as a summation of partial derivatives of error terms of each presented pattern.

$$\frac{\partial E_{total}}{\partial w_{ji}^{l}} = \sum_{p=1}^{P} \frac{\partial E_p}{\partial w_{ji}^{l}} \tag{2.43}$$

Using chain rule;

$$\frac{\partial E_p}{\partial w_{ji}^{l}} = \frac{\partial E_p}{\partial net_{pj}^{l}} \frac{\partial net_{pj}^{l}}{\partial w_{ji}^{l}} \tag{2.44}$$

The second term in the right-hand side of the equation 2.44 could be written as follows.

$$\frac{\partial net_{pj}^{l}}{\partial w_{ji}^{l}} = \frac{\partial}{\partial w_{ji}^{l}} \left(\sum_{k=1}^{N^{l-1}} w_{jk}^{l}\, o_{pk}^{l-1} + w_{j0}^{l} \right) = o_{pk}^{l-1} \tag{2.45}$$

The change in error with respect to weighted sum of the inputs is defined as the delta term;

$$\delta_{pj}^{l} = -\frac{\partial E_p}{\partial net_{pj}^{l}} \tag{2.46}$$

Substituting equations 2.45 and 2.46 back into equation 2.44 yields;

$$\frac{\partial E_p}{\partial w_{ji}^{l}} = -\delta_{pj}^{l} o_{pi}^{l-1} \tag{2.47}$$

For the solving of δ_{pj}^{l}, first term of the equation 2.44 could be further broken-down using chain rule as,

$$\frac{\partial E_p}{\partial net_{pj}^{l}} = \frac{\partial E_p}{\partial o_{pj}^{l}} \frac{\partial o_{pj}^{l}}{\partial net_{pj}^{l}} \tag{2.48}$$

The Second term in the right-hand side of the equation 2.48 could be written as,

$$\frac{\partial o_{pj}^{l}}{\partial net_{pj}^{l}} = \frac{\partial}{\partial net_{pj}^{l}} f_j^{l}\left(net_{pj}^{l}\right) = \dot{f}_j^{l}\left(net_{pj}^{l}\right) \tag{2.49}$$

When the first term in the right-hand side of ‘equation 2.48 is considered for the output layer, it could be written as;

$$\frac{\partial E_p}{\partial o_{pj}^L} = \frac{\partial}{\partial o_{pj}^l}\left[\frac{1}{2}\sum_{i=1}^{N^L}\left(t_{pi} - o_{pi}\right)^2\right] = -\left(t_{pi} - o_{pi}\right) \tag{2.50}$$

Considering the hidden layers (l < L)

$$\frac{\partial E_p}{\partial o_{pj}^l} = \sum_{k=1}^{N^{l+1}} \frac{\partial E_p}{\partial net_{pk}^{l+1}} \frac{\partial net_{pk}^{l+1}}{\partial o_{pj}^l} \tag{2.51}$$

According to equation 2.46 for l < L

$$\frac{\partial E_p}{\partial net_{pk}^{l+1}} = -\delta_{pk}^{l+1} \tag{2.52}$$

Second term in the right-hand side summation of the equation 2.51 could be written as;

$$\frac{\partial net_{pk}^{l+1}}{\partial o_{pj}^l} = \frac{\partial}{\partial o_{pi}^l}\left(\sum_{i=1}^{N^l} w_{ki}^{l+1}\, o_i^l + w_{k0}^{l+1}\right) = w_{kj}^{l+1} \tag{2.53}$$

Substituting equations 2.52 and 2.53 in equation 2.51 results in,

$$\frac{\partial E_p}{\partial o_{pj}^l} = -\sum_{k=1}^{N^{l+1}} \delta_{pk}^{l+1} w_{kj}^{l+1} \tag{2.54}$$

Now combining equations 2.46, 2.49 and 2.50, the equation 2.55 is obtained for the output layer. Further, equation 2.56 is obtained for hidden layers combining equations 2.46, 2.49 and 2.54.

$$\delta_{pj}^L = \dot{f}_j^L\left(net_{pj}^L\right)\left(t_{pi} - o_{pi}\right) \tag{2.55}$$

$$\delta_{pj}^l = \dot{f}_j^l\left(net_{pj}^l\right) \sum_{k=1}^{N^{l+1}} \delta_{pk}^{l+1} w_{kj}^{l+1} \tag{2.56}$$

The backpropagation process is started by calculating the delta terms for the output layer first using equation 2.55, then go backward to calculate delta terms for the hidden layer neurons using equation 2.56 [64].

To accomplish the final goal, which is to minimize the cost function, weights should be updated. This is achieved by subtracting a fraction of first order partial derivative of the cost function with respect to the weights from the weights used for the previous step. The fraction of first order partial derivative is calculated multiplying by η which is called the learning rate [64]. Therefore, the weight update equation could be written as;

$$\Delta w_{pji}^{l} = -\eta \frac{\partial E_p}{\partial w_{pji}^{l}} \tag{2.57}$$

Substituting $\frac{\partial E_p}{\partial w_{pji}^{l}}$ from equation 2.47 results in;

$$w_{pji}^{l} = \eta \delta_{pj}^{l} o_{pi}^{l-1} \tag{2.58}$$

Given learning rate decides how effectively the optimization algorithm converges to complete the training. If the value of η is set too small, it will take long time for the learning and updating the weights. Setting a too high value for η tends to oscillate the algorithm without converging and also it will result in poor model performance after the training. Further, a momentum term is added to equation 2.58, in order to reduce the chances of converging the algorithm to a local minima or oscillating around the error surface. The momentum term α in the equation 2.59 ensures that the weight keeps changing in the same direction even when a local minima is identified. This method is called the heavy ball method and it is illustrated in Figure 2.11 [64].

$$w_{pji}^{l}(n) = \eta \delta_{pj}^{l} o_{pi}^{l-1} + \alpha \Delta w_{pji}^{l}(n-1) \tag{2.59}$$

The range of α should be from 0 to 0.9. However, the range from 0.5 to 0.9 has been frequently used in the literature. $\Delta w(n)$ in the equation 2.59 represents the new weight update, and $\Delta w(n-1)$ represents the weight update in the previous iteration.

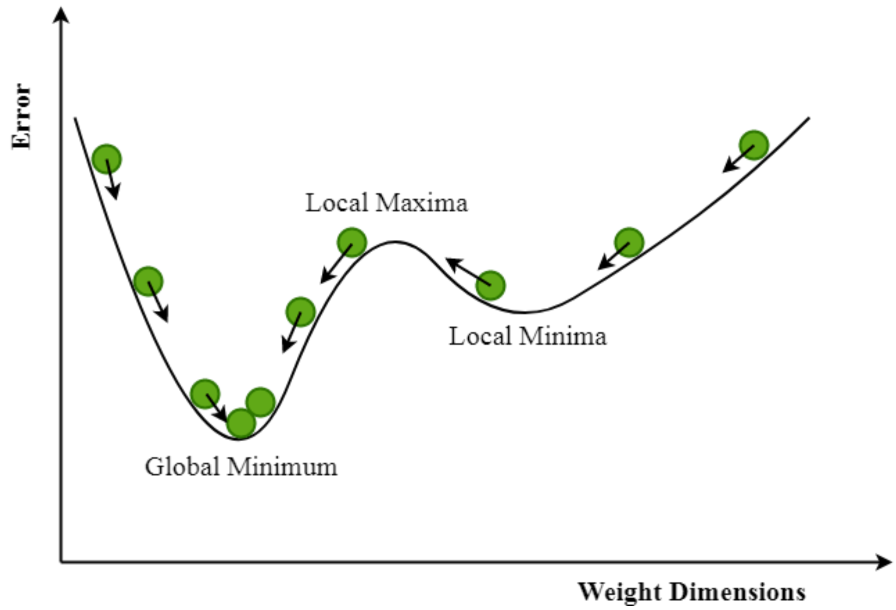


Figure 2-11: Heavy ball method - A heavy ball rolling downhill will roll up and over a small hill to reach to global minimum [64]

2.4.2 Conjugate Gradient Descent Method

Conjugate gradient descent method can be used as an optimization technique for a neural network model by adjusting the weights as used in the backpropagation. The method was developed by Magnus Hestenes and Eduard Stiefel in 1950's [64], [65] and subsequently several enhancements have been included.

The method generally provides faster convergence compared to gradient descent method, also it can work with large number of weights [65]. It searches and projects a line along the direction of steepest descent to locate the minimum and the weights are updated per every epoch. Then another point is searched in conjugate direction from the previous point by making sure that the minimization found in the previous step is not spoiled by the minimization in the next step and so on. This feature of 'not undoing the progress of previous search directions' results in a faster convergence with minimum number of steps. For quadratic error surfaces the convergence is achieved very rapidly and for non-quadratic error surfaces it takes few more additional steps [64], [65]. However, as the algorithm reaches closer to the minimum point, the error surface can be more approximated to a quadratic and the convergence is super accelerated [65].

2.4.3 Quasi - Newton Method

Feed forward neural networks and majority of configurations which work for backpropagation can be trained using Quasi-Newton optimization method. The method uses second order derivatives for the optimization process and convergence is generally faster than gradient descent method [64], [65]. Quasi - Newton is always a good option for nonlinear optimization problems. However, the requirement of memory and computational power to perform the operation of the algorithm is getting increased significantly as the number of weights in the model is high. When the error is closer to the minimum, the quasi -Newton method assumes that the error surface is quadratic, and the minimization is achieved by a single step from that point by using the second order partial derivatives of the error surface. In this process, the Hessian matrix H need to be computed to get the required second order derivative terms. Weight update is calculated multiplying the inverse of Hessian matrix by direction vector of the steepest descent g per every epoch as given in equation 2.60 [64].

$$\Delta w = -H^{-1} g \tag{2.60}$$

The computation of Hessian matrix is difficult and time consuming. One technique to speed up the computing is to use approximations such as Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula or Davidon–Fletcher–Powell (DFP) algorithm [64]. Usually, the approximation to Hession matrix is build up iteratively starting from the identity matrix. Another concern is that the Quasi - Newton technique has a tendency to stuck in a local minima than the other optimization techniques [64].

2.4.4 Levenberg–Marquardt Method

Levenberg-Marquardt (LM) is a nonlinear optimization algorithm which uses second order derivatives and Hessian matrix for the optimization process. Even though this method was introduced in 1960's, it was adapted for the optimization of feed forward neural network models with the aid of backpropagation in 1994 [64]. The LM algorithm can be considered as a combination of features of gradient descent backpropagation and Newton methods. It works based on summed square error functions and similar to the Quasi-Newton method, computational power and memory required for calculating and adjusting weights is significantly increases with the increase of number of weights in the model.

The algorithm tests the model with adjusted weights to check whether error is reduced. The adjusted weights are accepted only if there is an improvement on optimization process, otherwise the weight updates are rejected. During the optimization, Hessian matrix is computed by approximating that the product of Jacobian matrix (J) of first order partial derivatives and its transpose (J^T) which gives the Hessian matrix as given in the equation 2.61.

$$H \approx J^T J \tag{2.61}$$

The weight update equation can be written as;

$$\Delta w = -(J^T J + \mu I)^{-1} J^T e \tag{2.62}$$

where, I is the identity matrix, μ is the control parameter, and e is the minimized error matrix. When $\mu \approx 0$, The LM algorithm becomes similar to the Newton method with an approximated Hessian matrix, and for larger values of μ the algorithm is similar to gradient descent method. Comparing with the other methods, the convergence of the algorithm is rapid, and the converging process is continued until the desired error or number of iterations is reached.

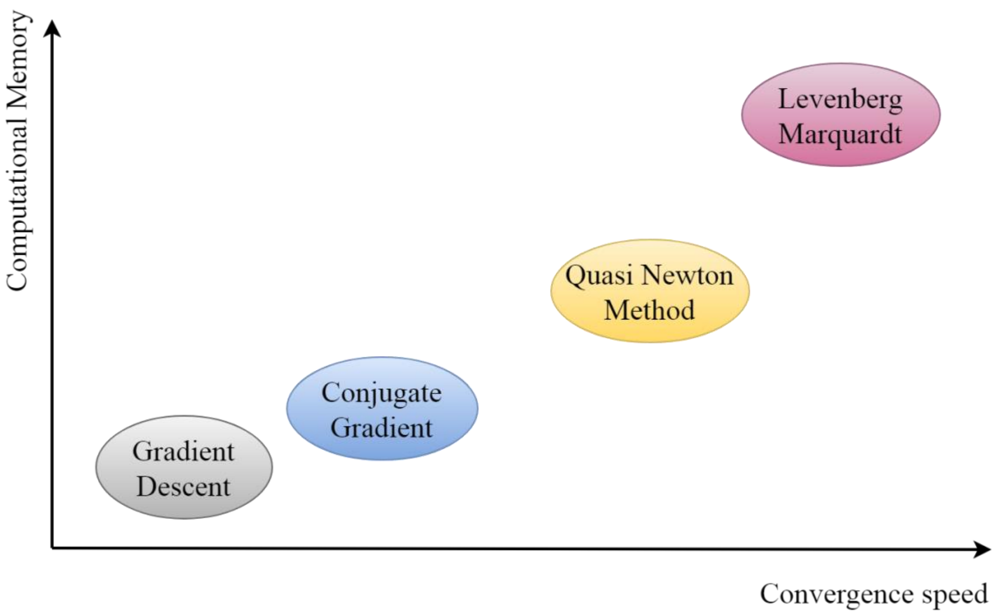


Figure 2-12: Performance comparison of optimization algorithms [65]

Figure 2.12 shows a comparison of computational memory required and convergence speed of each optimization algorithms discussed above. Gradient descent and conjugate gradient methods are the slowest algorithms to converge, but they use less computer memory. With the fastest convergence among discussed methods, Levenberg-Marquardt algorithm is the best option to use if there are many neural networks to train. However, the LM algorithm requires very high memory for the training of networks with as many as thousands of parameters.

2.5 Application of Forecasting Techniques in Solar Power Related Forecasting

The existing solar forecasting models can be found in different time horizons starting from short-term or ultra-short-term forecasting to long term forecasting. There is no globally agreed classification criterion for prediction periods and overlapping the prediction period among different time horizons can be observed in some of the available research work [66]. Commonly classified time horizons are listed in the Table 2.4.

Other than the classification listed in the Table 2.4, time horizons divided as intra-hour (span from a few seconds to an hour), intra-day (span from 1 to 6 hours) and day-ahead (span from 6 hours to 48 hours) are also found to be common in solar forecasting sector. Various solar forecasting models have been developed over the years by many researchers using various input parameters and forecasting techniques for different time horizons [67]–[69]. Also, solar forecasting models include both solar irradiance forecasting, and solar PV power forecasting as given in Figure 2.13.

Table 2-4: Different time horizons in solar forecasting

Time Horizon	Prediction period
Very short-term or ultra-short term	from few second to less than 30 minutes. (Sometimes, from 1 minute to several hours)
Short-term	From 30 minutes to 360 minutes (Sometimes, one to several hours, one day or up to seven days)
Medium-term	From 6 to 24 hours. (Sometimes, from one day, one week or up to one month)
Long term	More than 24 hours in advance (Sometimes from one month to a year)

Majority of the existing models are artificial neural network based models [70]–[72] whereas regression [73] and ARIMA [67], [74] based models are also common. Some of the models have been developed by combining two or more forecasting methods together. In solar irradiance forecasting, various input parameters such as solar irradiance, air temperature and sunshine duration have been used commonly. The list can be extended including the parameters such as atmospheric pressure, relative humidity, wind speed, wind direction, location coordinates, sky image, cloud cover, precipitation, zenith angle, month and day number.

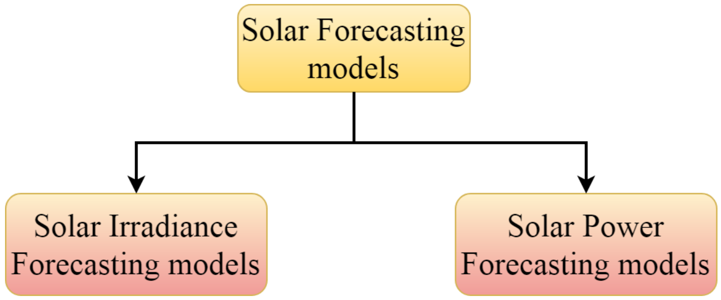


Figure 2-13: Solar forecasting models

The list is almost same for solar power forecasting, whereas solar irradiance forecast can also be used as an input [75]. Forecasting models developed by various researchers for different time horizons have been summarized in the Table 2.5. Performance of these models have been evaluated using various tools such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Coefficient of determination (R^2) etc. Among the available models, Machine learning based models are observed to be having better performance most of the times when compared with other type of models.

For example, [76] has done an assessment on the performance of three deferent forecasting models developed by using the techniques: Autoregressive integrated moving average (ARIMA), Artificial neural network (ANN) and Genetic algorithm-based artificial neural network (GA-ANN). The research reveals that GA-ANN model produced the best performance and ARIMA model had the worst performance. [77] also suggested that the ANN model developed by them is performing better than the ARIMA and Multiple linear regression (MLR) models developed for the forecasting of solar irradiance in long term average basis.

Even though there are already available models for solar power prediction, adopting an existing forecasting model to a different system is always a challenge, because the compatibility of the model depends on the factors associated with the new system such as required forecasting time horizon, availability of the data for input parameters and geographical location etc.

Table 2-5: Some of the available research work on solar forecasting

Author	Forecasting model	Input data	Forecast horizon
Antonio Sanfilippo et al. [67]	Autoregressive modeling (AR)	Global horizontal, diffuse horizontal and direct normal irradiance	1 min.
Amarasinghe,. and Abeygunawardane, [70]	Machine learning models	Solar power, Weather data	1 min.
Mashud Rana et al. [73]	Support vector regression (SVR)	Solar power, wind speed, pressure, irradiance, temperature, humidity	10 min.
Pedro, and Coimbra [71]	Artificial Neural Network (ANN)	Global horizontal irradiance	15 min.
Weerasekara, [72]	Artificial Neural Network (ANN)	Rainfall, Solar irradiation, Morning & Evening cloud cover, Hour number	1 hour
Wu, and Wang, [78]	Self-organizing map-based extreme learning machine (SOM-ELM)	Zenith angle, azimuth angle, extraterrestrial, diffuse, direct, and global solar radiation	1 hour
Munshi, and Yasser [79]	Bio-inspired clustering algorithm	PV power patterns, solar irradiance, ambient temperature	2 hours
McCandless et al [80]	k-means algorithm-based ANN	Temperature, wind speed, cloud cover, precipitation	2 hours
Vaz, et al. [74]	Autoregressive with exogenous inputs ANN	Solar power, solar radiation, ambient temperature	1 day
Prema, and Rao [68]	Triple exponential smoothing model	Time, temperature, humidity, solar irradiance	1 day
Sivaneasan, et al. [81]	ANN and fuzzy logic	Temperature, Dew point, Wind speed, Solar irradiance	Monthly
Huang, and Perry [69]	Gradient boosting machine (GBM)	Numerical weather data	Monthly

2.6 Evaluating a Forecasting Model Performance

Performance evaluation of a newly designed forecasting model is very important. This provides an opportunity to compare the new model with the existing similar type of models and also to assess how effectively the original goals of having a forecasting model are accomplished. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) are the most common statistical indicators used in model evaluations in the literature.

MAE estimates the error of the forecast data set by computing the average difference between the predicted and actual values of the entire test data sample. Generally, MAPE is used as a standard technique to demonstrate the accuracy and diversity of a forecast data set. On the other hand, MSE estimates the error by computing average square difference between predicted and actual observations. RMSE is the square root of MSE and it is more sensitive to the undesirable large deviations of predicted values from the real data set, allowing the researchers to identify and eliminate the outliers. If there are models with different scales, normalized MAPE (nMAPE) and normalized RMSE (nRMSE) are better options for comparison. The mathematical equations for MAE, MAPE, MSE, RMSE are as follows [66], [76].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|t_i - o_i| \tag{2.63}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|t_i - o_i|}{t_i} \times 100\% \tag{2.64}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - o_i)^2 \tag{2.65}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - o_i)^2} \tag{2.66}$$

$$nMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|t_i - o_i|}{\frac{1}{N}\sum_{i=1}^{N} t_i} \times 100\% \tag{2.67}$$

$$nRMSE = \frac{\sqrt{\sum_{i=1}^{N}(t_i - o_i)^2}}{\sqrt{\sum_{i=1}^{N} t_i^2}} \tag{2.68}$$

where, t_i and o_i are the actual and predicted values respectively and N is the number of samples in the test.

Coefficient of determination (R^2) is also a good measure in demonstrating how close the predicted values of the model to the actual values. R^2 gives the proportion of variation of the predicted output from the actual and, it can be calculated using the equation 2.69 [82].

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{2.69}$$

where SS_{res} and SS_{tot} are called residual sum of squares and total sum of squares, respectively which can be calculated using equations 2.70 and 2.71.

$$SS_{res} = \sum_{i=1}^{N}(t_i - o_i)^2 \tag{2.70}$$

$$SS_{tot} = \sum_{i=1}^{N}(t_i - \bar{t})^2 \tag{2.71}$$

where $\bar{t}$ is the mean of the actual values in the test data sample, which is defined by,

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i \qquad (2.72)$$

For a perfect prediction model SS_{res}= 0 and R^2 = 1. Any model having the coefficient of determination closer to 1 can be expected to perform better.

Except the above-mentioned statistical indicators to assess the accuracy, distribution of errors of the test sample can be analyzed in different ways. Dispersion of errors relative to mean error of the test sample can be demonstrated by computing the standard deviation (SD). Low standard deviation explicates high precision of the model [82].

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(e_i - \bar{e})^2}{N-1}} \qquad (2.73)$$

where, e_i represents each error value associated with the predicted value in the test sample and $\bar{e}$ is the mean error.

Furthermore, errors should be distributed randomly without any noticeable pattern and free of systematic errors for a better performance of the model. This can be graphically analyzed by plotting orderly arranged residuals. The other option is plotting error histogram, because randomly distributed error data should show a normal distribution with zero mean. Testing for autocorrelation of residuals Durbin-Watson Statistic (D) can be used, which is calculated using equation 2.74. D varies from 0 to 4, with values around 2 indicating weak or no correlation.

$$D = \frac{\sum_{i=2}^{N}(e_i - e_{i-1})^2}{\sum_{i=1}^{N} e_i^2} \qquad (2.74)$$

2.7 Summary of Literature Review

Power generation of a PV system could be successfully forecasted using GTI on PV panels and the other associated system parameters. GTI could be estimated substituting GHI and the basic solar parameters in a transposition model. Section 2.3 investigated different forecasting methods which could be considered for developing a solar power forecasting model. Any quantitative forecasting method is primarily classified as a time series method or a causal method. In time series methods, the future patterns are predicted based on the patterns observed in the historical data. Moving average, exponential smoothening, decomposition and Box-Jenkins methods are examined under time series forecasting. Furthermore, causal methods use a set of dependent variables possibly also including time series components to establish a statistical relationship with the forecasted variable. Regression analysis and artificial neural networks are discussed under causal forecasting methods. Artificial neural networks have been utilized to develop most of the existing solar power or solar radiation forecasting models, especially because they are observed to be performing well in handling large data sets and when forecasting in different time horizons. An artificial neural network could be trained using optimization algorithms such as gradient decent, conjugate gradient, Levenberg Marquardt and quasi-Newton. Levenberg Marquardt has faster convergence compared to the other algorithms, however the algorithm utilizes a high computational memory. Performance of a forecasting model could be evaluated using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) or Root Mean Square Error (RMSE). Coefficient of determination (R^2) is also a good measure in demonstrating how close the predicted values of the model to the actual values. Furthermore, standard deviation (SD) of the error terms indicates the precision of the model and Durbin-Watson Statistic (D) indicates any autocorrelation between error terms.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

The primary goal of the study was to develop a one day ahead energy output forecasting mechanism for a commercial or utility scale solar PV power plant which might be already running or still under construction. A schematic diagram of the proposed forecasting model is shown in Figure 3.1.

