NEURAL NETWORK MODEL FOR FORECASTING SOLAR ENERGY GENERATION AND ANALYSIS OF POWER OUTPUT CONTROLLING BY ENERGY STORAGE SCHEMES

Weerasekera W.B.M.U.T

(148367F)

Degree of Master of Engineering

Department of Mechanical Engineering

University of Moratuwa
Sri Lanka

May 2019

NEURAL NETWORK MODEL FOR FORECASTING SOLAR ENERGY GENERATION AND ANALYSIS OF POWER OUTPUT CONTROLLING BY ENERGY STORAGE SCHEMES

Weerasekera W.B.M.U.T

(148367F)

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

Department of Mechanical Engineering

University of Moratuwa
Sri Lanka

May 2019

Declaration

I declare that this is my own work and that this thesis doesn't include material from a previous submission of a third-party for a diploma or degree of any University or institute. This thesis does not contain previously published material from books, articles or research papers except where proper acknowledgement is made in the text.

Further, I hereby grant the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or part in print, electronic or other medium while retaining the right to use this content for my future endeavors.

Signature: ..... Date: .....

W.B M U T Weerasekera

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

Signature: ..... Date: .....

Prof. R.A Attalage

Abstract

Modeling of power fluctuations in a solar PV power plant using an Artificial Neural Network (ANN) was carried out in the study. The resulting model was used to evaluate the energy storage requirement to control fluctuations of the power output. The ANN was trained to model the output of a 300kW solar PV system installed in Colombo with an average hourly energy output of 90.55kWh and an average daily energy production of 1177 kWh. The ANN model proved to deliver forecasts with significant accuracy and generalizability. Correlation coefficients for training, validation and testing were 0.945 0.948 and 0.939 respectively. Further validation was done using an isolated data set of a time period of a month for which model was able to achieve a correlation coefficient of 0.93. Residual analysis confirmed the error was random and free of autocorrelation. Error terms had a normal distribution with mean 1.09kWh and standard deviation of 20.06kWh. A direct mapping was established between meteorological parameters and power output of a solar PV system, as oppose to estimating solar irradiance. Energy storage requirement was evaluated for two power output control schemes. First scheme specifies a ramp up, ramp down rate, and a continuous power delivery period. By means of an optimizing algorithm the combination of parameters corresponding to the least energy storage requirement was established and the result for energy requirement was approximately 15% of the average daily energy generation of the PV system. Variation of energy storage requirement under different operating conditions was analyzed further. The second output control scheme uses moving average smoothening to control power output. The calculated energy storage requirement for moving average scheme was approximately 8% of the average daily energy generation. The effect of imposing restriction on operating parameters of the schemes were examined in detail.

Acknowledgement

This research was carried out under the supervision of senior professor R.A Attalage, Department of Mechanical Engineering, University of Morata. I am very much grateful to Prof. R.A Attalage for providing me with the guidance, motivation and support throughout the research. I admire his kind hearted co-operation and encouragement extended at every stage of the study. I wish to thank Dr. Himan Punchihewa for the continuous engagement with the students to successfully complete the research. Finally, I would like to acknowledge all lecturers and teachers that imparted values and knowledge to us.

Table of Contents

Declaration i
Abstract ii
Acknowledgement iii
List of figures vi
List of Tables viii
Chapter 1 Introduction 1
1.1 Background of study 1
1.1.1 Solar Energy Potential in Sri Lanka 2
1.1.2 Sri Lankan context of Solar Energy Applications 6
1.2 Problem Statement 7
1.2.1 Aim 8
1.2.2 Objectives 8
1.2.3 Scope 8
1.2.4 Outcomes of the study 9
1.2.5 Outline of the Study 9
Chapter 2 Review of Literature 10
2.1 Fundamentals of Solar Radiation 10
2.1.1 Introduction to the solar radiation 10
2.1.2 Solar geometry 11
2.2 Introduction to the solar irradiation fluctuations 17
2.2.1 Weather seasons of Sri Lanka and climatic patterns 20
2.2.2 Effect of clouds on solar irradiation 24
2.3 Mathematical models for solar irradiation and PV power forecasting 35
2.3.1 Physical methods of forecast 35
2.3.2 Statistical methods of forecast 37
2.4 Artificial Neural Networks (ANN) 40
2.4.1 Introduction 40
2.4.2 Fundamentals of Neural Networks 41
2.4.3 Solution of a single hidden layer ANN 43
2.4.4 Training validation and testing an artificial neural network 49
2.4.5 Applications of neural networks in similar scope 51
2.4.6 Methods of evaluating model performance 52
Chapter 3 Research Approach 56
3.1 Introduction 56

3.2 Development and training of a neural network 56
3.2.1 Identification of input parameters 56
3.2.2 Establishing time span for neural network 64
3.2.3 Design of the artificial neural networks 66
3.2.4 Training neural network with data from existing solar PV project 69
3.2.5 Validation of model 73
3.2.6 Evaluation of adequacy and significance of the predictive model 75
3.3 Calculate energy storage requirement under different operating schemes 79
3.3.1 Operation under controlled ramp rate 79
3.3.2 Operation under moving average smoothening 83
Chapter 4 Results and Discussion 84
4.1 Evaluation of energy storage requirement. 87
Chapter 5 Conclusion and future work 93
References 96

List of figures

Figure 1-1 : Direct Normal Irradiation (DNI) data for horizontal plane.....3
Figure 1-2: Annual average daily total solar irradiance at surface tilted at latitude angle 4
Figure 1-3 Monthly average hourly global horizontal irradiation (W/m^2) of Colombo for the duration from year 2000 to 2003 6
Figure 1-4 installed Solar PV capacity of Sri Lanka and projected growth 7
Figure 1-5 Periodic change of Solar caused by relative motion of the sun 7
Figure 1-6 Short term random variation caused by weather.....7
Figure 1-7 Hourly variation of PV power.....9
Figure 2-1 Exaggerated depiction of the angle subtended on Earth by the sun due to the placement 10
Figure 2-2 Earth's orbit relative to the sun.....12
Figure 2-3 Variation of solar declination.....12
Figure 2-4 Variation of the solar declination angle throughout the year.....13
Figure 2-5 The Latitude angle (ϕ) and Hour angle (ω).....14
Figure 2-6: Altitude angle (α), Zenith angle (θ_z), and Azimuth angle (A_z).....15
Figure 2-7 : The position of the sun relative to a tilted plane.....16
Figure 2-8 Solar PV power curves under three sky conditions 17
Figure 2-9 Geographical smoothening of collective solar power curve 18
Figure 2-10 Effect of sky condition change on PV power within a day.....19
Figure 2-11 Rainfall of NE monsoon 1200mm-200mm.....20
Figure 2-12 Rainfall distribution of first inter-monsoon 700mm -100mm.....21
Figure 2-13 Rainfall distribution of SW monsoon 300mm-2000mm.....22
Figure 2-14 Rainfall distribution of inter-monsoon 1000mm-200mm 22
Figure 2-15 Variation of monthly average solar irradiance over the year 23
Figure 2-16 Cirrus clouds with fibrous texture source www.metoffice.gov.uk 24
Figure 2-17 Cirrocumulus clouds www.metoffice.gov.uk 25
Figure 2-18 Cirrostratus clouds www.metoffice.gov.uk 25
Figure 2-19 Altocumulus clouds www.metoffice.gov.uk.....26
Figure 2-20 Altostratus clouds www.metoffice.gov.uk.....27
Figure 2-21 Stratus Clouds www.metoffice.gov.uk 27
Figure 2-22 Stratocumulus Clouds www.metoffice.gov.uk 28
Figure 2-23 Nimbostratus Clouds www.metoffice.gov.uk.....28
Figure 2-24 Cumulus Clouds www.metoffice.gov.uk.....29
Figure 2-25 Cumulonimbus clouds www.metoffice.gov.uk.....30
Figure 2-26 Visual evaluation of cloud types, International Cloud Atlas WMO 31
Figure 2-27 Variation of irradiance with cloud cover and solar altitude angle 33
Figure 2-28 Image of the complete sky take from a TSI system 36
Figure 2-29 Single neuron of an artificial neural network.....41
Figure 2-30 Single hidden layer neural network structure.....43
Figure 2-31 Nomenclature of synaptic weights of a neural network.....43
Figure 2-32 Total sum of squares and regression sum of squares comparison 53
Figure 2-33 Linear model fit with R2>99% 53
Figure 2-34 Residuals appear random without a pattern 54

Figure 2-35 Histogram of errors with peak frequency at zero 55
Figure 3-1 Monthly solar PV energy output with monthly rainfall 57
Figure 3-2 Solar PV energy output with monthly average irradiation 59
Figure 3-3 Comparison of actual monthly average energy output with multivariable liner model 64
Figure 3-4 comparison of correlation coefficient of monthly models to annual model 65
Figure 3-5 comparison of mean error of monthly models to annual model 65
Figure 3-6 Correlation coefficient of training validation testing and overall model with number of hidden neurons 67
Figure 3-7 Structure of the artificial neural network for each month 68
Figure 3-8 Synaptic weights nomenclature of the monthly neural network models 69
Figure 3-9 Overall correlation coefficient variation with number of hidden layer neurons for annual model 70
Figure 3-10 Convergence of mean square error with training iterations 71
Figure 3-11 Error Histogram 71
Figure 3-12 Regression of Training Validation and Testing 72
Figure 3-13 Predicted power cycles to target cycles 73
Figure 3-14 Comparison of prediction to target data 73
Figure 3-15 Random distribution of error terms 74
Figure 3-16 Histogram of residuals for February 74
Figure 3-17 Normalized distribution of error 76
Figure 3-18 Comparison of error distribution to standard normal curve 77
Figure 3-19 Predicted monthly total energy comparison with actual monthly total energy and error percentage 78
Figure 3-20: Ramp rate control scheme 79
Figure 3-21 Ramp rate control compared to uncontrolled PV power output for an day 80
Figure 3-22 Potential ramp rate controlled paths identified in the optimization program 81
Figure 3-23 Optimum ramp rate controlled path and variation of SOC of the storage system 82
Figure 3-24 Operation under moving average output control scheme 83
Figure 4-1 Monthly energy output successfully modeled by identified variables 85
Figure 4-2 Comparison of model agreement to the training, validation, testing and overall data sets. 86
Figure 4-3 Energy storage requirement variation with time for ramp rate controlled scheme and moving average control scheme part 1 88
Figure 4-4 Energy storage requirement variation with time for ramp rate controlled scheme and moving average control scheme part 2 88
Figure 4-5 Mean energy storage requirement variation with averaging window size 89
Figure 4-6 Histogram of energy storage requirement for ramp rate controlled scheme 89
Figure 4-7 Variation of energy storage requirement 90
Figure 4-8 Cumulative probability of energy storage requirement under t2>0 and Rmax=40 condition 91
Figure 4-9 Variation of energy storage requirement for 95% of the days 92

List of Tables

Table 1-1 Monthly averaged insolation of Colombo ---------- 5
Table 2-1 Analysis of variation of solar radiation on degree of cloud cover and altitude angle by Dorota Matuszko in Krakow, Poland 2004-2007 ----------32
Table 2-2 Percentage of radiation intensity variation with cloud type and solar altitude angle for Krakow- Poland 2004-2007 ----------33
Table 2-3 Transmission of solar ratdiation by cloud genre analysis by Dorota Matuszko for Krakow poland 2004-2007----------34
Table 3-1 Monthly total energy variation with monthly rainfall average cloud cover and monthly insolation----------56
Table 3-2 Correlation coefficient and mean error of monthly neural network models---------64
Table 3-3 Correlation coefficient and mean error of annual neural network model----------65
Table 3-4 Variation of number of hidden layer neurons with alpha ----------67
Table 3-5 Correlation coefficient of training validation testing and overall model with number of hidden neurons ----------67
Table 3-6 Allocation of data Training Validation and Testing sets ----------70
Table 3-7 Mean square error and coefficient of correlation for training, validation and testing data sets ----------72
Table 3-8 Outlier data points ----------75
Table 3-9 Classed error terms ----------75
Table 3-10 Descriptive statistics of error term distribution ----------76
Table 3-11 Actual and predicted monthly total energy error percentage----------78

Chapter 1 Introduction

1.1 Background of study

Sri Lanka, is an upper middle income earning country with a rapidly increasing energy consumption owing to the economic growth rate. Aforementioned energy requirement is fulfilled by primary energy sources such as petroleum or by secondary sources such as electricity, generated using primary sources coal, petroleum, hydro power and biomass.

The expenditure for importing petroleum has caused severe strain to the economy. The effect was provoked further owing to the rapid fluctuations and increasing international oil prices. Reduced rainfall increased demand on petroleum imports. Under this circumstance, the country had to depend on expensive petroleum based thermal power generation to fulfill the rising energy demand. As a result of this unfavorable situation, the country had to spend a significant portion of its export earnings on petroleum importation. [1]

Under such conditions the Sri Lankan energy sector is compelled to exploring indigenous energy sources such as renewable energy derived from solar, wind, biomass, and hydro. Accommodating such available resources is favorable for the long term energy security in the country as well as to survive through the crisis at hand. Solar and wind are the most prominent renewable energy sources in terms of potential. Solar energy has significant advantages over other sources and probably is the most important renewable energy source available in present day. Most of other renewable energy sources are also derived directly or indirectly from solar energy.

Solar energy has advantages such as cleanliness, free availability, and accessibility from most of the urbanized geographical locations of the country throughout almost the entire year, capability to operate independently or in conjunction with traditional energy sources and being remarkably renewable. Sri Lanka experiences a high solar irradiance that can be a viable energy source for electricity production. During the recent times there has been a growing trend in accommodating solar electricity at domestic, industrial and utility scale.

Solar energy has the disadvantages of intermittency and variability. These ae common with all renewable sources. Accommodating solar energy is limited by the aforesaid inherent fluctuating nature of Photovoltaic electricity due to the high variability of the primary energy

source. As the solar electricity penetration increase, this fluctuation will affect the stability and controllability of the main utility grid. To mitigate this limitation it's absolutely vital to have the ability to predict solar generation and means of dampening the fluctuations.

1.1.1 Solar Energy Potential in Sri Lanka

Sri Lanka is in the equatorial belt where a significant amount of solar energy exist throughout the year. The solar energy resources in Sri Lanka has been estimated by several parties. Some estimations are based on the daily total solar radiation recorded at a number of weather stations throughout the country and some are based on the satellite image processing.

Climatological Solar Radiation (CSR) is a model developed by National Renewable Energy Laboratory (NREL) of Department of Energy United States capable of predicting monthly and annual average daily total solar radiation. The CSR was used to evaluate monthly average daily total solar radiation estimates for the seven years period from 1985 to 1991.

The results of this study show the distribution of solar resources measured in terms of annual average daily insolation (on tilted surface where the tilt angle equals to the local latitude). The insolation predicted is in the range of 4.5 to 6.0 kWh/m^2/day across the country. Lowest values appear in the hill country of the Central Province [2]. The seasonal variation of isolation is in the range between 4.5 and 6.5 kWh/m^2/day (on tilted surface where the tilt angle equals to the local latitude) [2]. The CSR predicts that the highest insolation is during the hot dry period between March and April during the first inter-monsoon season.

Following figures show the solar resource potential of direct normal irradiation on horizontal plane and annual average daily total irradiance at tilted surface where the tilt angle equal to the local latitude angle.

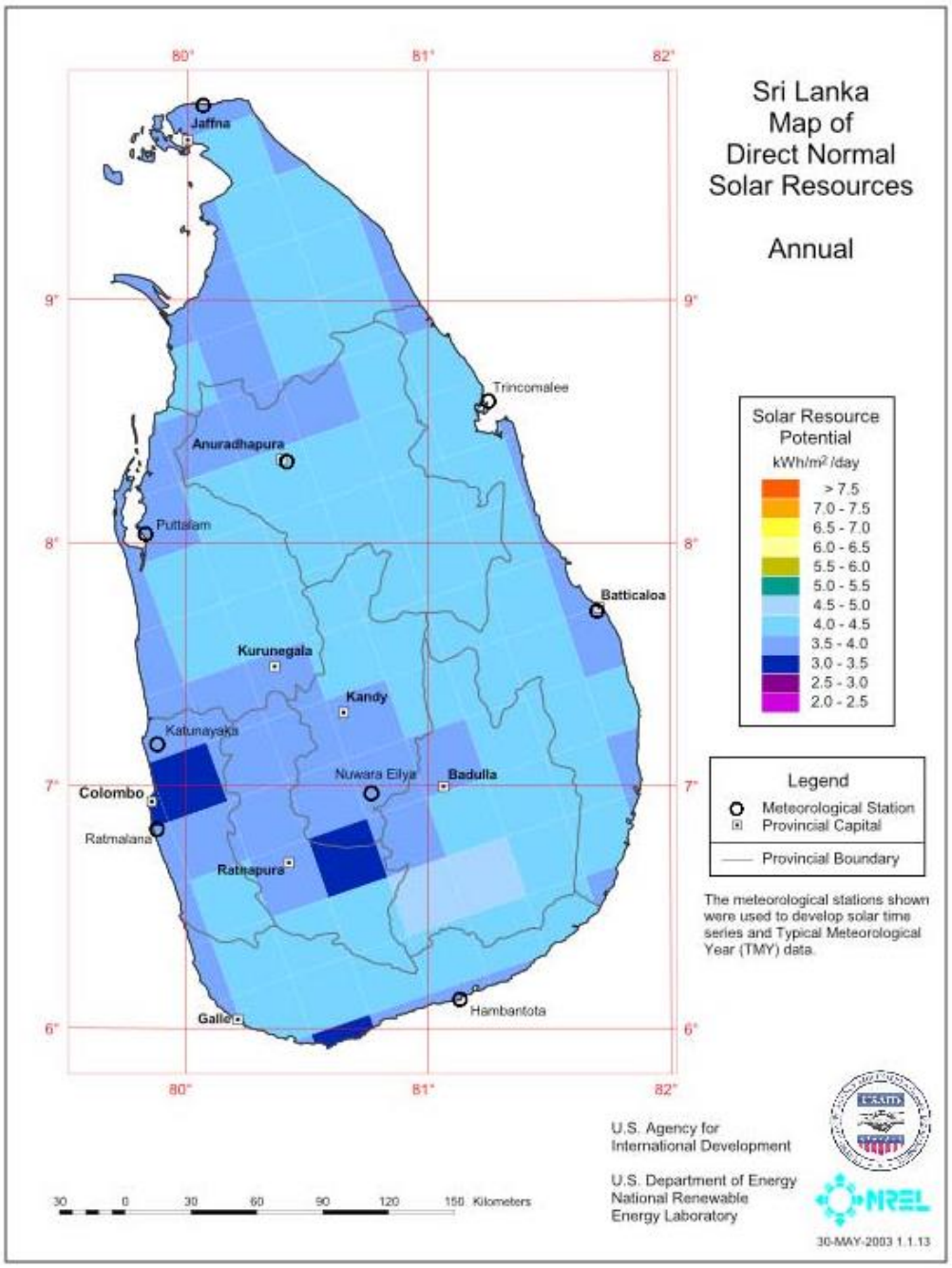


Figure 1-1 : Direct Normal Irradiation (DNI) data for horizontal plane.
Source: Solar Resource assessment for Sri Lanka and Maldives by NREL

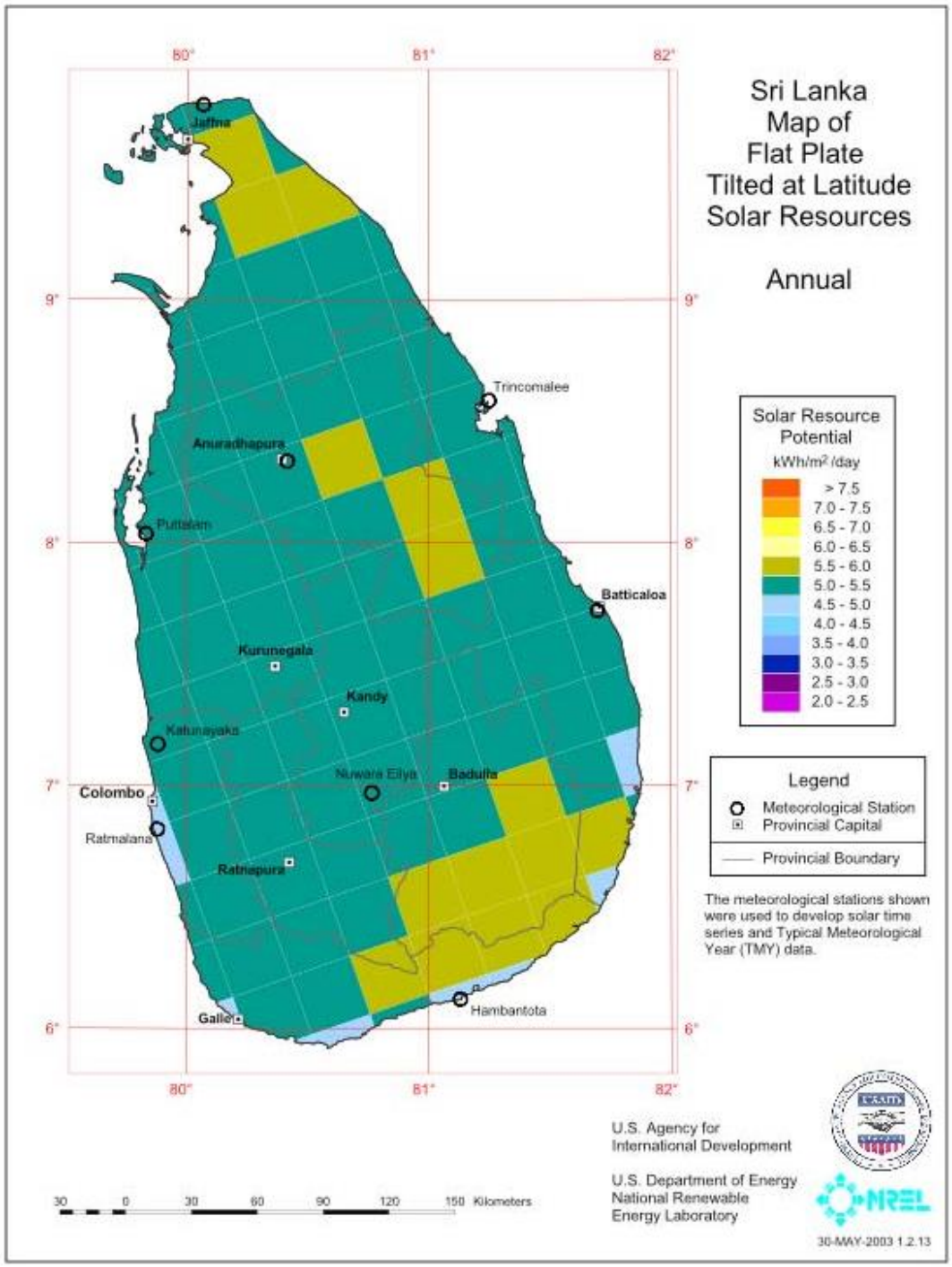


Figure 1-2: Annual average daily total solar irradiance at surface tilted at latitude angle
Source: Solar Resource assessment for Sri Lanka and Maldives by NREL

According to the figure, it can be concluded that harnessing of annual average daily solar radiation of 5.0-5.5 kWh/m2/day (on tilted surface with tilt angle to the local latitude angle) is possible in most parts of the country. Another conclusion drawn is that there exists a substantial solar energy potential in the dry zone.

NASA-SSE model is a satellite image based solar radiation model to estimate atmospheric data. NASA-SSE has estimated monthly average daily insolation for Colombo, averaged for the 22 years from 1983 to 2005. The result data is available at the Atmospheric Science Data Center web site of National Aeronautics and Space Administration (NASA). NASA-SSE model estimates 6.67 kWh/m2/day as the maximum insolation value for the month of March while 4.93 kWh/m2/day as the minimum insolation value for November [3]. The estimated annual average global horizontal solar insolation at Colombo is about 5.58 kWh/m2/day [3]. Monthly averaged results of NASA-SSE for insolation (kWh/m2/day) of Colombo averaged over 1983-2005 is as follow.

Table 1-1 Monthly averaged insolation of Colombo

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
5.50	6.27	6.67	5.96	5.29	5.30	5.40	5.65	5.65	5.29	4.93	5.16	5.58

Source: NASA Atmospheric Science Data Center

Employing solar energy for domestic and industrial applications in Sri Lanka, is economically favorable since the amply available energy source can minimize energy cost. According to the solar radiation data of Colombo for the period 2000-2003, solar energy is available with sufficient intensity for more than six hours in the day time throughout the year. Therefore solar photovoltaic modules and solar thermal collectors can function efficiently during major part of the day time. The figure below shows the sunshine hours along with the irradiation.

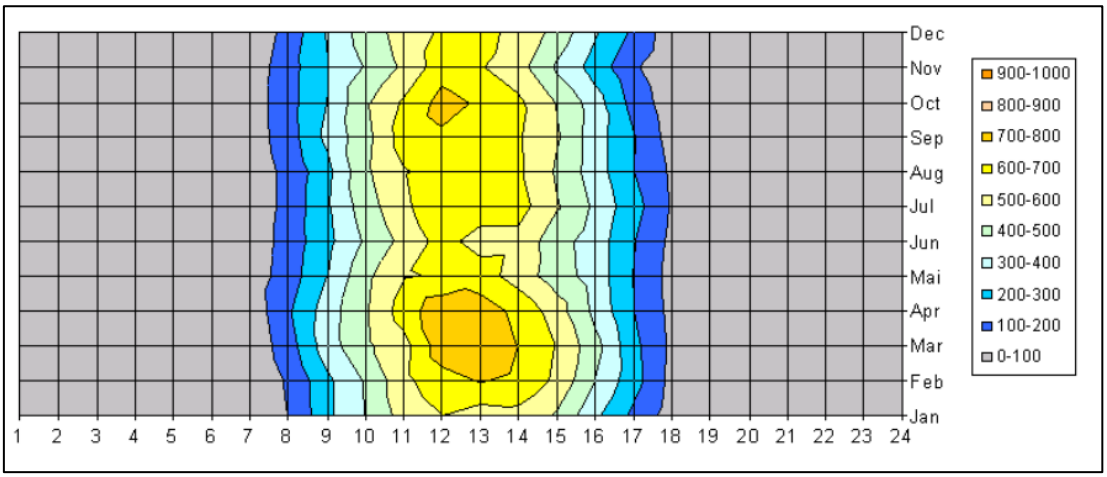


Figure 1-3 Monthly average hourly global horizontal irradiation (W/m^2) of Colombo for the duration from year 2000 to 2003
Source: Solar and Wind Energy Resource Assessment (SWERA) Report

1.1.2 Sri Lankan context of Solar Energy Applications

Solar energy is being used mainly as a source for electricity generation. The energy policy of Sri Lanka gives high priority to renewable energy [4]. Independent power producers are investing in the sector, resulting in a rapid growth.

On the user end both households and Industries are opening up to the technology especially with encouraging tariff schemes introduced by the Ceylon Electricity Board such as Net metering Net Accounting and Net plus. Households can drastically reduce their electricity bill while having capital cost recovery within the range of five to seven years. Industries while enjoying the same financial benefits as households can progress in their green and sustainability ratings. This is prominent in export industry where the carbon footprint of the product is meticulously calculated.

The current situation of solar energy system installation is in the range of 200MW$_p$ and could grow to double that amount [5].

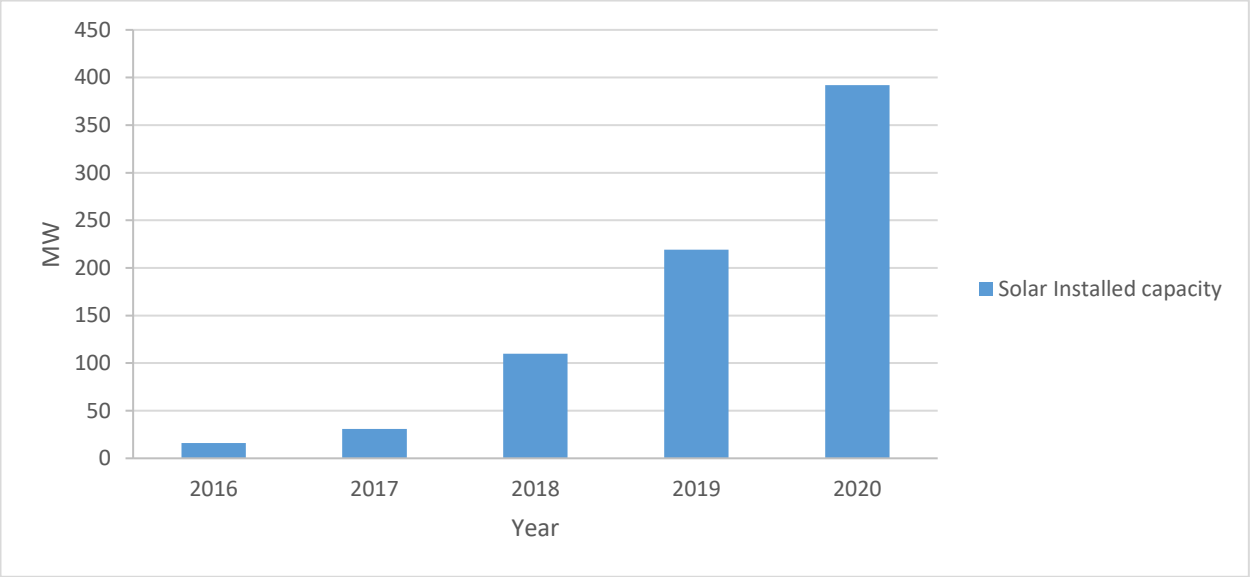


Figure 1-4 installed Solar PV capacity of Sri Lanka and projected growth
Source: Assessment of Sri Lanka power sector UNDP

1.2 Problem Statement

Solar energy is an exceptionally sustainable energy source for Sri Lanka. The only drawback in accommodating solar energy as an energy source is the characteristic intermittency and variability. These are common with all renewable sources.

