UNIVERSITI TEKNIKAL MALAYSIA MELAKA

SOLAR IRRADIANCE FORECASTING USING STATISTICAL AND MACHINE LEARNING METHODS

YEW POH LENG

MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

2023

Faculty of Electronic and Computer Engineering

SOLAR IRRADIANCE FORECASTING USING STATISTICAL AND MACHINE LEARNING METHODS

Yew Poh Leng

Master of Science in Electronic Engineering

2023

SOLAR IRRADIANCE FORECASTING USING STATISTICAL AND MACHINE LEARNING METHODS

YEW POH LENG

A thesis submitted
in fulfilment of the requirements for the degree of
Master of Science in Electronic Engineering

Faculty Electronic and Computer Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I declare that this thesis entitled "Solar Irradiance Forecasting using Statistical and Machine Learning Methods" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : leng ....

Name : YEW POH LENG ....

Date : 12-09-2023 ....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient interms of scope and quality for the award of Master of Science in Electronic Engineering.

Signature : ....

Supervisor's Name : ....

TS. DR. HO YIH HWA
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya
76100, Durian Tunggal, Melaka

Date : ....12 Sep 2023..........

DEDICATION

I dedicate my thesis to my beloved family who have encouraged me through this thesis. I am truly thankful for having my parents in my life that taught me to work hard for the things that I want to do. Besides, I am dedicating to my siblings who are my sister and brother that support me to continue my master degree.

ABSTRACT

The installed capacity of solar photovoltaic (PV) is continues to rise in the world and Malaysia throughout the year. In Malaysia, the average daily solar radiation is 4,000 to 5,000 Wh/m^2, with the average daily sunshine duration ranging from 4 to 8 hours. However, the output of solar energy is lack of stability due to weather variation. Solar irradiance forecasting is a crucial component in the effective integration of solar PV systems into the electrical grid. The variability of solar energy and the uncertainties associated with solar irradiance predictions pose significant challenges for grid operators and energy planners. This research project aims to develop advanced forecasting methods and methodologies for accurate and reliable solar irradiance prediction, considering the specific characteristics of local weather conditions. The study begins by analyzing the correlation between weather parameters and solar irradiance in the selected region, identifying the key variables that significantly impact solar irradiance. Quadratic regression methods are developed to forecast solar irradiance by leveraging the relationships between weather parameters. Additionally, artificial neural network (ANN), long-short term memory (LSTM), and seasonality autoregressive integrated moving average (SARIMA) methods are evaluated to determine their suitability for solar irradiance forecasting. Comparative analysis of the developed forecasting methods is conducted using evaluation metrics such as root mean square error (*RMSE*) and correlation of coefficient (*R*). The performance and suitability of different statistical and machine learning techniques for solar irradiance forecasting assisting grid operators, energy planners, and policymakers in effectively integrating solar PV systems into the electrical grid and optimizing the utilization of solar energy resources. Overall, this research project aims to advance the field of solar irradiance forecasting, enabling better planning, operation, and management of solar PV systems. By reducing uncertainties in solar energy generation, it contributes to the overall advancement of renewable energy integration and supports the transition towards a sustainable and clean energy future.