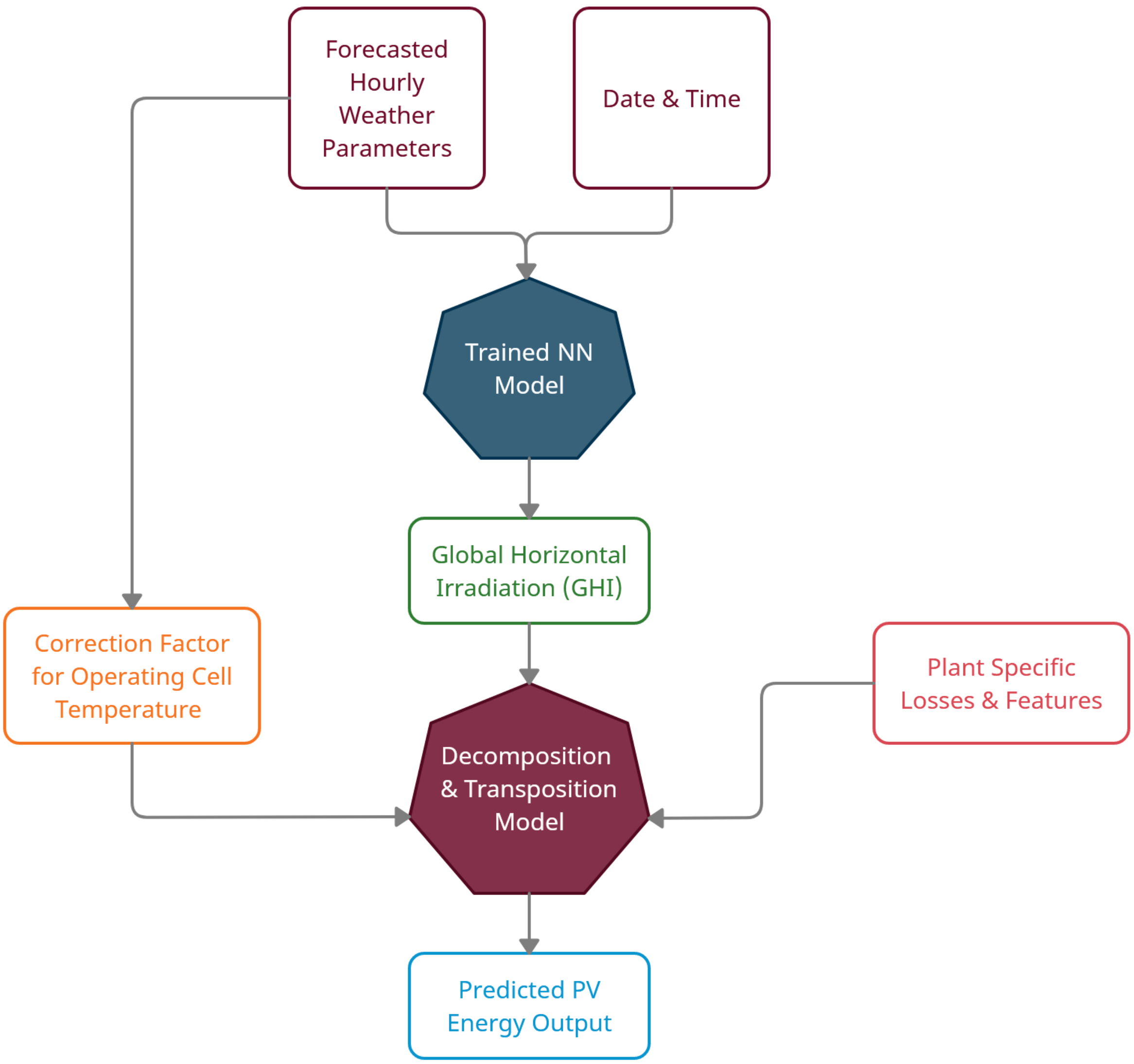


Figure 3-1: Schematic diagram of the forecasting model

The model consists of two sub-models: Trained Neural Network (NN), Decomposition & Transposition model, coupled together which require five different types of inputs to produce a forecast. First, the trained NN model requires forecasted hourly weather parameters for the next day together with relevant date and time for calculating the GHI as an output. Next, the output from the NN model is converted to Global Tilted Irradiance (GTI) by Decomposition and Transposition model and the predicted PV energy output is generated using the relevant plant specific features and losses together with a temperature adjustment for the operating PV modules. The proposed model has been implemented for an existing ground mounted 1 MW plant situated in Beliatta, Sri Lanka (Latitude: 6.08; Longitude: 80.74) which was used as a case study for the model training, validation and forecasting.

3.2 Collection of Data

As illustrated in Figure 3.2, four sets of data were collected for using in three different stages of the study. Historical hourly solar radiation (GHI) and hourly weather data were required for the training of NN model. Relevant plant specific features and losses were obtained through site visits, manufacturer catalogues and existing research work in similar applications, for the use in decomposition and transposition model. Finally, the generation data of the plant were collected for the use in validation process.

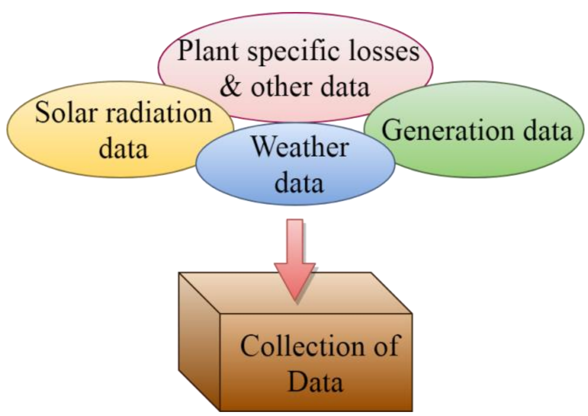


Figure 3-2: Collection of data for different stages of the study

3.2.1 Weather and Solar Radiation Data

For the collection of weather data, Sri Lanka Metrological Department has 410 regular measuring stations located island wide which provide hourly data for parameters such as rainfall, temperature, relative humidity, wind speed/direction and cloud type [83]. However, the number of data from the measuring stations which provide hourly solar radiation data is limited. Also, if the plant is situated away from a measuring station, observation values for the plant location should be extrapolated from observations taken from nearby stations. the accuracy of the calculated data can be reduced significantly when the number of measuring stations is less.

Because of this, alternative data sources with readily available weather data for any location were considered for the study. According to the literature, weather and solar radiation APIs (Application Programming Interfaces) are very good data sources which provide the access to large databases of historical weather and solar radiation information. Most referenced or used weather APIs in the recent past are shown in the Figure 3.3 [84]–[88].

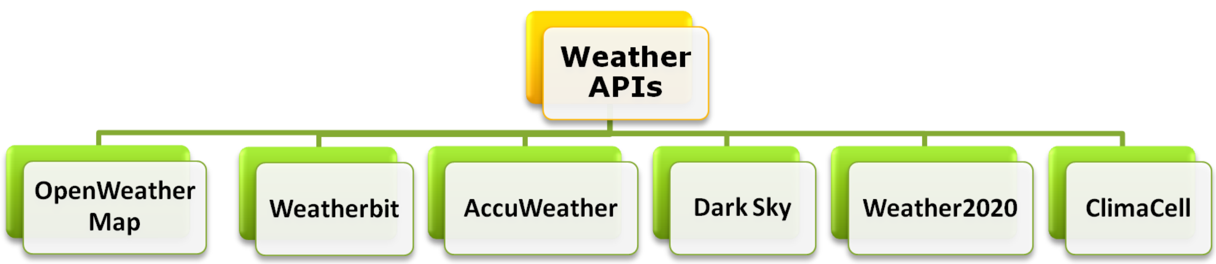


Figure 3-3: Different weather APIs

Generally, the hourly data in weather APIs are provided in Universal coordinated time (UTC) which is 05.30 hours behind the Sri Lankan time. For this study, 14 years of historical hourly weather data starting from 2007 January for the plant location were collected from OpenWeather Map API. Table 3.1 shows the weather parameters for which the data have been collected and the respective readings of the parameters as at 11.30 am on 14th of January 2021.

Table 3-1: Sample of hourly weather parameters on 14th of January 2021 at 11.30 am

Parameter	Cloud cover	Temperature	Relative Humidity	Rain	Wind Speed
Unit	%	^{0}C	%	mm for last hour	m/s
Sample	75	28.8	84	0	2.1

Similar to the above weather APIs, the solar radiation APIs illustrated in Figure 3.4 have been widely used by many researchers and professionals around the world in recent past [89]–[92].

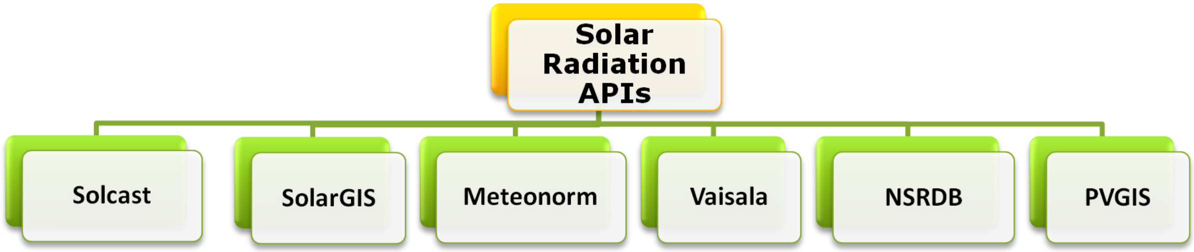


Figure 3-4: Different solar radiation APIs

Hourly solar radiation data for the parameters shown in the Table 3.2 were collected for the same time period used for the weather API.

Table 3-2: Sample of hourly solar radiation parameters on 14th of January 2021 at 11.30 am

Parameter	Global Horizontal Irradiance (GHI)	Direct Normal Irradiance (DNI)	Diffuse Horizontal Irradiance (DHI)	Ground Albedo (ρ_g)
Unit	W/m^2	W/m^2	W/m^2	-
Sample	544	14	532	0.09

3.2.2 Plant Specific Data

The solar PV plant consists of 3456 PV modules of rated capacity 340 W each, ground mounted, south facing with a fixed tilt angle of 8^0. The generated D/C power of the plant is converted to A/C through 10 numbers of 100 kW rated power (A/C) inverters. The total A/C power generated has been connected to the grid through a 400 V: 33 kV step up transformer. The basic layout plan of the plant is illustrated in the Figure 3.5.

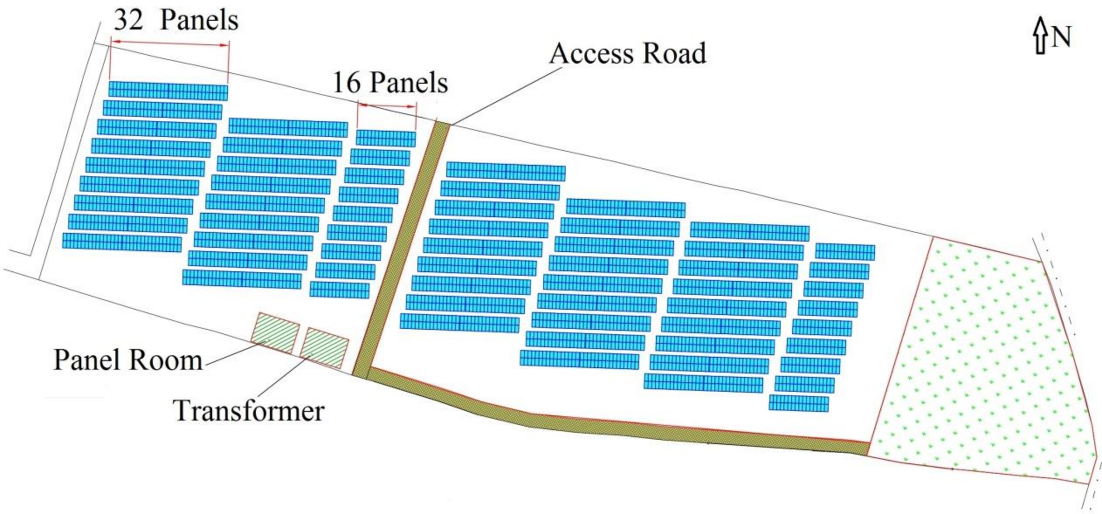


Figure 3-5: Layout plan - Beliatta ground mounted solar PV plant

Table 3-3: Manufacturer specifications of solar PV modules and inverters

1.	**Solar PV modules** [93]–[95]	
	Model	SPR-P17-340-COM
	Nominal power	340 W
	Power tolerance	+5-0%
	Efficiency	16.5%
	Power-temperature coefficient	-0.37%/^{0}C
	Operating temperature	-40 ^{0}C to 85 ^{0}C
	Solar cells	Multicrystalline
	Tempered glass	High transmission tempered anti-reflective
	Panel area	1.93 m^2 approx.
	Panel surface dimensions	2067 mm × 998 mm
	Degradation (1st year)	3 %
	Degradation (from 2nd year)	0.6%/year
	NOCT at 20^{0}C (Value +/- 2^{0}C)	47
	Module voltage mismatch loss	0.5%
	String current mismatch loss	0.3%
2.	**Inverters** [96]	
	Model	PVS-100/120-TL
	Rated DC input power	102 000W
	Rated AC power	100 000 W
	Maximum efficiency	98.4%

The details included in the Table 3.3 were obtained from manufacturer data sheets, product manuals and design guides.

3.2.3 Generation Data

Full operations of the plant were commenced on 1st of September 2019. Daily generation data (in kW h) were collected for the period from commencing date to 30th April 2021.

3.3 Identifying Input Parameters for the ANN Model

As illustrated in the Figure 3.1, the trained ANN model was used for the prediction of hourly GHI values. In order to train the model, suitable input parameters had to be decided considering the correlation of different parameters with the model output. Variation of GHI with some of the parameters could be clearly visualize from graphical representations. Also, calculating the correlation coefficient was more accurate option because it gives a quantitative estimate on how strongly two parameters have been correlated each other.

Correlation coefficient (r_{xy}) between two variables x, y could be calculated using the equation 3.1 [82].

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.1}$$

where, x_i, y_i are the values of *x* and *y* variables, $\bar{x}, \bar{y}$ are the mean values of *x*, *y*, respectively and *n* is the total number of data points.

In order to identify input parameters, collected weather and solar radiation data were analyzed using the correlation coefficient. Day Number, Hour Number, Temperature, Relative Humidity, Rain fall and Clouds percentage were identified as the input parameters based on the results obtained from the analysis. Details of the calculations of correlation coefficients are given in section 4.1.

3.4 Preparation of Data for the ANN Model

Collected weather and solar radiation data consisted of values for the entire 24 hours in hourly granules. However solar radiation is available only during the daytime. Because of this, both weather and solar radiation data were merged together and all the data sets from 6.30 p.m. to 6.30 a.m. were removed from the collection, assuming the solar radiation received during this period are negligible. Furthermore, time given in the data set were converted from coordinated universal time zone (UTC) to local time and temperatures were converted to ^{0}C from K. In addition to the weather parameters, date and time should be included in the list of input parameter in order to learn seasonal and daily GHI variation patterns when the model is trained. For the ease of handling the date and time as input parameters, following numbering method was used;

1. All the dates were given day numbers starting from 1^{st} of January, numbered as '1' and ending with 31^{st} of December, numbered as '365'. For a leap year (which comes every four years), 31^{st} of December was numbered as '366'.

2. All the hourly time periods were also given hour numbers so that hour starting at 6.30 am (6.30 am to 7.30 am) was numbered as '1' and hour starting at 5.30 pm (5.30 pm to 6.30 pm) was numbered as '12'

Span of the data set was found to be largely different from one parameter to the other. for example, temperature data lies within the range 20.8 - 33.3 ^{0}C while range of GHI is 0 -1075 W/m^2. This has been detailed in the Table 3.4

Table 3-4: Data span, Mean and Standard Deviation (SD) of the input - output parameters for NN model

	Day Number	**Hour Number**	**Temperature (^{0}C)**	**RH (%)**	**Rain (mm/hour)**	**Clouds (%)**	**GHI (W/m^2)**
Minimum	1	1	20.8	20	0	0	0
Maximum	366	12	33.3	100	23.5	100	1075
Mean	181	6.5	27.9	78.3	0.1	45.2	400
SD	106	3.5	1.5	11.4	0.5	27.6	301

The data span in substantially different ranges for different parameters could be caused difficulty in converging during the model training stage and also causes the NN model to perform poorly. In order to avoid these effects, data set was scaled using minimum- maximum scaling method,

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (3.2)$$

where, x_i is the original input, z_i is the scaled input, $\min(x)$ is the minimum value and $\max(x)$ is the maximum value of the relevant input/output parameter.

3.5 Training the NN Model

The NN model was designed with six input parameters described in section 3.3 and one output parameter which is GHI. Also, it included a scaling layer next to the input layer and an un-scaling layer before the output layer. Required number of hidden layers and hidden layer sizes were decided by training number of networks with different configurations and analyzing the performance of each. The basic architecture of the NN model which was used for training is shown in the Figure 3.22

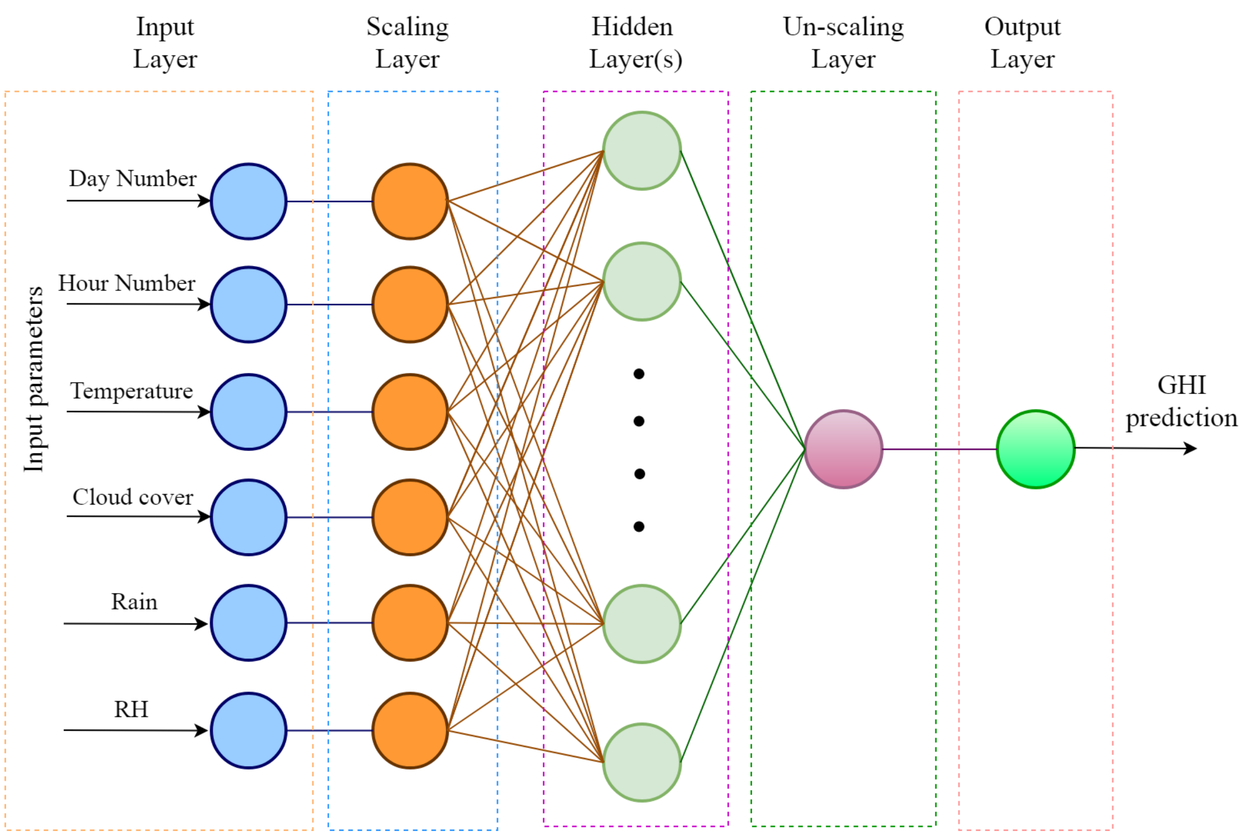


Figure 3-6: Neural Network (NN) model architecture

Before starting the training of the model, the scaled data set was divided into two sub-sets. First sub-set, which is from 1st of January 2007 to 31st December 2019 was used as training input/output for the NN model. The second sub-set which is from 1st of January 2020 to 28th February 2021 was kept as an 'unseen data set' for the model and was used to analyze the generalization ability of the model by presenting to the trained NN model after the training.

Customized codes generated from MATLAB Nftool were used for running the training process. For this, first data sub-set mentioned above was further divided in to 3 sets randomly as 'training set', 'validation set' and 'testing set' as shown in the Table 3.5. Synaptic weights for the model were initialized with random values and adjusted along with other parameters such as learning rate, using Levenberg-Marquardt backpropagation algorithm. Generally, the data in the training set are utilized for the weight adjustments whereas the data in the validation set are utilized for assessing the performance of the model while the training process and for adjusting the other parameters without over fitting. Finally, the data in the testing test are utilized for assessing the predictive power and the generalization capability of the trained model.

Table 3-5: Details of data allocation for training, validation and testing of the model

Data set	Number of observations	Percentage
Training	39884	70%
Validation	8546	15%
Testing	8546	15%

3.6 Decomposition & Transposition Model

As explained at the beginning of the chapter, The hourly GHI forecast obtained from the NN model was fed into the Decomposition and Transposition model (Fig. 3.7) in order to estimate the DHI (Diffuse horizontal irradiance) and EBH (Direct beam irradiance) components using the Erb's correlation. The DHI, EBH and GHI values are utilized in the HDKR model for the calculation of GTI (Global tilted irradiance) on the solar PV panels. It was observed, the ground albedo was slightly fluctuating between 0.08 and 0.1 for the location throughout the years according to the Solcast historical solar radiation related data. Therefore, average albedo value was taken as 0.09 for the calculations.

Table 3.6 includes the percentage losses observed for the solar plant. The values were decided based on the information available in manufacturer data sheets, published research work and current industrial practices.

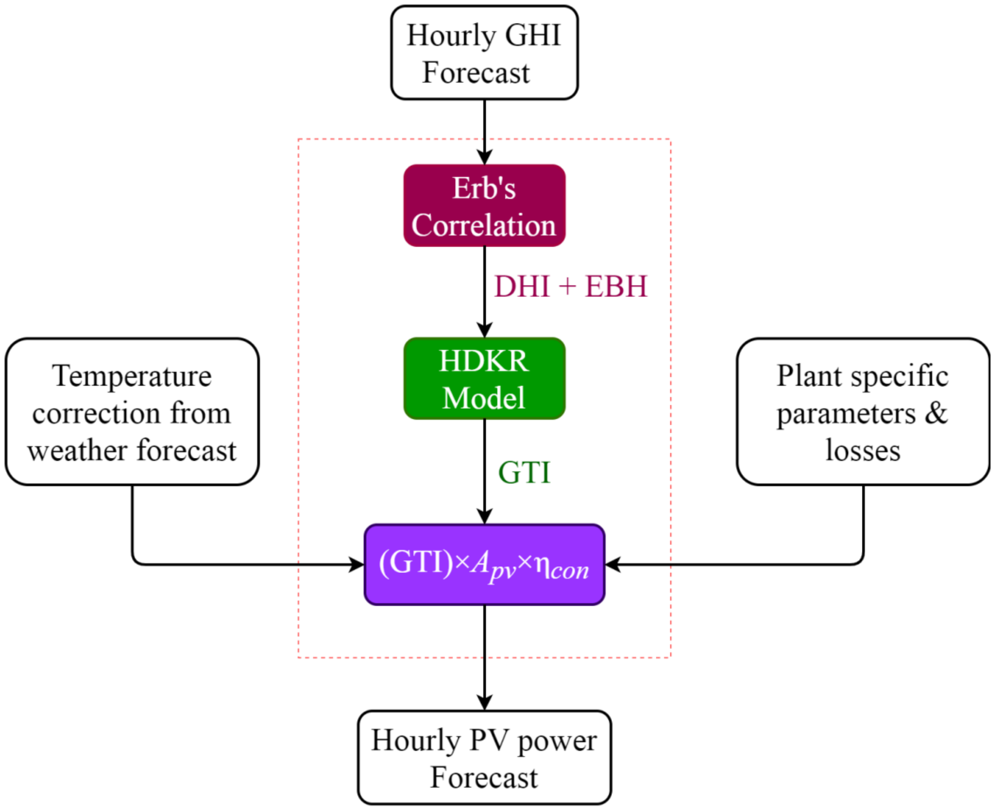


Figure 3-7: Schematic diagram of the Decomposition & Transposition model

Other than the losses given in the Table 3.6, module quality loss was taken as 1.5% based on the given manufacturer's tolerance and soiling loss was assumed to be increased linearly as the time goes on. Two major solar panel cleaning events have been undertaken in 2019 December and 2020 September, so the soiling loss is expected to be minimum in these months. The plant location is not an urbanized or highly populated area, however a considerable amount of sand and dust deposits and bird's dung were observed. Considering this, the soiling loss on 1st of January 2020 was taken as 0.1% with 0.08% daily increase. Soiling loss was again re-set to 0.05% on 17th September, after the completion of the plant cleaning event. Finally, a shading loss analysis was conducted for the site to obtain monthly average hourly shading losses.

