

MODELLING OF INTRADAY PHOTOVOLTAIC POWER PRODUCTION

Noor 'Adilah Ibrahim^{1a,b*}

^aFaculty of Science and Technology, Islamic Science University of Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, MALAYSIA. Email: nooradilah@usim.edu.my¹

^bDepartment of Mathematics, University of Oslo, PO Box 1053 Blindern, N-0316 Oslo, NORWAY. Email: nooradilah@usim.edu.my¹

*Corresponding author: nooradilah@usim.edu.my

Received: 31st Dec 2019

Accepted: 8th Jan 2021

Published: 30th Jun 2021

DOI: <https://doi.org/10.22452/mjs.vol40no2.8>

ABSTRACT Photovoltaic (PV) productions should occur within a time interval of sunlight. Time mismatches are detected between sunrise and first production hour as well as sunset and last production hour in a transmission system operator, Amprion, Germany. Hence, in this paper, we investigate this effect using an additive function of two seasonalities and a stochastic process. Both seasonalities are based on the mimicked locations, corrected by a weighing scale, depending on the first and last production hours' coordinates. The result shows that the proposed deterministic model could capture the effect of sunrise and sunset. Also, the dynamics of random components are sufficiently explained by an autoregressive process of order two. Finally, the Normal Inverse Gaussian distribution is shown as the best distribution in explaining noise behaviour, particularly heavy tails in the production's residuals, compared to the Gaussian distribution.

Keywords: Intraday PV production, Sunrise, Sunset, NIG Distribution, AR process

ABSTRAK Penghasilan fotovoltaik (PV) sepatutnya berlaku dalam selang waktu cahaya matahari. Berdasarkan kepada kajian di satu pengendali sistem penghantaran, Amprion di Jerman, berlaku ketidaksepadanan antara waktu matahari terbit dan penghasilan tenaga pertama serta waktu matahari terbenam dan penghasilan tenaga terakhir. Oleh itu, dalam makalah ini, kami cuba untuk menyelesaikan isu ketidaksepadanan ini dengan mengkombinasikan dua proses berketentuan (tetap) dan satu process tidak berketentuan (rawak). Kedua-dua proses berketentuan adalah berdasarkan kepada lokasi mimik, diselaraskan oleh satu skala pemberat. Lokasi mimik tersebut bergantung kepada koordinat waktu penghasilan tenaga yang pertama dan terakhir. Hasil kajian mendapati bahawa fungsi berketentuan yang dicadangkan mampu menjelaskan tentang kesan matahari terbit dan terbenam. Manakala, komponen tidak berketentuan pula memadai dijelaskan dengan proses autoregresif tahap dua. Akhir sekali, taburan Gaussang Normal (NIG) dikenalpasti sebagai taburan paling sesuai bagi menerangkan ciri-ciri rawak dalam residual penghasilan tenaga berbanding taburan Gaussian.

Kata kunci: Penghasilan fotovoltaik, matahari terbit, matahari terbenam, taburan Gaussang Normal, proses autoregresif.

1. INTRODUCTION

Fossil fuels are the dominant sources of electricity supply around the world. The utilization of fossil fuels has triggered debates across various fields, such as social, economics, politics, environments, and social sciences. Besides releasing greenhouse gases, such as carbon dioxide and methane, by combusting fossil fuels, depletion of sources has raised questions about their reliability. All these factors have motivated the deployment of alternative renewable energy resources, such as solar energy, which has grown exponentially since the last decade. However, the increasing penetration rates of solar power will continuously challenge the utilities and power system with various energy management issues, such as demand, storage, and forecasting techniques. Solar power critically depends on weather conditions. This is the major hurdle in the operating system since it may increase the variability and uncertainty in the production, leading to difficulty in dispatching the power. Therefore, to ensure power continuity and good management of ramp rates of the overall power system, it is vital to develop an accurate forecasting model, especially for the systems with a significant share of solar power. The best forecasting model will likely improve the economic dispatch decision (Martinez-Anido et al., 2016).

Essentially, there are two main types of forecasting models of PV production: parametric and nonparametric. The parametric model (also called deterministic or physical) summarizes the meteorological resources on the solar cells, such as solar irradiation and temperature (see Dolara et al., 2015a, Dolara et al., 2015b, Wolff et al., 2016, and Ogliari et al., 2017). In contrast, the nonparametric model (stochastic) does not presume any knowledge of the internal system. With

numerous advantages of the latter model, many researchers have successfully used nonparametric approaches in their studies, including Support Vector Machine (SVM) (see Felice et al., 2015, Abuella and Chowdhury, 2016, and Wolff et al., 2016), Numerical Weather Prediction (NWP) (see Almeida et al., 2015, and Larson et al., 2016), Partial Functional Linear Regression model (PFLRM) (Wang et al., 2016) and Multivariate Adaptive Regression Splines (MARS) (see Li et al., 2016, and Massida and Marrocu, 2017). Artificial Neural Network (ANN) has become the most effective and frequently used method in forecasting PV output with some theory development. A collection of research based on this model includes Do et al., (2016), Rana et al., (2016), Vaz et al., (2016), Cervone et al., (2017), Leva et al., (2017), Zhu et al., (2017), and Prakash et al., (2018).

Both the parametric and nonparametric models have their advantages. However, they are also not exempted from disadvantages. For instance, the parametric model requires several parameters, such as ambient and cell conditions (Dolara et al., 2015). The information of the PV panels' internal system is not often available and requires some assumptions and simplifications, leading to high uncertainty in the output. Besides, the deterministic forecasting model is not enough to convey the system's possible future conditions (van der Meer, 2018).

In contrast, the nonparametric models are often criticized due to historical data requirements, prone to complexity in computations, and the high probability of overfitting. To overcome their weaknesses, combining the unique features of these two groups may improve forecasting accuracy. Dolara et al., (2015) investigated this hybridization in their study extended by

Ogliari et al., (2017). As a result, they emphasized that the hybrid model turns out to be the best forecasting approach. In addition, solar power production is nonlinear and varying in time. Thus, a single forecasting model is insufficient to demonstrate the actual generation behaviour (Zhu et al., 2017). This hybridization is expected to contribute by introducing a simpler model but sophisticated enough to visualize forecasting PV production. A pile of the extensive review of forecasting PV production covering both spatial and temporal modelling has been studied by Antonanzas et al., (2016), Gandoman et al., (2016), Raza et al., (2016), Barbieri et al., (2017), Das et al., (2018), and van der Meer et al., (2018).

To choose a suitable approach, it is necessary to define the forecast time horizon. The intraday time-varying pattern is more informative than average daily data. The cloud cover changes highly caused by the variability of very short-term production (up to 30 minutes) (see Lipperheide et al., 2015) for a proposed forecasting model by computing the motion of clouds and day/night cycles. Forecasting the intraday production is more relevant for dispatching, regulatory, and load purposes. However, for electricity trading and operational planning, a 1-day ahead forecast is often required by the energy traders and transmission system operators (Wang et al., 2016). Many studies have constructed the forecasting model for 1-day ahead using ANN and NWP methods (see Vaz et al., 2016, Larson et al., 2016, Massida and Marrocu, 2017; Leva et al., 2017). However, using these approaches, the valuable information of minute data cannot be adopted since they used historical aggregated data. Moreover, using average data may cause a great loss of information and affect the forecasting accuracy. Compared to day-ahead forecasting, the

forecast for intraday production may lower the economic value. Nevertheless, the penetration of solar power substantially increases, boosting up market opportunities. Therefore, the accuracy of the intraday forecast should be improved.