Solar electricity is dependent primarily on the solar irradiance, which has periodic and random variations. The periodic variations are caused by the relative motion of the sun while random variations result mostly from cloud cover and seasonal changes. A predictive model would take these variations into account and estimate energy generation capacity of a solar power plant.

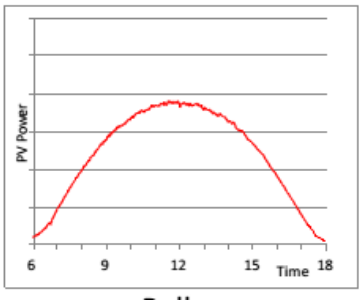


Figure 1-5 Periodic change of Solar caused by relative motion of the sun

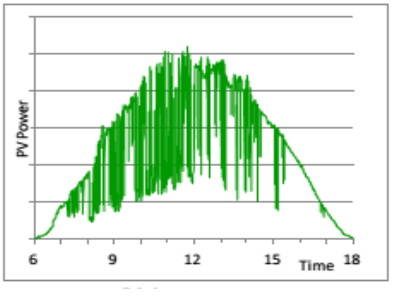


Figure 1-6 Short term random variation caused by weather.

As the solar electricity penetration increase, this fluctuation will affect the stability and controllability of the main utility grid. To mitigate this limitation it's absolutely vital to have the ability to predict solar generation and means of dampening the fluctuations.

1.2.1 Aim

Create a mathematical model to predict solar energy fluctuations, which will be used to optimally incorporate energy storage systems such that the electricity produced meets the acceptance criteria of the national grid.

1.2.2 Objectives

1. To explore solar electricity fluctuations and effects on the national grid
2. To identify and study meteorological patterns which cause fluctuations
3. To establish a mathematical model between variations in generated solar electricity and meteorological patterns
4. To Explore and evaluate energy storage schemes suited for mitigating the predicted variations

1.2.3 Scope

Fluctuation of solar energy can be of two types

1. Daily variation due to the change in position of the sun
2. Random variation

The first type can be closely expressed as a function of time using irradiation models and solar PV system characteristics. The second type of variation is random and chaotic. These random variations can be categorized by the time duration of fluctuations.

1. Very short duration flucuations
    a. Characterized by time spans of mS to S.
    b. Drops in power up to 95%.
    c. Smoothing can be done by super capacitors or by gerographical smoothnig
2. Medium duration fluctuations
    a. Characterized by time spans of 15 mis to 1hr.
    b. Captures effects of rains, seasonal cloud cover.
    c. Useful in planing and shceduling the power system in near future.
    d. Smoothing can be done by energy storage systems

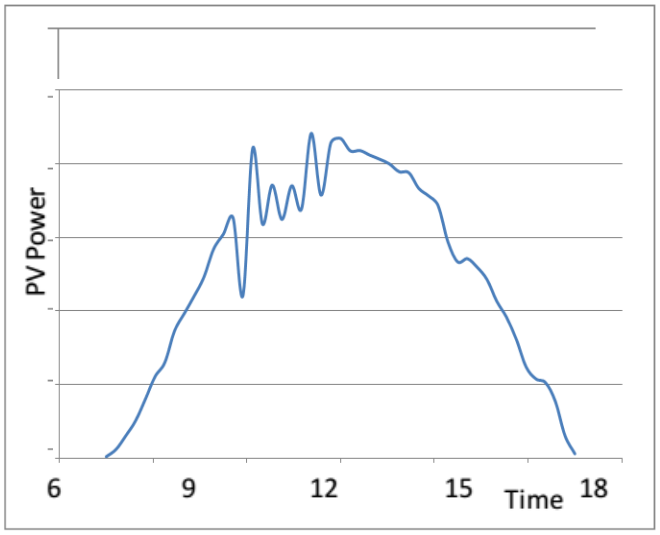


Figure 1-7 Hourly variation of PV power

The scope of the study is limited to the medium duration fluctuations. Neural networks are well equiped to modeling phenomena of this nature. The predictive model can be used to evaluated maximum power variations for sizing energy storage system as well as give the system controller the ability to plan electicity generation for the grid.

1.2.4 Outcomes of the study

1. Predictive mathematical model based on neural networks to estimate variations in solar electricity generation
2. Method to optimally size energy storage systems such that the electricity produced meets the acceptance criteria of the national grid.

1.2.5 Outline of the Study

First chapter of the thesis, describes characteristics of utilization of solar energy in Sri Lanka and the potential of harnessing of the energy resource. Present status of solar energy generation and its trends are also highlighted. Fundamental theories which describe solar geometry and basic concepts of solar radiation are discussed in Chapter 2. A literature review on correlations developed by various scientists and engineers is included in the same chapter. Chapter 3 involves the research methodology in detail. The results of the study are discussed in Chapter 4. The final chapter is reserved to present the discoveries and conclusions.

Chapter 2 Review of Literature

2.1 Fundamentals of Solar Radiation

2.1.1 Introduction to the solar radiation

Energy from the sun reaches earth by means of radiation. Solar radiation includes the total spectrum of the electromagnetic radiation of the sun. The sun is a star and center of the solar system. It is an extremely hot sphere mainly consisting of Hydrogen and Helium. The nuclear fusion reaction is the source of the immense energy emitted by the sun. .The diameter of the sun is 1.39×10^{9}m while earth has a diameter about 1.27×10^{7}m [6]. The average distance between the sun and the earth is approximately 1.5×10^{8}km [6]. Due to this large distance sun subtends a very small angle of 32 minutes on the earth's surface [6]. Therefore we can assume that the earth receives almost parallel beam radiation from the sun. The very significant distance implies that light traveling at speed of 3×10^{8} m/s in the vacuum takes 8 minutes and 20 minutes to reach earth from the sun [6].

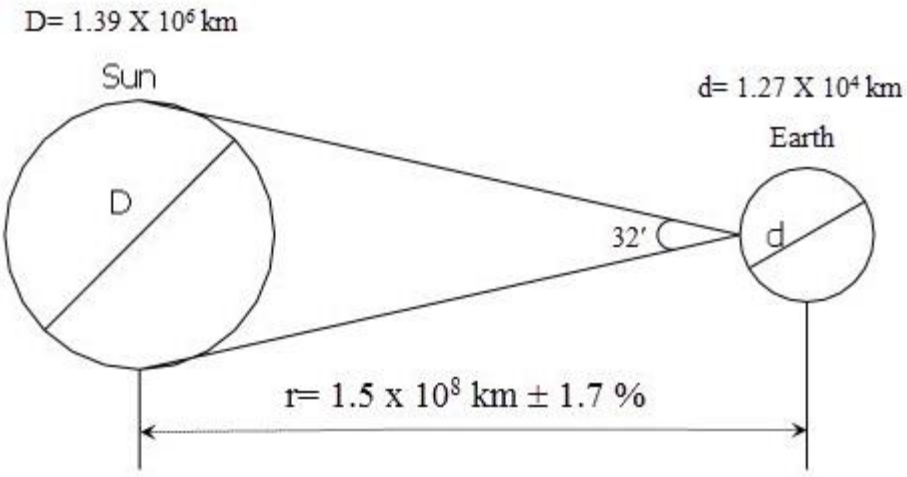


Figure 2-1 Exaggerated depiction of the angle subtended on Earth by the sun due to the placement

The fusion reaction of Hydrogen is facilitated by the immense pressure caused by gravity at the core of the sun. The sun emits this energy at a rate of 3.8×10^{20}MW, resulting in a per square meter rate of emission of 63 MW/m^2 of the sun's surface [6]. Energy radiates in all directions uniformly. The earth receives a very small portion of the total radiation emitted, which is equal to 1.7×10^{14}kW [6]. The brightness of the sun varies from its center to the edge. For engineering applications this variation is neglected. As a norm the solar disc is assumed to be of brightness.

Terrestrial radiation is defined as the radiation within the atmosphere, and extraterrestrial radiation is defined as radiation outside of the earth's atmosphere. Solar radiation is modeled as radiation of a blackbody surface with a temperature of 5762K [6]. The energy of the sun light measured on earth surface is in the range of 120 Wm^{-2}. Direct sunlight gives about 93 lux of illumination per watt of electromagnetic power including infrared, visible and ultraviolet radiation. Bright sunlight provides illumination of approximately 100 000 lux per square meter at the earth's surface [6].

2.1.2 Solar geometry

Solar geometry is the study of the position of the sun on the celestial globe. The position is defined by three angels. The amount of energy received on a specific location is dependent on the position of the sun. Therefore a clear understanding on solar geometry is essential. The energy received by a location largely determines the climate, weather, flora, fauna and seasonal and daily temperature variations. This is why the tropics have a very high bio diversity compared to the higher latitudes.

Orbit of earth around the sun is elliptic with the sun at one focus point. The earth orbits the sun every 365 ¼ days, while spinning about its axis every 24 hours. The diagram shows two characteristic lines, namely line of solstice and the line of apsides. The earth moves through six stages sequentially. Perihelion (nearest point to the Sun) January 2 to January 5. March equinox on March 19, 20, or 21, June solstice on June 20, 21, or 22, the aphelion (furthest point from the Sun) from July 3 to July 5, the September equinox on September 22, 23, or 24. December solstice on December 21, 22, or 23.

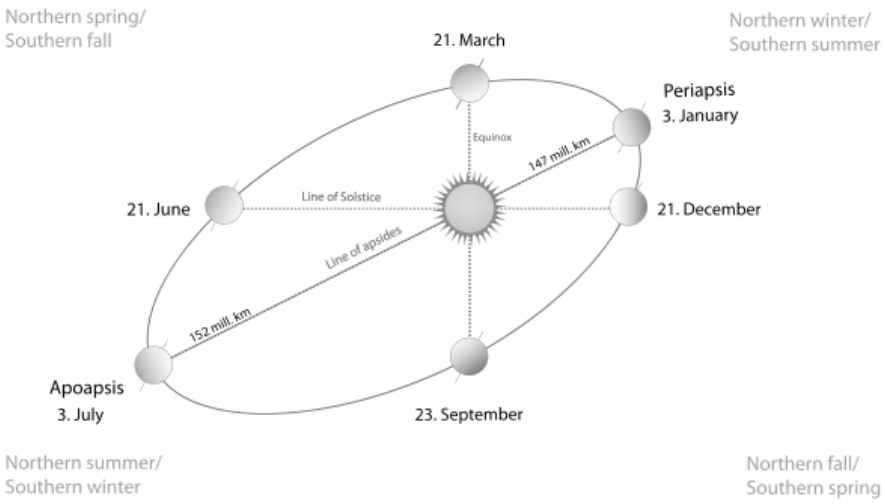


Figure 2-2 Earth's orbit relative to the sun

(Source: www.greenwitchmeantime.com)

The angle solar radiation makes with the surface affects the amount of energy received. This angle is dependent on the inclination of the surface to the tangent to the earth's surface, and the tilt of the earth. Solar declination angle (δ) illustrated in the figure is the angle between a plane perpendicular to incoming solar radiation and the rotational axis of the earth. It is a measure of the tilt of the earth. The earths' axis of rotation has an inclination of 23.5^0 from the ecliptic axis, normal to the ecliptic plane. The ecliptic plane is the plane of orbit of the earth around the sun. The solar declination angle varies from +23.45° on June 21 when the earths' axis tilts toward the sun, to -23.45° on December 21 when the earths' axis tilts away from the sun. The solar declination angle is 0° on equinox dates which are March 21 and September 21.

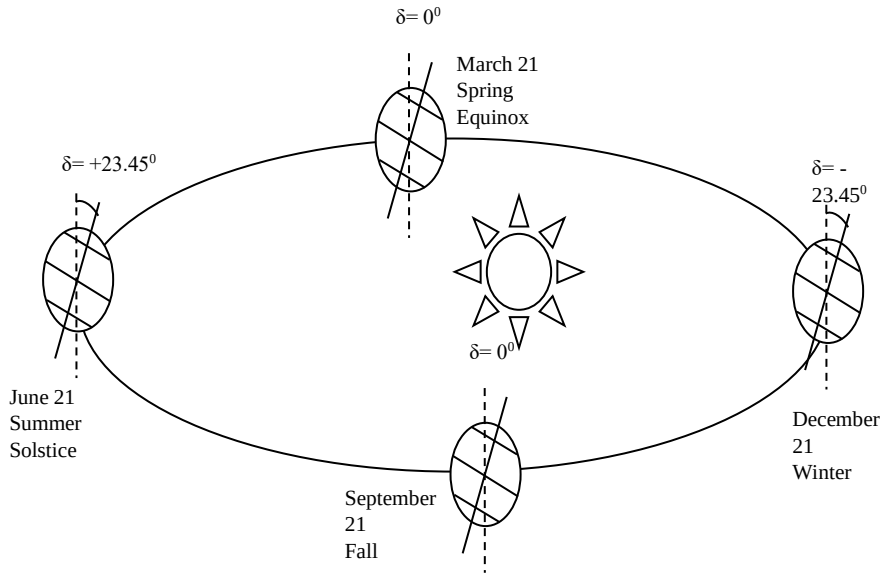


Figure 2-3 Variation of solar declination

Solar declination angle varies as the earth revolves around the sun. The variation causes cyclic changes in solar radiation. These variations contribute to cyclic weather changes causing seasons and monsoons. The variation of the solar declination throughout the year is shown in figure and the declination angle (δ) in degrees for any day of the year (N) can be calculated approximately by the Equation 2.1.

$$\delta° = 23.45 \sin\left[\frac{360}{365}(284 + N)\right] \tag{2.1}$$

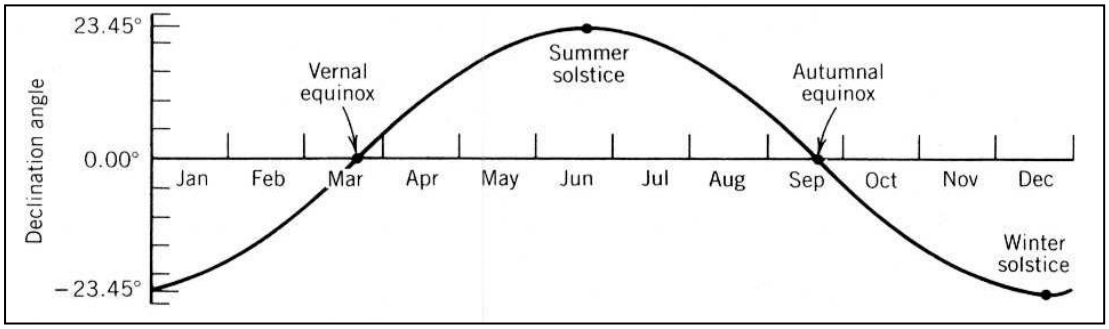


Figure 2-4 Variation of the solar declination angle throughout the year (Source: www.powerfromthesun.net)

A location on the surface of the earth can be specified using the Latitude (ϕ) and the Longitude (L). The Latitude angle (ϕ) is the angle between the radial line connecting the location to the center of the earth and its projection on the equatorial plane. The latitude angle indicates the placement of the location relative to the equatorial plane.

The earth spins about its axis every 24 hours. Therefore rotational speed is 15^0 per hour. The hour angle (ω) is defined as the angle measured on the earth's equatorial plane between the two projections on the equatorial plane of the line connecting center of earth and the location and the line connecting the center of the earth and sun. At solar noon $\omega=0$. Before noon $\omega<0$ *and* after noon $\omega>0$.

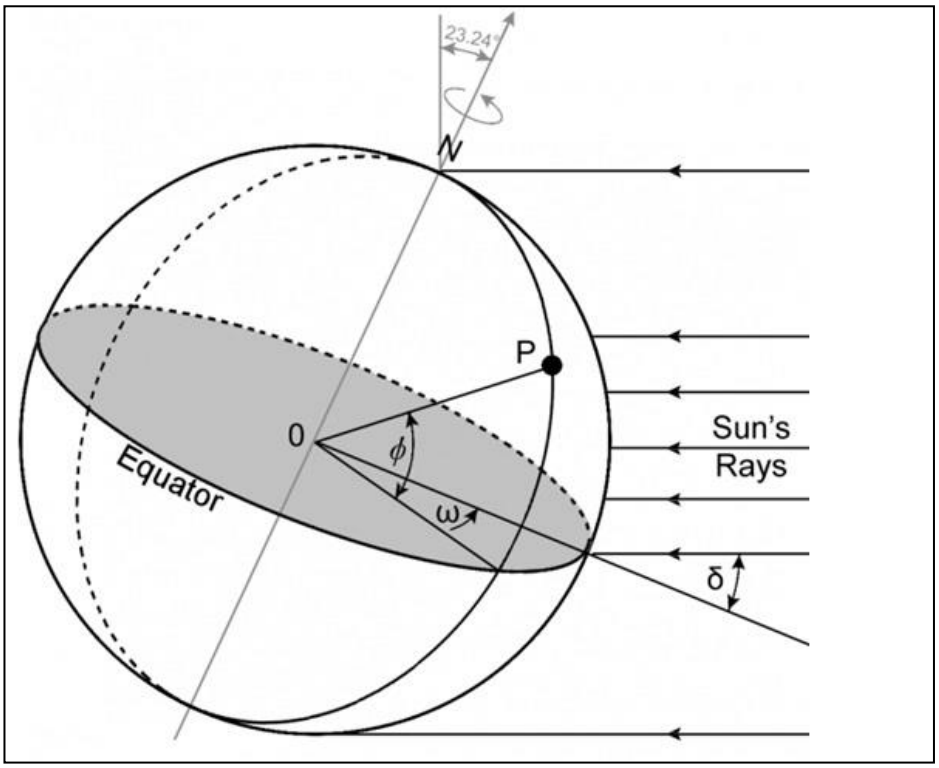


Figure 2-5 The Latitude angle (ϕ) and Hour angle (ω)
Source: www.itacanet.org

The Zenith Angle (θ) is the angle from the observers' zenith point to the sun's position in the sky. The zenith angle can be derived with solar declination angle the hour angle and latitude. The zenith angle is 90 degrees at sunrise and sunset and is minimum at noon

$$\cos\theta_Z = \sin\phi\sin\delta + \cos\phi\cos\delta\cos\omega \tag{2.2}$$

The solar altitude angle (α) is the angle of incidence of sun's rays on a horizontal plane at the specified location. The zenith angle and solar altitude angle are complementary.

$$\theta_Z + \alpha = \frac{\pi}{2} \tag{2.3}$$

The solar azimuth angle (A_z) is the angle is the angle on the tangential plane between the projection of sun rays and the due south on the tangential plane. Westward angle is designated as positive. The mathematical expression for the solar azimuth angle is as follow

$$\sin A_z = \frac{\cos\delta \sin\omega}{\cos\alpha} \tag{2.4}$$

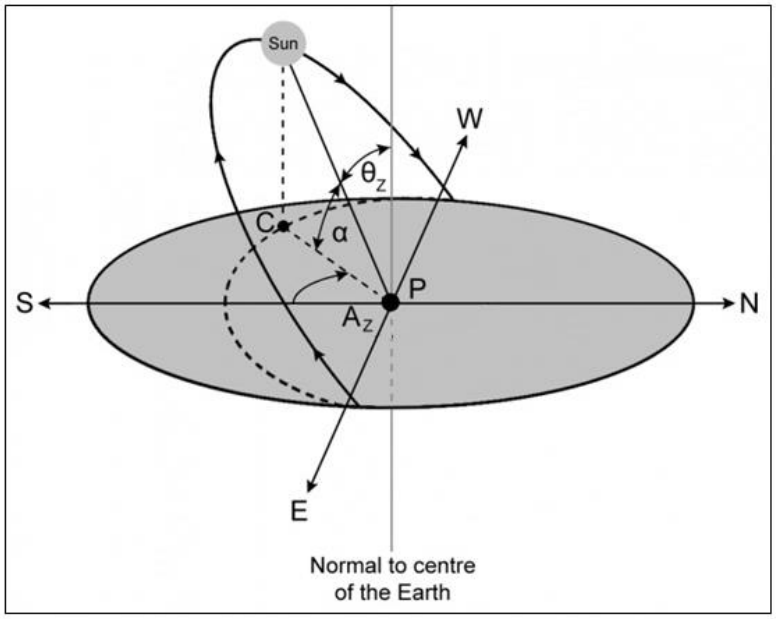


Figure 2-6: Altitude angle (α), Zenith angle (θ_z), and Azimuth angle (A_z)

Source: www.itacanet.org

The solar incidence angle (θ_i) is the angle between the sun's rays and the normal on the striking surface. The angle on incidence depends on the tilt of the surface.

For a horizontal plane angle of incidence angle (θ_i) = zenith angle (θ_z)

For tilted surfaces the angle of incidence can be derived as.

$$\begin{aligned}\cos\theta_i = {} & \sin\phi \sin\delta \cos\beta - \cos\phi \sin\delta \sin\beta \cos Azs \\ & + \cos\phi \cos\delta \cos\omega \cos\beta \\ & + \sin\phi \cos\delta \cos\omega \sin\beta \cos A_{zs} \\ & + \cos\delta \sin\omega \sin\beta \sin A_{zs}\end{aligned} \tag{2.5}$$

Where β is the surface tilt angle from the horizontal and A_{zs} is the surface azimuth angle, the angle between the normal to the surface and true south, with westward designated as positive. Figure 2.7 illustrates the inter relation of the above discussed sun-earth angles.

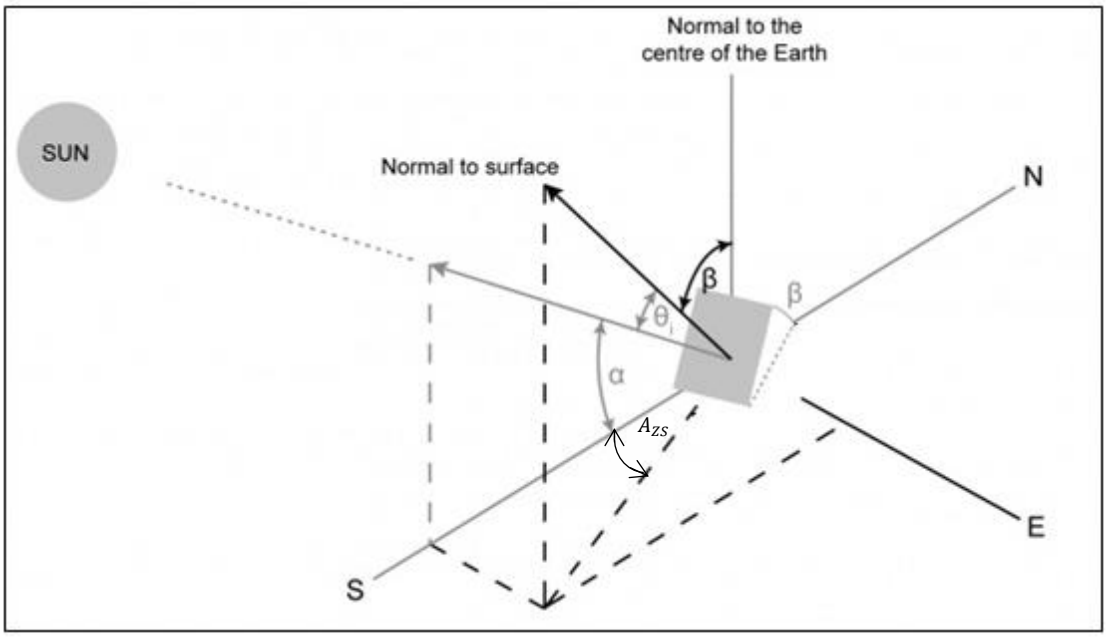


Figure 2-7 : The position of the sun relative to a tilted plane

2.2 Introduction to the solar irradiation fluctuations

Solar electricity is dependent primarily on the solar irradiance, which has periodic and random variations. The periodic variations are caused by the relative motion of the sun that can be explained and predicted using solar geometry and terrestrial solar insolation correlations. The random variations result mostly from cloud cover and seasonal changes. Power output of a solar electricity system will vary with irradiation fluctuations. The sky can be generally categorized to sunny without clouds, sunny with clouds and rainy (overcast).

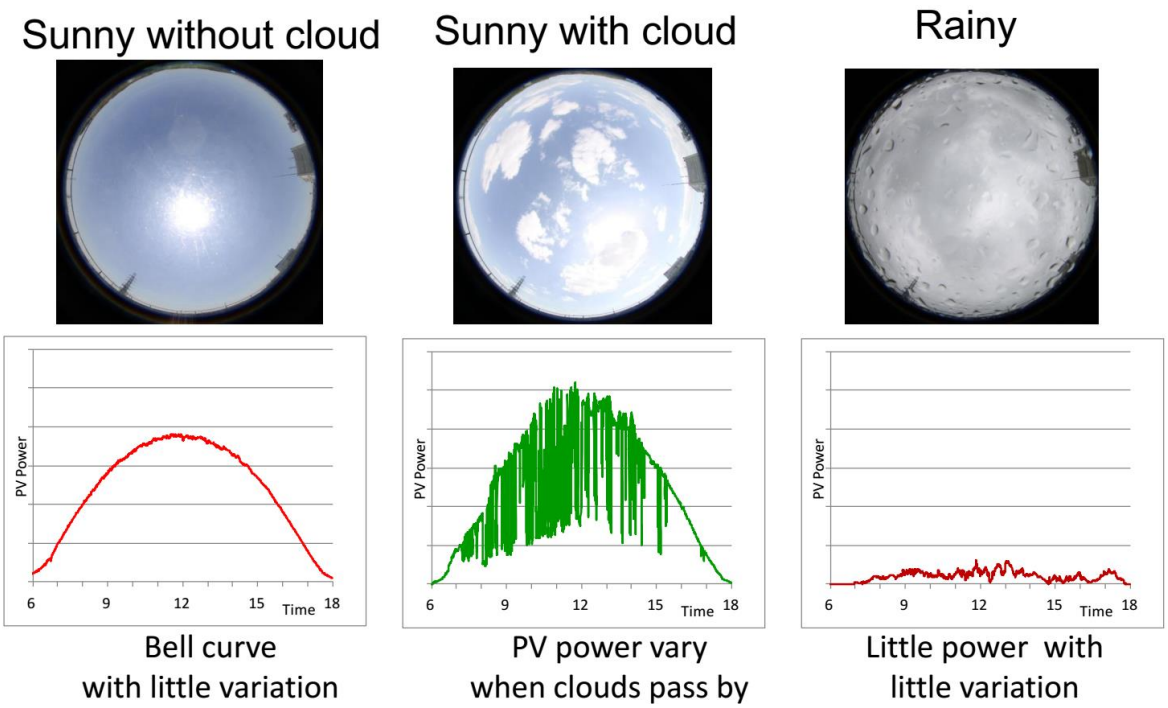


Figure 2-8 Solar PV power generation vs time curves under three sky conditions

The fluctuations are characterized by their time spans. Variations within a time span of a few seconds are caused by moving cloud cover. Though the fluctuations can reach 90% of the output the energy related to them is low due to the small time spans of the fluctuations. Only consistent fluctuations of energy that occur on a larger time period will transfer an effect to the hour scale. Further fluctuations of this type are reduced when number of geographically spread PV systems are connected. The output variability of 1000MW_p of dispersed 4kW residential PV systems have corresponded to a variability of 0.2% of the variability of a 1000MW_p plant

at a single location [7]. Similar results were observed in a study conducted in 2012 Nagoya Japan [8].

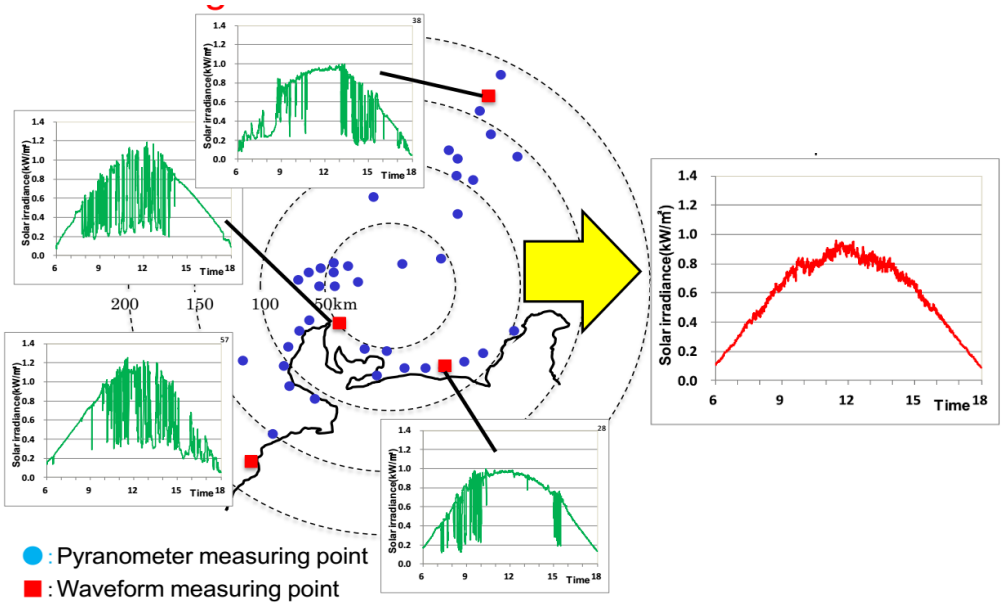


Figure 2-9 Geographical smoothening of collective solar power curve

The effect of overcast skies are rather significant in terms of instant power loss and time span. The effect is translated into a time span in the hour scale and changes the sinusoidal shape of the PV power curve.