Table 3-6: Percentage of different PV losses

	Types of losses	Percentage loss
1.	Degradation (1st year)	3.0%
2.	Reflection loss	2.5%
3.	Inverter loss	2.5 %
4.	Array over sizing loss	1.5%
5.	Electrical mismatch losses	1.5 %
6.	Loss due to low irradiance level	1.2%
7.	Loss in DC cables	1.0 %
8.	Loss in the Transformer	1.0%
9.	Spectral loss	0.7%
10.	Degradation (from 2nd year)	0.6%
11.	Loss in AC cables & Transmission	0.5%
12.	Auxiliary power loss	0.1 %

3.7 Validation of the Entire Forecasting Model

The system efficiency (η_{sys}) of the plant was calculated accounting all the losses described in the section 3.6. Then, the conversion efficiency (η_{con}) was calculated using equation 2.13 with the aid of temperature forecasts, and PV module characteristics specified by the manufacturer. Finally, the hourly PV power forecasts were generated in a way as illustrated in the Figure 3.7. However, validation of the model for hourly forecasts was not possible since the unavailability of actual historical hourly power generation data in the plant. Nevertheless, the plant authorities have maintained daily energy generation records in kW h. Therefore, the hourly power forecasts were converted to daily power generations and compared with the actual daily generation data collected for a period of 14 months, from 2020 January to 2021 February.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Analyzing the Variation of GHI with Different Weather Parameters

To identify the input parameters for the NN model, sensitivity of the output parameter (GHI) to the different weather parameters were analyzed as explained in the section 3.3. The results of the variation of GHI with different parameters are given in the subsequent sections.

4.1.1 Variation of GHI with the Day of a Year

To visually observe any seasonal variation of GHI throughout a year, the hourly solar radiation values were converted into monthly averaged daily insolation and plotted against a period of one year for the years from 2015 to 2020 as illustrated in Figure 4.1.

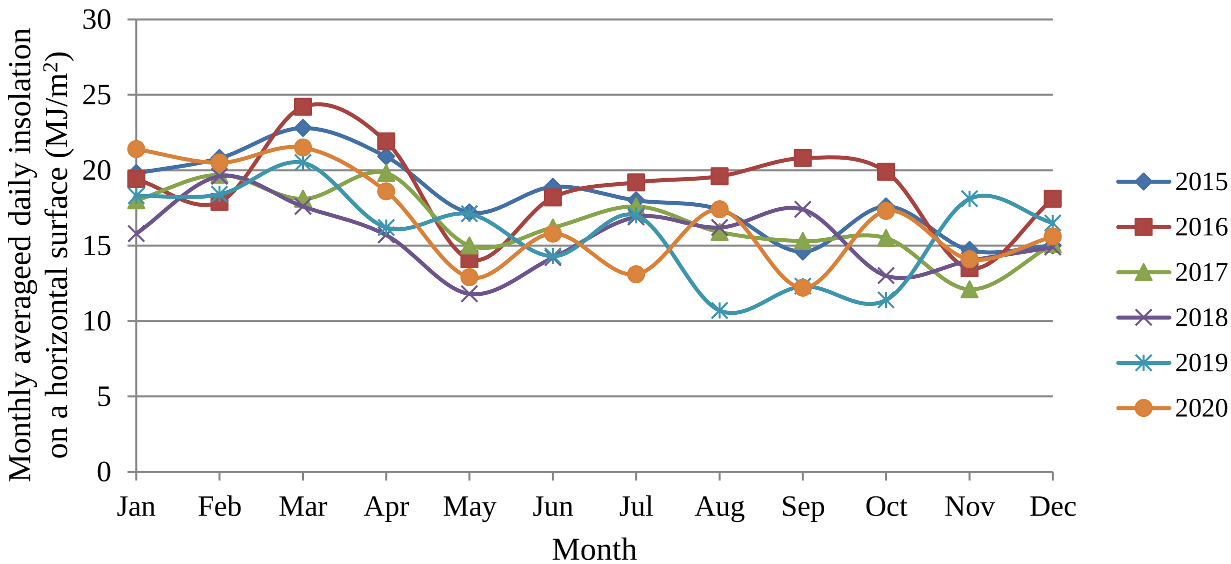


Figure 4-1: Variation of monthly averaged daily insolation on a horizontal surface over the year

According to Figure 4.1, the solar PV plant received more solar radiation during the period from January to April with a peak somewhere in March. Also, a significant drop in solar energy insolation was observed from April to May for most of the years and the variations were not that significant for the period from June to December.

4.1.2 Variation of GHI During a Day

Intensity of receiving solar radiation during a day gradually increases starting from the sunrise in the morning, reaches to a maximum somewhere in the middle of the day, then gradually reduces and reaches to zero with the sunset. This makes a bell-shaped variation of GHI during the daytime for a clear sunny day.

Figure 4.2 shows the variation of GHI during the daytime of a particular day in past 10 years. The graphs were plotted for the time duration from 6.30 a.m. to 5.30 p.m. on 10th March of each year. The date was taken randomly and considered the same date for all the years to remove any seasonal change. Proper bell-shaped variations could be observed for the given day in the years 2011and 2014.

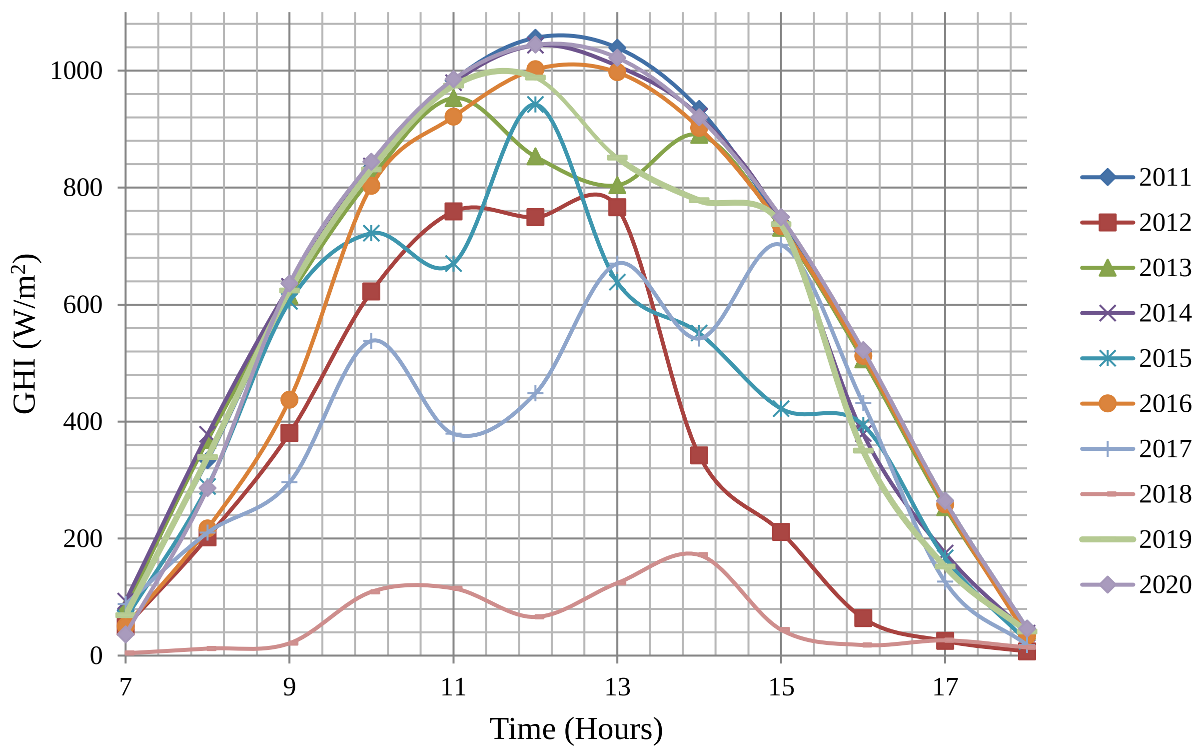
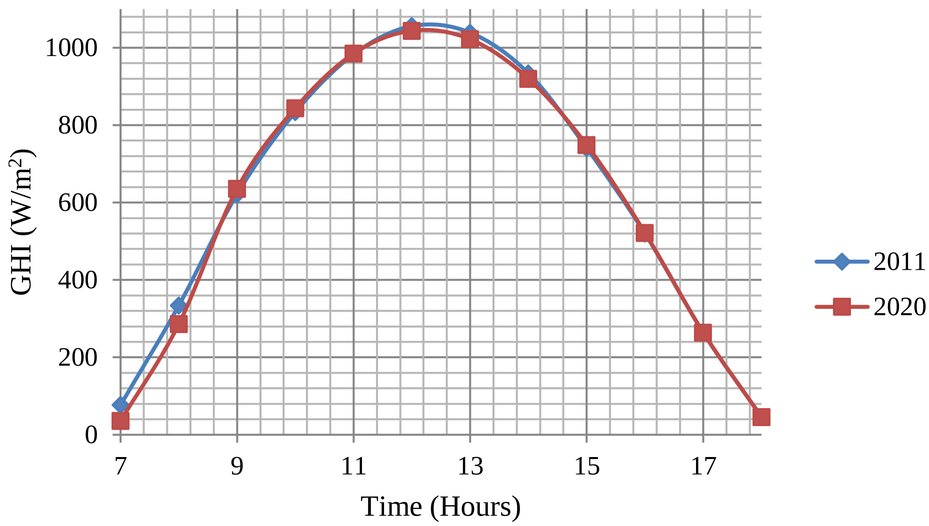


Figure 4-2: Variation of GHI during the daytime

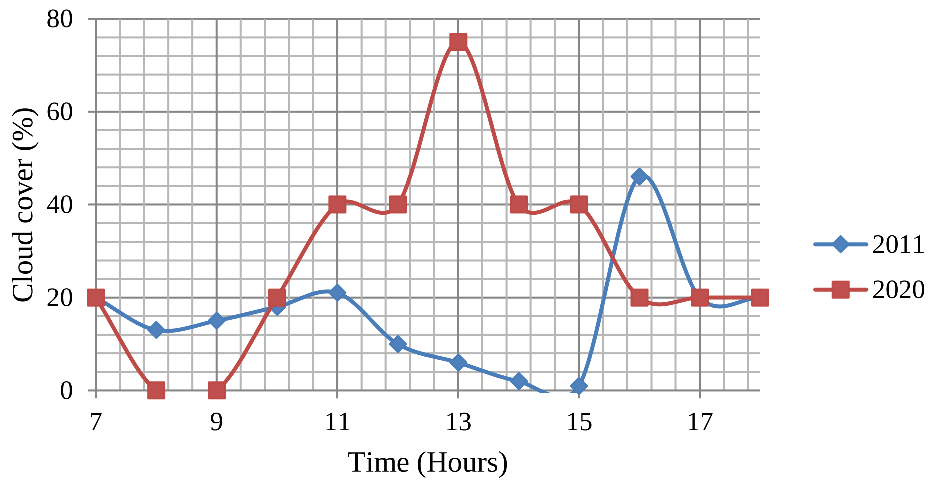
However, the regular variation of GHI could be significantly affected with the change of other parameters such as cloud percentage and rain, resulting different variations. This can be clearly seen for years such as 2017 and 2018 in Figure 4.2. Sensitivity of these parameters are discussed in subsequent sections.

4.1.3 Variation of GHI with Cloud Cover

To study the effect of different cloud cover percentages on the GHI, both cloud cover and GHI for several selected days were plotted as given in Figures 4.3 and 4.4. Same day (10th March) in each year was selected for the analysis to avoid any factors, seasonal variations effecting the results. Figure 4.3 (a) and (b) shows the variation of cloud cover percentage when the variations of GHI are similar in 10th of March 2011 and 2020. It was observed that the cloud cover percentage remained less than 40% for most of the time during for both the days except sudden increase of cloud cover around noon up to 75% in year 2020.



(a)



(b)

Figure 4-3: Similar variations of GHI in two different days (a) and respective variations of cloud cover (b) - plotted for 10th March 2011 and 10th March 2020

According to the Figure 4.3, it can be seen that when the cloud cover percentage is lower (i.e. lower than 40%), GHI is not significantly affected by the clod cover. However, Figure 4.4 illustrates the significant variation in GHI with the variation of cloud cover in year 2011 and 2018. On 10th March 2018, GHI level was less than 200 W/m^2 at any time and very high cloud cover percentage (80-90%) was observed throughout the day.

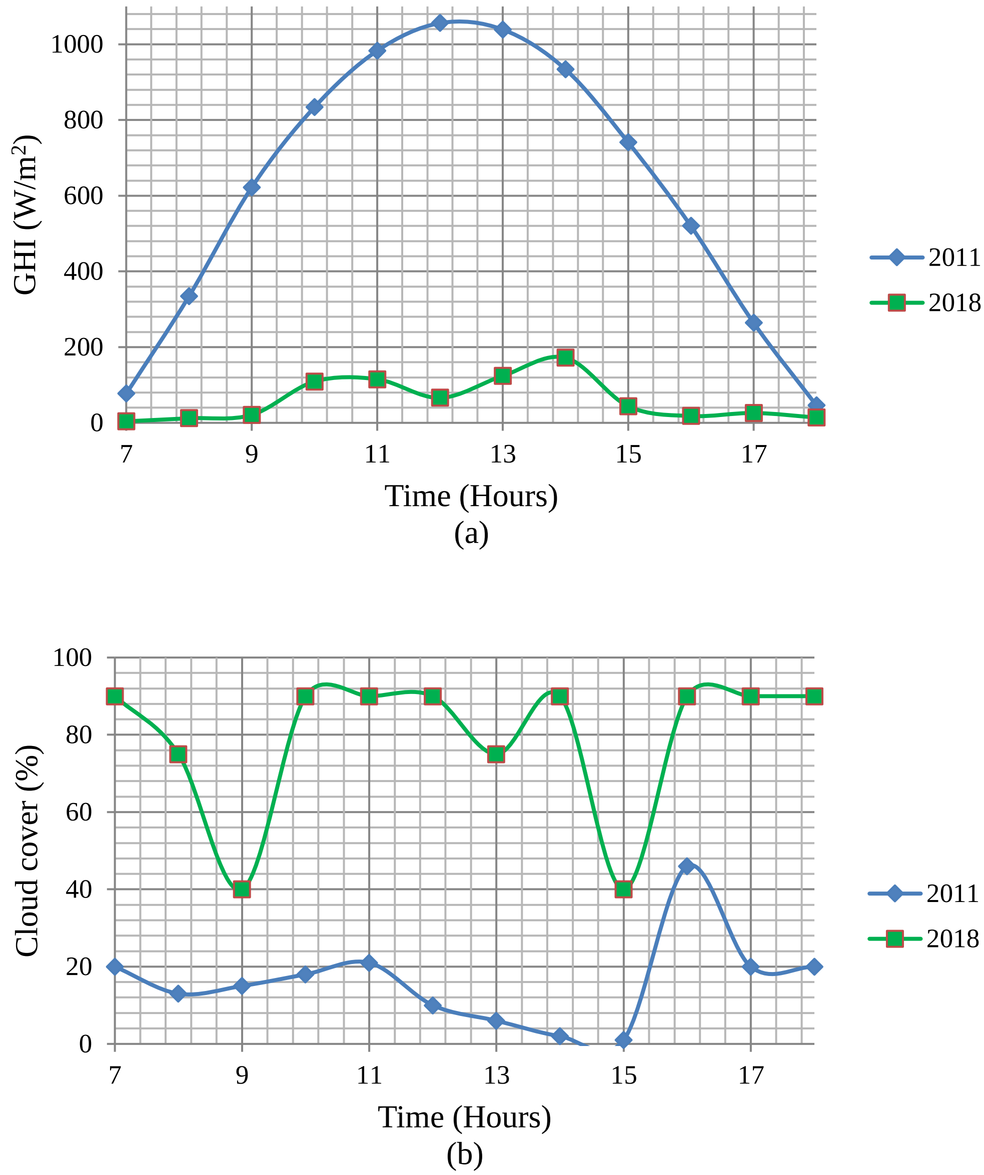


Figure 4-4: Different variations of GHI in two different days (a) and respective variations of cloud cover (b) - plotted for 10th March 2011 and 10th March 2018

The above evidence shows that there is dependence of GHI on the cloud cover percentage, specially at high cloud cover percentage. Hence, to verify the dependence, the variation of GHI with the cloud cover percentage was plotted and the correlation coefficient (r) was calculated. Year 2015 was randomly selected for plotting and time duration from 12.00 noon to 1.00 pm was considered to minimize the effect of other parameters on GHI.

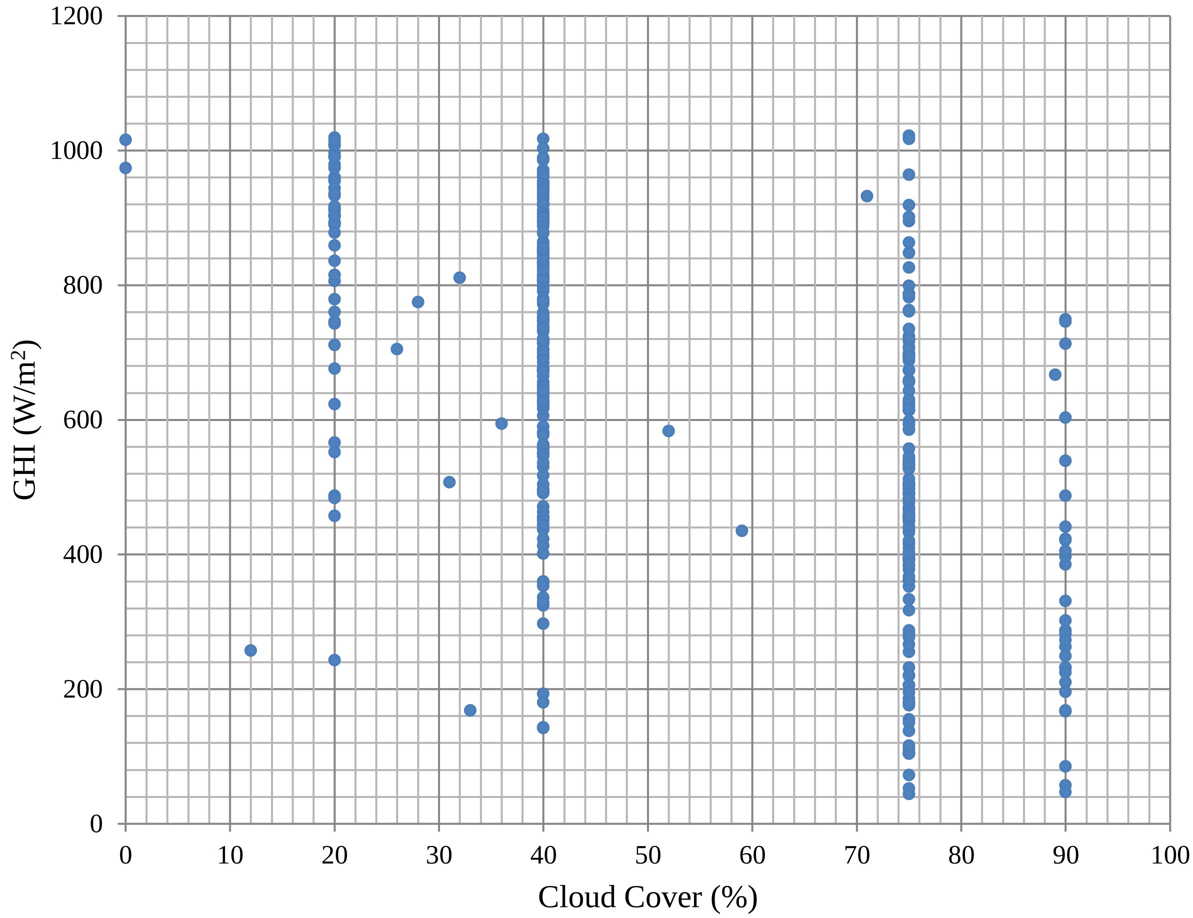


Figure 4-5: Variation of GHI with the cloud cover in the year 2015 from 12.00 noon to 1.00 pm

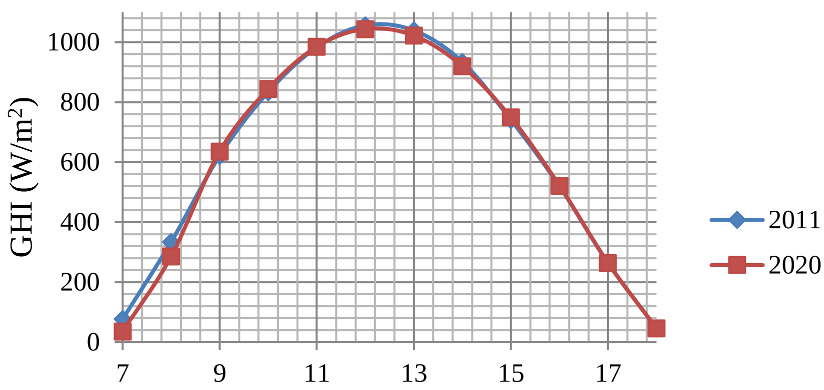
For the year 2015, calculated value for the correlation coefficient between GHI and cloud cover was -0.55. Table 4-1 shows correlation coefficients between GHI and cloud cover calculated similar manner for the years from 2015 to 2019. The five-year average (r_{avg}) of -0.48 demonstrates a moderate correlation between GHI and cloud cover.

Table 4-1: Correlation coefficient between cloud cover and GHI for different years

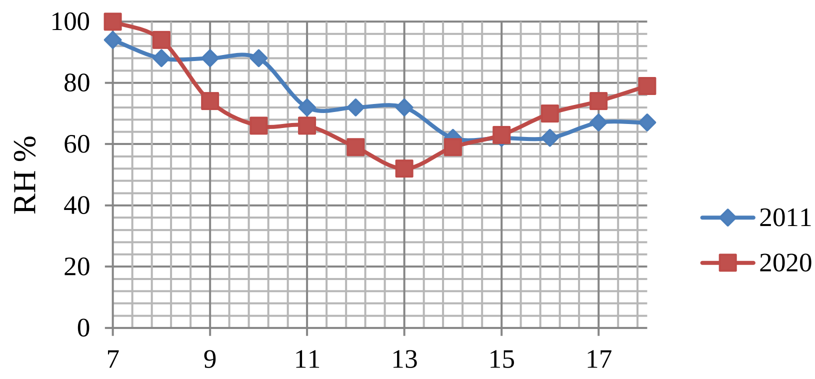
Year	2015	2016	2017	2018	2019
Correlation coefficient	-0.55	-0.47	-0.43	-0.35	-0.58

4.1.4 Variation of GHI with Relative Humidity

Generally high values of relative humidity (RH) could be expected in the morning of a day due to temperature reduction in the nighttime and reaching or nearly reaching the atmosphere to its dew point with the existing moisture content. RH value reduces with the time as more solar radiation receive in a clear sunny day and gradually increases again with the sun set and afterwards. This typical variation was observed in the Figure 4.6 which was plotted for two sunny days.



(a)



(b)

Figure 4-6: Similar variations of GHI in two different days (a) and respective variations of Relative humidity (b) - plotted for 10th March 2011 and 10th March 2020

However, when it comes to a day with low solar radiation as 10th March 2018 (due to high cloud cover) as shown in the Figure 4.7, a slight change in the variation of RH was observed. The RH tends to stay at high percentages rather than following the typical variation of a day as observed in sunny day.

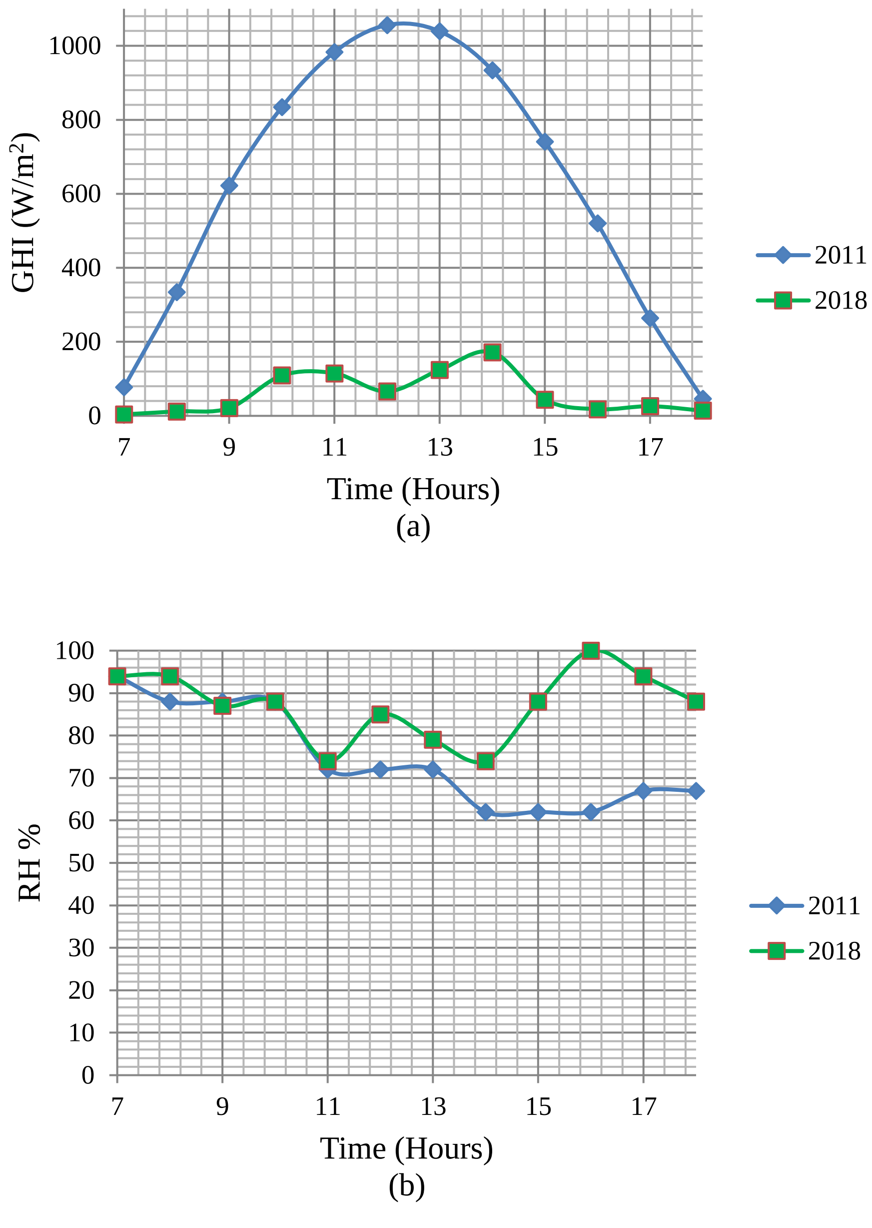


Figure 4-7: Different variations of GHI in two different days (a) and respective variations of Relative humidity (b) - plotted for 10th March 2011 and 10th March 2018

Based on the results shown in Figures 4.6 and 4.7, a variation of GHI with the variation of RH could be observed, however clear correlation cannot be established due to the RH interrelation with cloud cover. Hence, to further study the effect of RH, correlation between RH and GHI was further visualized by plotting the variation of two parameters as shown in Figure 4.8 and calculating correlation coefficient for independent year (year 2015). The calculated correlation coefficient was -0.33.