This paper concentrates on PV power in Germany, an extension of our previous work where we modelled maximum PV production assumed at 12:00 pm (see Benth and Ibrahim, 2017). By the end of 2016, Germany has been ranked as the third-highest PV installed, accumulating to 41.2 GW, lost after China and Japan (see IEA, 2017). Until recently, Wirth (2018) reported that Germany's PV-generated power has significantly covered 7.2% of net electricity consumption. These statistics show a remarkable contribution of PV in German's power system. To assure grid quality and stability, a good forecasting model is highly demanded. Motivated by previous studies mentioned above, our main aim in this paper is to introduce a hybrid model of the deterministic and stochastic model, emphasizing its significance by Dolara et al., (2015) and Ogliari et al., (2017) using a simpler model but sophisticated enough to capture both behaviours.

In particular, we will use the sun intensity function to capture the deterministic behaviour and autoregressive process (AR) to detect the random characteristics as in our previous paper (see Benth and Ibrahim, 2017). Our proposed deterministic model does not require any complex physical model such as a sky image system (see Chow et al., 2015). Indeed, it is a simple linear regression model with some estimated parameters that indirectly explain PV panels' physical behaviours. Instead of using a very complex stochastic model, such as ANN and NWP, we propose to use the AR model, which has been highly acknowledged in finance (see

Benth and Šaltytė Benth (2012), for extensive literature review on the AR process). Its ability to extract the statistical properties and detect how much past information affects today's value makes this model widely useful (Das et al., 2018). The main difference between this paper and our previous work is the time horizon. Previously, we picked the maximal production assumed at 12:00 pm. Since very short-term forecasting (less than 1 hour) is more relevant for scheduling the reserve and demand response (see Antonazas et al., 2016), we are now keen to construct a forecasting model of 15 minutes intervals *Intraday Production*.

This paper is divided into several sections. In Section 2, we describe the mismatch between productions and sunlight hours using sunrise/sunset functions. To overcome the mismatches, we mimic the original location such that the first/last productions coincide with the sunrise/sunset in Section 3.1. In Section 3.2, we propose an additive function of two seasonalities using the mimicked locations, corrected by a weighing scale. We capture the random effects in Section 3.3 using the autoregressive process, explaining the residuals' appropriate distribution. In Section 4, we present the results of the empirical analysis. We then conclude this paper in Section 5.

2. MISMATCH OF PRODUCTIONS AND SUNLIGHT HOURS

We analyze the intraday PV production based on a Transmission System Operator (TSO), called Amprion, located in Pullheim, Germany, with latitude 51° and longitude 6.8° . The data collected are based on quarter-hourly measurements for four years from 1st January 2012 to 31st December 2015. This generates 96 pieces of data in a day and 140160 data in total. We use linear interpolation to fill in all missing values, constituting approximately 0.21% of the data. The mean, standard deviation, skewness, and kurtosis are approximately 860.69MW, 1344.71MW, 1.75, and 1.63, respectively, while the maximum production is around 6630MW.

Before looking further at data analysis, we emphasize the key points of this paper. PV power can be produced with the presence of the sun. To analyse the intraday generations, it is crucial to investigate if the production occurs in the daytime. Thus, in this section, we look thoroughly into the variation of the first and last production hours in a day to ensure the productions are within the time interval of the sunlight hours. For this purpose, we use a fixed location of Amprion to determine the sunrise and sunset hours using the following functions given by Duffie and Beckman (2013),

$$\text{Sunrise} = 12\text{pm} - \frac{1}{15^\circ} \cos^{-1} [-\tan \phi \tan \delta], \quad (2.1)$$

and

$$\text{Sunset} = 12\text{pm} + \frac{1}{15^\circ} \cos^{-1} [-\tan \phi \tan \delta], \quad (2.2)$$

where ϕ and δ are latitude and declination angles, respectively. The declination angle is given as

$$\delta = 23.45^\circ \sin \left[\frac{360}{365} (284 + n) \right],$$

with n is the n th day of the year. The term $\cos^{-1} [-\tan \phi \tan \delta]$ is originally based on the zenith angle function, which is the main function of the sun intensity (will be explained in Section 3.2). Knowing the earth is rotating at 15° per hour, we can

compute the sunrise and sunset from noon. Since the sunrise and sunset are changing over time based on location, the latitude plays a crucial role in determining the precise hour.

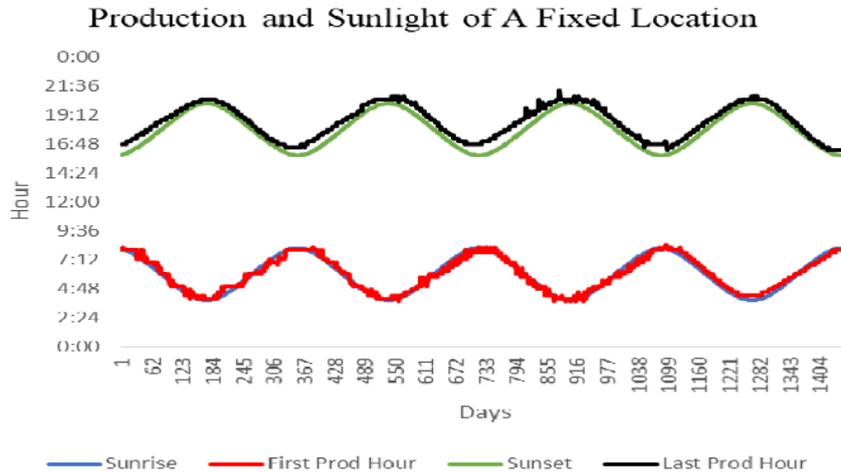


Figure 1. First production hour (red), last production hour (black), sunrise (blue), sunset (green)

Theoretically, the first and last productions should coincide with the sunrise and sunset hour obtained in equation (2.1) and (2.2) above. However, they are not exactly matched, as shown in Figure 1. The first production hours (red curve) seem to fit the sunrise (blue curve) rather well, with a 1-hour maximum difference. In contrast, the maximum difference of the last production hour is comparatively bigger being almost two hours, where the productions (black curve) continuously occur after sunset (green curve). These differences seem illogic as the productions should happen within 15 minutes intervals (since we use quarter-hourly data). In addition, the number of sunlight hours is different over the year. This leads to an inhomogeneous time series, which are certainly not easy to analyse. Thus, an alternative way is required to fix this problem, particularly the sunrise and sunset effects.