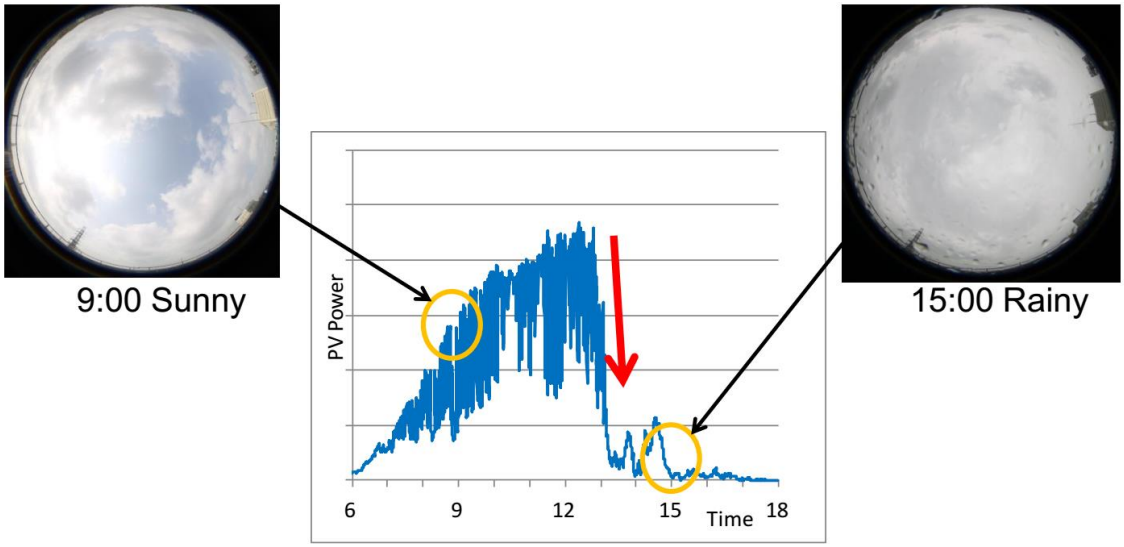


Figure 2-10 Effect of sky condition change on PV power within a day

2.2.1 Weather seasons of Sri Lanka and climatic patterns

Sri Lanka experience two monsoon seasons and two inter-monsoon periods. During all seasons an area of Sri Lanka gets rain. Annual rainfall is 2000mm with a range of 900-5000mm. The seasons have a very significant effect on the solar energy irradiation. It is important to be aware of them.

2.2.1.1 Northeast Monsoon (December to February)

This period is characterized by lesser rainfall than of the other monsoon season. Rainfall is prominent in the North, Eastern slopes of hill country and the Eastern slopes of the Knuckles range. Since the winds blow across Asian landmass it is considerably colder. Therefore holds less moisture and forms less clouds. This period is characterized by cloud-free skies, sunny days and colder nights. During this time the earth passes through its perihelion (Closest point to the sun), the day length is shorter compared to July and August. Colombo receives a low rainfall less than 400mm

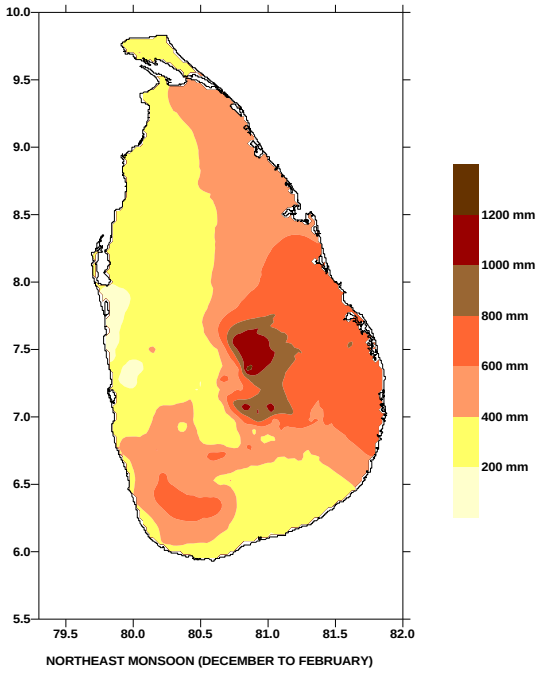


Figure 2-11 Rainfall of NE monsoon 1200mm-200mm

2.2.1.2 First inter-monsoon period (March to April)

Rainfall to Colombo slightly increases, and during this time Sri Lanka moves below the solar plane. Solar energy irradiation is high which cases increase in temperature during day. The high irradiance increases evaporation and promotes cloud formation. This movement effects the wind which were driven by the low pressure zone below Sri Lanka due to energy from the sun. Thunderstorm type of rain can be experienced during evening or afternoon. Rainfall is less than the NE monsoon and the rain in predominantly to the west and south west region of the country. Colombo experience slightly higher rain compared to the NE monsoon coupled with more cloud cover.

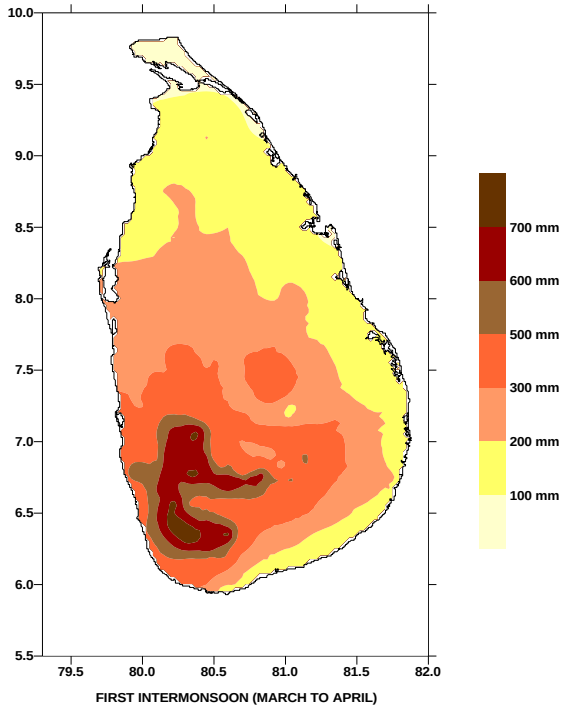


Figure 2-12 Rainfall distribution of first inter-monsoon 700mm -100mm

2.2.1.3 Southwest monsoon (May to September)

The solar plane is above Sri Lanka during this time. The energy influx causes low pressure and drives winds from the south towards the equator. The wind travels over the vast span of ocean, therefore it yields heights amount of rain. Rainfall is concentrated to the western and southwestern regions of the country. The rainfall starts with a peak and reduces exponentially towards the end of the period. Colombo experiences a rainfall in the range of 1500mm to 1000mm. This five month period has several important incidences related to the motion of the sun relative to earth. The sun moves furthest to the north at noon in June. Earth is furthest from the sun in July, and August the sun is directly above Sri Lanka.

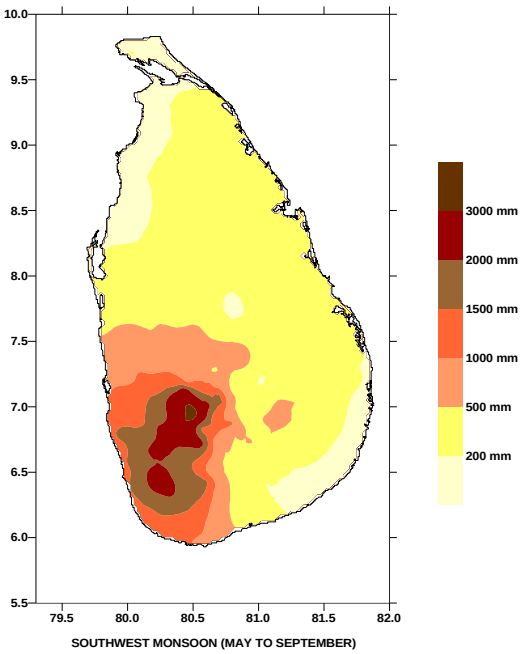


Figure 2-13 Rainfall distribution of SW monsoon 300mm-2000mm

2.2.1.4 Second inter-monsoon season (October to November)

The second inter-monsoon brings higher rainfall compared to the first inter-monsoon. Thunderstorm type of rain can be experienced during evening or afternoon. Rainfall is predominantly to the west and south west region of the country. Colombo experience rainfall in the range of 1500-1000mm, coupled with high cloud cover.

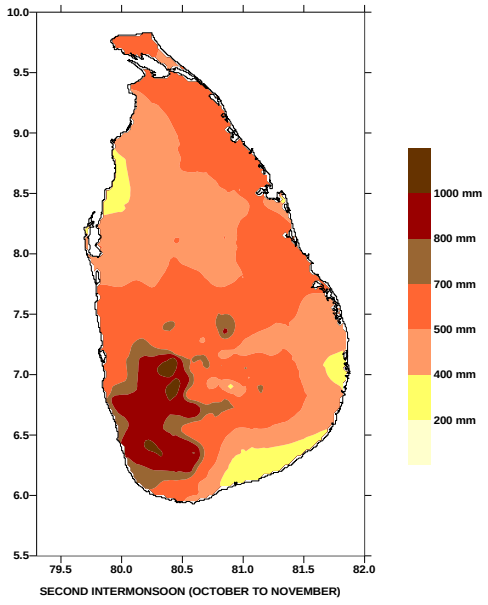


Figure 2-14 Rainfall distribution of inter-monsoon 1000mm-200mm

2.2.1.5 Seasonal Variation of Monthly Averaged Solar Irradiance

With the change in seasons and position of solar path relative to the country Sri Lanka experiences the maximum irradiance in the months of March and April. There is a downward trend in irradiation tom April onwards. During the downward trend there are two local peaks in July Aug period and in October. Each peak is less than the predecessor

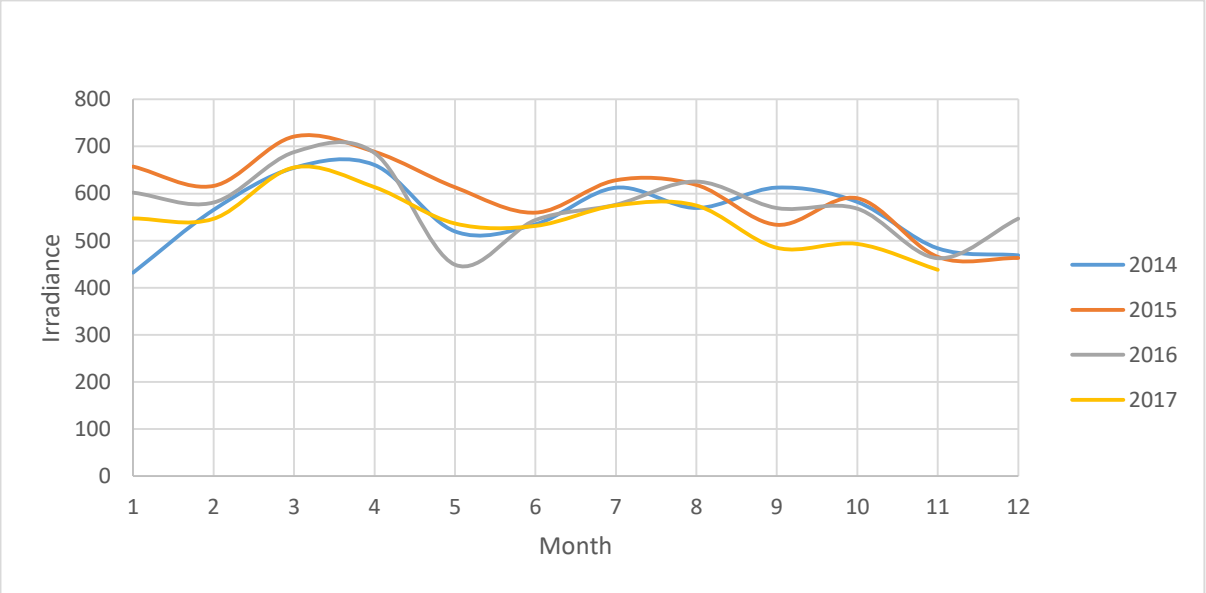


Figure 2-15 Variation of monthly average solar irradiance over the year

2.2.2 Effect of clouds on solar irradiation

2.2.2.1 Introduction to classification of clouds

A cloud is characterize by height, texture, shape, base color and appearance. The world Meteorological organization classifies clouds according to the distance to the distance from the sea level to the base of the cloud. This classification has three types of clouds [9].

2.2.2.2 High level clouds

The cloud base is above 16,500 ft. High level clouds are mostly consisted with ice owing to the low temperatures at that height [9].

2.2.2.2.1 Cirrus clouds

These are white in color and has a fibrous appearance. The strand shapes stretch across the sky. These clouds have low opacity and white in color. Cirrus clouds indicate fair weather in the future [10].

Figure 2-16 Cirrus clouds with fibrous texture source www.metoffice.gov.uk

2.2.2.2.2 Cirrocumulus

Cirrocumulus clouds consist large groups of white clumps called cloudlets that are neatly aligned. Cloudlets are made up of both water and ice. The water is at a temperature below zero degrees. Cirrocumulus clouds are formed when vertical currents meets a cirrus layer and condense creating cloudlets [10].

Figure 2-17 Cirrocumulus clouds www.metoffice.gov.uk

2.2.2.2.3 Cirrostratus (CS)

Cirrostratus are a transparent whitish veil of cloud. They cover a vast area of the sky up to hundred miles. Sun and moon can shine through creating a halo. Cirrostratus can be very small in thickness. Cirrostratus are formed in very stable conditions. Slow rising air caries vapor and condense to form Cirrostratus clouds. This type of rising air is generated by a forefront of a weather system. Their presence and movement of these clouds are an indication of moist weather within the 12-24 hours [9] [10].

Figure 2-18 Cirrostratus clouds www.metoffice.gov.uk

2.2.2.3 Mid-level clouds

The cloud base is in the range of 6500 to 16500 ft. These clouds comprises of mostly of water droplets [9].

2.2.2.3.1 Altocumulus:

Altocumulus are white or grey in color and appear as a sheet or layer of patched clouds with shading. There are composed of rounded masses that are regularly arranged. There is visible contrast within the cloud. These are contrasts are caused by the rounded masses of clouds that are called cloudlets. Altocumulus clouds can be formed by breaking up of Altostratus cloud layers, or under mild turbulent rising air currents. Altocumulus are arranged in an ordered fashion similar to Cirrocumulus. They can be differentiated by considering the fact that Altocumulus clouds have white or grey shaded masses within the cloud and Cirrocumulus are uniform in color [10].

Figure 2-19 Altocumulus clouds www.metoffice.gov.uk

2.2.2.3.2 Altostratus (As)

Altostratus clouds are mid-level clouds and spread as a uniform sheet or layer covering vast spans of the sky. Altostratus has a grey or bluish color and they vaguely lets sun light pass through. It doesn't cause any precipitation but indicates oncoming rain storms. They form from descending cirrostratus clouds that grow in thickness. Altostratus clouds have a tendency to absorb more moisture and grow into Nimbostratus clouds [10].

Figure 2-20 Altostratus clouds www.metoffice.gov.uk

2.2.2.4 Low level clouds

The cloud base is below 6500, contain water, block sunlight and can bring precipitation.

2.2.2.4.1 Stratus cloud (ST)

Stratus in Latin means layer or sheet. Stratus are a grey or white layer of cloud that covers the sky. They don't support a high moisture content. Therefore doesn't contribute to significant rain. Stratus clouds form in stable atmospheric conditions. The water vapor condense when breezes raise cool, moist air over colder land or ocean surfaces. These clouds have a variety of thicknesses and opacity [10].

Figure 2-21 Stratus Clouds www.metoffice.gov.uk

2.2.2.4.2 Stratocumulus (SC)

A stratocumulus cloud is a low level cloud formed by breaking up of a stratus cloud layer. Since stratus clouds ae formed in stable atmospheric conditions, the formation of stratocumulus clouds signifies an instability and the shape of the cloud is not uniform owing to being created from instability. (Base of these clouds are lumpy). Color varies within the cloud, usually grey or whitish sheet or layer composed of rounded masses or rolls. These clouds can produce only a slight precipitation. Rather they are a signal of a larger change on the horizon [10].

Figure 2-22 Stratocumulus Clouds www.metoffice.gov.uk

2.2.2.4.3 Nimbostratus (NS)

This is the standard rain cloud. Nimbus refers to rain and stratus refers to sheet or layer in Latin. They are formed very close to the ground level and base is well below the 6500 ft mark for low level clouds. Definite sign of precipitation. Nimbostratus are grey flat cloud with little color variation [10].

Figure 2-23 Nimbostratus Clouds www.metoffice.gov.uk

2.2.2.4.4 Cumulous

Cumulus clouds are separated from each other. They can be seen as individual cauliflower shaped heaps of clouds. These are the characteristic white fluffy clouds seen on sunny days. These clouds are formed by condensation of vapor in rising convection currents therefore have a heaped shape, which gives it the name cumulus which in Latin means heap. The top of cumulus clouds are brilliant white under the sun, while their base is relatively dark [10].

Figure 2-24 Cumulus Clouds www.metoffice.gov.uk

2.2.2.4.5 Cumulonimbus (CB)

Cumulonimbus loosely translates to heaped rain cloud. It is a very dense heavy cloud that grows vertically. Though base is below 6500ft its top reach to the upper part of the troposphere. The shape of the top take a characteristic anvil shape owing to the high winds at the top flattening the cloud head. Cumulonimbus are associated with heavy rain and thunderstorms. The base of the cloud is often dark like a cumulus cloud.

Cumulonimbus clouds are grown from smaller cumulus clouds. These clouds grow vertically through condensation of moisture in convection currents. The large energy dissipated in condensation adds up and these clouds contain massive amount of energy comparable to numerous atomic bombs. A Cumulonimbus brings isolated short duration heavy rain [9] [10].

Figure 2-25 Cumulonimbus clouds www.metoffice.gov.uk

2.2.2.4.6 Visual evaluation of cloud types

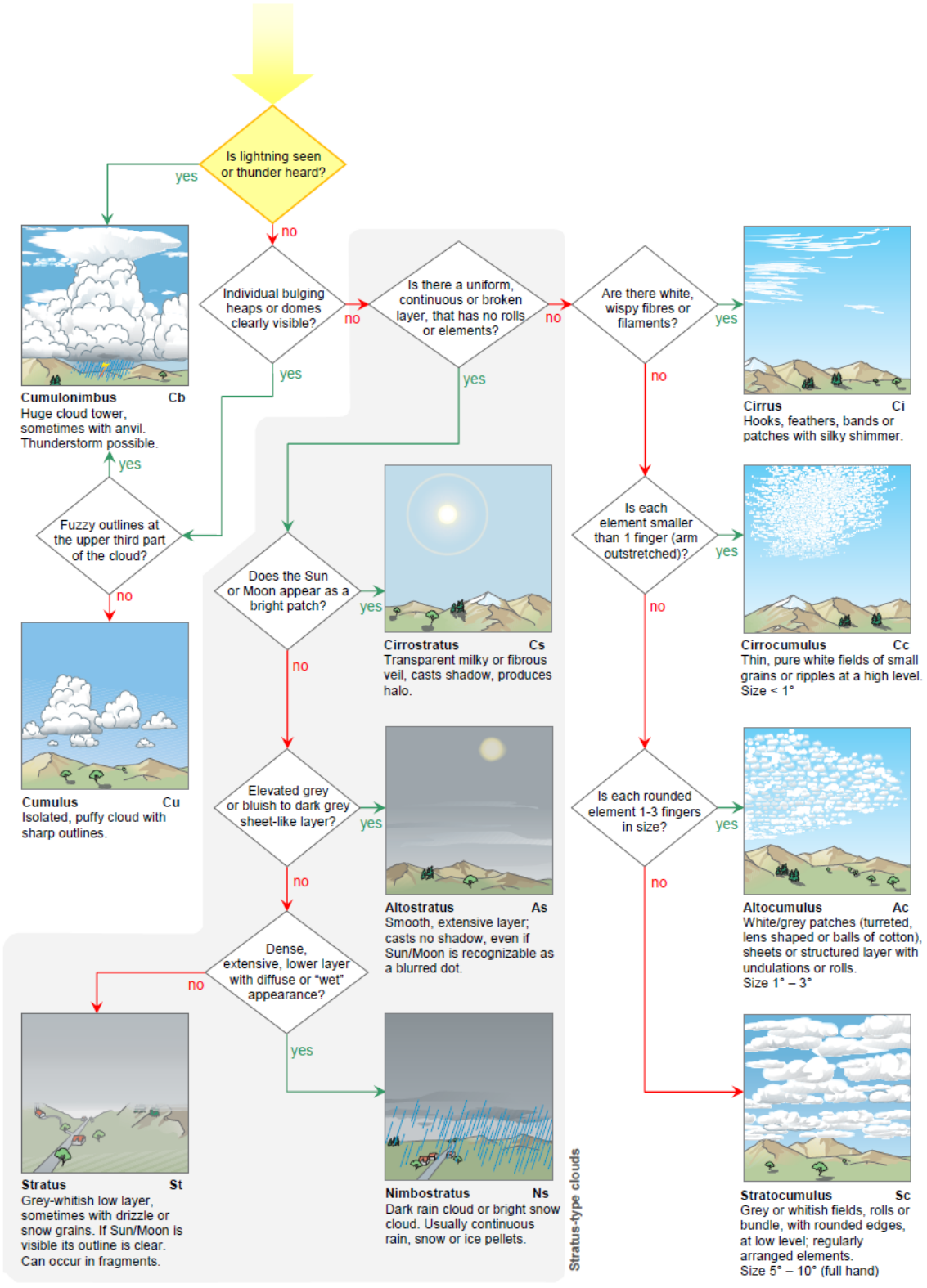


Figure 2-26 Visual evaluation of cloud types, International Cloud Atlas WMO

2.2.2.5 Cloud induced irradiation fluctuations

The effect of clouds on solar irradiance depends on

- Cloud cover
- Cloud type
- Cloud thickness
- Altitude angle of the sun

2.2.2.5.1 Effect of cloud cover on irradiation

Studies prove conclusively that solar irradiance is greatest at sun lit locations with partly clouded skies. It is absolutely vital that the location is not in the shaded area by the cloud. Following table compares the average effect of cloud cover (in Octas) and altitude angle on solar irradiance (W/m2). Such increments are caused by reflection of radiation by clouds. The reflection increases with cloud cover but decreases above 4 Octas as direct radiation passing through is reduced [11].

Table 2-1 Analysis of variation of solar radiation on degree of cloud cover and altitude angle by Dorota Matuszko in Krakow, Poland 2004-2007

Degree of cloudiness (in octas)	Sun's height above the horizon [h°]					
	$h° < 20$	21–30	31–40	41–50	51–60	$h° > 60$
0	235	354	531	668	751	819
1	247	399	525	653	775	825
2	230	301	502	623	725	807
3	228	295	467	612	714	767
4	199	287	423	546	590	752
5	155	272	408	483	582	624
6	152	233	359	425	574	599
7	129	169	311	367	387	475
8	62	109	178	196	219	270

At higher altitude angles and lower cloud cover there is little variation of irradiation. Further at higher altitudes the direct solar irradiation component is large. Therefore as cloud cover reaches 7-8 there is a sharp drop in global irradiance since a significant direct component is eliminated [11].

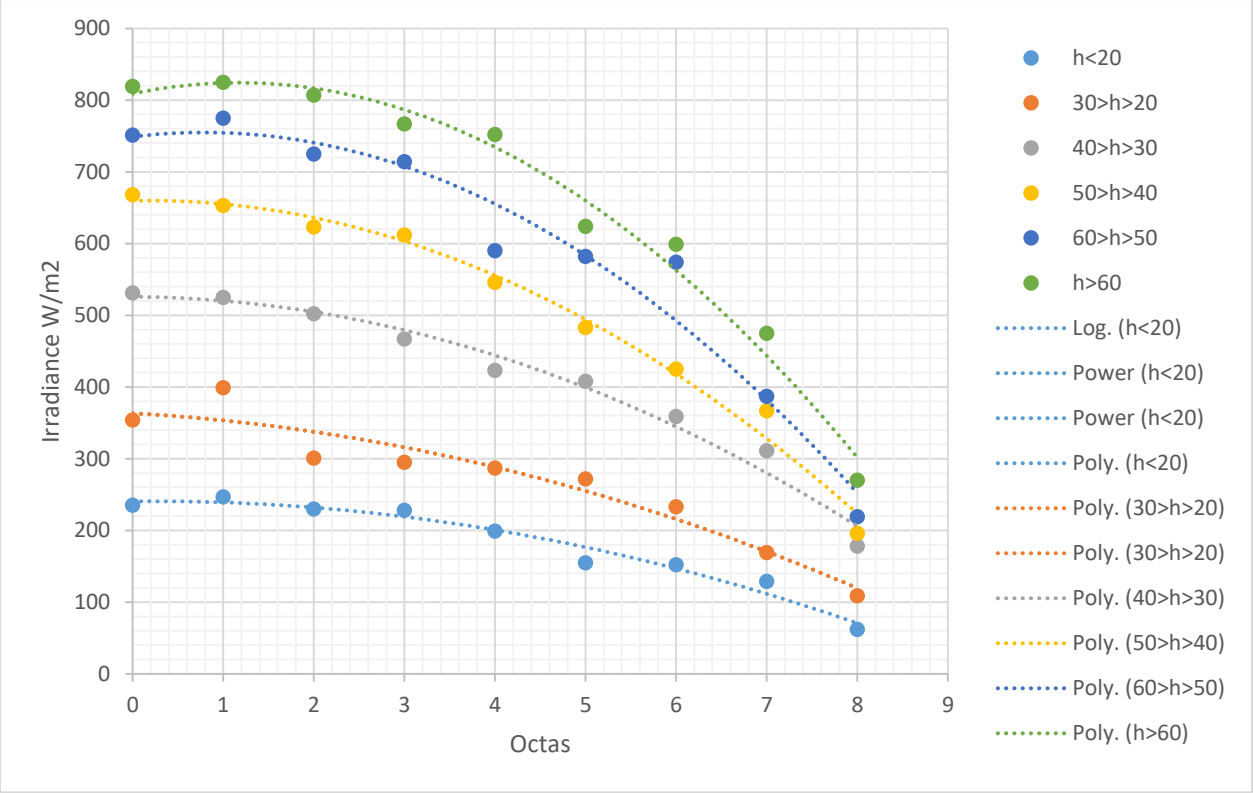


Figure 2-27 Variation of irradiance with cloud cover and solar altitude angle

2.2.2.5.2 Effect of cloud type

The effect of cloud genera on transmission of solar radiation differ and similar trends can be seen for data from around the globe, but exact comparisons and accurate interpretations have proved to be difficult [11].Yet the general trends are of useful knowledge in forecasting.

Table 2-2 Percentage of radiation intensity variation with cloud type and solar altitude angle for Krakow- Poland 2004-2007

Cloud genera	Sun's height above the horizon (h^o)					
	$h^o < 20$	21–30	31–40	41–50	51–60	$h^o > 60$
Cloudless	100	100	100	100	100	100
Ci	98	100	100	100	100	100
Ci + Cs + Cc	90	92	93	94	99	100
Ac	73	79	84	87	90	98
As	58	33	25	24		
Ns	18	24	21	30	7	23
Sc	32	34	28	18	26	12
St	18	22	27	18	15	12
Cu	94	93	84	93	92	93
Cb + Cu	28	25	57	35	60	49

Layered clouds types (stratus) such as Nimbostratus, Altostratus, and Stratocumulus cause the largest deviations in irradiance. Clouds with vertical growth such as cumulus and cumulonimbus clouds cause lower irradiance loss provided the sun's disk is not covered. Further cirrus clouds causes the least drop in global irradiance since they account for a large portion of scattered radiation [11]. To further examine the permeability of different cloud types 24h irradiance data for single genre clouded skies have been compared. Similar studies have been done over different time spans by Robinson P J in 1977; Estupinan J G and Raman S in 1996; Kuchinke C and Nunez M in 1999 [11]. Following table is the results of analysis by Dorota Matuszko for data from Krakow Poland. Permeability is compared to a similar day with equal cloud cover and equal cloud genre.