PERAMALAN SINARAN SURIA MENGGUNAKAN KAEDAH PEMBELAJARAN STATISTIK DAN PEMBELAJARAN MESIN

ABSTRAK

Kapasiti pemasangan fotovoltan (PV) suria terus meningkat di dunia dan Malaysia sepanjang tahun. Di Malaysia, purata sinaran suria harian ialah 4,000 hingga 5,000 Wh/m^2, dengan purata tempoh cahaya matahari harian antara 4 hingga 8 jam. Walau bagaimanapun, pengeluaran tenaga suria adalah kekurangan kestabilan disebabkan oleh variasi cuaca. Ramalan sinaran suria adalah komponen penting dalam penyepaduan berkesan sistem PV suria ke dalam grid elektrik. Kebolehubahan tenaga suria dan ketidakpastian yang berkaitan dengan ramalan sinaran suria menimbulkan cabaran penting bagi pengendali grid dan perancang tenaga. Projek penyelidikan ini bertujuan untuk membangunkan kaedah dan metodologi ramalan lanjutan untuk ramalan sinaran suria yang tepat dan boleh dipercayai, dengan mengambil kira ciri-ciri khusus keadaan cuaca tempatan. Kajian bermula dengan menganalisis korelasi antara parameter cuaca dan sinaran suria di rantau terpilih, mengenal pasti pembolehubah utama yang memberi kesan ketara kepada sinaran suria. Kaedah regresi kuadratik dibangunkan untuk meramalkan sinaran suria dengan memanfaatkan hubungan antara parameter cuaca. Selain itu, kaedah rangkaian saraf tiruan (ANN), ingatan jangka panjang-pendek (LSTM), dan kaedah purata bergerak bersepadu autoregresif bermusim (SARIMA) dinilai untuk menentukan kesesuaiannya untuk ramalan sinaran suria. Analisis perbandingan kaedah ramalan yang dibangunkan dijalankan menggunakan metrik penilaian seperti ralat purata kuasa dua akar (RMSE) dan korelasi pekali (R). Prestasi dan kesesuaian teknik statistik dan pembelajaran mesin yang berbeza untuk ramalan sinaran suria membantu pengendali grid, perancang tenaga dan penggubal dasar dalam menyepadukan sistem PV suria dengan berkesan ke dalam grid elektrik dan mengoptimumkan penggunaan sumber tenaga suria. Secara keseluruhannya, projek penyelidikan ini bertujuan untuk memajukan bidang peramalan sinaran suria, membolehkan perancangan, operasi dan pengurusan sistem PV suria yang lebih baik. Dengan mengurangkan ketidaktentuan dalam penjanaan tenaga suria, ia menyumbang kepada kemajuan keseluruhan integrasi tenaga boleh diperbaharui dan menyokong peralihan ke arah masa depan tenaga yang mampan dan bersih.

ACKNOWLEDGEMENTS

First of all, I am sincerely thank you to my supervisor, Dr. Ho Yih Hwa who had supervised me throughout this research. I had referred a lot of things towards him about what I had been confused upon this research. Without his patience, my thesis would have been frustrating and difficult to be done. Besides, I am grateful thank you to my faculty, Faculty of Electronics and Computer Engineering and Faculty of Electrical Engineering that allowed me to access and collect the local weather dataset from the weather station. Lastly, I also want to thank my parents and friends for their constant encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF SYMBOLS AND ABBREVIATIONS

a	–	Wien's Displacement constant
b_f	–	forget layer bias
b_i	–	input layer bias
b_o	–	bias output gate
c	–	speed of light
$\tilde{C}_t$	–	candidate value
C_t	–	new cell state
C_{t-1}	–	old cell state
E	–	intensity of radiation
e^{-x}	–	negative exponential
ε_t	–	Gaussian distribution with mean and variance
f_t	–	forget gate
h	–	Planck's constant
h_t	–	current time step hidden layer
h_{t-1}	–	previous time step hidden layer
i_t	–	input gate
L^s	–	seasonal length represented the "s"
σ	–	sigmoid layer
σ	–	Stefan-Boltzmann's constant
o_t	–	output gate layer
T	–	absolute temperature
μ	–	Observation value
W_f	–	weight of forget gate
W_{hh}	–	weight of hidden layer
W_{hx}	–	weight of input layer to hidden layer
W_{hy}	–	weight of hidden layer to output layer
W_i	–	weight of input gate
W_{ki}	–	weight of the feed-forward with backpropagation
W_o	–	weight of output gate
x	–	input variables
x_k	–	variable of new value
x_i	–	initial input variable
x_t	–	current time step input
y	–	target variable

