Short-term Solar Irradiance Forecasting Using Exponential Smoothing State Space Model

Zibo Dong∗, Dazhi Yang, Wilfred M. Walsh, Thomas Reindl, Armin Aberle
Solar Energy Research Institute of Singapore, National University of Singapore, Singapore

Abstract

We forecast high resolution solar irradiance time series using an exponential smoothing state space (ESSS) model. To stationarize the irradiance data before applying linear time series models, we propose a novel Fourier trend model and compare the performance with other popular trend models using residual analysis and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity test. Using the optimized Fourier trend, an ESSS model is implemented to forecast the stationary residual series of datasets from Singapore and Colorado, USA. To compare the performance with other time series models, autoregressive integrated moving average (ARIMA), linear exponential smoothing (LES), simple exponential smoothing (SES) and random walk (RW) models are tested using the same data. The simulation results show that the ESSS model has generally better performance than other time series forecasting models. To assess the reliability of the forecasting model in real-time applications, a complementary study of the forecasting 95% confidence interval and forecasting horizon of the ESSS model is proposed.

Keywords: Time series forecasting, Stationarity, Exponential smoothing state space model, Forecast horizon

Nomenclature

\(d\) Peak irradiance hour of the day
\(r_t\) Time series
\(\sigma\) Standard deviation of Gaussian curve
\(l_t\) Level component of ESSS forecasting equation
\(g_t\) Growth component of ESSS forecasting equation
\(s_t\) Seasonal component of ESSS forecasting equation
\(\alpha, \beta, \gamma\) Smoothing parameters
\(\phi\) Damping factor
\(L\) Maximum likelihood function
\(RW\) Random walk
\(SES\) Simple exponential smoothing

\(LES\) Linear exponential smoothing
\(ARIMA\) Autoregressive integrated moving average
\(ESSS\) Exponential smoothing state space model
\(KPSS\) Kwiatkowski-Phillips-Schmidt-Shin stationary test
\(AIC\) Akaike’s information criterion

1. Introduction

Solar photovoltaic (PV) electricity is set to be one of the major energy sources of the future, particularly in developing nations. The instantaneous power output of a PV system can vary substantially, depending on local irradiation conditions and the system performance, making it an intermittent type of renewable energy with high variability. Since more than 90% of the installed photovoltaic systems worldwide are grid-connected, this in turn raises concerns among grid operators regarding the influence of photovoltaic power on the performance of their electricity grid (EPIA).

Considering the power rather than the energy contribution of PV in a typical electricity grid, a 2%
share of the total energy generation, as seen today in Southern Germany, can easily correspond to a more than 10% instant share of the total power. This may occur for example on a bright, sunny weekend day when large industrial consumers are closed (EPIA).

In Singapore, future PV generation is predicted to contribute up to 30% of power generation (National Climate Change Strategy (NCCS)), so to avoid the risk of destabilising the electricity grid, it will be necessary to forecast the contribution of solar power on short (minute), medium (hour) and long (days) timescales. Forecasts allow the grid operator to implement on-the-fly grid management measures, such as up- or downscaling of the power output of other adjustable generation capacities, demand-side management, or forward buying at the electricity exchange at lower rates.

Many irradiance forecasting methods have been proposed. Although there are some outstanding forecasting models based on cloud motion analysis and numerical weather prediction (NWP) (Chow et al., 2011; Perez et al., 2010), statistical time series forecasting has been the most popular technique for short time scales. Linear time series analyses such as random walk (RW), autoregressive (AR), moving average (MA), simple exponential smoothing (SES) and autoregressive integrated moving average (ARIMA) method (Yang et al., 2012; Reikard, 2009), are widely used for modeling and prediction solar irradiance data. Nonlinear time series analyses like artificial neural network (ANN) (Mellit and Kalogirou, 2008; Mellit and Pavan, 2010a,b) and adaptive model (Mellit et al., 2010), have been suggested to be more accurate than simple time series analysis. It is also possible to combine different methods to improve overall forecasting accuracy (Mellit et al., 2005; Voyant et al., 2012).

