การปรับปรุงการพยากรณ์กำลังไฟฟ้าของเซลล์สุริยะจากการทำนายอากาศเชิงตัวเลข

นายศุภชัย สุขสโมสร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2561

AN IMPROVEMENT OF PHOTOVOLTAIC POWER FORECASTING FROM NUMERICAL WEATHER PREDICTION

Mr. Supachai Suksamosorn

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2018

Thesis Title	AN IMPROVEMENT OF PHOTOVOLTAIC POWER FORECASTING FROM NUMERICAL WEATHER PREDICTION
By	Mr. Supachai Suksamosorn
Field of Study	Electrical Engineering
Thesis Advisor	Assistant Professor Jitkomut Songsiri, Ph.D.
Thesis Co-Advisor	Associate Professor Naebboon Hoonchareon, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

..... Dean of the Faculty of Engineering
(Assoc. Prof. Supot Teachavorasinskun, D.Eng.)

THESIS COMMITTEE

..... Chairman
(Professor David Banjerdpongchai, Ph.D.)

..... Thesis Advisor
(Assistant Professor Jitkomut Songsiri, Ph.D.)

..... Thesis Co-Advisor
(Associate Professor Naebboon Hoonchareon, Ph.D.)

..... External Examiner
(Somphop Asadamongkol, Ph.D.)

ศุภชัย สุขสโมสร: การปรับปรุงการพยากรณ์กำลังไฟฟ้าของเซลล์สุริยะจากการทำนายอากาศเชิงตัวเลข
(AN IMPROVEMENT OF PHOTOVOLTAIC POWER FORECASTING FROM NUMERICAL WEATHER PREDICTION)
อ. ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร.จิตโกมุท ส่งศิริ,
อ. ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ.ดร.แนบบุญ หุนเจริญ, 106 หน้า

วิทยานิพนธ์ฉบับนี้นำเสนอแบบจำลองเพื่อปรับปรุงผลการทำนายกำลังผลิตของเซลล์สุริยะรายชั่วโมง ในระยะหนึ่งวันล่วงหน้าจากการทำนายอากาศเชิงตัวเลข โดยมีข้อกำหนดว่าแบบจำลองต้องให้ผลการทำนายภายในเวลา 14.00 น. ของทุกวัน เพื่อนำไปใช้เป็นแนวทางของ แผนกวางแผนปฏิบัติการผลิตระยะสั้น, กองวางแผนปฏิบัติการผลิตไฟฟ้า, ฝ่ายควบคุมระบบกำลังไฟฟ้า, การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย ในการวางแผนการเดินเครื่องของโรงไฟฟ้าในวันถัดไป ข้อมูลที่นำมาใช้ในการศึกษาในงานนี้ได้มาจาก 2 แหล่งข้อมูล คือ ข้อมูลที่ได้จากเครื่องมือวัดสภาพอากาศและเครื่องมือวัดกำลังผลิตของเซลล์สุริยะที่ถูกติดตั้งบน ตึกวิศวกรรมไฟฟ้า, คณะวิศวกรรมศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย และข้อมูลที่ได้มากจากผลการทำนายอากาศเชิงตัวเลข เนื่องจากผลการทำนายความเข้มแสงจากการทำนายอากาศเชิงตัวเลขยังมีความคลาดเคลื่อนสูงเพราะการทำนายตำแหน่งของเมฆมีความยาก ดังนั้นเราจึงนำเสนอแบบจำลองเพื่อใช้ปรับปรุงผลการทำนายความเข้มแสงจากการทำนายอากาศเชิงตัวเลขที่สามารถแบ่งออกได้เป็น 2 ประเภทตามขั้นเวลา คือ แบบจำลองที่มีขั้นเวลาของการพยากรณ์เป็นรายชั่วโมง และแบบจำลองที่มีขั้นเวลาของการพยากรณ์เป็นรายวัน แบบจำลองในแต่ละประเภทนั้น สามารถแยกออกได้เป็น 2 วิธีตามสมการคณิตศาสตร์ คือ แบบจำลองถดถอยเชิงเส้น กับแบบจำลองถดถอยเชิงเส้นที่ใช้ร่วมกับตัวกรองคาลมาน วิทยานิพนธ์นี้เสนอผลงานที่ต่างจากงานวิจัยที่ผ่านมาสองประเด็น คือ หนึ่ง การใช้หลักการเรียนรู้ทางสถิติเพื่อเลือกตัวแปรขาเข้าที่มีความสัมพันธ์ต่อความเข้มแสงเพื่อใช้ในแบบจำลองถดถอยเชิงเส้น อันได้แก่วิธี การวิเคราะห์สหสัมพันธ์บางส่วน การถดถอยแบบทีละขั้น การถดถอยแบบเซตย่อย ส่วนประเด็นที่สอง คือ การปรับปรุงค่าสัมประสิทธิ์ของแบบจำลองถดถอยเชิงเส้นด้วยตัวกรองคาลมาน ที่คำนึงถึงข้อจำกัดทางปฏิบัติด้านข้อมูลวัด ณ เวลาที่คำนวณค่าพยากรณ์ ผลการทดลองพบว่า แบบจำลองถดถอยเชิงเส้น และแบบจำลองถดถอยเชิงเส้นที่ใช้ร่วมกับตัวกรองคาลมาน มีค่าความคลาดเคลื่อนของความเข้มแสงแบบรากกำลังสองเฉลี่ยอยู่ที่ 160 และ 157 วัตต์ต่อตารางเมตร ตามลำดับ ในขณะที่แบบจำลองการทำนายอากาศเชิงเลขมีสมรรถนะค่าความผิดพลาดอยู่ที่ 250 วัตต์ต่อตารางเมตร เมื่อแปลงผลการปรับปรุงค่าพยากรณ์ความเข้มแสงเป็นค่าพยากรณ์กำลังไฟฟ้าด้วยแบบจำลองถอดถอยที่มีตัวแปรเข้าเป็นความเข้มแสงและอุณหภูมิ พบว่า แบบจำลองที่เสนอขึ้นที่มีสมรรถนะดีที่สุด ได้แก่ แบบจำลองถดถอยเชิงเส้นด้วยตัวกรองคาลมานที่มีขั้นเวลารายวัน แบบจำลองดังกล่าวให้ค่าความคลาดเคลื่อนรากกำลังสองเฉลี่ยที่ปรับให้เป็นมาตรฐานด้วยค่ากำลังผลิตติดตั้ง (8 กิโลวัตต์) เท่ากับ 13 % ซึ่งลดลงจากวิธีทำนายอากาศเชิงตัวเลขไป 9 %

ภาควิชา ..... วิศวกรรมไฟฟ้า	**ลายมือชื่อนิสิต** .....
สาขาวิชา ..... วิศวกรรมไฟฟ้า	**ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก** .....
ปีการศึกษา 2561	**ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม** .....

5970330821 : MAJOR ELECTRICAL ENGINEERING

KEYWORDS : SOLAR IRRADIANCE FORECASTING /
SOLAR POWER FORECASTING / NUMERICAL WEATHER PREDICTION /
MODEL OUTPUT STATISTICS / KALMAN FILTER / VARIABLE SELECTION

SUPACHAI SUKSAMOSORN : AN IMPROVEMENT OF PHOTOVOLTAIC POWER FORECASTING FROM NUMERICAL WEATHER PREDICTION.
ADVISOR : ASST. PROF. JITKOMUT SONGSIRI, Ph.D.,
CO-ADVISOR : ASSOC. PROF. NAEBBOON HOONCHAREON, Ph.D., 106 pp.

This thesis presents forecasting models for improving the hourly predicted solar power from Numerical Weather Prediction (NWP) in one day horizon. A practical constraint is that the models must provide the predicted solar power before 14.00 hrs daily so that the Short-term Operation Planning Section, Generation Operation Planning Department, Power System Control and Operation Division, Electricity Generating Authority of Thailand can use the prediction to plan dispatching of the next day. Available data used in this research are from two sources: local weather measurements and power meter installed at the top of Electrical Engineering Building, Faculty of Engineering, Chulalongkorn University and NWP outputs. The prediction from NWP still has a significant error due to a complexity of cloud movement, so we propose models to refine the predicted solar irradiance from NWP. The models can be separated into two types according to time steps of updating rules: hourly-step and daily-step models. Each of these two models can also be divided into two groups according to mathematical equations of updating rules: model output statistics (MOS) which is a linear regression model and MOS used with Kalman filter (MOS+KF). This thesis has two contributions. Firstly, we applied statistical methods including partial correlation analysis, stepwise regression, and subset regression to select relevant inputs of MOS. The second contribution is that the limitation of data at the time of forecast was taken into consideration to modify the forecast updating rules by MOS+KF. The experimental results showed that MOS and MOS+KF models provide the root mean squared error of solar irradiace of 160 and 157 kW/m^2, respectively, while the NWP error was 250 kW/m^2. When the irradiance forecasts were converted to the predicted solar power, the best among all proposed model was daily-step MOS+KF model which provided the normalized (by capacity of 8 kW) root mean squared error of 13 %, reduced from NWP by 9 %.

Department : Electrical Engineering	Student's Signature .....
Field of Study : Electrical Engineering	Advisor's Signature .....
Academic Year : 2018	Co-advisor's Signature .....

Acknowledgments

First and foremost, I would like to thank to my supervisor, Assistant Professor Jitkomut Songsiri for kindly supporting and providing helpful advice throughout my master's degree study. I would like to thank to my co-supervisor, Associate Professor Naebboon Hoonchareon for kindly supporting and providing helpful recommendation and computing resources. I would like to thank Professor David Banjerdpongchai and Dr.Somphop Asadamongkol for taking your valuable time to be committees of this thesis examination. I would like to thank Electricity Generating Authority of Thailand for giving me a scholarship to study in master's degree at Chulalongkorn University. I would also like to thank all students of the Control Systems Research Laboratory for supporting, suggestion, and good friendship. Last but not least, I would like to thank my family who has always supported me throughout my life.

Contents

List of Tables

List of Figures

Figure Page

Notations

Variable	Meaning	Unit
I	Solar irradiance	$\mathrm{W/m^2}$
I_{sc}	Solar constant	$\mathrm{W/m^2}$
I_0	Extraterrestrial solar irradiance	$\mathrm{W/m^2}$
I_{clr}	Solar irradiance from clear sky model	$\mathrm{W/m^2}$
$\hat{I}$	Predicted solar irradiance	$\mathrm{W/m^2}$
T	Temperature	°C
RH	Relative humidity	%
WS	Wind speed	m/s
WD	Wind direction	°
UV	UV index	–
tcc	Total cloud cover	–
$\cos\theta$	Cosine of solar zenith angle	–
$\hat{k}$	Predicted clear sky index	–
P	Electrical solar power	kW
$\hat{P}$	Predicted electrical solar power	kW

Variable symbols

- The variable $x(t)$ means the actual value of the measured variable and the variable $\hat{x}(t)$ means the predicted value of the measured variable.
- The variable $\hat{I}_{\mathrm{A}}(t)$ means the predicted solar irradiance from the method A.
- The notation $x(t)$ denote the variable x at time t and if the index of day is also used, then we use the notation $x^{(d)}(t)$ to refer x of day d at time t.

CHAPTER I

INTRODUCTION

Recently, the awareness of high dependence of electricity generation by natural gas in Thailand has been raised. Therefore, the government of Thailand plans to reduce the dependency of natural gas and finds new electricity sources. As in Power Development Plan (PDP) 2015, at year end 2036 the proportion of renewable energy will increase to 27.9% of installed generating capacity of Thailand and solar power plant has the largest proportion about 30.5% of a renewable energy type. However, Photovoltaic (PV) cells cannot generate stable power due to large solar irradiance fluctuations and this problem can reduce the power system reliability and stability. To deal with this problem, solar irradiance forecasting is necessary for operators in order to prepare the generation and voltage of the power system before problems occur. Moreover, solar irradiance forecasting can also be used in planning purpose to decrease the operating cost of conventional power plants. Typically, solar irradiance forecasting and solar power forecasting can be separated into three types by forecasting horizon, i) nowcasting (a minute up to an hour ahead), ii) very short-term (1 up to 6 hours ahead), and iii) short-term (6 hours up to several days ahead). In this work, we focus only in horizon of one day ahead to serve the planning of power plant for the next day.

Literature on solar forecasting can be found in many review articles recently [3–5]. Common techniques for solar irradiance forecasting and solar power forecasting in short-term horizon (or day-ahead) are statistical methods, physical methods, and hybrid methods [3]. The statistical methods use the historical data of the local weather measurements for modeling the forecasting model such as time series and machine learning. The physical methods use equations derived from physics to predict the weather variables and the most popular method is Numerical Weather Prediction (NWP). NWP models employ partial differential equations in order to model dynamics of Earth's atmosphere and the equations are solved to obtain the predicted weather variables [6]. Combinations of two or more existing methods for solar irradiance forecasting and solar power forecasting are called hybrid methods. There are various methods to predict solar irradiance and solar power, however, the popular methods for the day-ahead applications are physical methods and hybrid methods [4,5]. According to [4,5], they reported that there were a few studies that used NWP variables for intra-hour forecasts but there were many studies that used NWP variables for day ahead forecasts. Moreover, the hybrid methods which were the combination of NWP variables and neural networks usually outperformed the other methods because the hybrid methods use an advantage of one method to reduce a disadvantage of the other method. Besides, they also reported that if the forecast horizon is longer than one day, the NWP variables are recommended to use as inputs of the forecasting models. The benefit of NWP is that NWP can provide the future meteorological trends which are useful for explaining solar irradiance and solar power in day-ahead horizon.

However, the predicted solar irradiance from NWP models may contain some biases and its magnitudes is not significantly small. In [7], the various NWP models were used to

predict the solar irradiance from 1 day up to 7 days and the performance of global NWP models were better than regional NWP models. RMSE of the predicted solar irradiance one day in advance from global NWP models was about 90 to 140 W/m^2 and RMSE of regional NWP models was about 160 W/m^2. Therefore, some post-processing methods are needed. The post-processing method widely used to refine the predicted solar irradiance from NWP models is Model Output Statistics (MOS) [2]. MOS is a regression model that explains a variable of interest by relationships between the variable of interest and the weather variables. According to [8], MOS is used to refine the predicted wind speed from NWP and it can reduce the prediction error by half. Moreover, MOS can be used as a bias correction of the predicted solar irradiance from NWP [9, 10]. However, a disadvantage of MOS is that the regression coefficients are fixed and it may provide high forecasting error due to seasonality. Therefore, if a model needs local measurement data to adjust the parameters of MOS, then the Kalman filter (KF) can be used to serve this purpose. The implementation of KF based on MOS is called MOS+KF. In [11,12], MOS+KF was used to predict the solar irradiance and the results showed that MOS+KF models outperformed the persistence forecast and the predicted from NWP by 15 % of RMSE.

There are two goals for this work. The first goal is to refine the predicted solar irradiance from NWP by local weather measurements (MOS+KF method). However, MOS+KF models should include only the important variables to solar irradiance. From the previous studies, there is no attempt to considers the inputs of the model. Therefore, we use the statistical methods to identify the important variable to solar irradiance. We then include the relevant variables to the forecasting models. The second goal is to provide the predicted solar power from 7.00 to 16.00 hrs of the next day by 13.00 hrs daily. This goal is used to serve the Short-term Operation Planning Section (SOPS), Generation Operation Planning Department, Electricity Generating Authority of Thailand for dispatching the power plant of the next day in the afternoon. Though all our knowledge, the previous studies did not take this practical constraint into account for forecasting solar irradiance. Therefore, the contributions of this thesis is to present the solar irradiance forecasting models with take two goals in consideration into account. The proposed models can be separated into two types, the first type is hourly-step models which time index evolves in an hourly manner and the second types is daily-step models which time index evolves in daily manner. Hourly-step models have only one model but daily-step models consist of many sub-models which are equal to the number of hour of interest. Therefore, the daily-step models have more complexity than hourly-step models. Both hourly-step and daily-step models consist of MOS and MOS+KF methods to predict the solar irradiance. The inputs of all models are selected by statistical methods which are partial correlation, stepwise regression, and subset regression and are used in both hourly-step and daily-step models. In the forecasting procedure, MOS models just require the outputs from NWP models to provide the predicted solar irradiance but MOS+KF models require both the outputs from NWP models and the measured data to update the parameters and provide the predicted solar irradiance. As the constraint, we provide the predicted data by 13.00 hrs, therefore, the measured data between 14.00 to 16.00 hrs are still not available. We then modify the equations of Kalman filter for updating the parameters by relying on the MMSE estimator. The best forecast values are the conditional

mean of the value on the next day given the information of the previous day. The details of all proposed models are described in Chapter 7.

This thesis is organized as follows. The objective, scope of work, research methodology, and expected outcome of this thesis are described in Chapter 2. The basic knowledge about solar irradiance forecasting and the electrical equivalent circuit of PV cells are presented in Chapter 3. Common techniques for solar irradiance forecasting are described in Chapter 4. In Chapter 5, we present the post-processing methods that are widely used to refine the predicted weather variables from NWP models. The problem statement derived from practical goals and constraints of this work are presented in Chapter 6. The details of the proposed models that take into account the practical constraints are expressed in Chapter 7. In Chapter 8, we describe the sources of the data that we use in this thesis. Experimental results and discussion of this work are described in Chapter 9. Conclusion and suggestion of this work are summarized in Chapter 10.

CHAPTER II

THESIS OVERVIEW

2.1 Objective

This thesis aims to improve Photovoltaic (PV) power forecasting from Numerical Weather Prediction (NWP) in order to guide the Short-term Operation Planning Section, Generation Operation Planning Department, Power System Control and Operation Division, Electricity Generating Authority of Thailand for power plant dispatching of the next day.

2.2 Scope of work

1. The two PV sites with installed capacity of 8 and 15 kW and local weather measurements that used in this work are located at Electrical Engineering Building, Chulalongkorn University.
2. This work provide hourly forecasts with one-day horizon.

2.3 Research methodology

1. Review the background of solar irradiance and previous works of solar irradiance forecasting.
2. Collect and analyze the data from Chulalongkorn University's Building Energy Management System (CUBEMS) and Numerical Weather Prediction (NWP).
3. Improve the accuracy of predicted solar irradiance from Numerical Weather Prediction by regression models with time-varying coefficients.
4. Convert the refined predicted solar irradiance into Photovoltaic power using existing methods such as physical models or Artificial Neural Networks.
5. Analyze and compare the results between this work and the previous works with discussion about advantages and disadvantages.

2.4 Expected outcome

1. Forecasting models for improving hourly Photovoltaic power forecasts for one day in advance.
2. Computer codes of forecasting algorithms.

CHAPTER III

BACKGROUND ON SOLAR FORECASTING

This chapter describes about background of solar irradiance forecasting which are astronomical variables, geographical variables, clear sky model, forecasting performance, and solar irradiance to PV power conversion. Solar irradiance under clear sky conditions (I_{clr}), clear sky index (k), cosine of solar zenith angle ($\cos\theta$) and performance indices are calculated from the equations in this chapter. These variables are used to refine the predicted solar irradiance from NWP in Chapter 5.

3.1 Astronomical and geographical variables

The solar irradiance is related to a location of interest and the Sun's position. The Sun's position always change due to rotation of the Earth and the Earth orbit around the Sun. Therefore, the necessary variables for calculating clear sky model are latitude, declination angle, hour angle, and zenith angle which can be calculated as follows.

- Latitude (ϕ) is a latitude of a location of interest [13]. The latitude range is between $-90^\circ \leqslant \phi \leqslant 90^\circ$ where positive and negative sign stand for upper and lower equator respectively.

- Declination angle (δ) is degree angle between the Sun and equator plane at noon [13]. The declination angle range is between $-23.45^\circ \leqslant \delta \leqslant 23.45^\circ$ where positive and negative sign stand for north and south of equator plane respectively. Declination angle is defined by

$$\begin{aligned}\delta = {} & 0.006918 - 0.399912\cos\theta_0 + \sin\theta_0 - 0.006759\cos 2\theta_0 + 0.000907\sin 2\theta_0 \\ & - 0.002697\cos 3\theta_0 + 0.00148\sin 3\theta_0,\end{aligned} \tag{3.1}$$

  where θ_0 is the Sun's position angle which depends on day of the year (d_n) and $\theta_0 = 2\pi(d_n - 1)/365$.

- Hour angle ($\omega(t)$) is degree angle of the Sun's position east or west due to rotation of the Earth which is changed 15° per hour. ω is negative values in the morning, positive values in the afternoon, and equal to 0° at noon [13]. Hour angle is defined by

$$\omega(t) = 15^\circ|12 - \text{LAT}|, \tag{3.2}$$

  where LAT is the local apparent time (hour) determined by the local standard time in hour (LST), the equation of time in minutes (ET), the geographical longitude of the site in degrees (LS), and standard meridian in degrees (LSM) for the time zone:

$$\text{LAT} = \text{LST} + \frac{\text{ET}}{60} \pm \frac{\text{LSM} - \text{LS}}{15}. \tag{3.3}$$

The positive sign in the last term is for places west of Greenwich and the negative sign is for places east of Greenwich. The equation of time is defined by

$$\mathrm{ET} = 9.87\sin(2f) - 7.53\cos(f) - 1.5\sin(f), \tag{3.4}$$

where $f = 2\pi(n-81)/364$, n is Julian day.

Zenith angle ($\theta(t)$) is degree angle between the Sun's position and vertical line at time t [13] as shown in Figure 3.1 defined by

$$\cos\theta(t) = \cos\phi\cos\delta\cos\omega(t) + \sin\phi\sin\delta. \tag{3.5}$$

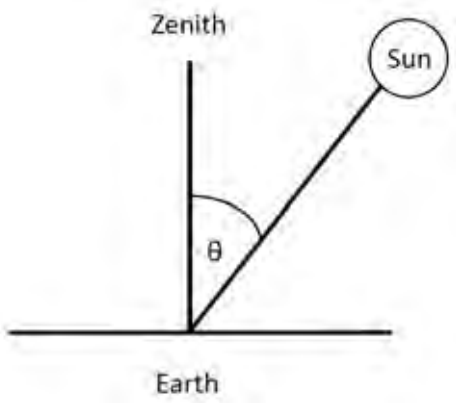


Figure 3.1: Solar zenith angle.

3.2 Solar irradiance

Solar irradiance is power in the form of electromagnetic from the Sun represent in W/m^2. A unit area of surface that perpendicular to the solar irradiance outside the Earth's atmosphere at mean Sun-Earth distance called solar constant (I_{sc}) = 1,367 W/m^2. A unit area of surface that perpendicular to the solar irradiance outside the Earth's atmosphere at any Sun-Earth distance called extraterrestrial solar irradiance ($I_0(t)$) [13]. The Sun-Earth distance ratio change with day of the year defined by

$$\frac{\overline{R}}{R} = \frac{1}{1 - 0.033\cos(2\pi d_n/Y)}, \tag{3.6}$$

where $\overline{R} = 1.49 \times 10^{11}$ m is the mean value of the Sun-Earth distance, R is the actual Sun-Earth distance, d_n is day of the year, and Y is the total day of year. Then, $I_0(t)$ is expressed as

$$I_0(t) = \left(\frac{\overline{R}}{R}\right)^2 I_{\mathrm{sc}}\cos\theta(t). \tag{3.7}$$

Solar irradiance can be classified into two categories, short wavelength ($0.3 - 3\mu$m) and long wavelength ($> 3\mu$m) [14]. PV cells generate electricity energy from short wavelength of solar irradiance only because the power in long wavelength of solar irradiance is not enough to stimulate the electrons of PV cell across the band gap as shown in Figure 3.2.

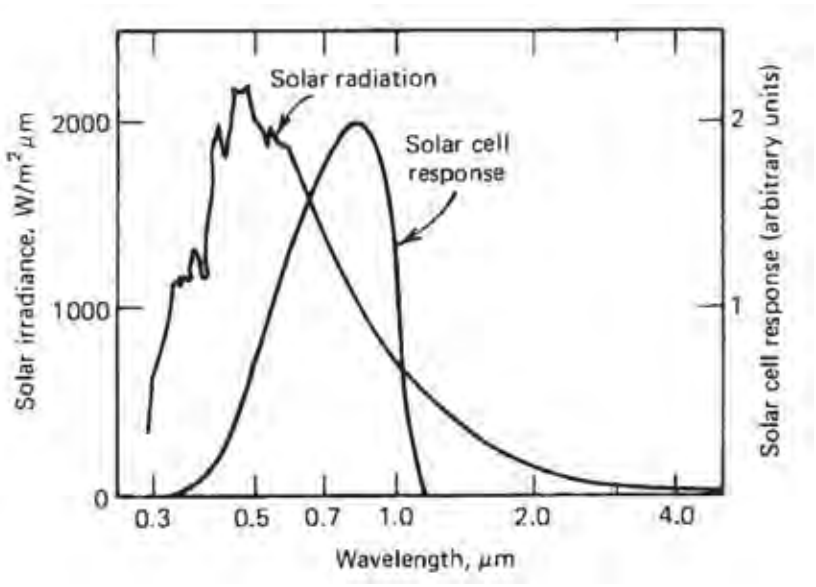


Figure 3.2: Spectral distribution of extraterrestrial solar radiation and spectral response of a silicon solar cell.

Source: Solar engineering of thermal processes [14]

According to [3], when solar irradiance incident on the Earth's atmosphere, it distinct into two parts as shown in Figure 3.3. The first part, Direct Normal Irradiance (DNI) is solar irradiance come in a straight line from the Sun. The second part, Diffuse Horizontal Irradiance (DHI) is scattering of solar irradiance by the Earth's atmosphere and the surroundings. Global Horizontal Irradiance (GHI) is the total irradiance from the Sun on a surface of interest and it is defined as

$$I(t) = \text{GHI}(t) = \text{DHI}(t) + \text{DNI}(t)\cos\theta(t). \tag{3.8}$$

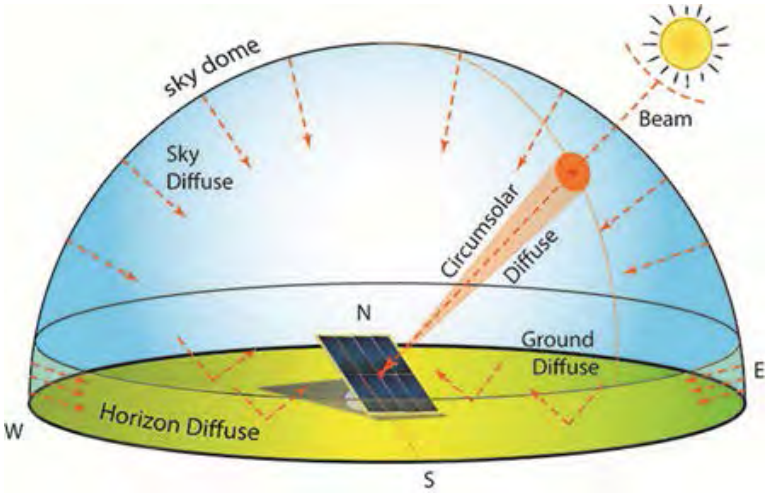


Figure 3.3: Solar irradiance components.

Source: `http://bdewilde.github.io/blog/blogger/2012/10/26/classification-of-hand-written-digits-3/`

The power of solar irradiance is reduced due to reflection and scattering of particles, gases, and fluid in the Earth's atmosphere until reach the Earth surface. The parameters that usually used for explaining the impact from the Earth's atmosphere are air mass and Linke turbidity. Both air mass and Linke turbidity are also necessary parameters for clear sky model which are described as follows.

Air mass (AM(t)) is the ratio of distance of solar irradiance reaching the Earth's surface [13]. The shortest distance occurs when the Sun's position is at vertical line and provide maximum value (AM(t)=1). AM(t) is defined as

$$\mathrm{AM}(t) = \frac{1}{\cos\theta(t) + 0.15(93.885 - \theta(t))^{-1.253}}. \tag{3.9}$$

Linke turbidity ($T_{\mathrm{L}}(t)$) is a coefficient to indicate the atmospheric turbidity. The coefficient is high when the atmospheric pollution is high. According to [15], Linke turbidity is defined as

$$T_{\mathrm{L}}(t) = \left(\frac{11.1\ln(bI_{\mathrm{sc}}/\mathrm{DNI}_{\mathrm{clr}})}{\mathrm{AM}(t)}\right) + 1, \tag{3.10}$$

where $b = 0.664 + 0.163e^{h/8000}$ is a coefficient, h is altitude of site in meters, and $\mathrm{DNI}_{\mathrm{clr}}$ is clear sky direct normal irradiance. The summary of the relevant variables to solar irradiance irradiance forecasting are shown in Table 3.1.

Table 3.1: Relevant variables in solar forecasting.

Variables	**Spatial**	**Temporal**	**Constant**
Latitude (ϕ)	$-90° \leqslant \phi \leqslant 90°$		
Declination angle (δ)		$-23.45° \leqslant \delta \leqslant 23.45°$	
Hour angle ($\omega(t)$)		change 15° per hour	
Zenith angle ($\theta(t)$)		$0° \leqslant \theta(t) \leqslant 90°$	
Solar constant (I_{sc})			1,367 W/m^2
Extraterrestrial solar irradiance (I_0)		depend on Sun-Earth distance	
Air mass (AM(t))		$1 \leqslant \mathrm{AM}(t)$	
Linke turbidity ($T_{\mathrm{L}}(t)$)		$1 \leqslant T_{\mathrm{L}}(t)$	

3.3 Clear sky model

Clear sky model is a mathematical equation to calculate solar irradiance on the Earth's surface under clear sky conditions ($I_{\mathrm{clr}}(t)$) [3]. There are various proposed clear sky models

and each clear sky model requires different meteorological variables such as altitude, gases, vapor, and Linke turbidity. For example, Kasten clear sky model [15] can be calculated by

$$I_{\text{clr}}(t) = 0.84 I_0(t) \cos(90 - \theta(t)) e^{-0.027(f_1 + f_2(T_L - 1))\text{AM}(t)}, \tag{3.11}$$

where h is height above mean sea level (in meters) at measured position, $f_1 = e^{-h/8000}$ and $f_2 = e^{-h/1250}$ (f_1 and f_2 are coefficients related to altitude of measured position). Kasten clear sky model depends more on AM(t), so $I_{\text{clr}}(t)$ is underestimate when AM(t) is high. In order to correct $I_{\text{clr}}(t)$, the new clear sky model was proposed called Ineichen clear sky model [15]. Ineichen clear sky model is defined as

$$I_{\text{clr}}(t) = a_1 I_0(t) \cos(90 - \theta(t)) e^{-a_2(f_1 + f_2(T_L - 1))\text{AM}(t)}, \tag{3.12}$$

where $a_1 = 5.09 \times 10^{-5} h + 0.868$, $a_2 = 3.92 \times 10^{-5} h + 0.0387$. Ineichen clear sky model provides accurate $I_{\text{clr}}(t)$ than Kasten clear sky model when AM(t) is high. $I_{\text{clr}}(t)$ is also used to calculate clear sky index. Moreover, $I_{\text{clr}}(t)$ and clear sky index are also used for persistence forecast.

Clear sky index ($k(t)$) is a ratio between measured solar irradiance and solar irradiance from clear sky model. $k(t)$ is defined as

$$k(t) = \frac{I(t)}{I_{\text{clr}}(t)}, \tag{3.13}$$

when $I(t)$ is solar irradiance at time t, $I_{\text{clr}}(t)$ is solar irradiance from clear sky model at time t. In addition, we define $\hat{k}(t)$, if the numerator is $\hat{I}(t)$. Moreover, the clearness index is also usually used to calculate persistence forecast.

Clearness index ($K(t)$) is a ratio between measured solar irradiance and solar constant corresponding with solar zenith angle. Clearness index is defined as

$$K(t) = \frac{I(t)}{I_0(t)}. \tag{3.14}$$

$I_{\text{clr}}(t)$ is always less than or equal to $I_0(t)$ because $I_{\text{clr}}(t)$ takes the turbidity of atmosphere into account but $I_0(t)$ does not. Therefore, $k(t)$ is also always higher than or equal to $K(t)$.

Persistence forecast is a simple forecasting method and is usually used to compare with other advanced forecasting methods. The predicted solar irradiance from this method can be calculated by assume that the conditions in the future are the same as the present conditions. According to [3], there are many ways to predict solar irradiance using persistence forecast as follows.

1. Persistence forecast using clear sky index: This method assumes that the clear sky index in the future is the same as the present and it is defined as

$$k(t + \Delta t) = k(t). \tag{3.15}$$

   Then, the solar irradiance in the future can be calculated by using (3.13) and (3.15)

$$\hat{I}(t + \Delta t) = k(t) I_{\text{clr}}(t + \Delta t). \tag{3.16}$$

2. Persistence forecast using clearness index: This method assumes that the clearness index in the future is the same as the present and it is defined as

$$K(t+\Delta t) = K(t). \tag{3.17}$$

Then, the solar irradiance in the future can be calculated by using (3.14) and (3.17)

$$\hat{I}(t+\Delta t) = K(t)I_0(t+\Delta t). \tag{3.18}$$

3. Persistence forecast using current solar irradiance: This method assumes that the solar irradiance in the future is the same as the present and it is defined as

$$\hat{I}(t+\Delta t) = I(t). \tag{3.19}$$

3.4 Forecasting performance index

Forecasting performance index is used to evaluate the forecast models. The forecasting models that provide low forecasting error will also provide forecasting performance index close to zero. Common forecasting performance indices for evaluating the solar irradiance forecast model are summarized as follows.

1. Mean Absolute Error (MAE): MAE is used to evaluate the forecasting performance by mean of accumulated prediction error thus we can see the overall prediction error from MAE.

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|\widehat{I}(t) - I(t)\right|. \tag{3.20}$$

2. Mean Absolute Percentage Error (MAPE): MAPE is MAE in the form of percentage which can be comparable with other works.

$$\text{MAPE} = \frac{100\%}{N}\sum_{t=1}^{N}\left|\frac{\widehat{I}(t) - I(t)}{I(t)}\right|. \tag{3.21}$$

3. Mean Bias Error (MBE): MBE is an averaged prediction error and we can see the type of the prediction error which are overestimate or underestimate.

$$\text{MBE} = \frac{1}{N}\sum_{t=1}^{N}(\widehat{I}(t) - I(t)). \tag{3.22}$$

4. Root Mean Square Error (RMSE): RMSE is a kind of performance index which is penalty more on the large prediction error.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\widehat{I}(t) - I(t))^2}. \tag{3.23}$$

5. Normalized Root Mean Square Error (NRMSE): NRMSE is RMSE in the form of percentage. This performance index is used to evaluate the forecasting model and it can be used to compare the performance to the other works. There are many ways to calculate NRMSE which can be expressed as follows.

(a) Normalized by the mean of $I(t)$ [11, 16–19]

$$\text{NRMSE} = 100\% \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{I}(t) - I(t))^2}}{\frac{1}{N}\sum_{t=1}^{N} I(t)}. \tag{3.24}$$

(b) Normalized by the range between maximum and minimum value of $I(t)$ [20]

$$\text{NRMSE} = 100\% \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{I}(t) - I(t))^2}}{I_{\max} - I_{\min}}. \tag{3.25}$$

(c) Normalized by the standard deviation [21]

$$\text{NRMSE} = 100\% \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{I}(t) - I(t))^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(I(t) - \bar{I})^2}}. \tag{3.26}$$

(d) Normalized by installed capacity of the PV site [22–25]

$$\text{NRMSE} = 100\% \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{I}(t) - I(t))^2}}{\text{PV}_{\text{cap}}}. \tag{3.27}$$

where $\hat{I}(t)$ is forecasted solar irradiance at time t, N is the number of data, $\bar{I}$ is mean value of I data, and PV_{cap} is the installed capacity of the PV site.

3.5 Conversion of GHI into electrical power

This section explains the details of various conversion models for converting GHI to solar power. The methods can be categorized into two main approaches, i) based on equivalent circuit models and ii) based on grey-box models such as neural network.

Manufacturer's datasheet of PV cell

In order to determine the parameters of PV cell, a datasheet from the manufacturer is needed. The information in the datasheet is obtained under the Standard Test Conditions (STC) as follows: $I_{\text{stc}} = 1000$ W/m^2, $T_{c,\text{stc}} = 25$ °C, and $\text{AM}_{\text{stc}} = 1.5$ where T_c is a temperature of PV cell. The important parameters in manufacturer's datasheet are open-circuit voltage ($v_{\text{oc,stc}}$), short-circuit current ($i_{\text{sc,stc}}$), maximum power ($P_{\text{max,stc}}$), voltage at maximum power ($v_{\text{mpp,stc}}$), current at maximum power ($i_{\text{mpp,stc}}$), temperature coefficient of open-circuit voltage (K_v), and temperature coefficient of short-circuit current (K_i). Moreover, the $i - v$ and $P - v$ characteristics of PV cell from datasheet are also obtained under STC. The information in this part will be used to determine the variables and parameters in equivalent electrical circuit part.