θ	–	zenith angle
λ	–	wavelength
λ_m	–	maximum wavelength of the radiation
y_t	–	current time step output
$(1-\emptyset_p L)$	–	non-seasonal AR
$(1-\Phi_P L^P)$	–	seasonal AR
$(1-L)^d$	–	non-seasonal I
$(1-L^s)^D$	–	seasonal I
$(1+\theta_q L)$	–	non-seasonal MA
$(1+\Theta_Q L^Q)$	–	seasonal MA
ACF	–	Auto-Correlation Function
AIC	–	Akaike’s Information Criteria
ANN	–	Artificial Neural Network
AR	–	Autoregressive
BIC	–	Bayesian Information Criteria
BR	–	Bayesian Regularization
DHI	–	Diffuse Horizontal Irradiance
DNI	–	Direct Normal Irradiance
GHI	–	Global Horizontal Irradiance
I	–	Integrated
LM	–	Levenberg-Marquardt
LSTM	–	Long-Short Term Memory
MA	–	Moving Average
P	–	Pressure
PACF	–	Partial Auto-Correlation Function
PV	–	Photovoltaic
R	–	Correlation
RH	–	Relative Humidity
RMSE	–	Root Mean Square Error
RNN	–	Recurrent Neural Network
S	–	Seasonality
SARIMA	–	Seasonal Autoregressive Integrated Moving Average
SI unit	–	International System of Unit
SR	–	Solar Irradiance
Temp	–	Temperature
VIF	–	Variance Inflation Factor
WS	–	Wind Speed

LIST OF PUBLICATIONS

Yew, P.L. and Ho, Y.H., 2022. Solar Irradiance Forecasting for Malaysia using Multiple Regression and Artificial Neural Network. *Science & Technology Research Institute for Defence (STRIDE)*, vol. 15, no. 1, pp. 83-90.

CHAPTER 1

INTRODUCTION

1.1 Introduction

For the first time in history, solar photovoltaic (PV) overtook the wind in terms of the new installed capacity in 2016. Solar PV set a new record of 71GW, while wind generated 51GW (IRENA, 2017). Solar PV in 2016 is nearly half the size of the solar PV in 2020, which is 126GW. As a result, the world's and Malaysia's new installed capacity for solar PV is gradually increasing (IRENA, 2017, IRENA, 2021). Against the backdrop of the declining prices of coal, it is indeed interesting to note that plans for coal-fired power plants actually dropped by almost half in 2020. Malaysia ratified the Paris Climate Agreement in November 2016 along with the deposition of instrument with the UN Headquarters. Our Nationally Determined Contribution (NDC) is to reduce our greenhouse gas (GHG) emissions intensity of Gross Domestic Product (GDP) by 45% by the year 2030 relative to 2005 levels. This consists of 35% on an unconditional basis while a further 10% is conditional upon the receipt of support from the developed countries in terms of climate finance, technology transfer and capacity building.

With an increasing PV fraction of total power generation ("penetration"), the solar irradiance variability becomes important and grid operators and regulators alike need to understand the impact on the electric power system (technical considerations), but also on electricity markets (economic considerations). PV grid integration has been already a matter of immediate concern in some locations. For example, in the parts of Australia, Hawaii, Germany, which is as PV power continues to enjoy policy support such as feed-in tariffs or has reached price parity with existing generation technologies, there is a tendency that PV systems get deployed in large numbers very rapidly. In Germany, for example, 3 GW of new PV capacity was installed in the month of December 2011 alone, which is equivalent to about 4% of peak load. Malaysia also has taken steps to further clean energy deployment by mandating adoption of a renewable energy Feed-in Tarif (FiT) mechanism under the country's Renewable Energy Act 2011. At the end of 2016, the Authority has approved 232.0434 MW of RE installed capacities. Meanwhile, in terms of approved RE installed capacities, solar PV consisting of 98.139 MW.

Therefore, the increase in contribution of renewable energy sources into the grid is part of smart grid initiatives. The integration of renewables such as solar energy into the electrical network in the world and in Malaysia to reduce the carbon footprint. In 2020, the country of Asia that most contributes to solar energy is China then followed by India, Japan, and Korea which keep forward to green energy world. The ASEAN country that most contributes to solar energy is Vietnam, which is 16.5MW (Nam and Burke, 2021). However, the integration of solar energy is also facing the challenge for grid operators because of its intermittent due to weather variations. Despite this, the installed capacity of solar PV globally continues to increase. Thus, forecasting is becoming an important tool for system grid operators to manage solar photovoltaic (PV) energy production and satisfy the demand of energy consumers (Lawin et al., 2019).