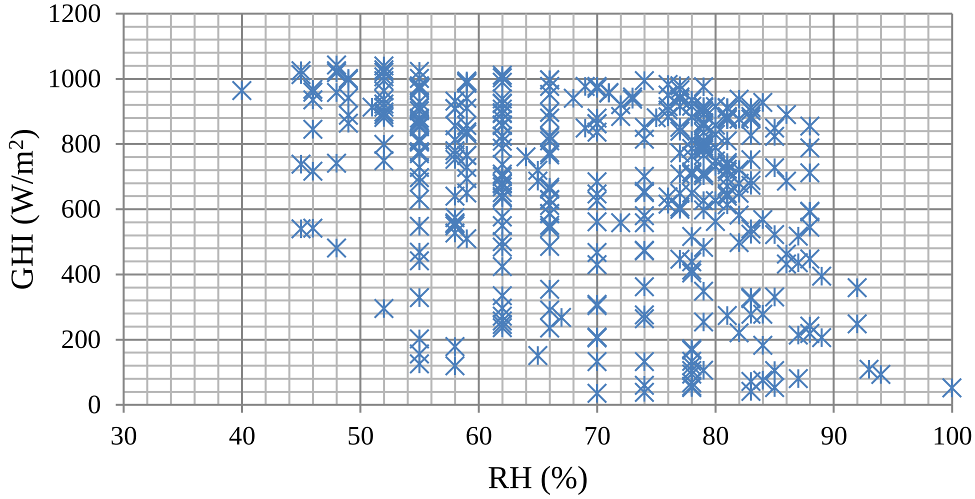


Figure 4-8: Variation of GHI with RH in the year 2015 from 12.00 noon to 1.00 pm

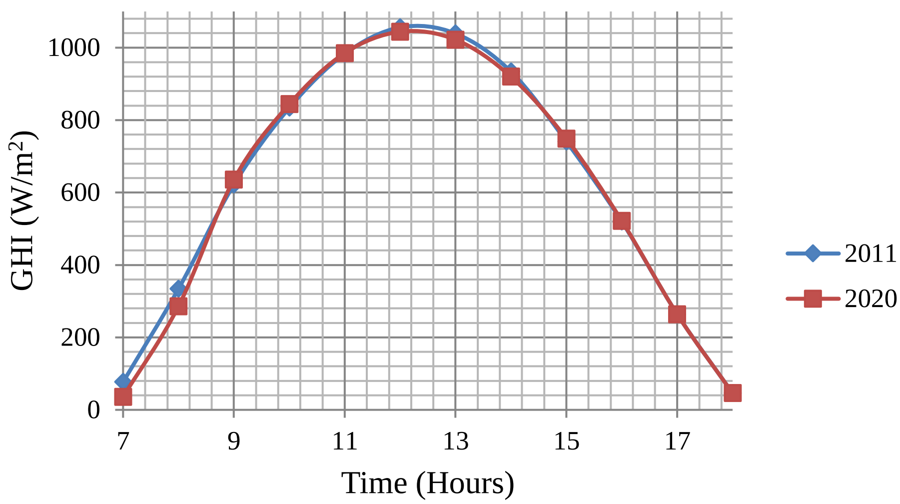
Table 4-2 shows the correlation coefficients between relative humidity and GHI calculated for different years, resulting five-year average of -0.46 which is a moderate correlation between the two parameters.

Table 4-2: Correlation coefficients between Relative humidity and GHI for different years

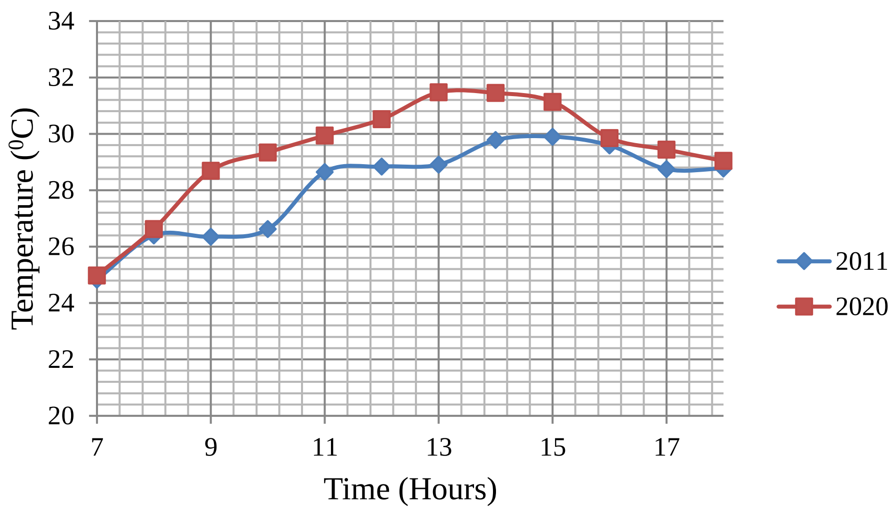
Year	2015	2016	2017	2018	2019
Correlation coefficient	-0.33	-0.33	-0.42	-0.57	-0.67

4.1.5 Variation of GHI with Temperature

High temperature values could be expected as more solar radiation receives during a day. This was observed by plotting GHI and temperature variation throughout the day for two similar sunny days in 2011 and 2020 as shown in Figure 4.9. As can be observed from the Figure 4.9, the temperature goes up with the increase of receiving solar radiation and reduces nearing the sunset.



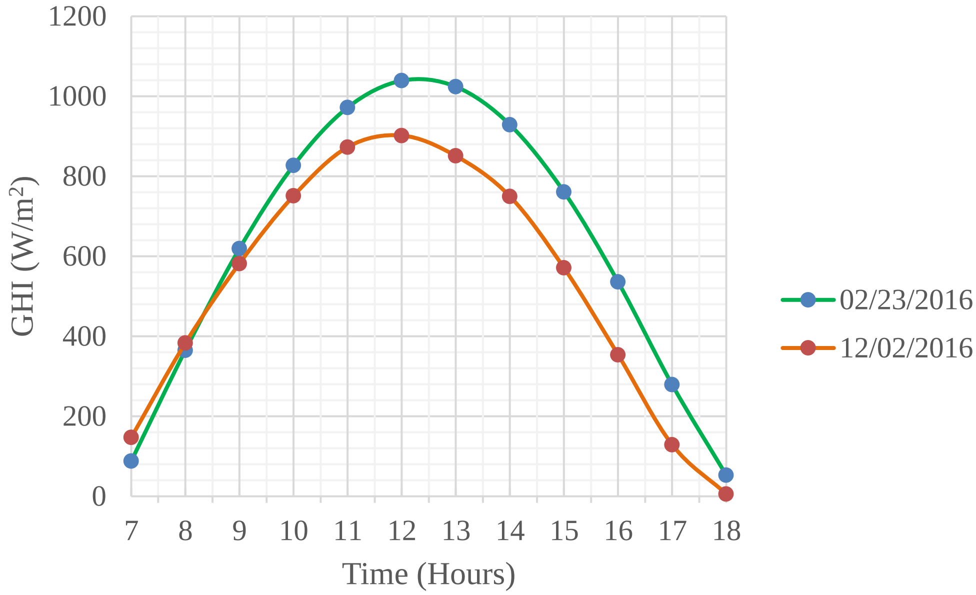
(a)



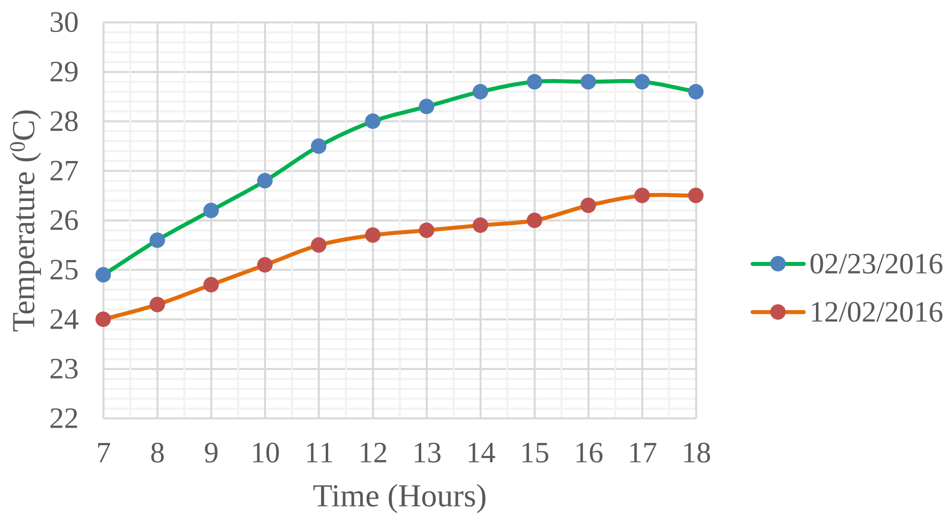
(b)

Figure 4-9: Similar variations of GHI in two different days (a) and respective variations of Temperature (b) - plotted for 10th March 2011 and 10th March 2020

Figure 4.10 shows variation of GHI in two sunny days in 2016, however in different time of the year. It is clear from the Figure 4.10; more solar radiation has been received and temperature has been remained high on 23rd February compared to that on 2nd December 2016.



(a)



(b)

Figure 4-10: Different variations of GHI in two different days (a) and respective variations of Temperature (b) - plotted for 23rd February 2016 and 02nd December 2016

The Figures 4.9 and 4.10 indicates a possible correlation between GHI and temperature. Furthermore, this has been confirmed by calculating the correlation coefficient between GHI and temperature for five consecutive years and averaging the values as below.

Table 4-3: Correlation coefficients between Temperature and GHI for different years

Year	2015	2016	2017	2018	2019
Correlation coefficient	0.54	0.54	0.53	0.45	0.49

The five-year average of correlation coefficient is 0.51 which indicates a moderate correlation between GHI and temperature. Figure 4.11 shows the variation of GHI with the temperature for the year 2015.

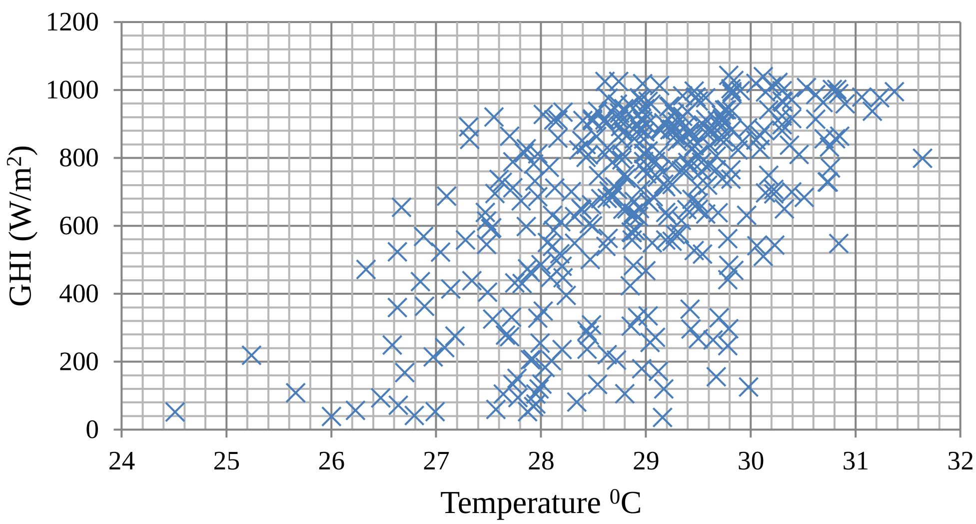
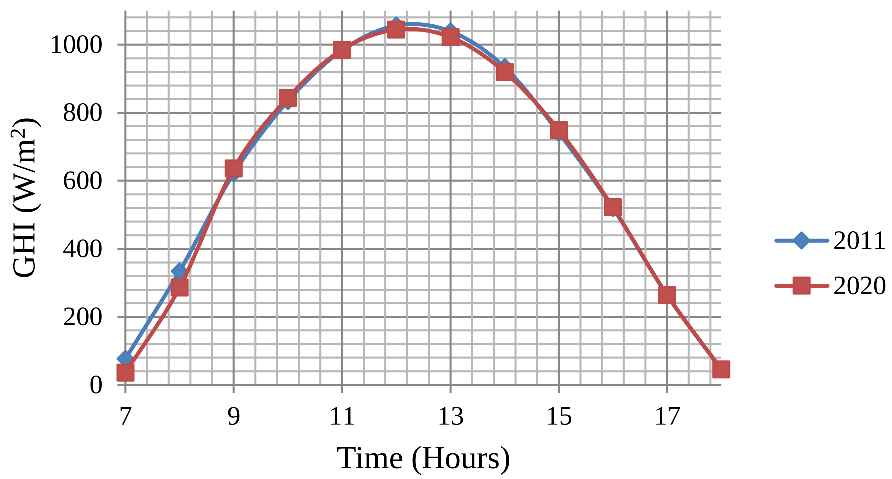


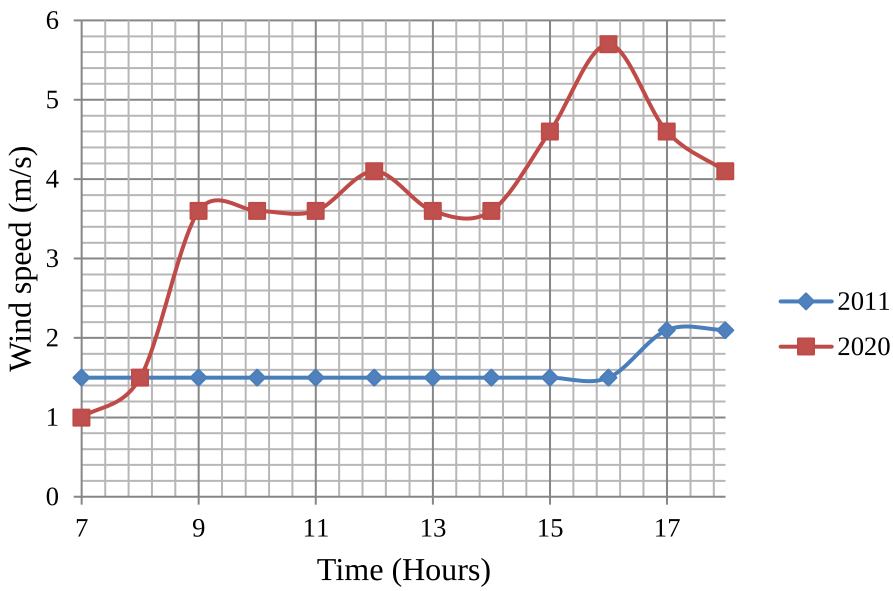
Figure 4-11: Variation of GHI with Temperature in the year 2015 from 12.00 noon to 1.00 pm

4.1.6 Variation of GHI with Wind Speed

It was difficult to identify any noticeable similarity of the variation of wind speed for two similar GHI variations shown in Figure 4.12. Wind speed was steady and around 1.5 m/s during most of the time on 10th March 2011, but it was fluctuating and relatively high in the same day in 2020.



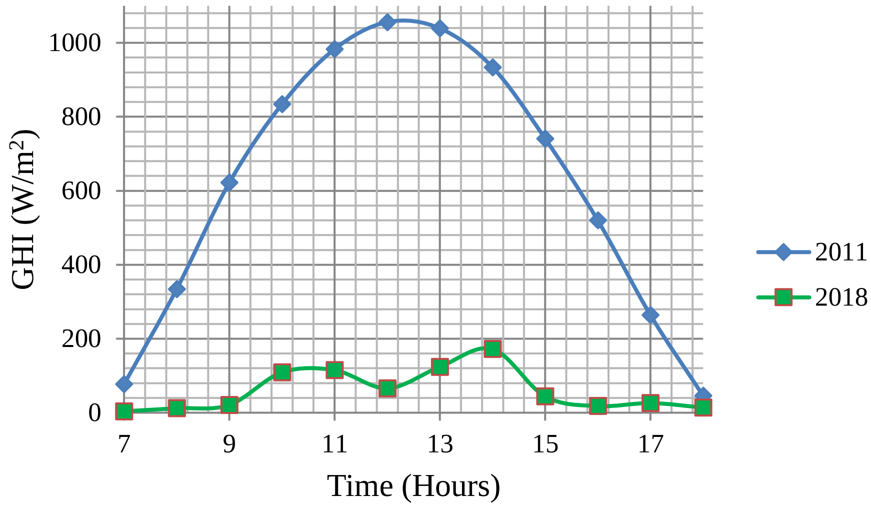
(a)



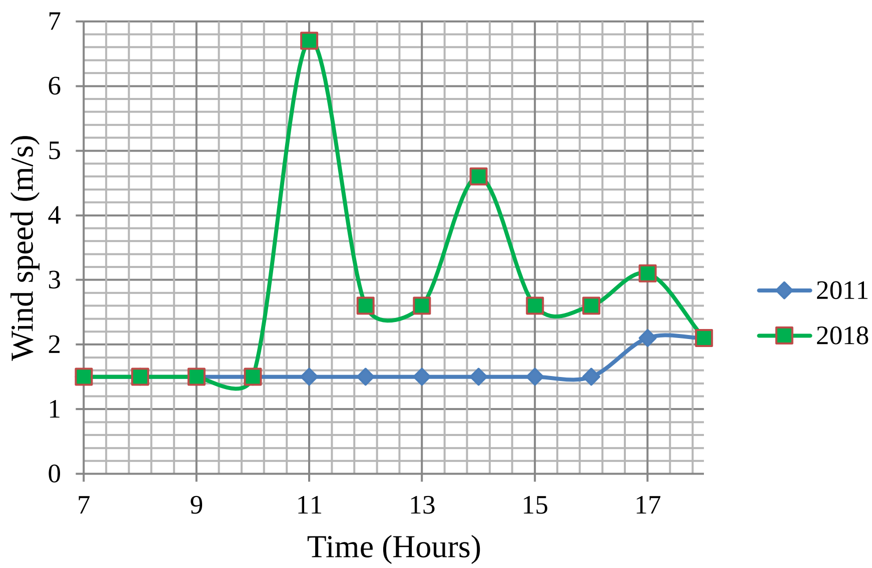
(b)

Figure 4-12: Similar variations of GHI in two different days (a) and respective variations of Wind speed (b) - plotted for 10th March 2011 and 10th March 2020

In 10th of March 2018, GHI is low throughout the daytime whereas, wind speed was observed to be relatively high, and highly fluctuating in a random way compared to the same day in 2011 as shown in Figure 4.13. On the other hand, noticeable difference in the behaviors of the wind speed in 2018 (in Figure 4.13 b) and 2020 (in Figure 4.12 b) could not be observed even though GHI variations are completely different.



(a)



(b)

Figure 4-13: Different variations of GHI in two different days (a) and respective variations of Wind speed (b) - plotted for 10th March 2011 and 10th March 2018

Correlation between GHI and Wind speed was further investigated by plotting GHI against Wind speed as in the Figure 4.14. However, no significant correlation was observed, and the calculated correlation coefficient was -0.06 which is very weak.

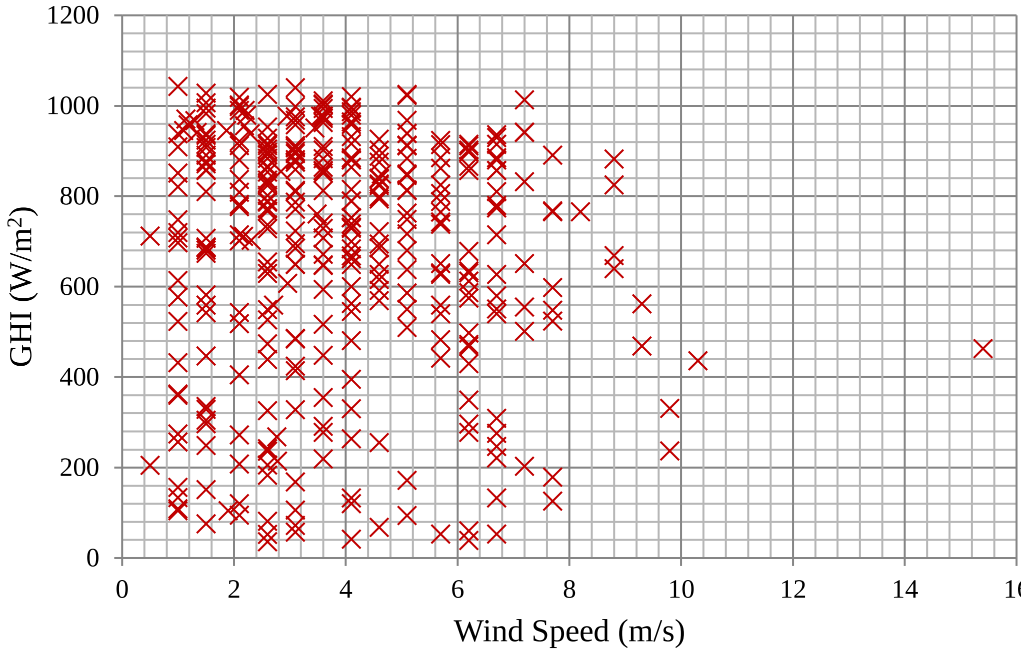


Figure 4-14: Variation of GHI with Wind speed in the year 2015 from 12.00 noon to 1.00 pm

Furthermore, the calculated correlation coefficients for five consecutive years are shown in the Table 4.4 which gives five-year averaged value of -0.05. These observations suggest there is no clear correlation with GHI and wind speed.

Table 4-4: Correlation coefficients between Wind speed and GHI for different years

Year	2015	2016	2017	2018	2019
Correlation coefficient	-0.06	0.03	-0.08	0.05	-0.18

4.1.7 Variation of GHI with Rain Fall

Figure 4.15 illustrates how rainfall affected on the typical variation of GHI during a day of the year for three years: 2011, 2016, 2019. As can be seen from Figure 4.15, on 28th October 2011, a drop in GHI and coincident rainfall in the afternoon could be observed. Similarly, received GHI level on 28th October 2019 was quite low because of the rainfall throughout the day. The correlation between GHI and rainfall was further analyzed by plotting solar insolation received from 2015 January to 2021 February for rainy days during 10.30 am to 1.30 pm against the rainfall during the same three-hour period. This has been shown in the Figure 3.21 and calculated correlation coefficient was -0.32.