Equivalent electrical circuit

A PV model is used to convert the energy from sunlight into the electricity. Typically, the PV model is modeled by an equivalent electrical circuit which consists of source, diode, and resistance. According to [26], the source stands for photocurrent and depends on

solar irradiance. The diode represents the dark current losses and the effect of generation-recombination on the space charge region. The resistance is used for modeling the losses of the PV cell such as heat. To convert the solar power into the electrical power, the equivalent electrical circuit with single and two diode are usually used as shown in Figures 3.4 and 3.5 respectively. From the equivalent electrical circuit, the generated current from the PV cell can be calculated by Kirchhoff's law and the generated current of two- and single-diode models can be expressed as (3.28) and (3.29) respectively [26]. The generated current of the two-diode model can be calculated by

$$i = i_{\text{ph}} - i_{01}e^{\frac{v+iR_s}{V_{t1}}-1} - i_{02}e^{\frac{v+iR_s}{V_{t2}}-1} - \frac{v+iR_s}{R_p}, \tag{3.28}$$

where

- i is the terminal current of the PV cell,
- i_{ph} is the photocurrent,
- i_{01} and i_{02} are the diode reversed saturation current of the diode D1 and D2 respectively,
- v is the terminal voltage of the PV cell,
- R_s is the series resistance and R_p is the parallel resistance,
- $V_{t1} = A_1 k_b T_c/q$ and $V_{t2} = A_2 k_b T_c/q$ are the junction thermal voltage of the diode D1 and D2 respectively,
- A_1 and A_2 are the ideality factor of the diode D1 and D2 respectively,
- $k_b = 1.38 \times 10^{-23}$ Joule/Kelvin is the Boltzmann constant,
- $q = 1.6 \times 10^{-19}$ Coulomb is the elementary charge,
- T_c is the cell temperature.

To simplify the model, one can neglect a diode in the two-diode model and become the single-diode model. The generated current of single-diode model can be calculated by

$$i = i_{\text{ph}} - i_0 e^{\frac{v+iR_s}{V_t}-1} - \frac{v+iR_s}{R_p}. \tag{3.29}$$

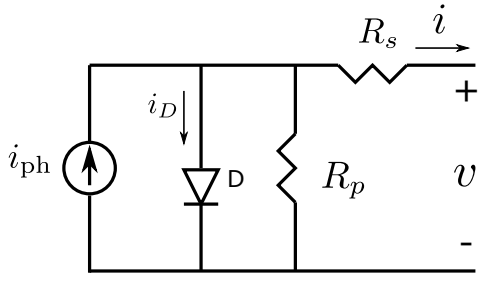


Figure 3.4: The single-diode model equivalent circuit of a PV cell.

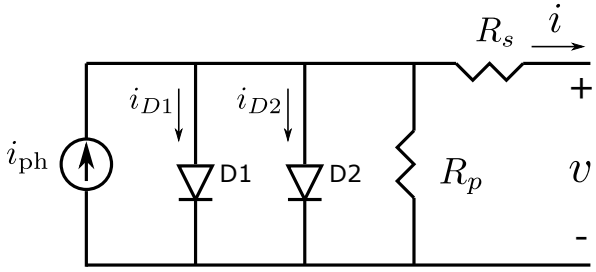


Figure 3.5: The two-diode model equivalent circuit of a PV cell.

For simplicity, only the single-diode model is explained in this section. According to (3.29), the general equations to describe current and voltage at the short-circuit, open-circuit, and maximum power point condition are expressed as follows.

- The equation for short-circuit condition ($v = 0$): According to the single-diode model and (3.29), the short-circuit current is calculated by Kirchhof's law and can be expressed as
$$i_{\text{sc}} = i_{\text{ph}} - i_0 e^{\frac{i_{\text{sc}} R_s}{V_t} - 1} - \frac{i_{\text{sc}} R_s}{R_p}, \tag{3.30}$$
where i_{sc} is short-circuit current.

- The equation for open-circuit condition ($i = 0$): According to the single-diode model and (3.29), the open-circuit voltage is calculated by Kirchhof's law and can be expressed as
$$0 = i_{\text{ph}} - i_0 e^{\frac{v_{\text{oc}}}{V_t} - 1} - \frac{v_{\text{oc}}}{R_p}, \tag{3.31}$$
where v_{oc} is open-circuit voltage.

- The equation for maximum power point condition ($i = i_{\text{mpp}}$ and $v = v_{\text{mpp}}$): According to (3.29), at the maximum power point condition, the relationships between maximum current and maximum voltage of the single-diode model can be expressed as
$$i_{\text{mpp}} = i_{\text{ph}} - i_0 e^{\frac{v_{\text{mpp}} + i_{\text{mpp}} R_s}{V_t} - 1} - \frac{v_{\text{mpp}} + i_{\text{mpp}} R_s}{R_p}, \tag{3.32}$$
where i_{mpp}, and v_{mpp} are the current and voltage at the maximum power point respectively.

However, the PV model still requires variables and parameters that are not given in the manufacturer's datasheet and need to be determined. From (3.29), the unknown variables are i_{ph} and i_0. The unknown parameters are R_s, R_p, and A. These unknown variables and parameters are described as follows.

- The photocurrent (i_{ph}): Typically, i_{ph} highly depends on solar irradiance. The general equation of i_{ph} which derived from physics is non-linear in I and T_c. There are also many proposed equations to calculate i_{ph} and the example equations are described as follows.

– According to [26–32], i_{ph} depends on I and T_c and can be expressed as

$$i_{\text{ph}} = (i_{\text{ph,stc}} + K_i(T_c - T_{c,\text{stc}}))\frac{I}{I_{\text{stc}}}. \tag{3.33}$$

– According to [33, 34], i_{ph} can be calculated under the open-circuit condition. According to (3.31), i_{ph} can be expressed as

$$i_{\text{ph}} = i_0 e^{\frac{v_{\text{oc}}}{n_s V_t}} + \frac{v_{\text{oc}}}{R_p}, \tag{3.34}$$

where n_s is the number of PV cells in series.

– According to [35], i_{ph} can be calculated under the short-circuit condition (3.30) with the assumption that $i_0 \equiv 0$. Therefore, i_{ph} can be expressed as

$$i_{\text{ph}} \approx \frac{R_p + R_s}{R_p} i_{\text{sc}}. \tag{3.35}$$

- The diode reversed saturation current (i_0): i_0 is the diode reversed saturation current which is the small amount of current flowing in the diode when the reverse bias is applied on a diode. In addition, the current flowing depends on the depletion region of the diode and the depletion region also varies according to temperature. Therefore, i_0 highly depends on T_c. There are many proposed equations which derived from physics to calculate i_0 and some examples equations are described as follows.

– According to [28, 32, 36], i_0 at any operating point can be expressed as

$$\frac{i_0}{i_{0,\text{stc}}} = \left(\frac{T_c}{T_{c,\text{stc}}}\right)^3 \exp\left(\frac{E_g}{A_{\text{stc}}}\left(1 - \frac{T_{c,\text{stc}}}{T_c}\right)\right), \tag{3.36}$$

where E_g is the band gap energy of the semiconductor, $i_{0,\text{stc}}$ and A_{stc} are derived from physics and can be expressed as follows [28].

$$i_{0,\text{stc}} = \left(i_{\text{sc,stc}} - \frac{v_{\text{oc,stc}}}{R_{p,\text{stc}}}\right) \exp\frac{-v_{\text{oc,stc}}}{A_{\text{stc}}}, \tag{3.37}$$

$$A_{\text{stc}} = \frac{v_{\text{mpp,stc}} + i_{\text{mpp,stc}} R_{s,\text{stc}} - v_{\text{oc,stc}}}{\log(i_{\text{sc,stc}} - \frac{v_{\text{mpp,stc}}}{R_{p,\text{stc}}} - i_{\text{mpp,stc}}) - \log(i_{\text{sc,stc}} - \frac{v_{\text{oc,stc}}}{R_{p,\text{stc}}}) + \left(\frac{i_{\text{mpp,stc}}}{i_{\text{sc,stc}} - \frac{v_{\text{oc,stc}}}{R_{p,\text{stc}}}}\right)} \tag{3.38}$$

– According to [27], this work aims to improve the equation to get better at explain i_0 in the open-circuit voltage condition based on the equation of $i_{0,\text{stc}}$. This work uses $i_{0,\text{stc}} = i_{\text{sc,stc}}/(\exp(v_{\text{oc,n}}/AV_{\text{t,stc}}) - 1)$. The improved equation of i_0 is obtained from the equation of $i_{0,\text{stc}}$ by including in the equation the current (K_i) and voltage coefficients (K_v) and can be express as

$$i_0 = \frac{i_{\text{sc}} + K_i(T_c - T_{c,\text{stc}})}{\exp((v_{\text{oc}} + K_v(T_c - T_{c,\text{stc}}))/AV_t) - 1}. \tag{3.39}$$

– According to [26], the equation of i_0 is inferred from [37] and can be expressed as

$$i_0 = C_0 T_c^3 \exp\left(-\frac{E_g}{k_b T_c}\right) \tag{3.40}$$

where C_0 is a constant depending on material parameters and can be expressed by

$$C_0 = \frac{i_{0,\text{stc}}}{T_{c,\text{stc}}^3 \exp(-E_g/k_b T_{c,\text{stc}})}. \tag{3.41}$$

For unknown parameters, there are many studies that explore the way to determine the unknown parameters (R_s and R_p). To determine the unknown parameters, it can be classified into two approaches, i) derived the equation from the datasheet and ii) obtained from experimental results.

1. Determination the parameters from derived equations: In [26], this work uses (3.29) under STC and assumes that $R_p \to \infty$ in order to obtain R_s. A is also obtained by solving (3.29) and assumes $i = 0$ under STC, $R_p \to \infty$, and $v_{\mathrm{oc,stc}} = (Ak_bT_c/q)\log(i_{\mathrm{ph,stc}}/i_{0,\mathrm{stc}})$, then A is determined by $\frac{dv_{\mathrm{oc}}}{dT_c}|_{T_c=T_{c,\mathrm{stc}}} \equiv K_v$. After that, A has been used to determine R_p by (3.30). According to [36], this study investigates the effect on $i-v$ curve when varying the parameters R_s and R_p. The results show that R_s impacts the shape of $i-v$ curve near the maximum power point (smaller R_s provides higher output power) and R_s impacts the slope of $i-v$ curve near the short circuit conditions (largeer R_p provides higher output power). From the impact of R_s and R_p, the slope of $i-v$ curve is usually used to determine R_s and R_p in many studies as follows. In [28, 38], R_p can be estimated from (3.29) under the conditions of $v \to 0$, $i_0 \ll i_{\mathrm{ph}}$and $R_s \to R_p$. Then, the $i-v$ curve is assumed as linear and it can be expressed as

$$i \approx i_{\mathrm{ph}} - \frac{v}{R_p}. \tag{3.42}$$

From (3.42), the derivative of i respect to v provides:

$$\left.\frac{di}{dv}\right|_{i=i_{\mathrm{sc,stc}}} = -\frac{1}{R_p}. \tag{3.43}$$

Thus, R_p can be estimated from the slope of $i-v$ curve near short-circuit condition. R_s is obtained from (3.29) under the conditions of $i \to 0$ and $v \to v_{\mathrm{oc}}$ based on Taylor series expansion and gives $i = i_{\mathrm{ph}}\frac{V_t}{i_0R_s} - \frac{v}{R_s}$. Then, they take the derivative of i respect to v yield:

$$\left.\frac{di}{dv}\right|_{v=v_{\mathrm{oc,stc}}} = -\frac{1}{R_s}. \tag{3.44}$$

From [32], R_p is obtained from (3.43) by take into account the i_0 term, therefore, the relationships between R_p and R_s can be expressed as

$$R_p = \frac{R_s}{i_0\exp(i_{\mathrm{sc}}R_s/A)}. \tag{3.45}$$

Then, R_s and R_p can be obtained by a numerical method. In [31], R_s and R_p can be obtained by $\frac{dP}{dv}|_{\mathrm{mpp,stc}} = 0$ and (3.43) respectively, where

$$\left.\frac{dP}{dv}\right|_{\mathrm{mpp,stc}} = \left.\frac{d(iv)}{dv}\right|_{\mathrm{mpp,stc}} = i_{\mathrm{mpp}} + v_{\mathrm{mpp}}\left.\frac{di}{dv}\right|_{\mathrm{mpp,stc}} = 0. \tag{3.46}$$

According to [35], R_s can be obtained from (3.46). A non-linear equation of R_p can be derived from (3.32) and it is non-linear in R_s. Using (3.35), (3.32), (3.46) and an optimization method in order to obtained the optimal R_s, R_p, and i_{ph} that provide the lowest sum of square error. In [34], the parameters i_{ph}, i_0, R_s, and R_p can be estimated

from (3.46), (3.30) ,(3.31), and (3.32) respectively by Newton method. The parameter A of this work is adjusted in order to reduce the error between $i_{\mathrm{sc,stc}}$ from datasheet and i_{sc} from (3.30). Typically, the parameter A is a constant which vary around 1 to 1.5.

2. Determination the parameters from experimental results: According to [30], this work uses the two-diode model to determine A_1 and A_2 (ideality factor of diode 1 and diode 2) with neglected R_s and R_p. This study uses (3.39) and (3.28) under open-circuit condition to derive the equation to determine A_2 in terms of A_1. Then, the pair of A_1 and A_2 are used to find P_{max}. After that, the error between i_{max} and $i_{\mathrm{max,stc}}$ is calculated. The optimal A_1 and A_2 should provide the lowest error between i_{max} and $i_{\mathrm{max,stc}}$. According to [27], (3.33) and (3.39) are substituted in (3.32) to find i_{mpp}. Then, the maximum power can be obtained by $P_{\mathrm{max}} = i_{\mathrm{mpp}} v_{\mathrm{mpp}}$. The relationships between R_s and R_p can be derived from equation of P_{max}. After that, this study finds P_{max} by varying R_s. The optimal R_s and R_p should provide $P_{\mathrm{max}} = P_{\mathrm{max,stc}}$.

After the five variables and parameters are obtained, one can find the generated power of the PV cell at any time. However, the power output of PV cell is direct current while almost electrical appliances are using alternating current. Thus, direct current from PV cell must be converted into alternating current by inverter. The inverter usually has the efficiency about 90-95% in order to convert direct current into alternating current. To calculate the power output of PV cell by equivalent electrical circuit method, it's always involved the non-linear equations. Thus, the method of machine learning which usually deal with non-linear can also be used to determine the unknown variables and parameters and we will explain in next section.

Machine learning

Machine learning methods such as ANN and genetic algorithm (GA) can also be used to determine the unknown variables and unknown parameters of PV model, moreover, these methods can also be used to convert the relevant variables to solar power into the power output of PV cell. According to [39], this study proposes a model to explain i of PV cell and uses GA to determine the parameters of the proposed model. The summary of the studies that investigate the PV model by neural network are summarized in Table 3.3.

Note that, the time parameter (t) in [40] is the value between -0.5 to 0.5 corresponding to 5.00 to 20.00 hrs.

From the literature reviews about PV model, we can conclude that solar irradiance and temperature are the most important variables to estimate the solar power of PV cells. Therefore, we use these two variables for modeling the PV conversion model which the details are described in Chapter 6.

Table 3.3: Studies of PV model using machine learning methods.

Reference	**Tool**	**Hidden layer /nodes**	**Input**	**Target**	**Performance**
[40]	ANN	1/6	$I(t)$, T(t), WS(t), and Time parameter (t)	$P_{\max}(t)$	MAPE of $P_{\max}(t) < 5\%$
[41]	ANN	2/6, 12	$I(t)$, $I(t-1)$, $I(t-2)$, $I(t-3)$, $I(t-4)$, and T(t)	$v(t)$ and $i(t)$	MAE of $v(t) <$ 0.35 Volt and MAE of $i(t) <$ 0.063 Amp
[42]	ANN	1/20	$I(t)$ and T(t)	$i_{\text{ph}}(t)$, $i_0(t)$, $R_s(t)$, $R_p(t)$, and $A(t)$	MAPE of $P_{\max}(t) <$ 2.1%
[43]	RNN	1/4	$I(t)$, T(t), and WS(t)	$P(t)$	RMSE of $P(t) <$ 0.3 kW

CHAPTER IV

SOLAR FORECASTING TECHNIQUES

According to [5, 44–46], solar forecasting methods can be classified into three common methods which are statistical methods, physical methods, and hybrid methods. Statistical methods aim to find a model that explain relationship between input and forecast values and to estimate the model parameter from statistical concept. Examples of statistical methods are time series and machine learning. Physical methods use the equations which are derived from physics to explain the dynamic of weather variables such as Numerical Weather Prediction (NWP). The hybrid methods are the combination two or more existing methods in order to improve the forecasting accuracy. These forecasting methods are categorized as shown in Figure 4.1 and the detail of each method is described as follows.

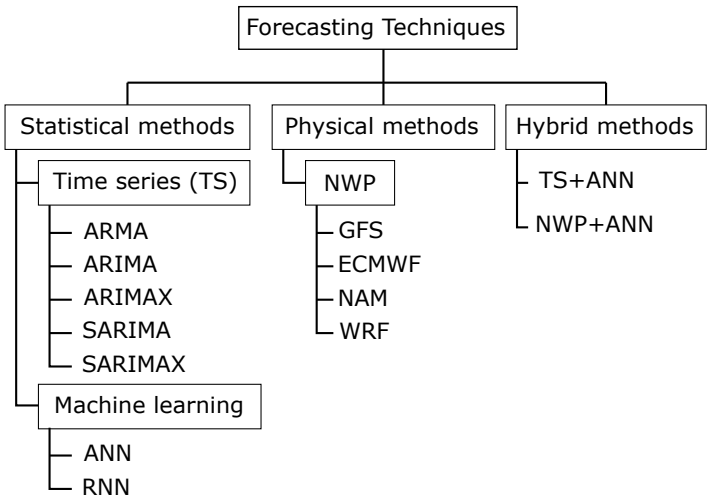


Figure 4.1: Solar irradiance forecasting techniques.

4.1 Statistical methods

This methods use the historical data of solar irradiance and relevant variables to solar irradiance to train the forecasting model. The common methods are Time series models and Machine learning. These methods are described as follows.

Time series models

Time series models are used to describe the characteristics of solar irradiance from historical solar irradiance data. The solar irradiance data is always non-stationary and has periodic cycle. Thus, time series models which usually use for solar irradiance forecasting is Seasonal Auto-Regressive Integrated Moving Average with exogenous term (SARIMAX)

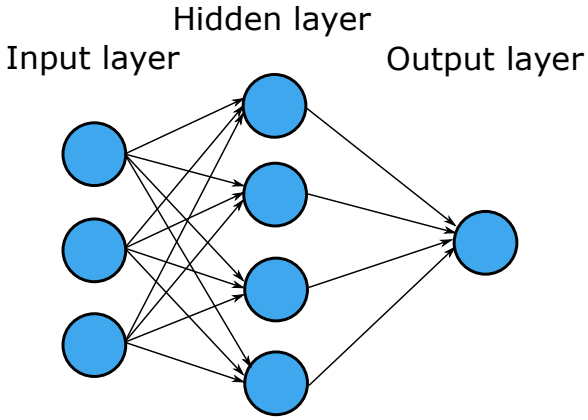
models. SARIMAX models are used to deal with non-stationary data in time series due to seasonal, trends, drift, etc. To solve the problem, a seasonal integrated and integrated term have been used to transform the process until stationary before apply SARMAX models. The exogenous term includes the relevant variables to solar irradiance for explaining the characteristics of solar irradiance. In [47], SARIMAX $(p,d,q)(P,D,Q)_T$ model is defined as

$$\tilde{A}(L^T)A(L)(1-L^T)^D(1-L)^d y(t) = B(L)u(t) + \tilde{C}(L^T)C(L)v(t) \tag{4.1}$$

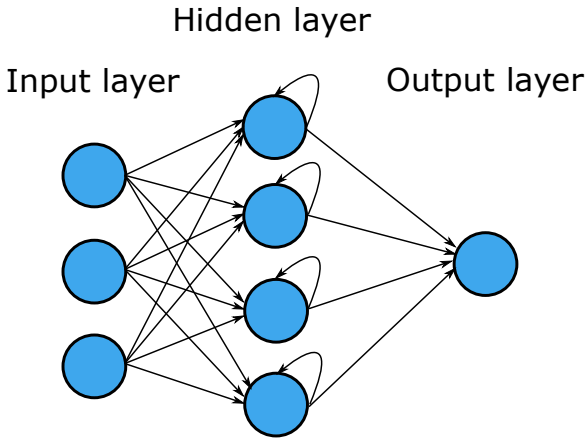
where L is lag operator, $y(t)$ is the solar irradiance, T is seasonal period, $v(t)$ is white noise, $(1-L)^d$ and $(1-L^T)^D$ are integrated term using d order and seasonal integrated term using D order respectively, $A(L)$ and $C(L)$ are Auto-Regressive and Moving Average polynomials: $A(L) = I - (a_1L^1 + \cdots + a_pL^p), C(L) = I + c_1L^1 + \cdots + c_qL^q$, $\tilde{A}(L)$ and $\tilde{C}(L)$ are seasonal Auto-Regressive and seasonal Moving Average polynomials respectively, $B(L) = B_1L^1 + B_2L^2 + \cdots + B_mL^m$, and $u(t)$ is relevant variable(s) of solar irradiance such as relative humidity, temperature, wind speed, air pressure. Then, SARIMAX (p,d,q) $(P,D,Q)_T$ is used to described the characteristic of solar irradiance in the future based on the previous observed or forecasted values. There are some works that use time series method to predict solar irradiance. In [48], the SARIMAX model was used to predict the solar irradiance 5, 15, and 30 minutes in advance and SARIMAX model outperforms the other time series models such as neural networks and regression models.

Machine learning

Machine learning also use the historical solar irradiance and the relevant variables to solar irradiance data to train the forecasting model. The common methods for solar irradiance forecasting are support vector machine (SVM), support vector regression (SVR), and neural networks. However, the neural networks method is the most widely used for solar forecasting. The simplest model of neural network is the Artificial Neural Networks (ANN) that uses the historical data and optimization algorithm to train and adjust the weights and biases of the model. ANN consists of input layer, hidden layer, and output layer as shown in Figure 4.2a. The inputs from input layer are fed into the hidden layer with the multiplication of weights. The hidden layer can be used to learn complex characteristics of the variables of interest by optimization algorithms such as batch gradient descent, stochastic gradient descent, and mini-batch gradient descent. The output layer is the sum of product of the outputs from hidden layer and weights. The error between output and the target is fed back to the network in order to adjust weights in all layers in training process. Moreover, there are some other neural networks such as Radial Basis Function Neural Networks (RBFNN) and Recurrent Neural Networks (RNN). RBFNN has the same structure as ANN, however, the activation function in hidden layer is Gaussian function. RNN also has three layers as same as ANN but the outputs from hidden layer are fed back to hidden layer for taking into account the sequence of the outputs. RNN structure is shown in Figure 4.2b.



(a) Diagram of a basic structure of ANN.



(b) Diagram of a basic structure of RNN.

Figure 4.2: Diagram of a structure of neural networks.

According to [49], solar irradiance forecasting up to six hours from ANN and time series were investigated. The results showed that the performance of ANN outperformed time series models, especially for forecasting horizon greater than one hour. The eight algorithms of ANN were investigated in [50] in order to find the best algorithm for solar forecasting in India. Moreover, many studies of solar forecasting by neural networks methods were reported in [51] and all results showed that neural networks methods outperformed the other methods such as time series methods and persistence forecast. The machine learning for solar forecasting and other renewable energy are summarized in Table 4.1.

Table 4.1: Summary of renewable energy forecasting by machine learning.

Reference	**Tool**	**Hidden layer/ node**	**Input**	**Target**
[18]	ANN	1/3	I	$\hat{I}$
[52]	ANN	1/10	I	$\hat{I}$
		1/1460	I	$\hat{I}$
[53]	ANN	2/11,17	mean daily I, mean daily T, and day of the month	$\hat{I}$
[54]	ANN	1/3	I and T	$\hat{P}$
[55]	ANN	1/3	I_{clr}, PR, T, RH, WS, WD, cloudy height class, peak wind speed, sunshine duration, and rain precipitation	$\hat{I}_{\text{clr}}$
[56]	ANN	1/8	I and T	$\hat{P}$
	PHANN, Wavelet+ANN, and ANFIS	1/12	I, T, and I_{clr}	$\hat{P}$
[57]	Ensembles of ANN	1	P	$\hat{P}$
[58]	Hierarchical ANN and SVR	-	P and predicted weather variables from weather forecast: T, WS, and WD	$\hat{P}$
[59]	ANN with ELM	1/100	P and T	$\hat{P}$
[60]	Wavelet+ANN	-	Load, Weekday index, Wind-chill, T, RH, WS, Cloud cover, and Precipitation	Predicted load
[61]	ANN with ELM	1/50	mean daily I, T, T_c, WS, and P	$\hat{P}$
	SVR	-		
	ANN	1/35		
[20]	K-means+ANN	-	I	$\hat{I}$
[62]	EMD+ANN	1	WS	$\widehat{\text{WS}}$

[63]	Wavelet+SVM	-	I, zenith angle, tcc, opaque cloud, dry bulb temp., dew point temp., RH, PR, WS, WD, visibility, ceiling height, precipitation, aerosol optical depth, and albedo	$\hat{I}$
	Wavelet+ANN	1/10		
[20]	Transformation based K-means+ANN	-	I	$\hat{I}$
[64]	Wavelet+Diagonal RNN	-	I, hour, cloud cover, and day of year	$\hat{I}$
[65]	Wavelet+GTSOM+ BNN	1/5	I, T , WS, and WD	$\hat{I}$
[66]	PHANN	2/9,7	physical data: I_{clr}, sunshine duration, day, and hour and predicted weather variables from weather forecast: RH, T, WS, PR, and tcc	$\hat{P}$
[67]	PHANN	1	I_{clr} and predicted weather variables from weather forecast: T and RH	$\hat{P}$
[68]	PHANN	2/11,5	predicted weather variables from weather service: I, I on plane of array, I_{clr}, T, WS, WD, PR, tcc, cloud type, and precipitation	$\hat{P}$
[22]	RNN	1/5	P	$\hat{P}$
			P and I	
			P, I, and T	
[69]	Wavelet+DCNN	-	P	$\hat{P}$
[70]	SVM	-	Sunshine duration	$\hat{I}$
[71]	Wavelet+SVM	-	I, max/min/average T, RH, water vapour, and sunshine duration	$\hat{I}$
[72]	SVM based on weather classification	-	P and predicted max/min/mean T from weather report	$\hat{P}$
[73]	SVR	-	I, hours T, and RH	$\hat{P}$

[74]	SOM+LVQ+ Fuzzy+SVR	-	I and predicted weather variables from weather forecast: T and probability of precipitation	$\hat{P}$
[75]	SVR+KF	-	Load	Predicted load

Other statistical approaches

Other statistical approaches for solar forecasting and other renewable energy are summarized in Table 4.2.

Table 4.2: Summary of renewable energy forecasting by other statistical approaches.

Reference	**Tool**	**Input**	**Output**
[76]	Volterra filter	I	$\hat{I}$
[77]	Dynamic Harmonic Regression	I	$\hat{I}$
[78]	Partial Functional Linear Regression models	PR, max/min/mean T, dew temperature, RH, insulation, WS, and precipitation	$\hat{P}$
[79]	Multi-linear Adaptive Regression Splines	max/min/mean T, dew temperature, RH, isolation duration, PR, WS, and precipitation	$\hat{P}$
[80]	Hammerstein model	WS	$\widehat{\text{WS}}$

4.2 Physical methods

Physical methods use equations derived from physics to predict the weather variables. The most popular method for predict the weather variables is Numerical Weather Prediction (NWP) models. NWP models [6] use non-linear partial differential equations which derived from physics to model the dynamics of the Earth's atmosphere such as air pressure, temperature, relative humidity, wind speed, and solar irradiance. Then, the predicted variables are obtained by solve the equations. The resolution of NWP forecasting is 1-28 km^2 which is too wide to predict cloud motion accurately [81]. Therefore, the low accuracy of predicted

cloud motion cause the low accuracy of predicted solar irradiance. However, NWP models can yield better solar irradiance prediction with long time horizon (6 hours up to several days) than other methods. In order to run a prediction, NWP models need an initialization data which is the collected data from ground stations and satellites around the world. NWP can be classified into two main categories.

1. Global NWP models are used to forecast meteorological variables around the world. The initialization data of global NWP models are collected from ground stations, radar, radiosonde, and satellite.

   (a) Global Forecast System (GFS): GFS model runs by the National Oceanic and Atmospheric Administration (NOAA) four times daily with a resolution of 28×28 km^2. The time step of GFS is 3-hourly and forecast up 10 days ahead and then the time step has changed to 12-hourly and GFS continues to forecast up to 16 days in advance.

   (b) European Centre for Medium-Range Weather Forecasts (ECMWF): ECMWF model is supported by most of the nations of Europe. The ECMWF runs twice a day. The time step of ECMWF is 3-hourly and forecast up 10 days in advance.

2. Regional NWP models are used to forecast meteorological variables only a sub-domain of the global space. The initialization of regional NWP models are the data from the output of global NWP models.

   (a) North American Mesoscale (NAM): NAM model runs by the National Centers for Environmental Prediction (NCEP) four times daily with a resolution of 12×12 km^2. NAM is used to forecast the meteorological variables over the area of North America with time step of 3-hourly and forecast up to 84 hours in advance.

   (b) Weather Research and Forecasting (WRF): WRF model is created by NOAA and the National Center for Atmospheric Research (NCAR). WRF model is flexible to users for design a forecasting area and resolution. WRF software is a free ware supported by many developers, so there are many schemes of predictions to use.

There are many studies that compare the performance of NWP models for solar irradiance forecasting [7, 82, 83]. The results in those studies show that the global NWP models outperform the regional NWP models because the global NWP models provide an averaged forecast values over a large area while regional NWP models provide the forecasted values in more specific area which have high fluctuation. However, the regional NWP models require less resources and also computing time than global NWP models. Thus, we use WRF model which is an open source software to predict solar irradiance for this work. WRF model requires initialization data, forecasting domain, microphysics scheme, cumulus parameterization scheme, surface physics scheme, planetary boundary layer physics scheme, longwave radiation scheme, and shortwave radiation scheme before solve the non-linear partial differential equations in order to forecast the meteorological variables. The forecasting error of NWP outputs can be reduced by spatial averaging method which is described in Chapter 5. We summarize the works that benchmark the NWP models for predicting solar irradiance one day ahead in Table 4.3.

Table 4.3: Summary of the performance of NWP models for predicting solar irradiance one day in advance.

Reference	**Country**	**Model**	**Resolution** (km^2)	**Time step** (minutes)	**RMSE** (W/m^2)	**NRMSE** (%)
[11]	USA	GEM	15 × 15	7.5	16.7 to 43.6	
[10]	USA	ECMWF	25 × 25	180	96.4 to 157.6	
		GFS	50 × 50	180	83.2 to 129.4	
		NAM	11 × 11	60	93.5 to 149	
[7]	USA	GEM	15 × 15	7.5		33
		ECMWF	25 × 25	180		33
		WRF	5 × 5	60		44
		MASS	5 × 5	60		55
		ARPS	5 × 5	60		56
		NDFD	5 × 5	180		40
[83]	Germany	ECMWF	16 × 16	180		40
		DWD	7.5 × 7.5	60		43
		DMI	3 × 3	60		45
[84]	French	WRF	3 × 3	60	189.81	35.2
[85]	Brazil	WRF	5 × 5	60		20.5 to 58.9
[86]	Italy	ECMWF	13.5 × 13.5	180	114.3 to 123.2	29 to 33
[86]	Italy	ECMWF	13.5 × 13.5	180	114.3 to 123.2	29 to 33

4.3 Hybrid methods

This section describes about the combination of two or more existing methods for solar irradiance forecasting. From the literature, there are many combinations of the forecasting models which can be described as follows.

1. Extension on adaptive parameters: In these methods, one has a forecasting model containing parameters that are initially trained in an off-line manner so that the model

fits well with historical data. The idea is to apply an adaptive rule to adjust model parameters when the new data sample arrives in an on-line manner. Examples of adaptive scheme are Kalman filter, [75] used the Extended Kalman filter (EKF) to update the parameter in Radial Basis Function Neural Network (RBFNN).

2. Extension on adding non-linear models: Conventional techniques based on linear models are time series approach. More examples, though non-linear, such as deterministic clear-sky models, they are too simple to capture the real characteristics of solar irradiance. Therefore, including more complex nonlinearity in the model feature is a research direction here. Including nonlinearity can be done in many ways and we present three main schemes:

   - Cascade: The output of model 1 will be the input of the (typically non-linear) model 2. Examples of non-linear models are neural network and its variants.

     $$y = G_1 G_{\text{nonlinear}} u,$$

     where G_1 represents system 1 and $G_{\text{nonlinear}}$ represents the non-linear model in consideration.

   - Bias correction: The output of model 1 is compared to the measurement and results in the residual error (e). This error is subsequently fed as an input to the model 2 with an assumption that the residual error still contain unexplained non-linear dynamic of the model. As a result, the output of model 2 is linearly added with the output of model 1.

     $$y = \hat{y}_1 + e, \quad \hat{y}_2 = G_{\text{nonlinear}} e, \quad y = \hat{y}_1 + \hat{y}_2$$

   - Variants: There are a certain number of research work who proposed different configurations of how two methods are combined. These methods typically consider residual errors from the model 1, similarly to the bias correction approach. However, the combination with the model 1 is distinguished from one work to another. We will explain in detail here.

Later this section explain examples of non-linear models ($G_{\text{nonlinear}}$) and G_1 (the first model), input u, and output y from the literature.

Adaptive approach. According to [75], the RBFNN with dual EKF is used for load forecasting. The parameters of RBFNN are updated by dual EKF iteratively. However, this work uses only the historical load data where the forecasting model can be improved by adding the relevant variables to the model. For example, [87] uses the RBFNN for on-line load forecasting 24 hours ahead. The unit centers, widths and weights of RBFNN are updated iteratively by K-mean clustering algorithm, nearest-neighbor, and lest-squares technique respectively. The input of RBFNN are day of the month, $P(t)$, $\hat{I}_{\text{nwp}}(t+1)$, $\widehat{\text{WS}}_{\text{nwp}}(t+1)$, $\widehat{\text{T}}_{\text{nwp}}(t+1)$, and $\widehat{\text{RH}}_{\text{nwp}}(t+1)$. The output are $P(t+1, \ldots, t+24)$. The adaptive approach takes a short time in both predict the variables and update the model parameters which is usually used in an on-line forecasting.

Cascade models. The cascade models are updated the model parameters in an off-line manner. The diagram of the cascade models is shown in Figure 4.3.

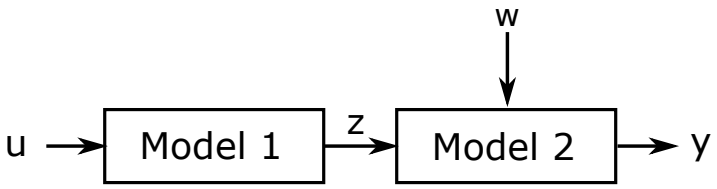


Figure 4.3: Diagram of two cascade models.

There are many studies that use a hybrid forecasting method having two cascade models for many purposes and are summarized in Table 4.4.

Table 4.4: Summary of a hybrid forecasting method having two cascade models.