1.2 Problem Statement

Solar irradiance forecasting plays a vital role in optimizing the integration of solar photovoltaic (PV) systems into electrical grid. The variability of solar energy and the inherent uncertainty in solar irradiance predictions pose significant challenges for grid operators and energy planners. The lack of accurate and reliable solar irradiance forecasts can lead to imbalances between electricity supply and demand, necessitating the development of additional backup power resources and negatively impacting the overall operational efficiency of the grid.

Furthermore, the intermittent nature of solar energy due to weather variations introduces significant fluctuations in solar irradiance, making it difficult to accurately estimate the amount of electricity that can be generated by solar PV systems at specific time intervals. This uncertainty in solar irradiance forecasts hinders effective grid management and prevents the optimal utilization of solar energy resources.

In fact, reliable forecast information is needed, while it can offer a better quality of service (Zhang et al., 2018). The fluctuations in solar energy and the uncertainty associated with solar energy forecasts indirectly necessitate the presence of operating reserves within electric systems. These reserves are crucial for reconciling discrepancy between energy demand and production. For example, the large plants cannot follow the existence of operating reserves in electric system when the variability of the solar resources caused ramp events (Yang et al., 2018).

Forecasting of irradiance is needed to ensure that the solar power generation consumption to residential is in good working order and that there is no gap between switching from the solar power to fuel power. Forecasting solar irradiance requires different techniques for different time horizons. For short-term forecasting, such as the next few minutes (“now-casting”) to sub-hourly predictions, the statistical and machine learning time series models that are commonly used (Abuella and Chowdhury, 2015; Jawaid and NazirJunejo, 2016; Chong et al., 2018; Fernando et al., 2019). For example, in the mediumterm forecasting that is a few hours ahead ("intra-day"), meteorological models relying on combinations of weather data-derived data with ground observations (such as sky cameras) that produces the greatest results. While, for example, in the long-term forecasts is day(s) ahead, which require the use of complex numerical weather prediction (NWP) models (Massidda and Marrocu, 2017) such as WRF (Weather Research and Forecasting) or ECMWF (European Centre for Medium-Range Weather Forecasts).

While various statistical and machine learning methods have been proposed for solar irradiance forecasting, there is a need for further research to identify the most suitable and reliable techniques that can effectively address the challenges associated with accurate solar irradiance forecast. The development of robust and accurate forecasting methods that can capture the complex relationship between local weather measurements and solar irradiance patterns is crucial.

Therefore, this study aims to investigate and develop advanced forecasting methodologies for solar irradiance prediction, considering the specific characteristics of the local weather conditions. The research explores the potential of statistical and machine learning techniques, such as regression learning, artificial neural networks, long-short term memory, and seasonality autoregressive integrated moving average, to improve the accuracy and reliability of solar irradiance forecasts. The outcomes of this research provides valuable insights and tools for grid operators, energy planners, and policymakers to enhance the integration of solar PV systems into the electrical grid and promote the effective utilization of solar energy resources.

1.3 Research Contributions

The research on solar irradiance forecasting makes several key contributions to the field:

1. Advanced forecasting methods

   – The research aims to improve the accuracy and reliability of solar irradiance forecasting by incorporating the specific characteristics of local weather conditions. This contributes to the development of robust forecasting methods that can capture the complex relationships between weather parameters and solar irradiance. Before evaluating the regression learning methods and artificial neural network methods, the correlation between weather parameters and solar irradiance is analyzed.

2. Evaluation of Statistical and Machine Learning Methods

   – The research evaluates the performance of various statistical and machine learning methods that include regression learning, artificial neural network, long-short term memory, and seasonality autoregressive integrated moving average. By comparing and assessing the effectiveness of these methods, the research provides valuable insights into the most suitable approaches for solar irradiance forecasting.

3. Addressing Uncertainty and variability

   – The research acknowledges the inherent uncertainty and variability associated with solar energy and solar irradiance forecasts. The research aims to reduce the uncertainties and fluctuations in solar irradiance forecasting by developing accurate and reliable forecasting methods. This contribution enhances the ability of grid operators and energy planners to manage the integration of solar PV systems effectively and optimize the utilization of solar energy resources.

4. Practical Applications and Implications