3. METHODOLOGY

As aforementioned, we are keen to build a simple model but sophisticated enough to explain intraday production behavior. Let us recall step by step procedures used in previous papers, as mentioned above (one also can refer to Veraart and Zdanowicz (2016) for a PV analysis). First, we applied a suitable seasonality function to explain the deterministic behaviour of PV production. Next, we eliminated all trend and seasonality components to see the random effects using the AR process. Finally, we then specified an appropriate distribution for the residuals by deducting all deterministic and stochastic factors. For this paper, the modelling approach is similar. However, we emphasize the quarter-hourly variations. Thus, our general model for quarter-hourly measurement at time $t = 1, 2, \dots, 140160$ is written as

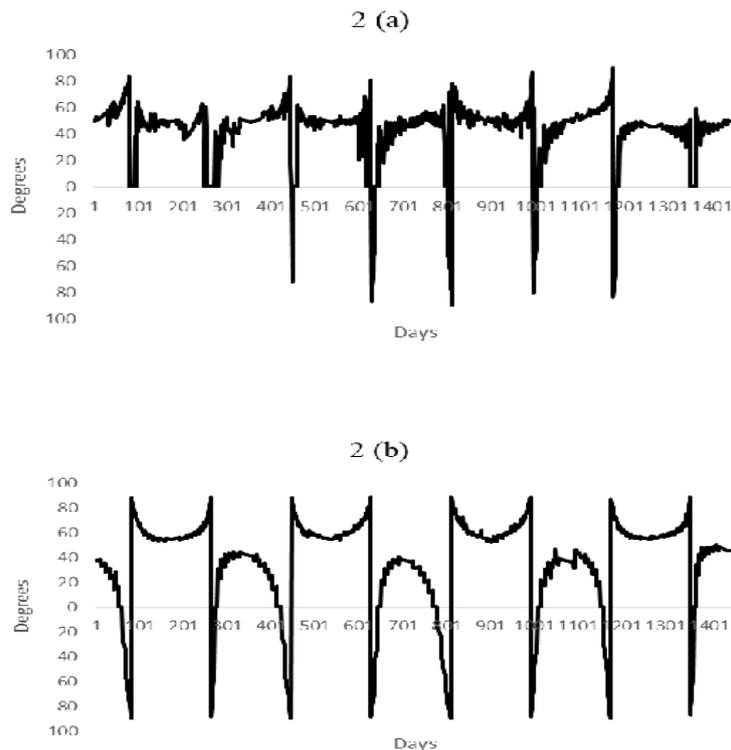
$$\ln(PV(t)) = \Lambda(t) + X(t), \tag{3.1}$$

where $\Lambda(t)$ is to capture the deterministic pattern and $X(t)$ is the stochastic process to seize the dynamics of random effects, explained in the next section. We use a logarithmic transformation to have more organized data.

3.1 Mimicking location

To overcome the mismatch problem of sunrise/sunset and the first/last production, we suggest mimicking the original location by finding locations of the first and last productions. To specify the corresponding productions' exact locations, we use equations (2.1) and (2.2) above by assuming the first and last production hours as the sunrise and sunset hour, respectively. As expected, the

latitude and longitude obtained vary over time, where the locations based on equations (2.1) and (2.2) do not coincide, as demonstrated in Figure 2. Note that Figures 2(a) and 2(b) are the latitude of the first and last productions, respectively. In contrast, Figure 2(c) and Figure 2(d) display the longitude. These plots show seasonality patterns and significantly deviate far from the original location of Amprion, where the latitude is 51° and longitude 6.8° . In contrast, the mimicked locations of the first production are $-90^\circ \leq \text{latitude} \leq 90^\circ$ (as in Figure 2(a)) and $10.9^\circ \leq \text{longitude} \leq 13.2^\circ$ (in Figure 2(c)), and the locations of the last production hour are $-90^\circ \leq \text{latitude} \leq 90^\circ$ (in Figure 2(b)) and $14.94^\circ \leq \text{longitude} \leq 15.1^\circ$ (in Figure 2(d)).



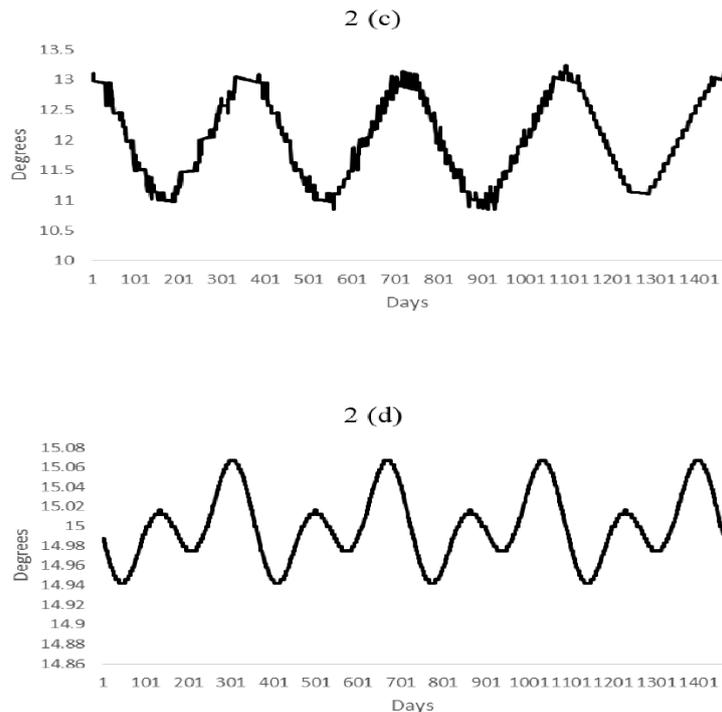


Figure 2. Mimicked locations of productions. 2(a) Latitude of the first production; 2(b) Latitude of the last production; 2(c) Longitude of the first production; 2(d) Longitude of the last production.

For a better understanding, let us pick one point as an example. As shown in Figure 2(a), the latitude of the first production hour on 22nd March 2014 (which is day-752) is -90° , the lowest part of the southern hemisphere. Clearly, this point of latitude is far outside the original location. On this date, the declination angle, δ , is approximately 0° . This means the northern and southern hemispheres' locations are in spring and autumn equinoxes, respectively (refer to Honsberg and Bowden, 2016). On the equinoxes, the sunlight rays directly to the equator and the length of day and night are nearly equal. Hence, we can conclude that the first PV production hour in the Amprion area on 22nd March 2014 is the sunrise in the South Pole hemisphere and autumn equinox. At 90° or -90° latitude, the longitude is undefined. Thus, we use linear interpolation to determine an approximate point of longitude. However, this new location seems unreasonable since Germany is in the North East. For this paper, we put aside this

issue since our main objective is to mimic the original location to determine a good estimate of the first and last production.