Table 2-3 Transmission of solar radiation by cloud genre analysis by Dorota Matuszko for Krakow Poland 2004-2007

Day	Degree of cloudiness (octas)	Date	T	R	S (%)
B		22.09.2005	0.7		
Cumulonimbus	7	27.09.2004	0.5	0.2	71.4
B		17.01.2005	0.6		
Nimbostratus	8	18.01.2006	0.1	0.5	16.7
B		21.02.2004	0.7		
Altostratus	8	13.02.2004	0.3	0.4	42.8
B		16.12.2004	0.6		
Stratocumulus	7	08.12.2004	0.1	0.5	16.7
B		01.11.2005	0.6		
Stratus	8	12.11.2004	0.2	0.4	33.3
B		06.03.2004	0.8		
Cumulus	7	10.03.2005	0.5	0.3	62.5
B		01.09.2005	0.7		
Altocumulus	3	29.09.2006	0.4	0.3	57.1
B		12.05.2006	0.7		
Cirrus	2	22.05.2005	0.6	0.1	85.7

2.3 Mathematical models for solar irradiation and PV power forecasting

The distributed nature of solar energy plants coupled with the impact of weather such as rain and clouds have on the consistency of the solar resource, creates unique challenges. The distributed solar generation is invisible to system control. While the load can be forecasted with significant accuracy the inaccessibility to real time energy production data and time variability of the production has caused a rapid rise in solar resource and PV energy production forecasting. There are many companies that have made commercial institutions that have made this their prime venture.

The final outcome of forecasting is the solar energy of a power plant. Depending on the algorithm used this could be a single step process or two step process.

- Direct forecast of solar power output.
- Two step process of solar irradiance forecast and solar power forecast in sequence

Solar power prediction methods are generally characterized as follow Physical models and Statistical models. Each has its merits and complications.

2.3.1 Physical methods of forecast

Physical models explicitly model physical atmospheric phenomenon to predict irradiance using numerical weather prediction (NWP) models or sky images.

2.3.1.1 Numerical weather prediction (NWP)

NWP method uses several global weather models to generate forecasts of atmospheric variables such as rainfall, wind, temperature etc. These parameters are then used to forecast the solar irradiation. NWP has a forecast horizon of 6 hours. NWP require very powerful computers as three-dimensional models of the atmosphere and oceans are used to predict the weather based on current weather. Due to high computational loads the NWP models have relatively coarse resolution, with grid spacing in 50km-90km range. Mesoscale NWP models cover limited area but have higher resolution. Daily weather forecasts are from NWP tools. Since these are numerical models, the result validation is very important. This is done by comparing multiple NWP results. Combining NWP results in know and ensemble forecast and improves accuracy. [12]

2.3.1.2 Cloud imagery

The second group of physical solar power prediction uses cloud images to predict cloud movement, and the predict irradiance into the future at a particular geographic site. Cloud imagery has better performance over NWP forecasts in shorter timeframe under six hours ahead. Total sky imaging (TSI) devices and satellite cloud images are used as methods of gathering cloud data.

A TSI takes and overhead image which is processed to generate a forecast using velocity, opacity, size of shadow of the cloud, etc. A sky imager can cover 5-15 km^2 of sky depending on the cloud presence, therefore can cover multiple solar plants. TSI has a prediction range below 30 mins, and cannot detect clouds beyond this time horizon. TSI method is best for short term power fluctuations. [13]

Figure 2-28 Image of the complete sky take from a TSI system

TSI has better spatial resolution and higher sampling frequency compared to satellite images. Given that TSI images are limited by locations, images from geostationary satellites can be used which are available for the entire globe. The opacity of the clouds in satellite images and positioning in consequent images are used to estimate irradiance. Satellite based forecasting is the best for intraday timeframes of 1 to 6 hrs [13].

Physical models can be further improved with statistical analysis. Historical Similar Mining method is one such method .Numerical weather predictions (NWP) tools coupled with a database of historical weather data and power production of a photovoltaic system is used in historical similar mining method. The model uses the weather forecasting for a defined time frame and corresponding historical real power production to find similar conditions within the

database. The similar historical conditions are used to arrive at a spot prediction for the PV plant output. Recent advancements in NWP have aided in the success of this method

2.3.2 Statistical methods of forecast

Statistical approaches predict irradiance based on past data. The database is used for training and statistical analysis. A persistence forecast is a basic statistical approach to predict the future output from a solar plant. Past time series data is used to forecast solar plant power output in the future, with minor adjustments based on the sun's position in the sky. Purely statistical forecasting is not widely regarded as a modern solar power prediction scheme. However, hybrid approaches make use machine learning or advance statistical techniques to increase performance of past data based forecasting methods.

Statistical approaches to solar power predictions directly forecast power output rather than going through the two steps of forecasting irradiation followed by forecasting power. Their starting point is a training dataset that contains a variety of inputs that could include NWP outputs, ground station or satellite data, historic solar plant production data etc. The dataset is used to train models such as autoregressive or artificial intelligence models that produce a forecast of solar plant output under future conditions. The future conditions could be from physical models like NWP forecasts [13].

Best practices adopt a hybrid of the two approaches. Physical models are used for irradiance forecasting, which is then used as input to solar plant simulation to predict power output. Using past data this forecast is subjected to statistical post-processing to improve performance and suitability to a specific location or power plant. Having multiple stages could create a cascading of forecasting error in to the next. But that is not the case. Each step has multiple forecasts from several models (i.e. weather from several NWP models) and a combined forecast is created, which is passed to the next step. This ensures higher accuracy and removes model deficiencies.

2.3.2.1 Persistence method

This approach is considered as a native predictor. It is extensively used for forecasting meteorological parameters [14], [15].This is also known as moving average method. The parameter in the next time step is considered to be the average of the values within a chosen time window.

For n number of time steps of size k (Sec) total time is T=nk

$$P_{t+k} = \left(\frac{1}{n}\right)(P_t + P_{t-k} + P_{t-2k} + \cdots + P_{t-k} + P_{t-(n-1)k})$$

$$P_{t+k} = \left(\frac{1}{n}\right)(P_t + P_{t-k} + P_{t-2k} + \cdots + P_{t-k} + P_{t-(T-k)})$$

$$P_{t+k} = \left(\frac{1}{n}\right)\sum_{i=o}^{T-k} P(t-i)$$

Moving average is considered to be very effective in predicting solar energy production in short forecasting windows of below few hours [16].

The model accuracy will decrease drastically as the forecasting horizon increase. Moving average is considered the base predictor, and the success of other prediction schemes are evaluated compare to the persistence approach. [17]

2.3.2.2 Auto Regression model (AR)

This model presumes the forecasted parameter values of the previous time steps can have a linear relationship with the parameter value of the immediate time step. The number of time steps can be defined. The regression is similar to multivariable regression. F statistic is evaluated to establish the significant of the model while t statistic is evaluated for the significance of each coefficient of the variables in the linear relationship

$$P_t = a_0 + a_{t-1}P_{t-1} + a_{t-2}P_{t-2} + a_{t-3}P_{t-3} + \cdots + a_{t-k}\,P_{t-k}$$

Moving average model (MA)

Value of a parameter can be expressed linearly using the value of the previous time step. This linear expression would have an error.

$$P_t = \epsilon_1 + a_{t-1}P_{t-1}$$

$$P_{t-1} = \epsilon_2 + a_{t-2}P_{t-2}$$

$$P_{t-2} = \epsilon_3 + a_{t-2}P_{t-2}$$

If the error values are random (normal distribution) these error terms can be used for forecasting. Random error terms are known as white noise terms.

$$P_t = \mu + \theta_1\epsilon_1 + \theta_2\epsilon_2 + \theta_3\epsilon_3 + \cdots + \theta_k\epsilon_q$$

2.3.2.3 Auto regressive moving average (ARMA)

ARMA has two elementary sections namely Auto regressive (AR) and moving average (MA). The predicted variable is considered to be correlated to its past values hence auto regressive. The second portion is the MA method. The p,q constants are known as order of AR model and order of MA model. Which dictates how many previous values are considered in each case.

$$P_t = \left(a_0 + a_{t-1}P_{t-1} + a_{t-2}P_{t-2} + \cdots + a_{t-k}\,P_{t-p}\right) + (\mu + \theta_1\epsilon_1 + \theta_2\epsilon_2 + \cdots + \theta_k\epsilon_q)$$

ARMA is suited for forecasting parameters that vary uniformly about the mean value which is time independent (Stationary time series). This is proved to have better performance when an underlying correlation is present in the time series. ARMA shows better performance than persistence models [17].

2.3.2.4 Auto Regressive Moving Average with Exogenous inputs. (ARMAX)

ARMA method is not facilitated to process exogenous inputs i.e. weather data at each time step. ARMAX method provides this facility it improves the performance of the scheme significantly [18].

Simulations done in USA with actual PV output data as exogenous inputs have yielded satisfactory results in short forecasting horizons of 1-2 hr [19].

Along with AR and MA components another exogenous input is introduced. Following is the formula for the ARMAX method. The input is d_i and number of time steps considered for the exogenous input is considered as b.

$$P_t = \left(a_0 + a_{t-1}P_{t-1} + a_{t-2}P_{t-2} + \cdots + a_{t-k}\,P_{t-p}\right) + \left(\mu + \theta_1\epsilon_1 + \theta_2\epsilon_2 + \cdots + \theta_k\epsilon_q\right) + (\eta_0 + \eta_1 d_1 + \eta_2 d_2 \ldots + \eta_b d_b)$$

2.4 Artificial Neural Networks (ANN)

2.4.1 Introduction

A neural network is a mean of establishing a correlation between several governing and governed parameters based on a database of both predicted and corresponding governing data. The mathematical model takes cue from nature and functions analogous to a network of neurons. Artificial Neural Networks are a powerful tool that brings the learning ability of machines to establishing complex correlations. Most of recent advances in machine learning applications attribute to the advances in processing power of computers that enabled sophisticated neural networks.

The idea initiated in 1943 when mathematician Walter Pitts and neurologist Warren McCulloch published their work on a model of the functioning of neurons. The intention was to understand the functioning of the brain and a simple neural network was modeled as an electric circuit considering the analogues behavior.

Donald Hebb theorized in his book “The Organization of Behavior” (1949) that learning is the consequent strengthening of connections between neurons with repeated practice. As the processing power of computers increased, it became possible to run a model of a neural network.

First real world application was from Stanford University in 1959. A neural network named MADALINE. It was an adaptive filter that could eliminate echoes on phone lines. This technology is still functional to date.

Though research was slow on neural networks, the interest was reinvigorated in 1982 by a scientist named John Hopfield form Caltech with his publication at the National Academy of Sciences USA. He was instrumental in the mathematical analysis of the ANNs. The power of neural networks was identified internationally and research were expedited. During the last decade machine learning and artificial intelligence have grown in leaps and bounds, giving machines the ability to communicate, learn and invent. Notable landmarks in the recent past were the accomplishments of Google AI, Facebook AI and Tesla AI. From simple features like recognizing a picture on social media to self-learning and self-driving cars, ANNs have proved to be a defining point in human civilization.

Key advantage of ANNs is the ability to establish nonlinear correlations with high degree of accuracy [20].

The resulting model is adaptable and will modify with the availability of real-time data. The optimizing functions related to ANN training is based on numerical methods thus suits computer based processing.

2.4.2 Fundamentals of Neural Networks

The figure shows a basic make up of a neuron. The inputs come from synapses and each synapse has a weight. The function of the neuron body is to multiply each input with the corresponding weight and sum up the values. The bias is an input with a unit weight. The result is passed through an activation function. The resulting output can act as the input to a consequent neuron.

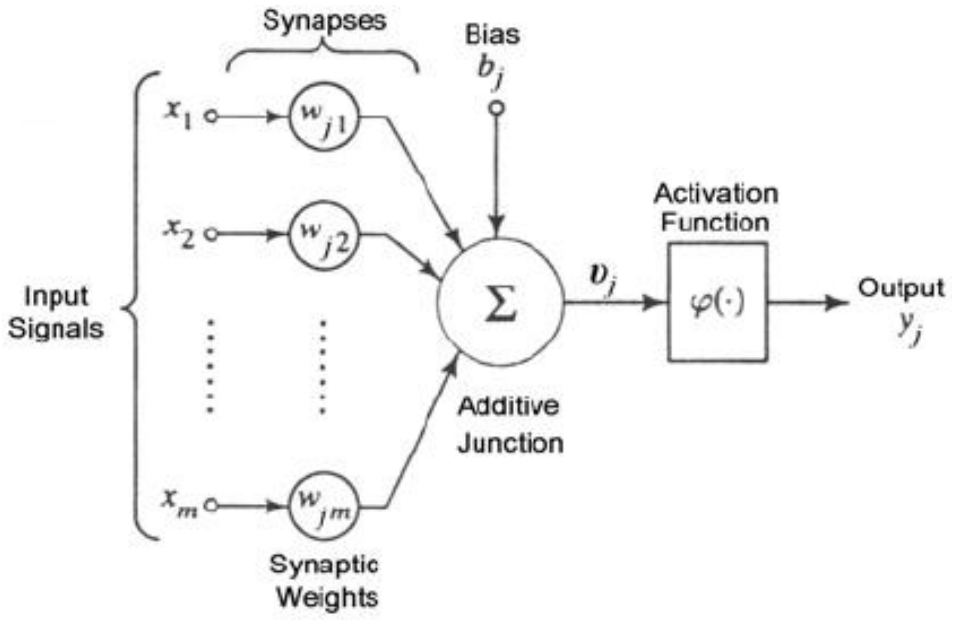


Figure 2-29 Single neuron of an artificial neural network

$$v_j = \sum_i x_i W_{ji} + b_j$$

$$y_j = \phi\,(\mathrm{v_j})$$

The activation function could have several forms

Threshold Function

$$v_j \geq 0 \rightarrow y_j = 1$$

$$v_j < 0 \rightarrow y_j = 0$$

The response in binary in nature. This is also known as the McCulloch-Pitts model named after the scientist who first published the concept of neural network model.

Sign function

$$v_j \geq 0 \rightarrow y_j = 1$$

$$v_j < 0 \rightarrow y_j = -1$$

Liner function

$$v_j \geq l_1 \rightarrow y_j = 1$$

$$l_2 < v_j < l_1 \rightarrow y_j = kv_j$$

$$v_j < l_2 \rightarrow y_j = -1$$

Hyperbolic function

$$y_j = \tanh(av_j)$$

Sigmoid function

$$y_j = \frac{1}{1 + e^{(-av_j)}}$$

$$0 < a < \infty$$

A neural network has an input neuron layer and an output neuron layer. In-between the input and output layers are hidden layers. The number of hidden layers define the type of neural network. An ANN with no hidden layers can only represent linear separable functions. Single hidden layer ANNs can model function with continuous mapping between inputs and outputs. Multi-layer ANNs facilitate deep learning.

2.4.3 Solution of a single hidden layer ANN

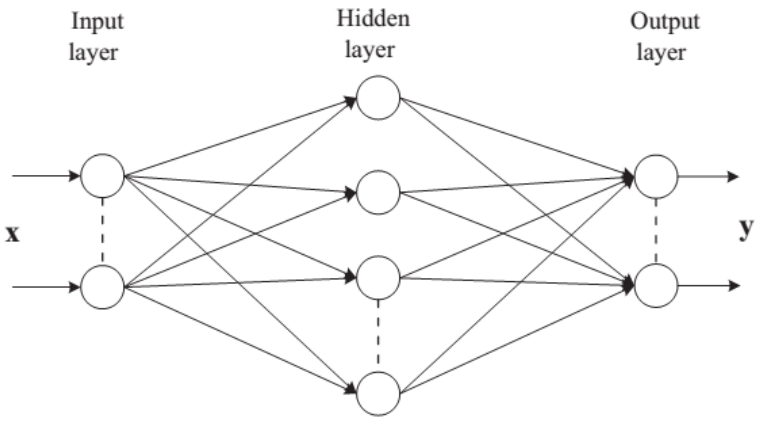


Figure 2-30 Single hidden layer neural network structure

Numbering for an ANN with 2 neuron input layer, 12 neuron hidden layer and a 1 neuron output layer.

$$W^{(layer\ number)}_{staring\ neuron-ending\ neuron}$$

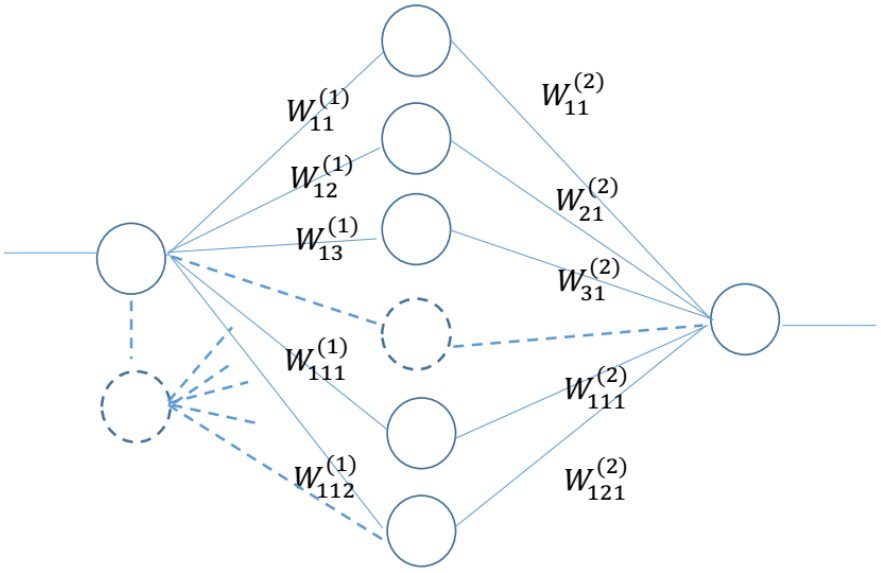


Figure 2-31 Nomenclature of synaptic weights of a neural network

A neural network has ***n*** number of inputs and results in one output. The data set associated with the ANN has an input data matrix which would have ***m instances of n inputs*** and a corresponding target variable data matrix of ***m*** instances. The hidden layer has ***p*** number of neurons

Input data

$$[X]_{m\,X\,n} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ X_{31} & X_{32} & \cdots & X_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}$$

Target data

$$[Y]_{\,mX\,1} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_m \end{pmatrix}$$

Each neuron of the input layer is connected to each neuron of the hidden layer. Resulting in np number of connections, equal number of weights. The weights are expressed in a matrix such that weights related to the neurons of the first layer appear in the same column. The purpose of this is to facilitate the activity of the hidden layer can be expressed as the matrix multiplication of the input matrix and weight matrix.

$$W^{(layer\ number)}_{staring\ neuron-ending\ neuron}$$

$$\left[W^{(1)}\right]_{nX\,p} = \begin{pmatrix} W_{11}^{(1)} & W_{12}^{(1)} & W_{13}^{(1)} & .. & W_{1p}^{(1)} \\ W_{21}^{(1)} & W_{22}^{(1)} & W_{23}^{(1)} & .. & W_{1p}^{(1)} \\ W_{31}^{(1)} & W_{32}^{(1)} & W_{33}^{(1)} & \cdots & W_{1p}^{(1)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ W_{n1}^{(1)} & W_{n2}^{(1)} & W_{n3}^{(1)} & \cdots & W_{np}^{(1)} \end{pmatrix}$$

Activity of the hidden layer is [$Z^{(2)}$]. Each column of the Z matrix has the activity of each neuron of the hidden layer, while each row corresponds to each observation of the input data set.

$$\left[Z^{(2)}\right]_{mXp} = [X]_{m\ X\ n} X \left[W^{(1)}\right]_{nX\ p}$$

Output of the hidden layer is gained by passing the activity matrix Z through activation function

$$\left[a^{(2)}\right]_{mX\ p} = \phi\left(\left[Z^{(2)}\right]_{mXp}\right)$$

Similar steps can be taken for the output layer for which the output matrix $\boldsymbol{[a^{(2)}]_{mxp}}$ is the input. The weight matrix will be of size px1 taking into account that there is only one neuron in the output layer.

$$\left[W^{(2)}\right]_{pX\ 1} = \begin{pmatrix} W_{11}^{(2)} \\ W_{21}^{(2)} \\ W_{31}^{(2)} \\ \vdots \\ W_{p1}^{(2)} \end{pmatrix}$$

Activity of the output layer is [$Z^{(3)}$]. Since there is only one output neuron Z is a column matrix. Each row of [$Z^{(3)}$] corresponds to each observation of the input data set.

$$\left[Z^{(3)}\right]_{mX1} = \left[a^{(2)}\right]_{m\ X\ p} X \left[W^{(2)}\right]_{pX\ 1}$$

The i $^{\boldsymbol{th}}$ term of $Z^{(3)}$ is

$$Z_i^{(3)} = \sum_{k=1}^{p} a^{(2)}(i,k) * W^{(2)}(k,1)$$

Output of the Output layer is gained by passing the activity matrix Z through activation function

$$\left[\hat{Y}\right]_{mX\ 1} = \phi\left(\left[Z^{(3)}\right]_{mX1}\right)$$

We can see that the output of the network is dependent of the weights of each synapse. The ANN has ***(np+p)*** number of weights. By optimizing these weights we can reduce the error between the output of the ANN [$\hat{Y}$] and the target data [Y]. The error is calculated as the

difference between the output and target data. This error is used to create a cost function, which can be optimized with respect to the values of the weights.

$$J = \sum_{i=1}^{m} \frac{\left(Y_i - \widehat{Y}_i\right)^2}{2}$$

With the large number of weights, the optimization can be very tedious and processing power consuming. Even modern super computers would not be able to follow a brute force algorithm by considering all the combinations between the weights to evaluate the one with the lowest error. Therefore other algorithms have been used to achieve the optimizations. Gradient decent method is the most widely used algorithm for this purpose.

$$dJ = \frac{\partial J}{\partial W^{(1)}} dW^{(1)} + \frac{\partial J}{\partial W^{(2)}} dW^{(2)}$$

$$\frac{\partial J}{\partial W_{11}^{(2)}} = \frac{\partial \sum_i^m \frac{\left(Y_i - \widehat{Y}_i\right)^2}{2}}{\partial W_{11}^{(2)}} = -\sum_{i=1}^{m} (Y_i - \widehat{Y}_i) \frac{\partial \widehat{Y}_i}{\partial W_{11}^{(2)}}$$

Since

$$\left[\widehat{Y}\right]_{mX\,1} = \phi\left(\left[Z^{(3)}\right]_{mX1}\right)$$

$$\frac{\partial J}{\partial W_{11}^{(2)}} = -\sum_{i=1}^{m} (Y_i - \widehat{Y}_i) \frac{\partial \phi(Z_i^{(3)})}{\partial Z_i^{(3)}} \frac{\partial Z_i^{(3)}}{\partial W_{11}^{(2)}}$$

The vertical position of Z elements correspond to each observations. Therefore ith term stands for the ith observation.

Since the i th term of $Z^{(3)}$ is

$$Z_i^{(3)} = \sum_{k=1}^{p} a^{(2)}(i,k) * W^{(2)}(k,1)$$

For $W_{11}^{(2)}; k = 1$

$$\frac{\partial Z_i^{(3)}}{\partial W_{11}^{(2)}} = a^{(2)}(i,1)$$

$$\frac{\partial J}{\partial W_{11}^{(2)}} = -\sum_{i=1}^{m}(Y_i - \widehat{Y_\iota})\frac{\partial \phi(Z_i^{(3)})}{\partial Z_i^{(3)}} a^{(2)}(i,1)$$

Similarly for W_{21} $^{(2)}$

$$\frac{\partial J}{\partial W_{21}^{(2)}} = -\sum_{i=1}^{m}(Y_i - \widehat{Y_\iota})\frac{\partial \phi(Z_i^{(3)})}{\partial Z_i^{(3)}}\frac{\partial Z_i^{(3)}}{\partial W_{21}^{(2)}}$$

$$\frac{\partial J}{\partial W_{21}^{(2)}} = -\sum_{i=1}^{m}(Y_i - \widehat{Y_\iota})\frac{\partial \phi(Z_i^{(3)})}{\partial Z_i^{(3)}} a^{(2)}(i,2)$$

For W_{p1} $^{(2)}$

$$\frac{\partial J}{\partial W_{p1}^{(2)}} = -\sum_{i=1}^{m}(Y_i - \widehat{Y_\iota})\frac{\partial \phi(Z_i^{(3)})}{\partial Z_i^{(3)}} a^{(2)}(i,p)$$

$$\begin{pmatrix} \frac{\partial J}{\partial W_{11}^{(2)}} \\ \frac{\partial J}{\partial W_{21}^{(2)}} \\ \frac{\partial J}{\partial W_{31}^{(2)}} \\ \vdots \\ \frac{\partial J}{\partial W_{p1}^{(2)}} \end{pmatrix} = (-)\begin{pmatrix} (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} a^{(2)}(1,1) + (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} a^{(2)}(2,1) + \cdots + (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} a^{(2)}(m,1) \\ (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} a^{(2)}(1,2) + (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} a^{(2)}(2,2) + \cdots + (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} a^{(2)}(m,2) \\ (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} a^{(2)}(1,3) + (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} a^{(2)}(2,3) + \cdots + (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} a^{(2)}(m,3) \\ (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} a^{(2)}(1,4) + (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} a^{(2)}(2,4) + \cdots + (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} a^{(2)}(m,4) \\ (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} a^{(2)}(1,5) + (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} a^{(2)}(2,5) + \cdots + (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} a^{(2)}(m,5) \\ \vdots \\ (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} a^{(2)}(1,p) + (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} a^{(2)}(2,p) + \cdots + (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} a^{(2)}(m,p) \end{pmatrix}$$

Let [δ] be a mx1 matrix with that indicates a scaled error for each observation

$$[\delta]_{mX1} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{pmatrix} = -\begin{pmatrix} (Y_1 - \widehat{Y_1})\,\phi_1^{(3)\prime} \\ (Y_2 - \widehat{Y_2})\,\phi_2^{(3)\prime} \\ \vdots \\ (Y_m - \widehat{Y_m})\,\phi_m^{(3)\prime} \end{pmatrix}$$

$$\begin{pmatrix} \frac{\partial J}{\partial W_{11}^{(2)}} \\ \frac{\partial J}{\partial W_{21}^{(2)}} \\ \frac{\partial J}{\partial W_{31}^{(2)}} \\ \vdots \\ \frac{\partial J}{\partial W_{p1}^{(2)}} \end{pmatrix} = \begin{pmatrix} a^{(2)}(1,1) & a^{(2)}(2,1) & a^{(2)}(3,1) & \dots & a^{(2)}(m,1) \\ a^{(2)}(1,2) & a^{(2)}(2,2) & a^{(2)}(3,2) & \dots & a^{(2)}(m,2) \\ a^{(2)}(1,3) & a^{(2)}(2,3) & a^{(2)}(3,3) & \dots & a^{(2)}(m,3) \\ \vdots & \vdots & \vdots & & \vdots \\ a^{(2)}(1,p) & a^{(2)}(2,p) & a^{(2)}(2,p) & \cdots & a^{(2)}(m,p) \end{pmatrix} \begin{pmatrix} (Y_1 - \widehat{Y_1})\,\phi_1' \\ (Y_2 - \widehat{Y_2})\,\phi_2' \\ \vdots \\ (Y_m - \widehat{Y_m})\,\phi_m' \end{pmatrix}$$

$$\begin{pmatrix} \frac{\partial J}{\partial W_{11}^{(2)}} \\ \frac{\partial J}{\partial W_{21}^{(2)}} \\ \frac{\partial J}{\partial W_{31}^{(2)}} \\ \vdots \\ \frac{\partial J}{\partial W_{p1}^{(2)}} \end{pmatrix} = \left[\frac{\partial J}{\partial W^{(2)}}\right] = \left(\left[a^{(2)}\right]_{mXp}\right)^T \left([\delta]_{mX1}\right)$$

A similar approach can be taken to evaluate differential coefficient $\left[\frac{\partial J}{\partial W^{(1)}}\right]$ of the cost function for the hidden layer.

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = -\sum_{i=1}^{m} (Y_i - \widehat{Y}_i) \frac{\partial \phi(Z_i^{(3)})}{\partial W_i^{(1)}}$$

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = -\sum_{i=1}^{m} (Y_i - \widehat{Y}_i) \frac{\partial \phi(Z_i^{(3)})}{\partial Z_i^{(3)}} \frac{\partial Z_i^{(3)}}{\partial W^{(1)}}$$

$$Z_i^{(3)} = \sum_{k=1}^{p} a^{(2)}(i,k) * W^{(2)}(k,1)$$

$$\frac{\partial Z_i^{(3)}}{\partial a_i^{(2)}} = {W^{(2)}}^T$$

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = \delta^{(3)} \frac{\partial Z_i^{(3)}}{\partial a_i^{(2)}} \frac{\partial a_i^{(2)}}{\partial W^{(1)}}$$

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = \delta^{(3)} W^{(2)^T} \frac{\partial a_i^{(2)}}{\partial W^{(1)}}$$

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = \delta^{(3)} W^{(2)^T} \frac{\partial a_i^{(2)}}{\partial Z^{(2)}} \frac{\partial Z_i^{(2)}}{\partial W^{(1)}}$$

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = \delta^{(3)} W^{(2)^T} \frac{\partial \phi(Z_i^{(2)})}{\partial Z_i^{(2)}} \frac{\partial Z_i^{(2)}}{\partial W^{(1)}}$$

$$\left[\frac{\partial J}{\partial W^{(1)}}\right] = X^T \delta^{(3)} W^{(2)^T} \frac{\partial \phi(Z_i^{(2)})}{\partial Z_i^{(2)}}$$

Using $\left[\frac{\partial J}{\partial W^{(1)}}\right]$ and $\left[\frac{\partial J}{\partial W^{(2)}}\right]$ the gradient of the cost function at each $W^{(1)}$ and $W^{(2)}$ combination can be assessed. The starting point is an arbitrary set of values for the weights. The initial values are incremented in the direction of the maximum descent of the error. The gradient is evaluated for the new set of weights and the process is repeated until the minimum error is reached.