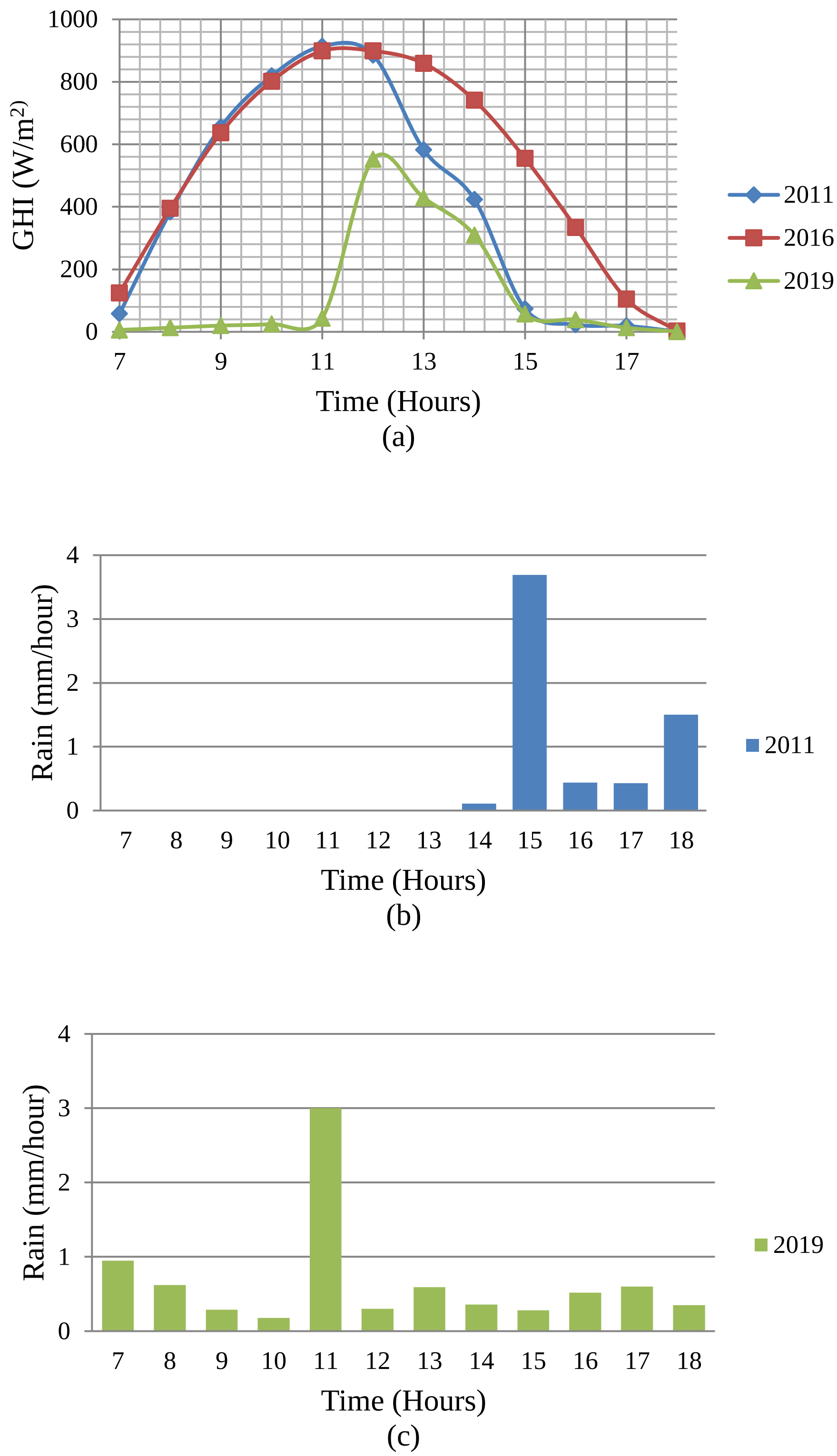


Figure 4-15: Variation of GHI - 28th October 2011,2016 and 2019 (a), Hourly Rainfall - 28th October 2011 (b), Hourly Rainfall - 28th October 2019 (c)

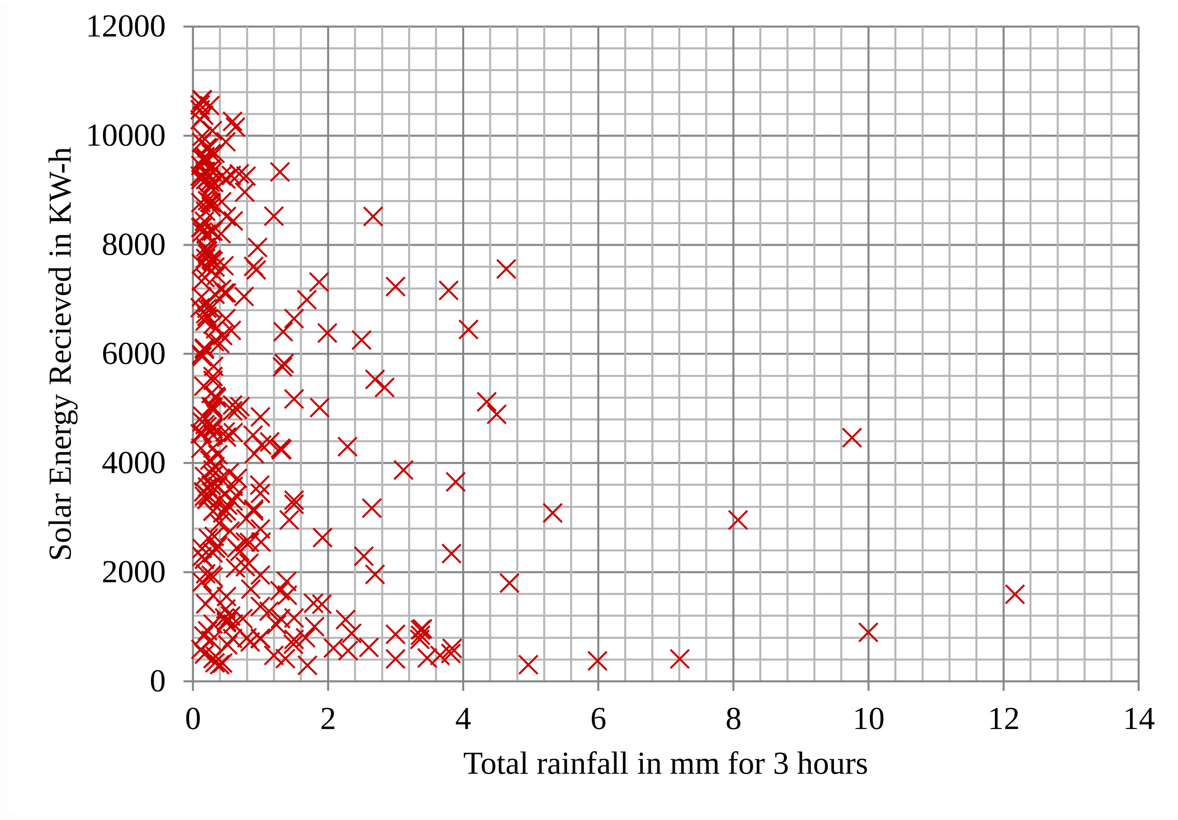


Figure 4-16: Variation of solar insolation received from 2015 January to 2021 February for rainy days during 10.30 am to 1.30 pm

From the above sensitivity analysis, it can be concluded that all the parameters discussed here other than wind speed could be considered as ANN model input parameters. Generally, correlation coefficient greater than 0.7 indicates strong correlation between the two variables. Further, coefficient between 0.5 - 0.7 indicates moderate correlation and between 0.3 - 0.5 indicates low correlation. Correlation coefficient less than 0.3 is considered as very weak or no correlation. All the weather parameters except wind speed were found to be having absolute value of correlation coefficients above 0.3. Parameters such as cloud cover, humidity and rain fall got negative values for the coefficient with GHI, which means one parameter decrease when the other increases. Summary of the analysis is given in the Table 4.5.

Table 4-5: Summary of the analyzed input weather parameters

	Parameter	Conclusion
1	Cloud cover	Remains at high, for low GHI values. Moderate correlation coefficient (r_{avg} = - 0.48)
2	Relative humidity	Remains high, for low GHI values. Moderate correlation coefficient (r_{avg} = - 0.46)
3	Temperature	Moderate correlation with GHI (r_{avg} = 0.51)
4	Wind speed	Correlation coefficient is very weak. Not considered as a model input parameter
5	Rain	GHI drops with rainy environments. Low correlation coefficient (r = - 0.33)

4.2 Training of NN Models with Different Configurations

4.2.1 Single Hidden Layer NN Model

More than 100 different models were trained by limiting the number of hidden layers to 1 and changing the hidden layer size (number of neurons) from 5 to 700, to analyze how different models are performing specially for unseen data set. As discussed in section 3.5, data from year 2007-2019 was used for training and testing of the model and data from year 2020 – 2021 was used for the validation. The Figure 4.17 depicts the mean square error (MSE) of the model predictions for the data set.

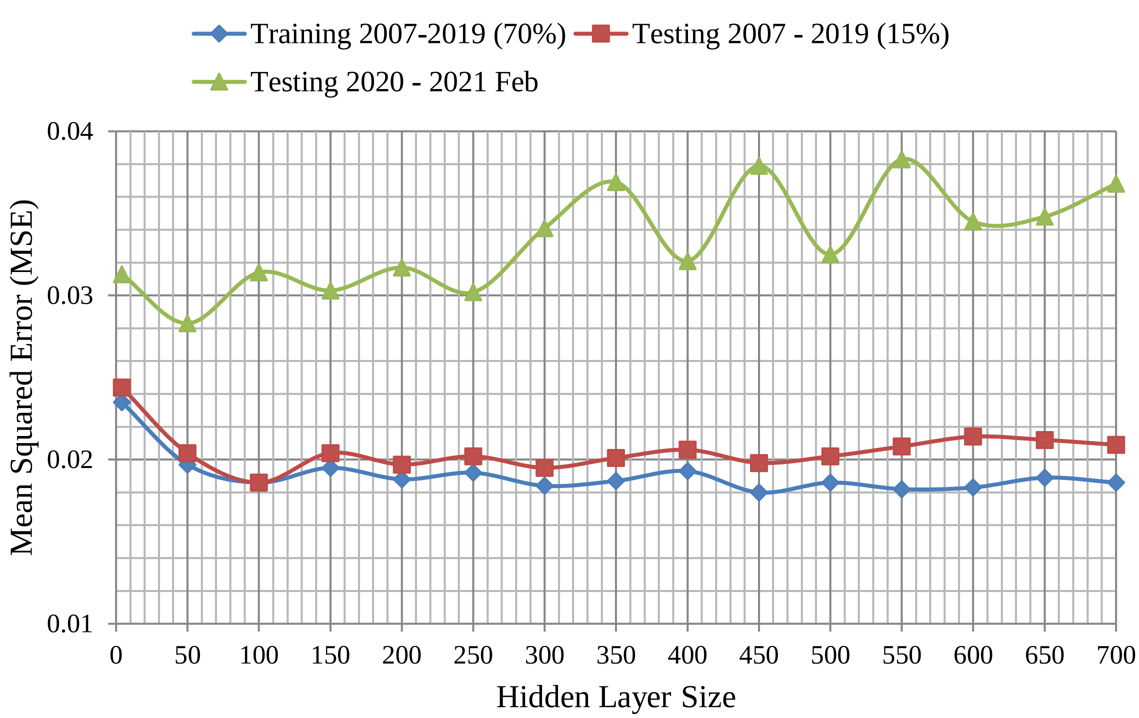


Figure 4-17: Variation of MSE of the trained model with hidden layer sizes from 5 to 700

As can be seen from Figure 4.17, mean squared error of the predicted scaled output with respect to the targeted scaled output for different hidden layer sizes have been plotted. The model is performing well for the hidden layer size around 100, for the testing data set from 2007 to 2019 which was randomly separated from the entire data set presented for the training of the model. However, MSE is observed to be always high for the unseen data set (Testing 2020 - 2021 Feb) and the best performance of the model is found for the hidden layer size around 50. Therefore, at this point, the model has good generalization capability with relatively low MSE for all plotted data sets. Figure 4.18 was plotted for a precise analyze on the hidden layer size around 50.

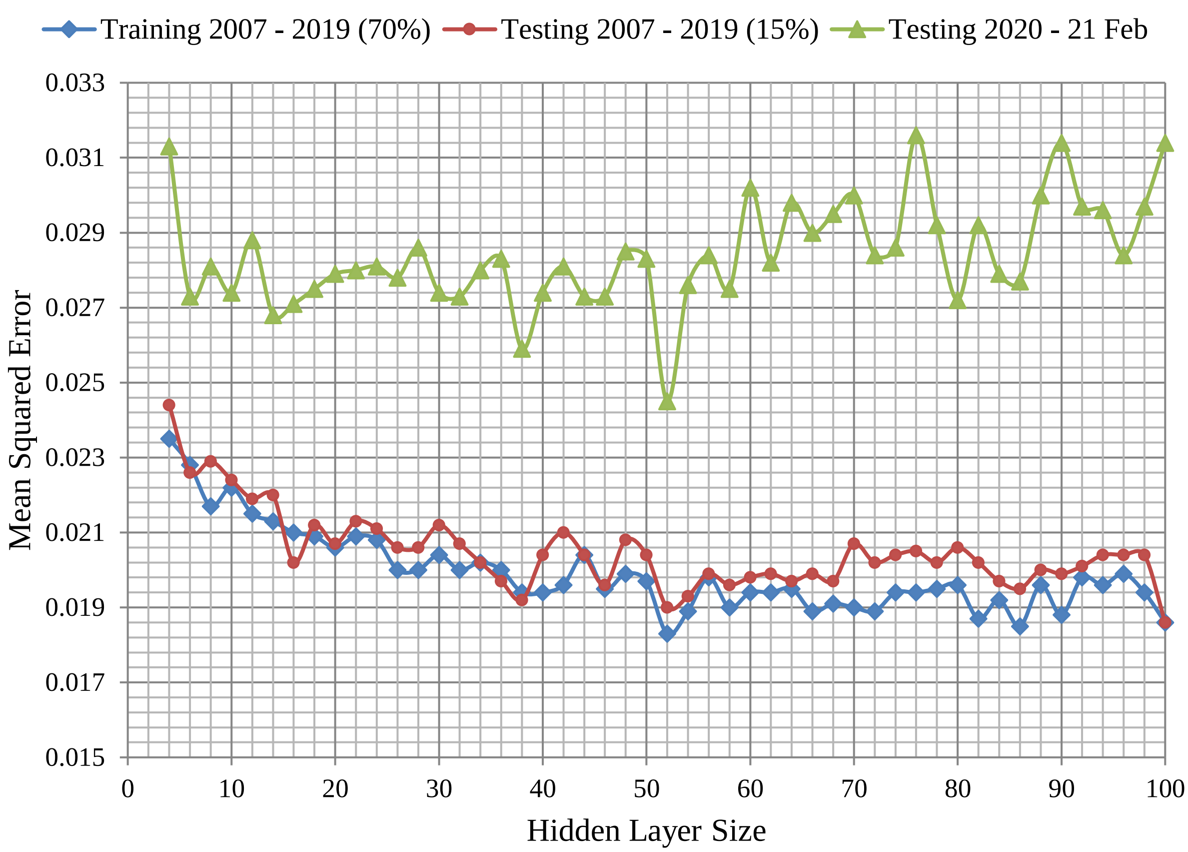


Figure 4-18: Variation of MSE of the trained model with hidden layer sizes from 4 to 100

Figure 4.18 shows that the model performs better in all the aspects when the hidden layer size is 52. It was observed that the model converges and shows the best performance after 414 epochs during the training.

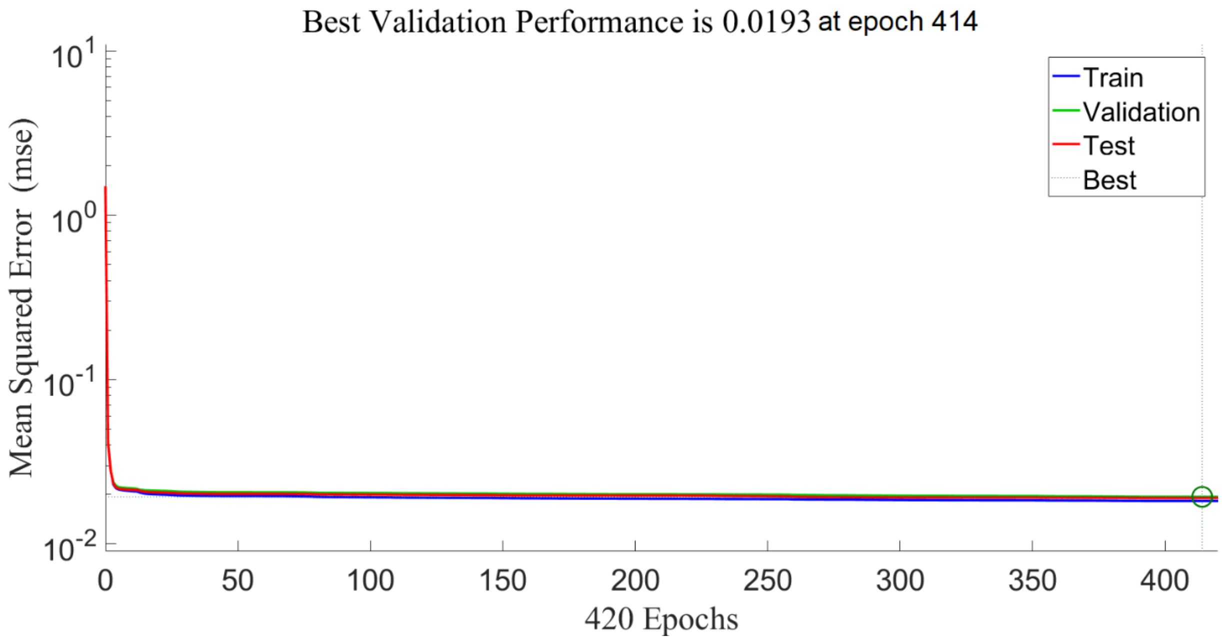


Figure 4-19: Process of convergence of mean squared error with the number of epochs

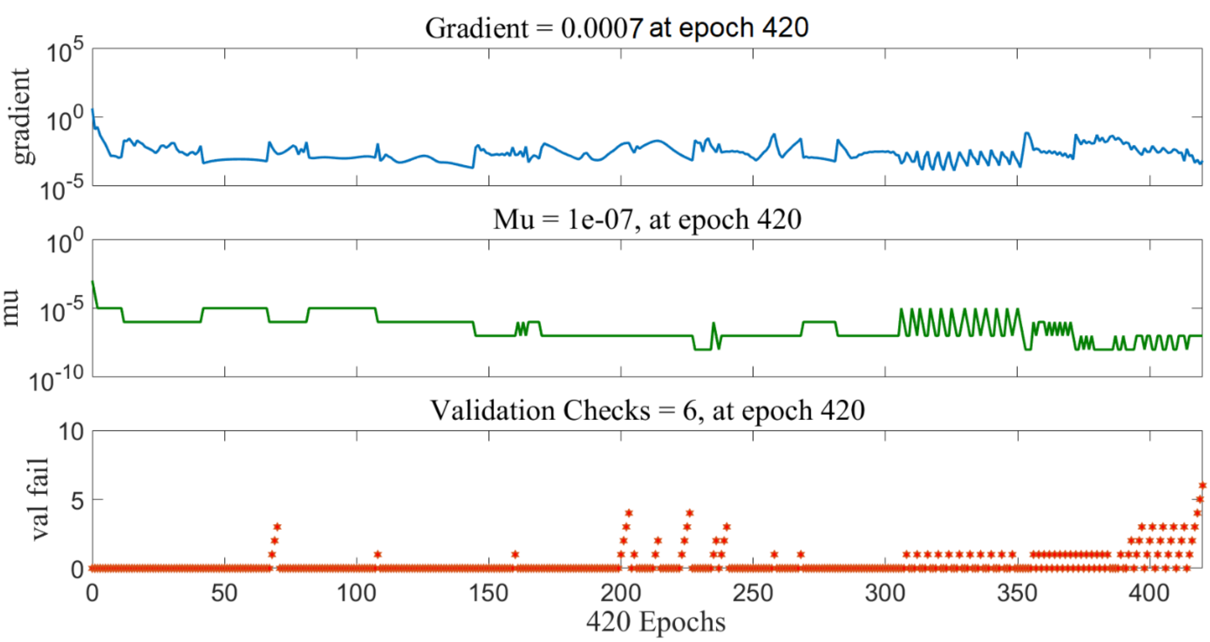


Figure 4-20: Variation of gradient, controlling parameter μ and process of validation with the number of epochs in the training process of 52 hidden layer size NN model

Number of failed validations checks during the training process was set to six in order to minimize the possibility of sticking the model in a local minima. This follows the concept of heavy ball method which was described in the section 2.2.1 and the process has been illustrated in the Figure 4.20, with the variation of control parameter μ and the updated gradient after each epoch.

4.2.2 Two Hidden Layer Models

Numerous different configurations could be considered when the model has two or more hidden layers. However, this is practically very difficult and time consuming since the evaluation of each, and every model should be done on trial-and-error basis after the training process. Therefore, a limited number of different configurations were tested for observing any significant difference in performance. Table 4.6 shows mean square error values for two hidden layer models with different configurations.

Table 4-6: Mean squared errors for two hidden layer models with different configurations

Configuration Number	**Size of hidden layer 1**	**Size of hidden layer 2**	**MSE (Training :2007 - 2019)**	**MSE (Testing :2007 - 2019)**	**MSE (Testing :2020 - 21 Feb)**
1	52	05	0.0197	0.0204	0.0276
2	52	10	0.0191	0.0194	0.0299
3	52	15	0.0187	0.0198	0.0312
4	52	20	0.0189	0.0204	0.0285
5	52	25	0.0179	0.0199	0.0345
6	52	30	0.0191	0.0196	0.0294
7	52	35	0.0197	0.0202	0.0282
8	52	40	0.0192	0.0204	0.0279
9	52	45	0.0190	0.0204	0.0298
10	52	52	0.0188	0.0207	0.0294
11	46	06	0.0197	0.0198	0.0301
12	40	12	0.0174	0.0193	0.0316
13	34	18	0.0193	0.0203	0.0283
14	28	24	0.0194	0.0203	0.0288
15	26	26	0.0184	0.0202	0.0280
16	22	30	0.0191	0.0199	0.0294
17	16	36	0.0189	0.0201	0.0295
18	10	42	0.0191	0.0205	0.0269
19	04	48	0.0208	0.0201	0.0263

Configuration 12 has the minimum MSE values for the training and testing (2007 - 2019) data sets. However, in terms of the generalization ability, configuration 19 is the best model since its MSE value for the unseen testing data set (2020 - 2021 Feb) is minimum.

However, the MSE values for the single hidden layer model with 52 hidden neurons discussed in the previous section were 0.0183 and 0.019 for training and testing (2007 - 2019) data set respectively. Also, the model's MSE for the 2020 - 2021 unseen data set was 0.0245. These numbers indicate a better performance than any of the two hidden layer models listed in the table 4.2. Based on this, the 52 hidden layer size trained model was adopted for the use in the forecasting model.

4.2.3 Performance of the 52 Hidden Layer Size NN Model

As described in the previous section, the 52 hidden layer size model has the minimum MSE values for all the data sets. However, illustrating the behavior of the model in several different ways helps to recognize the performance of the model under different situations. Hence, target - output plots and error histograms were obtained for the available data sets separately. Figure 4.21 shows the target - output performance of the trained NN model together with regression value (R) for training, validation and testing data sets which were presented to the model at the training stage. In order to illustrate the performance for unseen data, 2020 - 21 data set was presented and the same target - output graph was plotted as shown in the Figure 4.22. Error histograms for training, validation and testing data sets are shown in the Figure 4.23. The hourly forecasted GHI values could be visualized after un-scaling outputs by reversing the equation 3.2 as discussed in the section 3.4.