One may raise a question: Why did the first and last productions not coincide with the sunrise and sunset? This is likely due to several factors, such as the incidence angle of PV panels, type of PV cells, ambient temperature, site location, and solar spectrum (see Mambrini et al., 2015). In fact, it is rather complicated to quantify such factors. Furthermore, based on the discussions by Chattopadhyay (2017), the bell shape diurnal curve of PV power production is strongly dependent on the position of the sun and the configurations of the PV modules. Moreover, the average European peak load is primarily in the evening. Hence west-facing PV modules are suitable for matching the daily load curve. By assuming that the majority of the PV panels in the area of Amprion coverage are facing west, we can deduce that the PV panels require some time after the sun rises

to generate power. This is one of the major reasons for the mismatched curves between the first/last production hours with the sunrise/sunset. In addition, the thickness of the cloud cover also may influence the intraday productions. Despite the sun intensity being virtually similar for the whole plant area, the PV panels covered by thicker clouds will have lower production. Measuring the thickness of the clouds at various locations of PV power plants requires highly complex modelling. For this paper, to avoid complexity, we consider this a random component, easily modelled stochastically, as discussed in Section 3.3.

As shown in Figures 3(a) and 3(b), the mimicked locations obtained are dissimilar for the first and last production hours. The first production hour in Figure 3(a) (the red curve) and the last production hour in Figure 3(b) (the black curve) show that the sunrise and sunset hours properly coincide with the corresponding first and last production hours. However, the other two plots, sunset (green curve in Figure 3(a)) and sunrise (blue curve in Figure 3(b)), are not properly matched with the last/first production hour. Thus, to overcome this issue, we propose a function that can capture these differences in the next subsection.

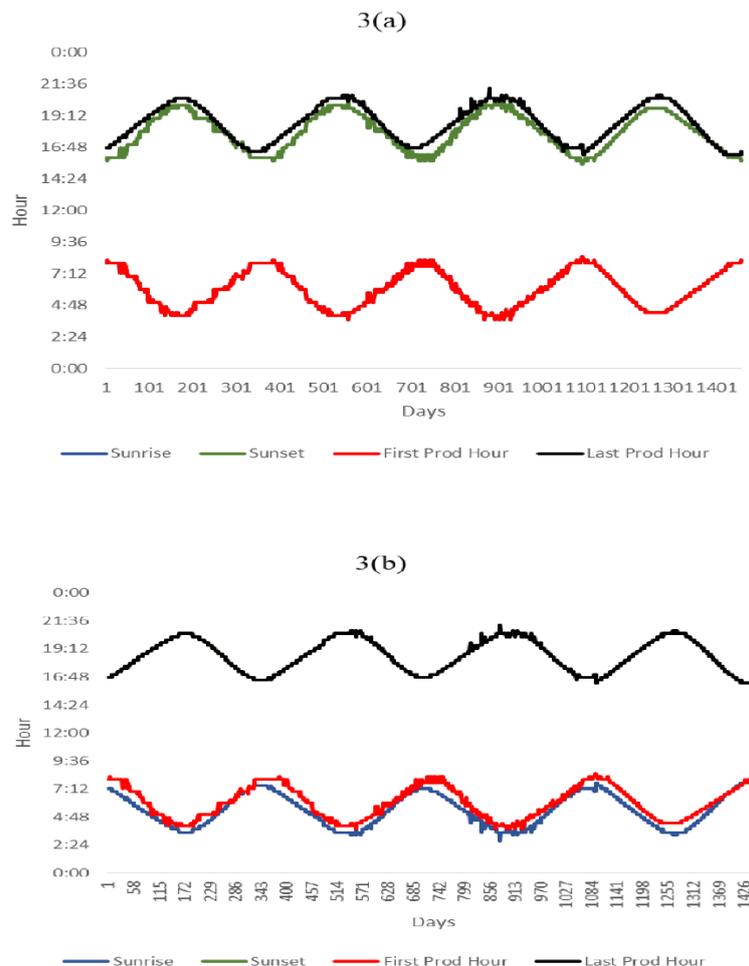


Figure 3. First production hours (red) vs. sunrise (blue) and last production hours (black) vs. sunset (green), based on the mimicked location of the first production (Figure 3(a)) and the last production (Figure 3(b)).

3.2 Capturing the seasonality

In the previous section, two different locations were obtained based on the first and last production hours. In this

section, we investigate how to exploit these two locations series to capture the effect of sunrise and sunset. We let $\Lambda(t)$ be a composition of two seasonality functions as follows

$$\Lambda(t; \theta) = \lambda(t)\Lambda_m(t) + (1 - \lambda(t))\Lambda_e(t), \tag{3.2}$$

where

$$\Lambda_m(t) := \Lambda(t; x_m, y_m, \theta_m),$$

and

$$\Lambda_e(t) := \Lambda(t; x_e, y_e, \theta_e).$$

The seasonality of the first and last production hours are respectively denoted as $\Lambda_m(t)$ and $\Lambda_e(t)$ with weighing

scale $\lambda(t)$, latitude x , longitude y , and the fitted parameters θ . The seasonality function is defined as

$$Q(t) = a + bt + c \ln(I(t)), \tag{3.3}$$

where $Q(t)$ is a generic notation for $\Lambda_m(t)$ and $\Lambda_e(t)$ while a, b , and c are the estimated parameters representing the average level, trend, and amplitude of the mean, respectively. Also, the $\ln(I(t))$ denotes logarithmic of the sun intensity.

Before we go deep on our proposed model, let us first review the sun intensity function, as defined by Duffie and Beckman (2013) and Honsberg and Bowden (2016), given by

$$I(t) = 1.353 \times 0.7^{AM^{0.678}}, \tag{3.4}$$

where 1.353 represents the solar constant, while 0.7 (equivalent to 70%) is the percentage of transmission of the solar radiation to the earth, and 0.678 is the

empirical fit to the observed data and non-uniformities in the atmospheric layers. The term AM denotes the air mass function as follows

$$AM = \frac{1}{\cos(\theta) + 0.50572(96.07995 - \theta)^{-1.6364}}, \tag{3.5}$$

which is a crucial component to measure the reduction of light power passing through the atmosphere and is absorbed by the surroundings. The main ingredient to compute the AM is zenith angle, θ , measured from a vertical line to the sun's angle. The bigger the zenith angle, the bigger the value of air mass. For instance, if the sun is directly overhead, then $\theta = 0$

and, therefore, the $AM \approx 1$. It is noteworthy to mention that since θ varies over time (its value should be between 0° and 90°), the AM is time-dependent.

We can compute the zenith angle, θ , as

$$\cos(\theta) = \cos \phi \cos \delta \cos \omega + \sin \phi \sin \delta, \quad (3.6)$$

where ϕ is the latitude. The declination angle, δ , is given by

$$\delta = 23.45^\circ \sin \left[\frac{360}{365} (284 + n) \right],$$

to represent the tilt of the earth. It should vary seasonally between $+23.45^\circ$ and -23.45° . Meanwhile, the hour angle, ω , can be computed as

$$\omega = 15^\circ (LST - 12pm),$$

whereby the Local Solar Time, LST , is not the same as local time, LT . Thus, we need to add a correction term, TC , given by

$$LST = LT + TC,$$

in which

$$TC = 4^\circ (LSTM - Longitude) + E.$$

Local Standard Time Meridian, $LSTM$, is expressed as

$$LSTM = 15^\circ \Delta T_{GMT},$$

and Equation of Time, E , in minutes, is given by

$$E = 229.2(0.000075 + 0.001868 \cos B - 0.032077 \sin B - 0.014615 \cos 2B - 0.04089 \sin 2B),$$

where

$$B = (n - 1) \frac{360}{365}.$$

The notation ΔT_{GMT} is the difference between the local time from Greenwich Mean Time (GMT) and n represents the number of the day of the year.