2.4.4 Training validation and testing an artificial neural network

A neural network can learn in many methods. Such as Error correlation learning (Back propagation by gradient decent method, Memory based learning, Hebbian learning, competitive learning, Boltzmann learning [21] . Each method has types of problems they are suited for. Back propagation of error is the most widely used method for forecasting.

Training or learning phase of a neural network is the process of optimizing the weights of the synapses based on past data, such that the error is minimum. The method is well defined and the current computer processors can execute the numerical methods at great speed. A common misconception is that accuracy of a model is the most difficult aspect to achieve. In reality with sufficient well defined data an ANN can reach 99% of accuracy with ease. This is so prevalent that model overfitting the data is extreme that should be avoided in the training process. A model that over fits the data is less likely to give and accurate value for a new data point. The ability to extend the model to new data accurately is known as generalizability of a model.

Generalizability is a very important characteristic of performance of an ANN, and ensures the purpose of its existence.

Validation of an ANN is assuring generalizability of the model. There are many ways of attaining this purpose. The most common method of maintaining generalizability is by using a randomly chosen portion of the historical data to evaluate the model at each optimizing cycle of synapse weights. This portion of data is known as the validation data set and it's not used in the process of training. With each optimization cycle the error for the training data and validation data will drop. Overfitting is identified when the training error drops and validation error increase for a set number of cycles. The default in Matlab neural networks is six cycles.

The k-fold method is an extension of the random selection of validation data. The data is first partitioned into k number of randomly chosen sets of equal size. As in the random method one set is used to validate and (k-1) number of sets are used to train. This process is repeated k times until energy subset has become the validation set.

Another set of randomly chosen data is set aside as test data. This data set is not used for training or validation. It is used to establish the goodness of fit. If the test data and model output are significantly different there are some parameters that should be included as inputs

2.4.5 Applications of neural networks in similar scope

Artificial neural networks have been successfully implemented in the scope of solar energy forecasting [22] .A study was carried out to predict solar irradiation for upcoming 24 hrs based on the irradiation of last 24 hrs. This is a direct forecasting of irradiance using past irradiance as inputs. A multi-layer (two hidden layers) ANN was trained on NREL Solar Radiation Research Laboratory (SRRL) data. The model is able to avoid overfitting and predict irradiation accurately [23]

Forecasting power output of grid connected solar plants is done by predicting irradiance. A study was carried out to use an ANN with multilayer perceptron model which had a forecast horizon of 24hrs. Inputs to the model were solar mean irradiation and ambient temperature of the past hours. Data set was in the time frames of July 1st 2008 to May 23rd 2009 and November 23rd 2009 to January 24th 2010. The location of the study Trieste, Italy (latitude 45 40'N, longitude 13 46'E). The model reached correlation coefficients of 98-99% for sunny days and 94-96% for cloudy days [24].

There is research carried out on including the Aerosol Index in the input data to a back propagation ANN along with irradiation data of previous hours. It has been shown that with the aerosol index the ANN can produce better results for a forecasting horizon of 24 hrs [25].

The relationship between meteorically parameters and solar irradiance could be different in different types of weather. This concept has been tried out where three separate neural networks were trained for three weather conditions i.e. Sunny, partly cloudy and overcast. The data test site was from southern Italy,

Advance neural networks like Bayesian neural networks (BNN) have been incorporated to estimating daily global solar irradiation. The input set comprises of weather data such as ambient temperature, humidity, sunshine hours etc. BNN have shown higher performance over the classical neural networks and empirical correlations [26]. Wavelet analysis is used to preprocess the sample data set for the ANN. Accuracy of forecast can be improved by wavelet analysis prior to training of the neural network [27].

2.4.6 Methods of evaluating model performance

There are numerous methods of evaluating performance of a model. The predominantly used steps are

- Test for correlation coefficient
- Test for error percentage distribution statistics
- Test for randomness of error

2.4.6.1 Test for correlation coefficient and coefficient of determination (R^2)

The regression coefficient is a very important and basic statistic used for this purpose. It is a measure of the amount of variance captured by the model compared to the actual.

Observation -y_i

Model fit -f_i

Mean of observed data:

$$\bar{y} = \frac{1}{n}\sum_{1}^{n} y_i$$

Sum of squares (SS)

Total sum of squares (SS_{tot}):

$$SS_{tot} = \sum_{1}^{n} (y_i - \bar{y})^2$$

Reggression sum of squares (SS_{reg}):

$$SS_{reg} = \sum_{1}^{n} (\bar{y} - f_i)^2$$

Residual sum of squares (SS_{res}):

$$SS_{res} = \sum_{1}^{n} (y_i - f_i)^2$$

$$SS_{tot} = SS_{reg} + SS_{res}$$

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - (SS_{res}/SS_{tot})$$

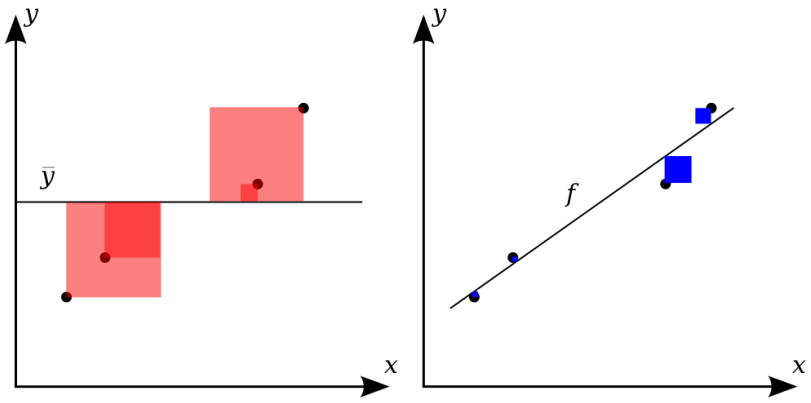


Figure 2-32 Total sum of squares and regression sum of squares comparison

The coefficient of determination in following ranges imply the nature of the correlation [28]

$R^2 < 0.3$ *None or very werak correlation*

$0.3 < R^2 < 0.5$ *weak or low correlation*

$0.5 < R^2 < 0.7$ *Moderate correlatoin*

$R^2 > 0.7$ *Strong correlation*

Following diagram shows the successful implementation of a linear model for the observed data.

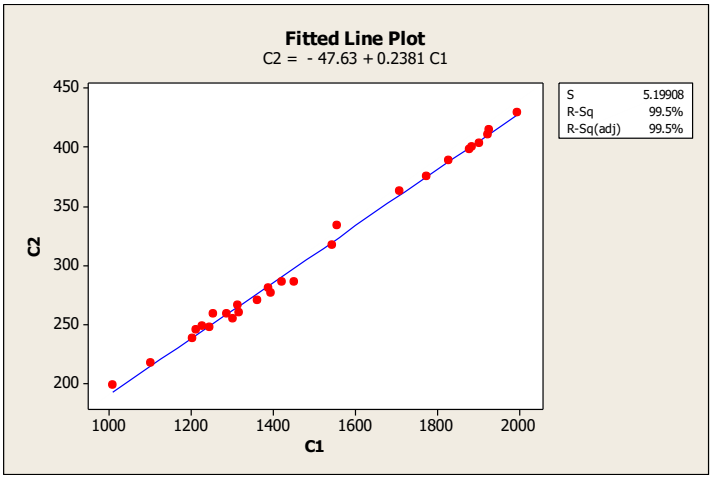


Figure 2-33 Linear model fit with R2>99%

2.4.6.2 Test for error percentage distribution statistics

Percentage error is a measure of accuracy of the model. For large data groups mean error percentage is considered.

Residual = Observation – model

Error% = Residualx100/observed

$-10\% < Mean\ Error\ \% < 10\%$

2.4.6.3 Test for randomness of errors

If the model is exhaustive of all relationships between the input parameters and output parameters there should not be any correlation among the residual terms. If the residual terms are showing a pattern i.e. sinusoidal, it means a sinusoidal relationship has been left out of the model, therefore making the model incomplete. The first method to graphically analyze the error terms. They should be random without any noticeable pattern.

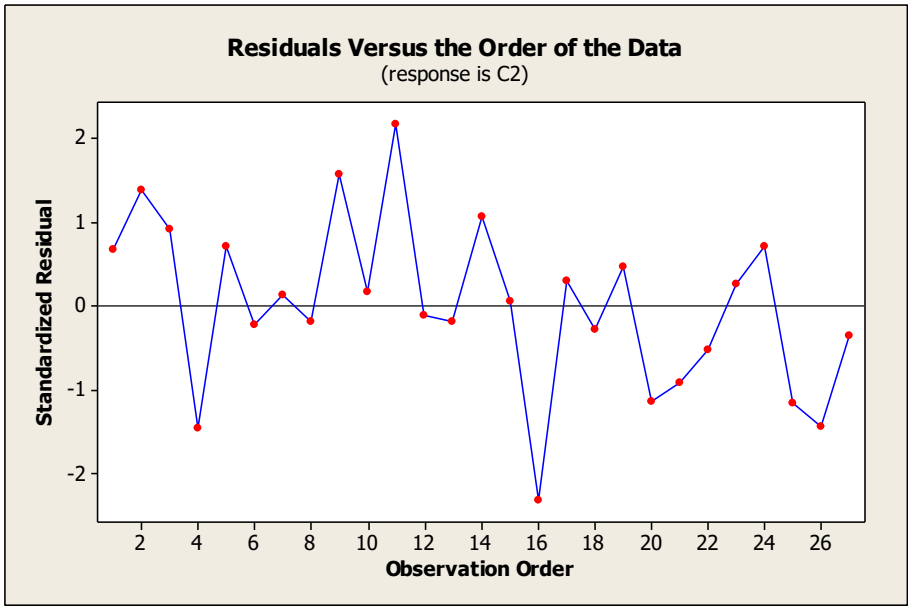


Figure 2-34 Residuals appear random without a pattern

Randomness of error suggests that there is no autocorrelation within the error terms.

Durbin-Watson statistic is a method of establishing the presence of autocorrelation in a data set.

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n}(e_i)^2}$$

Durbin Watson statistic is larger than 1 and closer to 2 suggest the errors are not auto correlated.

Randomly distributed data would show a normal distribution. Therefore the mean of residuals will be zero and the histogram should be symmetric about zero.

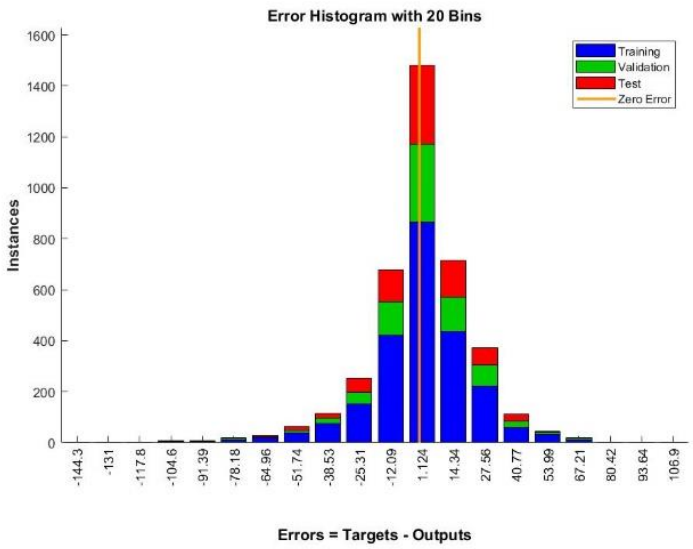


Figure 2-35 Histogram of errors with peak frequency at zero

The QQ plot compares the Z score of a distribution with the Z scores of a standard normal distribution. It a measure to evaluate if the error distribution is normal. The Z scores of the error terms closely follow the standard normal distribution values.

Chapter 3 Research Approach

3.1 Introduction

The study is aimed at developing a relationship between selected meteorological parameters to solar electricity generation of a system. An artificial neural network was employed to predict the fluctuations caused by weather conditions in solar electricity generation. Initial step of identifying inputs to the model was by statistically analyzing the correlation between meteorological parameters and solar electricity generation. Data was collected from Meteorological Department of Sri Lanka for Colombo (6.9° N, 79.8° E). The data available were hourly measurements. Therefore the complete model is in hourly based variations. Training and validation of the neural network was done on a data set from 2016 January to December.

3.2 Development and training of a neural network

3.2.1 Identification of input parameters

The fluctuations in hourly scale in solar electricity is attributed to irradiance, rainfall and cloud cover. The significance of each input is analyzed statistically.

3.2.1.1 Statistical analysis

Table 3-1 Monthly total energy variation with monthly rainfall average cloud cover and monthly insolation

Month	Energy kWh	Rain mm	Avg cloud cover	Solar radiation(MJ/m2)
Jan	36586.80	848.90	4.21	582.27
Apr	42036.02	2416.70	5.00	688.61
May	29220.16	9781.20	6.48	457.09
Jun	34395.80	1623.70	6.26	531.89
Jul	36652.23	639.60	6.00	579.70
Aug	38214.81	9.50	5.74	625.67
Sep	36657.91	377.00	5.95	584.41
Oct	31311.99	3040.70	5.77	523.56
Nov	31661.73	5262.40	6.28	463.60
Dec	36411.91	2146.30	4.76	555.34

3.2.1.1.1 Correlation between monthly total rain fall to monthly total solar energy

Regression Statistics	
Multiple R	0.713985202
R Square	0.509774869
Adjusted R Square	0.448496727
Standard Error	2793.585968
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	64922751.61	64922751.61	8.319032806	0.020377438
Residual	8	62432980.5	7804122.562		
Total	9	127355732.1			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	37691.1166	1207.942748	31.2027343	1.21027E-09	34905.59563	40476.63757
X Variable 1	-0.908812688	0.315092425	-2.884273358	0.020377438	-1.635417123	-0.182208252

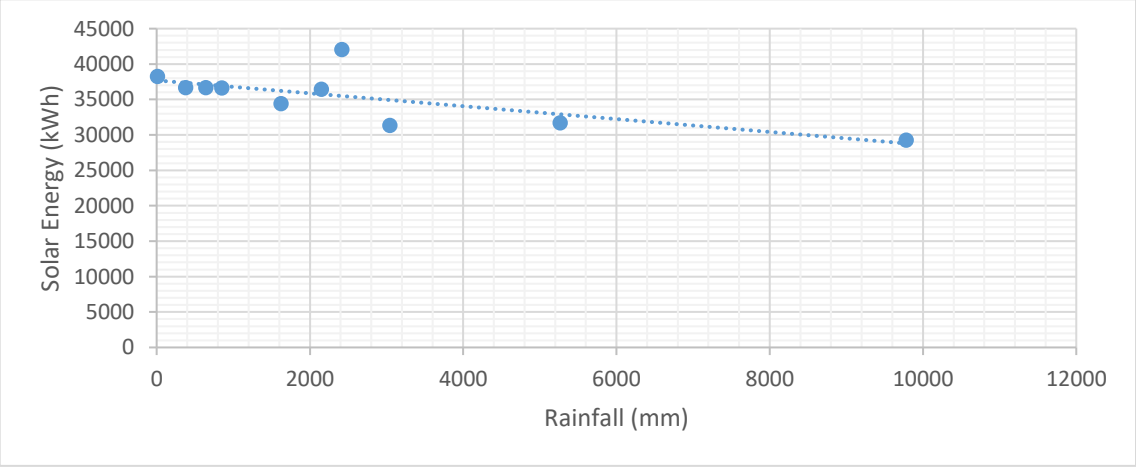


Figure 3-1 Monthly solar PV energy output with monthly rainfall

The correlation coefficient is 0.71. Which is sufficient to assume a correlation between monthly total solar energy to the monthly total rain. Therefore an analysis of variance is done and the F statistic is calculated.

H_0 : Coefficients do noy significantly differ from zero Regression model is not significant

H_1 : Coefficients significantly differ from zero. Regression model is significant

H_0 : Model is not significant

H_1 : Model is significant

From the Anova table it's evident that the significance of the calculated F statistic is 0.02 which less than 5%. Therefore, the null hypothesis is rejected. It can be concluded that under a confidence level of 95% there is a significant correlation between the two variables.

3.2.1.1.2 Correlation between monthly total irradiation to monthly total solar energy

Regression Statistics	
Multiple R	0.962829055
R Square	0.927039789
Adjusted R Square	0.917919762
Standard Error	1077.723359
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	118063831	118063831	101.6488054	7.98523E-06
Residual	8	9291901.113	1161487.639		
Total	9	127355732.1			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	6462.017033	2882.017072	2.242185549	0.05523676	-183.9262537	13107.96032
X Variable 1	51.59548565	5.117531978	10.08210322	7.98523E-06	39.79443575	63.39653555

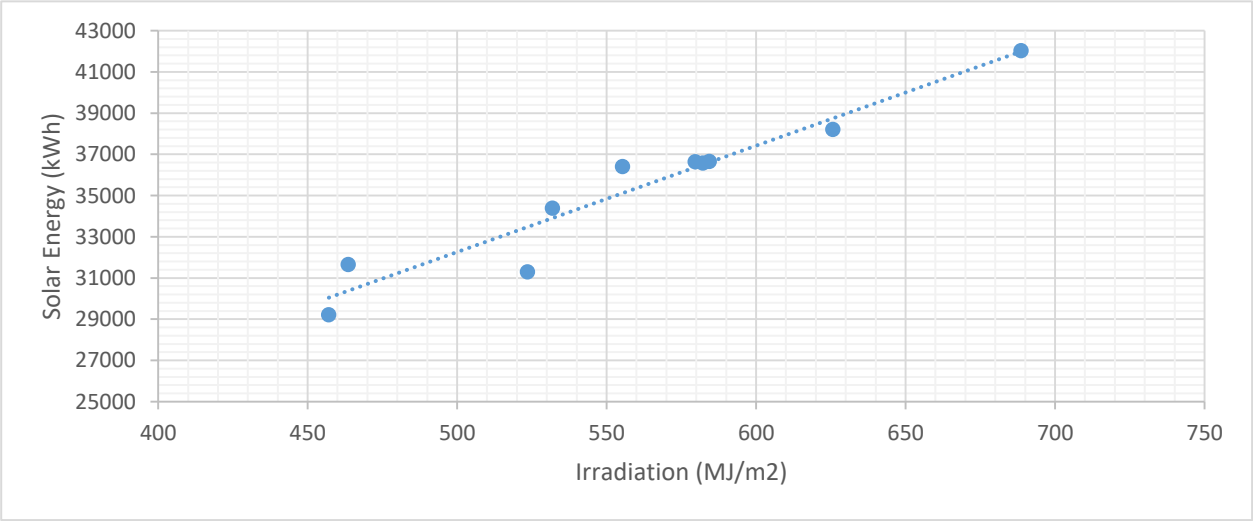


Figure 3-2 Solar PV energy output with monthly average irradiation

The correlation coefficient is 0.96. Which is sufficient to assume a correlation between monthly total solar energy to the monthly total rain. Therefore an analysis of variance is done and the F statistic is calculated.

H_0 : Model is not significant

H_1 : Model is significant

From the Anova table it's evident that the significance of the calculated F statistic is 7.98523E-06 which less than 5%. Therefore, the null hypothesis is rejected. It can be concluded that under a confidence level of 95% there is a significant correlation between the two variables.

3.2.1.1.3 Correlation between cloud cover and daily solar irradiation

The effect on monthly solar electricity output obtained from data logged by inverters is not significant compared to other variables. The effect of cloud cover is at lower time scale. Cloud cover is a main cause of hourly and daily solar irradiance fluctuations.

Regression Statistics	
Multiple R	0.726482058
R Square	0.52777618
Adjusted R Square	0.5114926
Standard Error	0.182851528
Observations	31

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.08367008	1.0836701	32.411557	3.70793E-06
Residual	29	0.96960576	0.0334347		
Total	30	2.05327584			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.587283099	0.173825469	9.131476	4.973E-10	1.231770097	1.942796102
X Variable 1	-0.14987741	0.026326082	-5.6931149	3.708E-06	-0.20372029	-0.096034526

The correlation coefficient is 0.72. Which is sufficient to assume a correlation between cloud cover and solar irradiation. Therefore an analysis of variance is done and the F statistic is calculated.

H_0 : Model is not significant

H_1 : Model is significant

From the Anova table it's evident that the significance of the calculated F statistic is 3.70793E-06 which less than 5%. Therefore, the null hypothesis is rejected. It can be concluded that under a confidence level of 95% there is a significant correlation between the two variables.

3.2.1.1.4 Correlation between cloud cover and daily solar energy

Regression Statistics	
Multiple R	0.70447
R Square	0.496277
Adjusted R Square	0.478908
Standard Error	272.3127
Observations	31

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2118686	2118686	28.57136135	9.71949E-06
Residual	29	2150471	74154.18		
Total	30	4269157			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2301.385	258.8706	8.890099	8.87191E-10	1771.935201	2830.834652
X Variable 1	-209.566	39.20626	-5.34522	9.71949E-06	-289.7518701	-129.3802395

The correlation coefficient is 0.7. Which is sufficient to assume a correlation between monthly total solar energy to the monthly total rain. Therefore an analysis of variance is done and the F statistic is calculated.

H_0 : Model is not significant

H_1 : Model is significant

From the Anova table it's evident that the significance of the calculated F statistic is 9.71949E-06 which less than 5%. Therefore, the null hypothesis is rejected. It can be concluded that under a confidence level of 95% there is a significant correlation between the two variables.

3.2.1.1.5 Correlation between total solar energy and meteorological variables

Regression Statistics	
Multiple R	0.936904621
R Square	0.877790269
Adjusted R Square	0.876551657
Standard Error	90.73046271
Observations	300

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	17501808.88	5833936.294	708.6885741	1.0052E-134
Residual	296	2436676.992	8232.016864		
Total	299	19938485.87			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	225.5754009	44.53032766	5.065657783	7.17436E-07	137.9392383	313.2115634
X Variable 1	-0.032390189	0.020347002	-1.591890041	0.112476748	-0.072433307	0.007652928
X Variable 2	-5.70160047	3.972514966	-1.435262175	0.152268206	-13.51955254	2.116351598
X Variable 3	52.92568861	1.482517096	35.69988417	2.8135E-109	50.00807906	55.84329815

The correlation coefficient is 0.93. Which is sufficient to assume a correlation between monthly total solar energy to the monthly total rain. Therefore an analysis of variance is done and the F statistic is calculated.

H_0 : Model is not significant

H_1 : Model is significant

From the Anova table it's evident that the significance of the calculated F statistic is less than 5%. Therefore, the null hypothesis is rejected. It can be concluded that under a confidence level of 95% there is a significant correlation between the two variables.

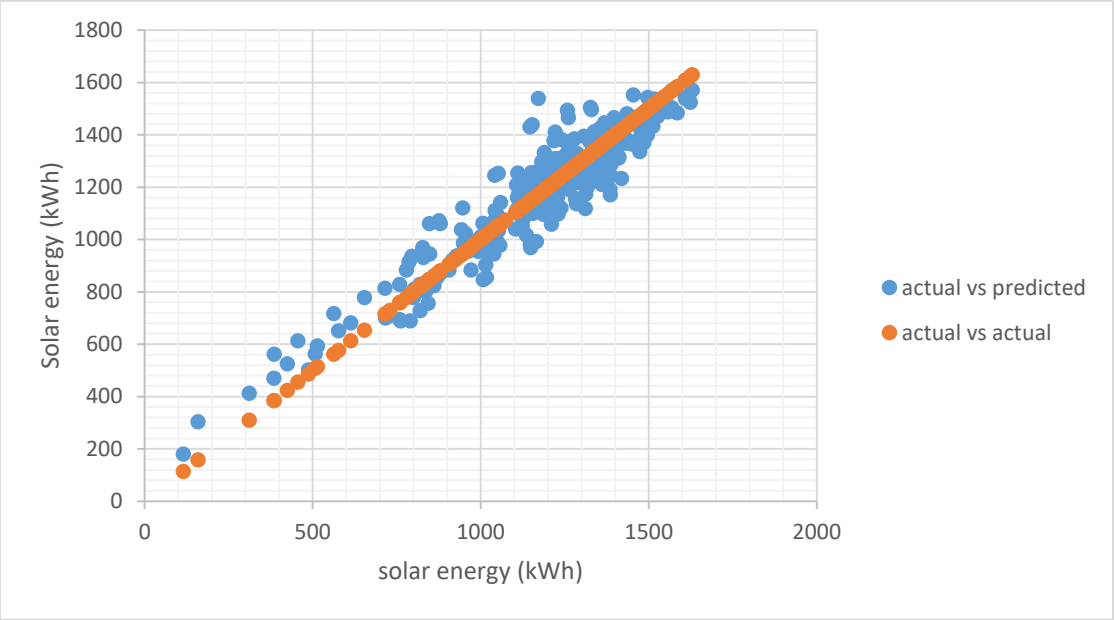


Figure 3-3 Comparison of actual monthly average energy output with multivariable liner model

3.2.2 Establishing time span for neural network

Training and validation of the neural network was done on a data set from 2016 January to December. Neural networks were created and trained for the span of the whole year and another twelve for individual months. The aim was to establish if the monthly models could capture monsoon seasonal effects better than the annual model.

Table 3-2 Correlation coefficient and mean error of monthly neural network models

Time span	R	e_mean
January	95.06948	-1.3213
February	95.22771	0.60817
March	92.2588	-9.89884
April	94.04515	0.852014
May	91.19639	-2.23707
June	93.7885	-2.18534
July	94.88098	1.551123
August	96.70551	-0.93861
September	96.96935	-0.11287
October	90.01435	-0.25667
November	95.17279	0.479164
December	96.48266	1.55862

Performance statistics for the annual model

Table 3-3 Correlation coefficient and mean error of annual neural network model

Time span	R	Mean error
Annual	94.26625	0.794578

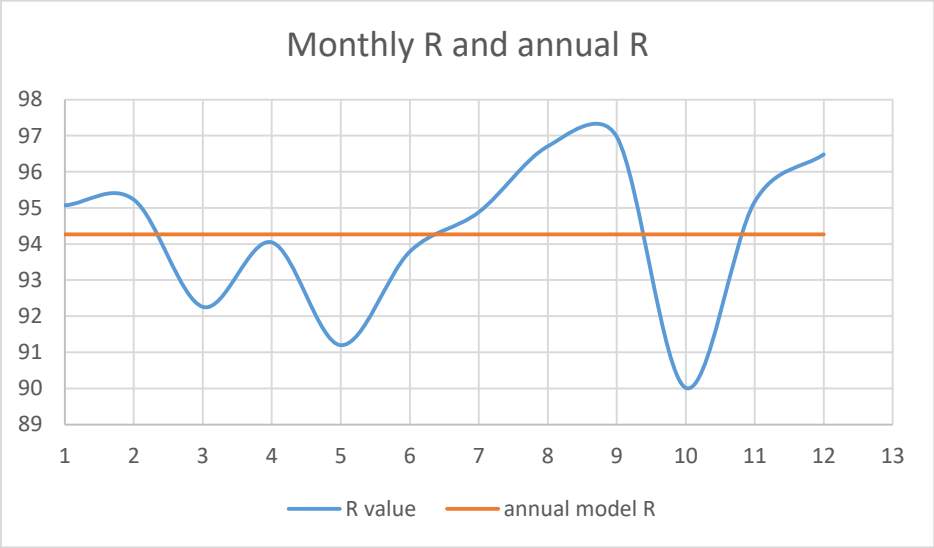


Figure 3-4 comparison of correlation coefficient of monthly models to annual model

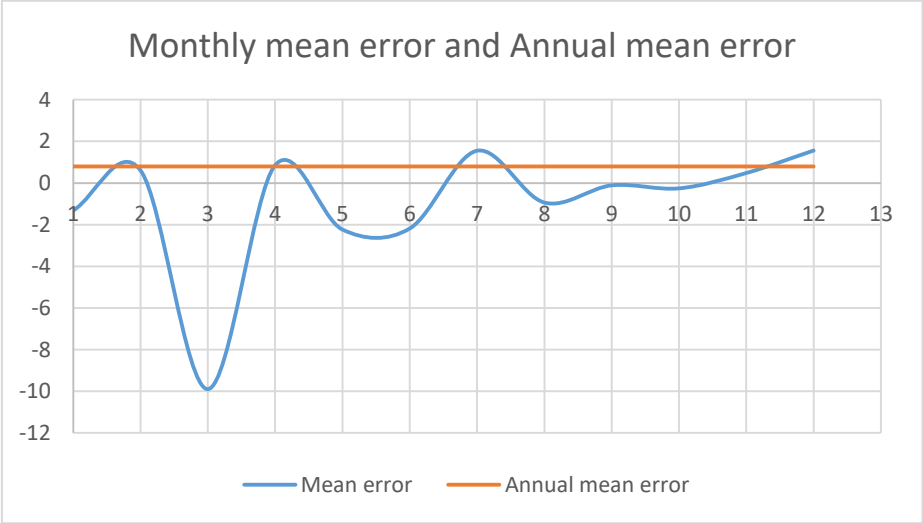


Figure 3-5 comparison of mean error of monthly models to annual model

The results of the model was compared to the actual data on following parameters

- Error variation with time
- Regression coefficient

Though monthly models had better predictions they weren't significantly accurate than the annual model. A model trained on a span of a year is more generalized and more advantageous in usability and ease of analysis. Therefore it was decided to continue the study with a single annual model.