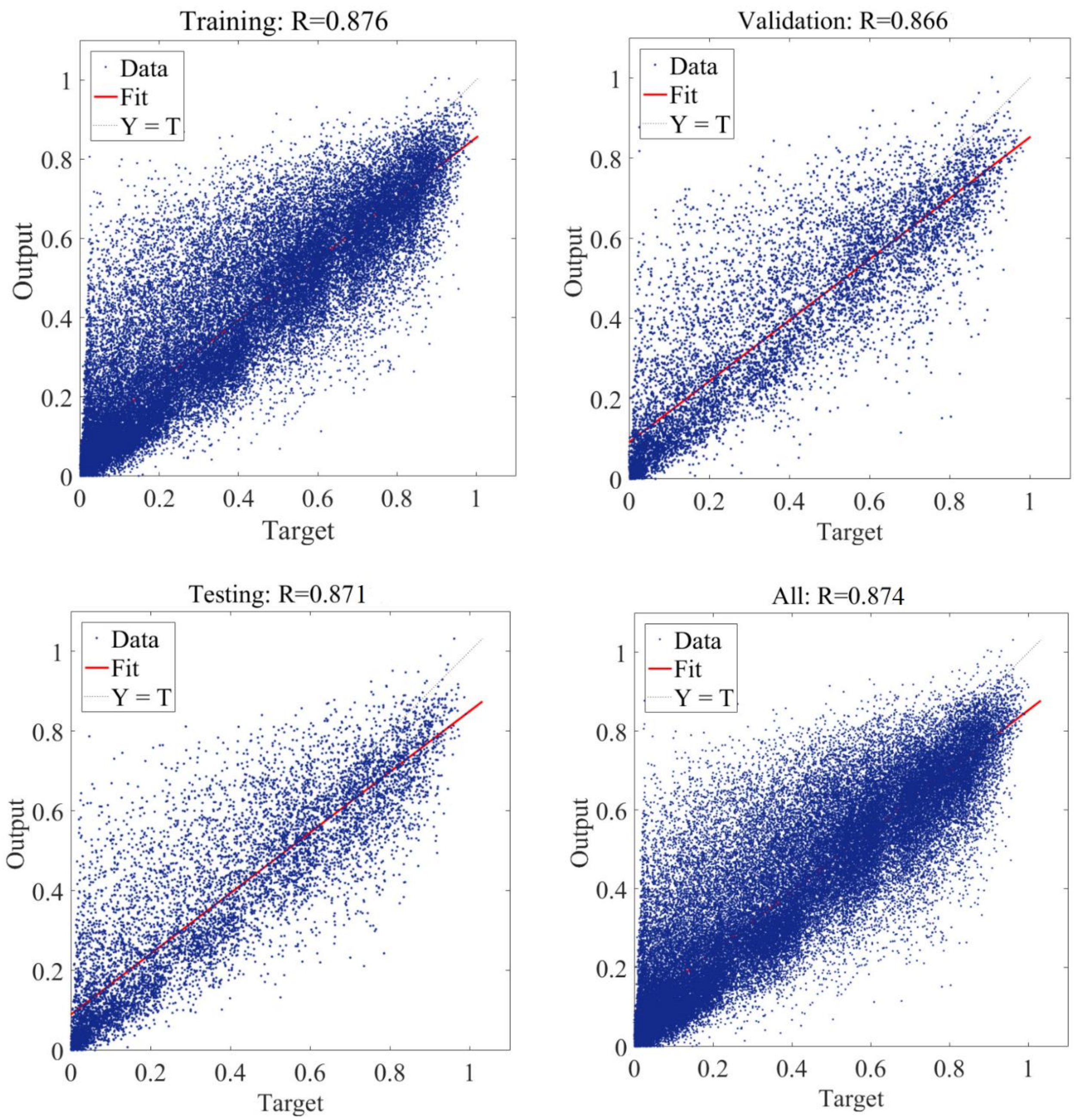


Figure 4-21: Target -output plots with R values obtained for the NN model after training

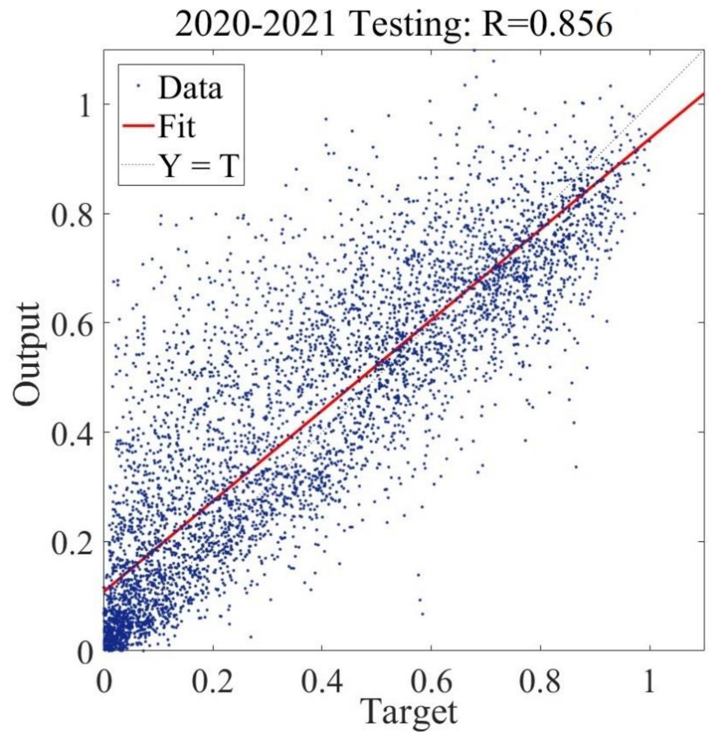
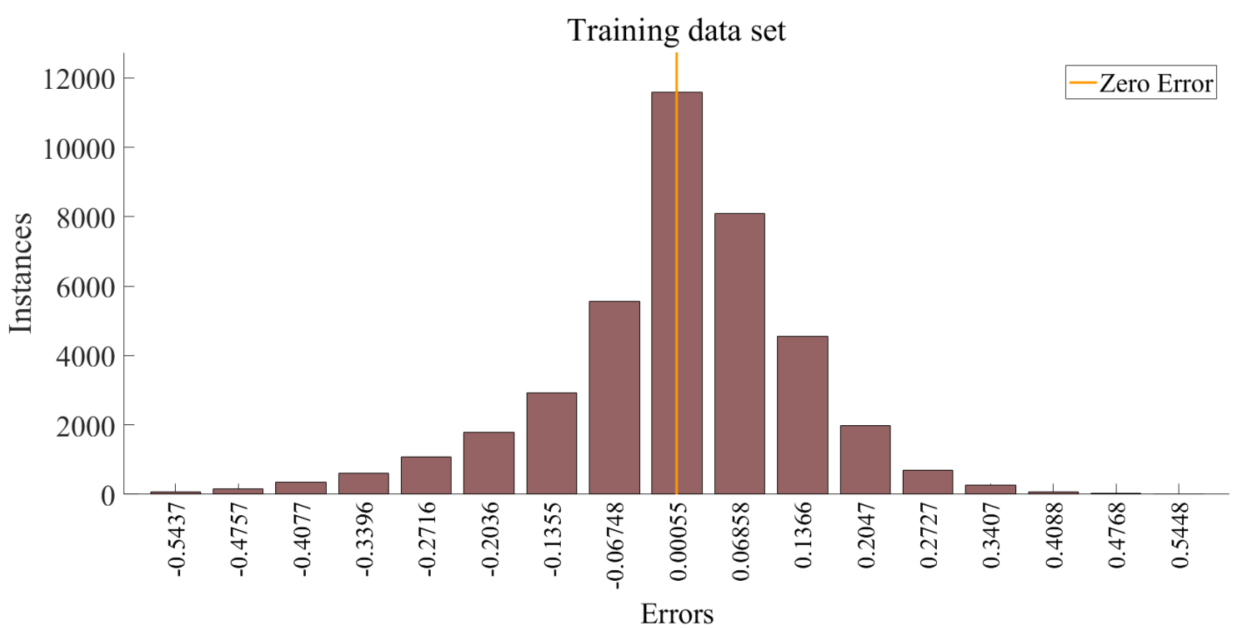
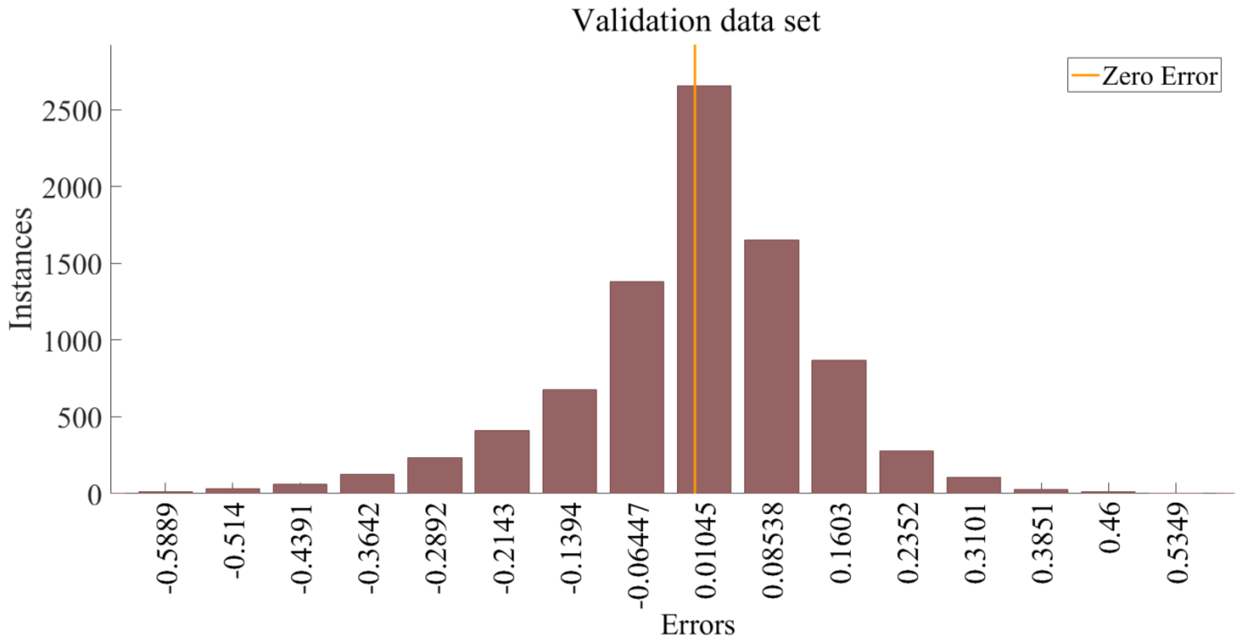


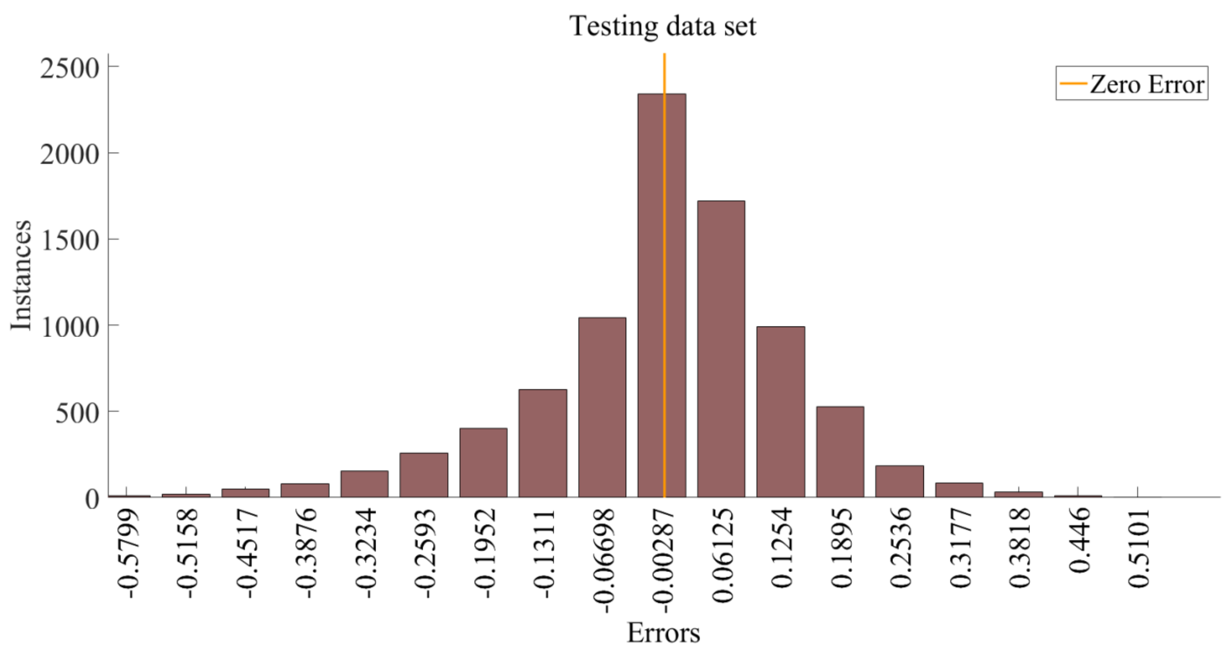
Figure 4-22: Target -output plot with R value for 2020 -21 data set



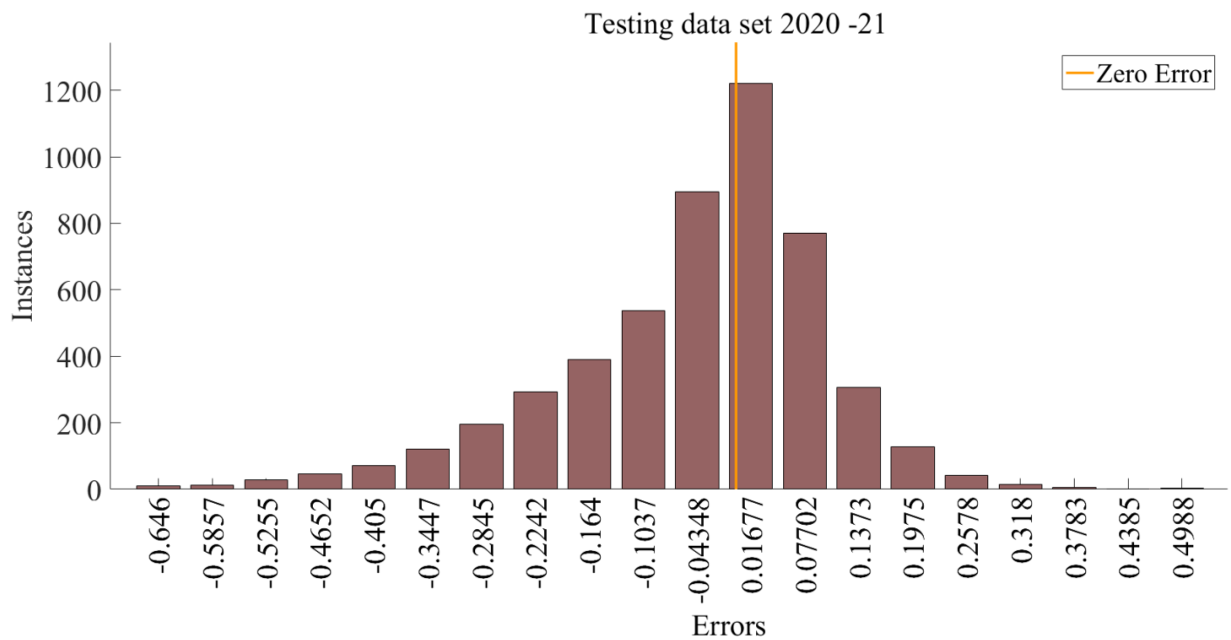
(a)


Validation data set
Zero Error
2500
2000
1500
1000
500
0
Instances
-0.5889
-0.514
-0.4391
-0.3642
-0.2892
-0.2143
-0.1394
-0.06447
0.01045
0.08538
0.1603
0.2352
0.3101
0.3851
0.46
0.5349
Errors

(b)


Testing data set
Zero Error
2500
2000
1500
1000
500
0
Instances
-0.5799
-0.5158
-0.4517
-0.3876
-0.3234
-0.2593
-0.1952
-0.1311
-0.06698
-0.00287
0.06125
0.1254
0.1895
0.2536
0.3177
0.3818
0.446
0.5101
Errors

(c)



(d)

Figure 4-23: Error histograms for (a) Training data set; (b) Validation data set; (c) Testing data set; (d) Testing data set 2020 - 21

The Figure 4.24 illustrates how the actual GHI values and forecasted values are changing with the time. The graphs were plotted for randomly selected consecutive 7 days (From 04/05/2020 to 10/05/2020).

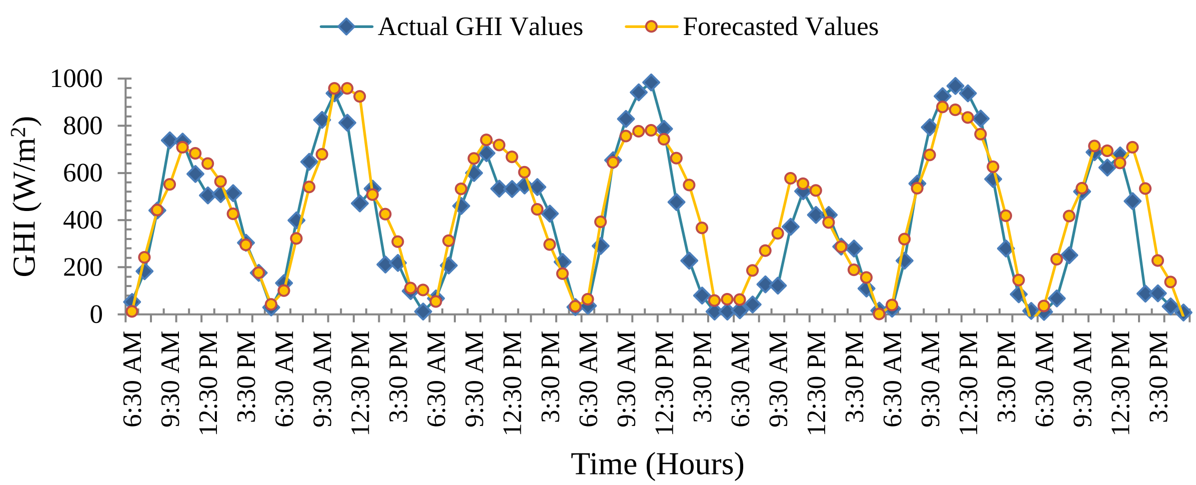


Figure 4-24: Variation of actual and forecasted GHI values for a period of one week (From 04/05/2020 to 10/05/2020)

4.3 Shading Loss Simulation

Shading losses associated in a PV system are highly site specific and they change throughout the year with the seasonal changes of sun path and/or with any change of existing shading sources. In order to find out hourly shading losses, a 3D representation of the plant was constructed with the aid of topological information collected through site visits, architectural drawings, and Google map. Simulation was run in PVsyst which is a software package used for study, sizing, and data analysis of photovoltaic systems.

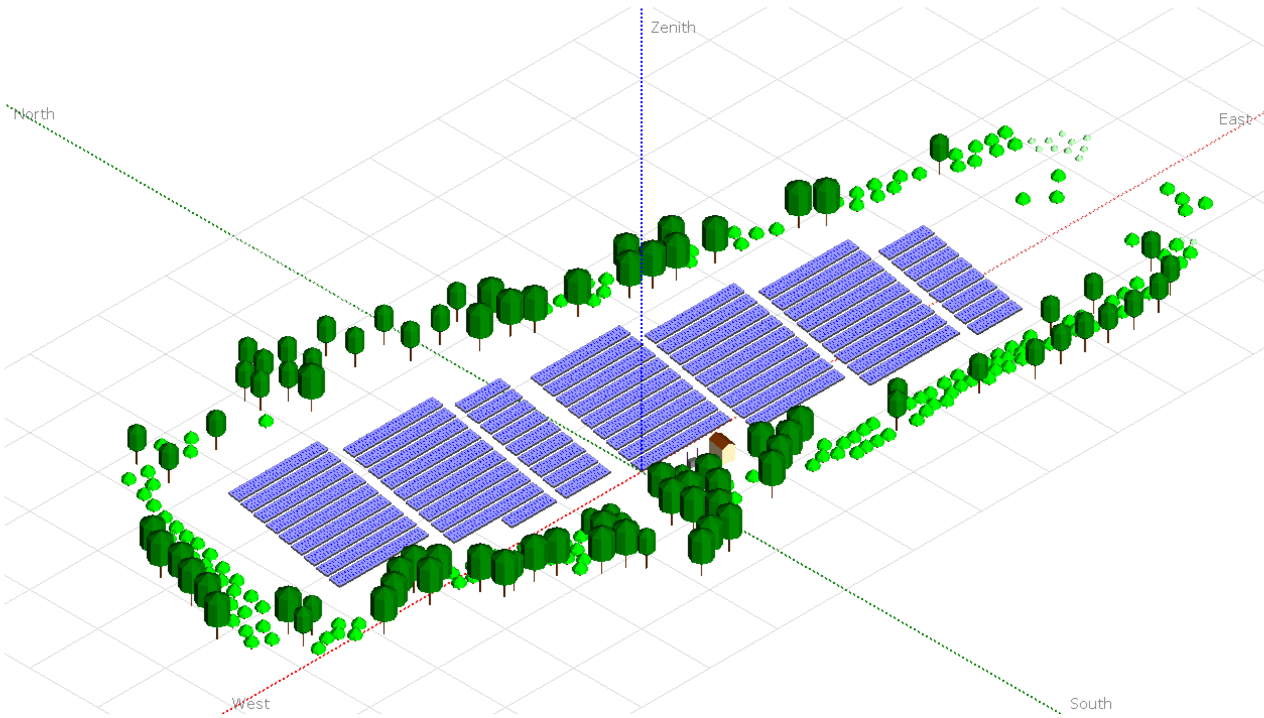


Figure 4-25: 3D representation of the plant (Modeled in PVsyst)

Cylindrical sun path chart shown in Figure 4.26 was generated through the 3D shading simulation and the percentage shading losses at the middle of each hour were used for tabulating hourly shading losses for the specified dates in the chart. The tabulated hourly shading losses were considered as average hourly shading losses for a given month.

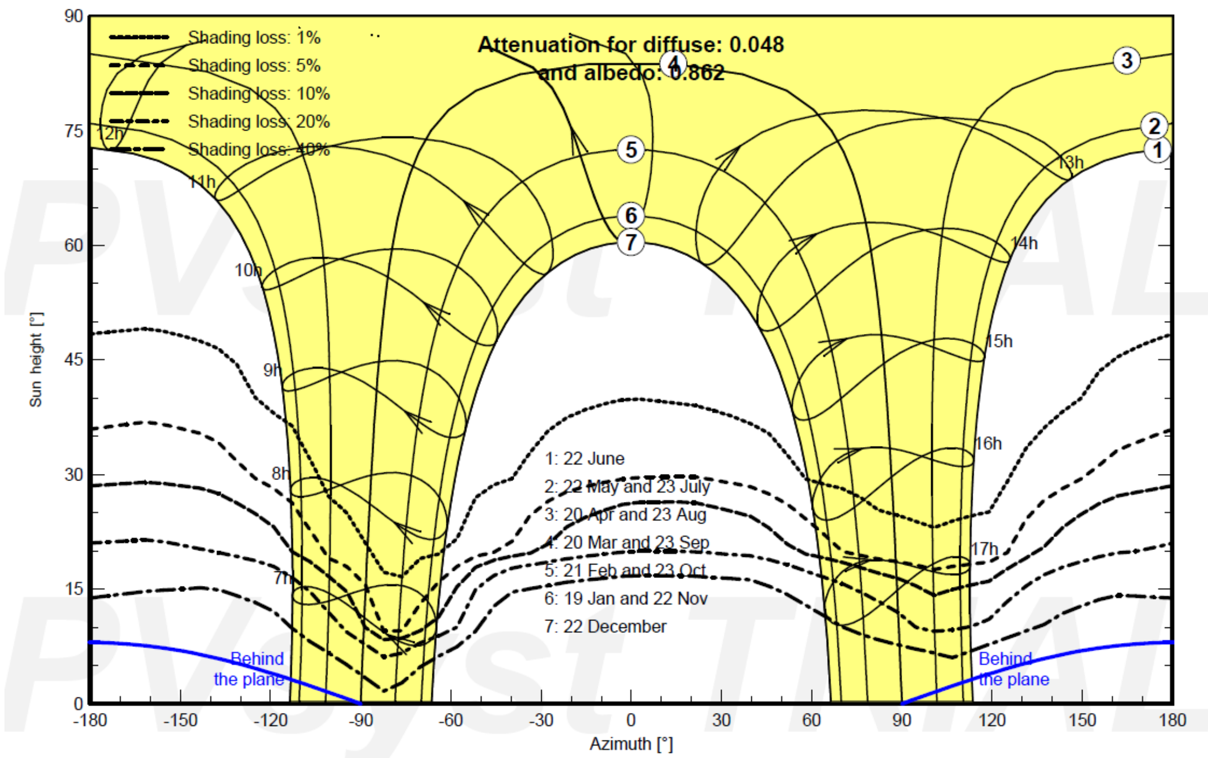


Figure 4-26: Cylindrical sun path chart for the plant location with iso shading curves

Table 4-7: Hourly shading losses for the specified dates in the cylindrical sun path diagram

Month	**Percentage shading losses in each hour**										
	6.30-7.30	**7.30 - 8.30**	**8.30 - 9.30**	**9.30 - 10.30**	**10.30-11.30**	**11.30-12.30**	**12.30-13.30**	**13.30-14.30**	**14.30-15.30**	**15.30-16.30**	**16.30-17.30**
January	21	0	0	0	0	0	0	0	0	1	20
February	10	0	0	0	0	0	0	0	0	0	7
March	10	0	0	0	0	0	0	0	0	0	5
April	15	0	0	0	0	0	0	0	0	0	6
May	20	3	0	0	0	0	0	0	0	0	7
June	25	4	0	0	0	0	0	0	0	0	5
July	25	7	0	0	0	0	0	0	0	0	4
August	17	1	0	0	0	0	0	0	0	0	9
September	5	0	0	0	0	0	0	0	0	0	13
October	2	0	0	0	0	0	0	0	0	2	32
November	10	0	0	0	0	0	0	0	0	4	40
December	20	0	0	0	0	0	0	0	0	3	36

4.4 Validation Results of the Forecasting Model

An example for a one-day hourly power forecast of the model is shown in the Figure 4.27. As discussed in the section 3.7, the hourly power forecasts were converted to daily power generations and compared with the actual daily generation data, in order to validate the model.