Ref.	u	z	w	G_1	$G_{\text{nonlinear}}$	y	RMSE (W/m^2)	NRMSE (%)
[88]	WRF initialization data	mean daily I_{nwp}	day of year, T, tcc, and mean daily I on the plane of PV cell	NWP	ANN	$\hat{I}$ on the plane of PV cell	145 - 149	31 - 31.9
		I_{nwp}	day of year					
		I_{nwp}	day of year, T, tcc, and mean daily I on the plane of PV cell					
[86]	WRF initialization data	I_{nwp}	I, T, K, and day of year	NWP	ANN	$\hat{I}$	96.4 - 112.6	24 - 29
[85]	WRF initialization data	RH_{nwp}, T_{nwp}, WS_{nwp}, and tcc_{nwp}	I_{clr}	NWP	ANN	$\hat{I}$		15 - 51.5
[89]	WRF initialization data	I_{nwp}	P, AVG value of P over day at the same time, and I_{clr}	NWP	ANN	$\hat{P}$		11.26 - 11.4
[90]	WRF initialization data	T_{nwp} and predicted sky conditions	I, T, and sky conditions	NWP	Fuzzy+ANN	$\hat{I}$		
[91]	WRF initialization data	I_{nwp}, T_{nwp}, tcc_{nwp}, solar azimuth, and solar elevation	time	NWP	ANN, AnEn, ANN+AnEn	$\hat{P}$		8.09 - 8.66

[92]	WRF initialization data	I_{nwp}, RH_{nwp}, T_{nwp}, WS_{nwp}, tcc_{nwp}, and PR_{nwp}	day	NWP	ANN with Genetic Swarm Optimization and Back Propagation algorithms	$\hat{P}$		
[93]	WRF initialization data	I_{nwp}	P	NWP	ARIMA, SVM, ANN, ANFIS with GA algorithm	$\hat{P}$		3.4 - 6.6
[94]	predicted amount of cloud and relative humidity from weather report	$\hat{I}$	-	Fuzzy	RNN	$\hat{P}$		
[87]	WRF initialization data	I_{nwp}, RH_{nwp}, T_{nwp}, and WS_{nwp}	hour and P	NWP	RBFNN	$\hat{P}$		
[95]	WRF initialization data	I_{nwp}, T_{nwp}, WS_{nwp}, PR_{nwp}, and tcc_{nwp}	$\cos\theta$	NWP	Multi-linear Adaptive Regression Splines	$\hat{P}$		
[96]	WRF initialization data	Predicted surface sensible heat flux, surface latent heat flux, surface downward shortwave radiation, surface downward longwave radiation, top outgoing shortwave radiation, top outgoing longwave radiation, and T	day of year and hour	NWP	RNN	$\hat{P}$		

Bias correction approach. The bias correction approach also updates the parameters in an off-line manner. The diagram of bias correction approach is shown in Figure 4.4.

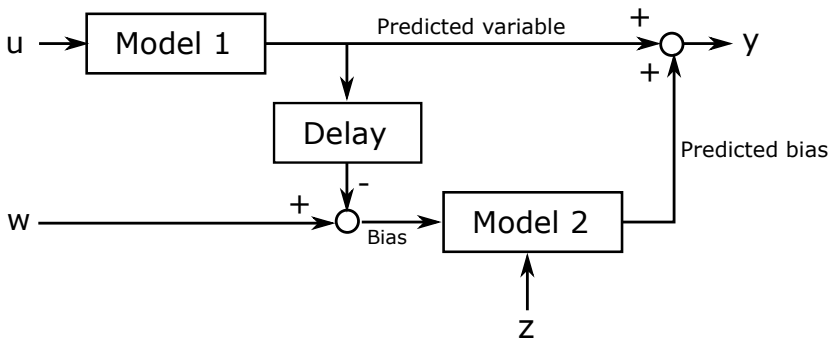


Figure 4.4: Diagram of the bias correction approaches.

There are many studies that use the bias correction approach in forecasting applications and are summarized in Table 4.5.

Other hybrid approaches. There are some studies that combine the model 1 and model 2 in different ways. According to [1], this work uses the combination of ARIMA model and ANN model. The inputs of ANN model are the residual error and the differencing terms of ARIMA model. The diagram of this work is shown in Figure 4.5.

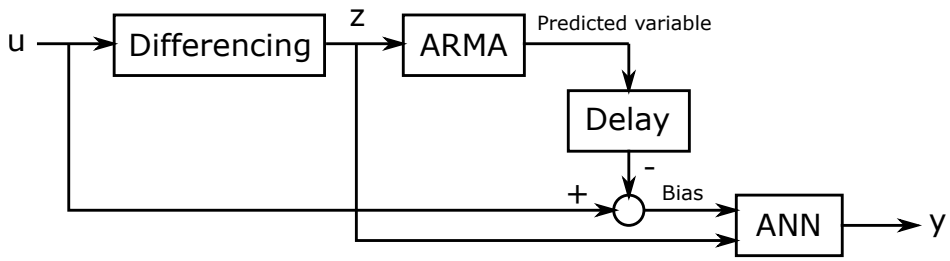


Figure 4.5: Diagram of the hybrid model of [1].

In [105], the forecasting model is similar to [1] but this work also includes the predicted value from ARIMA model as the input of ANN model. Moreover, there is the method that uses the model 1 and model 2 in parallel to forecast the solar irradiance and the diagram of the parallel model is shown in Figure 4.6.

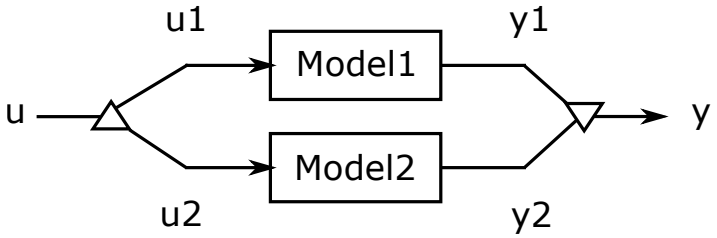


Figure 4.6: Diagram of the parallel models.

According to [106], this work also uses parallel model and requires only the measured

solar irradiance as the input of the solar forecasting model. The model 1 is an ANN model which is used to forecast the solar irradiance around noon. The model 2 is the average measured solar irradiance over day which is used to forecast the solar irradiance both in the morning and in the evening. However, the model 2 of this parallel model provides high forecasting error because the solar irradiance is not the same as the previous day. According to [107], the parallel model is used to forecast the solar irradiance. The model 1 and model 2 are the ANN models. The inputs of model 1 are the measured solar irradiance and the measured relevant variables to solar irradiance to forecast the solar irradiance 1 to 3 hours ahead. The input of model 2 is only the measured solar irradiance to forecast the solar irradiance 4 to 6 hours ahead.

Table 4.5: Summary of the hybrid methods using bias correction for solar irradiance and PV power forecasting.

Reference	u	w	z	G_1	$G_{\text{nonlinear}}$	y
[84]	WRF initialization data	I	$\cos\theta$ and k	NWP	ANN	$\hat{I}$
[97]	WRF initialization data	I	-	NWP	Wavelet+ RNN	$\hat{I}$
[98]	I	I	-	ARMA	ANN	$\hat{I}$
[99]	P	P	-	ARIMA	SVM	$\hat{P}$
[100]	WS	WS	-	ARIMA	ANN	$\widehat{\text{WS}}$
	WS, T, PR, Precipitation	WS	-	ARIMAX	ANN	$\widehat{\text{WS}}$
	WS	WS	-	Holt-Winters	ANN	$\widehat{\text{WS}}$
[101]	Sunspot data and Money exchange rate data	Sunspot data and Money exchange rate data	-	ARIMA	ANN	Predicted sunspot and money exchange rate
[102]	Stock price	Stock price	-	ARIMA	SVM	Predicted stock price
[103]	Canadian Lynx data	Canadian Lynx data	-	ARIMA	RNN	Predicted Canadian Lynx
[104]	Sunspot data and Money exchange rate data	Sunspot data and Money exchange rate data	-	ARIMA	PNN	Predicted sunspot and money exchange rate

CHAPTER V

NUMERICAL WEATHER PREDICTION AND POST-PROCESSING

This chapter describes about the methods for improving the accuracy of NWP outputs which are usually called post-processing methods. Common methods are spatial averaging, Model Output Statistics (MOS), and Kalman filter (KF). The spatial averaging method uses mean of the forecasted value from NWP instead of the nearest forecasted value of the site of interest for reducing the variance of forecast values. MOS uses the predicted weather variables from NWP model and the measured weather variables to predict the weather variable of interest. KF is usually applied to MOS for updating the coefficients of MOS over time. However, MOS should include only the important variables for explaining the characteristics of the weather variable of interest. Therefore, the statistical methods such as partial correlation, stepwise regression, and subset regression are used to specify the important variables to the weather variable of interest. The details of each are described as follows. Note that, the Numerical Weather Prediction (NWP) model that we use in this work is Weather Research and Forecasts (WRF) model which is a kind of regional NWP model as described in Section 4.2.

5.1 Spatial averaging

The NWP outputs are reported in the form of latitude and longitude grid points as shown in Figure 5.1.

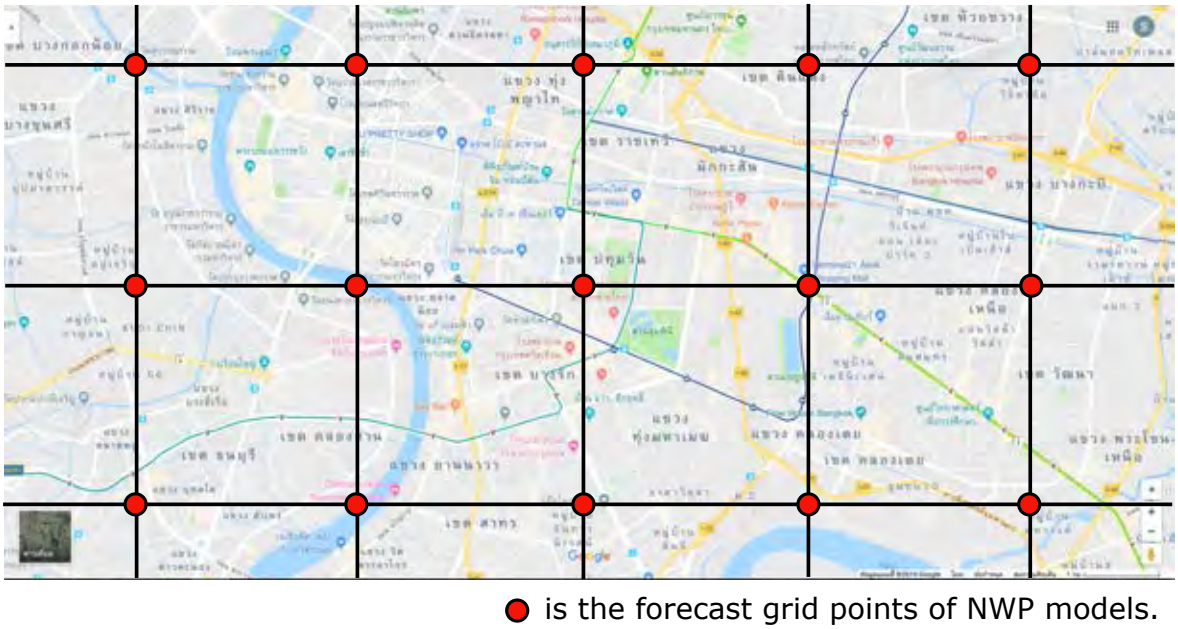


Figure 5.1: Coordinate grid points of NWP forecast.

Therefore, the method that uses the averaged value of nearby forecasted grid points instead of only the nearest grid point called spatial averaging [9]. Spatial averaging is used to reduce the variation of NWP outputs. For example, the equation of spatial averaging for solar irradiance forecasting can be expressed as

$$\hat{I}_{\text{spatial}}(t) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{I}_{\text{nwp}}(p_i, q_j)(t), \tag{5.1}$$

where $\hat{I}_{\text{spatial}}$ is the averaged value of predicted solar irradiance, $\hat{I}_{\text{mwp}}$ is the predicted solar irradiance from NWP model, m is the number of different latitude, n is the number of different longitude, (p_i, q_j) is latitude and longitude coordinates, and t is time index. According to [9], the spatial averaging can be used to reduce the forecasting error of multiple sites by half with spatial area of $3° \times 3°$.

5.2 Model Output Statistics (MOS)

MOS is firstly introduced by [2]. MOS is a regression model for modeling the relationships between a predictand and predictors to refine the predicted variable from NWP models. The general equation of MOS can be expressed by

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n, \tag{5.2}$$

where y is the refined weather variable of interest from NWP models, $\beta_1, \beta_2, \ldots, \beta_n$ are regression coefficients, and $x_1, x_2, \ldots, x_n$ are the measured relevant variables to weather variable of interest, including the predicted weather variables of interest from NWP models. The flowchart of MOS is shown in Figure 5.2.

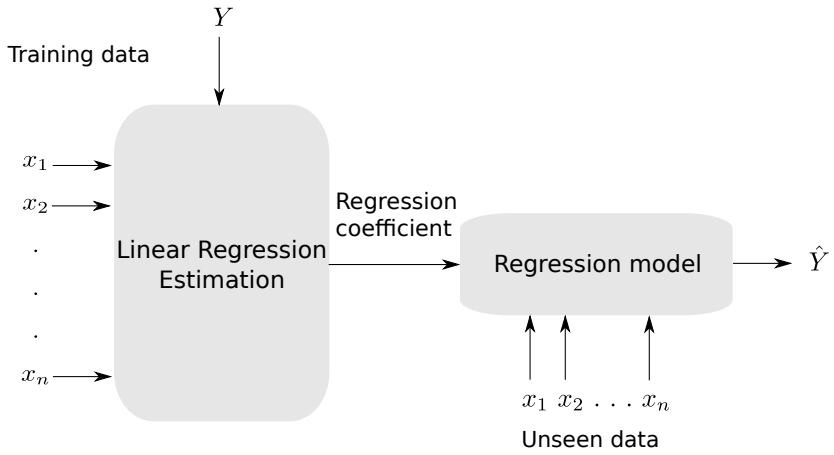


Figure 5.2: Diagram of general Model Output Statistics of [2].

According to [7,82,83], the solar irradiance forecasting performances from various NWP models are compared. Moreover, these works are also investigated in MOS and the results show that all the NWP models with MOS provide lower prediction error than the original NWP output. Moreover, MOS can also be used for bias correction where $\widehat{\text{Bias}}(t) = \hat{I}_{\text{nwp}}(t) - I(t)$, then the $\widehat{\text{Bias}}(t)$ is regressed on the relevant variables. The estimated bias is used to correct the predicted solar irradiance by $\hat{I}_{\text{mos}}(t) = \hat{I}_{\text{nwp}}(t) - \widehat{\text{Bias}}(t)$ [9,10,16]. In [108], MOS with the satellite data as input and MOS with the ground measurement data as input are compared. The results show that MOS with ground measurement data as input provides lower prediction error than MOS with the satellite data as input in both hourly and daily forecast horizon. According to [109], MOS and ANN are cascaded in order to reduce the prediction error.

MOS is a simple method that is used to reduce the prediction error from NWP models. However, MOS should includes only the important variables to solar irradiance in order to explain the characteristics of solar irradiance. The important variables to solar irradiance can be specified by statistical methods such as partial correlation, stepwise regression, and subset regression which are described as follows.

Partial correlation: A partial correlation is a measure of relations between two variables by taking into account the effect of the third variable that may jointly influence the two variables of interest. Specifically, let X and Y be the two variables and Z is another common variable. The partial correction between X and Y is the correlation between two residual errors of X and Y after regressing the effect from Z. If the partial correlation is zero, then the two variables are not correlated (even after conditioning the remaining variables).

Stepwise regression: Stepwise regression [110] is used to select relevant variables from hypothesis testing of estimated regression coefficients. For this work, only forward and backward stepwise regression have been applied.

- Forward stepwise regression: There are no variables in the regression model at the beginning. Then solar irradiance is regressed on each variable individually and apply

hypothesis testing of estimated coefficients. A variable with the lowest p-value will be added to the regression model and repeat again until there are no variables that meet the criterion of hypothesis testing (95 % confidence interval).

- Backward stepwise regression: All variables are included in the regression model at the beginning. Then solar irradiance is regressed on all variables and apply hypothesis testing of estimated coefficients. A variable with the highest p-value will be removed from the regression model and repeat again until there are no variables that meet the criterion of hypothesis testing (95 % confidence interval).

Subset regression: This method performs regression of solar irradiance on all combinations of relevant meteorological variables. Therefore, the total combinations of regression model is 2^p, where p is the number of the relevant meteorological variables. To evaluate the performance of each model, we use Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [47] in order to find the optimal model. The model which provides the lowest AIC or BIC score will be selected. AIC and BIC can be expressed by

$$\mathrm{AIC} = N + N\log(2\pi) + N\log(\mathrm{SSE}/N) + 2d, \tag{5.3}$$

$$\mathrm{BIC} = N + N\log(2\pi) + N\log(\mathrm{SSE}/N) + d\log(N), \tag{5.4}$$

where N is the number of observed data, d is the number of parameters, $\mathrm{SSE} = \sum_{t=1}^{N}(I(t) - \hat{I}(t))^2$, $I(t)$ is the measured solar irradiance at time t and $\hat{I}(t)$ is the refined solar irradiance at time t. The important variables which are frequently selected from the three statistical methods will be considered to include in MOS model.

5.3 MOS and Kalman filter

In MOS model, the regression coefficients are fixed and may not be suitable to use to predict the solar irradiance which changes over time. Therefore, the regression coefficients of MOS should be updated and Kalman filter is a method that can be used to update the regression coefficients over time. Normally, KF is used to estimate unobservable variables from the previous state of relevant observable variables recursively. KF is an optimal estimator for conditional mean of a random variable [111]. The state equation and observation equation can be expressed by

$$\text{State equation : } z(t+1) = A(t)z(t) + w(t), \tag{5.5}$$

$$\text{Observation equation : } y(t) = C(t)z(t) + v(t), \tag{5.6}$$

where $z(t)$ is the estimated variable at time t, $A(t)$ is the state transition matrix at time t, $w(t)$ is white Gaussian process noise with zero mean and covariance W at time t, $y(t)$ is observed variable at time t, $C(t)$ is observation matrix at time t, and $v(t)$ is white Gaussian measurement noise with zero mean and covariance V at time t. In realistic, $z(t)$ in state equation is unknown, so it changes over time based on the dynamical of state transition matrix $A(t)$ plus process noise $w(t)$ which is also unknown. Then, the estimated $z(t)$ can be obtained by conditional mean given measurement $y(t)$ called $\hat{z}(t|t)$, where the notation of $x(a|b)$ means the variable x at time a given observe variable from time $0, \ldots, b$. The

error covariance of $\hat{z}(t|t)$ can also be obtained from conditional mean called $P(t|t)$ where $P(t|t) = \mathbf{E}(z(t) - \hat{z}(t|t))(z(t) - \hat{z}(t|t))^T$ and it is necessary for calculate the Kalman gain. In order to run KF, $A(t)$, $C(t)$, W, and V must be known. Moreover, KF also needs an initial conditions $\hat{z}(0|0)$ and initial conditions of error covariance matrix $P(0|0)$ in order to predict and update $\hat{z}(t|t)$ and $P(t|t)$ recursively which called time update and measurement update respectively. Time update and measurement update are described as follows [111].

Time update

$$\hat{z}(t+1|t) = A(t)\hat{z}(t|t), \tag{5.7}$$

$$P(t+1|t) = A(t)P(t|t)A(t)^T + W(t), \tag{5.8}$$

Measurement update

$$\hat{z}(t|t) = \hat{z}(t|t-1) + P(t|t-1)C(t)^T(C(t)P(t|t-1)C(t)^T + V)^{-1}(y(t) - C(t)\hat{z}(t|t-1)), \tag{5.9}$$

$$P(t|t) = P(t|t-1) - P(t|t-1)C(t)^T(C(t)P(t|t-1)C(t)^T + V)^{-1}C(t)P(t|t-1), \tag{5.10}$$

where $\hat{z}(t+1|t)$ is the conditional mean of $z(t+1)$ given $y(0), \ldots, y(t)$, $\hat{z}(t|t)$ is the conditional mean of $z(t)$ given $y(0), \ldots, y(t)$, $P(t+1|t)$ is the error covariance of $\hat{z}(t+1|t)$, $P(t|t)$ is error covariance of $\hat{z}(t|t)$, and

$$P(t|t-1)C(t)(C(t)^T P(t|t-1)C(t) + V)^{-1} \tag{5.11}$$

is the Kalman gain. Then, the predicted output can be obtained by $\hat{y}(t) = C(t)\hat{z}(t|t-1)$.

KF based on MOS is also widely used to refine the predicted weather variables from NWP. The regression coefficients (β) of MOS are obtained from least squares method which are fixed values. Then, the purposes of KF are prediction and correction β over time. In order to refine the predicted weather variables from NWP by KF, typically, we do not know exactly what $A(t)$, $C(t)$, W, and V should be. Then, those parameters are usually assumed. According to [112–115], these works aim to refine the predicted wind speed and temperature from NWP models by MOS+KF. All of those studies determine the state equation as random walk and noise covariances can be estimated by residual error of previous N iterations. In [11], this work aims to refine the predicted solar irradiance from NWP models by MOS+KF. The input of this model is only the predicted solar irradiance from NWP. The state equation is assumed to be random walk and the noise covariances are estimated from the latest residual error of 30 days. From [12], the predicted solar irradiance from NWP and solar zenith angle are included in MOS+KF model to predict the solar irradiance. The noise covariances of this work are constant values. Moreover, [116] uses 24 KF models (one model for one specific time) to refine the predicted solar irradiance from NWP model one day in advance and the inputs of these model are the predicted solar irradiance from NWP model and the measured solar irradiance.

CHAPTER VI

PROBLEM STATEMENT AND FORMULATION

This chapter describes about problem statement derived from practical goal and constraints. We proposed a solar forecast scheme and describe the model as follows. Note that, our proposed solar forecast scheme requires the predicted weather variables from NWP model. For more clearly, we use WRF model which is a kind of regional NWP model to predict the weather variables for this work.

6.1 Framework

This work aims to predict PV power one day in advance. The predicted information can help the Short-term Operation Planning Section (SOPS), Generation Operation Planning Department, Electricity Generating Authority of Thailand for more efficient dispatching. SOPS usually starts to plan the generation of the next day after 13.00 hrs daily. Therefore, the forecasting model should be run at 13.00 hrs (forecasting time (t_f) = 13.00 hrs) and provide the predicted PV power of day $d+1$ between 7.00 to 16.00 hrs by 13.00 hrs of day d as shown in Figure 6.1.

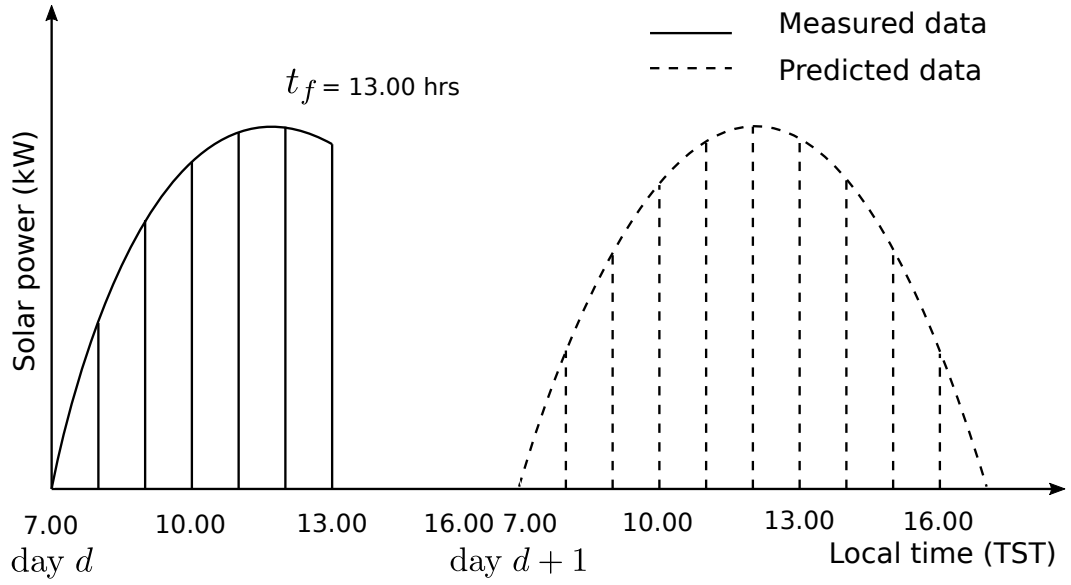


Figure 6.1: Forecasting time of this work for providing the predicted solar power one day ahead.

The resolution of solar power forecasts that SOPS uses to plan the generation is 30 minutes but we use the resolution of 1 hour in this work because currently, we have limitation on computing resource to run WRF in the resolution of 30 minutes. The scheme of this work is shown in Figure 6.2.

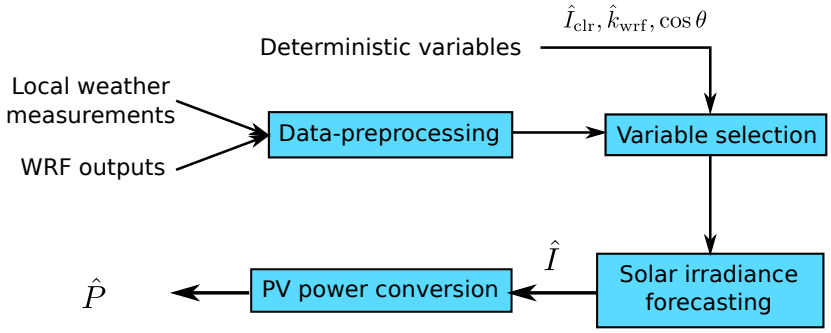


Figure 6.2: Proposed scheme of PV power forecasting with one day horizon.

From Figure 6.2, at the beginning, both local weather measurements and predicted weather variables from WRF must be available. The measured I and P usually contain the missing values, so the measurement data need to be cleaned first. Typically, the WRF data are not contain the missing data, however, the data are reported in the coordinate grid points then we use the spatial averaging method to average the forecasted values for reducing the variance of forecast values. The averaged values are used in the later process instead of a forecasted value that is closest to the PV site of interest. Then, all meteorological variables that relate to solar irradiance are selected with statistical methods in order to determine the important variables to solar irradiance. After that, the important variables to solar irradiance are used in the proposed model to predict the solar irradiance one day in advance. Then, the predicted solar irradiance values are converted to predicted PV power. In what follows, we describe the proposed solar irradiance forecasting model.

6.2 Solar irradiance forecasting model

The content from now on will describe about the proposed models. The proposed solar forecasting models can be classified into two types, i) hourly-step models and ii) daily-step models. The details of these two models are described as follows.

1. Hourly-step models: This type of the model has only one model that is used to predict the solar irradiance. The inputs of this model are the important variables to solar irradiance and then the output will be converted to PV power as shown in Figure 6.3.

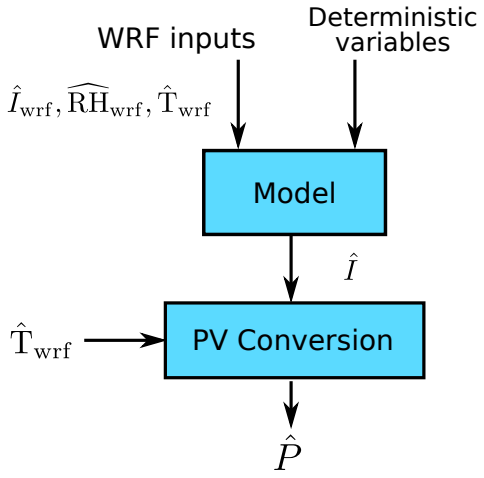


Figure 6.3: Proposed scheme of PV power forecasting (hourly-step models).

2. Daily-step model: This type of the model contains h sub-models; each of which gives the forecasts at specific hour. The h forecast values are then arranged as a forecast profile of the next day and converted to PV power as shown in Figure 6.4.

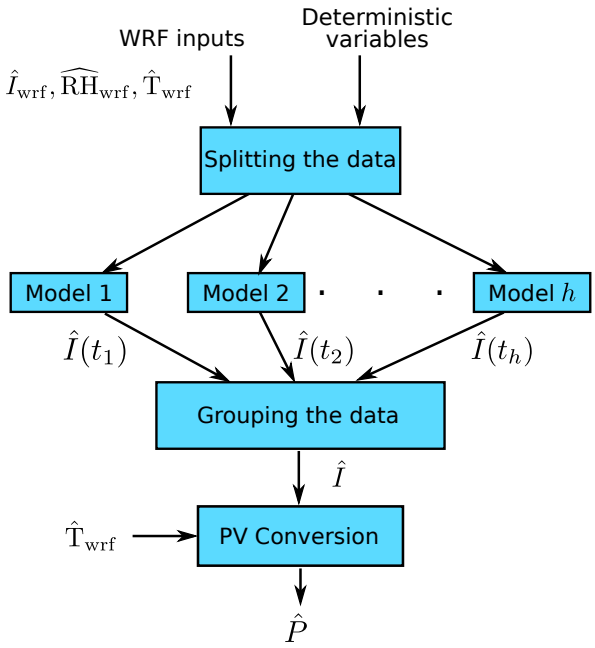


Figure 6.4: Proposed scheme of PV power forecasting (daily-step models).

The details of the proposed models both hourly-step and daily-step are described in-Chapter 7.

6.3 Solar irradiance power conversion

From the literature review of PV cells in Section 3.5, the important variables for calculating the power of the PV cells are the solar irradiance and temperature. We then propose

the model for estimating the generated power of the PV cells. The proposed model is a regression model and the inputs of the model are solar irradiance and temperature [117]. The equation of solar irradiance conversion model is expressed as

$$P(t) = \beta_1 I(t) + \beta_2 \mathrm{T}(t) + \beta_3 I(t)\mathrm{T}(t). \tag{6.1}$$

However, the variable $\mathrm{T}(t)$ is not available when we use this model to convert $\hat{I}(t)$ into $\hat{P}(t)$ of the next day. We then use the predicted temperature from WRF ($\widehat{\mathrm{T}}_{\mathrm{wrf}}$) instead of $\mathrm{T}(t)$.

CHAPTER VII

MODEL DESCRIPTION

This chapter describes the details of the model for solar irradiance forecasting which are presented in Section 6.2. The proposed models for solar irradiance forecasting are based on the regression model which can be expressed as

$$\hat{I}(t) = \beta_1 x_1(t) + \beta_2 x_2(t) + \cdots + \beta_p x_p(t) = X\beta, \tag{7.1}$$

where $x_1, \ldots, x_p$ are predictors (the important variables to solar irradiance) and β is the parameter that obtained by least-squares method. From the variable selection results in Section 9.3, we conclude that the important variables to solar irradiance are $\hat{I}_{\text{wrf}}$, $\widehat{\text{RH}}_{\text{wrf}}$, $\widehat{\text{T}}_{\text{wrf}}$, and $\cos\theta$, so these variables will be the inputs of the proposed models. Moreover, Kalman filter requires prior knowledge about parameters from least-squares (LS) fitting that are approximated noise covariance from residual error ($\hat{\sigma}^2$) and approximated error covariance of the estimated regression coefficients ($P(0|0)$) from (7.1). The approximated noise covariance from residual error is expressed as

$$\hat{\sigma}^2 = \frac{1}{N-n}\Sigma_{t=1}^{N} e(t)^2, \tag{7.2}$$

where N is the number of samples, n is the number of predictors, and e is residual error. The approximated error covariance of the estimated regression coefficients is expressed as

$$P(0|0) = \hat{\sigma}^2 (X^T X)^{-1}. \tag{7.3}$$

The models can be classified into two types, i) hourly-step models and ii) daily-step models, as described in Section 6.2. Each type of the model consists of two methods which are Model Output Statistic (MOS) and Kalman filter applied on MOS (MOS+KF). The details of the forecasting models are described as follows.

7.1 Hourly-step models

Hourly-step models consist of two methods which are MOS and MOS+KF. There are one proposed MOS model and three proposed MOS+KF models. The difference in MOS+KF models is the assumption for updating the parameters of KF. All hourly-step models are described as follows.

1. **MOS**: The inputs of this model are the important variables to solar irradiance. This model can be expressed as

$$\hat{I}(t) = \beta_1 \hat{I}_{\text{wrf}}(t) + \beta_2 \widehat{\text{RH}}_{\text{wrf}}(t) + \beta_3 \widehat{\text{T}}_{\text{wrf}}(t) + \beta_4 \cos\theta(t). \tag{7.4}$$

   The diagram of proposed hourly-step MOS is shown in Figure 7.1.

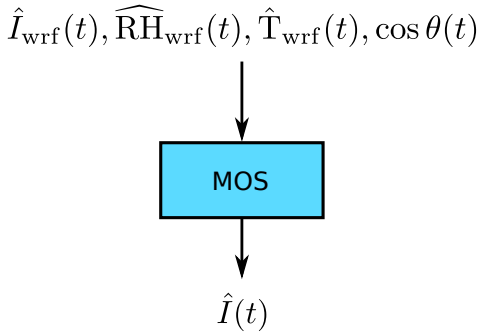


Figure 7.1: Proposed scheme of solar irradiance forecasting (hourly-step MOS model).

There are three proposed MOS+KF models and the diagram of general MOS+KF models is shown in Figure 7.2.

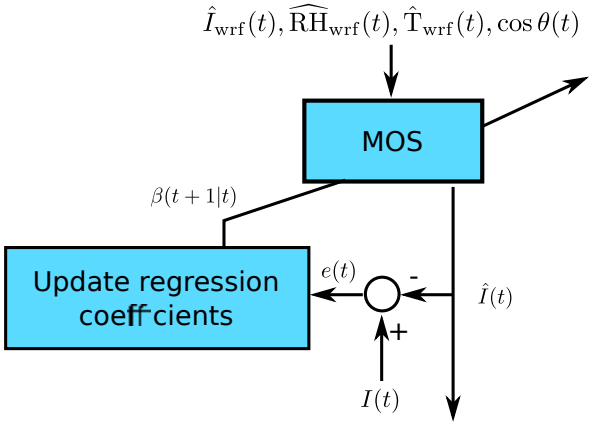


Figure 7.2: Proposed scheme of solar irradiance forecasting (hourly-step MOS+KF model).

2. **MOS+KF1**: The regression coefficients of MOS (β) are updated by the Kalman filter processes. The state-space of the dynamical system is assumed as a random walk equation,

$$z(t+1) = z(t) + w(t), \tag{7.5}$$

$$y(t) = \left[\hat{I}_{\mathrm{wrf}}(t) \quad \widehat{\mathrm{RH}}_{\mathrm{wrf}}(t) \quad \hat{\mathrm{T}}_{\mathrm{wrf}}(t) \quad \cos\theta(t)\right] z(t) + v(t), \tag{7.6}$$

where $z(t) = \beta(t)$ and $y(t) = I(t)$. The initial conditions of MOS+KF1 are chosen as follows.

$$\begin{aligned} A = I, \quad C(t) &= \left[\hat{I}_{\mathrm{wrf}}(t) \quad \widehat{\mathrm{RH}}_{\mathrm{wrf}}(t) \quad \hat{\mathrm{T}}_{\mathrm{wrf}}(t) \quad \cos\theta(t)\right], \\ V = \hat{\sigma}^2, \; W &= 10^{-6}\,\mathbf{diag}(\hat{\beta}_{\mathrm{ls}}), \\ \hat{z}(0|0) = \hat{\beta}_{\mathrm{ls}}, \quad P(0|0) &= \hat{\sigma}^2 (X^T X)^{-1}, \end{aligned} \tag{7.7}$$

where $\hat{\beta}_{\mathrm{ls}}$ is obtained by least squares and has the same value as β of (7.4), $\hat{\sigma}^2$ is obtained from (7.2), $P(0|0)$ is the error covariance matrix of $z(0|0)$, V is the approximated variance from the residual error of (7.4), and W is chosen to be $10^{-6} z(0|0)$. To calculate

the forecast values, in this work, the forecasting time (t_f) is 13.00 hrs of day d and the forecasted values are the forecasted solar irradiance of day $d+1$ between 7.00 to 16.00 hrs ($h = 10$ hours). Therefore, we have three hours left until the end of day d and we need to estimate the parameters of 7.00 hrs of day $d+1$ which is equal to the forth step ahead of the forecasting time (t_f+4). However, the parameters are updated up to time t_f, then the parameters that we need to estimate are $\hat{z}(t_f+4|t_f), \ldots, \hat{z}(t_f+13|t_f)$ and can be obtained by measurement update process as follows.

$$\hat{z}(t_f+4|t_f) = A^4\hat{z}(t_f|t_f), \hat{z}(t_f+5|t_f) = A^5\hat{z}(t_f|t_f), \ldots, \hat{z}(t_f+13|t_f) = A^{13}\hat{z}(t_f|t_f). \quad (7.8)$$

Moreover, since $A = I$, (7.8) reduces to

$$\hat{z}(t_f+4|t_f) = \hat{z}(t_f|t_f), \hat{z}(t_f+5|t_f) = \hat{z}(t_f|t_f), \ldots, \hat{z}(t_f+14|t_f) = \hat{z}(t_f|t_f). \quad (7.9)$$

Note that, the regression coefficients are not updated since t_f.