3.2.3 Design of the artificial neural networks

The network has five inputs and one output (electricity generation in kWh). The number of neurons in the hidden layer is decided on following empirical formula, which states that the number of training observations must be of a factor of alpha of the number of weights in the network.

$$Number\ of\ input\ neurons\ N_i$$
$$Number\ of\ output\ neurons\ N_o$$
$$Numer\ of\ hidden\ neurons\ N_h$$
$$number\ of\ weights = N_i * N_h + N_h * N_o = N_h(N_i + N_o)$$
$$Number\ of\ observations = \alpha *\ number\ of\ weights$$
$$Number\ of\ observations = \alpha * N_h(N_i + N_o)$$
$$\alpha \sim (2 - 10)$$

Alpha is in the range of 2-10. As alpha increase the model becomes more generalizable but the accuracy will reduce [21] . Exact value for alpha is experimental on each case. The starting value for alpha is chosen as $\alpha = 3$.

A typical month has 30 days and 390 observation points (13 per day from 6am to 6 pm). Since 30% of the data is used for validation and testing 273 points of observations are available for training of the ANN. The design of the ANN would differ with alpha as follow

Table 3-4 Variation of number of hidden layer neurons with alpha

α	N_i	N_o	N_h
3	5	1	16
4	5	1	12
5	5	1	10
6	5	1	8
7	5	1	7
8	5	1	6
9	5	1	6
10	5	1	5

The following is a test on the number of neurons needed for modeling data from January 2016. The number of observations for training is 283. The table has correlation coefficient values for each of training validation testing an overall stages of modeling.

Table 3-5 Correlation coefficient of training validation testing and overall model with number of hidden neurons

Neurons	5	6	7	8	10	12	16	18	25	30
Training	90.92	95.97	95.58	91.65	93.93	95.3	96.21	96.37	97.57	97.01
Validation	93.12	94.61	96.51	94.12	90.32	95.5	95.45	88.66	94.55	90.63
Testing	86.25	91.91	93.26	89.56	90.69	94.85	93.81	94.27	92.8	90.98
Overall	90.58	95.17	95.38	91.7	92.78	95.17	95.69	95.01	96.46	95.23

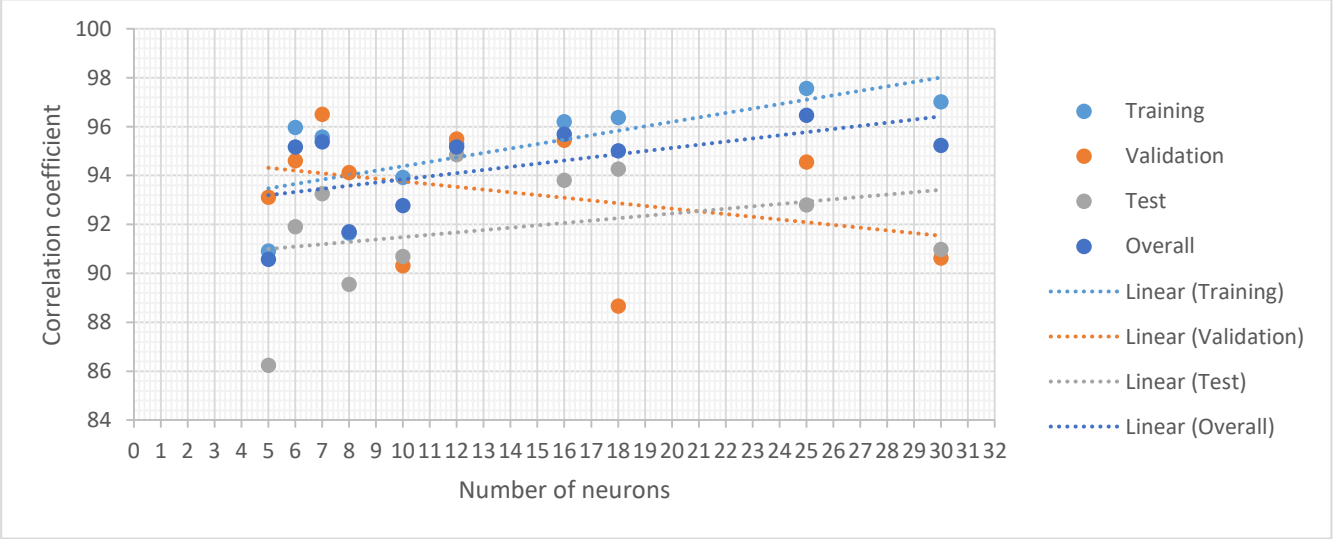


Figure 3-6 Correlation coefficient of training validation testing and overall model with number of hidden neurons

The correlation coefficient of validation decreases as the number of neurons increase owing to the overfitting of the ANN. Further this cause the training correlation to increase as well. The best combination of accuracy and generalizability is at 12 neurons where alpha is equal to 4. This analysis was done on all months and they too resulted in optimum correlation coefficients at 12 neurons. Therefore a two-layer feed-forward network with 12 sigmoid hidden neurons and linear output neurons is used to for mapping using the available data. The first layer consists of input neurons, which connects via synapses to the second layer of neurons (hidden layer) and then via more synapses to the third layer, which includes the output neuron.

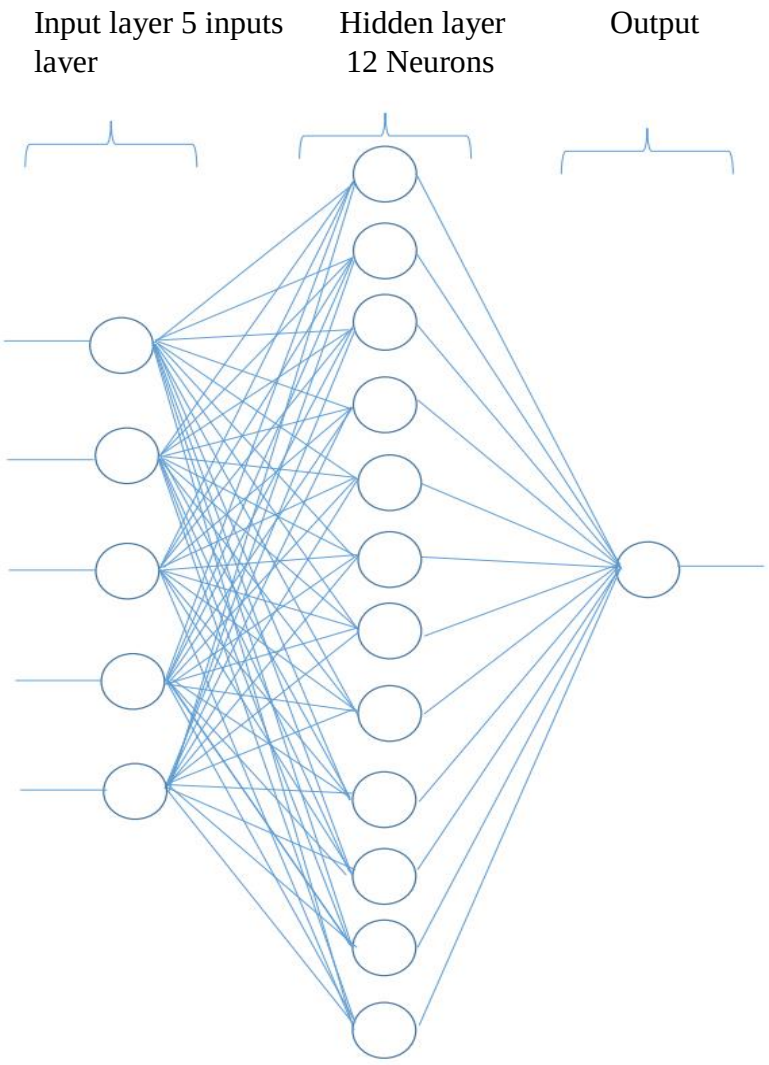


Figure 3-7 Structure of the artificial neural network for each month

Each neuron is connected with a synapse which has a weight. Notation is such that

$$W^{(layer\ number)}_{staring\ neuron-ending\ neuron}$$

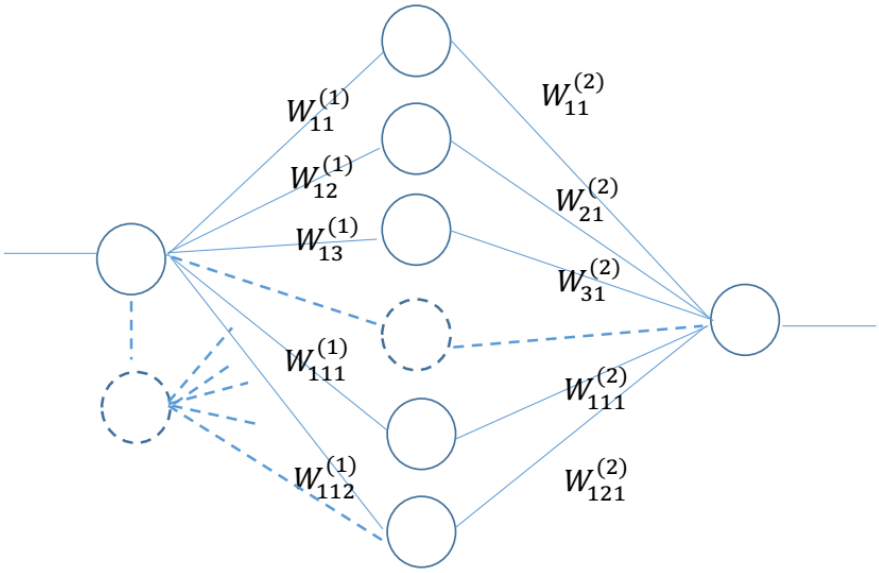


Figure 3-8 Synaptic weights nomenclature of the monthly neural network models

The neural networks take in five inputs

1. Rainfall (mm)
2. Irradiation(MJ/m2)
3. Morning cloud cover (Octas)
4. Evening cloud cover (Octas)
5. Hour number (6 am =1 6 pm=13)

The five inputs are mapped to the output of a 300kW solar PV system.

3.2.4 Training neural network with data from existing solar PV project

The network was trained with Levenberg-Marquardt backpropagation algorithm where the data is divided into three groups as follow to ensure accuracy coupled with generalizability. The annual hourly data set of 2016 has 3900 observation points.

Table 3-6 Allocation of data Training Validation and Testing sets

Data group	Percentage	Number of observations
Training data	60%	2340
Validation data	20%	780
Testing data	20%	780

To ensure the generalizability of the model a considerable percentage (40%) of data was used to validate and test the model. Selecting data to these two groups is done randomly.

The size of the hidden layer is decided by training multiple neural networks. The network with the best performance is selected. The hidden layer size was 15 for the best performing neural network.

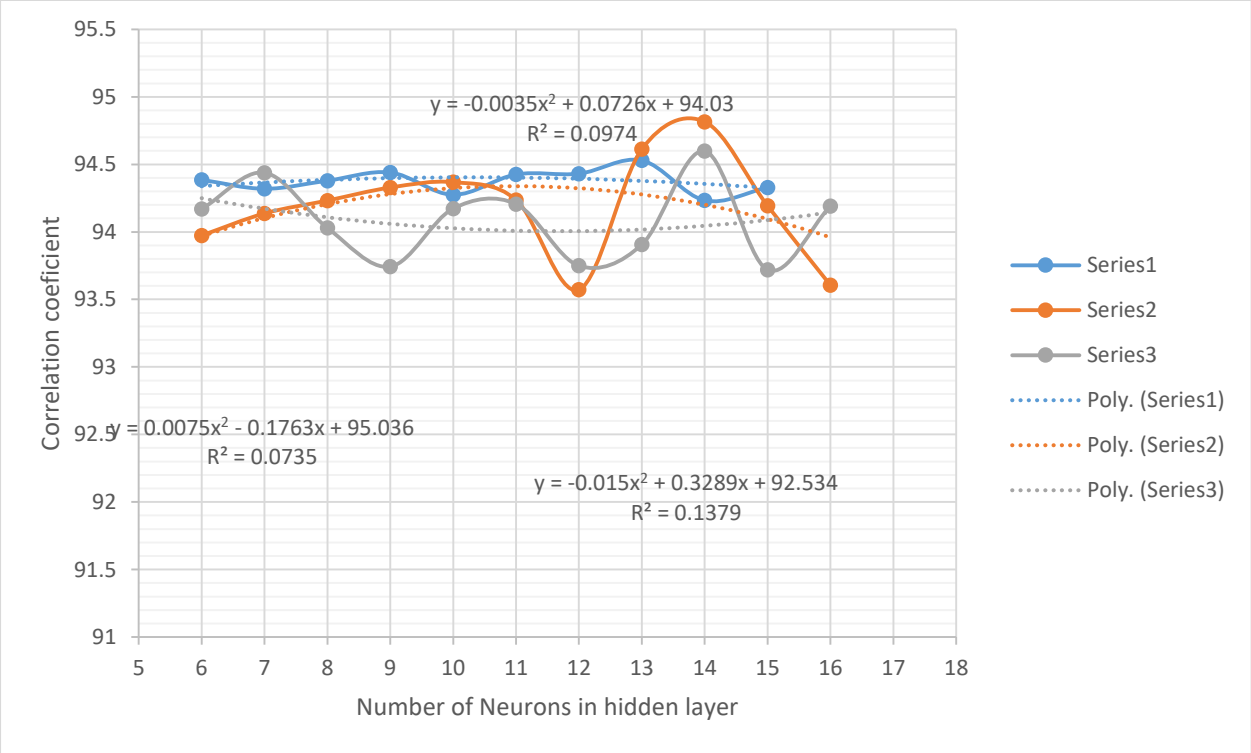


Figure 3-9 Overall correlation coefficient variation with number of hidden layer neurons for annual model

Training of a neural network is optimizing the synapse weights such that the Mean Squared Error over the whole input data is minimized. The optimization is done by gradient decent method. The solution converges and training is stopped at six consecutive increments of the validation error. The following figure shows the error variation with optimization iterations

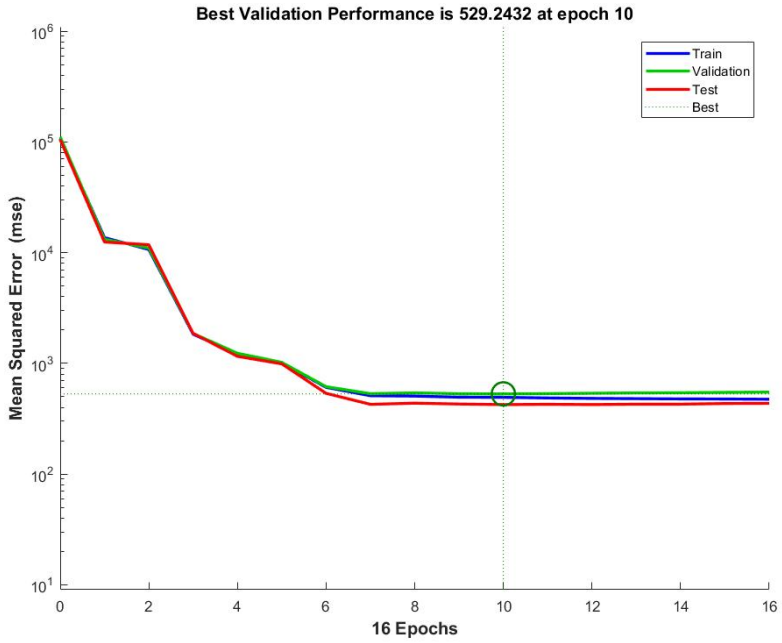


Figure 3-10 Convergence of mean square error with training iterations

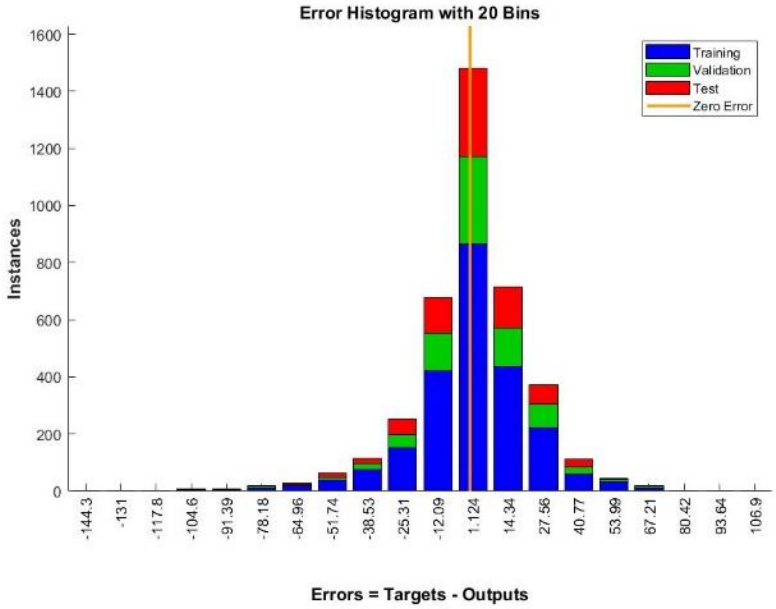


Figure 3-11 Error Histogram

At convergence the MSE and correlation coefficient is as follow.

Table 3-7 Mean square error and coefficient of correlation for training, validation and testing data sets

Data group	MSE	R
Training data	492.91	.944
Validation data	529.24	.937
Testing data	423.81	.951

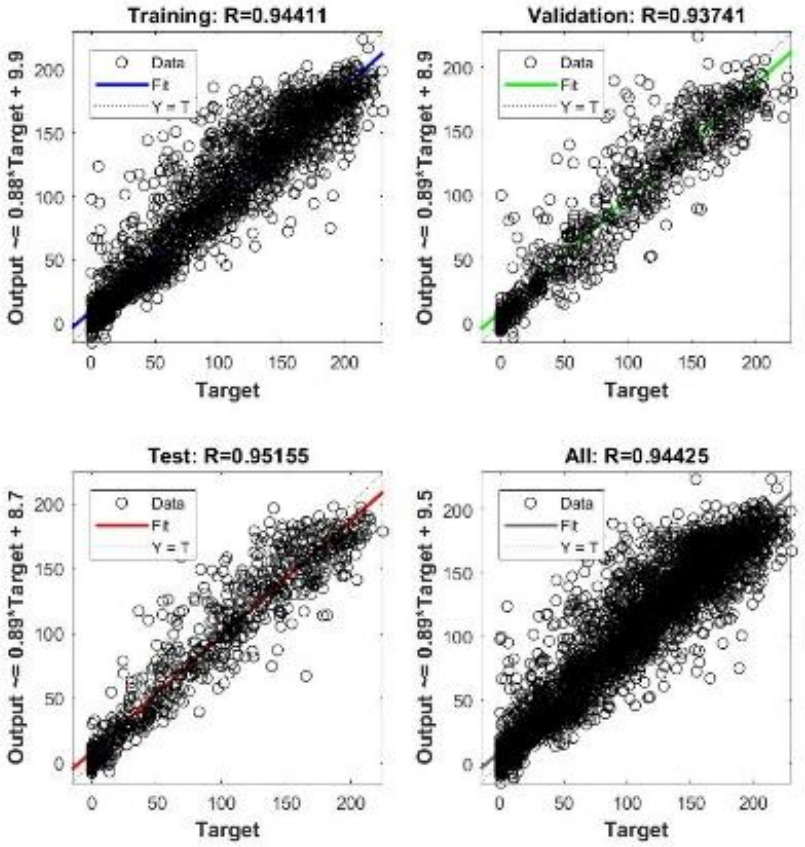


Figure 3-12 Regression of Training Validation and Testing

3.2.5 Validation of model

Data from February of 2016 was not used in any of the training processes. They were left out to further test the generalizability of the neural network after training. The prediction using the trained model for February is as follow. The correlation coefficient between the target data and predicted is 93.92%. The first of the following graphs depicts that the model output closely follows the actual daily power cycles.

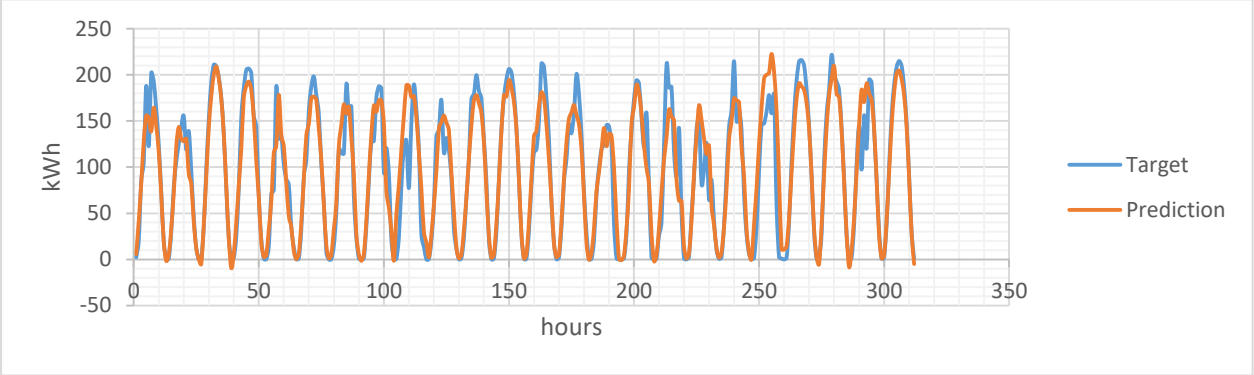


Figure 3-13 Predicted power cycles to target cycles

The following graph clearly shows the model output is in significant agreement with the target data.

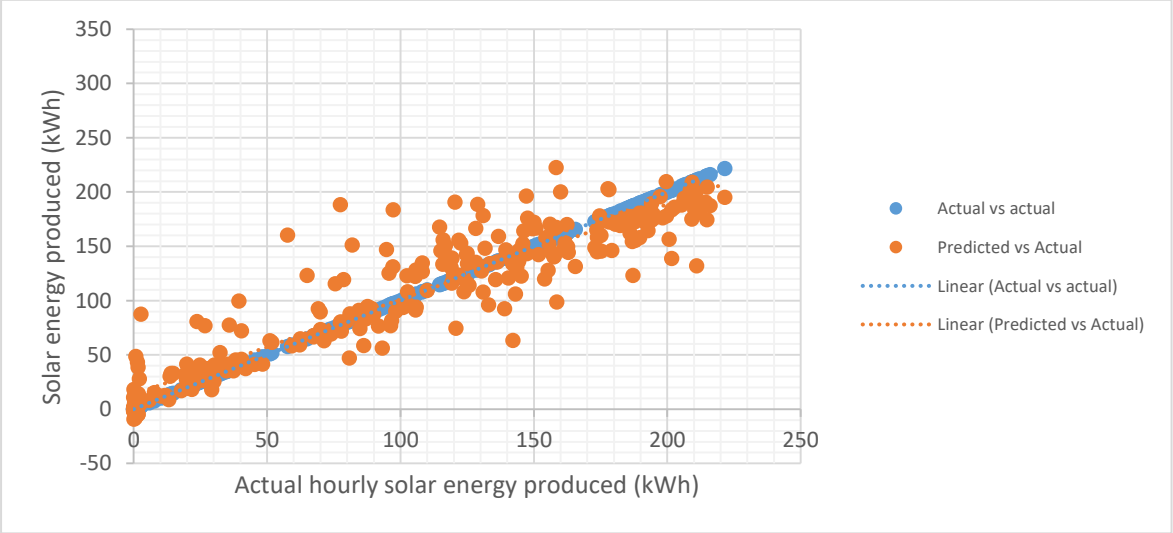


Figure 3-14 Comparison of prediction to target data

Analysis of error terms reveal that error terms are randomly distributed.

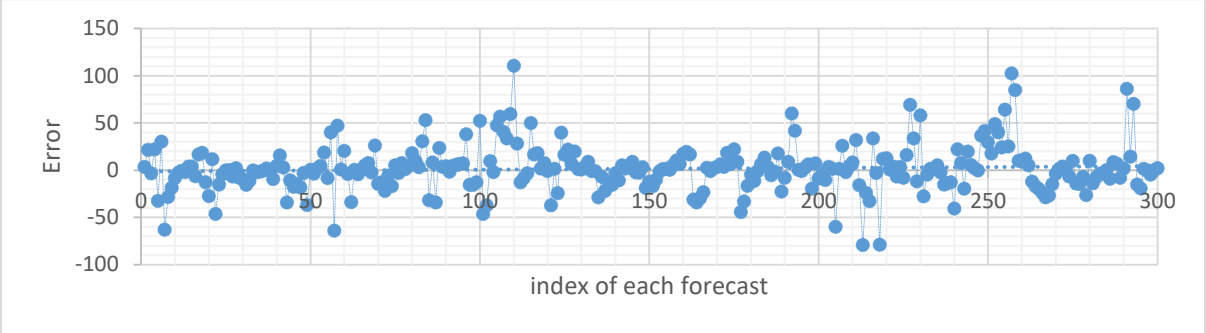


Figure 3-15 Random distribution of error terms

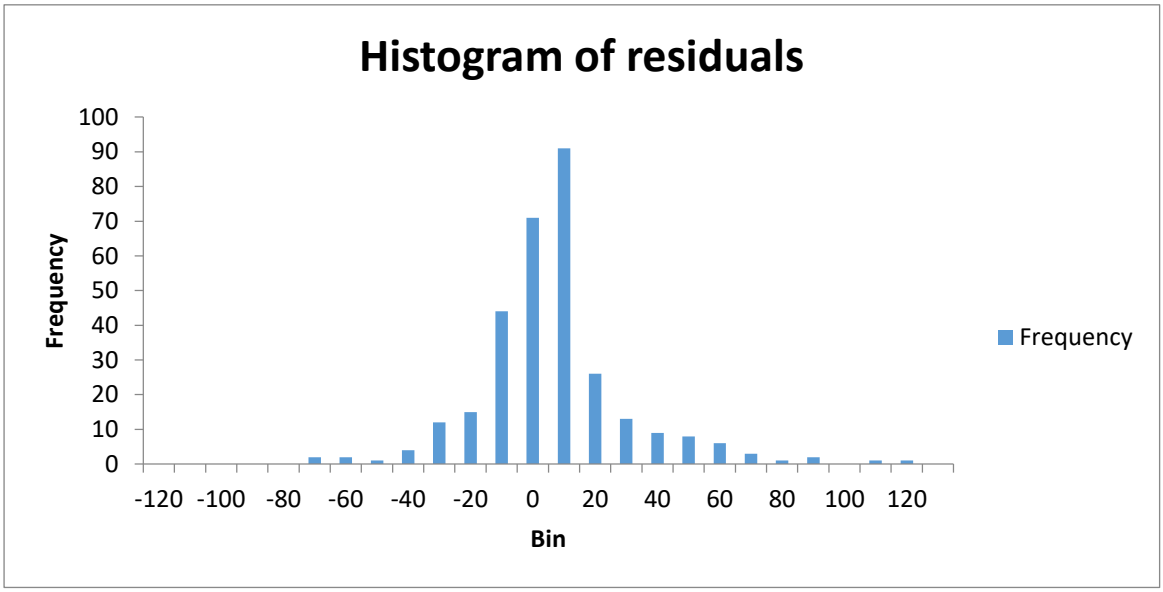


Figure 3-16 Histogram of residuals for February

3.2.6 Evaluation of adequacy and significance of the predictive model

3.2.6.1 Analysis of the hourly error

3.2.6.1.1 Hourly error distribution

Outliers caused by maintenance outages, inverter tripping and power outages were found and eliminated. The eliminated data points are

Table 3-8 Outlier data points

Sorted e	Day	time
-148.52	4/10/2016	11:00
-125.64	5/30/2016	13:00
-122.03	8/22/2016	11:00
-116.52	10/18/2016	13:00
-115.09	10/9/2016	11:00
-112.97	1/15/2016	11:00
-106.80	10/12/2016	11:00
-100.80	7/29/2016	14:00

Error between prediction and actual values were calculated. The error terms were classed as follow for further analysis.

Table 3-9 Classed error terms

Class low	class upper	Freq	class mark	PDF	Normal Z
-113.04	-95.55985727	1	-104.2999	0.000257	-5.11475
-95.5599	-78.0798097	17	-86.81983	0.004369	-4.26644
-78.0798	-60.59976212	30	-69.33979	0.00771	-3.41812
-60.5998	-43.11971455	82	-51.85974	0.021074	-2.5698
-43.1197	-25.63966697	197	-34.37969	0.05063	-1.72148
-25.6397	-8.159619394	603	-16.89964	0.154973	-0.87316
-8.15962	9.320428182	1820	0.580404	0.467746	-0.02484
9.320428	26.80047576	808	18.06045	0.207659	0.823477
26.80048	44.28052333	267	35.5405	0.06862	1.671796
44.28052	61.76057091	51	53.02055	0.013107	2.520114
61.76057	79.24061849	11	70.50059	0.002827	3.368433
79.24062	96.72066606	4	87.98064	0.001028	4.216752

The class mark of error terms were normalized and plotted to obtain the distribution of error. It is observed that the error terms have a symmetric distribution closely resembling a normal distribution. Further tests were done to establish this fact.