The next step is to specify the weighing scale, $\lambda(t)$. Since we know the first and last production hours, we can define $\lambda(t)$ as follows

$$\lambda(t) = \begin{cases} 1, & t = \text{first production hour} \\ 0, & t = \text{last production hour} \end{cases} \quad (3.7)$$

Meanwhile, the value between the first and last production hour will follow a simple linear function

$$\lambda(t) = m(t) + b, \quad (3.8)$$

where m is the slope and b is the intercept. The slope can be obtained by

$$m = \frac{\lambda(t_2) - \lambda(t_1)}{t_2 - t_1},$$

and its value will be different for different numbers of the production's quarter, j . Note that the value of $\lambda(t)$ is zero before the first production starts and after the last

production ends. We plot four examples of $\lambda(t)$ with $j = 37, 58, 65,$ and 46 in Figure 4 representing winter, spring, summer, and fall, respectively.

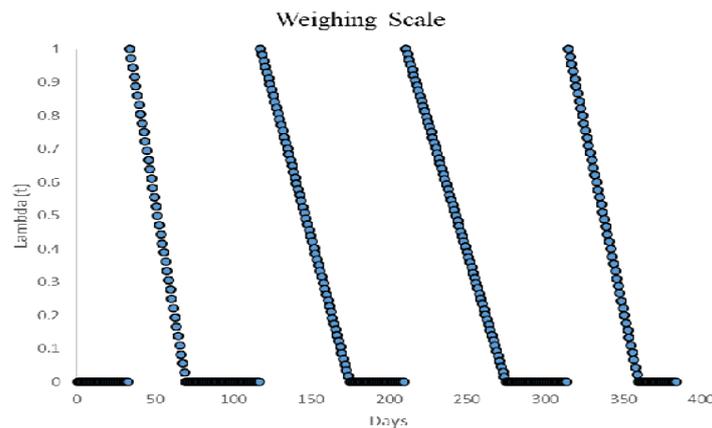


Figure 4. Example of weighing scale, $\lambda(t)$

3.3 Autoregressive process

We move on to the stochastic modelling of the deseasonalized data, which is obtained by subtracting logarithmic PV production, with $\ln(PV(t))$, with $\Lambda(t)$. Similar to our previous paper (see Benth and Ibrahim, 2017), we also apply the AR process to explain stochastic behaviour. The discrete-time AR processes are a parametric family of stationary processes expressed in linear difference equations with constant coefficients (Brockwell and Davis, 1991). Its continuous-time version, CAR, processes are highly needed in pricing the derivatives. Since we are not valuing any derivatives in this paper, the AR processes

are sufficient to explain the dynamics' random characteristics.

According to Sfetsos and Coonick (1999), the AR process alone, including ARMA and ARIMA, tend to a higher prediction error since heteroskedasticity is rarely considered. To overcome this problem, many previous studies have combined the ARMA process with another process, such as GARCH and neural network as in Sun et al., 2015; Benmouiza and Cheknane, 2016; Wu and Chan, 2011; and David et al., 2016, where the volatility was modelled seasonally. However, in this paper, we assumed that the volatility is constant. Therefore, we will remain with our choice of AR process defined as

$$X_q = \sum_{t=1}^p \beta_t X_{q-t} + \varepsilon_q, \tag{3.9}$$

where q denotes a quarter-hour of deseasonalized intraday PV production, p is the order of the AR process, residuals

are denoted as ε_q and the coefficient β_t 's are the estimated regression coefficients.

To describe the residuals' random behaviour, we suggest using Normal Inverse Gaussian (NIG), initially introduced by Barndorff-Nielsen, (1998), to fit the logarithmic returns of financial

data. It belongs to the class of generalized hyperbolic distributions and has four parameters $\alpha, \beta, \delta, \mu$. Its probability density function is given as follows:

$$f(x; \alpha, \beta, \delta, \mu) = \frac{\alpha \beta}{\pi} \exp\left(\delta \sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)\right) \frac{K_1(\alpha \sqrt{\delta^2 + (x - \mu)^2})}{\sqrt{\delta^2 + (x - \mu)^2}}, \quad (3.10)$$

where α represents tails heaviness, β is the skewness, δ denotes scale parameter, μ refers to the location parameter and K_1 is the modified Bessel function of the second kind with order 1. Note that $\alpha \geq |\beta|$ and $\delta > 0$.

4. RESULTS AND DISCUSSION

In this section, we present the findings based on our proposed model. First, the

estimated parameters for average level \hat{a} , trend \hat{b} , and amplitude of the mean \hat{c} are reported in Table 1. According to the results obtained, the trend coefficients seem to be very close to zero. Still, their values are as significant as other parameters with p -values much less than 0.05 significance level. This implies that all parameters introduced in the seasonality function are equally important for explaining PV production's cyclical pattern.

Table 1. Fitted regression parameters of seasonality function

	\hat{a}	\hat{b}	\hat{c}
$\Lambda_m(t)$	7.302	2.26e-06	0.795
$\Lambda_e(t)$	6.916	3.27e-06	0.874

Since all factors are highly significant, then we can proceed with our empirical analysis to the next step. Using equation (3.2), we fit the exponential of logarithmic PV production, as shown in Figure 5. For a better view, we zoomed in the fitted plots of a week production for each season. The plots show that the fitted values (red curve) are almost similar on each of the four plots, but the productions (black curve) vary over the estimation period. Moreover, the fitted curves are

rather weak in explaining the seasonality pattern during the day since the daily productions seem to be much higher than the red curve predictions. However, the fitted curve captured the sunrise and sunset of production very well. Since our main focus is to capture the effect of sunrise and sunset, we can conclude that we managed to explain the variabilities in the first and last production hours by mimicking the location.