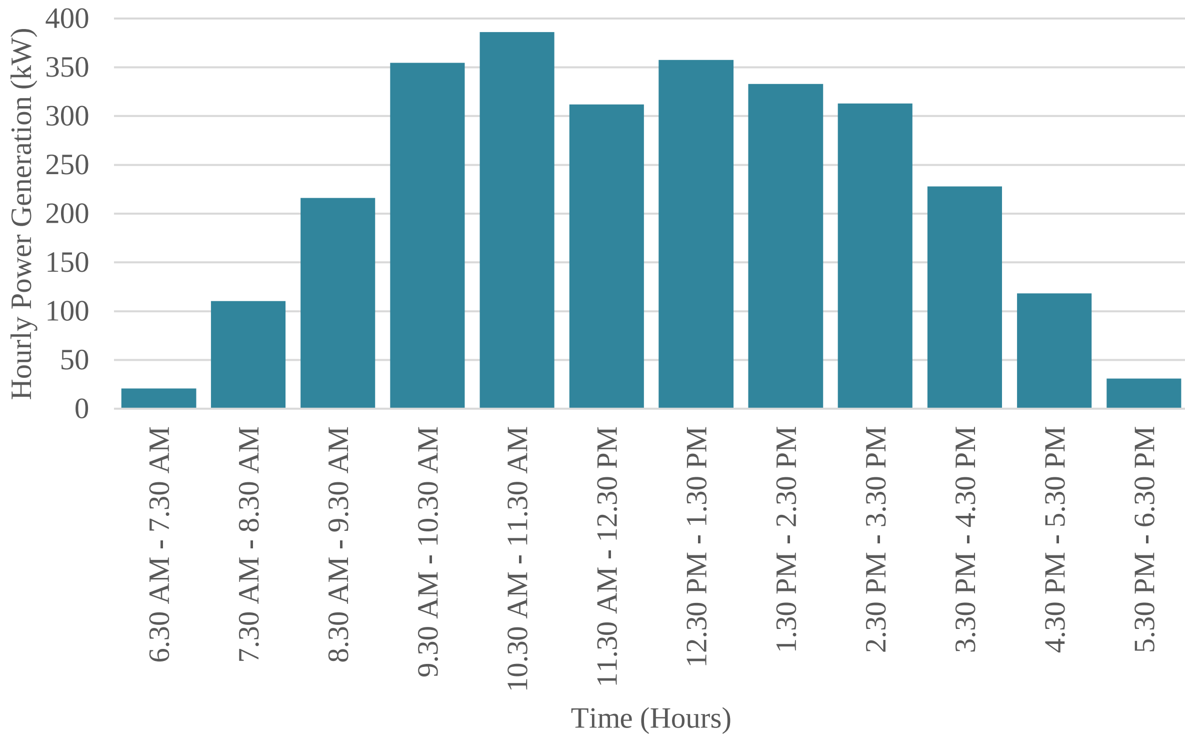


Figure 4-27: A one day hourly power forecast of the model (Plotted for the date 15th January 2021)

The comparison of daily actual generation and the forecasted generation from the developed model for 14 months period from 2020 January is shown in the Figure 4.28. The mean absolute percentage error (MAPE) for the model was computed using equation 2.64 and it was observed to be 14%. Further, standard deviation for the error terms was observed to be 633 kW h/day. Figure 4.29 shows the variation of model output values with target or actual daily energy generations. Error histogram for the entire model shown in the Figure 4.30 demonstrates how the error values have been distributed.

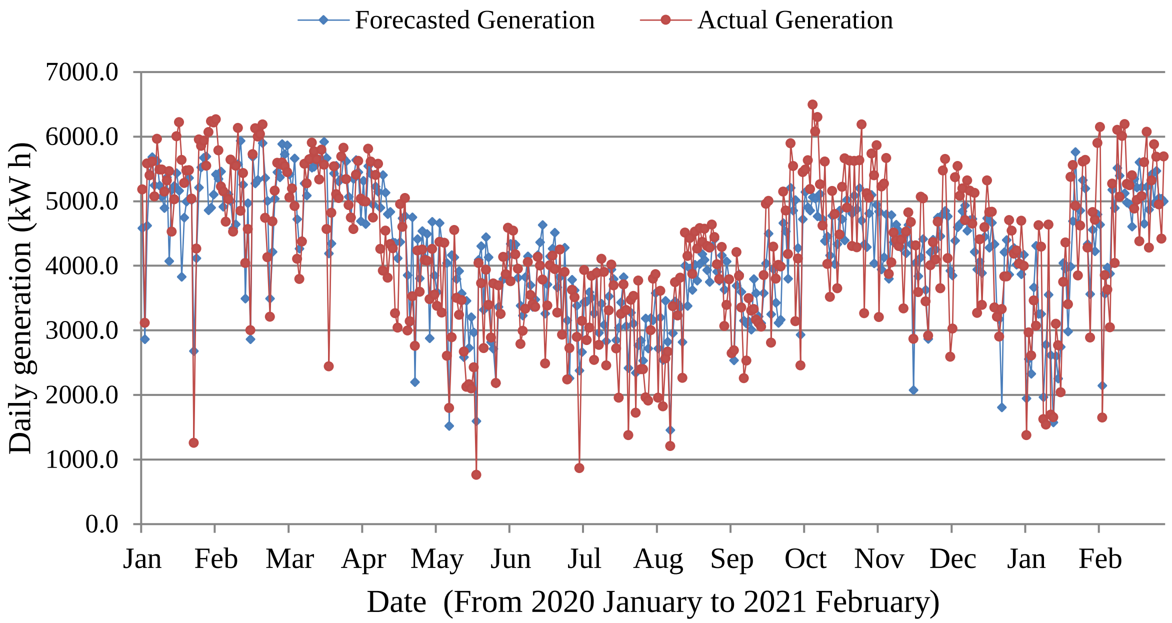


Figure 4-28: Comparison of actual energy generation of the PV plant for the period of January 2020 to February 2021 with model predictions

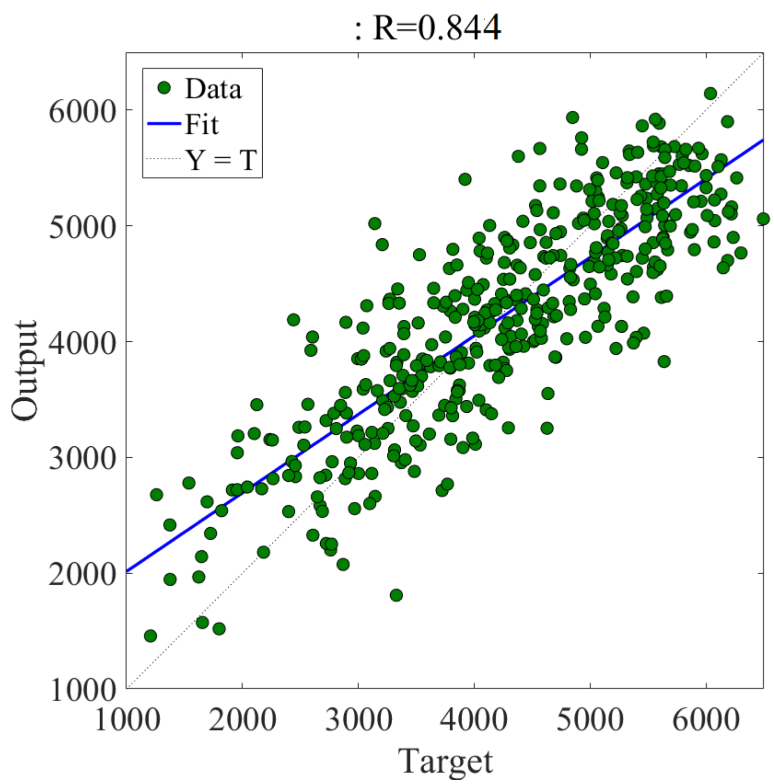


Figure 4-29: Target - output plot with R value for the entire model (2020 - 21 data set)

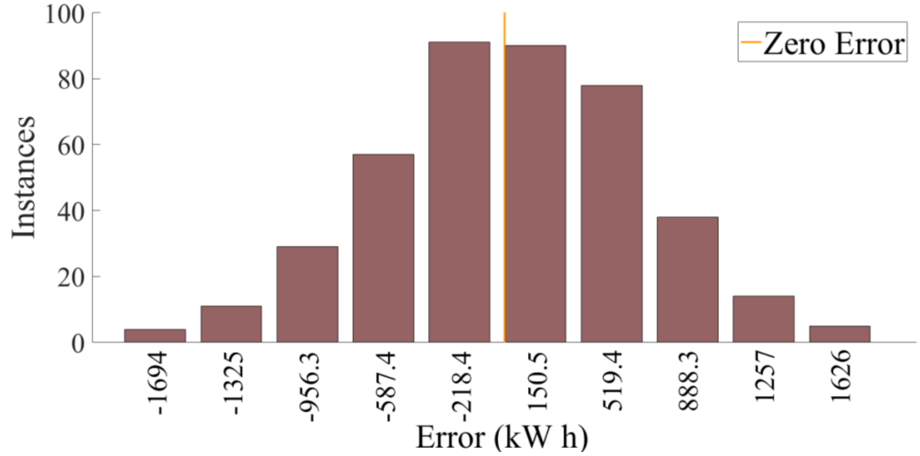


Figure 4-30: Error histogram for the entire model (2020 - 21data set)

In order to observe any noticeable autocorrelation of residuals, Durbin-Watson Statistic (D) for the errors was computed using the equation 2.74. The value for D was observed to be 1.7 which indicate very weak autocorrelation. Further, error terms of each day for the period of 2020 - 21 February were plotted against time axis, to identify the randomness of errors or any seasonal pattern of errors and the plot is shown in the Figure 4.31.

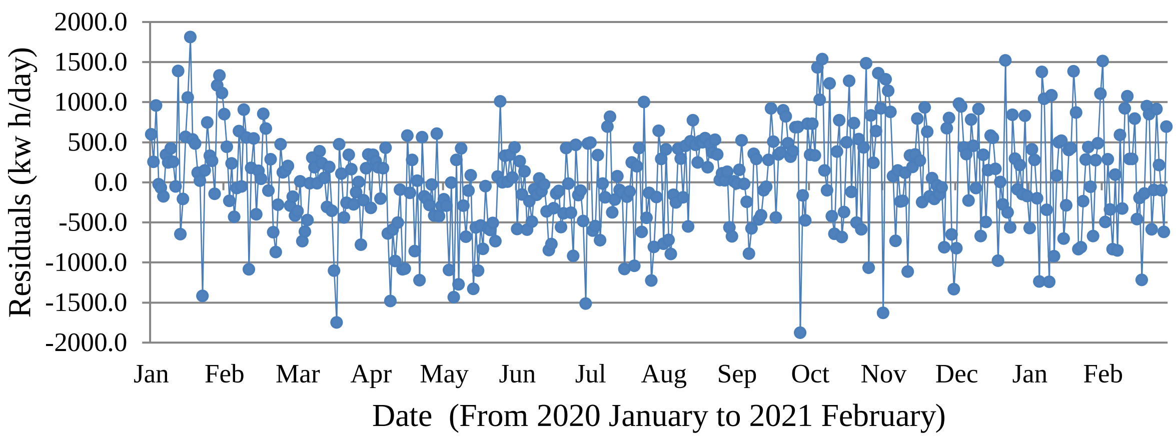


Figure 4-31: Daily residual distribution throughout the 14 months period

4.5 Further Discussion of the Model Performance

The model has a mean forecasting accuracy of 86% with 12% standard deviation, based on the MAPE calculated in the previous section. However, this was observed to be slightly varying from one month to the other as shown in the Table 4.8. The least accuracies could be found in the months May, July, and November whereas the accuracies in February and March are very high.

Table 4-8: Monthly mean forecasting accuracies of the model

Month	Mean Forecasting accuracy (%)	Standard deviation (%)
January	87	8
February	92	6
March	92	13
April	87	11
May	79	18
June	84	9
July	81	19
August	87	10
September	87	12
October	86	8
November	84	11
December	87	12

Figure 4.32 which gives the dependence of model forecast with cloud cover and rain fall, gives a good explanation on why the model performance is bit vulnerable in the months May, July and November. In the months May and July, the cloud percentage is observed to be high throughout most of the time and a regular daytime rainfall is observed in November. Therefore, a noticeable accuracy drop in the model could be observed especially when the cloud percentage exceeds 60%. This observation is further confirmed by the scatter plot shown in the Figure 4.33, which depicts the correlation of model accuracy and cloud cover percentage. It illustrates that the reduction of frequency which hits the accuracy more than 90%, when the cloud cover is higher than 60%.

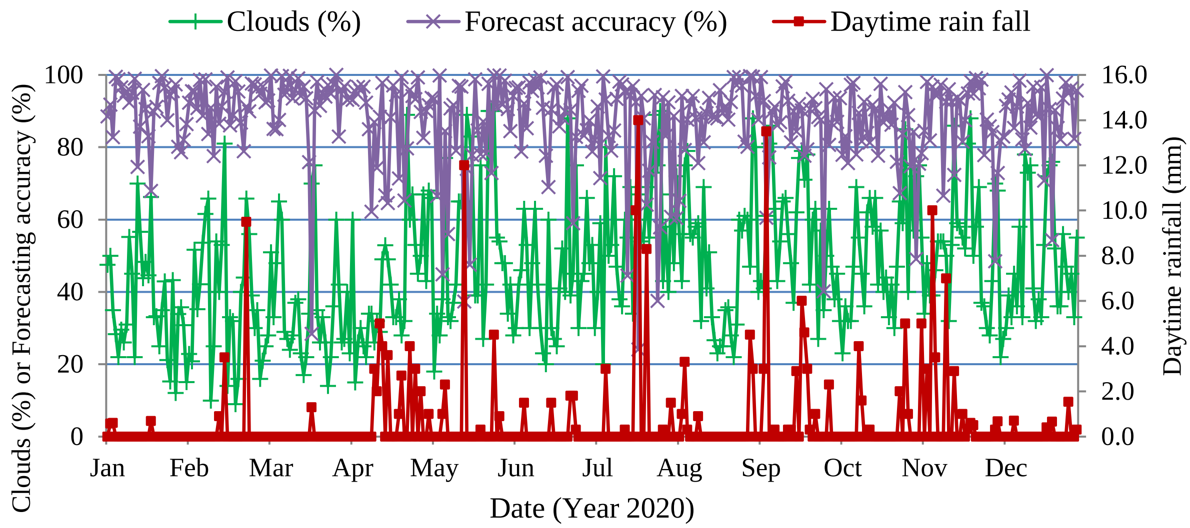


Figure 4-32: Dependence of the forecasting accuracy with cloud cover and rain fall

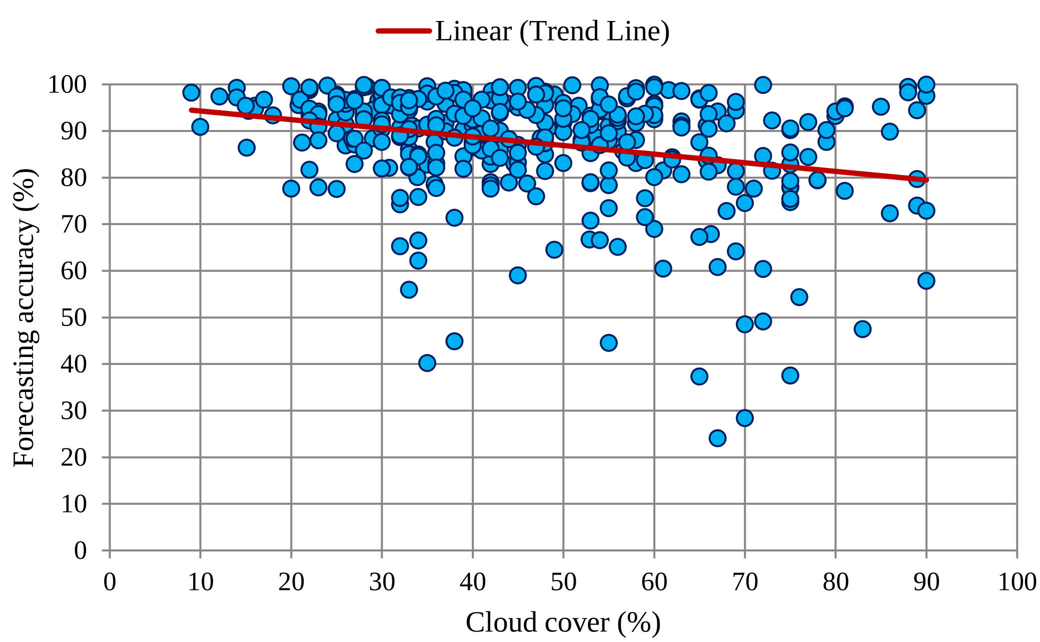


Figure 4-33: Scatter plot - Cloud cover Vs Forecasting accuracy

However, the forecasting accuracy drops below 70% rarely even in high cloud percentage conditions. According to 2020 records, the model has produced forecasts with less than 70% accuracy only in 28 times from all daily forecasts. So, it is a chance of less than 8% to drop the accuracy of the forecast below 70%.

As a slightly different approach, forecasting accuracies of the model for different daily energy generations of the plant have been plotted in the Figure 4.34. High cloud cover always reduces the incident solar radiation on PV modules and leads to low generation of electricity in the plant. Figure 4.34 illustrates again the drop of forecasting accuracy for low energy generations probably because of low incident solar radiation.

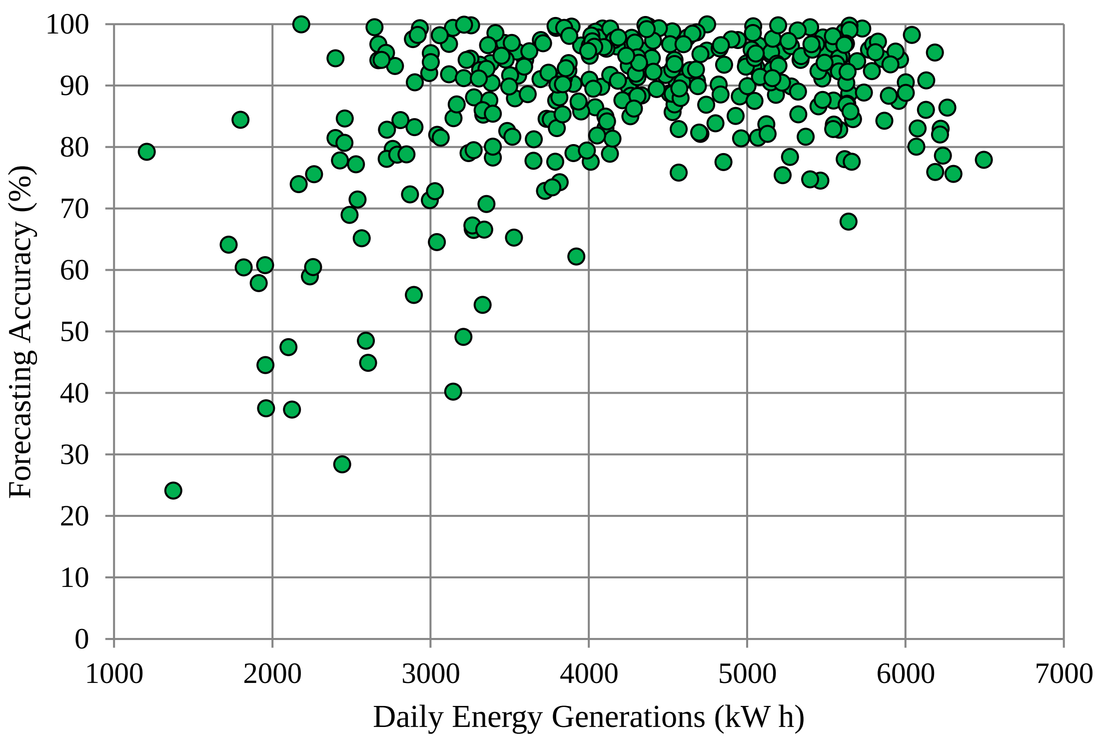


Figure 4-34: Daily Energy Generations Vs Forecasting Accuracy

However, it cannot be guaranteed that high cloud cover or rainy conditions are the major reason to reduce forecasting accuracy all the time. Sudden faults or breakdown occurring in some PV arrays or modules result to reduce power generation and lead to increase forecasting error. Six such incidents were recognized through the plant maintenance records for the year 2020 and they were not considered for the analysis. However, there is a possibility for these incidences going unnoticed or unrecorded. Another reason for the reduced forecasting accuracy in high cloud condition or at low power generating times is the various PV losses considered for the final calculation. Most of the loss factors were considered as flat values, however these values might change in extreme conditions. For example, the inverter loss was set to a flat value of 2.5% for all radiation levels. However, this might increase even beyond 5% in low irradiance levels as shown in the efficiency curves in the Figure 4.35 which has been published by the manufacturer.

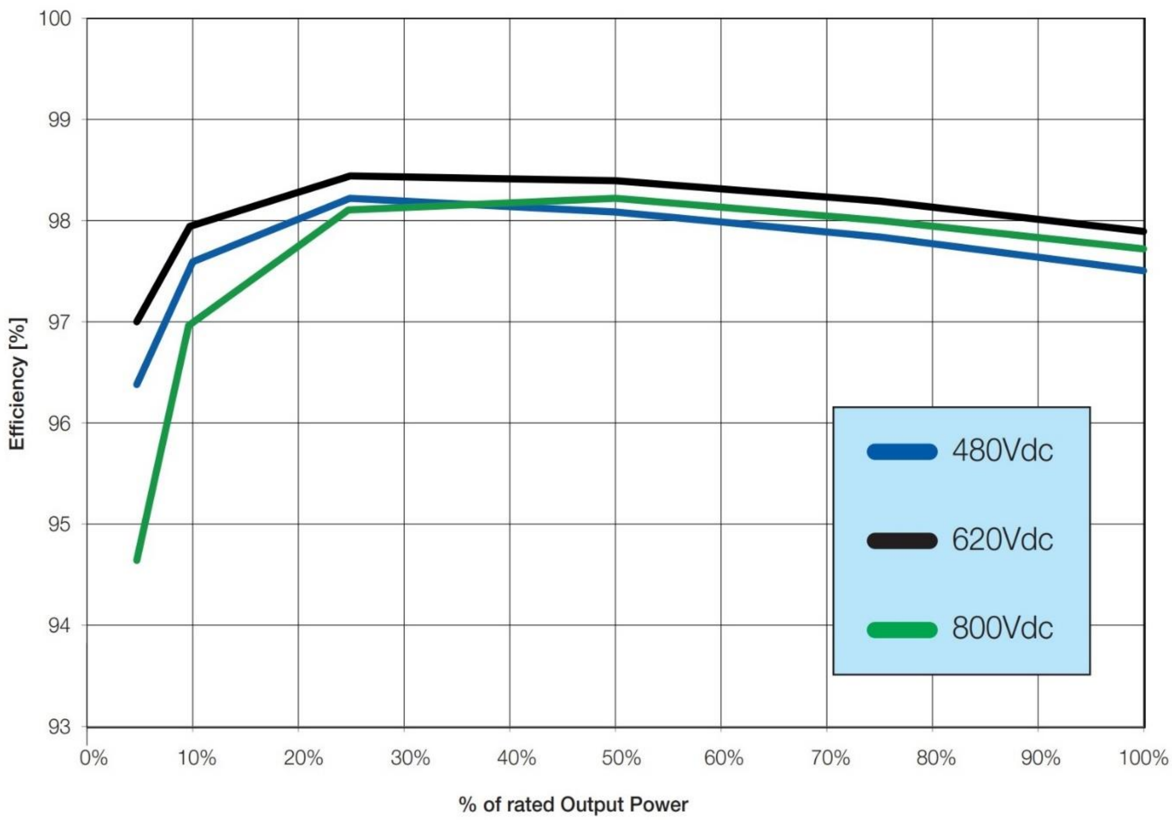


Figure 4-35: Efficiency curves - PVS-100-TL inverters [96]

Figure 4.36 shows a comparison of forecasting accuracy with energy conversion efficiency of the entire plant. Here, energy conversion efficiency was computed by taking the ratio of actual daily energy generation to daily solar energy insolation on the PV panels. Variation of energy conversion efficiency demonstrates the non-linear relationship between the incident solar radiation and PV power generated. As non-linearity increases with extreme conditions, the forecasting accuracy could be decreased. For example, a steady energy conversion rate could be observed from January to somewhere in the mid of April, other than few spikes at times. The forecasting accuracy remains high most of the time and Figure 4.32 confirms low cloud cover levels and low rainfall during this period.