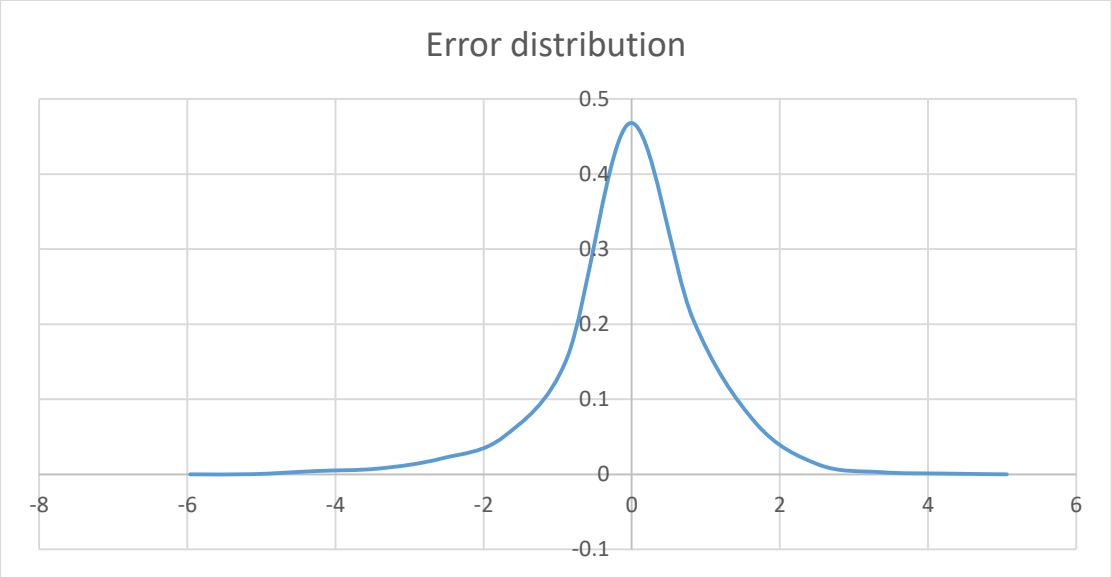


Figure 3-17 Normalized distribution of error

Descriptive statistics of the error distribution were calculated. The values suggest a normal distribution with Mean error percentage is 3.1%. Which is a favorable value for a significant model

Table 3-10 Descriptive statistics of error term distribution

Descriptive statistics of error	
Mean	1.092280668
Standard Error	0.330333768
Median	1.418776841
Mode	0.288470185
Standard Deviation	20.60552038
Sample Variance	424.5874703
Kurtosis	2.588083893
Skewness	-0.52386842
Range	187.159949
Minimum	-95.71766717
Maximum	91.44228183
Sum	4250.06408
Count	3891

3.2.6.1.2 Randomness of error

Randomness of error suggests that there is no autocorrelation within the error terms. Durbin Watson statistic is evaluated for this purpose

Durbin Watson statistic

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n}(e_i)^2}$$

$$\sum_{i=1}^{3900}(e_i)^2 = 1{,}666{,}450$$

$$\sum_{i=2}^{3900}(e_i - e_{i-1})^2 = 2563970$$

$$d = 1.54$$

Since Durbin Watson statistic is larger than 1 and closer to 2, it can be determined that the errors are not auto correlated, therefore can be considered random.

3.2.6.1.3 Normal distribution of error

The QQ plot compares the Z score of a distribution with the Z scores of a standard normal distribution. It a measure to evaluate if the error distribution is normal. The Z scores of the error terms closely follow the standard normal distribution values.

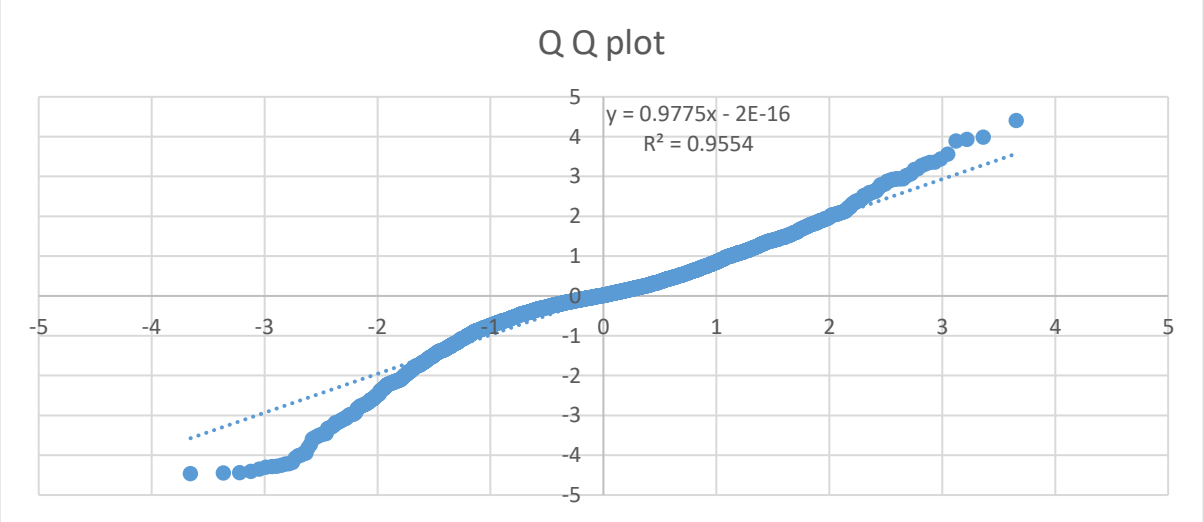


Figure 3-18 Comparison of error distribution to standard normal curve

3.2.6.2 Comparison of predicted monthly total energy to actual monthly total energy

Monthly total energy of the prediction and the actual were compared to establish the deviation. The results show a deviation between 2% to -4%. This error percentage is very favorable and is better than estimations done by commercial software.

Table 3-11 Actual and predicted monthly total energy error percentage

Month	E_pre	E_act	(Epre- Eact)/Eact %
Jan	37254.48	36586.80	1.8
Apr	40355.08	42036.02	-4.0
May	30778.61	30351.80	1.4
Jun	34488.07	34903.67	-1.2
July	35935.31	36652.23	-2.0
Aug	37835.03	38214.81	-1.0
Sep	35896.97	36657.91	-2.1
Oct	31404.96	31311.99	0.3
Nov	30840.86	31661.73	-2.6
Dec	36758.59	36411.91	1.0

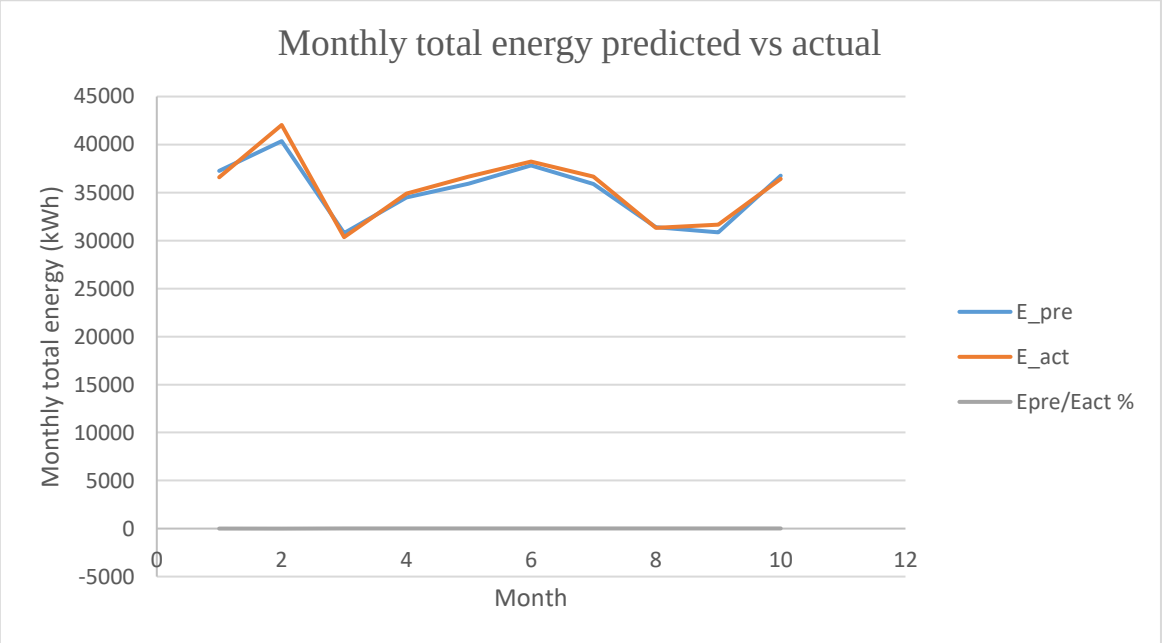


Figure 3-19 Predicted monthly total energy comparison with actual monthly total energy and error percentage

3.3 Calculate energy storage requirement under different operating schemes

Upon creating a successful prediction scheme for solar electricity output, the energy storage requirement is calculated for an Energy Storage System (ESS) to mitigate fluctuations. Determining the energy storage requirement the ESS was done under the following smoothening schemes

1. Controlled ramp rate scheme(Constant ramp, constant power, constant de-loading ramp)
2. Moving average smoothening scheme

3.3.1 Operation under controlled ramp rate

This scheme controls the energy output of a system in three segmented as follow

- Ramp up stage: Output increases at constant rate to reach a maximum
- Constant stage: Output remains constant at maximum power
- Ramp down stage: Output decreases at constant rate to reach a minimum

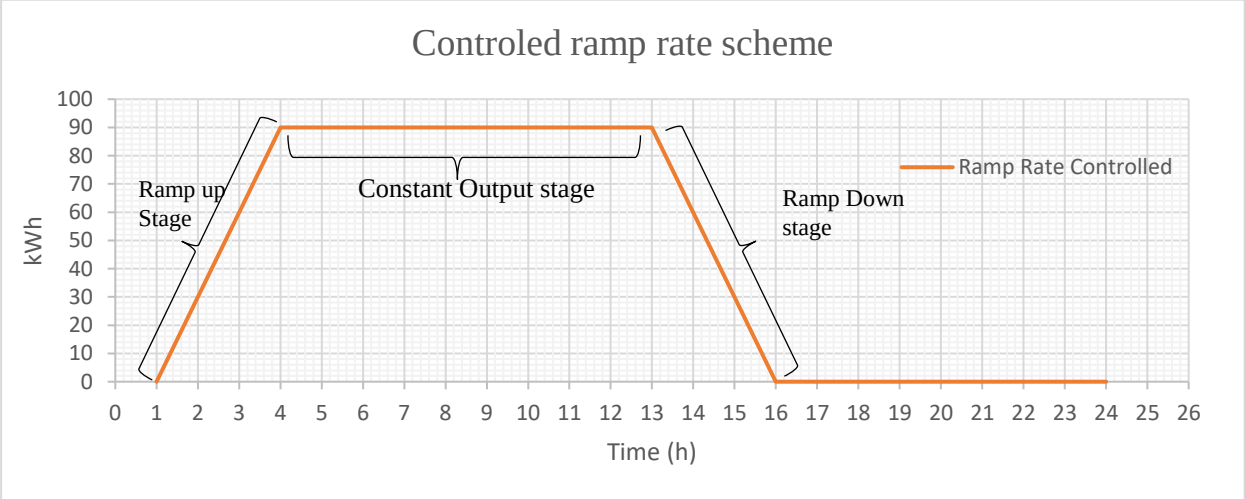


Figure 3-20: Ramp rate control scheme

Ramping (kW/min) rate is an important parameters which effects the stability of the system. The frequency controlling machines respond to a ramping of an energy source in the system by de-loading and vice versa. The amount of power change required to change the system frequency by 1Hz for the Sri Lankan power system has an average value of 85MW/Hz. The value of the parameter is dependent on the combination of power plants dispatched at the time. Operating at a known constant ramp rate aids the stability of the power system. A controlled ramp rate has the following variables

1. Ramp rate (R)
2. Ramping time (t_1)
3. Constant power time (t_2)
4. Ramp down time (t_3)

Conditions of the ramp rate control method studied in the

1. Constant ramping rate
2. Ramp up time and ramp down time is assumed to be equal.
3. Cycle time of the process is limited to 24 hr (Daily cycled)
4. Starting time of the system operation is 6 am.
5. Daily total energy of the controlled output is equal to the energy of the uncontrolled daily output

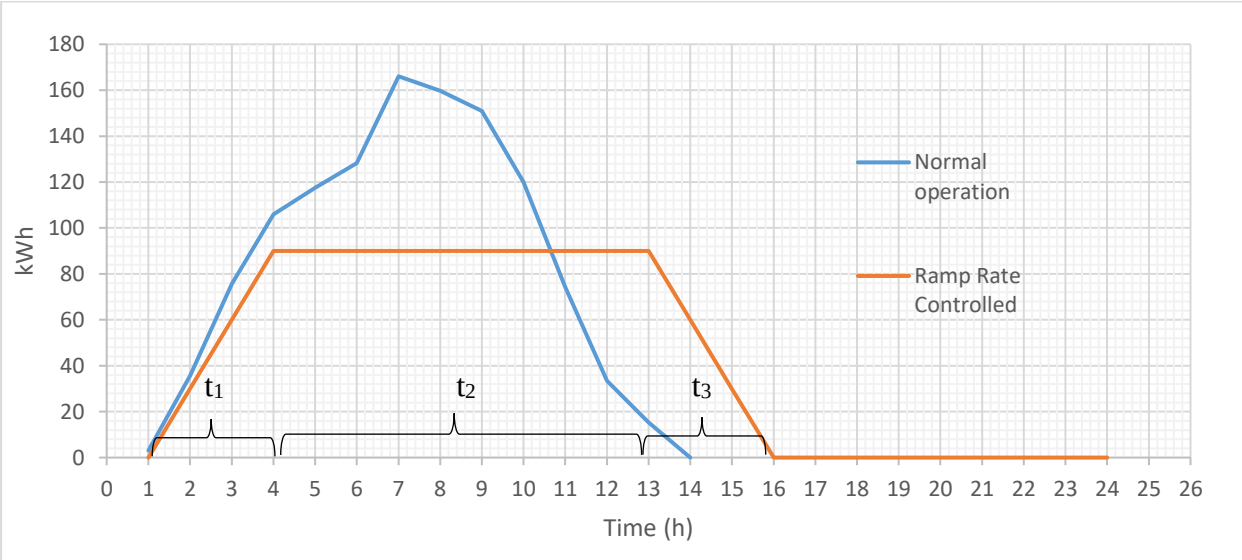


Figure 3-21 Ramp rate control compared to uncontrolled PV power output for an day

When the normal output (Blue) is greater than the delivered power of the ramp rate controlled scheme, the excess energy is stored in an Energy Storage System (ESS). Similarly when the energy output requirement is less than the produced energy the ESS is discharged. When a zero reference is considered for the state of charge (SOC) of the ESS, charging results in positive SOC and discharging results in negative SOC. The maximum of the charging SOC and minimum SOC is the required energy storage capacity. (The study is limited to calculating the energy storage capacity and does not move ahead to designing an actual ESS, which would depend heavily on the chosen technology.)

$$ESS_{Capacity} = SOC_{max} - SOC_{min}$$

The parameters of ramp rate control scheme can be optimized such that the required the ESS capacity is minimized. Optimization limits are as follow

1. Maximum output of RRC scheme <1.5x maximum normal output.
2. Time of constant power output >0 to avoid a sudden shift from ramping up to ramping down stage
3. Energy storage system is cycled every 24 hrs

Following figure is an optimization process carried out for the 1st of August 2016. From the different paths a ramp rate controlled scheme can take the path which requires the least of energy storage is selected.

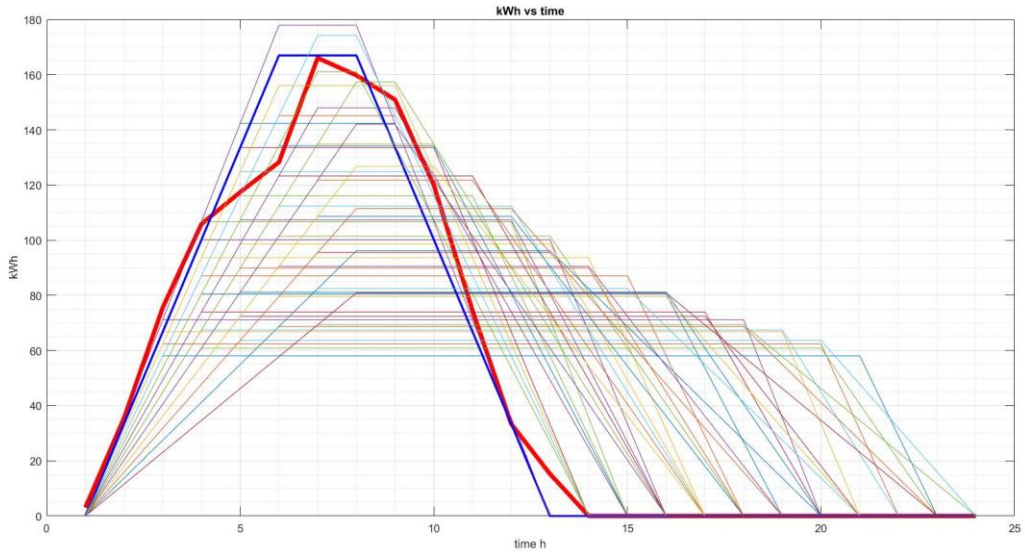


Figure 3-22 Potential ramp rate controlled paths identified in the optimization program

The graph in red color represents the actual output of the system. The numerous trapezes are outputs of the optimizing iterations. For each of them the energy storage requirement is calculated, and the path which offers the smallest storage requirement is selected as the optimal path.

The optimal path is the graph in blue color. The figure below shows the variation of the state of charge (SOC) of the energy storage system when following the optimum path. SOC is shown by the light blue graph and it has a reference of zero.

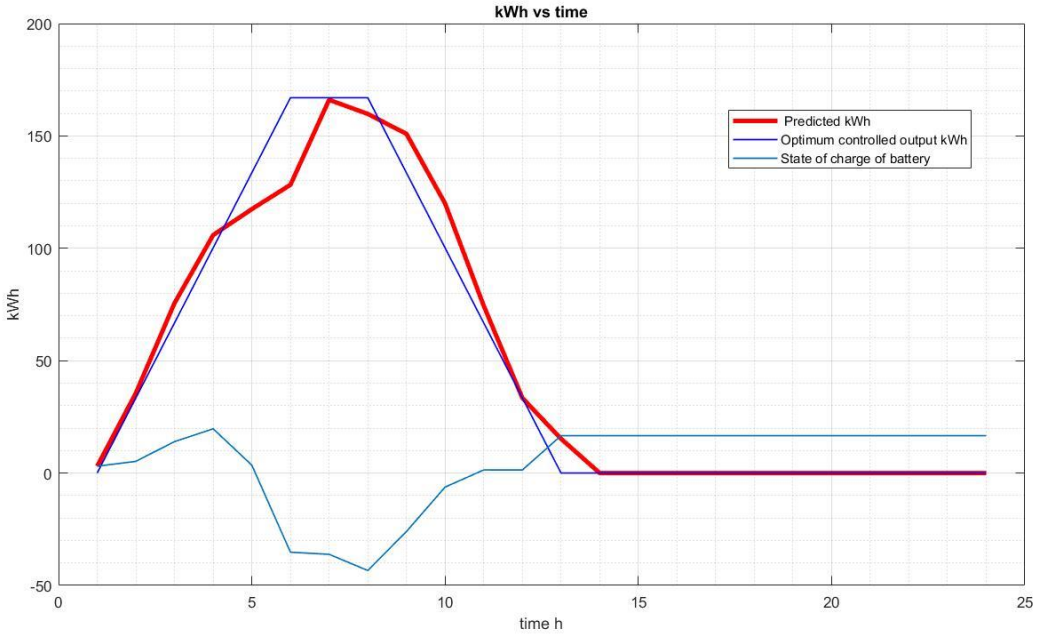


Figure 3-23 Optimum ramp rate controlled path and variation of SOC of the storage system

3.3.2 Operation under moving average smoothening

This power output control scheme controls the energy output of a system by smoothening based on moving average. The output is not as smooth as the ramp rate controlling scheme. Moving average window is selected such that the output energy profile is constrained by an upper limit to the hourly fluctuations. Actual power output typically shows fluctuations above 20%. The fluctuations are limited to 15% by moving average smoothening.

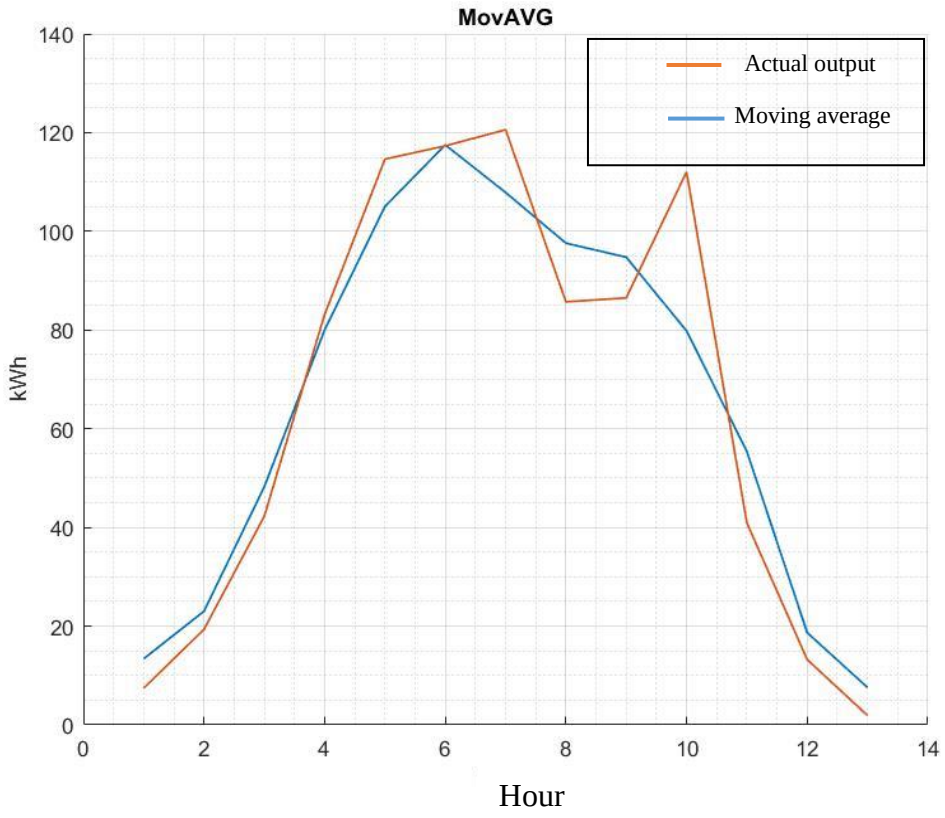


Figure 3-24 Operation under moving average output control scheme

Chapter 4 Results and Discussion

The research was aimed at establishing a neural network model to forecast the weather induced fluctuations in power output of a Solar PV power plant. The first step is to identify the input parameters. Power output of a solar electricity system is dependent on meteorological parameters.

Through statistical analysis a linear correlations between monthly solar energy output and monthly rain was established with 95% confidence interval. The regression shows there is a negative correlation between monthly energy output (kWh) with monthly rain fall (mm)

$$Y = -0.9\,X_1 + 37691$$

A positive correlation exists between Monthly energy output (kWh) with monthly irradiation(kW/m2). The linear correlations established at 95% confidence interval. Correlation is as follow

$$Y = 51.59X_1 + 6462.01$$

The effect of clouds on monthly solar energy output is not significant compared to other variables. The effect of cloud cover is at lower time scale. Cloud cover is a main cause of hourly and daily solar irradiance fluctuations and fluctuations in solar energy output of PV systems. The correlation is negative and is as follow.

$$Y = -209.56X_1 + 2301$$

The effect of all three variables on solar energy output was established as

$$Y = 52.92X_1 - 5.70X_2 - 0.03X_3 + 225$$

Y : Solar energy output

X_1 : Monthly Solar irradiation

X_2 : Monthly rainfall

X_3 : Monthly cloud cover

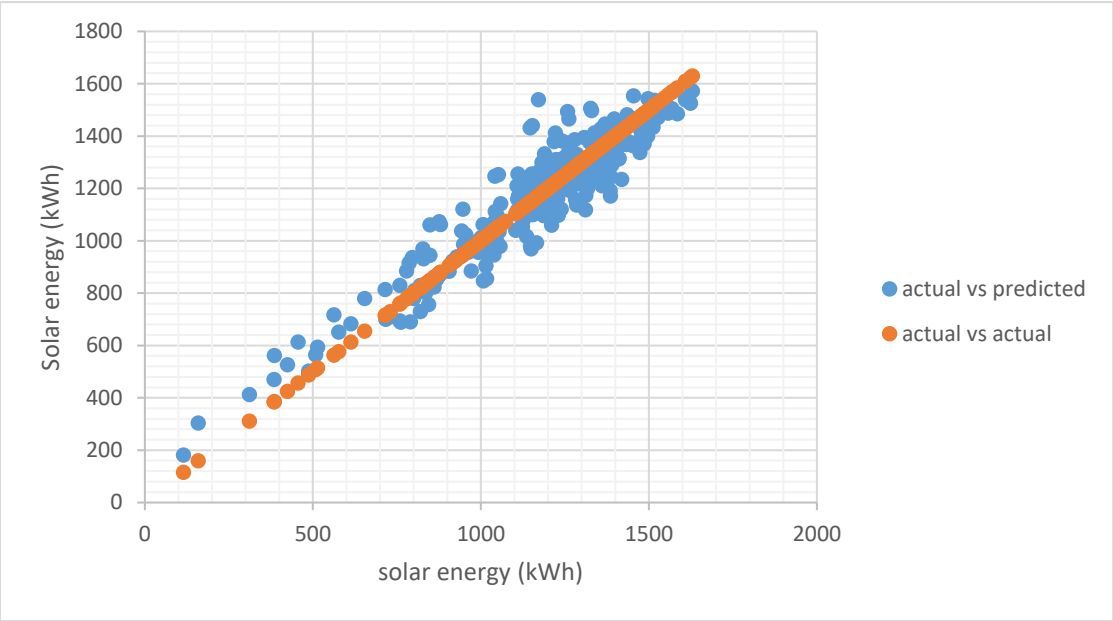


Figure 4-1 Monthly energy output successfully modeled by identified variables

The artificial neural network was built and trained for a 300kW solar electricity system in Colombo for the year of 2016. The model fits the data with significant accuracy. Two separate methods of analysis were carried out. The first was to develop a separate ANN for each month. The results were compared to a model made for the complete year. The annual model is more generalizable since it's trained validated and tested on a larger data set and easier to use in analysis. Further there is no significant advantage of the monthly models in terms of regression coefficient and mean error. Therefore developing a model for the complete year was selected. Performance of the neural network for the complete year of 2016 graphs.

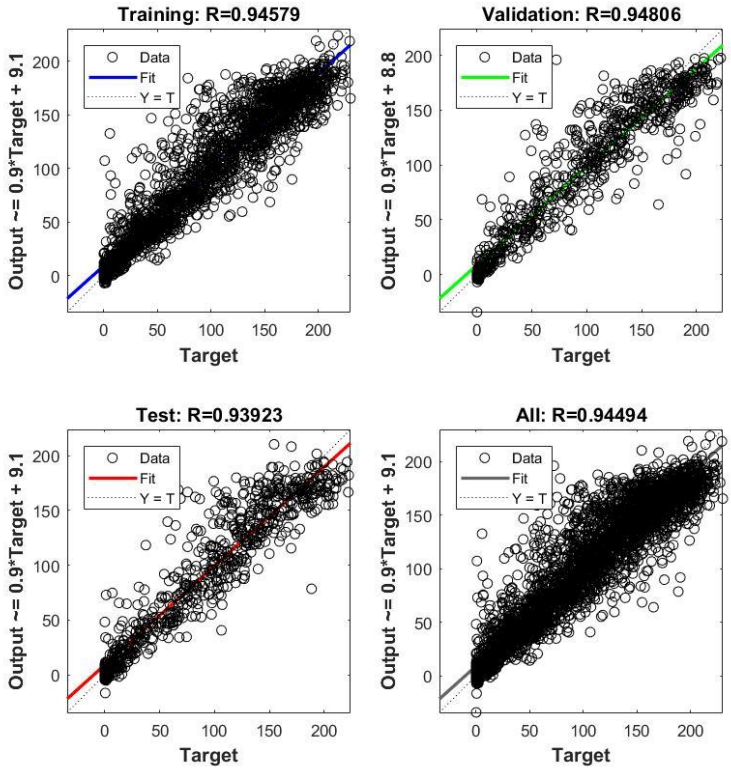
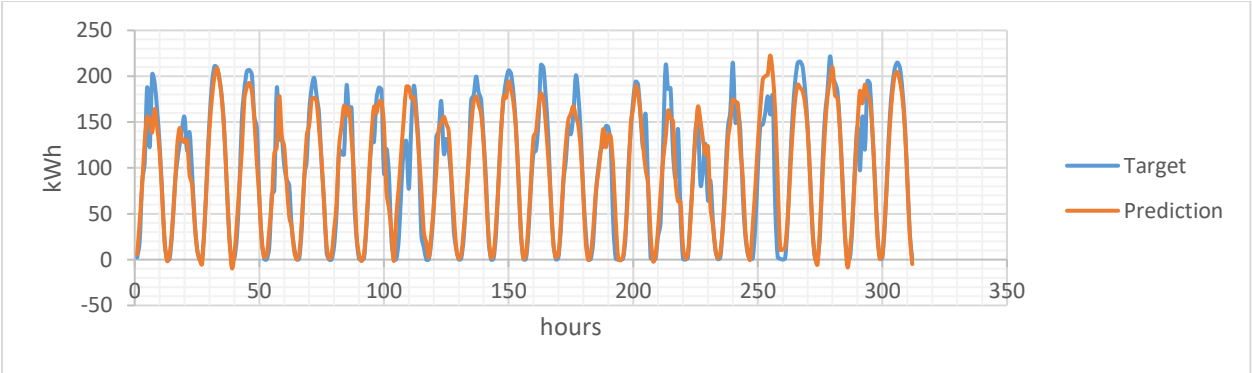


Figure 4-2 Comparison of model agreement to the training, validation, testing and overall data sets.