3. **MOS+KF2**: This model is modified from MOS+KF1. MOS+KF1 updates the parameters β based only on the forecasting error at time t_f and the parameters are not changed until time t_f of day $d+1$. Therefore, MOS+KF2 use the averaged forecasting error of h hours on day d for updating the parameters [117]. Then, we define the output y and the state variable z as follows.

$$y(t) = \begin{bmatrix} I(t) \\ \frac{1}{h}\sum_{k=0}^{h-1} I(t-k) \end{bmatrix},$$

$$z(t) = (\beta(t), \beta(t-1), \ldots, \beta(t-h+1)).$$

We assume that the regression coefficients evolve as random walk process ($\beta(t+1) = \beta(t)+w(t)$). Then, the state-space of the dynamical system can be expressed as follows.

$$z(t+1) = \begin{bmatrix} I & 0 & & & 0 \\ I & 0 & & & 0 \\ 0 & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix} z(t) + \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w(t), \quad (7.10)$$

$$y(t) = \begin{bmatrix} \tilde{C}(t) & 0 & \cdots & 0 \\ \frac{\tilde{C}(t)}{h} & \frac{\tilde{C}(t-1)}{h} & \cdots & \frac{\tilde{C}(t-h+1)}{h} \end{bmatrix} z(t) + v(t), \quad (7.11)$$

where $\tilde{C}(t) = \begin{bmatrix} \hat{I}_{\text{wrf}}(t) & \widehat{\text{RH}}_{\text{wrf}}(t) & \widehat{\text{T}}_{\text{wrf}}(t) & \cos\theta(t) \end{bmatrix}$. The initial conditions of MOS+KF2 are chosen as follows.

$$A = \begin{bmatrix} I & 0 & & & 0 \\ I & 0 & & & 0 \\ 0 & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix}, \quad C(t) = \begin{bmatrix} \tilde{C}(t) & 0 & \cdots & 0 \\ \frac{\tilde{C}(t)}{h} & \frac{\tilde{C}(t-1)}{h} & \cdots & \frac{\tilde{C}(t-h+1)}{h} \end{bmatrix},$$

$$\tilde{C}(t) = \begin{bmatrix} \hat{I}_{\text{wrf}}(t) & \widehat{\text{RH}}_{\text{wrf}}(t) & \widehat{\text{T}}_{\text{wrf}}(t) & \cos\theta(t) \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \frac{1}{h} & \frac{1}{h} & \cdots & \frac{1}{h} \end{bmatrix}, \tag{7.12}$$

$$V = H \begin{bmatrix} R_e(0) & R_e(1) & \cdots & R_e(h-1) \\ R_e(1) & R_e(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_e(1) \\ R_e(h-1) & \cdots & R_e(1) & R_e(0) \end{bmatrix} H^T, \quad W = \begin{bmatrix} 10^{-6}\hat{\beta}_{\text{ls}} & 0 \\ 0 & 0 \end{bmatrix},$$

$$\hat{z}(0|0) = (\hat{\beta}_{\text{ls}}, \hat{\beta}_{\text{ls}}, \ldots, \hat{\beta}_{\text{ls}}), \quad P(0|0) = \mathbf{1}_{h\times h} \otimes \hat{\sigma}^2 (X^T X)^{-1}.$$

From (7.12), $\hat{z}(0|0)$ can be chosen as $\hat{\beta}_{\text{ls}}$ at the first iteration. Then, $P(0|0) = \mathbf{cov}(\hat{z}(0|0))$ is a block matrix which consists of $\mathbf{cov}(\hat{\beta}_{\text{ls}})$. To determine the process noise covariance W, the process noise is of form $\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T w(t)$. Therefore, all blocks of W are zero except block (1,1) and $W(1,1)$ is chosen to be $10^{-6} z(0|0)$. For determining the measurement noise (V), we assume that the measurement noise ($v(t)$) can be expressed as follows.

$$\hat{v}(t) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \frac{1}{h} & \frac{1}{h} & \cdots & \frac{1}{h} \end{bmatrix} \begin{bmatrix} e(t) \\ e(t-1) \\ \vdots \\ e(t-h+1) \end{bmatrix} \triangleq H \left(\begin{bmatrix} I(t) \\ I(t-1) \\ \vdots \\ I(t-h+1) \end{bmatrix} - \begin{bmatrix} \hat{I}_{\text{mos}}(t) \\ \hat{I}_{\text{mos}}(t-1) \\ \vdots \\ \hat{I}_{\text{mos}}(t-h+1) \end{bmatrix} \right).$$

Then, $V = \mathbf{cov}(v(t)) \approx \mathbf{cov}(\hat{v}(t))$ as shown in (7.12). The forecasted values of MOS+KF2 can be obtained from the first element of the output vector y.

4. **MOS+KF3**: This model is also modified from MOS+KF1. MOS+KF3 uses the forecasting error of all hours at day d to update the parameters [117]. Then, we define the output y and the state variable z as follows.

$$y(t) = (I(t), I(t-1), \ldots, I(t-h+1)),$$

$$z(t) = (\beta(t), \beta(t-1), \ldots, \beta(t-h+1)).$$

The state-space of the dynamical system can be expressed as follows.

$$z(t+1) = \begin{bmatrix} I & 0 & & & 0 \\ I & 0 & & & 0 \\ 0 & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix} z(t) + \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w(t), \tag{7.13}$$

$$y(t) = \begin{bmatrix} \tilde{C}(t) & & & \\ & \tilde{C}(t-1) & & \\ & & \ddots & \\ & & & \tilde{C}(t-h+1) \end{bmatrix} z(t) + v(t), \tag{7.14}$$

where $\tilde{C}(t) = \begin{bmatrix} \hat{I}_{\text{wrf}}(t) & \widehat{\text{RH}}_{\text{wrf}}(t) & \widehat{\text{T}}_{\text{wrf}}(t) & \cos\theta(t) \end{bmatrix}$. The initial conditions of MOS+KF3 are chosen as follows.

$$A = \begin{bmatrix} I & 0 & & & 0 \\ I & 0 & & & 0 \\ 0 & I & & & \\ & & \ddots & & \\ & & & I & 0 \end{bmatrix}, \quad C(t) = \mathbf{diag}(\tilde{C}(t), \tilde{C}(t-1), \ldots, \tilde{C}(t-h+1)),$$

$$\tilde{C}(t) = \begin{bmatrix} \hat{I}_{\text{wrf}}(t) & \widehat{\text{RH}}_{\text{wrf}}(t) & \widehat{\text{T}}_{\text{wrf}}(t) & \cos\theta(t) \end{bmatrix}, \tag{7.15}$$

$$V = \begin{bmatrix} R_e(0) & R_e(1) & \cdots & R_e(h-1) \\ R_e(1) & R_e(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_e(1) \\ R_e(h-1) & \cdots & R_e(1) & R_e(0) \end{bmatrix}, \quad W = \begin{bmatrix} 10^{-6}\hat{\beta}_{\text{ls}} & 0 \\ 0 & 0 \end{bmatrix},$$

$$\hat{z}(0|0) = (\hat{\beta}_{\text{ls}}, \hat{\beta}_{\text{ls}}, \ldots, \hat{\beta}_{\text{ls}}), \quad P(0|0) = \mathbf{1}_{h\times h} \otimes \hat{\sigma}^2 (X^T X)^{-1}.$$

From (7.15), $\hat{z}(0|0)$ can be chosen as $\hat{\beta}_{\text{ls}}$ at the first iteration. Then, $P(0|0) = \mathbf{cov}(\hat{z}(0|0))$ is a block matrix which consists of $\mathbf{cov}(\hat{\beta}_{\text{ls}})$. To determine the process noise covariance W, the process noise is of form $\begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T w(t)$. Therefore, all blocks of W are zero except block (1,1) and $W(1,1)$ is chosen to be $10^{-6} z(0|0)$. For determining the measurement noise (V), we assume that the measurement noise ($v(t)$) can be expressed as follows.

$$\hat{v}(t) = \begin{bmatrix} e(t) \\ e(t-1) \\ \vdots \\ e(t-h+1) \end{bmatrix} \triangleq \begin{bmatrix} I(t) \\ I(t-1) \\ \vdots \\ I(t-h+1) \end{bmatrix} - \begin{bmatrix} \hat{I}_{\text{mos}}(t) \\ \hat{I}_{\text{mos}}(t-1) \\ \vdots \\ \hat{I}_{\text{mos}}(t-h+1) \end{bmatrix}.$$

Then, $V = \mathbf{cov}(v(t)) \approx \mathbf{cov}(\hat{v}(t) = \mathbf{cov}(e(t))$ which is the Toeplitz matrix as shown in (7.15). The forecasted values of MOS+KF3 can be obtained from the first element of the output vector y.

7.2 Daily-step models

Daily-step models consist of h sub-models to forecast the solar irradiance of specific h hours. Daily-step models also consist of two methods which are MOS and MOS+KF. There are one proposed MOS model and two proposed MOS+KF models. The difference in MOS+KF models is the assumption for updating the parameters of KF. All Daily-step models are described as follows.

1. **MOS**: The inputs of this model are the important variables to solar irradiance as hourly-step MOS model. However, daily-step MOS consists of h sub-models where $h = 10$ is the number of hour between 7.00 to 16.00 hrs. Each sub-model is used to predict the solar irradiance at specific hour, *i.e.* sub-model 1 is used to predict the solar irradiance at 7.00 hrs, sub-model 2 is used to predict the solar irradiance at 8.00 hrs, and so on. This model can be expressed as

$$\hat{I}^{(d+1)}(t) = \beta_1(t)\hat{I}^{(d+1)}_{\text{wrf}}(t) + \beta_2(t)\widehat{\text{RH}}^{(d+1)}_{\text{wrf}}(t) + \beta_3(t)\widehat{\text{T}}^{(d+1)}_{\text{wrf}}(t) + \beta_4(t)\cos\theta^{(d+1)}(t), \quad (7.16)$$

where $t \in \{t_1, \ldots, t_h\}$. The flowchart of daily-step MOS is shown in Figure 7.3.

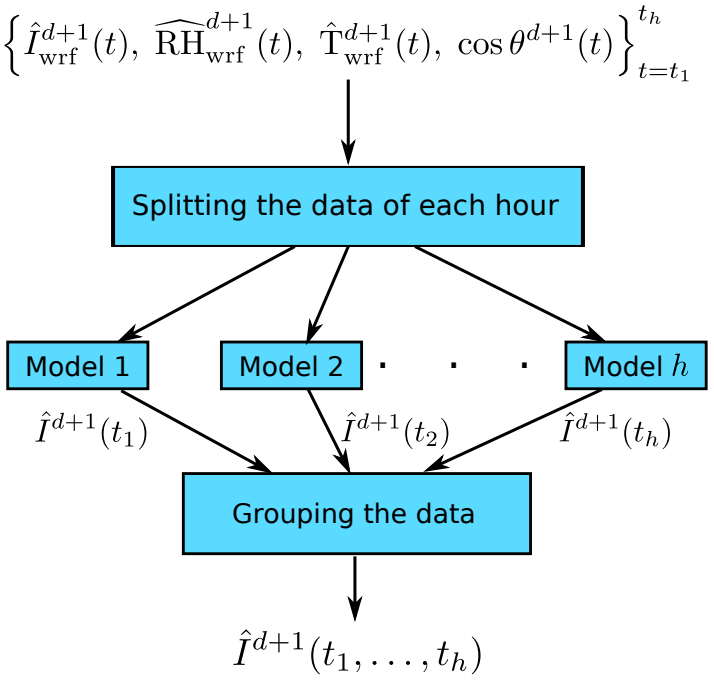


Figure 7.3: Proposed scheme of solar irradiance forecasting (daily-step MOS model).

2. **MOS+KF1**: We apply Kalman filter to daily-step MOS model called daily-step MOS+KF1. The regression coefficients (β) of sub-models are updated based on forecasting error of day d. β is assumed to evolve as a random walk process and can be expressed as

$$\beta^{(d+1)}(t) = \beta^{(d)}(t) + w^{(d)}(t).$$

We then determine the state variable and the output as

$$z^{(d)}(t) = \beta^{(d)}(t), \quad y^{(d)}(t) = I^{(d)}(t),$$

where d is day and t is hour. The state-space of the dynamical system can be expressed as

$$\begin{aligned} z^{(d+1)}(t) &= z^{(d)}(t) + w^{(d)}(t), \\ y^{(d)}(t) &= \left[\hat{I}^{(d)}_{\text{wrf}}(t) \quad \widehat{\text{RH}}^{(d)}_{\text{wrf}}(t) \quad \hat{\text{T}}^{(d)}_{\text{wrf}}(t) \quad \cos\theta^{(d)}(t)\right] z^{(d)}(t) + v^{(d)}(t), \end{aligned} \tag{7.17}$$

for $t \in \{t_1, t_2, \ldots, t_h\}$. The time index of dynamic equation of (7.17) is d. The diagram of the daily-step MOS+KF is shown in Figure 7.4

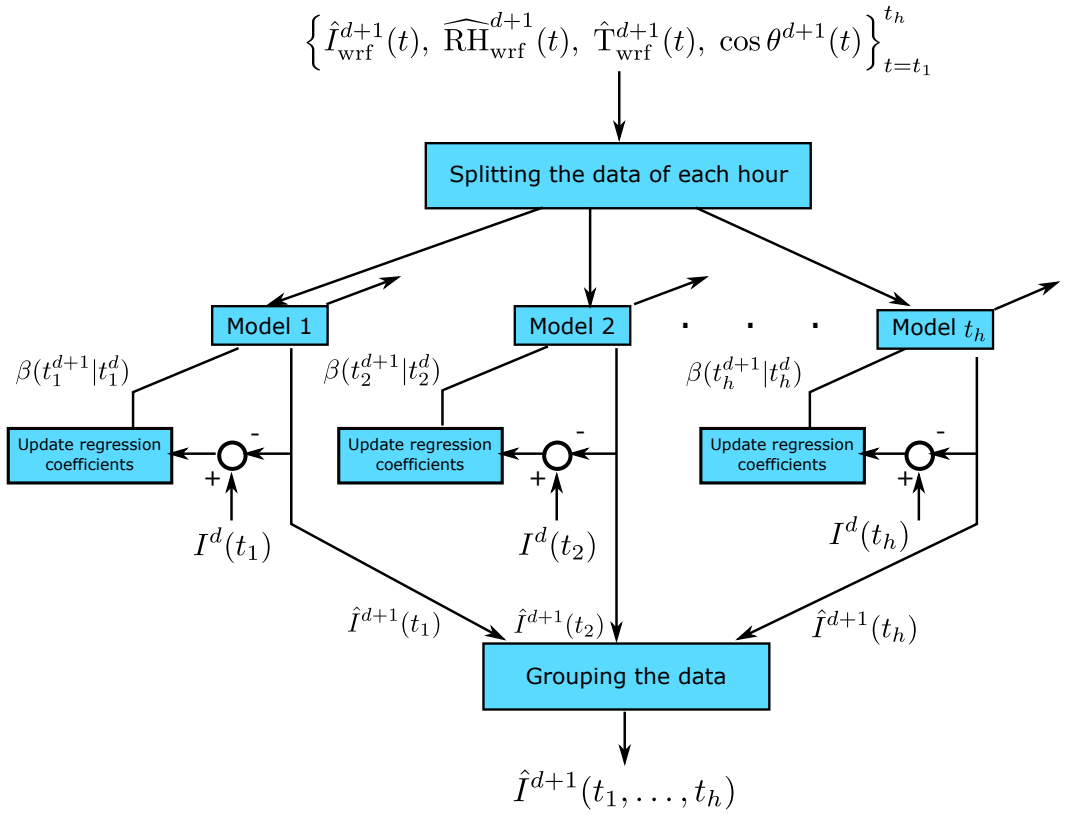


Figure 7.4: Proposed scheme of solar irradiance forecasting (daily-step MOS+KF model).

The daily-step models consists of h sub-models, however, the h sub-models can be rewritten into single vector equation as

$$z^{(d+1)} = z^{(d)} + w^{(d)}, \quad y^{(d)} = C^{(d)} z^{(d)} + v^{(d)},$$

where

$$\begin{aligned} z^{(d+1)} \triangleq \begin{bmatrix} z^{(d+1)}(t_1) \\ z^{(d+1)}(t_2) \\ \vdots \\ z^{(d+1)}(t_h) \end{bmatrix} &= \begin{bmatrix} z^{(d)}(t_1) \\ z^{(d)}(t_2) \\ \vdots \\ z^{(d)}(t_h) \end{bmatrix} + \begin{bmatrix} w^{(d)}(t_1) \\ w^{(d)}(t_2) \\ \vdots \\ w^{(d)}(t_h) \end{bmatrix}, \\ y^{(d)} \triangleq \begin{bmatrix} I^{(d)}(t_1) \\ I^{(d)}(t_2) \\ \vdots \\ I^{(d)}(t_n) \end{bmatrix} &= \begin{bmatrix} C^{(d)}(t_1) & & & \\ & C^{(d)}(t_2) & & \\ & & \ddots & \\ & & & C^{(d)}(t_h) \end{bmatrix} \begin{bmatrix} z^{(d)}(t_1) \\ z^{(d)}(t_2) \\ \vdots \\ z^{(d)}(t_h) \end{bmatrix} + \begin{bmatrix} v^{(d)}(t_1) \\ v^{(d)}(t_2) \\ \vdots \\ v^{(d)}(t_h) \end{bmatrix}. \end{aligned} \tag{7.18}$$

According to (7.18), the parameter $A = I$ and $C(t)$ are block diagonal matrices. From the Kalman filter process, if the parameters $A = I, C, W, V, P(0|0)$ are block diagonal matrices, then $K^{(d)}, P^{(d|d)}, P^{(d|d-1)}$ are also block diagonal matrices. Therefore, $\hat{z}^{(d|d)}$ and $\hat{z}^{(d+1|d)}$ are separately updated in each t_k.

Modification of KF According to the constraint, the predicted solar power must be provided by 13.00 hrs daily. Therefore, the data between 14.00 to 16.00 hrs are not available to use in update parameters process in Kalman filter, then the parameters in Kalman filter according to the missing data should not be updated. To deal with this constraint, we derive the equations in update parameter process of Kalman filter at time t_f when measurement data are partially missing [118]. We use a matrix $F = \begin{bmatrix} I_{r\times r} & 0_{r\times(h-r)} \end{bmatrix}$; where r is the number of hours between t_1 and t_f, to represent as a transformation of unavailable data y in update parameter process of Kalman filter as follows.

- Measurement update at t_f:

$$\bar{K}^{(d)}(t) = P(t|t-1)C(t)^T F^T (FC(t)P(t|t-1)C(t)^T F^T + FVF^T)^{-1} \quad (7.19)$$

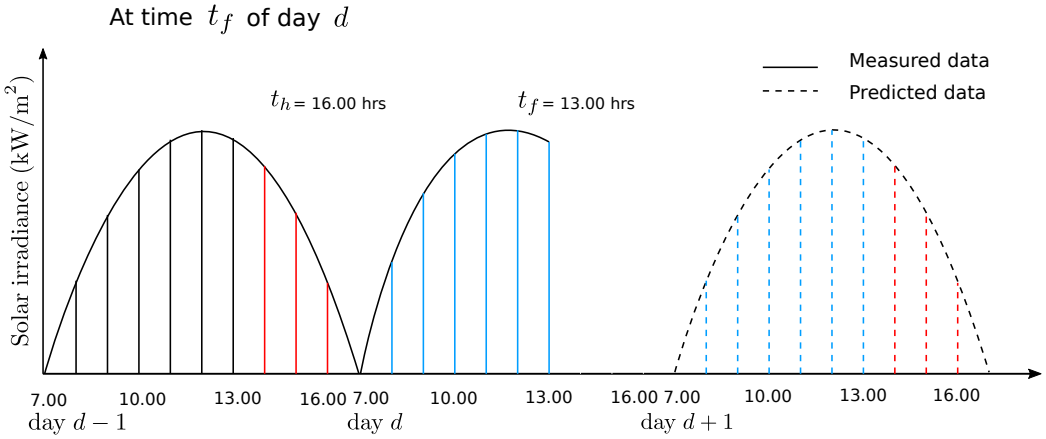
$$\bar{z}^{(d|d)}(t) = \hat{z}^{(d|d-1)}(t) + \bar{K}^{(d)}(t)(Fy^{(d)}(t) - FC^{(d)}(t)\hat{z}^{(d|d-1)}(t)), \quad (7.20)$$

  where $\bar{K}^{(d)}(t)$ is Kalman gain which is updated at time t_f of day d and $\bar{z}^{(d|d)}(t)$ is the regression coefficients which is updated at time t_f of day d. Since $A = I$ and C are block diagonal matrix, then $\bar{K}^{(d)}(t) = \begin{bmatrix} X \\ 0 \end{bmatrix}$ is that the parameters according to $t > t_f$ will not be updated.

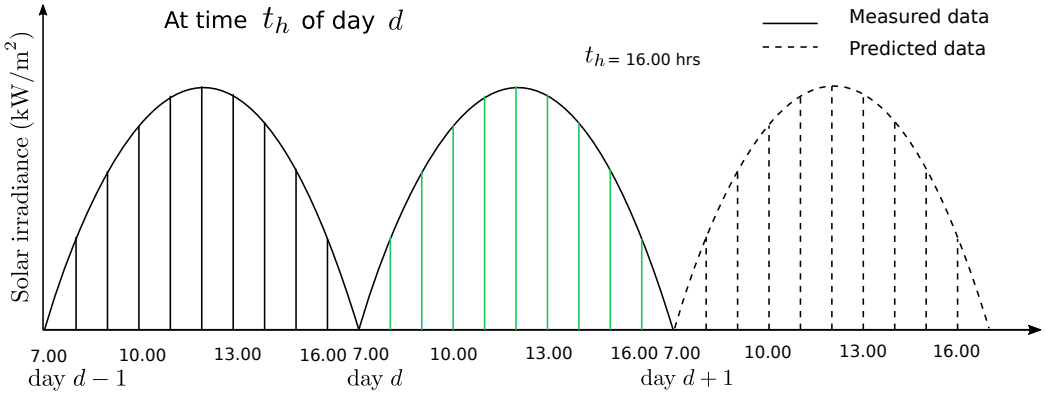
- Time update at t_f:

$$\bar{z}^{(d+1|d)}(t) = A\bar{z}^{(d|d)}(t). \quad (7.21)$$

We then determine $y^{d+1}(t) = C^{d+1}(t)\bar{z}^{(d+1|d)}(t)$ for $t \in \{t_1, t_2, \ldots, t_h\}$. For more clearly, the parameters of Kalman filter for $t \in \{t_1, t_2, \ldots, t_f\}$ are updated and are used to predict solar irradiance as blue line in Figure 7.5a. The parameters of Kalman filter for $t \in \{t_f + 1, t_f + 2, \ldots, t_h\}$ are not updated and still use the same parameters as the previous update at time t_h of day $d-1$ as red line in Figure 7.5a. At the end of the day at time t_h, the local measurement data are already collected. We then update the parameters of the Kalman filter using normal procedure again for the next iteration as green line in Figure 7.5b.



(a) Updating parameters at time t_f.



(b) Updating parameters at time t_h.

Figure 7.5: Updating time of the parameters in Kalman filter.

The initial parameters $\beta(t)$ of daily-step MOS+KF1 are the same as $\hat{\beta}_{\text{ls}}(t)$ of daily-step MOS which are obtained from least-squares method. We determine $\{e^{(d)}(t)\}_{d=1}^{N}$ is the residual error of daily-step MOS model of time t, variance of $\{e^{(d)}(t)\}_{d=1}^{N}$ is $\sigma^2(t) = \frac{1}{N-p}\sum_{d=1}^{N}|e^{(d)}(t)|^2$. Then, the initial parameters for Kalman filter of model t_k can be chosen as follows.

$$\begin{aligned} \hat{z}^{(0|0)} &= (\hat{\beta}_{\text{ls}}(t_1), \hat{\beta}_{\text{ls}}(t_2), \ldots, \hat{\beta}_{\text{ls}}(t_h)), \quad \mathbf{cov}(\hat{\beta}_{\text{ls}}(t)) = \hat{\sigma}^2(t)(X^T X)^{-1}, \\ P^{(0|0)} &= \mathbf{diag}(\mathbf{cov}(\hat{\beta}_{\text{ls}}(t_1)), \mathbf{cov}(\hat{\beta}_{\text{ls}}(t_2)), \ldots, \mathbf{cov}(\hat{\beta}_{\text{ls}}(t_h)), \\ W &= 10^{-4}\,\mathbf{diag}(\hat{z}^{(0|0)}). \end{aligned} \tag{7.22}$$

For estimating the noise covariance V, we estimate V from the residual errors of daily-step MOS. The residual errors of daily-step MOS can be expressed by

$$\hat{v}^{(d)} = \begin{bmatrix} e^{(d)}(t_1) \\ e^{(d)}(t_2) \\ \vdots \\ e^{(d)}(t_h) \end{bmatrix}.$$

Moreover, there are two assumptions for estimating the noise covariance V as follows.

(a) $e^{(d)}(t_j)$ and $e^{(d)}(t_k)$ are not correlated.

(b) $e^{(d)}(t_j)$ and $e^{(d)}(t_k)$ are correlated.

We then propose two MOS+KF models based on MOS+KF1 called MOS+KF1a and MOS+KF1b. The difference between MOS+KF1a and MOS+KF1b is the assumption of the noise covariance. The proposed MOS+KF1a and MOS+KF1b are described as follows.

(a) **MOS+KF1a**: $e^{(d)}(t_j)$ and $e^{(d)}(t_k)$ *are not correlated* and the noise covariance V can be expressed by

$$V = \mathbf{diag}(\mathbf{cov}(e^{(d)}(t_1)), \mathbf{cov}(e^{(d)}(t_2)), \ldots, \mathbf{cov}(e^{(d)}(t_h))). \tag{7.23}$$

Note: V and $P^{(0|0)}$ in (7.22) are block diagonal matrices, then the h sub-models of Kalman filter are updated *independently*.

(b) **MOS+KF1b** [117]: $e^{(d)}(t_j)$ and $e^{(d)}(t_k)$ *are correlated* and the noise covariance V can be expressed by

$$V = \mathbf{cov}(\hat{v}^{(d)}) = \begin{bmatrix} \mathbf{cov}(e^{(d)}(t_1)) & \mathbf{cov}(e^{(d)}(t_1), e^{(d)}(t_2)) & \cdots & \mathbf{cov}(e^{(d)}(t_1), e^{(d)}(t_h)) \\ \mathbf{cov}(e^{(d)}(t_2), e^{(d)}(t_1)) & \mathbf{cov}(e^{(d)}(t_2)) & & \mathbf{cov}(e^{(d)}(t_2), e^{(d)}(t_h)) \\ \vdots & & \ddots & \\ \mathbf{cov}(e^{(d)}(t_h), e^{(d)}(t_1)) & \cdots & & \mathbf{cov}(e^{(d)}(t_h)) \end{bmatrix}, \tag{7.24}$$

where the covariance (assume mean is removed) can be expressed as

$$\mathbf{cov}(e^{(d)}(t_j), e^{(d)}(t_k)) = \frac{1}{N}\sum_{d=1}^{N} e^{(d)}(t_j)e^{(d)}(t_k).$$

For this choice, V is not diagonal matrix then the Kalman gain is also not diagonal matrix. If Kalman gain is not diagonal matrix, then the residual error of $y^{(d)}(t_k)$ will effect on state variable $z^{(d|d)}(t_j)$.

In conclusion, all the solar irradiance forecasting models of this work are listed as follows.

1. Hourly-step MOS
2. Hourly-step MOS+KF1
3. Hourly-step MOS+KF2
4. Hourly-step MOS+KF3
5. Daily-step MOS
6. Daily-step MOS+KF1a
7. Daily-step MOS+KF1b

The details of the inputs data are describes in Chapter 8.

CHAPTER VIII

DATA DESCRIPTION

This chapter describes about the detail of the data. The data can be separated into three sources, i) measurement ii) WRF model and iii) deterministic variables and only the data from local measurements and WRF model are described in this chapter as follows.

8.1 Measurement data

The measurement data that used in forecasting model are collected from the sensors which are installed on the top of the Electrical Engineering Building, Chulalongkorn University. The sensors that installed at Electrical Engineering Building are Pyranometer (for measuring the solar irradiance), Energy meter (for measuring the energy that generate from PV cells), Thermometer (for measuring the temperature), Hygrometer (for measuring the relative humidity), wind speed sensor, wind direction sensor, and UV index sensor. However, we just use only two measurement data that are solar irradiance (I) and power (P) to train the forecasting model. The details of the data are summarized in Table 8.1.

Table 8.1: Summary of the data description of this work.

Variables	Chulalongkorn University Lat.13.7365 Lon.100.5321			Note
	Source	Sampling period	Starting from	
I (W/m^2)	CUBEMS	3 minutes	15/12/16	Contains missing data
$\hat{I}_{\text{wrf}}$ (W/m^2)	Supachai	60 minutes	1/1/17	-
$\widehat{\text{RH}}_{\text{wrf}}$ (%)	Supachai	60 minutes	1/1/17	-
$\widehat{\text{T}}_{\text{wrf}}$ (Degree Celsius)	Supachai	60 minutes	1/1/17	-
P (kWm) (8kW)	CUBEMS	1 minute	15/12/16	Contains missing data
P (kWm) (15kW)	CUBEMS	1 minute	24/03/17	Contains missing data

Measurement data are typically missing, so we need to clean the data first. The processes to clean the measurement data of this work are shown in Figure 8.1 and the details are described as follows.

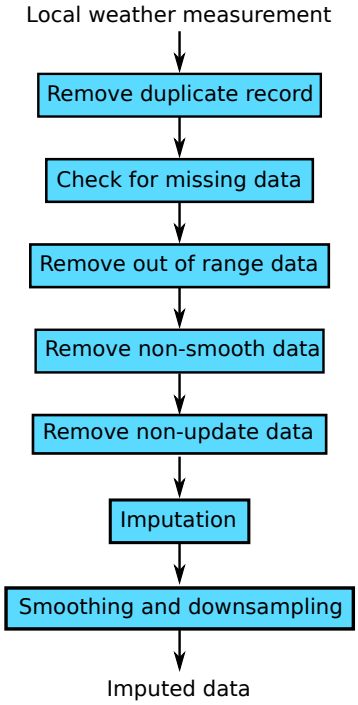


Figure 8.1: Proposed scheme of data pre-processing.

1. Duplicate records: The latest timestamp and value of the duplicate record are stored and the previous timestamp and value of the duplicate record will be removed.

2. Missing values: The missing values are automatically treated as erroneous value by MATLAB.

3. Out-of-range values (limit from sensor spec, practical range of variables): The range of solar irradiance is $0 < I < 1366\text{W/m}^2$, the range of solar power of 8 kW PV site is $0 < P < 8\text{kW}$, and the range of solar power of the 15 kW PV site is $0 < P < 15\text{kW}$. If the value exceeds the range, then treat that value as erroneous value.

4. Non-smooth values: We define the maximum rate of change of each variable as follows. The maximum rate of change of solar irradiance is 133.3 W/m^2. The maximum rate of change for 8 kW PV system is 2.5 kW and the maximum rate of change for 15 kW PV system is 5 kW. If the rate of change of the measured variables is more than the defined maximum rate of change, then the value is treated as erroneous value.

5. Non-updated values: If the value is not equal to zero and the values are not changed for one hour, then treat those values as erroneous values.

6. Imputation: The imputation method can be classified into two types, i) short consecutive missing values and ii) long consecutive missing values. Short consecutive missing values contain the consecutive missing values less than *one hour* and we use the linear interpolation method to impute the missing values. Long consecutive missing values contain the consecutive missing values that longer than *one hour* and we impute the

missing values by averaged value of the previous and the next 10 days at the same time that missing value occur.

7. Smoothing and downsampling: The imputed data usually fluctuate and we do the experiments by using just hourly data. Then we apply smoothing that using a moving average filter on the imputed data for reducing variance of the data with time span of 15 minutes before and after current value. After that, we do downsampling of the variables I and P (from the sampling period of 3 minutes and 1 minute respectively) into 60 minutes by choosing tha data that are collected at minute 0 of each hour.

The imputed data are used to train the model or used as the inputs of the forecasting model.

8.2 WRF data

In this work, the predicted weather variables are obtained from WRF model. The parameters of WRF model that need to be determined are described as follows.

- The number of spatial domain of the area of interest: We use 2 domains to predict the weather variables from WRF model. The first domain is used to forecast in a large area which consist of Western, Central, and Eastern region of Thailand. The second domain is used to forecast only in Central region of Thailand.
- Resolution of the area (km^2): The first domain has the resolution of 9×9 km^2 and the second domain has the resolution of 3×3 km^2.
- Sampling of the forecast value: The sampling of the forecast value is hourly data. Moreover, the sampling can be changed to 15 or 30 minutes. However, it consumes a lot of time to run WRF (depend on computing resources).
- Date and time of the prediction: One day ahead prediction.
- Schemes of the WRF model: The forecasting schemes of the WRF model should be consistent in location of interest. Therefore, we follow Thai Meteorological Department (TMD) to use the same forecasting schemes which may be the most suitable for Thailand. The schemes that we use are shown in Table 8.2.

The input of the WRF model is the output of GFS model. The output of GFS model is provided four times daily by the National Oceanic and Atmospheric Administration (NOAA), USA. GFS model runs daily at *0.00, 6.00, 12.00, and 18.00 Universal Time Co-ordinated (UTC) (7.00, 13.00, 19.00, and 1.00 Thailand Standard Time (TST)*, respectively) and GFS model takes 4 hours to run the prediction. Therefore, the input data of WRF model are available at *11.00, 17.00, 23.00, 5.00 TST*. From Figure 8.2, we illustrate the time schedule of the input data of WRF model. We can start to download the input data of WRF model at 11.00, 17.00, 23.00, and 5.00 TST (green dots). Our PC (CPU: Intel®Xeon®Processor E5-2620 v4 2.10GHz upto 3.00GHz 8Cores 16Threads 20MB SmartCache 8 GT/s QPI, RAM: 32GB Module - DDR4 2400 ECC Registered) takes 2 hours to run WRF model to predict the

Table 8.2: Summary of the physics options for running WRF model of this work.

Physics options	**Schemes**
Micro Physics Options	WRF Single–moment 6–class Scheme
Planetary Boundary Layer Physics Options	Mellor–Yamada Nakanishi Niino (MYNN) Level 2.5 Scheme
Cumulus Parameterization Options	Grell 3D Ensemble Scheme
Shortwave and Longwave Options	RRTMG Shortwave and Longwave Schemes
Land Surface Options	Unified Noah Land Surface Model
Surface Layer Options	Revised MM5 Scheme

weather variables one day in advance, therefore, the predicted variables data are obtained at red dots. The first predicted weather variable value always start at blue dots.

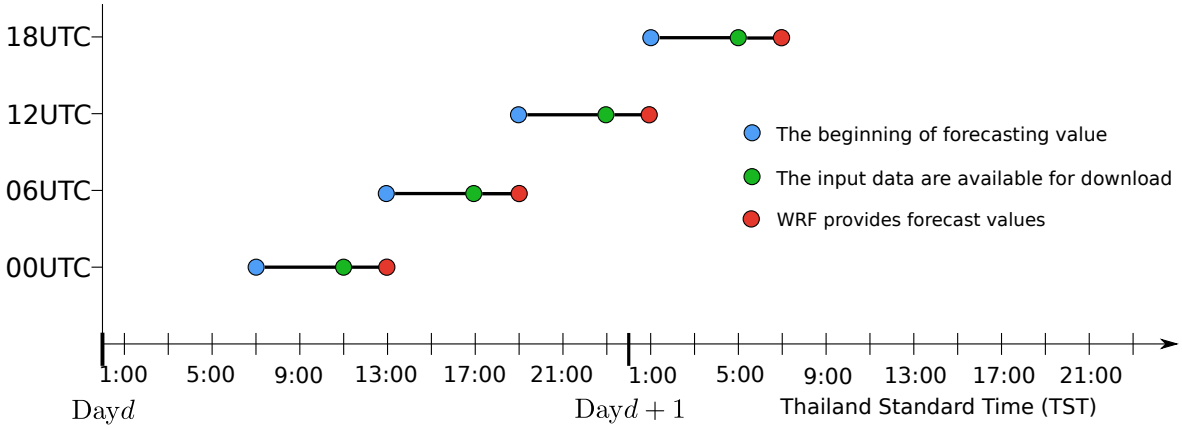


Figure 8.2: Time schedule of the input data of WRF model (the input data of 12UTC are used in this work).

If we would like to obtain WRF outputs by 13.00 hrs to compute the predicted solar power of day $d+1$, then it cannot be achieved by running WRF one day in advance. To deal with this constraint, we recommend that WRF should be run to predict the weather variables *two days* in advance with the input data of 12UTC as shown in Figure 8.3.