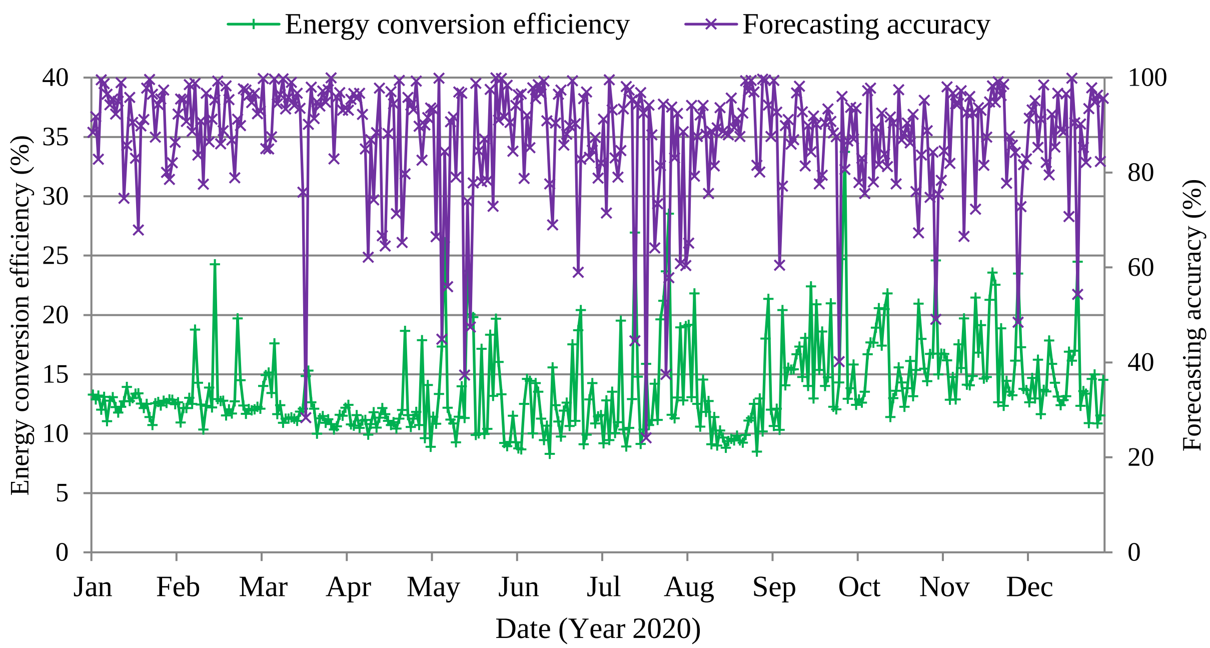


Figure 4-36: Comparison of plant energy conversion efficiency with forecasting accuracy

On the other hand, relatively high fluctuations of conversion efficiency and forecasting accuracy could be observed in the months of May and July. This confirms by the standard deviation percentages 4.71 in March and 5.21 in July listed in the Table 4.9. Even though a high standard deviation has been reported for September, it was not considered here, assuming the panels cleaning event conducted in the month has been affected significantly to the conversion efficiency. Another important observation of energy conversion efficiency curve in the Figure 4.36 is that the slight negative trend of the curve towards the end of August, starting from January. One reason for this could be the degradation of the modules with time. However, it should be very small trend since the manufacturer's specification shows 0.6% degradation per year. The second reason could be the soil and dirt accumulation on the PV modules with the time. This further confirms with the quick boost of conversion efficiency which is observable in the first week of September. Because the efficiency improvement is exactly coinciding with the major solar modules cleaning event conducted in the plant during the same week. A linear soiling loss rate was introduced in the model, in order to reduce the forecasting error as described in section 3.6 in the previous chapter. However, the rate of soiling loss could be affected by various other factors. Rainfall is a major factor among those, because it cleans PV modules. Some research revealed that rainfall of 5 mm or more is sufficient to clean the modules effectively [50].

Table 4-9: Monthly mean standard deviation of the energy conversion efficiency

Month	**SD (%)**
January	0.69
February	2.88
March	1.74
April	2.01
May	4.71
June	3.08
July	5.21
August	3.50
September	4.94
October	2.88
November	3.39
December	3.10

CHAPTER 5

CONCLUSIONS

The research focused on formulating a one day ahead energy output forecasting mechanism for a commercial or utility scale solar PV plant using readily available data sources. Energy generation or power output of a solar PV plant primarily depends on incident solar radiation on individual PV modules. Therefore, the forecasting accuracy of solar power output mainly depends on how accurately the amount of incident solar radiation could be predicted. Since the amount of incident solar radiation changes invariably with different weather conditions, solar radiation forecasting models always associated with weather parameters.

In the process of formulating the model, collection of historical weather and solar radiation data for adequate span of years was a major challenge. Weather and solar radiation APIs were a very good solution for this, in terms of availability for the desired location, available span of years, granularity and accuracy. Among the weather parameters, cloud cover, rainfall, temperature and relative humidity were observed to be having a considerable correlation with incident solar radiation on PV panels. An artificial neural network model could be effectively used to predict the incident solar radiation, based on above mentioned weather parameters particularly with the number of available observations. The model was trained using historical weather and solar radiation data of 14 years to predict hourly global horizontal irradiation (GHI). The trained model produced outputs having the correlation coefficient (R) 0.86 with the target values for an unseen weather parameter data set, reflecting a good generalization ability.

Decomposition and transposition model had to be used for the conversion of forecasted GHI to global tilted irradiation (GTI), in order to calculate solar power output and energy generation. The entire model was validated against 14 months daily generation data collected from the plant. The value of correlation coefficient (R) drops slightly from 0.86 to 0.84 when it comes to the power forecast model compared to the Nural Network (NN) based GHI forecast model. However, R value for the power forecast model was calculated based on around 420 target - output observations compared to more than 5100 such observations were contributed for the value of R of the NN model.

The power forecasting model produced predictions with mean forecasting accuracy of 86% and 12% standard deviation which were computed based on the mean average percentage error. Even though the model could be used for the forecasting of power output or energy generation for 5 days ahead, one week ahead or even further extended duration, weather forecasting accuracy drops significantly as the duration is extended. According to NOAA (National Oceanic and Atmospheric Administration), accuracy of weather predictions for one day ahead period is around 95%. However, accuracy reduces to around 90% for 5 days ahead, around 80% for one week ahead forecasts and so on. Therefore, using of the model is recommended for producing of one day ahead forecasts.

Accuracy of the model is also affected by the uncertainty of various losses associated with solar power generation. Most of these losses were considered as flat percentages, but observations suggest that they change specially with extreme conditions. The research can be further extended by finding ways to model the non-linearity of these losses.

References

[1] B. Fattouh, R. Poudineh, and R. West, "The rise of renewables and energy transition: what adaptation strategy exists for oil companies and oil-exporting countries?" *Energy Transitions*, vol. 3, no. 1–2, pp. 45–58, Dec. 2019, doi: 10.1007/s41825-019-00013-x.

[2] M. Agnoletti, S. Neri, and S. Editors, "Environmental History 4," 2014. [Online]. Available: http://www.springer.com/series/10168

[3] R. Hannah, "Energy Mix," 2020. https://ourworldindata.org (accessed Aug. 19, 2021).

[4] Enerdata, "Global Energy Statistical Yearbook 2020," 2021. https://yearbook.enerdata.net (accessed Aug. 19, 2021).

[5] World Nuclear Association, "Heat values of various fuels." https://world-nuclear.org/information-library/facts-and-figures/heat-values-of-various-fuels.aspx (accessed Aug. 19, 2021).

[6] F. Barbir, T. N. Veziroglu, and H. J. Plass, "Environmental Damage Due to Fossil Fuels Use," 1990.

[7] I. Tiseo, "Global historical CO2 emissions 1750 - 2020," *Statista*, 2021. https://www.statista.com/statistics/264699/worldwide-co2-emissions/ (accessed Jul. 02, 2021).

[8] M. Jaganmohan, "Annual anomalies in global ocean surface temperature from 1880 to 2021, based on temperature departure," *Statista*, Mar. 23, 2022. https://statista.com/statistics/ 736147/ocean-temperature-anomalies-based-on-temperature-departured#professional (accessed Mar. 29, 2022).

[9] R. S. Vose *et al.*, "Improved historical temperature and precipitation time series for U.S. climate divisions," *Journal of Applied Meteorology and Climatology*, vol. 53, no. 5, pp. 1232–1251, 2014, doi: 10.1175/JAMC-D-13-0248.1.

[10] C. Böhringer, "The Kyoto Protocol: A Review and Perspectives," 2003. [Online]. Available: http://oxrep.oxfordjournals.org/

[11] A. Savaresi, "The Paris agreement: A new beginning?" *Journal of Energy and Natural Resources Law*, vol. 34, no. 1, pp. 16–26, 2016, doi: 10.1080/02646811.2016.1133983.

[12] United Nations, "Report of the conference of the parties on its twenty-sixth session, held in Glasgow from 31 October to 13 November 2021," Mar. 2022. Accessed: Apr. 19, 2022. [Online]. Available: https://unfccc.int/sites/default/files/resource/cp2021_12_adv. pdf?

[13] BP, "BP Statistical Review of World Energy 2020," 2020. Accessed: Aug. 15, 2021. [Online]. Available: www.bp.com/statisticalreview.

[14] D. H. Meadows, D. L. Meadows, J. • Rgen, R. William, and W. Behrens Ill, "The Limits to Growth. A report for the Club of Rome's project on the predicament of mankind.," 1971.

[15] B. Pillot, M. Muselli, P. Poggi, and J. B. Dias, "Historical trends in global energy policy and renewable power system issues in Sub-Saharan Africa: The case of solar PV," *Energy Policy*, vol. 127, pp. 113–124, Apr. 2019, doi: 10.1016/j.enpol.2018.11.049.

[16] M. Guangul and G. Chala, "Solar Energy as Renewable Energy Source: SWOT Analysis," 2019.

[17] X. Wang and A. Barnett, "The evolving value of photovoltaic module efficiency," *Applied Sciences (Switzerland)*, vol. 9, no. 6, 2019, doi: 10.3390/app9061227.

[18] T. Andre *et al.*, "Renewables 2021 global status report," Paris, 2021.

[19] International Renewable Energy Agency (IRENA), *Renewable Capacity Statistics 2021 Statistiques De Capacité Renouvelable 2021 Estadísticas De Capacidad Renovable 2021*. Abu Dhabi, 2021. [Online]. Available: www.irena.org

[20] International Renewable Energy Agency (IRENA), *Future of Solar Photovoltaic Deployment, investment, technology, grid integration and socio-economic aspects A Global Energy Transformation paper About IRENA*. Abu Dhabi, 2019. [Online]. Available: www.irena.org/publications.

[21] P. R. Wolfe, “Utility-scale solar power,” in *McEvoy’s Handbook of Photovoltaics: Fundamentals and Applications*, Elsevier Inc., 2018, pp. 1073–1093. doi: 10.1016/B978-0-12-809921-6.00030-6.

[22] International Energy Agency (IEA), “Renewable Energy Market Update - Outlook for 2021 and 2022,” 2021. [Online]. Available: www.iea.org/t&c/

[23] International Energy Agency (IEA), “Renewables 2019,” 2019. [Online]. Available: www.iea.org/renewables2019

[24] International Energy Agency (IEA), “Renewables 2020 - Analysis and forecast to 2025,” 2020.

[25] BP, “Full report – Statistical Review of World Energy 2021,” London, 2021.

[26] International Renewable Energy Agency (IRENA), “Energy Profile Sri Lanka,” Abu Dhabi, 2020.

[27] Ceylon Electricity Board (CEB), “Annual Report 2011,” Colombo 02, Sri Lanka, 2011.

[28] Sri Lanka Sustainable Energy Authority (SLSEA), “Annual Report 2018 - Tomorrow’s Energy Today,” Colombo 07, Sri Lanka, 2018.

[29] Ceylon Electricity Board, “Annual Report 2016,” Colombo 02, Sri Lanka, 2016.

[30] Ceylon Electricity Board (CEB), “Long Term Generation Expansion Plan,” Colombo 02, Sri Lanka, 2018.

[31] F. Ueckerdt, R. Brecha, and G. Luderer, “Analyzing major challenges of wind and solar variability in power systems,” *Renewable Energy*, vol. 81, pp. 1–10, Sep. 2015, doi: 10.1016/j.renene.2015.03.002.

[32] Solcast, “Solar irradiance data,” 2021. https://solcast.com (accessed Mar. 11, 2021).

[33] Openweathermap, “Historical weather data collection,” 2021. https://openweathermap .org/api (accessed Mar. 11, 2021).

[34] B. N. Stram, “Key challenges to expanding renewable energy,” *Energy Policy*, vol. 96, pp. 728–734, Sep. 2016, doi: 10.1016/j.enpol.2016.05.034.

[35] M. Joshi, S. Rehaman, D. Palchak, S. C. Saxena, and S. R. Narasimhan, “Ramping up the ramping capability - India’s power system transition,” Golden, Colorado, Sep. 2020.

[36] C. Brancucci Martinez-Anido *et al.*, “The value of day-ahead solar power forecasting improvement,” *Solar Energy*, vol. 129, pp. 192–203, May 2016, doi: 10.1016/j.solener.2016.01.049.

[37] A. Kaur, L. Nonnenmacher, H. T. C. Pedro, and C. F. M. Coimbra, “Benefits of solar forecasting for energy imbalance markets,” *Renewable Energy*, vol. 86, pp. 819–830, 2016, doi: 10.1016/j.renene.2015.09.011.

[38] International Renewable Energy Agency (IRENA), *Advanced forecasting of variable renewable power generation: Innovation Landscape Brief*. Abu Dhabi, 2020. [Online]. Available: www.irena.org

[39] Agora Energiewende, “Flexibility in thermal power plants - with a focus on existing coal-fired power plants,” Berlin, Germany, Jun. 2017.

[40] International Renewable Energy Agency (IRENA), *Innovation landscape brief: Flexibility in conventional power plants*. Abu Dhabi, 2019. [Online]. Available: www.irena.org

[41] J. A. Duffie and W. A. Beckman, *Solar engineering of thermal processes*. Wiley, 2013.

[42] M. H. Chung, “Estimating Solar Insolation and Power Generation of Photovoltaic Systems Using Previous Day Weather Data,” *Advances in Civil Engineering*, vol. 2020, 2020, doi: 10.1155/2020/8701368.

[43] Solar Design Guide, “Solar Fundamentals,” 2021. https://solardesignguide.com/stc-and-noct-solar-panel-test-conditions-explained/ (accessed Sep. 05, 2021).

[44] R. K. Satpathy and V. Pamuru, “Solar PV Power: Design, Manufacturing and Applications from Sand to Systems,” 2020.

[45] C. Y. Park *et al.*, “Inverter efficiency analysis model based on solar power estimation using solar radiation,” *Processes*, vol. 8, no. 10, pp. 1–19, Oct. 2020, doi: 10.3390/pr8101225.

[46] J. Bruce, “Solar PV System Losses: How To Calculate Solar Panel Efficiency ,” 2021. https://diysolarshack.com/10-solar-pv-system-losses-their-impact-on-solar-panel-output/ (accessed Sep. 09, 2021).

[47] N. Martín and J. M. Ruiz, “Annual angular reflection losses in PV modules,” *Progress in Photovoltaics: Research and Applications*, vol. 13, no. 1, pp. 75–84, Jan. 2005, doi: 10.1002/pip.585.

[48] E. A. Sjerps-Koomen, E. A. Alsema, and W. C. Turkenburg, “A Simple Model for PV Module Reflection Losses Under Field Conditions,” 1996.

[49] M. R. Maghami, H. Hizam, C. Gomes, M. A. Radzi, M. I. Rezadad, and S. Hajighorbani, “Power loss due to soiling on solar panel: A review,” *Renewable and Sustainable Energy Reviews*, vol. 59. Elsevier Ltd, pp. 1307–1316, Jun. 01, 2016. doi: 10.1016/j.rser.2016. 01.044.

[50] A. Kimber, L. Mitchell, S. Nogradi, and Wenger H, “The Effect of Soiling on Large Grid-Connected Photovoltaic Systems in California and the Southwest Region of the United States,” 2006.

[51] A. al Mansur, M. Ruhul Amin, and K. K. Islam, “Performance comparison of mismatch power loss minimization techniques in series-parallel PV array configurations,” *Energies (Basel)*, vol. 12, no. 5, 2019, doi: 10.3390/en12050874.

[52] N. M. Kumar, R. P. Gupta, M. Mathew, A. Jayakumar, and N. K. Singh, “Performance, energy loss, and degradation prediction of roofintegrated crystalline solar PV system installed in Northern India,” *Case Studies in Thermal Engineering*, vol. 13, Mar. 2019, doi: 10.1016/j.csite.2019.100409.

[53] F. Mavromatakis, F. Vignola, and B. Marion, “Low irradiance losses of photovoltaic modules,” *Solar Energy*, vol. 157, pp. 496–506, 2017, doi: 10.1016/j.solener.2017.08.062.

[54] W. G. J. H. M. van Sark, N. H. Reich, A. Armbruster, K. Kiefer, and C. Reise, "Review of PV Performance Ratio Development," 2012.

[55] T. Parker, "How to maximize solar project value using inverter clipping," 2019. https://www.solarpowerworldonline.com/2019/12/how-to-maximize-solar-installation-value-using-inverter-clipping/ (accessed Sep. 09, 2021).

[56] J. Polo, M. Alonso-Abella, J. A. Ruiz-Arias, and J. L. Balenzategui, "Worldwide analysis of spectral factors for seven photovoltaic technologies," *Solar Energy*, vol. 142, pp. 194–203, 2017, doi: 10.1016/j.solener.2016.12.024.

[57] J. J. John, "Characterization of Soiling Loss on Photovoltaic Modules, and Development of a Novel Cleaning System," 2015.

[58] T. Ishii and A. Masuda, "Annual degradation rates of recent crystalline silicon photovoltaic modules," *Progress in Photovoltaics: Research and Applications*, vol. 25, no. 12, pp. 953–967, Dec. 2017, doi: 10.1002/pip.2903.

[59] D. Sharma and A. Khare, "Comparative Analysis of Solar PV Modules," *International Journal on Emerging Technologies*, vol. 5, no. 2, pp. 22–26, 2014, [Online]. Available: http://eosweb.larc.nasa.gov/cgi-

[60] K. D. Lawrence, R. K. Klimberg, and S. M. Lawrence, *Fundamentals of forecasting using Excel*. Industrial Press, 2009.

[61] S. Wheelwright, S. Makridakis, and R. J. Hyndman, *Forecasting methods and applications*. John Wiley & Sons, 1998.

[62] United States Census Bureau, "X-13 ARIMA - SEATS Seasonal Adjustment Program," 2021. https://www.census.gov/data/software/x13as.html (accessed Aug. 26, 2021).

[63] C. Chatfield and D. L. Prothero, "Box-Jenkins Seasonal Forecasting: Problems in a Case-Study," *Journal of the Royal Statistical Society*, vol. Series A (General), no. 136 (3), pp. 295–315, 1973.

[64] Keller, Paul E, Priddy, and Kevin L, "Artificial Neural Networks," 2005.

[65] Artificial Intelligence Techniques, “Data Science and Machine Learning Blog,” 2021. https://www.neuraldesigner.com/blog/5_algorithms_to_train_a_neural_network (accessed Aug. 22, 2021).

[66] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, “A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization,” *Renewable and Sustainable Energy Reviews*, vol. 124. Elsevier Ltd, May 01, 2020. doi: 10.1016/j.rser.2020.109792.

[67] A. Sanfilippo, L. Martin-Pomares, N. Mohandes, D. Perez-Astudillo, and D. Bachour, “An adaptive multi-modeling approach to solar nowcasting,” *Solar Energy*, vol. 125, pp. 77–85, Feb. 2016, doi: 10.1016/j.solener.2015.11.041.

[68] V. Prema and K. Uma Rao, “Development of statistical time series models for solar power prediction,” *Renewable Energy*, vol. 83, pp. 100–109, Nov. 2015, doi: 10.1016/j.renene. 2015.03.038.

[69] J. Huang and M. Perry, “A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 1081–1086, 2016, doi: 10.1016/j.ijforecast.2015.11.002.

[70] P. A. G. M. Amarasinghe and S. K. Abeygunawardane, “Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka; Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka,” 2018.

[71] H. T. C. Pedro and C. F. M. Coimbra, “Short-term irradiance forecastability for various solar micro-climates,” *Solar Energy*, vol. 122, pp. 587–602, 2015, doi: 10.1016/j.solener. 2015.09.031.

[72] T. Weerasekara, “Neural Network Model for Forecasting Solar Energy Generation and Analysis of Power Output Controlling by Energy Storage Schemes,” Moratuwa, 2019.

[73] M. Rana, I. Koprinska, and V. G. Agelidis, “Univariate and multivariate methods for very short-term solar photovoltaic power forecasting,” *Energy Conversion and Management*, vol. 121, pp. 380–390, Aug. 2016, doi: 10.1016/j.enconman.2016.05.025.

[74] A. G. R. Vaz, B. Elsinga, W. G. J. H. M. van Sark, and M. C. Brito, “An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands,” *Renewable Energy*, vol. 85, pp. 631–641, Jan. 2016, doi: 10.1016/j.renene.2015.06.061.

[75] M. Yesilbudak, M. Colak, R. Bayindir, and N. Haci Bektas, “What are the Current Status and Future Prospects in Solar Irradiance and Solar Power Forecasting?” 2018.

[76] H. T. C. Pedro and C. F. M. Coimbra, “Assessment of forecasting techniques for solar power production with no exogenous inputs,” *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, Jul. 2012, doi: 10.1016/j.solener.2012.04.004.

[77] R. C. Deo and M. Şahin, “Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland,” *Renewable and Sustainable Energy Reviews*, vol. 72. Elsevier Ltd, pp. 828–848, 2017. doi: 10.1016/j.rser.2017.01.114.

[78] Y. Wu and J. Wang, “A novel hybrid model based on artificial neural networks for solar radiation prediction,” *Renewable Energy*, vol. 89, pp. 268–284, Apr. 2016, doi: 10.1016/j.renene.2015.11.070.

[79] A. A. Munshi and Y. A. R. I. Mohamed, “Comparisons among Bat algorithms with various objective functions on grouping photovoltaic power patterns,” *Solar Energy*, vol. 144, pp. 254–266, 2017, doi: 10.1016/j.solener.2017.01.034.

[80] T. C. McCandless, S. E. Haupt, and G. S. Young, “A regime-dependent artificial neural network technique for short-range solar irradiance forecasting,” *Renewable Energy*, vol. 89, pp. 351–359, Apr. 2016, doi: 10.1016/j.renene.2015.12.030.

[81] B. Sivaneasan, C. Y. Yu, and K. P. Goh, “Solar Forecasting using ANN with Fuzzy Logic Pre-processing,” in *Energy Procedia*, 2017, vol. 143, pp. 727–732. doi: 10.1016/j.egypro. 2017.12.753.

[82] D. T. Larose and W. H. Freeman, *Discovering the Fundamentals of Statistics*. New York: W.H. Freeman & Company, 2014. [Online]. Available: www.whfreeman.com

[83] Department of Metrology, “Time forecasting system.” https://www.meteo.gov.lk/ (accessed Sep. 12, 2021).

[84] R. F. Rahamat, F. R. Nasution, M. F. Syahputra, and Sitompul O.S., “Implimentation of Bayesian model averaging on the weather data forecastiong applications utilizing Openweathermap,” *IOP Conference Series*, vol. 309, 2018.

[85] M. M. U. Hussain, B. Avci, G. Trajcevski, and P. Scheuermann, “Incoporating weather updates for public transportation users of recommended systems,” *17th IEEE International Conference on Mobile Data Managment (MDM)*, vol. Vol. 1, pp. 333–336, 2016.

[86] H. Du, C. F. Bandera, and I. Chen, “Nowcasting Methods for Optimising Building Performance,” 2019.

[87] A. Roy, E. Simelton, and C. Quinn, “Which forecast represents the local weather best? Preliminary case study findings from My Loi village, Northcentral Vietnam,” 2017.

[88] G. Demirezen and A. S. Fung, “Application of artificial neural network in the prediction of ambient temperature for a cloud-based smart dual fuel switching system,” in *Energy Procedia*, 2019, vol. 158, pp. 3070–3075. doi: 10.1016/j.egypro.2019.01.992.

[89] S. Bhatti and A. Williams, “Estimation of surplus energy in off-grid solar home systems,” *Renewable Energy and Environmental Sustainability*, vol. 6, p. 25, 2021, doi: 10.1051/rees/2021024.

[90] O. Asowata, J. Swart, and C. Pienaar, “Correlating the power conversion of a PV panel to the solar irradiance obtained from Meteonorm,” *IEEE International Conference on Industrial Technology (ICIT)*, no. IEEE, pp. 684–689, 2013.

[91] B. Marion, “Measured and satellite-derived albedo data for estimating bifacial photovoltaic system performance,” *Solar Energy*, pp. 321–327, 2021.

[92] R. Abdullah *et al.*, “Estimation of solar irradiation and optimum tilt angles for south facing surfaces in the United Arab Emiratates: A case study using PVGIS and PVWatts,” *In Recent Advances in Renewable Energy Technologies*, pp. 3–39, 2022.

[93] SunPower Corporation, “SunPower Performance Series P17,” 2016. https://us.sunpower .com/sites/default/files/media-library/data-sheets/ds-sunpower-p17-355-commercial-solar-panels.pdf (accessed Sep. 12, 2021).

[94] SunPower Corporation, “P series commercial solar panels: Supplementary technical specifications.” https://us.sunpower.com/sites/default/files/sp-p-series-commercial-solar-panels-supplimentary-technical-spec.pdf (accessed Sep. 14, 2021).

[95] SunPower Corporation, “SunPower Performance Series PVsyst Modeling Guide 2016,” 2016. https://us.sunpower.com/sites/default/files/media-library/manuals/mn sunpower-p-series-modeling-guide.pdf (accessed Sep. 14, 2021).

[96] ABB, “ABB solar inverters Product manual PVS-100/120-TL.” https://new.abb.com/products/6AGC076172/3q179909000a (accessed Sep. 15, 2021).