Model is validated by two methods. The first is by a randomly selected validation data set. The data and shows a very significant correlation coefficient of 0.94. The second method was by the excluded data set of February from the main data set . This data was not used for training and it is completely alien to the model. The correlation coefficient between the target data and predicted is 93.92%. Error percentage between predicted and actual total monthly energy generation is 1.6%. Therefore model is validated for generalizability.



The ANN model is evaluated for adequacy by analysis of error terms distribution. Error terms are distributed normally. Autocorrelation within error terms was evaluated by Durbin Watson statistic. Durbin Watson statistic for the error term distribution is close to 2. Therefore suggests error terms are random.

Comparison between predicted total monthly energy and actual total monthly energy for 2016 reviled the percentage difference was between -4% to 1.8% . Under these criteria it is evident that the neural network model is significantly accurate and generalizable.

4.1 Evaluation of energy storage requirement.

Forecasts of the neural networks were used to evaluate the energy storage requirement of two power output controlling schemes.

1. Controlled ramp rate scheme(Constant ramp, constant power, constant de-loading ramp)
2. Moving average smoothening scheme

In the ramp rate control scheme numerous combinations of ramping rate (kw/h), ramping time and constant power output time were assessed to find viable combinations. The viable combinations were evaluated for the energy storage requirement of each and the one with the least energy storage requirement was selected as the optimum path for the day. This optimization was carried out over all the days of the year to create a vector of least energy requirement for each day of the year.

Under the moving average smoothening scheme the energy storage requirement for each day was calculated with specified averaging window size.

The following figures show the energy storage requirement for each day of the year under the ramp rate controlling scheme and moving average smoothening scheme after optimizations. The data is split into two graphs for ease. First graph consists 152 days of January April, May, June and July. Second graphs consists 148 days of August, September, October November and December. Maximum energy storage requirement for ramp rate control scheme is 146.2 kWh and Maximum energy storage requirement for moving average control is 62.2kWh

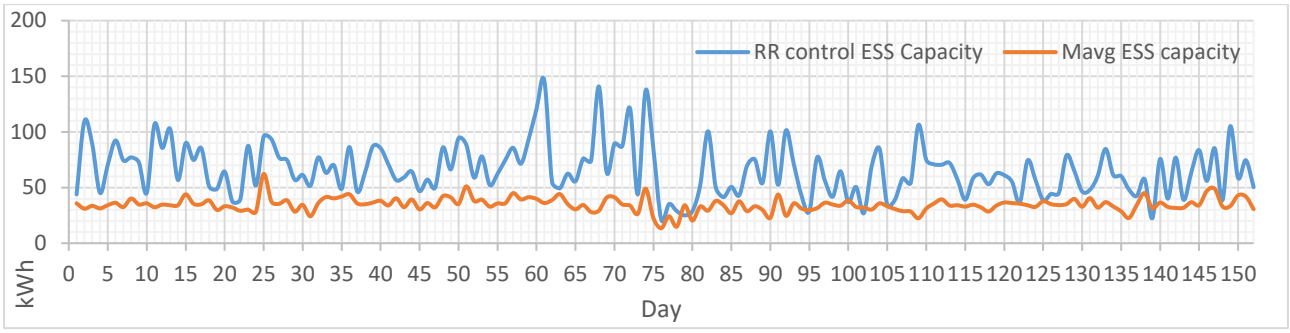


Figure 4-3 Energy storage requirement variation with time for ramp rate controlled scheme and moving average control scheme part 1

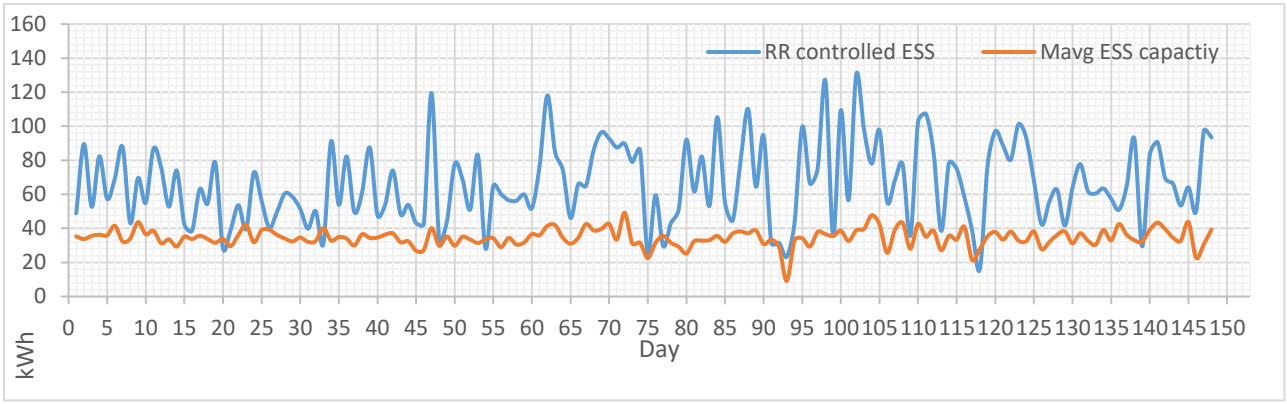


Figure 4-4 Energy storage requirement variation with time for ramp rate controlled scheme and moving average control scheme part 2

It is evident that the moving average smoothening scheme requires less energy storage than the ramp rate controlled method. Variations in the moving average method is less. Standard deviation of energy storage requirement is 5.6 kWh from a mean of 35kWh. The energy storage requirement is dependent on the averaging window size. Following graph shows that there is a linear relationship between averaging window size and energy storage requirement.

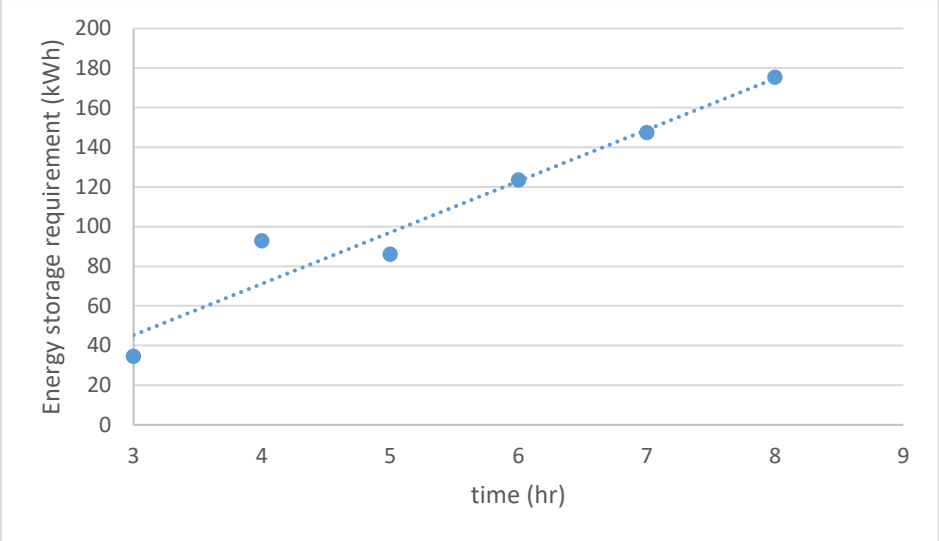


Figure 4-5 Mean energy storage requirement variation with averaging window size

The ramp rate control method energy storage requirement has more dispersion and variations in its data. Therefore a frequency distribution analysis is more suited over selecting the maximum value as the design criteria.

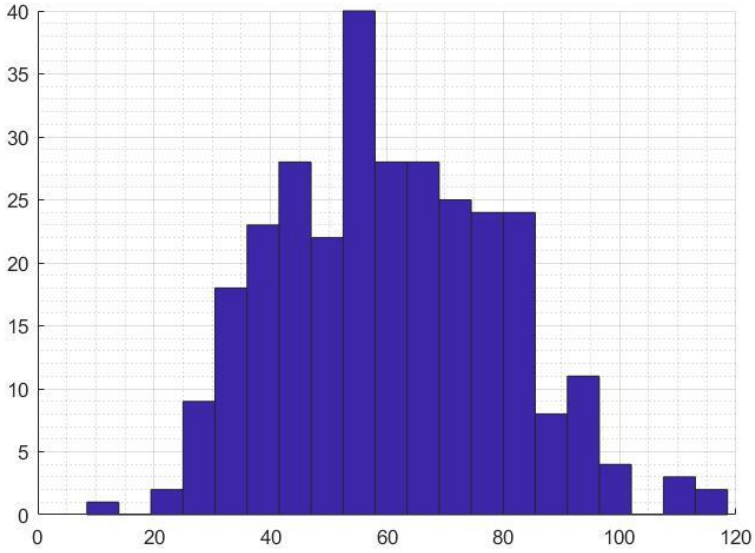


Figure 4-6 Histogram of energy storage requirement for ramp rate controlled scheme

The energy storage requirement is dependent on the following characteristics of the scheme.

- Maximum allowable ramp rate (RR_{max})
- Minimum allowed time of continuous power delivery(t_{2min})

The following diagram shows the variation of required energy storage capacity under variable combinations of RR_{max} and t_{2min}. The variation of energy storage requirement with t_{2min} is quadratic under constant maximum ramp rate. As t_{2min} increase there is more storage requirement and the system performs as a load shifting scheme.

As the maximum ramp limit is decreased there is a limit imposed on the variability of the system. This limitation must be enforced by the energy storage system. Therefore when RR_{max} decrease the energy storage requirement increase, resulting in the curve moving upwards. There is a change in curvature a well. As the maximum allowed ramp rate increase the effect of t_{2min} diminish.

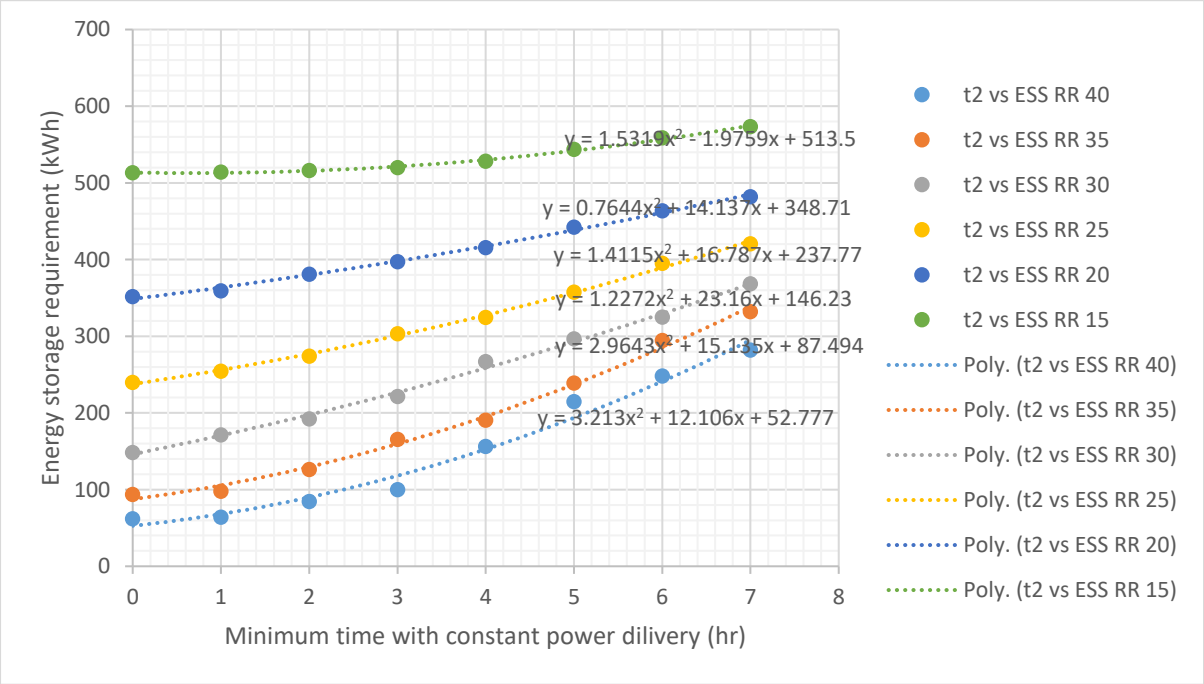


Figure 4-7 Variation of energy storage requirement

From the frequency distribution of $t_2>0$ and $RR_{max}=40$ it was observed that by considering only up to 95% of the days the energy storage requirement can be reduced by 25%. This is an important design criteria considering the high cost of energy storage.

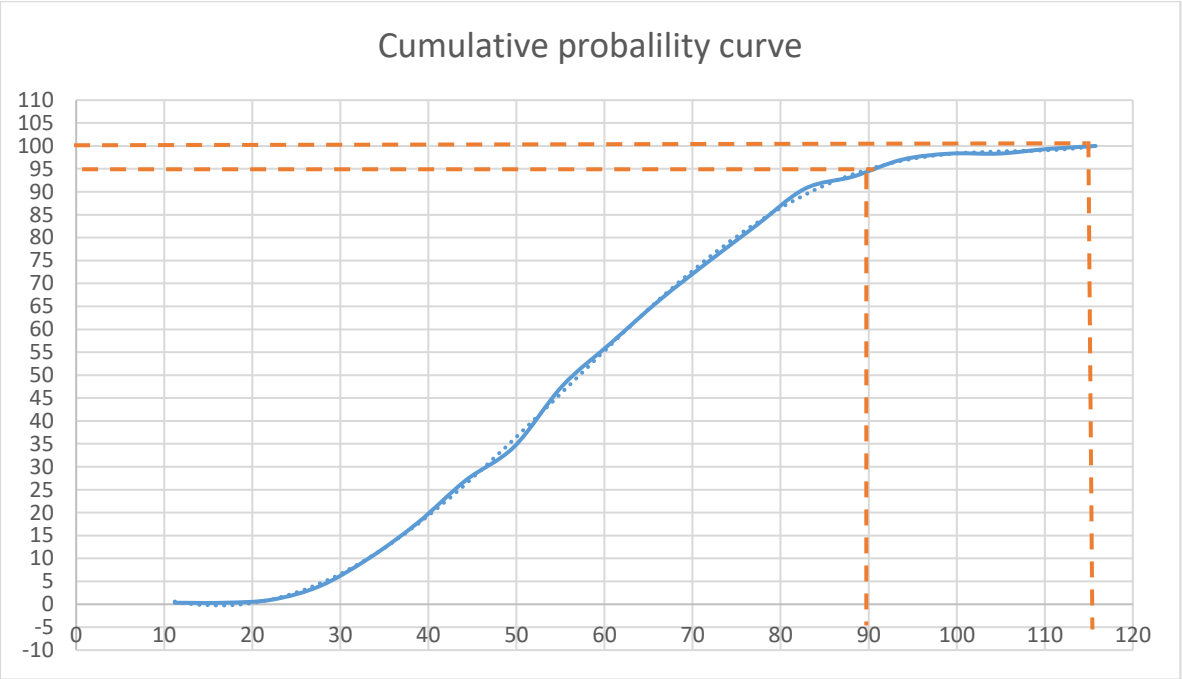


Figure 4-8 Cumulative probability of energy storage requirement under t2>0 and Rmax=40 condition

This is a common tendency under all Rmax and t_{2min} conditions. It is important to note that the gain in reduction of energy storage requirement comes at the cost of imposing operating limits on the system as well as reducing availability i.e if the storage requirement is selected for 95% of cumulative probability, there is a loss of 5% in availability.

The relationship between Rmax and t_{2min} was evaluated on the mean of the distribution at 95% cumulative probability mark. The curves shows similar characteristics to the previous curves.

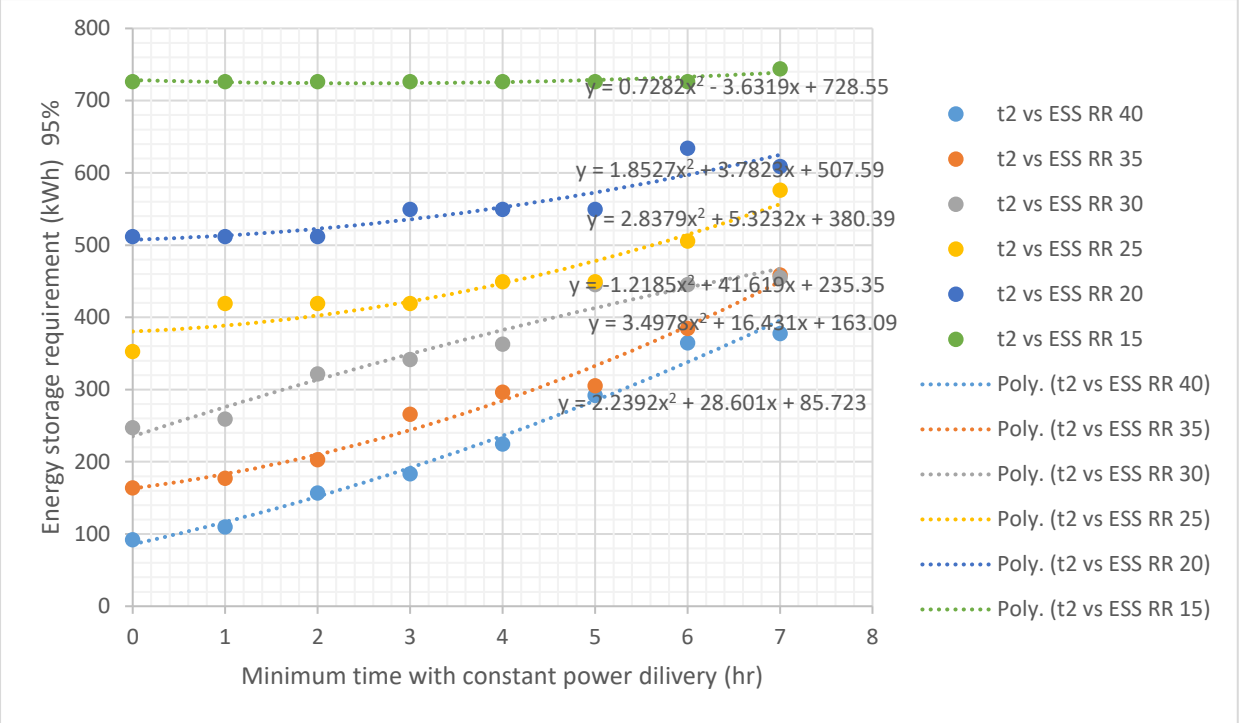


Figure 4-9 Variation of energy storage requirement for 95% of the days

Chapter 5 Conclusion and future work

Artificial neural networks (ANN) can be used to successfully model and predict the output energy of a solar PV system using five input parameters. The resulting model was used to evaluate the energy storage requirement to control fluctuations of the power output. The neural network was trained to model the output of a 300kW solar PV system installed in Colombo. The mean of hourly energy output of the PV system during the period was 90.55kWh which accounted to a mean daily energy production of 1177 kWh. The ANN model proved to deliver forecasts with significant accuracy and generalizability. Correlation coefficients for training, validation and testing were 0.945, 0.948 and 0.939 respectively. Further validation was done using an isolated data set of a month which was not used in the process of training. Model was able to achieve a correlation coefficient of 0.93 for the isolated data set. Residual analysis confirmed the error terms were random and free of autocorrelation. Error terms were normally distributed with mean 1.09kWh and standard deviation of 20.06kWh.

A direct mapping was established between meteorological parameters and power output of a solar PV system, as oppose to estimating solar irradiance. The training process maps physical properties of the system to the network provided a suitable structure is provided for the neural network. Therefore including inverter function, PV panel performance, orientation etc. in the model is not required in this approach.

The model was used to evaluate energy storage requirement for two power output control schemes. First scheme evaluated was controlling of ramp rate. This method has a specified ramp up and ramp down rate, and a continuous power delivery period. By means of an optimizing algorithm the combination of parameters corresponding to the least energy storage requirement was established and the mean energy requirement was 15% of the average daily energy generation. The optimization was carried out in the rages of 15kW/h-40kW/h for ramp rate and 0-8 hr for constant power output time. Curves for energy storage requirement were developed under each operating condition. The second output control scheme evaluated is moving average smoothening. The energy storage requirement for moving average scheme is 8% of the daily energy generation.

Ramp rate control while having better performance requires twice the energy storage capacity compared to the moving average scheme. It must be noted that imposing strict conditions on

maximum ramp rate and minimum continuous power delivery time will increase this percentage up to 50% of the daily energy generation. The effect of imposing restriction on operating parameters of the schemes were examined in detail.

A common trend was identified by analyzing cumulative probability curves of energy storage requirements. For a small drop in availability of the energy storage a significant saving can be made on the energy storage capacity i.e. for a drop of 5% of availability a 25% reduction in energy storage requirement can be achieved under some operating conditions . This is a common tendency under all ramp rate (RRmax) limits and constant power delivery time limits (t_{2min}) . It is important to note that the gain in reduction of energy storage requirement comes at the cost of imposing operating limits on the system as well as reducing availability i.e. if the storage requirement is selected for 95% of cumulative probability, there is a loss of 5% in availability.

The forecasting model is depended on meteorological parameters. With the advancement of weather prediction technologies, ANN can have more accurate inputs which can increase the accuracy of forecast. Considering a solar PV plant in operation, there is a limited number of years of power output data available to evaluate potential fluctuations, but weather data is available for a much longer time period. A neural network can use this data to estimate corresponding power output, which will give a wider time frame to evaluate potential fluctuations characteristic to the location.

Knowledge of potential fluctuations can assist to control the output of the solar power plant. The study focuses on two control schemes. Simple ramp rate control scheme assures a known ramp up rate, ramp down rate and a constant power output time period. This knowledge is critical for the power system operator to stabilize the system. The second scheme uses moving average to smoothen the output by reducing drastic fluctuations.

Neural networks are adapting systems to new developments and changes in the system such as degradation of inverter and panel efficiency or climate changes. With a continuously updating data stream the ANN can form a self-learning forecasting model. Operation of an energy storage system (ESS) can be assisted through a predictive model as well. Depending on the requirement of the system operator and the prediction of the model, an operator can take decisions on selecting the smoothening scheme i.e. Ramp rate control or Moving average and setting limits to maximum ramp rate, ramp up time, constant power delivery, ramp down rate and moving average window size. A group of power plants with the knowledge of the potential

output for the next 24 hrs can decide on specific output control schemes and use a shared energy storage optimally. A cluster of power plants that mitigate each other's fluctuations have a higher inertia against instability.

The results of the ANN model can be compared to other solar PV systems in close proximity to the 300 kW_p solar PV system used for the research. By comparison adjustments for differences in static properties i.e. inverter characteristics, PV panel performance, orientation and etc. can be evaluated. Results of such a study would improve the versatility of the ANN modeling approach. It is important to note that artificial neural networks are data driven. There is a clear lack of access to meteorological data in Sri Lanka, which effects the performance of the model. Increasing the time period of input data will greatly improve the accuracy and generalizability. Clouds are a major factor in determining fluctuations in solar energy. The type of cloud and cloud cover on an hourly basis (cloud profile) can be set as an input to increase accuracy. Incorporating satellite data, sky imaging data as inputs is an area where further study should be extended to.

This study is structured to focus on a single solar PV system. The research can be further extended to modeling the combined effect of power fluctuations in multiple solar PV power systems geographically dispersed. A research of such nature can shed light to the effort of quantifying limits and effects of solar PV integration to the power system.

References

[1] Sri Lanka Sustainable Energy Authority, "Sri Lanka Energy Balance 2012," 2015. [Online]. Available: http://www.energy.gov.lk. [Accessed 12 10 2017].

[2] R. George, C. Gueymard, D. Heimiller, B. Marion and D. Renne, "Solar Resource Assesment for Sri Lanka and Maldives," National Renewable Energy Laboratory, 2003.

[3] "Surface meteorology and Solar Energy," 2005. [Online]. Available: https://eosweb.larc.nasa.gov/sse/. [Accessed 02 12 2014].

[4] CEB, "Power Generation Plan 2018-2037," Ceylon Electricity Board, 2018.

[5] UNDP, "ASsesment of Sri Lnka's power sector," Asian Development Bank and UNDP, Colombo, 2017.

[6] S. A. Kalogirou, Solar Energy Engineering Processes and Systems, 2 ed., Kidlington: Acedemic Press, 2014.

[7] R. P. Thomas E. Hoff, "Quantifying PV Power Output Variability," *Solar Energy* , vol. 84, no. 10, pp. 1782-1793, 2010.

[8] Electric Power Research and Development Center CHUBU Electric Power Company, "Analysis and Forecast of PV Power Variation," CHUBU Electric Power Company, Nagoya, 2012.

[9] World Meteorological Organization, "International Cloud Atlas," 22 March 2017. [Online]. Available: www.cloudatlas.wmo.in. [Accessed 10 September 2018].

[10] Department of Meteorology United Kingdom, "MetOfiice," [Online]. Available: www.metoffice.gov.uk. [Accessed 10 September 2018].

[11] D. Matuszko, "Influence of the extent and genera of cloud cover on solar radiation intensity," *International Journal of Climatology,* vol. 32, pp. 2403-2414, 2014.

[12] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker, D. Heinemann, J. Remund, S. Mu¨ller, W. Traunmuller, G. Steinmauer, D. Pozo, J. Ruiz-Arias, V. Lara-Fanego, L. Ramirez-Santigosa, M. Gaston-Romero and L. Pomare, "Comparison of neumerical weather prediciotn solar irradiance forecasts in U.S, Canada and Europe," *Solar Energy,* vol. 94, pp. 605-326, 2013.

[13] S. Pelland, J. Remund, J. Kleissl, Oozeki and K. De Brabandere, "Photovoltaic and Solar Forecasting: State of the Art. IEA PVPS Task 14 subtask 3.1," IEA-PVPS, 2013.

[14] S. Campbell and F. Diebold, "Weather forecasting for weather derivatives," *Journal of the American statistical Association,* vol. 100, no. 469, pp. 6-16, 2005.

[15] H. Madsen, P. Pinson, G. Kariniotakis, A. Nielsen H and T. Neilsen, "Standardizing the performnace evaluation of short-term wind power prediction models," *Wind Energy,* vol. 29, no. 6, pp. 475-489, 2005.

[16] R. Perez, K. J. Schlemmer, K. Hemker Jr, D. Renne and T. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy,* vol. 84, no. 12, pp. 2161-2172, 2010.

[17] R. Huang, T. Huang, R. Gadh and N. Li, "Solar generation prediction using the ARMA model in a laboratory-level micro-grid," in *Smart Grid 2012 IEEE Third International Conference Communications (SmartGridComm)* , 2012.

[18] Y. Li, Y. Su and L. Shu, "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy,* vol. 66, pp. 78-79, 2014.

[19] Y. Chen, A. A. Thatte and L. Xie, "Multitime-scale data-driven spatiotemporal forecast of photovoltaic generation," *IEEE Transactions on Sustainable Energy,* vol. 6, no. 1, pp. 104-112, 2015.

[20] J. ,. Hornik, M. Stinchcombe and H. White, "Multilayer feedforwardnetworks are universal approximators," *Neural Networks,* vol. 2, no. 5, pp. 359-366, 1989.

[21] J. Anderson, Introduction to Neural Networks, Massachusetts: MIT Press, 1997.

[22] J. Mubiru, "Predicting total solar irradiation values using artificial neural networks,," *Renewable Energy,* vol. 33, no. 10, pp. 2329-2332, 2008.

[23] Z. Wang, Wang.F and S. Su, "Solar irradiance short term prediction model based on BP neural network," *Energy Procedia,* vol. 12, pp. 488-494, 2011.

[24] A. Mellit and A. Pavan, "A 24-h forecast of solar irradiance using artificial neural network:Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy,* vol. 84, pp. 807-821, 2010.

[25] J. Liu, W. Fang, X. Zhang and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Transaction on Sustainable Energy,* vol. 6, no. 2, pp. 434-442, 2015.

[26] R. Yacef, M. Benghanem and A. Mellit, "Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study," *Renewable Energy,* vol. 48, no. 15, pp. 146-154, 2012.

[27] J. Cao and S. Cao, "Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis," *Energy,* vol. 31, no. 15, pp. 3435-3445, 2006.

[28] D. Moore, W. Notz and M. Flinger, The basic practice of statistics, New York: W H Feeman and Company, 2013.