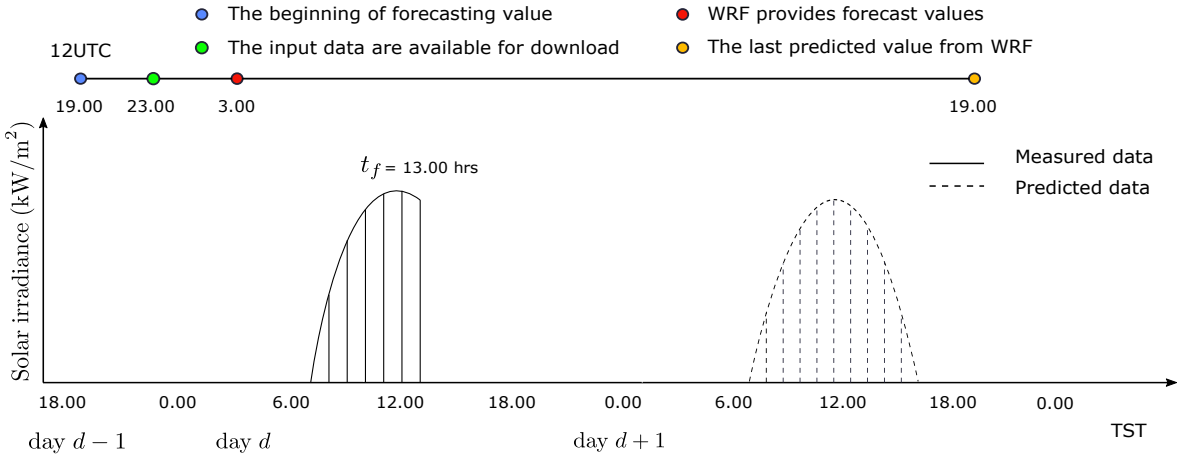


Figure 8.3: Recommended time schedule and input data for running WRF to deal with the practical constraint.

However, WRF model takes a lot of time (at least 4 hours based on our PC) to predict weather variables two days in advance, therefore, we just use WRF model to predict the weather variables *one day* in advance for this work. At the beginning, we run WRF model twice a day as TMD with the input data of 00UTC and 12UTC. Then, we compare the forecasting performance of solar irradiance between input data of 00UTC and 12UTC. From the results that will be shown in Section 9.1, we conclude that the predicted solar irradiance from WRF model with the input data of 12UTC provides lower forecasting error than the input data of 00UTC. Therefore, we run WRF model with the input data of 12UTC only.

Moreover, the forecast values from WRF are reported in the form of coordinate grid points and the forecasting error can be reduced by spatial averaging method. From the experimental result in Section 9.2, we will see that the forecasting error decreases as the spatial area increases. However, the forecasting error does not significantly reduce when the spatial area is too large. According to the experimental result in Section 9.2, we conclude that the spatial averaging with 7×7 grid points is the most suitable, so we use the spatial averaged value instead of the nearest grid point in forecasting process.

CHAPTER IX

EXPERIMENTAL RESULTS

Main experimental results and conclusions of the experiments are shown and described in this chapter. We explore the following topics in the experments.

1. Comparison between WRF with input data of 00UTC and 12UTC: This experiment compares the predicted solar irradiance from WRF model between the input data of 00UTC and 12UTC. The input data which provides the lowest RMSE is chosen as the input data for the rest of this work.

2. Improvement on spatial averaging: This experiment investigates the improvement of spatial averaging method with various spatial areas. The most suitable spatial area will be chosen and the averaged values will be used instead of the closest predicted value from WRF model to the PV site.

3. Selection of influential variables to solar irradiance: This experiment aims to determine the important variables to solar irradiance by statistical methods which are partial correlation, stepwise regression, and subset regression. The important variables to solar irradiance are the variables that are frequently selected by the statistical methods.

4. Solar irradiance forecasting: The performance of all proposed models are compared. The proposed model which provides the lowest RMSE is chosen as the best of proposed solar irradiance forecasting model for this work. Moreover, we also compare the performance between our proposed models and the models from previous work. In this experiment, we use k-fold cross validation in order to see the averaged forecasting error from each model.

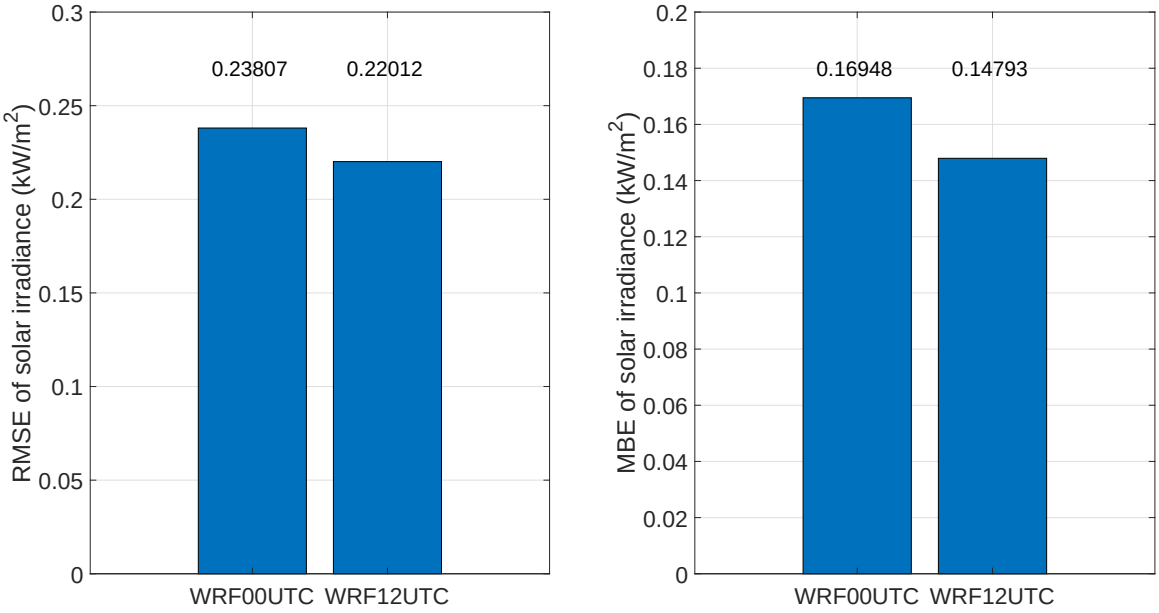
5. Solar irradiance to PV power conversion: The predicted solar irradiance from previous work and the proposed models are converted into predicted solar power. The predicted solar power data are evaluated by k-fold cross validation of the measured solar power from the PV sites of 8 kW and 15 kW installed capacity.

Moreover, there are some minor experiments that are described in appendix such as Linke turbidity estimation, imputation methods for missing data, and noise covariance estimation for Kalman filter. The results and discussion of the main experiments are shown as follows.

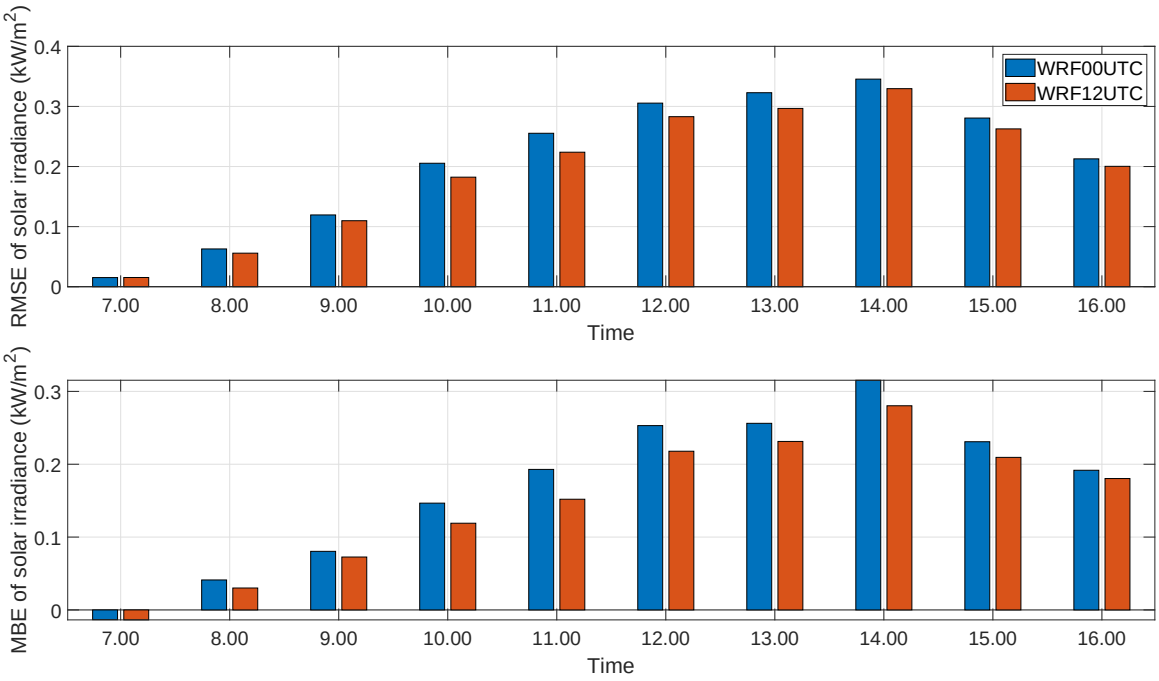
9.1 Comparison between WRF with input data of 00UTC and 12UTC

The input of WRF model is the output of GFS model which provided four times a day. However, we try to use the input of WRF model the same as TMD, therefore, output of GFS at 00UTC and 12UTC (7.00 hrs and 19.00 hrs Thailand time, respectively) are used

to forecast the solar irradiance. The GFS input that provide the lowest forecasting error in RMSE will be used in later process. In this experiment, we compare the predicted solar irradiance data of 00UTC and 12UTC on the measurement data between 1 Dec 2017 to 31 Jan 2018. The results are shown in Figure 9.1.



(a) RMSE and MBE of the predicted solar irradiance of WRF 00UTC and 12UTC.



(b) RMSE and MBE of the predicted solar irradiance of WRF 00UTC and 12UTC with specific hour.

Figure 9.1: RMSE and MBE of the predicted solar irradiance of WRF 00UTC and 12UTC of the data between 1 Dec 2017 to 31 Jan 2018.

From the results, we can see that the predicted solar irradiance of GFS at 12UTC provides lower forecasting error than GFS at 00UTC. From the results of MBE of specific hour,

we can see that the forecasting errors around noon of the GFS at 00UTC provide significantly high compared to the GFS at 12UTC. The GFS at 00UTC provides high forecasting error than the GFS at 12UTC because the forecasted values that we are interested are too close to the initial time. This problem may affect to the accuracy of the solar irradiance forecasting. Therefore, we use the predicted solar irradiance values from the GFS at **12UTC** only in subsequent experiments.

9.2 Improvement on spatial averaging

The solar irradiance always fluctuates due to cloud positions which are difficult to predict. Therefore, the nearest forecasted solar irradiance value from WRF model to the PV site of interest may provides too high forecasting error. The idea of spatial averaging can be used to reduce the forecasting error from WRF model. Therefore, this section aims to explore the improvement of the spatial averaging on predicted solar irradiance from WRF model. The spatial areas that we explore are 3×3, 5×5, and 7×7 km^2. The performance indices (RMSE and MBE) are used to justify which spatial area is most suitable. The most suitable spatial area is used to calculate the spatial averaged value of the predicted solar irradiance from WRF which is called $\hat{I}_{\text{wrf}}$ in following experiments. The data between 1 Jan 2017 to 30 Jun 2018 are used in this experiment. The results of the spatial averaging are shown in Figure 9.2

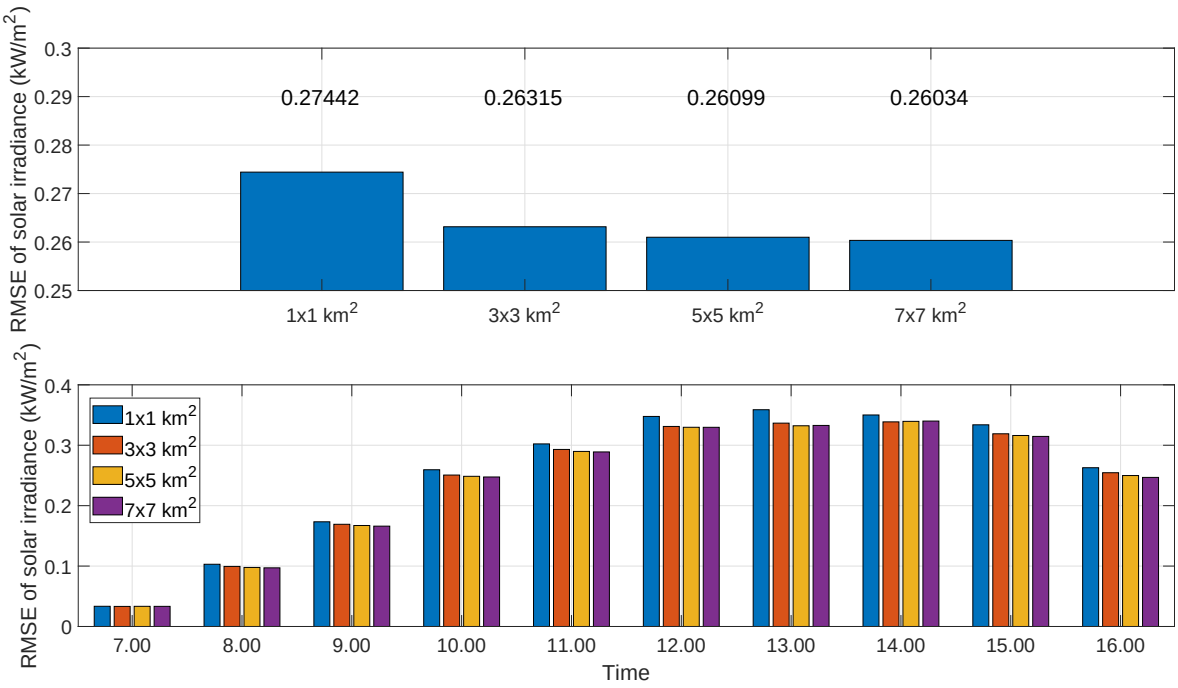


Figure 9.2: RMSE of the predicted solar irradiance from WRF with various spatial areas.

Figure 9.2 shows that RMSE decreases as spatial area is higher. However, the RMSE does not significantly reduce when the spatial area is too large. The results are the same as [9]. Therefore, we use the WRF forecast with spatial area of 7×7 km^2 instead of the single nearest grid point for all predicted variables from WRF model. Moreover, we also plot

histogram of the residual error of the predicted solar irradiance from WRF with spatial area of 7×7 km^2. The residual errors are the data between 1 Jan 2017 to 31 Dec 2018. The histogram plots of residual error are shown in Figures 9.3 and 9.4.

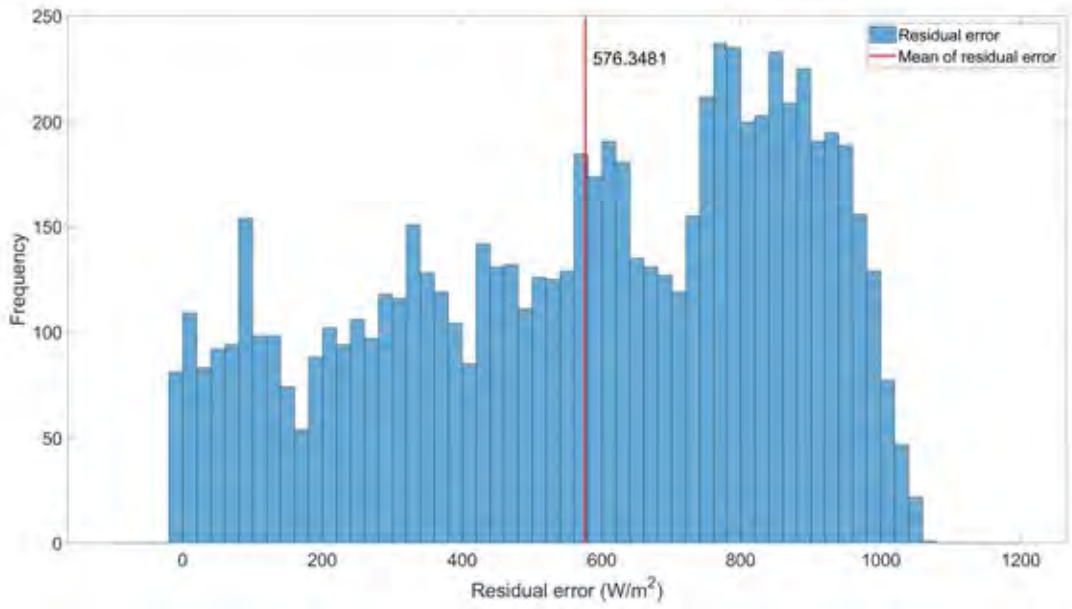


Figure 9.3: Histogram plot of residual errors of the predicted solar irradiance from WRF model.

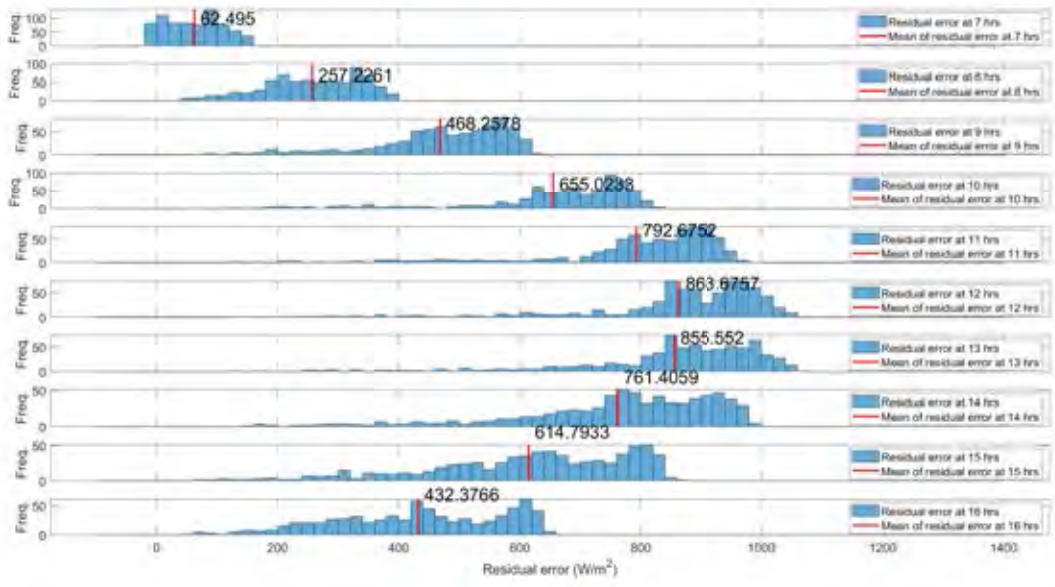


Figure 9.4: Histogram plot of residual errors of the predicted solar irradiance from WRF model in specific hour.

From Figures 9.3 and 9.4, we can see that WRF model usually overestimates the mea-

sured solar irradiance especially around noon.

9.3 Selection of influential variables to solar irradiance

The aim of this section is to analyze the influential variables to solar irradiance by using statistical methods which are partial correlation, stepwise regression, and subset regression. Selecting an important variable is based on a hypothesis testing on the regression coefficient using p-value approach to justify if the regression coefficient is different from zero. p-value is the probability of the random variable (associated with the test statistic distribution) is greater than the test statistic value computed from the data. If the p-value is a smaller than a significance level (α and determined by a user) then the variable in question is significant. The variables that we choose for testing are

$$\hat{I}_{\text{wrf}}, \widehat{\text{RH}}_{\text{wrf}}, \widehat{\text{T}}_{\text{wrf}}, I_{\text{clr}}, \cos\theta, \hat{k}_{\text{wrf}}.$$

Note that, we calculate I_{clr} from Ineichen clear sky model (3.12) with the estimated Linke turbidity value which is obtained from least-squares method. The estimated Linke turbidity for top of the Electrical Engineering Building is 4.8597, see Chapter 10 for more details. The data that we use in this experiment are the data between 1 Jan 2017 to 30 Jun 2018. Once we can determine the important variables, performance on forecasting correction is reported and the seasonal effect on the improvement can be observed.

Partial correlation

The results are shown in Table 9.1 and the selected variables from partial correlation are

$$\hat{I}_{\text{wrf}}(t), \widehat{\text{T}}_{\text{wrf}}(t), \cos\theta(t).$$

Table 9.1: Partial correlation coefficients and p-value of correlation coefficients of the selected weather variables.

	$\hat{I}_{\text{wrf}}(t)$	$\widehat{\text{RH}}_{\text{wrf}}(t)$	$\widehat{\text{T}}_{\text{wrf}}(t)$	$I_{\text{clr}}(t)$	$\cos\theta(t)$	$\hat{k}_{\text{wrf}}(t)$
Correlation	0.2155	-0.0181	-0.1657	-0.0186	0.0353	-0.0256
p-value	2.5708×10^{-58}	0.1820	6.9053×10^{-35}	0.1688	0.0091	0.0584

Stepwise regression

- Forward stepwise regression: The processes of forward stepwise regression are shown below and the estimated coefficients and p-value of forward stepwise regression are shown in Table 9.2.
  - Step 1: Adding $\hat{I}_{\text{wrf}}(t)$, p-value = 0

- Step 2: Adding $I_{\text{clr}}(t)$, p-value $= 1.758 \times 10^{-59}$
- Step 3: Adding $\widehat{\text{RH}}_{\text{wrf}}(t)$, p-value $= 0.013$
- Step 4: Adding $\widehat{\text{T}}_{\text{wrf}}(t)$, p-value $= 3.239 \times 10^{-10}$
- Step 5: Adding $\cos\theta(t)$, p-value $= 6.030 \times 10^{-10}$

Table 9.2: Estimated coefficients and p-value of the selected weather variables from forward stepwise regression.

Variables	Estimated coefficients	Standard Error	p-value
$\hat{I}_{\text{wrf}}(t)$	0.532	0.020	3.346×10^{-141}
$\widehat{\text{RH}}_{\text{wrf}}(t)$	0.868	0.190	4.745×10^{-6}
$\widehat{\text{T}}_{\text{wrf}}(t)$	-6.559	0.741	1.132×10^{-18}
$I_{\text{clr}}(t)$	-0.597	0.134	9.023×10^{-6}
$\cos\theta(t)$	924.28	149.06	6.03×10^{-10}

The selected variables from forward stepwise regression are

$$\hat{I}_{\text{wrf}}(t), \widehat{\text{RH}}_{\text{wrf}}(t), \widehat{\text{T}}_{\text{wrf}}(t), I_{\text{clr}}(t), \cos\theta(t).$$

- Backward stepwise regression: The processes of backward stepwise regression are shown below and the estimated coefficients and p-value of backward stepwise regression are shown in Table 9.3.
  - Step 1: Removing $\hat{k}_{\text{wrf}}(t)$, p-value $= 0.103$

Table 9.3: Estimated coefficients and p-value of the selected weather variables from backward stepwise regression.

Variables	Estimated coefficients	Standard Error	p-value
$\hat{I}_{\text{wrf}}(t)$	0.532	0.020	3.346×10^{-141}
$\widehat{\text{RH}}_{\text{wrf}}(t)$	0.868	0.190	4.745×10^{-6}
$\widehat{\text{T}}_{\text{wrf}}(t)$	-6.559	0.741	1.132×10^{-18}
$I_{\text{clr}}(t)$	-0.597	0.134	9.023×10^{-6}
$\cos\theta(t)$	924.28	149.06	6.03×10^{-10}

The selected variables in backward stepwise regression are the same as the results from forward stepwise regression.

Subset regression

This method performs regression of I on all combinations of relevant variables and we have six relevant variables, so the total is 2^6=64 different models with different goodness of fit and model complexity. Therefore, we use AIC and BIC in order to choose the optimal model which is the model that has the lowest AIC or BIC score from (5.3) and (5.4). In this experiment, the data are separated into two sets which are training data set and validation data set. The training data set is the data between 1 Jan 2017 to 31 Dec 2017 and this data set is also used to train the model. The validation data set is the data between 1 Jan 2018 to 30 Jun 2018 and this data set is used to evaluate the trained models. The top five models that provide low AIC and BIC score in both training and validation set are shown in Table 9.4.

The important variables to solar irradiance from each method are summarized in Table 9.5.

Table 9.5: Summary of the selected weather variables from partial correlation, stepwise regression, and subset regression.

Methods	Predictor variables					
	$\hat{I}_{\text{wrf}}(t)$	$\widehat{\text{RH}}_{\text{wrf}}(t)$	$\widehat{\text{T}}_{\text{wrf}}(t)$	$I_{\text{clr}}(t)$	$\cos\theta(t)$	$\hat{k}_{\text{wrf}}(t)$
Partial correlation	•		•		•	
Forward stepwise	•	•	•	•	•	
Backward stepwise	•	•	•	•	•	
Subset regression						
AIC training	•	•	•			
AIC validation	•	•	•		•	
BIC training	•		•	•	•	
BIC validation	•	•	•		•	

Mostly, the important variables to solar irradiance are the variables which selected by the statistical methods. From Table 9.5, we can conclude that the important variables to solar irradiance are

$$I_{\text{wrf}}, \text{RH}_{\text{wrf}}, \text{T}_{\text{wrf}}, \text{and}\cos\theta.$$

These important variables are used as the input of MOS and MOS+KF to predicted the solar irradiance in later process. Therefore, the equation of proposed MOS which is used in the rest of the experiments can be expressed as

$$\hat{I}_{\text{mos}}(t) = \beta_1\hat{I}_{\text{wrf}}(t) + \beta_2\widehat{\text{RH}}_{\text{wrf}}(t) + \beta_3\widehat{\text{T}}_{\text{wrf}}(t) + \beta_4\cos\theta(t). \tag{9.1}$$

The statistical test of (9.1) is shown in Table 9.6.

Table 9.4: Top five models which provide the lowest AIC/BIC score evaluated on training/validation data set.

Criterion	Ranking	Score	Regression variables					
			$\hat{I}_{\mathrm{wrf}}(t)$	$\widehat{\mathrm{RH}}_{\mathrm{wrf}}(t)$	$\widehat{\mathrm{T}}_{\mathrm{wrf}}(t)$	$I_{\mathrm{clr}}(t)$	$\cos\theta(t)$	$\hat{k}_{\mathrm{wrf}}(t)$
AIC training	1	4.7500×10^4	•	•	•	•	•	
	2	4.7502×10^4	•	•	•	•	•	•
	3	4.7503×10^4	•		•	•	•	
	4	4.7504×10^4	•		•	•	•	•
	5	4.7515×10^4	•	•	•		•	•
AIC validation	1	2.3701×10^4	•	•	•		•	
	2	2.3702×10^4	•	•	•	•		
	3	2.3707×10^4	•	•	•	•		•
	4	2.3707×10^4	•	•	•		•	•
	5	2.3708×10^4	•	•	•			
BIC training	1	4.7527×10^4	•		•	•	•	
	2	4.7531×10^4	•	•	•	•	•	
	3	4.7535×10^4	•		•	•	•	•
	4	4.7539×10^4	•	•	•	•	•	•
	5	4.7545×10^4	•	•	•		•	
BIC validation	1	2.3723×10^4	•	•	•		•	
	2	2.3724×10^4	•	•	•			
	3	2.3724×10^4	•	•	•	•		
	4	2.3729×10^4	•				•	
	5	2.3732×10^4	•			•		

Table 9.6: The regression coefficients and p-value of the regression coefficients of the proposed MOS model.

Variables	Regression coefficients	Standard Error	p-value
$\hat{I}_{\text{wrf}}(t)$	0.522	0.020	1.545×10^{-137}
$\widehat{\text{RH}}_{\text{wrf}}(t)$	0.183	0.190	2.279×10^{-9}
$\widehat{\text{T}}_{\text{wrf}}(t)$	-4.280	0.536	1.613×10^{-15}
$\cos\theta(t)$	267.84	19.901	1.214×10^{-40}

9.4 Solar irradiance forecasting

The performance of the solar irradiance forecasting methods are compared and summarized in this section. The experiments are separated into three topics. The first topic is the performance comparison between WRF forecast, persistence forecast, and MOS. The second topic is the performance comparison of proposed methods of MOS and MOS+KF. The last topic is the performance comparison between the proposed methods and the previous works from literature reviews. We use k-fold cross validation in order to see the average of the forecasting error of the proposed models. The data from 1 Jan 2017 to 31 Dec 2018 are split into 10 folds and each fold contains 73 days of data. Ninety percent of the data are used for training the models called training fold and the rest of the data is used to evaluate the performance of the models called test fold. Therefore, we have 10 folds of the data which contain training fold and test fold. All 10 folds of the data are used to train the proposed models and provide the predicted solar irradiance. The performance of the proposed models are evaluated with the test fold. Then the average performance indices of the proposed models are calculated. We apply 10-fold cross validation in order to evaluate the performance of the forecasting models in all following experiments. The model that provides the lowest RMSE or MBE is the best model for solar irradiance prediction.

WRF forecast, Persistence forecast, and MOS: This experiment aims to see the improvement of MOS over the persistence forecast and the spatial averaging of predicted solar irradiance from WRF model.
Note that, the persistence forecast is modified as daily-step to predict the solar irradiance of the next day. We assume that the clear sky index of day $d+1$ is the same as day d. The modified equation of persistence forecast can be expressed as follows.

$$\hat{I}^{d+1}_{\text{persist}}(t) = \hat{k}^{d}(t)\hat{I}^{d+1}_{\text{clr}}. \tag{9.2}$$

The results of this experiments are shown in Figure 9.5.

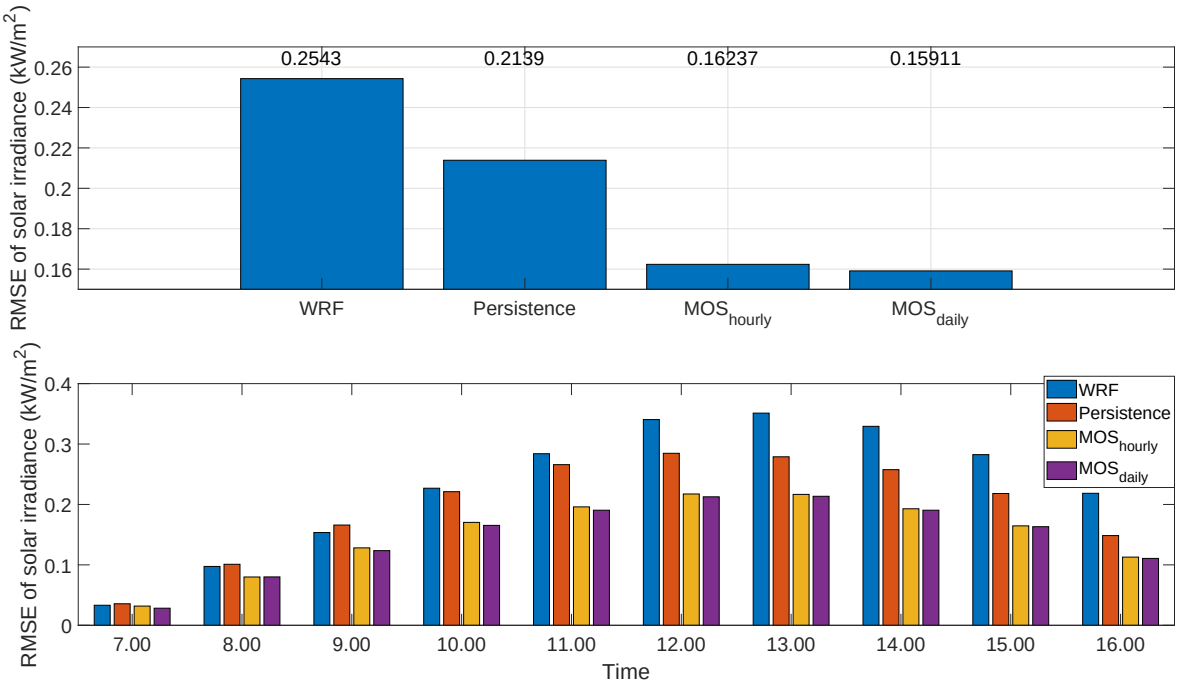
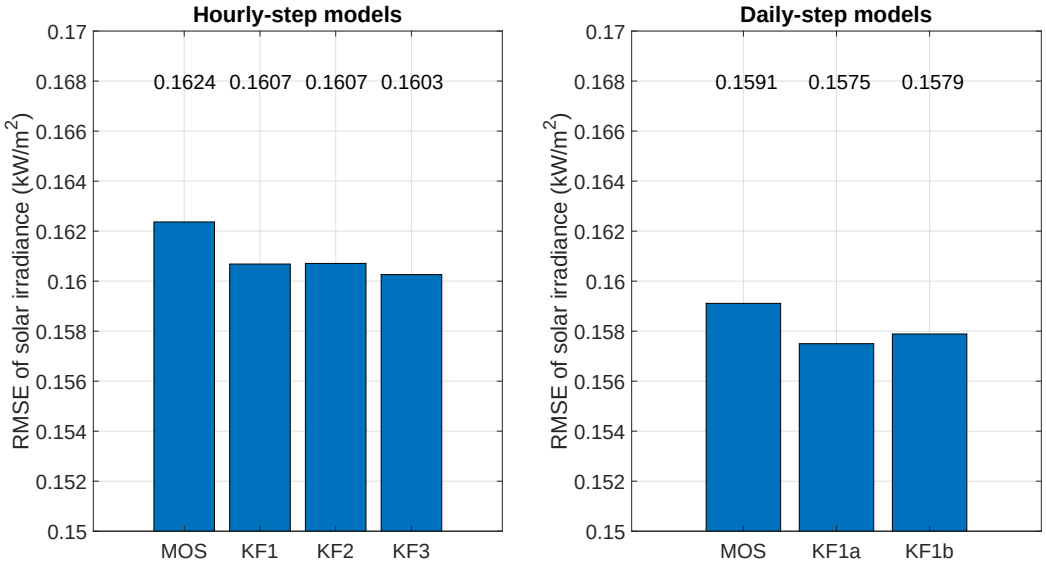


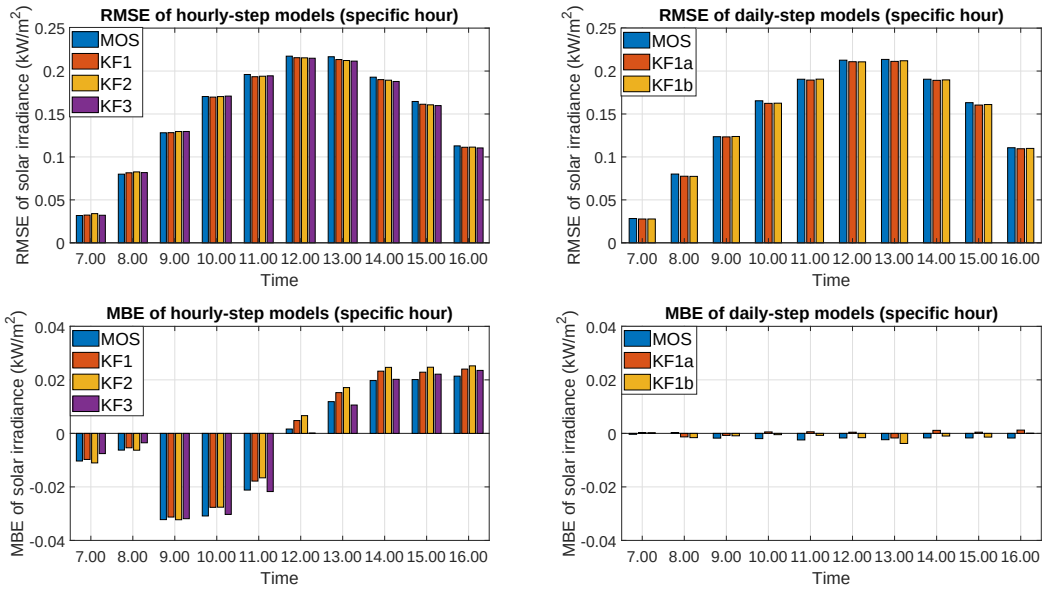
Figure 9.5: The averaged RMSE of the predicted solar irradiance of WRF, persistence forecast, hourly-step MOS, and daily-step MOS from 10-fold cross validation.