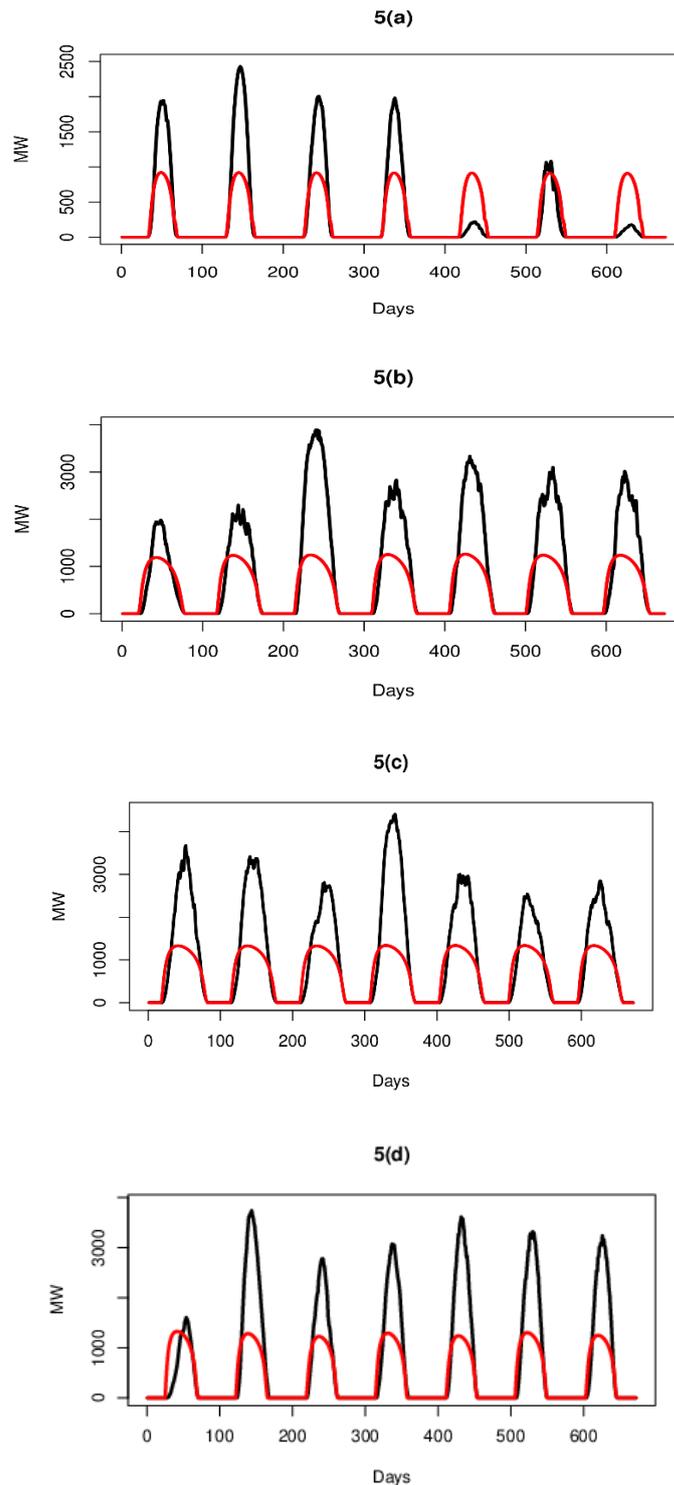


Figure 5. Fitted PV production: Production (Black), Fitted (Red). 5(a) Winter, 5(b) Spring, 5(c) Summer, 5(d) Fall.

Before presenting the AR model's significant order, we first show the plot of the autocorrelation function (ACF) of deseasonalized data in Figure 6. The plot shows exponentially decaying ACF,

implying that the proposed deterministic function in equation (3.2) explains seasonality behaviour. Care must be taken when calculating lags of ACF since the current deseasonalized intraday production

depends on the previous quarter-hour on the same day. Thus, the maximum lag

should not be greater than the total quarter hours of the day.

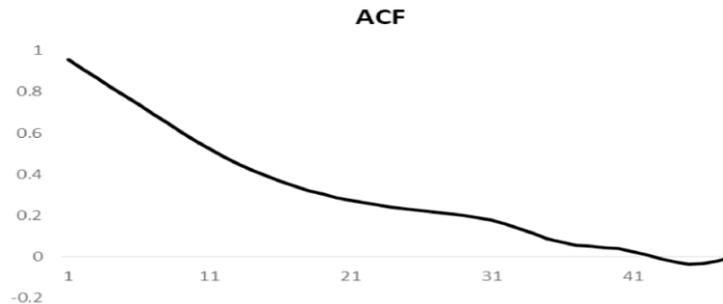


Figure 6. ACF of deseasonalized intraday PV production

In fact, it is difficult to determine the significant lags by looking at the ACF plot. The alternative way is by examining its partial autocorrelation functions (PACF), as shown in Figure 7. With a 95% confidence interval, we can conclude that AR with order 2 is enough to explain the

autocorrelation in deseasonalized intraday PV production. The estimated regression parameters, β_t are reported in Table 2. Since the moduli of the autoregressive polynomials' roots are outside the unit circle, we can conclude that the fitted AR(2) is in stationarity condition.

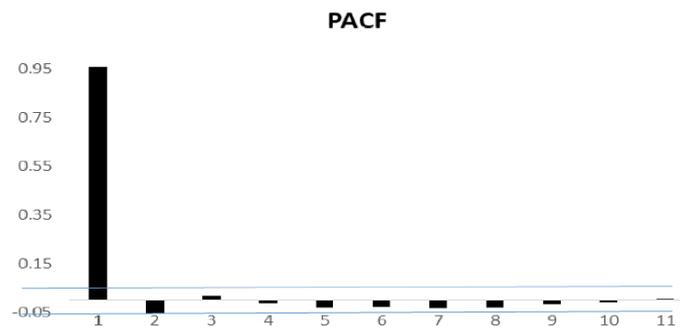


Figure 7. PACF of deseasonalized intraday PV production

Table 2. Regression parameters of AR(2) processes

β_1	β_2
1.228	-0.303

The last step is eliminating AR components to examine the residuals. We show a time series plot of residuals in Figure 8 and its ACF in Figure 9. Strict positive values in the ACF plot tell us that the current random effect is positively correlated with its previous values.

Referring to a 95% confidence interval, we can conclude that most ACF values do not depend on time. Some values lie outside the confidence band, but it can be ignored since the autocorrelation values are very small (less than 0.08).

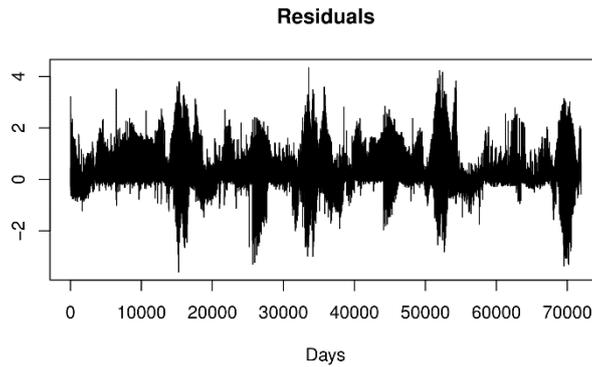


Figure 8. Time series of residuals

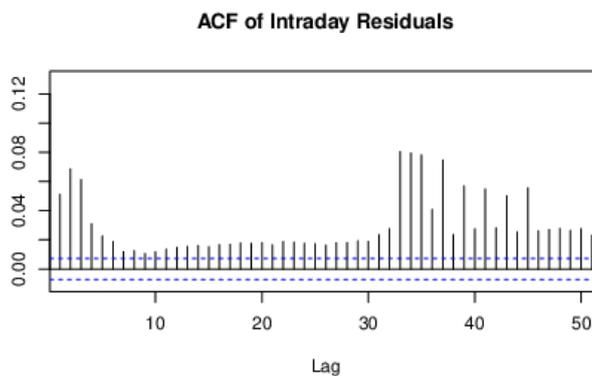


Figure 9. ACF of residuals

Before describing an appropriate distribution of random behaviour, let us briefly summarize the residuals' descriptive statistics. The kurtosis is very high, approximately 41.18, where the skewness is positive at around 3.05. For better demonstration, a quantile plot is displayed in Figure 10, which shows heavy

tails on both sides. This gives an idea that the residuals are certainly not following the normal distribution. Apart from these, Kolmogorov-Smirnov (K-S) 0.285 has a p -value lower than $2.2e-16$. This shred of evidence supports a stylized fact of rejecting the null hypothesis of normality.