Recently, the exponential smoothing method has attracted attention. This method was developed by Robert G. Brown in a tracking model of military fire-control equipment in the USA Navy (Gass and Harris, 2000). Brown’s simple exponential smoothing models have been extended to exponential smoothing with trends and seasonality. Brown’s two books, Statistical Forecasting for Inventory Control (Brown, 1959), and Smoothing, Forecasting, and Prediction of Discrete Time Series (Brown, 1963), describe the fundamental methodology of exponential smoothing. In the same period, Charles C. Holt independently developed a similar exponential smoothing method with a different approach to smoothing seasonal data, but this was not published until 2004 (Holt, 2004). Winters Winters (1960) further developed Holt’s method and their work is known as Holt-Winters forecasting system.

Muth (1960) was the first to demonstrate that exponential smoothing can forecast an optimal random walk with noise. Since then, many authors have worked to develop exponential smoothing within a statistical framework. Among them, Box et al. (1970), Roberts (1982), and Abraham and Ledolter (1983) pointed out that some linear exponential smoothing methods are special cases of ARIMA models.

Although substantial research had been done on the exponential smoothing method Gardner (1985), the method was considered an ad hoc forecasting approach since there was no appropriate underlying stochastic formulation until 2002 when Hyndman et al. (2002) proposed the state space framework for exponential smoothing.

Exponential smoothing consists of a total of fifteen methods, which are summarized in Table 1 (adapted from Hyndman et al. (2002)). Some of these methods were first published with different names: cell NN is simple exponential smoothing (SES) and cell AN is linear exponential smoothing (LES).

This paper is organized as follows. Section 2 describes the data we have used: solar irradiance time series measured from stations in both Singapore and Unites States. Section 3 provides the detailed description of our forecasting methodology using an exponential state space model (ESSS), including our innovative Fourier trend model to stationarize the time series, the exponential smoothing forecast equations, state space model equations and the model selection method. Section 4 includes final results and comparisons with other time series methods, showing that the ESSS model has the best forecasting performance of the models tested. The confidence interval generated by the ESSS model indicates the reliability of the predictions. Section 5 concludes and suggests future work.

2. Solar irradiance time series data

The first irradiance time series dataset we have used is from the rooftop station of Solar Energy Research Institute of Singapore (SERIS) located at (1.30°N, 103.77°E). The pyranometer used in this station is a meteorological class Delta-T SPN1 Sunshine Pyranometer.
## Table 1: 15 Exponential Smoothing Methods

<table>
<thead>
<tr>
<th>Trend</th>
<th>Seasonal</th>
<th>N (None)</th>
<th>A (Additive)</th>
<th>M (Multiplicative)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N (None)</td>
<td>N (None)</td>
<td>N (None)</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td></td>
<td>M (Multiplicative)</td>
<td>M (Multiplicative)</td>
<td>M (Multiplicative)</td>
</tr>
<tr>
<td>M (Multiplicative)</td>
<td></td>
<td>M (Multiplicative)</td>
<td>M (Multiplicative)</td>
<td>M (Multiplicative)</td>
</tr>
</tbody>
</table>

1. Located just 1.0 degrees north of the equator, Singapore enjoys a tropical climate with an average daily temperature between 25 degrees Celsius and 31 degrees Celsius, and relative humidity of 70% − 80%. Rainfall occurs almost every day but usually only for a short period of time. The high temperature and high humidity leads to high evaporation which causes frequent cloud formation and dense cloud cover. Therefore the solar irradiance variability is high in Singapore.