From Figure 9.5, the results show that the daily-step persistence forecast provides lower RMSE than the spatial averaged of solar irradiance from WRF model about 0.04 kW/m^2. Moreover, MOS can further reduce the RMSE about 0.05 kW/m^2 from Persistence forecast. We can conclude that MOS is simple and efficient method to forecast the solar irradiance. Besides, more complexity of MOS which is the daily-step MOS can be used to provide lower RMSE than hourly-step MOS but it is not significant. From the bar plot of specific hour, we can see that the RMSE of WRF and persistence forecast are too high around noon but the RMSE of MOS is significantly lower than WRF and persistence forecast. In conclusion, the proposed MOS outperforms the other basic solar forecasting methods. Moreover, the daily-step MOS provides lower RMSE than hourly-step MOS because of the complexity of the model. In the next topic, we apply Kalman filter on MOS which should provide better performance than MOS only.

MOS and MOS+KF: The Kalman filter that applied to MOS adjusts the regression coefficients over time. Therefore, the results of all proposed MOS+KF models should provide lower forecasting errors than MOS only. Therefore, this experiment aims to see the improvement of MOS+KF over MOS. Moreover, the daily-step models should provide lower forecasting errors than the hourly-step models because the daily-step models have high complexity than the hourly-step models. The performance indices that are used to evaluate the forecasting models are RMSE and MBE. The results are shown in Figure 9.6.



(a) RMSE.



(b) RMSE and MBE in specific hour.

Figure 9.6: The averaged RMSE and MBE of 10-fold cross validation of the proposed models.

From the results, we can see that the daily-step models provide lower RMSE than hourly-step models. In training set, the performance of MOS and MOS+KF are not significantly different in both hourly-step and daily-step models. However, the results in testing data set show that MOS+KF outperforms MOS in both hourly-step and daily-step models. We can conclude that MOS+KF can be used to provide the suitable regression coefficients of MOS when the time pass by. From hourly-step MOS+KF models, MOS+KF2 and MOS+KF3 provide lower RMSE than MOS+KF1. Therefore, we can conclude that the residual error from the previous steps can be used in Kalman filter to provide more suitable regression coefficients than using only one residual error of the previous step. However, the performance of MOS+KF1, MOS+KF2, and MOS+KF3 are not significantly different. From daily-step MOS+KF models, the results show that MOS+KF1a is better than MOS+KF1b. Therefore, we can conclude that the relationships of the residual error between $t_1, \ldots, t_h$ of

day d of MOS+KF1b cannot be used to provide more suitable regression coefficients than using independent residual error between $t_1, \ldots, t_h$ of day d of MOS+KF1a. From the bar chart of MBE, we can see that the MBE of hourly-step models are significantly high compared to daily-step models. The reasons that daily-step models are better than hourly-step models because daily-step models have higher complexity than hourly-step models. Moreover, the regression coefficients of daily-step MOS+KF models at time $t_1, \ldots, t_h$ of day $d+1$ are evolved based on the residual error at time $t_1, \ldots, t_h$ of day d. Therefore, daily-step MOS+KF models provide the suitable regression coefficients than hourly-step MOS+KF models. On the other hand, the regression coefficients at time $t_1, \ldots, t_h$ of day $d+1$ of hourly-step MOS+KF models are evolved based on the residual error at time t_f (13.00 hrs) of day d. Therefore, hourly-step MOS+KF models usually underestimate in the morning and overestimate in the after noon as shown in MBE of the specific hour.

Performance comparison with previous works: The previous works that we choose to compare the results with our proposed models are the models of [9, 11, 12] (MOS of Lorenz, MOS+KF of Pelland, and MOS+KF of Diagne respectively). We choose these three previous works because our models and the previous works are based on the same methods which are MOS and MOS+KF. Moreover, the models of these three previous works are identified as *hourly-step models*. However, the inputs of the models and the initial conditions of MOS+KF models of the previous works are different from our models. Therefore, we need to compare the performance between our models and the relevant previous works. The details of the model from previous works are described as follows.

- Model of [9] (MOS of Lorenz): This work use MOS to predict the forecasting error of the predicted solar irradiance from the NWP model called bias correction ($\text{Bias}(t) = \hat{I}_{\text{wrf}}(t) - I(t)$). $\text{MOS}_{\text{Lorenz}}$ is a fourth order polynomial regression model which the predictors are the clear sky index and the solar zenith angle. This model can be expressed as

$$\widehat{\text{Bias}}(t) = \beta_1 \hat{k}(t)^4 + \beta_2 \cos\theta(t)^4 + \beta_3 \hat{k}(t)^3 + \beta_4 \cos\theta(t)^3 + \cdots + \beta_7 \hat{k}(t) + \beta_8 \cos\theta(t). \quad (9.3)$$

  The predicted solar irradiance ($\hat{I}_{\text{Lorenz}}$) can be obtained by $\hat{I}_{\text{Lorenz}}(t) = \hat{I}_{\text{wrf}}(t) + \widehat{\text{Bias}}(t)$.

- Model of [11] (MOS+KF of Pelland): This work apply Kalman filter on bias correction. The input is $\hat{I}_{\text{wrf}}$. The state-space is determined as follows.

$$\begin{aligned} z(t+1) &= z(t) + w(t), \\ \text{Bias} &= y(t) = \begin{bmatrix} 1 & \hat{I}_{\text{wrf}}(t) \end{bmatrix} z(t) + v(t). \end{aligned}$$

  The initial conditions of this model are expressed as follows.

$$\begin{aligned} A = I, \qquad C(t) &= \begin{bmatrix} 1 & \hat{I}_{\text{wrf}}(t) \end{bmatrix}, \\ V(0) = 0.01, \quad W(0) &= 10^{-5} I, \\ \hat{z}(0|0) = 0, \quad P(0|0) &= 5 \times 10^{-5} I. \end{aligned} \quad (9.4)$$

$W(t)$ and $V(t)$ of this model are estimated by adaptive noise covariance as follows.

$$W(t) = \frac{1}{N-1}\sum_{t=1}^{N}\left\{(\hat{w}(t)-\bar{w})(\hat{w}(t)-\bar{w})^T - \left(\frac{N-1}{N}\right)[A(t)P_{t-1|t-1}A(t)^T - P_{t|t}]\right\}, \quad (9.5)$$

$$V(t) = \frac{1}{N-1}\sum_{t=1}^{N}\left\{(\hat{v}(t)-\bar{v})(\hat{v}(t)-\bar{v})^T - \left(\frac{N-1}{N}\right)[C(t)P_{t|t-1}C(t)^T]\right\}, \quad (9.6)$$

where $\hat{w}(t) = \hat{z}_{t|t} - \hat{z}_{t|t-1}$ is the error between measurement update and time update, $\bar{w} = \frac{1}{N}\sum_{t=1}^{N}\hat{w}(t)$, $\hat{v}(t) = y(t) - C(t)\hat{z}_{t|t-1}$ is the error between measurement and predicted value, $\bar{v} = \frac{1}{N}\sum_{t=1}^{N}v(t)$, and N is the number of sample data which are the latest data of 30 days. The predicted solar irradiance can be obtained by $\hat{I}_{\text{Pelland}}(t) = \hat{I}_{\text{wrf}}(t) + \widehat{\text{Bias}}(t)$.

- Model of [12] (MOS+KF of Diagne): This work apply Kalman filter on bias correction based on the following state-space equation.

$$\begin{aligned} z(t+1) &= z(t) + w(t), \\ \text{Bias} &= y(t) = \begin{bmatrix} 1 & \hat{I}_{\text{wrf}}(t) & \cos\theta(t) \end{bmatrix} z(t) + v(t). \end{aligned}$$

The initial conditions of this model are expressed as follows.

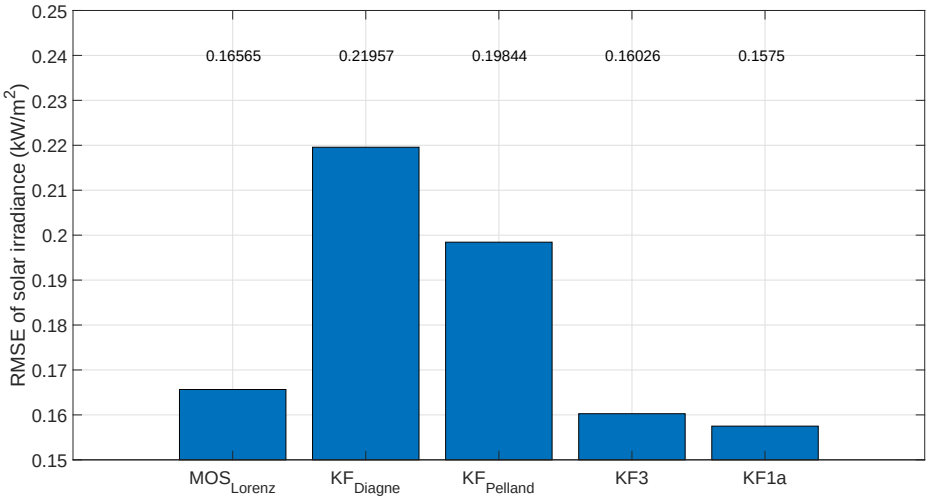
$$\begin{aligned} A = I, \qquad & C(t) = \begin{bmatrix} 1 & \hat{I}_{\text{wrf}}(t) & \cos\theta(t) \end{bmatrix}, \\ V = 0.01, \quad & W = I, \\ \hat{z}(0|0) = 0, \quad & P(0|0) = 5I. \end{aligned} \quad (9.7)$$

Note that, the predicted solar irradiance ($\hat{I}_{\text{Diagne}}$) can be obtained by $\hat{I}_{\text{Diagne}}(t) = \hat{I}_{\text{wrf}}(t) + \widehat{\text{Bias}}(t)$.

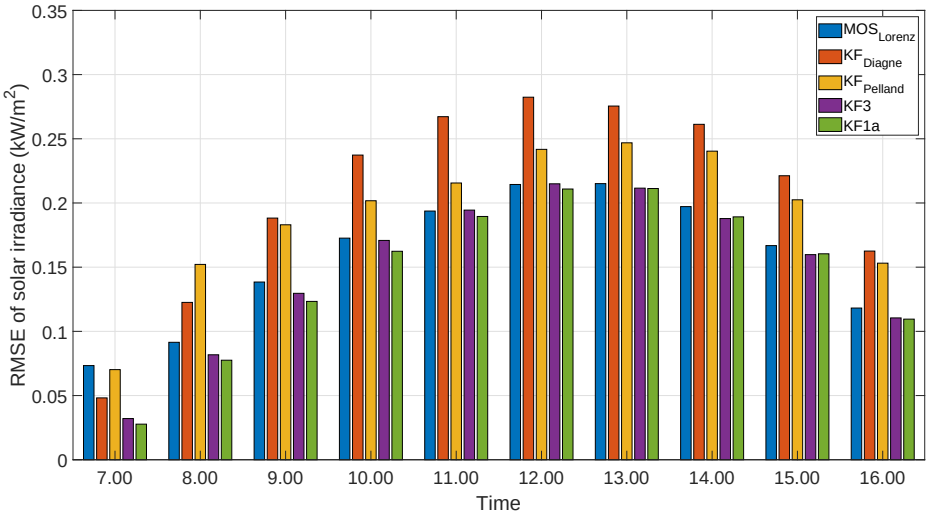
The results are shown in Figures 9.7 and 9.8. From these results, we can see that our proposed MOS+KF models outperform the others in both hourly-step and daily-step models. Among the works from literature review, MOS of Lorenz provides the lowest RMSE. However, this model is still not better than our proposed MOS in both hourly-step and daily-step models. MOS of Lorenz is worse than our proposed MOS models because the model just includes only one important variables to solar irradiance in Thailand which is solar zenith angle but our proposed MOS models have four important variables to solar irradiance in Thailand. For MOS+KF of Diagne, the RMSE is very high because noise covariance W is too high and the model includes only one important variables to solar irradiance. For MOS+KF of Pelland, the result is better than MOS+KF of Diagne, however, it is still not better than MOS of Lorenz because the adaptive noise covariance method may provide inappropriate estimated noise covariance. Moreover, this model just includes two important variables to solar irradiance.

From the results, daily-step MOS+KF1a provides the lowest performance indices which RMSE and NRMSE are 157.5 W/m^2 and 37.7 %, respectively. However, the performance indices of daily-step MOS+KF1a are not significantly different when compared to the works

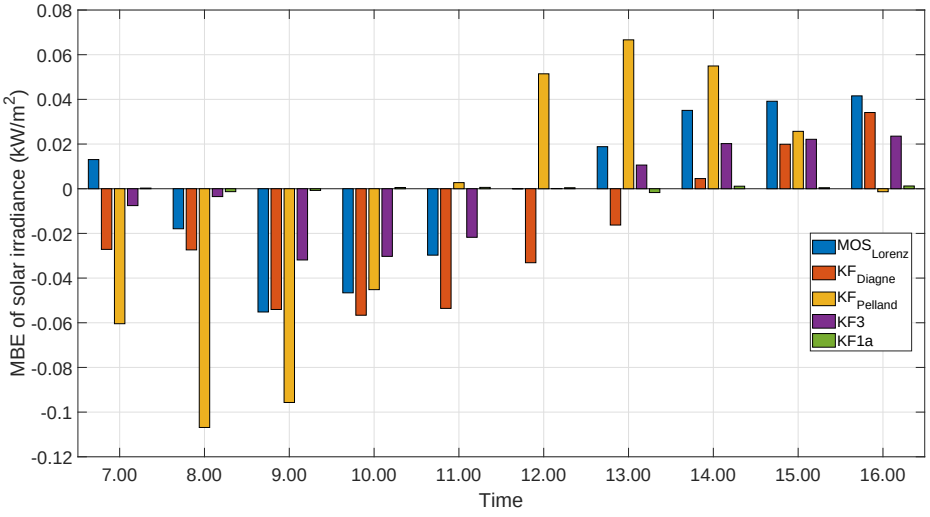
from literature review. In [9, 11], NRMSE (normalized by mean of solar irradiance) of these works were 37 % and 16.8 to 35 %, respectively. Accoreding to [12], RMSE and NRMSE (normalized by mean of solar irradiance) of this work were 119.8 to 154 W/m^2 and 22.2 to 35.5 %, respectively. The example plots of the predicted solar irradiance are shown in Figure 9.9.



(a) RMSE.



(b) RMSE in specific hour.



(c) MBE in specific hour.

Figure 9.7: RMSE and MBE of 10-fold cross validation of the predicted solar irradiance of the proposed MOS+KF models and the models from previous work.

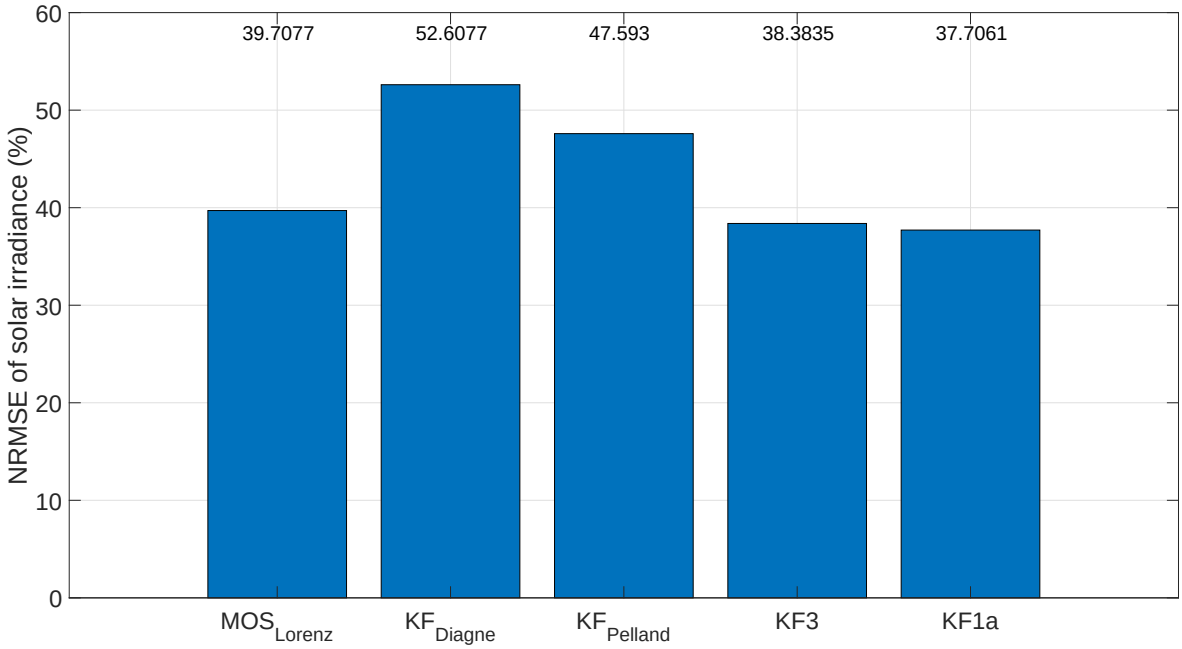
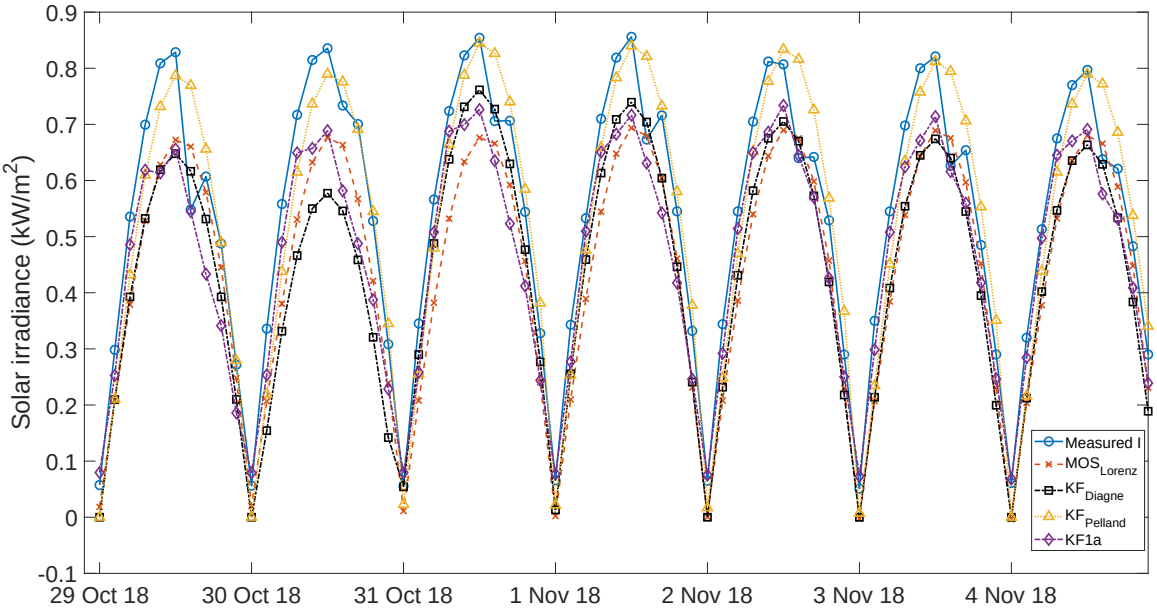
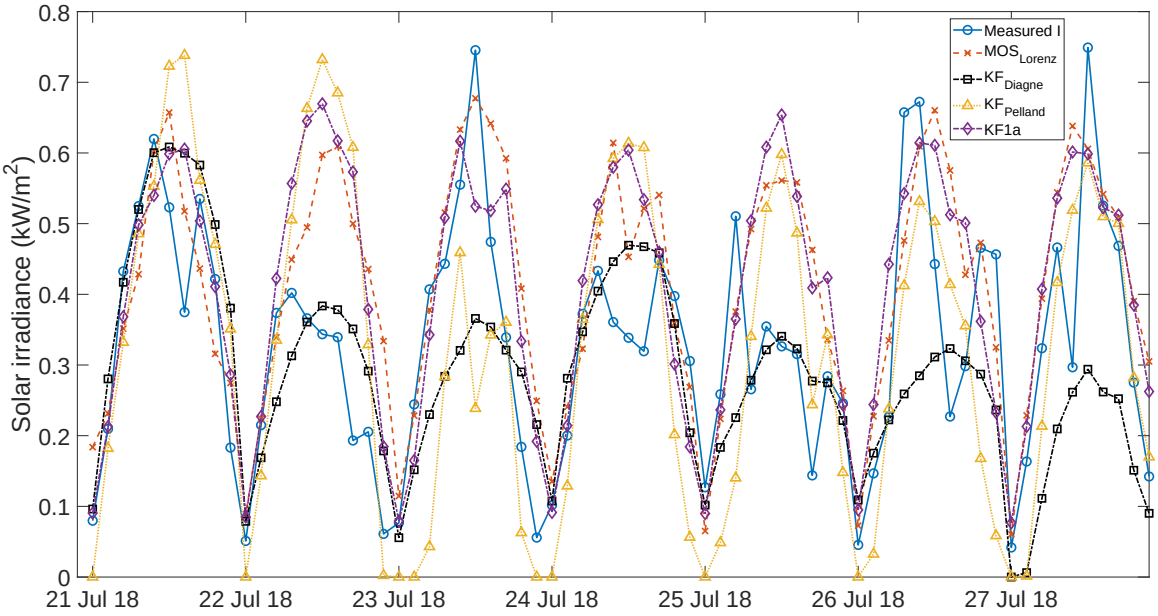


Figure 9.8: NRMSE (normalized by mean of solar irradiance) of 10-fold cross validation of the predicted solar irradiance of the proposed MOS+KF models and the models from previous work.



(a) Example plots of the predicted solar irradiance under good weather conditions.



(b) Example plots of the predicted solar irradiance under bad weather conditions.

Figure 9.9: Example plots of the predicted solar irradiance from the proposed models and the models from previous work

Note that, the solar irradiance under good condition (high magnitude in solar irradiance and bell-shaped curve) rarely occur during 1 Jul 2018 to 31 Dec 2018. Therefore, MOS+KF1a outperforms the other models because the regression coefficients of MOS+KF1a are more proper than the other models.

Discussion From the experimental results of solar irradiance forecasting, we can conclude that WRF model always provides the highest RMSE. The error of WRF model is very high around noon. Therefore, we apply MOS to reduce the forecasting error of predicted solar irradiance from WRF model. The proposed MOS models can be used to reduce the RMSE of

WRF model around 92 W/m^2. Moreover, the proposed MOS+KF models can be used to further reduce the RMSE of the proposed MOS models by 2 to 3 W/m^2. The best model among the proposed models is daily-step MOS+KF1a. MOS+KF1a has high complexity of the model and the regression coefficients are updated properly over time. Moreover, MOS+KF1a also outperforms the models from the previous work. The reasons are described as follows.

- MOS of Lorenz: This model is a regression model which is the fourth order polynomial of $\hat{k}(t)$ and $\cos\theta(t)$. However, this model contains only one important variable to solar irradiance which is $\cos\theta$. Therefore, the RMSE of this model is higher than even the proposed hourly-step MOS. We can conclude that the relevant variables to solar irradiance are the important things to explain the characteristic of the solar irradiance.

- MOS+KF of Diagne: This model contains two relevant variables to solar irradiance which are $\hat{I}_{\text{wrf}}$ and $\cos\theta$. Moreover, this model also updates the regression coefficients of MOS by Kalman filter. Therefore, this model should provide better performance than MOS of Lorenz. However, the result show that the RMSE of this model is significantly high compared to MOS of Lorenz. Therefore, we try to apply MOS of Diagne to see the performance of MOS without the Kalman filter. The RMSE of MOS of Diagne is lower than MOS of Lorenz about 5 W/m^2. Therefore, we can conclude that the Kalman filter makes this model worse. We then try to vary the noise covariance and see the performance. We found that the noise covariance W of this model is too high which makes the regression coefficients too adaptive. Moreover, the regression coefficients of this model are updated as hourly-step MOS+KF models. Therefore, the regression coefficients are updated depend on the information at time t_f (13.00 hrs). In conclusion, this model is worse than our proposed models because the author choose too high W. Moreover, this model lacks the relevant variables to solar irradiance and also low complexity.

- MOS+KF of Pelland: This model contains only one relevant variable to solar irradiance which is $\hat{I}_{\text{wrf}}$ and the regression coefficients are updated by Kalman filter. Therefore, this model should provides worse performance than MOS+KF of Diagne. However, the result show that this model is better than MOS+KF of Diagne. We then try to apply MOS of Pelland to see the performance of MOS without the Kalman filter. The RMSE of this model is higher than both Lorenz model (about 1 W/m^2) and Diagne model. Therefore, we can conclude that the Kalman filter makes this model worse. According to MOS+KF of Pelland, this model use the adaptive noise covariance method to estimate the noise covariance W and V in each iteration. We then try to use the fixed noise covariance W and V instead of adaptive noise covariance method. We found that the regression coefficients are too adaptive when the noise covariances are estimated from adaptive noise covariance method. Moreover, this model also updates the regression coefficients with the same process as the hourly-step MOS+KF models. In conclusion, this model is worse than our proposed models because the adaptive noise covariance method provides inappropriate noise covariance. Besides, this model lacks the relevant variables to solar irradiance and also low complexity. However, this model provides better performance than MOS+KF of Diagne because the adaptive noise covariance

method provides more suitable noise covariance than fixed noise covariance in MOS+KF of Diagne.

We can conclude that the suitable initial condition and noise covariance V of Kalman filter can be estimated from the residual error of MOS. The details of the experiments on noise covariance of the MOS+KF model are shown in Chapter 10. After we obtain the predicted solar irradiance then we convert the predicted solar irradiance into the predicted solar power of the PV site which is described in Section 9.5.

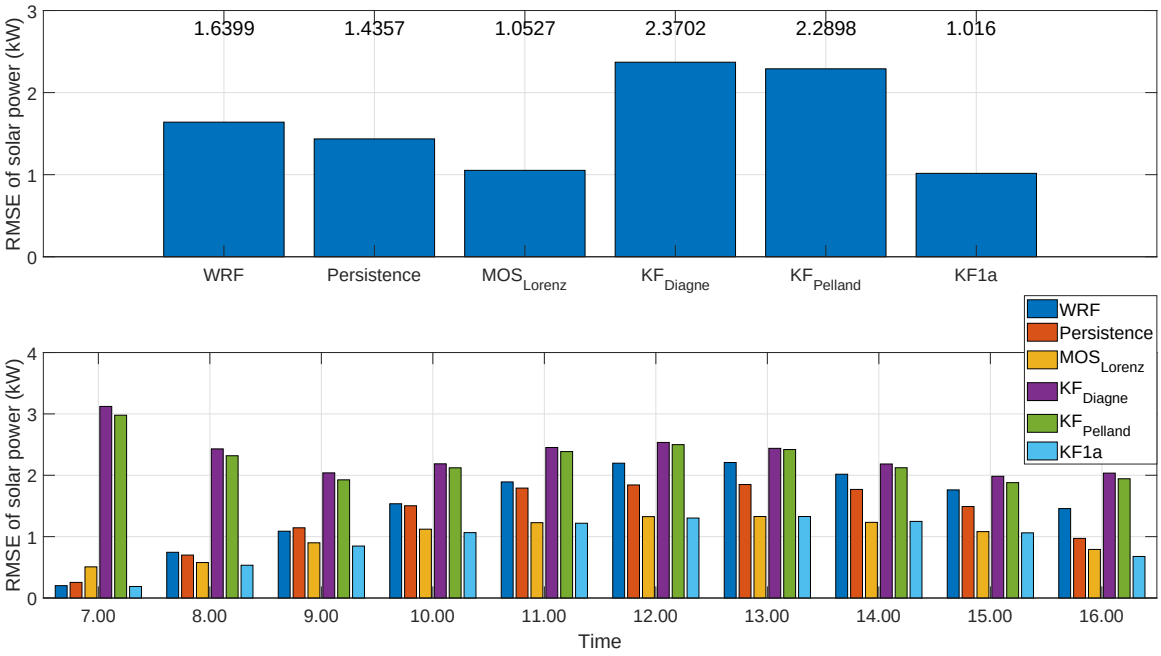
9.5 Solar irradiance to PV power conversion

The predicted solar irradiance is converted into the predicted electrical power from PV cells by PV power conversion. There are two PV sites of interest that are used to evaluate the predicted solar power. The installed capacity of the first PV site is 8 kW and the installed capacity of the second PV site is 15 kW. Both PV sites are installed at top of Electrical Engineering building, Faculty of Engineering, Chulalongkorn University. From the PV conversion equation (6.1), the regression coefficients of the conversion model for converting the solar irradiance into solar power of 8 kW and 15 kW PV systems are shown in Table 9.7 respectively.

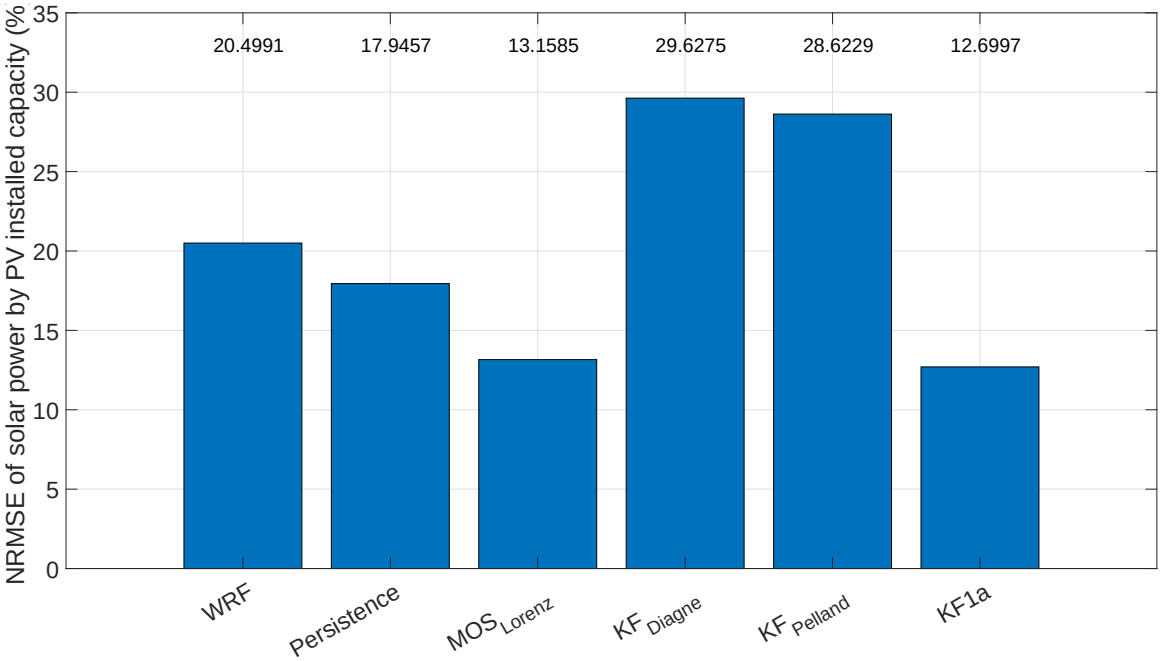
Table 9.7: The regression coefficients of PV conversion models for 8 kW and 15 kW PV systems.

PV site	Regression coefficient of the predictor		
	$\hat{I}_{\text{wrf}}$	$\widehat{\text{T}}_{\text{wrf}}$	$\hat{I}_{\text{wrf}}\widehat{\text{T}}_{\text{wrf}}$
8 kW	9.4451	−0.0031	−0.0929
15 kW	17.9051	−0.0226	−1.4267

The results of predicted solar power of 8 kW and 15 kW PV systems are shown in Figures 9.10 and 9.11 respectively.

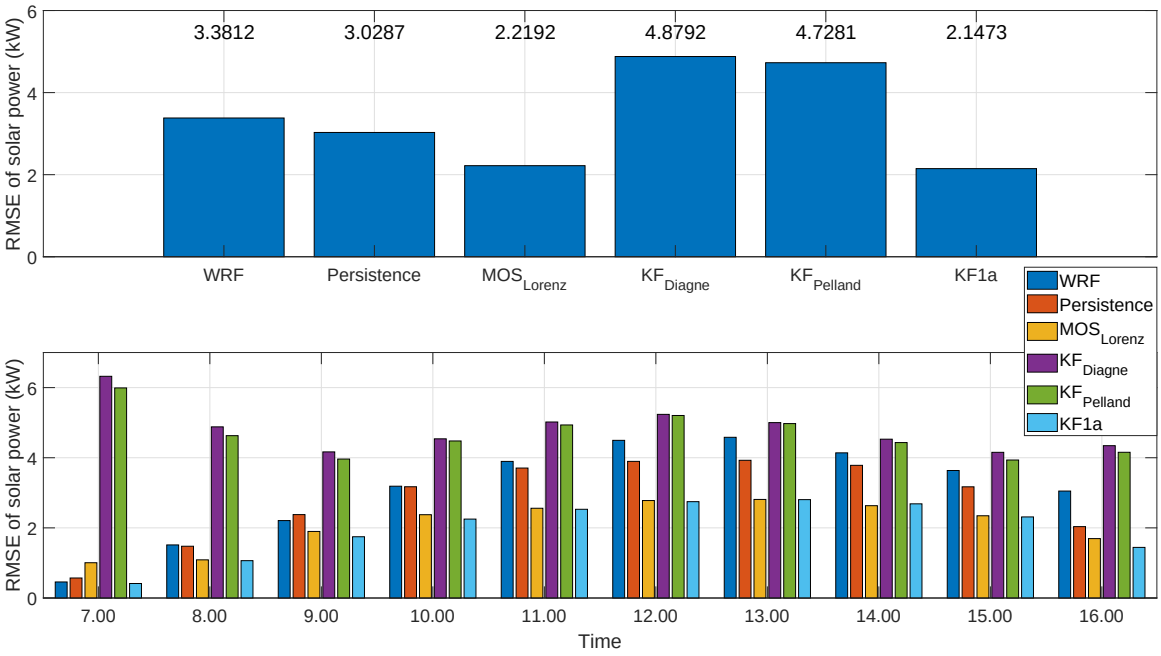


(a) RMSE and RMSE in specific hour.

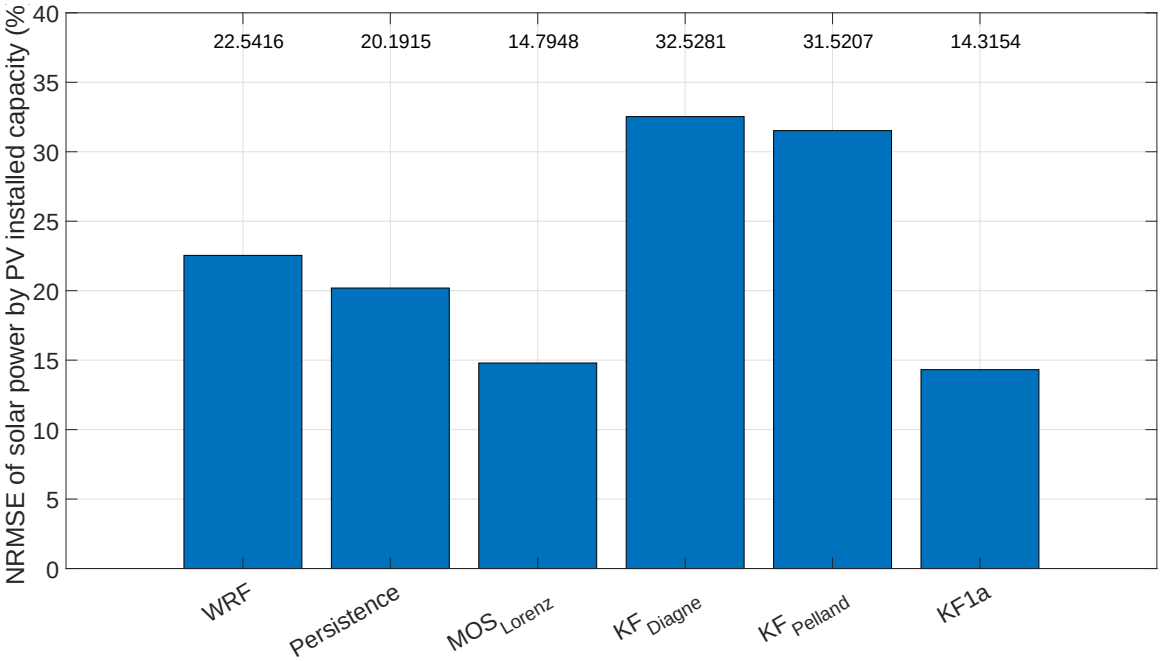


(b) NRMSE (normalized by install capacity).

Figure 9.10: RMSE and NRMSE of the predicted solar power from the proposed MOS+KF model and the models from previous work for **8 kW** PV system.