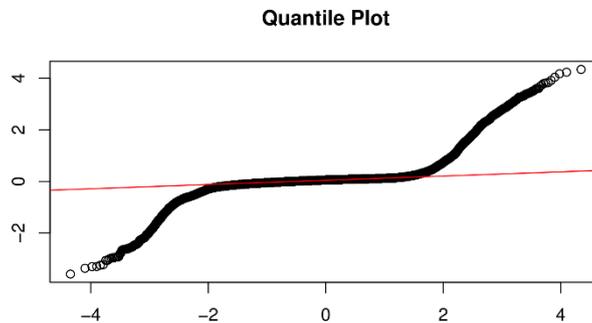


Figure 10. Quantile plot of residuals

Figure 11 compares the normal distribution (red curve) and NIG distribution (black curve). Looking at the plot, we suggest that the residuals follow NIG distribution rather well since it outperforms the normal distribution. We

fit the NIG distribution to the residuals by maximum likelihood estimation (MLE) using `nigFit` command in R programming. The estimated NIG parameters are reported in Table 3.

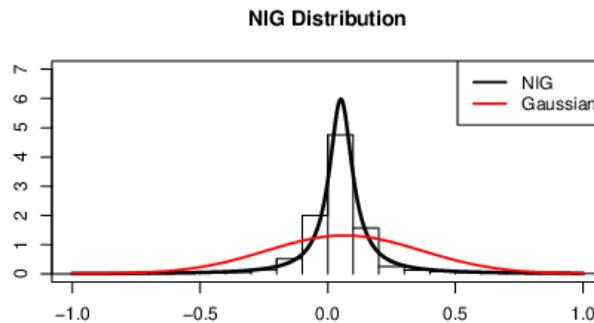


Figure 11. NIG distribution of residuals

Table 3: NIG estimated parameters

α	β	δ	μ
0.657	0.153	0.0550	0.051

There are voluminous studies that have used NIG distribution, such as fitting the residuals of temperature data (Benth and Šaltytė Benth (2005); Benth and Che Taib (2013)), quantifying risk in stocks market (Bølviken and Benth (2000)), and modelling the energy spot prices (Benth and Šaltytė Benth, 2004, and Benth & Henriksen, 2011). It is worth mentioning

5. CONCLUSION

PV power can be produced with the sun's presence and it should occur within a time interval of sunrise and sunset hour. However, the empirical analysis based on four years of quarter-hourly PV production of Amprion, Germany, shows that the first and last production hours are not coinciding with the sunrise and sunset at the corresponding location. Different total sunlight hours over the year leads to complexity in modelling the inhomogeneous time series. To overcome this problem, we came out with a

that the random components are not very well explained, as shown in the quantile plot in Figure 10. We leave this for future research since it requires a more sophisticated model to explain the randomness in intraday PV production, including the effect between days (interday) in the modelling function.

modelling approach suggesting an additive model of two seasonality functions and a stochastic process. We find the locations of the first and last productions to mimic the original location. The result shows that our proposed approach can capture sunrise and sunset effects. However, its ability to explain the seasonality between the first and last productions are rather weak. Additionally, we can also explain the time dependency in the deseasonalized production stochastically with order two's autoregressive process. Heavy tails in the residuals led the NIG distribution to outperform the Gaussian distribution.

With these findings, there are some limitations observed. First, we managed to find locations to mimic the location of Amprion. However, the locations obtained are too far away from Amprion. We put this aside since our main purpose is to find the locations that match the first and last hour of production. This results in a good fit between sunrise-first production as well as sunset-last production. Second, our proposed deterministic function only captures the effect of sunrise and sunset. The seasonality between the first and last productions is not well explained. This is a tremendous and challenging task since we require a model that can capture every quarter-hour of production, which is very sensitive to weather conditions. Despite these drawbacks and limitations, we emphasize that our model could control the effect of sunrise and sunset over the whole dataset.

Finally, there are a few suggestions for future research. First, one may construct a pricing formula of the Quanto option. This acts as a hedging strategy for non-renewable energy producers against high-volume PV production and the low electricity price. This paper used a discrete autoregressive process to describe the short-term random fluctuation in PV production. However, to price the derivatives, it is more convenient to use the continuous-time stochastic process called CARMA processes. Many applications use CARMA processes in energy markets, such as a futures contract on electricity spot and temperature futures on accumulated temperature, CAT. Thus, the CARMA processes might contribute well in pricing the Quanto options too. Second, to be more useful, one may also compute the hedging ratios and measure the effectiveness of the strategy's ex-post performances. In fact, there are few ways to compute the hedging ratios, including minimizing the variance and maximizing profit from hedging. Thus, one needs to specify which types of risk management

they are in. Lastly, one may also include the interday effects in the model. As shown in our findings, the quantile plot of the residuals still shows extreme heavy tails. This gives the idea that some factors remain unexplained, where a more sophisticated model is required. We believe that the intraday PV production is influenced by the previous hours and the previous days. This is certainly not an easy task, but it may overcome the extreme behaviour of the residuals.

6. ACKNOWLEDGEMENT

The author thanks Professor Fred Espen Benth for valuable guidance and discussions. This project is supported by the Malaysian Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS) with code FRGS/1/2020/STG06/USIM/02/1.

7. REFERENCES

- Abuella, M., Chowdhury, B. (2016). Solar power forecasting using support vector regression. *Proceedings of the American Society for Engineering Management International Annual Conference*.
- Almeida, M. P., Perpiñán, O., Narvarte, L. (2015). PV power forecast using a nonparametric PV model. *Solar Energy*, 115, pp. 354-368.
- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136, pp. 78-111.
- Barbieri, F., Rajakaruna, S., Ghosh, A. (2017). Very short-term photovoltaic power forecasting with cloud modelling: A review. *Renewable and Sustainable Energy Reviews*, 75, pp. 242-263.

- Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance & Stochastics*, 2(1), 41-68.
- Benmouiza, K., Chekane, Ali. (2016). Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. *Theoretical and Applied Climatology*, 124, 945-958.
- Benth, F. E., Šaltytė Benth, J. (2004). The normal inverse Gaussian distribution and spot price modelling in energy markets. *International Journal of Theoretical and Applied Finance*, 7(2), 177-192.
- Benth, F. E., Šaltytė Benth, J. (2005). Stochastic modelling of temperature variations with a view towards weather derivatives. *Applied Mathematical Finance*, 12(1), 53-85.
- Benth, F. E., Henriksen, P. N. (2011). Pricing of basket options using univariate normal inverse Gaussian approximates. *Journal of Forecasting*, 30, 355-376.
- Benth, F. E., Šaltytė Benth, J. (2012). *Modelling and Pricing in Financial Markets for Weather Derivatives*. World Scientific, Singapore.
- Benth, F. E., Che Taib, C. M. I. (2013). On the speed towards the mean for continuous time autoregressive moving average processes with applications to energy markets. *Energy Economics*, 40, 259-268.
- Benth, F. E., Ibrahim, N. A. (2017). Stochastic modelling of photovoltaic power generation and electricity prices. *Journal of Energy Markets*, 10 (3), 1-33.
- Bølviken, E., Benth, F. E. (2000). Quantification of risk in Norwegian stocks via the normal inverse Gaussian distribution. R. Norberg et al. (eds) Proceedings of the 10th AFIR Colloquium in Tromsø. AFIR, 87-98.
- Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.
- Cervone, G., Clemente-Harding, L., Alessandrini, S., Monache, L. D. (2017). Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renewable Energy*, 108, pp. 274-286.
- Chattopadhyay, F. K. (2017). *Optimization of spatial balancing and storage needs for large-scale power system integration of fluctuating solar energy* (Doctoral Dissertation, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany). Retrieved from <http://oops.unioldenburg.de/3217/1/chopt17.pdf>.
- Chow, C. W., Belongie, S., Kleissl, J. (2015). Cloud motion and stability estimation for intrahour solar forecasting. *Solar Energy*, 115, pp. 645-655.
- Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., van Dementer, W., Horan, B., Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization. *Renewable and Sustainable Energy Reviews*, 81, pp. 912-928.
- David, M., Ramahatana, F., Trombe, P. J., Lauret, P. (2016). Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy* 133, 55-72.
- Do, M-T., Soubdhan, T., Robyns, B. (2016). A study on the minimum duration of training data to provide a high accuracy forecast for PV generation between two different climatic zones. *Renewable Energy*, 85, pp. 959-964.
- Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., Ogliaresi, E. (2015a). A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies*, 8, pp. 1138 - 1153. (doi:10.3390/en8021138).

- Dolara, A., Leva, S., Manzolini, G. (2015b). Comparison of different physical models for PV power output prediction. *Solar Energy*, 119, pp. 83 - 99.
- Duffie, J.A., Beckman, W.A. (2013). *Solar Engineering of Thermal Processes*, 4th edn. Wiley (<http://doi.org/b993>).
- Felice, M. D., Petitta, M., Ruti, P. M. (2015). Short-term predictability of photovoltaic production in Italy. *Renewable Energy*, 80, pp. 197 - 204.
- Gandoman, F. H., Raeisi, F., Ahmadi, A. (2016). A literature review on estimating of PV-array hourly power under cloudy weather conditions. *Renewable and Sustainable Energy Reviews*, 63, pp. 579-592.
- Honsberg, C., Bowden, S. (2016). Photovoltaic education network. Web page. URL:www.pveducation.org.
- IEA PVPS. (2017). *Snapshot of Global Photovoltaic Markets* (Report No. T1-31:2017). Retrieved from [http://www.iea-pvps.org/_leadadmin/dam/public/report/statistics/IEA-PVPS-A Snapshot of Global PV-1992-2016 1.pdf](http://www.iea-pvps.org/_leadadmin/dam/public/report/statistics/IEA-PVPS-A%20Snapshot%20of%20Global%20PV-1992-2016%201.pdf).
- Larson, D. P., Nonnenmacher, L., Coimbra, C. F. M. (2016). Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. *Renewable Energy*, 91, pp. 11-20.
- Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., Ogliari, E. (2017). Analysis and validation of 24 hours ahead of neural network forecasting of photovoltaic output power. *Mathematics and Computers in Simulation*, 131, pp. 88-100.
- Li, Y., He, Y., Su, Y., Shu, L. (2016). Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines. *Applied Energy*, 180, pp. 392-401.
- Lipperheide, M., Bosch, J. L., Kleissl, J. (2015). Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, 112, pp. 232-238.
- Mambrini, T., Dubois, A. M., Longeaud, C., Badosa, J., Hae_elin, M., Prieur, L., Radivoniuk, V. (2015). Photovoltaic yield: Correction method for the mismatch between the solar spectrum and the reference ASTM G 173-03 spectrum. *EPJ Photovoltaics*, 6, 60701. Doi: 10.1051/epjpv/2014011.
- Martinez-Anido, C. B., Botor, B., Florita, A. R., Draxl, C., Lu, S., Hamann, H. F., Hodge, B-M. (2016). The value of day-ahead solar power forecasting improvement. *Solar Energy*, 129, pp. 192 – 203.
- Massida, L., Marrocu, M. (2017). Use of multilinear adaptive regression splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany. *Solar Energy*, 146, pp. 141-149.
- Ogliari, E., Dolara, A., Manzolini, G., Leva, S. (2017). Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renewable Energy*, 113, pp. 11 - 21.
- Prakash, S., Gopinath, N. P., Suganthi, J. (2018). Wind and solar energy forecasting system using artificial neural networks. *International Journal of Pure and Applied mathematics*, 118(5), pp. 845-854.
- Rana, M., Koprinska, I., Agelidis, V. G. (2016). Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Conversion and Management*, 121, pp. 380-390.
- Raza, M. Q., Nadarajah, M., Ekanayake, C. (2016). On recent advances in PV output power forecast. *Solar Energy*, 136, pp. 125-144.

- Sfetsos, A., Coonick, A. H. (1999). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy* 68(2), 169-178.
- Sun, H., Yan, D., Zhao, N., Zhou, J. (2015). Empirical investigation on modeling solar radiation series with ARMA-GARCH models. *Energy Conversion and Management* 92, 385-395.
- van der Meer, D. W., Widén, J., Munkhammar, J. (2018). Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews*, 81, pp. 1484-1512.
- Vaz, A. G. R., Elsinga, B., van Sark, W. G. J. H. M., Brito, M. C. (2016). An artificial neural network to assess the impact of neighbouring photovoltaic system in power forecasting in Utrecht, the Netherlands. *Renewable Energy*, 85, pp. 631-641.
- Veraart, A. E. D and Zdanowicz, H. (2016). Modelling and predicting photovoltaic power generation in the EEX Market. Working paper, Social Science Research Network (<http://doi.org/b964>).
- Wang, G., Su, Y., Shu, L. (2016). One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models. *Renewable Energy*, 96, pp. 469-478.
- Wirth, H. (2018). *Recent Facts about Photovoltaics in Germany*. Press and Public Relation: Germany. Retrieved from <https://www.ise.fraunhofer.de/content/dam/ise/en/documents/publications/studies/recentfacts-about-photovoltaics-in-germany.pdf>.
- Wol, B., Kühnert, J., Lorenz, E., Kramer, O., Heinemann, D. (2016). Comparing support vector regression for PV power forecasting to a physical model approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135, pp. 197 - 208.
- Wu, Ji., Chan, C. K. (2011). Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Solar Energy* 85, 808-817.
- Zhu, H., Lian, W., Lu, L., Dai, S., Hu, Y. (2017). An improved forecasting method for photovoltaic power based on adaptive BP neural network with a scrolling time window. *Energies*, 10(1542). doi:10.3390/en10101542.