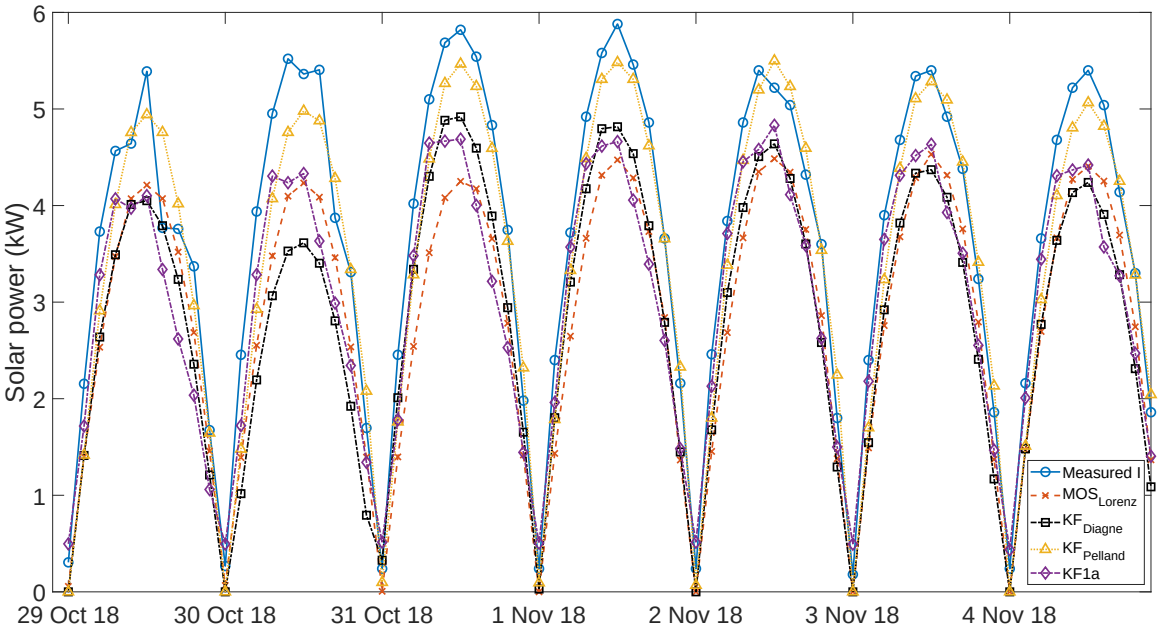
(a) RMSE and RMSE in specific hour.



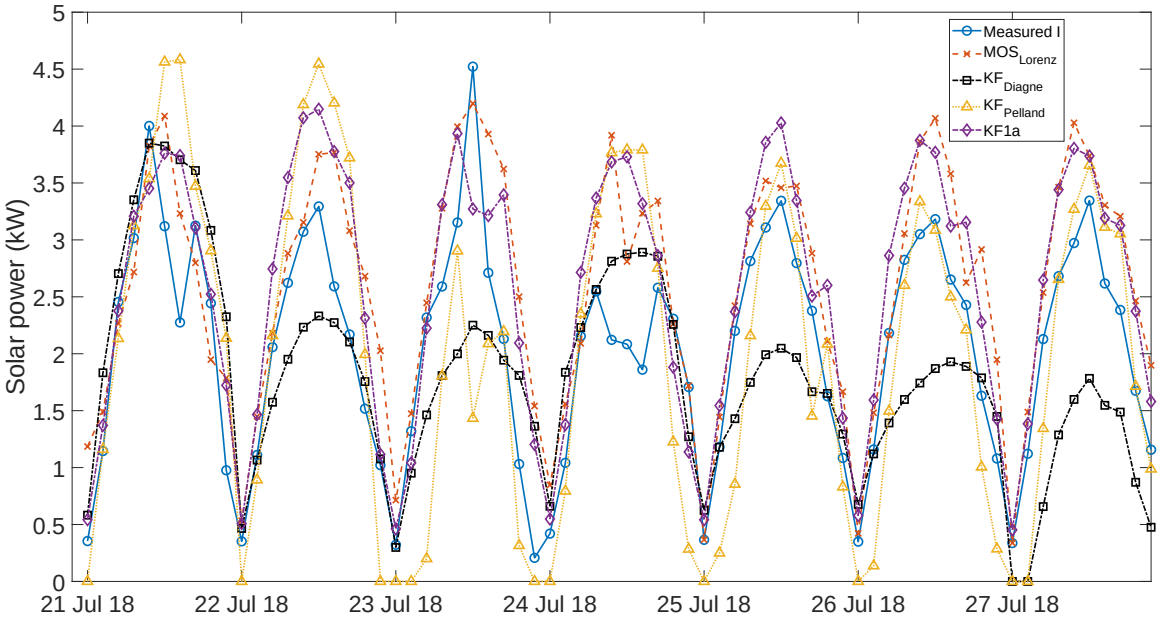
(b) NRMSE (normalized by install capacity).

Figure 9.11: RMSE and NRMSE of the predicted solar power from the proposed MOS+KF model and the models from previous work for **15 kW** PV system.

From the result, daily-step MOS+KF1a outperforms the other models in both 8 kW and 15 kW PV systems. we can see that the trend of RMSE of the predicted solar power is the same as RMSE of the predicted solar irradiance. According to Table 9.7, we can see that the weight of the predicted solar irradiance in PV conversion model fitting is the highest. Therefore, we can conclude that the predicted solar power is linearly dependent to the predicted solar irradiance. The example plots of the predicted solar power of 8 kW and 15 kW PV systems are shown in Figures 9.12 and 9.13.

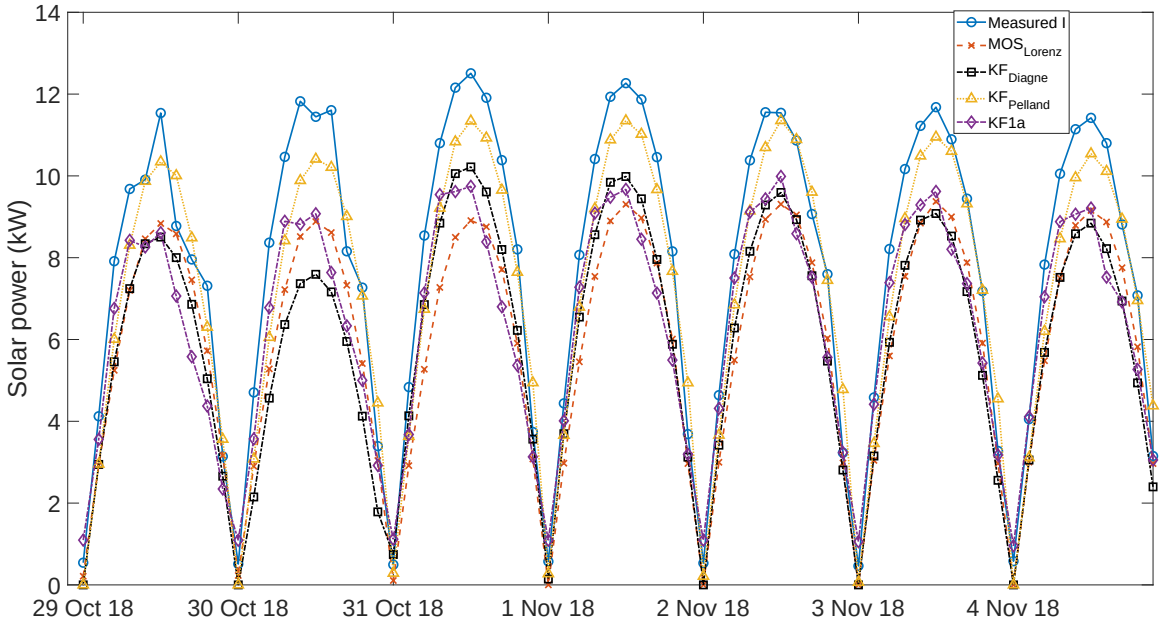


(a) Example plots of the predicted solar power under good weather conditions for 8 kW PV system.

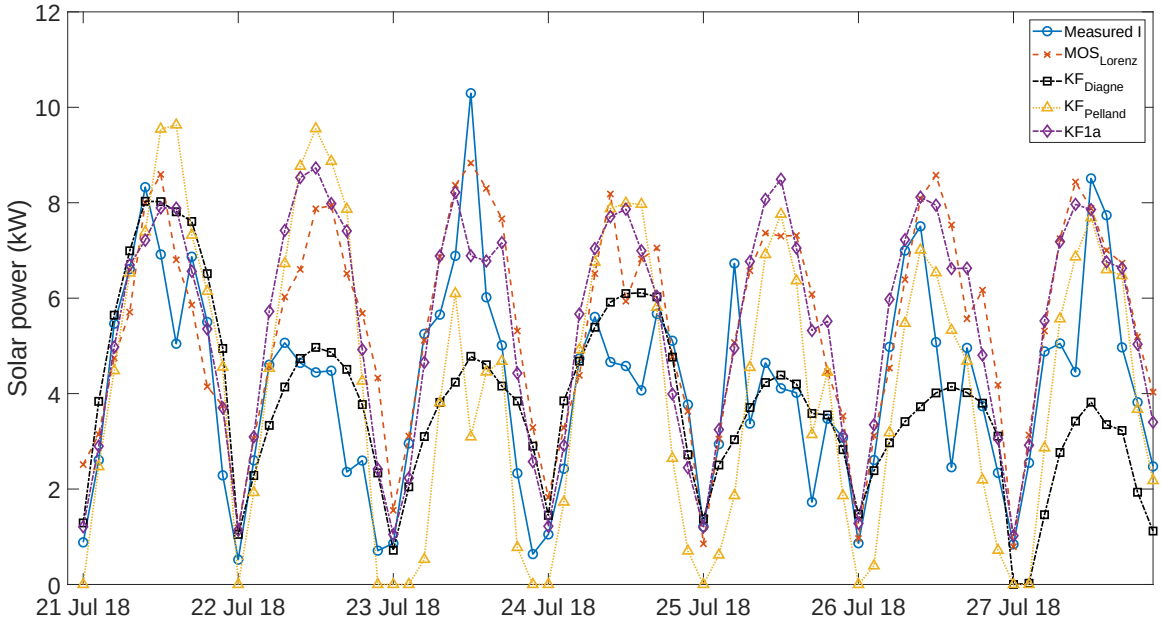


(b) Example plots of the predicted solar power under bad weather conditions for 8 kW PV system.

Figure 9.12: Example plots of the predicted solar power from the proposed MOS+KF model and the models from previous work for **8 kW** PV system.



(a) Example plots of the predicted solar power under good weather conditions for 15 kW PV system.



(b) Example plots of the predicted solar power under bad weather conditions for 15 kW PV system.

Figure 9.13: Example plots of the predicted solar power from the proposed MOS+KF model and the models from previous work for **15 kW** PV system.

A comparison of the performance of the predicted PV output between this work and the other works is not easy. The formula for calculating some performance indices such as NRMSE and MAPE in other works are not the same. Moreover, there are various installed capacity of PV site, therefore, some performance indices of the other works such as RMSE, MBE, and MAE cannot be compared. To deal with this problem the performance index with normalized term should be used. The performance index that we prefer to use in order to compare our work to the previous works is normalized RMSE (NRMSE) by installed capacity of the PV site. We summarize the results and the details of the predicted solar power for one day ahead from previous works in Table 9.8. From Figure 9.10, NRMSE by installed capacity

Table 9.8: Summary of the predicted solar power one day ahead from previous works.

Reference	**Forecasting model**	**Condition**	**NRMSE** (%)
[22]	ANN	-	23.99
[23]	ANN	Clear	12.5
		Partially cloudy	24
		Cloudy	36.9
[24]	Wavelet+ Fuzzy ARTMAP + Firefly	Winter	12.51
		Spring	13.13
		Summer	12.11
		Fall	12.82
[25]	Wavelet+ANN	Clear	7.193
		Cloudy	16.817
		Overcast	17.607
		Rainy	19.663

of 8 kW PV system of our daily-step MOS+KF1a is about 13 %. Moreover, our daily-step MOS+KF1a is evaluated with unclassify data. However, the results from the previous work in Table 9.8 usually evaluate with the classified data. Therefore, there is only work from [22] that can be compared with our daily-step MOS+KF1a and our model is better than the model from [22] about 11 % of NRMSE.

CHAPTER X

CONCLUSION

This work aims to predict the solar power from the PV cells one day in advance by 13.00 hrs daily to serve Short-term Operation Planning Section (SOPS) of EGAT. The predicted solar power of the next day can help SOPS for efficient dispatching. However, the solar power mostly depends on the solar irradiance and temperature of the PV panels as described in Section 3.5. Therefore, the predicted solar irradiance and temperature are needed. NWP models can provide forecasts of both variables with moderate accuracy up to several days in advance. Therefore, we use WRF model which is a kind of the regional NWP models to predict the solar irradiance and the weather variables. However, we found that the predicted solar irradiance from WRF model usually overestimated the actual solar irradiance. We then apply the post-processing method to reduce the prediction error of the predicted solar irradiance from WRF model.

WRF model provides the predicted weather variables according to longitude and latitude coordinates. Therefore, we used spatial averaging method to reduce the forecasting error of WRF model before we use the predicted weather variables in the forecasting procedures. From Section 9.2, the results showed that the spatial averaging of the predicted solar irradiance of 7×7 km^2 was the most suitable spatial area compared to 3×3 km^2 and 5×5 km^2. The spatial area of 7×7 km^2 reduced the RMSE of the predicted solar irradiance from WRF model without spatial averaging by 14 W/m^2. Moreover, the relevant variables to solar irradiance can be used as the input of MOS model to further reduce the prediction error of the predicted solar irradiance from WRF model with spatial averaging.

MOS is a regression model and uses the relationships between solar irradiance and relevant variables to solar irradiance in order to explain the characteristics of solar irradiance. Therefore, the first contribution of thesis is to define the important variables to solar irradiance for using in MOS model. The important variables to solar irradiance can be decided by statistical methods. The statistical methods that we use in this work were partial correlation, stepwise regression, and subset regression. As the results in Section 9.3, we concluded that the important variables to solar irradiance in Thailand were $\hat{I}_{\text{wrf}}(t)$, $\widehat{\text{RH}}_{\text{wrf}}(t)$, $\hat{\text{T}}_{\text{wrf}}(t)$, and $\cos\theta(t)$. From Section 9.4, we found that the proposed MOS could be used to further reduce the RMSE of the predicted solar irradiance from WRF model with spatial averaging by 90 W/m^2 in both hourly-step and daily-step MOS. However, the regression coefficients of MOS are fixed then it may not suitable to use all the time. We then apply Kalman filter to MOS in order to update the regression coefficients over time. As the forecasting time constraint, some data are not available to use in update parameters process of Kalman filter. We then modify the equations of Kalman filter to take into account the missing measurement data and this is the second contribution of this thesis. We found that MOS+KF could further reduce the RMSE of MOS about 2 to 3 W/m^2 in both hourly-step and daily-step models. Moreover, daily-step MOS+KF1a could further reduce the RMSE of hourly-step MOS+KF3

by 4 W/m^2. In conclusion, daily-step MOS+KF1a outperforms the other models because of high complexity of the model and the adaptive regression coefficients from Kalman filter. Besides, daily-step MOS+KF1a also outperforms the models from previous work. After that, the predicted solar irradiance is converted into the predicted solar power by PV conversion model which is a regression model. From Section 9.5, we found that the accuracy of the predicted solar power was related to the accuracy of the predicted solar irradiance since the regression coefficient of the predicted solar irradiance in PV conversion model fitting was the highest magnitude. Therefore, the solar irradiance forecasting is the most important process that should provides accurate predicted solar irradiance values.

The outcomes of this work is the predicted solar power. From the results in Section 9.5, the NRMSE (normalized by capacity of the PV site) of the predicted solar power from daily-step MOS+KF1a is about 13 to 14 %. Nowadays, the install capacity of PV in Thailand is about 3,000 MW. Therefore, if we use the predicted solar irradiance from WRF model to predict the solar power, then the operators should prepare the spinning reserved of 600 MW in daytime to deal with the fluctuation of the PV generation. Besides, if we apply MOS+KF1a to predict the solar power, then the operators just prepare the spinning reserved of 360 MW. Normally, the National Control Center (NCC) always prepares the spinning reserved in the range of 700 to 1500 MW for preventing blackouts when the power plant with the biggest installed capacity tripped. Therefore, NCC can still operate in normal conditions for both the predicted solar power from WRF model and MOS+KF1a. However, if we assume the cost of spinning reserved is 0.2 baht per kWh, then our model can further reduce the cost of spinning reserved of the predicted solar power from WRF model by 48,000 baht per hour. This cost will reflect the operating cost of power plant and also the electricity bill.

A limitation of this work is that KF models need the measured solar irradiance to provide the forecast values. Moreover, MOS+KF models still cannot provide a high accuracy of the predicted solar irradiance more than one day in advance due to lack of measured data of day $d+1$ for updating the parameters. Therefore, MOS models may be the most suitable to provide the predicted solar irradiance more than one day in advance because MOS requires only the outputs of WRF model. In this work, we just use the available weather variables in the forecasting models, so the model still can be improved by adding more important variables to solar irradiance to the forecasting models. Moreover, our proposed models are linear models and solar irradiance characteristics may be non-linear, therefore, the machine learning methods may further reduce the forecasting error of the linear models.

REFERENCES

[1] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Systems with Applications*, vol. 37, no. 1, pp. 479–489, 2010.

[2] H. Glahn and D. Lowry, "The Use of model output statistics (MOS) in objective weather forecasting," *Journal of Applied Meteorology*, vol. 11, no. 8, pp. 1203–1211, 1972.

[3] R. Inman, H. Pedro, and C. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, 2013.

[4] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[5] S. Sobri, S. Koohi-Kamali, and N. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, vol. 156, pp. 459–497, 2018.

[6] W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, M. Duda, X. Huang, W. Wang, and J. Powers, "A Description of the advanced research WRF version 3," tech. rep., DTIC Document, 2008.

[7] R. Perez, M. Beauharnois, K. Hemker, S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, and G. V. Knowe, "Evaluation of numerical weather prediction solar irradiance forecasts in the US," in *Proceedings of the American Solar Energy Society Annual Conference*, pp. 1–8, 2011.

[8] L. Lazić, G. Pejanović, M. Živković, and L. Ilić, "Improved wind forecasts for wind power generation using the Eta model and MOS (Model Output Statistics) method," *Energy*, vol. 73, pp. 567–574, 2014.

[9] E. Lorenz, J. Hurka, D. Heinemann, and H. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, 2009.

[10] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States," *Solar Energy*, vol. 85, no. 5, pp. 967–977, 2011.

[11] S. Pelland, G. Galanis, and G. Kallos, "Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model," *Progress in photovoltaics: Research and Applications*, vol. 21, no. 3, pp. 284–296, 2013.

[12] M. Diagne, M. David, J. Boland, N. Schmutz, and P. Lauret, "Post-processing of solar irradiance forecasts from WRF model at Reunion Island," *Solar Energy*, vol. 105, pp. 99–108, 2014.

[13] V. Badescu, *Modeling solar radiation at the earth's surface*. Springer, 2014.

[14] J. Duffie and W. Beckman, *Solar Engineering of Thermal Processesl, Fourth Edition*. Wiley, 2013.

[15] P. Ineichen and R. Perez, "A New airmass independent formulation for the Linke turbidity coefficient," *Solar Energy*, vol. 73, no. 3, pp. 151–157, 2002.

[16] R. Verzijlbergh, P. Heijnen, S. de Roode, A. Los, and H. Jonker, "Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications," *Solar Energy*, vol. 118, pp. 634–645, 2015.

[17] P. Bacher, H. Madsen, and H. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[18] C. Paoli, C. Voyant, M. Muselli, and M. Nivet, "Forecasting of preprocessed daily solar radiation time series using neural networks," *Solar Energy*, vol. 84, no. 12, pp. 2146–2160, 2010.

[19] C. Voyant, G. Notton, S. Kalogirou, M. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[20] R. Azimi, M. Ghayekhloo, and M. Ghofrani, "A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting," *Energy Conversion and Management*, vol. 118, pp. 331–344, 2016.

[21] H. Pedro and C. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2012.

[22] M. D. Giorgi, P. Congedo, and M. Malvoni, "Photovoltaic power forecasting using statistical methods: impact of weather data," *IET Science, Measurement & Technology*, vol. 8, no. 3, pp. 90–97, 2014.

[23] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Mathematics and Computers in Simulation*, vol. 131, pp. 88–100, 2017.

[24] A. Haque, M. Nehrir, and P. Mandal, "Solar PV power generation forecast using a hybrid intelligent approach," in *2013 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2013.

[25] H. Zhu, X. Li, Q. Sun, L. Nie, J. Yao, and G. Zhao, "A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks," *Energies*, vol. 9, no. 1, p. 11, 2016.

[26] M. Paulescu, E. Paulescu, P. Gravila, and V. Badescu, *Weather modeling and forecasting of PV systems operation*. Springer Science & Business Media, 2012.

[27] M. Villalva, J. Gazoli, and E. R. Filho, "Comprehensive approach to modeling and simulation of photovoltaic arrays," *IEEE Transactions on Power Electronics*, vol. 24, no. 5, pp. 1198–1208, 2009.

[28] A. Chouder, S. Silvestre, N. Sadaoui, and L. Rahmani, "Modeling and simulation of a grid connected PV system based on the evaluation of main PV module parameters," *Simulation Modelling Practice and Theory*, vol. 20, no. 1, pp. 46–58, 2012.

[29] C. Qi and Z. Ming, "Photovoltaic module Simulink model for a stand-alone PV system," *Physics Procedia*, vol. 24, pp. 94–100, 2012.

[30] B. Babu and S. Gurjar, "A novel simplified two-diode model of photovoltaic (PV) module," *IEEE Journal of Photovoltaics*, vol. 4, no. 4, pp. 1156–1161, 2014.

[31] A. Dolara, S. Leva, and G. Manzolini, "Comparison of different physical models for PV power output prediction," *Solar Energy*, vol. 119, pp. 83–99, 2015.

[32] M. Slimani, M. Amirat, I. Kurucz, S. Bahria, A. Hamidat, and W. Chaouch, "A detailed thermal-electrical model of three photovoltaic/thermal (PV/T) hybrid air collectors and photovoltaic (PV) module: Comparative study under Algiers climatic conditions," *Energy conversion and management*, vol. 133, pp. 458–476, 2017.

[33] D. Sera, R. Teodorescu, and P. Rodriguez, "PV panel model based on datasheet values," in *Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on*, pp. 2392–2396, IEEE, 2007.

[34] N. Tong and W. Pora, "Parameter extraction for a solar cell model with three measured points only," in *International Conference on Electronics, Information, and Communications (ICEIC)*, pp. 1–4, IEEE, 2016.

[35] V. Sangsawang and S. Chaitusaney, "Modeling of photovoltaic module from commercial specification in datasheet," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on*, pp. 1–4, IEEE, 2012.

[36] W. D. Soto, S. Klein, and W. Beckman, "Improvement and validation of a model for photovoltaic array performance," *Solar Energy*, vol. 80, no. 1, pp. 78–88, 2006.

[37] C. Sah, *Fundamentals of solid state electronics*. World Scientific Publishing Company, 1991.

[38] H. Wei, J. Cong, X. Lingyun, and S. Deyun, "Extracting solar cell model parameters based on chaos particle swarm algorithm," in *International Conference on Electric Information and Control Engineering (ICEICE)*, pp. 398–402, IEEE, 2011.

[39] O. Gergaud, B. Multon, and H. Ahmed, "Analysis and experimental validation of various photovoltaic system models," in *Electrimacs*, p. 6p, 2002.

[40] T. Hiyama and K. Kitabayashi, "Neural network based estimation of maximum power generation from PV module using environmental information," *IEEE Transactions on Energy Conversion*, vol. 12, no. 3, pp. 241–247, 1997.

[41] M. Balzani and A. Reatti, "Neural network based model of a PV array for the optimum performance of PV system," in *Research in Microelectronics and Electronics, 2005 PhD*, vol. 2, pp. 123–126, IEEE, 2005.

[42] E. Karatepe, M. Boztepe, and M. Colak, "Neural network based solar cell model," *Energy Conversion and Management*, vol. 47, no. 9-10, pp. 1159–1178, 2006.

[43] J. Klomklao, W. Thangjitwisuth, W. Preechakul, D. Banjerdpongchai, and P. Rachdawong, "Performance assessment of solar energy generation using artificial intelligence," 2018.

[44] M. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in PV output power forecast," *Solar Energy*, vol. 136, pp. 125–144, 2016.

[45] F. Barbieri, S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 242–263, 2017.

[46] U. Das, K. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Idris, W. V. Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.

[47] A. Metcalfe and P. Cowpertwait, *Introductory time series with R.* Springer, 2009.

[48] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, no. 3, pp. 342–349, 2009.

[49] P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi, "A benchmarking of machine learning techniques for solar radiation forecasting in an insular context," *Solar Energy*, vol. 112, pp. 446–457, 2015.

[50] Y. Kashyap, A. Bansal, and A. Sao, "Solar radiation forecasting with multiple parameters neural networks," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 825–835, 2015.

[51] R. Kumar, R. Aggarwal, and J. Sharma, "Comparison of regression and artificial neural network models for estimation of global solar radiations," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1294–1299, 2015.

[52] M. Bou-Rabee, S. Sulaiman, M. Saleh, and S. Marafi, "Using artificial neural networks to estimate solar radiation in Kuwait," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 434–438, 2017.

[53] 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, no. 5, pp. 807–821, 2010.

[54] A. Mellit, S. Sağlam, and S. Kalogirou, "Artificial neural network-based model for estimating the produced power of a photovoltaic module," *Renewable Energy*, vol. 60, pp. 71–78, 2013.

[55] C. Voyant, M. Muselli, C. Paoli, and M. Nivet, "Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation," *Energy*, vol. 36, no. 1, pp. 348–359, 2011.

[56] G. Graditi, S. Ferlito, G. Adinolfi, G. Tina, and C. Ventura, "Energy yield estimation of thin-film photovoltaic plants by using physical approach and artificial neural networks," *Solar Energy*, vol. 130, pp. 232–243, 2016.

[57] M. Rana, I. Koprinska, and V. Agelidis, "Forecasting solar power generated by grid connected PV systems using ensembles of neural networks," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, pp. 1–8, IEEE, 2015.

[58] Z. Li, S. Rahman, R. Vega, and B. Dong, "A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting," *Energies*, vol. 9, no. 1, p. 55, 2016.

[59] T. Teo, T. Logenthiran, W. Woo, and K. Abidi, "Forecasting of photovoltaic power using regularized ensemble Extreme Learning Machine," in *Region 10 Conference (TENCON), 2016 IEEE*, pp. 455–458, IEEE, 2016.

[60] Y. Chen, P. Luh, C. Guan, Y. Zhao, L. Michel, M. Coolbeth, P. Friedland, and S. Rourke, "Short-term load forecasting: similar day-based wavelet neural networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 322–330, 2010.

[61] M. Hossain, S. Mekhilef, M. Danesh, L. Olatomiwa, and S. Shamshirband, "Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems," *Journal of Cleaner Production*, vol. 167, pp. 395–405, 2017.

[62] Z. Guo, W. Zhao, H. Lu, and J. Wang, "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model," *Renewable Energy*, vol. 37, no. 1, pp. 241–249, 2012.

[63] L. Lyu, M. Kantardzic, and E. Arabmakki, "Solar irradiance forecasting by using wavelet based denoising," in *Computational Intelligence for Engineering Solutions (CIES), 2014 IEEE Symposium on*, pp. 110–116, IEEE, 2014.

[64] J. Cao and X. Lin, "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Conversion and Management*, vol. 49, no. 6, pp. 1396–1406, 2008.

[65] M. Ghayekhloo, M. Ghofrani, M. Menhaj, and R. Azimi, "A novel clustering approach for short-term solar radiation forecasting," *Solar Energy*, vol. 122, pp. 1371–1383, 2015.

[66] A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "Hybrid model analysis and validation for PV energy production forecasting," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, pp. 1957–1962, IEEE, 2014.

[67] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "A physical hybrid artificial neural network for short term forecasting of PV plant power output," *Energies*, vol. 8, no. 2, pp. 1138–1153, 2015.

[68] E. Ogliari, A. Dolara, G. Manzolini, and S. Leva, "Physical and hybrid methods comparison for the day ahead PV output power forecast," *Renewable Energy*, vol. 113, pp. 11–21, 2017.

[69] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, and W. Liu, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Conversion and Management*, vol. 153, pp. 409–422, 2017.

[70] J. Chen, G. Li, and S. Wu, "Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration," *Energy Conversion and Management*, vol. 75, pp. 311–318, 2013.

[71] K. Mohammadi, S. Shamshirband, C. Tong, M. Arif, D. Petković, and S. Ch, "A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation," *Energy Conversion and Management*, vol. 92, pp. 162–171, 2015.

[72] J. Shi, W. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.

[73] M. Abuella and B. Chowdhury, "Solar power forecasting using support vector regression," *arXiv preprint arXiv:1703.09851*, 2017.

[74] H. Yang, C. Huang, Y. Huang, Y. Pai, *et al.*, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, 2014.

[75] C. Ko and C. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," *Energy*, vol. 49, pp. 413–422, 2013.

[76] L. Ma, K. Khorasani, Y. Xiao, N. Yorino, and F. Wahyudi, "Hourly solar radiation forecasting using LS-based Volterra filters," in *Proceedings of the 2017 2nd International Conference on Power and Renewable Energy (ICPRE)*, pp. 930–936, IEEE, 2017.

[77] J. Trapero, N. Kourentzes, and A. Martin, "Short-term solar irradiation forecasting based on dynamic harmonic regression," *Energy*, vol. 84, pp. 289–295, 2015.

[78] G. Wang, Y. Su, and L. Shu, "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models," *Renewable Energy*, vol. 96, pp. 469–478, 2016.

[79] Y. Li, Y. He, Y. Su, and L. Shu, "Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines," *Applied Energy*, vol. 180, pp. 392–401, 2016.

[80] O. Maatallah, A. Achuthan, K. Janoyan, and P. Marzocca, "Recursive wind speed forecasting based on Hammerstein Auto-Regressive model," *Applied Energy*, vol. 145, pp. 191–197, 2015.

[81] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013.

[82] E. Lorenz, J. Remund, S. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. Ruiz-Arias, V. Fanego, L. Ramirez, M. Romeo, *et al.*, "Benchmarking of different approaches to forecast solar irradiance," in *Proceedings of the 24th European Photovoltaic Solar Energy Conference*, pp. 4199–4208, Hamburg, Germany, 2009.

[83] E. Lorenz, J. Kühnert, D. Heinemann, K. Nielsen, J. Remund, and S. Müller, "Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions," *Progress in Photovoltaics: Research and Applications*, vol. 24, no. 12, pp. 1626–1640, 2016.

[84] P. Lauret, M. Diagne, and M. David, "A neural network post-processing approach to improving NWP solar radiation forecasts," *Energy Procedia*, vol. 57, pp. 1044–1052, 2014.

[85] F. Lima, F. Martins, E. Pereira, E. Lorenz, and D. Heinemann, "Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks," *Renewable Energy*, vol. 87, pp. 807–818, 2016.

[86] C. Cornaro, M. Pierro, and F. Bucci, "Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast," *Solar Energy*, vol. 111, pp. 297–312, 2015.

[87] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Solar Energy*, vol. 85, no. 11, pp. 2856–2870, 2011.

[88] C. Cornaro, F. Bucci, M. Pierro, F. D. Frate, S. Peronaci, and A. Taravat, "Twenty-four hour solar irradiance forecast based on neural networks and numerical weather prediction," *Journal of Solar Energy Engineering*, vol. 137, no. 3, p. 031011, 2015.

[89] E. Kardakos, M. Alexiadis, S. Vagropoulos, C. Simoglou, P. Biskas, and A. Bakirtzis, "Application of time series and artificial neural network models in short-term forecasting of PV power generation," in *Power Engineering Conference (UPEC), 2013 48th International Universities'*, pp. 1–6, IEEE, 2013.

[90] S. Chen, H. Gooi, and M. Wang, "Solar radiation forecast based on fuzzy logic and neural networks," *Renewable Energy*, vol. 60, pp. 195–201, 2013.

[91] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. D. Monache, "Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble," *Renewable Energy*, vol. 108, pp. 274–286, 2017.

[92] D. Quan, E. Ogliari, F. Grimaccia, S. Leva, and M. Mussetta, "Hybrid model for hourly forecast of photovoltaic and wind power," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pp. 1–6, IEEE, 2013.

[93] Y. Wu, C. Chen, and H. A. Rahman, "A novel hybrid model for short-term forecasting in pv power generation," *International Journal of Photoenergy*, vol. 2014, 2014.

[94] A. Yona, T. Senjyu, T. Funabashi, and C. Kim, "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 527–533, 2013.

[95] L. Massidda and M. Marrocu, "Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany," *Solar Energy*, vol. 146, pp. 141–149, 2017.

[96] L. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces, M. Mendoza-Villena, E. Garcia-Garrido, P. Lara-Santillan, E. Zorzano-Alba, and P. Zorzano-Santamaria, "Short-term power forecasting system for photovoltaic plants," *Renewable Energy*, vol. 44, pp. 311–317, 2012.

[97] M. Kratzenberg, S. Colle, and H. Beyer, "Solar radiation prediction based on the combination of a numerical weather prediction model and a time series prediction model," in *Proceedings of the EUROSUN 2008, 1st International Congress on Heating, Cooling, and Buildings*, 2008.

[98] J. Wu and C. Chan, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.

[99] M. Bouzerdoum, A. Mellit, and A. Pavan, "A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant," *Solar Energy*, vol. 98, pp. 226–235, 2013.

[100] H. do Nascimento Camelo, P. S. Lucio, J. V. L. Junior, D. von Glehn dos Santos, and P. C. M. de Carvalho, "Innovative hybrid modeling of wind speed prediction involving time-series models and artificial neural networks," *Atmosphere*, vol. 9, no. 2, p. 77, 2018.

[101] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[102] P. Pai and C. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[103] C. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," *Applied Mathematics Letters*, vol. 22, no. 9, pp. 1467–1470, 2009.

[104] M. Khashei, M. Bijari, and G. Ardali, "Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs)," *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 37–45, 2012.

[105] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Applied Soft Computing*, vol. 11, no. 2, pp. 2664–2675, 2011.

[106] B. Amrouche and X. L. Pivert, "Artificial neural network based daily local forecasting for global solar radiation," *Applied Energy*, vol. 130, pp. 333–341, 2014.

[107] F. Gutierrez-Corea, M. Manso-Callejo, M. Moreno-Regidor, and M. Manrique-Sancho, "Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations," *Solar Energy*, vol. 134, pp. 119–131, 2016.

[108] D. Heinemann, E. Lorenz, and M. Girodo, "Forecasting of solar radiation," *Solar energy resource management for electricity generation from local level to global scale. Nova Science Publishers, New York*, 2006.

[109] M. Pierro, F. Bucci, C. Cornaro, E. Maggioni, A. Perotto, M. Pravettoni, and F. Spada, "Model output statistics cascade to improve day ahead solar irradiance forecast," *Solar Energy*, vol. 117, pp. 99–113, 2015.

[110] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.

[111] B. Anderson and J. Moore, *Optimal Filtering*. Dover Publications, 2005.

[112] G. Galanis and M. Anadranistakis, "A one-dimensional Kalman filter for the correction of near surface temperature forecasts," *Meteorological Applications*, vol. 9, no. 4, pp. 437–441, 2002.

[113] G. Galanis, P. Louka, P. Katsafados, I. Pytharoulis, and G. Kallos, "Applications of Kalman filters based on non-linear functions to numerical weather predictions," *Annales Geophysicae*, vol. 24, no. 10, pp. 2451–2460, 2006.

[114] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 12, pp. 2348–2362, 2008.

[115] F. Cassola and M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output," *Applied Energy*, vol. 99, pp. 154–166, 2012.

[116] C. Lynch, M. O'Mahonyl, and R. Guinee, "A novel 24 Kalman filter bank estimator for solar irradiance prediction for PV power generation," in *Photovoltaic Specialist Conference (PVSC), 2015 IEEE 42nd*, pp. 1–7, IEEE, 2015.

[117] J. Songsiri. private communication, September 28, 2018.

[118] T. Cipra and R. Romera, "Kalman filter with outliers and missing observations," *Test*, vol. 6, no. 2, pp. 379–395, 1997.

[119] K. Myers and B. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Transactions on Automatic Control*, vol. 21, no. 4, pp. 520–523, 1976.

APPENDIX

APPENDIX A
ESTIMATION OF LINKE TURBIDITY

According to the Ineichen and Kasten clear sky models, $T_{\mathrm{L}}(t)$ and $I_{\mathrm{clr}}(t)$ are the variables that we have no data. Therefore, we assume $I(t)$ of the most clear sky day as $I_{\mathrm{clr}}(t)$ and we estimate $T_{\mathrm{L}}(t)$ from Ineichen and Kasten clear sky models using least squares method. The processes for estimating $T_{\mathrm{L}}(t)$ are described as follows.

Ineichen clear sky model (3.12):

$$\begin{aligned} I_{\mathrm{clr}}(t) &= a_1 I_{\mathrm{sc}} \cos\theta(t) e^{-a_2 \mathrm{AM}(t)(f_1 + f_2(T_{\mathrm{L}}(t)-1))} \\ \log \frac{I_{\mathrm{clr}}(t)}{a_1 I_{\mathrm{sc}} \cos\theta(t)} &= -a_2 \mathrm{AM}(t)(f_1 + f_2 T_{\mathrm{L}}(t) - f_2)) \\ a_2 \mathrm{AM}(t) f_2 T_{\mathrm{L}}(t) &= -\log \frac{I_{\mathrm{clr}}(t)}{a_1 I_{\mathrm{sc}} \cos\theta(t)} - a_2 \mathrm{AM}(t)(f_1 - f_2) \end{aligned}$$

where $f_1 = e^{-35/8000}$, $f_2 = e^{-35/1250}$, $a_1 = 35{\times}5.09{\times}10^{-5}{+}0.868$, $a_2 = 35{\times}3.92{\times}10^{-5}{+}0.0387$ then perform least squares to estimate $T_{\mathrm{L}}(t)$.

Kasten clear sky model (3.11):

$$\begin{aligned} I_{\mathrm{clr}}(t) &= 0.84 I_{\mathrm{sc}} \cos\theta(t) e^{-0.027 \mathrm{AM}(t)(f_1 + f_2(T_{\mathrm{L}}(t)-1))} \\ \log \frac{I_{\mathrm{clr}}(t)}{0.84 I_{\mathrm{sc}} \cos\theta(t)} &= -0.027 \mathrm{AM}(t)(f_1 + f_2 T_{\mathrm{L}}(t) - f_2) \\ 0.027 \mathrm{AM}(t) f_2 T_{\mathrm{L}}(t) &= -\log \frac{I_{\mathrm{clr}}(t)}{0.84 I_{\mathrm{sc}} \cos\theta(t)} - 0.027 \mathrm{AM}(t)(f_1 - f_2) \end{aligned}$$

We then perform least squares to estimate $T_{\mathrm{L}}(t)$. The measured solar irradiance data that we assume to be $I_{\mathrm{clr}}(t)$ are the data between 9 to 15 Apr 17 from 7.00 to 16.00 hrs with the time step of 3 minutes as shown in Figure A.1. The measured solar irradiance after 16.00 hrs is not considered because the sunlight have been blocked by the building.

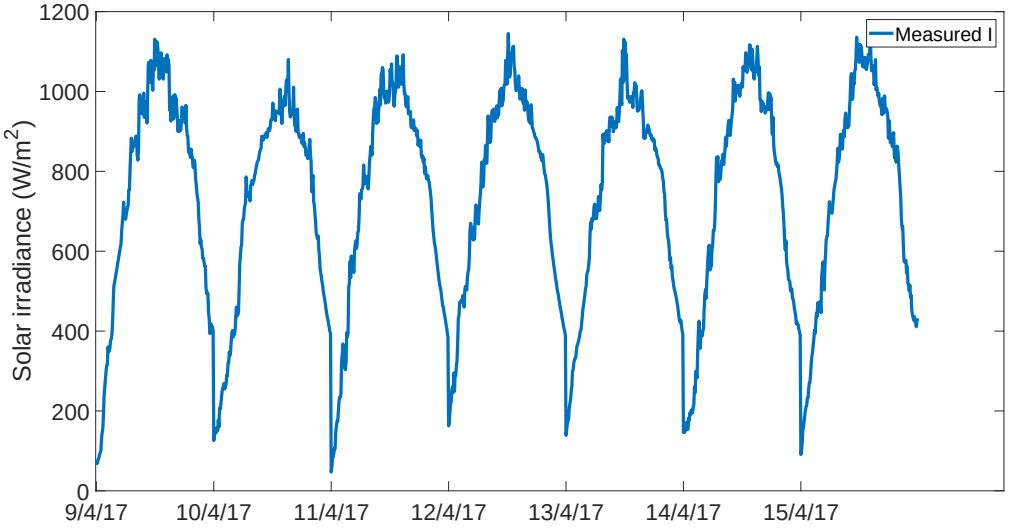


Figure A.1: Measured solar irradiance from CUBEMS between 9 to 15 Apr 2017 (7.00 to 16.00 hrs).

The results of estimated $T_{\mathrm{L}}(t)$ are shown in Table A.1

Table A.1: Estimation of $T_{\mathrm{L}}(t)$

Data	Ineichen clear sky model	Kasten clear sky model
9-15 Apr 2017	4.8597	6.4993

The solar irradiance data that we use to validate the estimated $T_{\mathrm{L}}(t)$ value are data between 28 to 30 Apr 17. In Figure A.2, we plot the measured solar irradiance and the solar irradiance under clear sky conditions between 28 to 30 Apr 17 using Ineichen clear sky model with $T_{\mathrm{L}}(t) = 4.86$ and Kasten clear sky model with $T_{\mathrm{L}}(t) = 6.49$. The MBE and RMSE of Ineichen clear sky model is 23.15 W/m^2 and 60.52 W/m^2 respectively. The MBE and RMSE of Kasten clear sky model is 15.34 W/m^2 and 61.27 W/m^2 respectively. Thus, we use $I_{\mathrm{clr}}(t)$ from Ineichen clear sky model for the later process beacuse it provides the lowest error in RMSE.

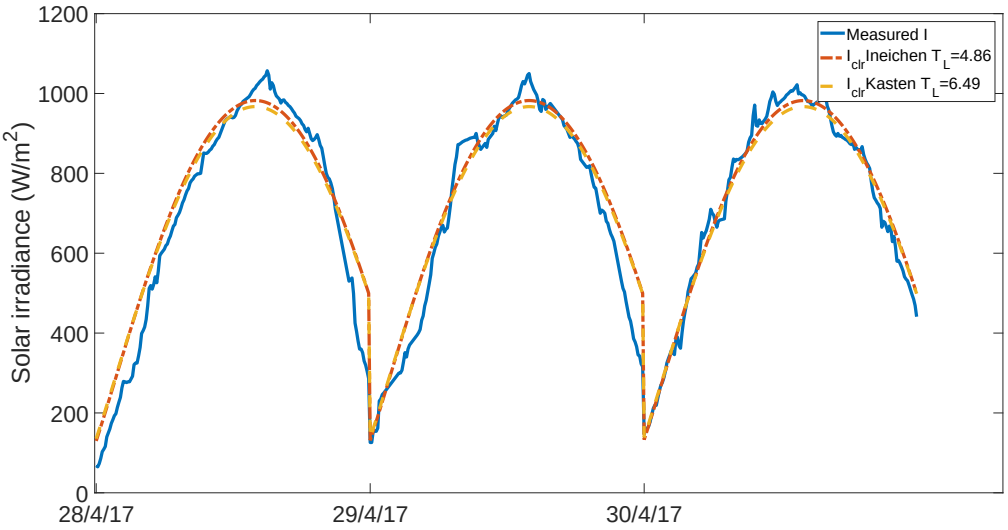


Figure A.2: Measured solar irradiance and solar irradiance from Ineichen and Kasten clear sky models.

APPENDIX B
DATA PRE-PROCESSING

The measured data are contaminated with missing values. If the missing values are not longer than one hour, the linear interpolation is used to impute the missing values, otherwise, the averaged value of 10 days before and after the missing value occurs at the same time is used to impute the missing value. Due to high sampling rates of local weather data, the smoothing method is also used to reduce the variation of data by moving average method with time span of 15 minutes before and after the current value. We then downsampling to hourly data. The processes of imputation long consecutive missing data are described as follows.

1. Missing data imputation: Measured data usually contains long consecutive missing values as shown in Figure B.1. There are many ways to impute the data and we try to impute the missing data by some simple methods such as linear interpolation, spline interpolation, and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP). The results of simple imputation methods are shown in Figures B.2 to B.6. We can see that these methods yield bad results. In addition, we also try to impute with moving average method but the results are worse than those previous methods as shown in Figure B. 7. In order to impute long consecutive missing values, we use the averaged value of 10 days before and after the missing value occurs at the same time to impute the missing value. The equation can be expressed as

$$Y_{\text{impute}}(t) = \frac{1}{20}\sum_{i=1}^{10} Y(t-24i) + Y(t+24i) \tag{B.1}$$

where $Y_{\text{impute}}(t)$ is the imputed data at time t, i is the number of days, and $Y(t)$ is the measured data at time t. The results of our proposed method are better than the other methods for imputing the long consecutive missing values as shown in Figure B.8. Therefore, we impute long consecutive missing values by this method in all variables.

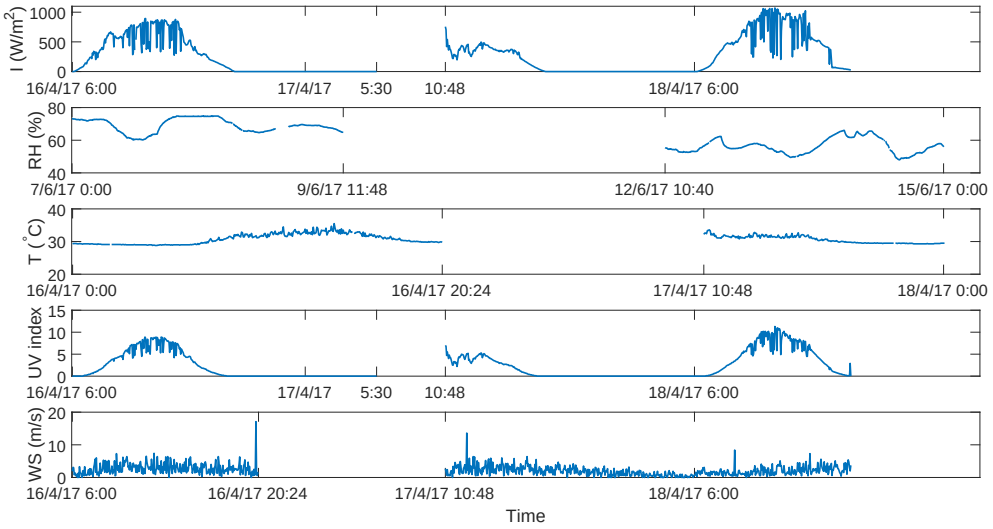


Figure B.1: Missing measurements data.

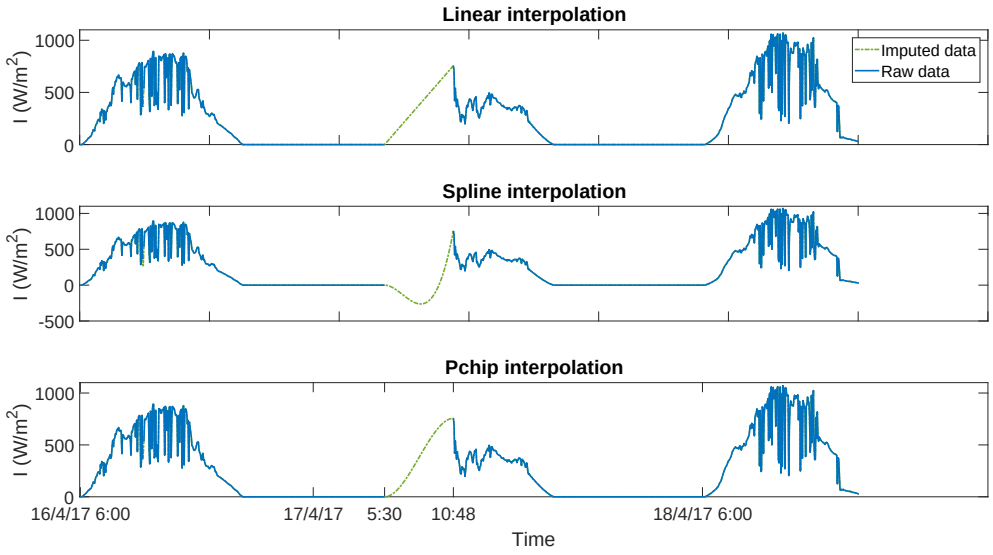


Figure B.2: Imputed solar irradiance data with simple methods.

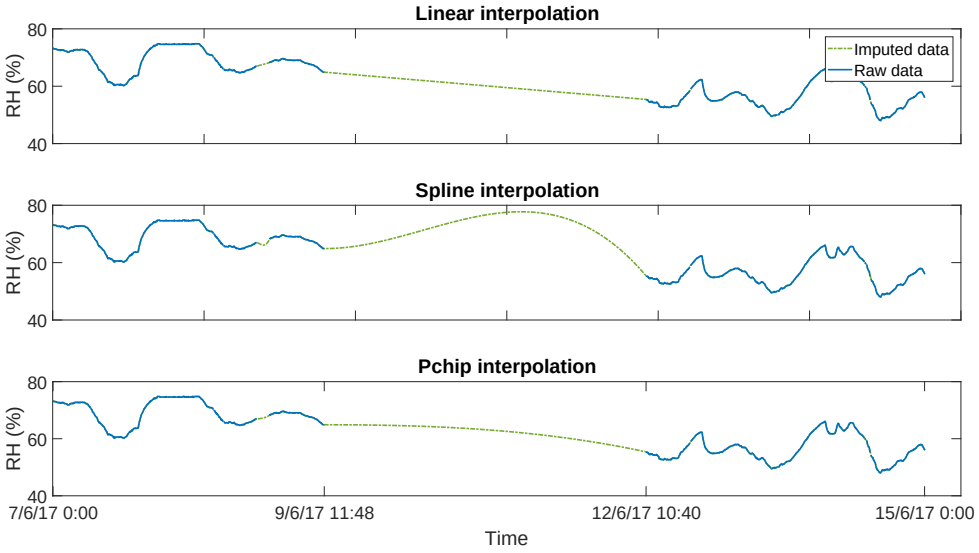


Figure B.3: Imputed relative humidity data with simple methods.

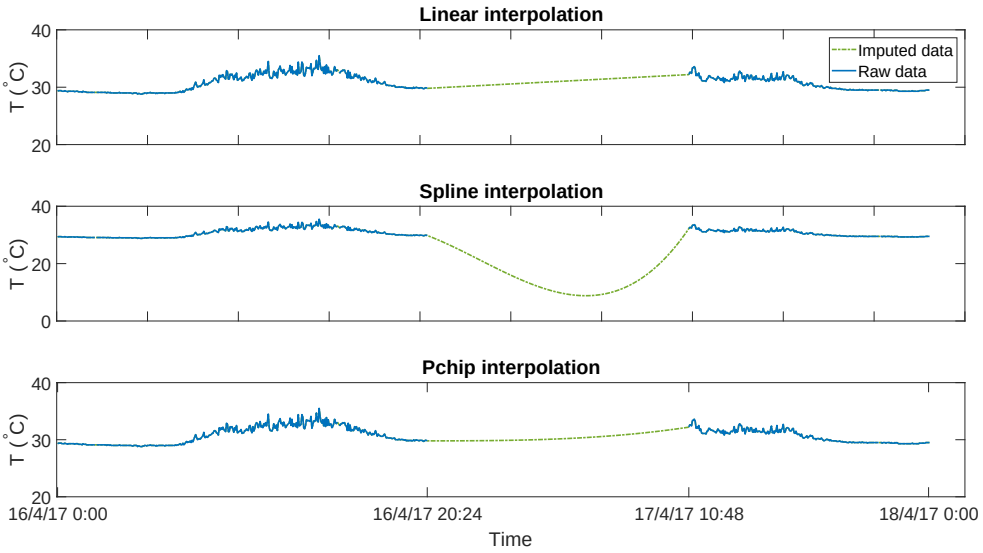


Figure B.4: Imputed temperature data with simple methods.

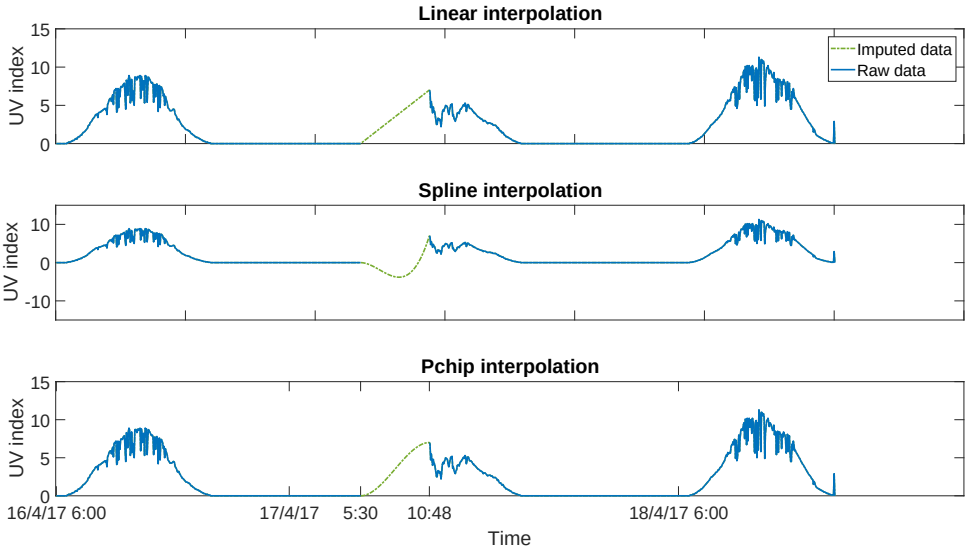


Figure B.5: Imputed UV index data with simple methods.

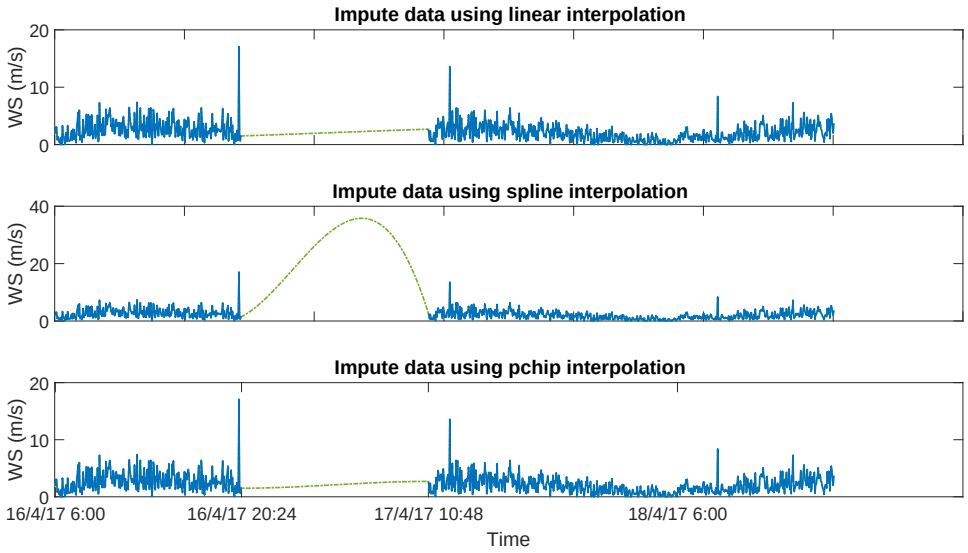


Figure B.6: Imputed wind speed data with simple methods.

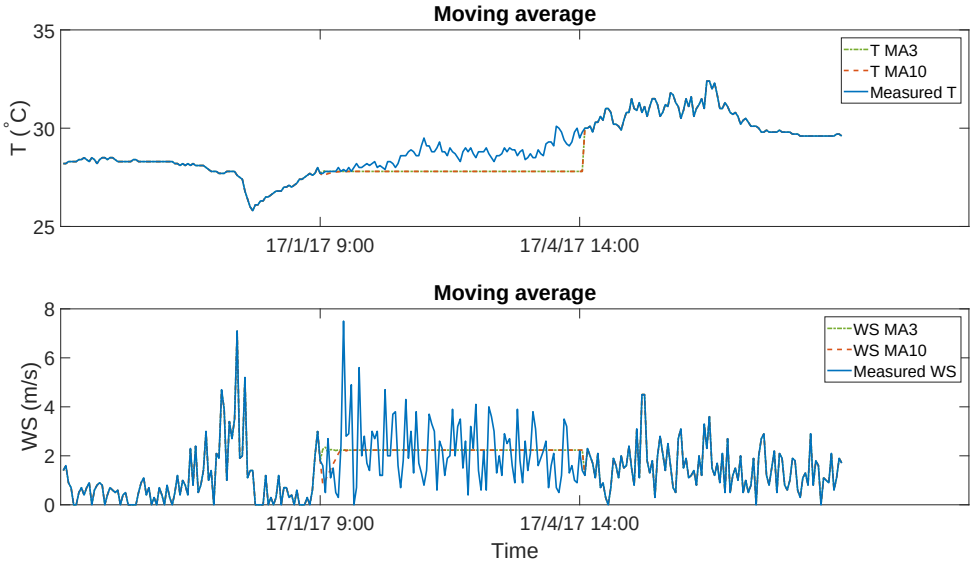


Figure B.7: Imputed temperature and wind speed data by 3rd and 10th order moving avrage.

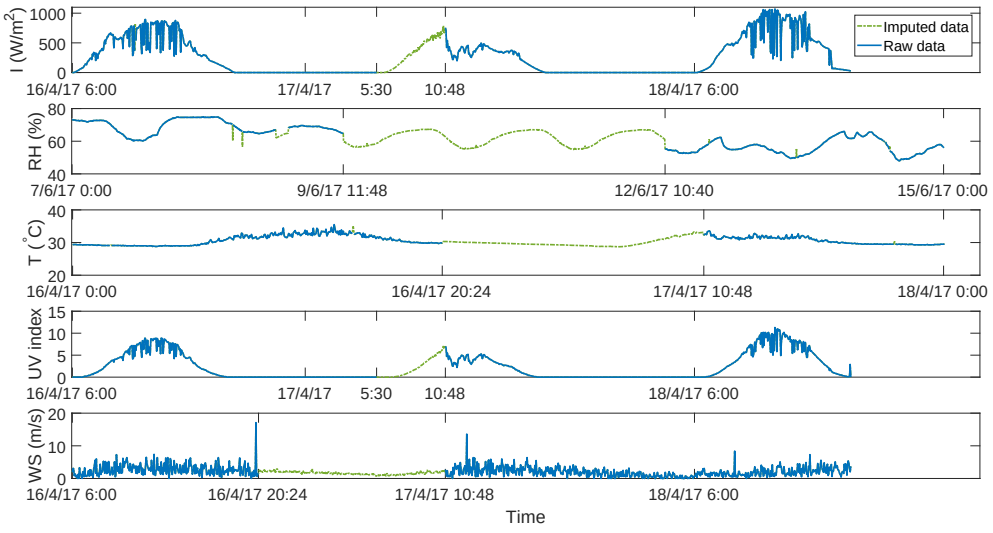


Figure B.8: Imputed data using mean of 10 days before and after the missing value occur.

2. Smoothing: From Figure B.8, we can see that the imputed data fluctuate all the time and we need to do the experiments using just hourly data. Then we apply smoothing method to the imputed data in order to reduce the fluctuation of the data by moving average method with time span of 15 minutes before and after current value.

APPENDIX C
EXPERIMENT ON NOISE COVARIANCE OF KALMAN FILTER

This experiment aims to determine the most suitable method to estimate noise covariance that is used in Kalman filter. We explore three methods to determine the noise covariance which described as follows.

1. KF with fixed noise covariance: All conditions of this method is the same as hourly-step MOS+KF1.

2. KF with fixed noise covariance from batch estimation [119]: This method uses the output of MOS+KF1 to estimate the noise covariance and it can be expressed by

$$W = \frac{1}{N-1}\sum_{t=1}^{N}\left\{(\hat{w}(t)-\bar{w})(\hat{w}(t)-\bar{w})^T - \left(\frac{N-1}{N}\right)[A(t)P_{t-1|t-1}A(t)^T - P_{t|t}]\right\}$$
$$V = \frac{1}{N-1}\sum_{t=1}^{N}\left\{(\hat{v}(t)-\bar{v})(\hat{v}(t)-\bar{v})^T - \left(\frac{N-1}{N}\right)[C(t)P_{t|t-1}C(t)^T]\right\}$$

   where $\hat{w}(t) = \hat{z}_{t|t} - \hat{z}_{t|t-1}$ is the error between measurement update and time update, $\bar{w} = \frac{1}{N}\sum_{t=1}^{N}\hat{w}(t)$, $\hat{v}(t) = y(t) - C(t)\hat{z}_{t|t-1}$ is the error between measurement and predicted value, $\bar{v} = \frac{1}{N}\sum_{t=1}^{N}v(t)$, and N is the number of training data.

3. KF with adaptive noise covariance [112]: At the begining, this method uses the same condition as MOS+KF1. After that this method estimates the noise covariance by

$$W(t) = \frac{1}{N-1}\sum_{i=1}^{N}\{(\hat{w}(t-i)-\bar{w})(\hat{w}(t-i)-\bar{w})^T\}$$
$$V(t) = \frac{1}{N-1}\sum_{i=1}^{N}\{(\hat{v}(t-i)-\bar{v})(\hat{v}(t-i)-\bar{v})^T\}$$

   where $\hat{w}(t) = \hat{z}_{t|t} - \hat{z}_{t|t-1}$ is the error between measurement update and time update, $\bar{w} = \frac{1}{N}\sum_{t=1}^{N}\hat{w}(t)$, $\hat{v}(t) = y(t) - C(t)\hat{z}_{t|t-1}$ is the error between measurement and predicted value, $\bar{v} = \frac{1}{N}\sum_{t=1}^{N}v(t)$, and N is the number of the data that used in the process of estimate the noise covariance.

The results of the three models based on MOS+KF1 are evaluated by RMSE. The method that provide the lowest RMSE will be selected to use in the proposed model. The results are shown in Figure C.1.

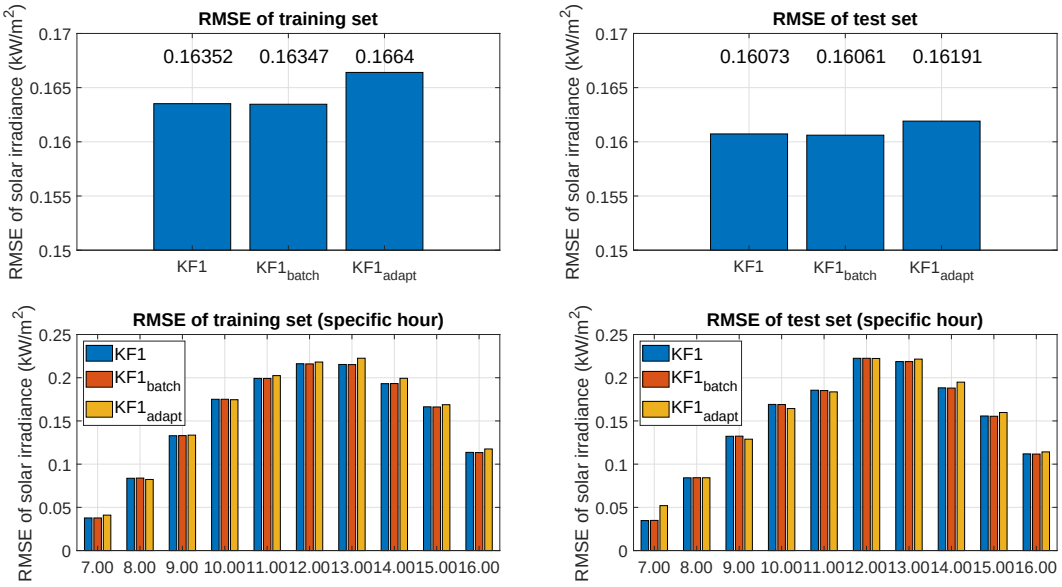


Figure C.1: RMSE of the predicted solar irradiance from various estimate noise covariance methods.

We can see that MOS+KF1 with batch estimation provides the lowest RMSE. However, the result is not significant difference from the other models. MOS+KF1 with adaptive noise covariance is the worst because at the beginning there is too little data for estimating the noise covariance. Then, MOS+KF with adaptive noise covariance provides inaccurate forecasted values at the beginning. We can see the regression coefficients of $\hat{z}(t+1|t)$ in Figure C.2.

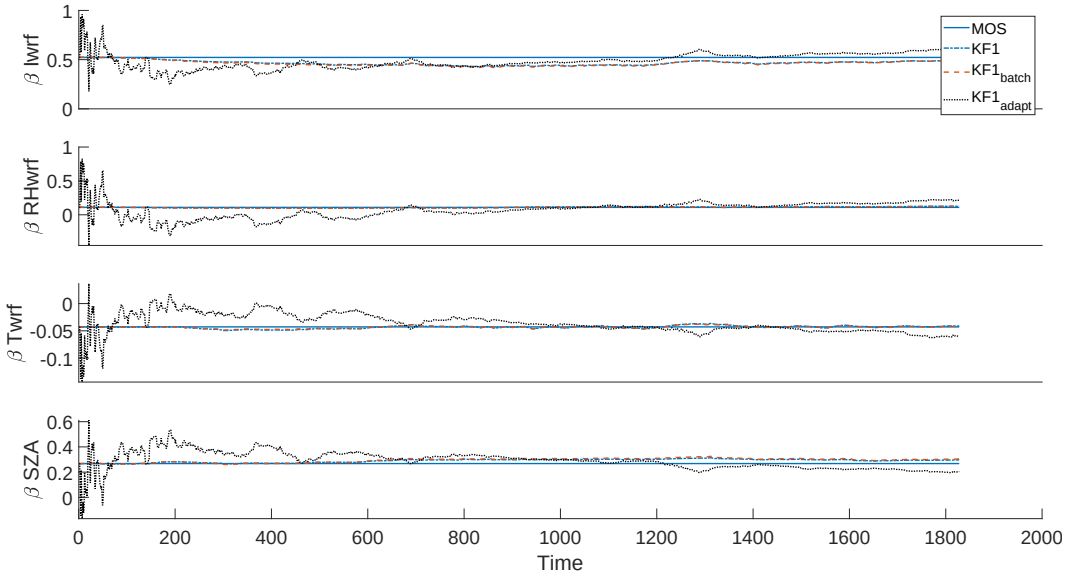


Figure C.2: Regression coefficients from various estimate noise covariance methods.

The regression coefficients of MOS+KF1 and MOS+KF1 with batch estimation have almost the same trend which varies around the regression coefficients of MOS. The regression coefficients of MOS+KF1 with adaptive noise covariance are too adaptive especially at the

beginning due to lack of data. This problem also affects the error covariance and Kalman gain as shown in Figures C.3 and C.4.

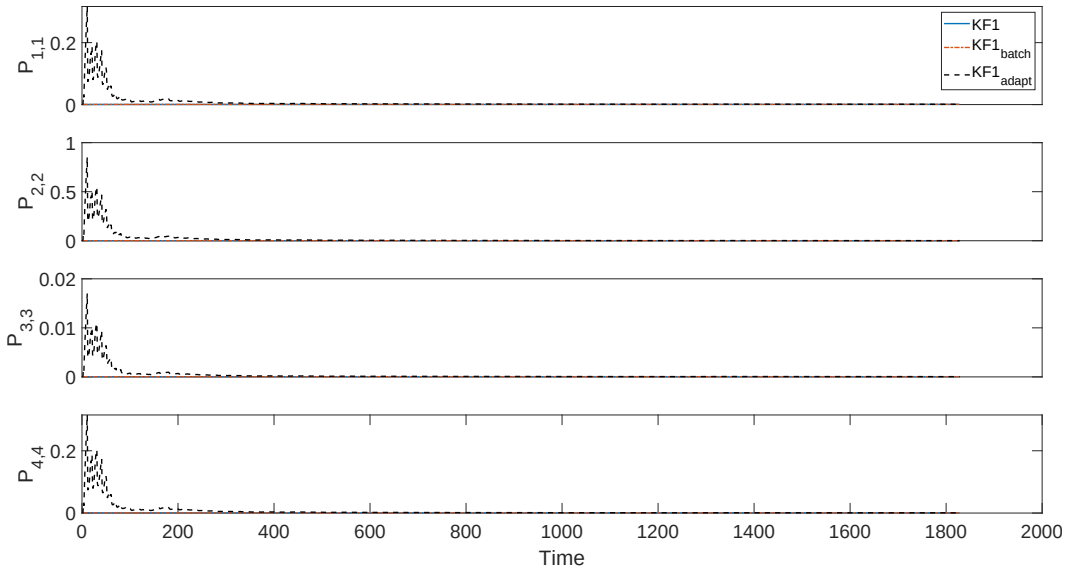


Figure C.3: Error covariance from various estimate noise covariance methods.

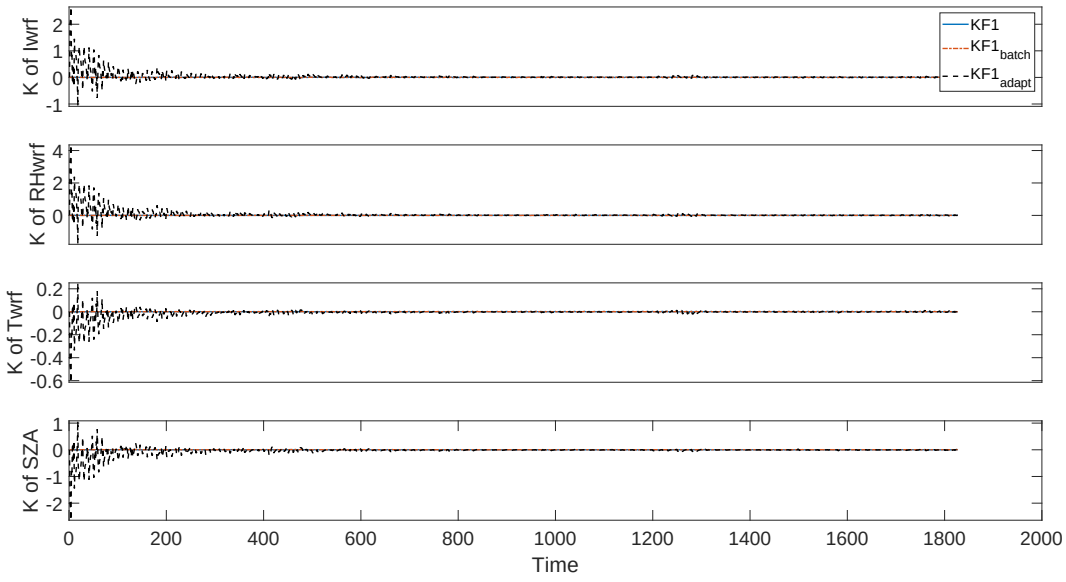


Figure C.4: Kalman gain from various estimate noise covariance methods.

In conclusion, MOS+KF1 is the simple way to estimate the noise covariance and the model still provides good results. Although MOS+KF with batch estimation is the best, however, the performance is a little bit improved. Therefore, we use MOS+KF1 which estimates noise covariance from the residual error of MOS in this work.

APPENDIX D
COMPUTER CODES

This section provides the computer codes of MATLAB which are used in this work. The computer codes are uploaded in github (`https://github.com/jitkomut/solarnwpmos`) and the uploaded computer codes are listed in Table D.2.

Table D.2: List of computer codes.

Name	**Description**
impute.m	Imputation for measurements data
prepare.m	Preparation of the input data for proposed MOS and MOS+KF both hourly and daily models
run_mos.m	Run MOS
run_kf.m	Run MOS+KF
perf_plot.m	Performance plot for solar irradiance forecasting

Biography

Supachai Suksamosorn was born in Bangkok, Thailand, in 1989. He received his Bachelor's degree in electrical engineering from Kasetsart University, in 2012. He has been granted a scholarship from Electricity Generating Authority of Thailand for master's degree at Chulalongkorn University, in 2016. His research interests are in the area of smart grid and renewable energy, especially solar irradiance forecasting by Numerrical Weather Prediction.

List of Publications

1. S. Suksamosorn, N. Hoonchareon and J. Songsiri, "Influential Variable Selection for Improving Solar Forecasts from Numerical Weather Prediction," *Proc. 15th ECTI Conference*, Chiang Rai, Thailand, (2018): 1–4.

2. V. Layanun, S. Suksamosorn and J. Songsiri, "Missing-data imputation for solar irradiance forecasting in Thailand," *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, Kanazawa, Japan, (2017): 1–6.