The second set of irradiance time series data is from a rooftop station in South Park, Colorado, USA located at (39.16°N, 105.37°W) (Measurement and Instrumentation Data Center (MIDC), NREL). The pyranometer used in this station is a LI-200 LI-COR Pyranometer which compares favorably with first class thermopile-type pyranometers. The climate of South Park is semi-arid with low humidity and low precipitation. The solar irradiance variability is moderate and typical for the temperate zone. This is thus an interesting location to compare with Singapore.

### 3. Forecasting methodology

This section details the steps in implementing our prediction model. First, our Fourier trend model is introduced and compared with other trend models. Second, the exponential smoothing forecasting equations are elaborated. In order to generate 95 percent confidence interval, the state space models of the exponential smoothing technique are then introduced. Finally, the model selection method is presented to identify the best model in each step forecast.

#### 3.1. Trend model

The foundation of time series analysis is stationarity. A time series \( \{r_t\} \) is said to be strictly stationary if the joint distribution of \( (r_{t_1}, \ldots, r_{t_k}) \) is identical to that of \( (r_{t_1+t}, \ldots, r_{t_k+t}) \) for all \( t \), where \( k \) is an arbitrary positive integer and \( (r_{t_1}, \ldots, r_{t_k}) \) is a collection of \( k \) positive integers (Box et al., 1970). This simply means the joint distribution of the time series remains unchanged under time shift. However, this condition is hard to verify empirically: thus a weaker version of stationarity is usually applied. A time series \( \{r_t\} \) is considered weakly stationary if both the mean of \( r_t \) and the covariance between \( r_t \) and \( r_{t-l} \) are time invariant, where \( l \) is an arbitrary integer. In real applications, with \( N \) observed data points, weak stationarity implies that the time plot of the data would show that these \( N \) values fluctuate with constant variation around a fixed level. Weak stationarity is generally sufficient to predict future observations (Brown, 1963).

Since our global horizontal irradiance (GHI) data is non-stationary with a clear daily trend, this daily trend must be removed to obtain stationarity. The most commonly-used trend models include high order polynomials (Al-Sadah et al., 1990), cosine functions (Kaplanis, 2006) and the Gaussian function model (Baig et al., 1991). However, we have found that the performance of these models are not satisfactory and thus we have developed our own Fourier trend model.

#### 3.1.1. Fourier trend model

By observing the weekly mean GHI data, we note that the diurnal cycle is not the only cycle existing in the trend. Several other cycles at higher frequency are also associated with the trend. After carrying out the spectral analysis of the GHI time series, these higher frequency cycles are clearly identified. The spectral density of the GHI time series of year 2010 is shown in Figure 1 using the Singapore station 5.0 minutes data.

The highest peak in the figure appears at the frequency of \( 1.16 \times 10^{-5} \) Hz, which corresponds to the 24-hour daily cycle. Other peaks in the figure are also
indicated. After obtaining these significant frequency peaks, the Fourier trend is generated based on these peak frequencies:

\[ r_t = \mu + \sum_{i=1}^{n} [a_i \cos(2\pi f_i t) + b_i \sin(2\pi f_i t)] \]  

(1)

where \( t \) is time, \( r_t \) is the fitted trend at different time, \( n \) is the number of significant frequencies and \( a_i, b_i \) are the regression coefficients. By fitting the sinusoids, we eliminate the significant cyclic behaviors (trend) from the GHI time series.

The cosine function model was first used by Kaplanis (2006):

\[ r_t = a + b \cos\left(\frac{2\pi(t - d)}{24}\right) \]  

(3)

where \( a \) and \( b \) are the regression parameters and \( d \) is the peak irradiance hour of the day.

The Gaussian function model (Baig et al., 1991) tries to fit the GHI series with a Gaussian function:

\[ r_t = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-d)^2}{2\sigma^2}\right) \]  

(4)

where \( \sigma \) is the standard deviation of the Gaussian curve (the regression parameter that should be decided by the actual data) and \( d \) is the peak irradiance hour of the day which corresponds to the expected mean of the Gaussian distribution.

We take a random week’s data (the first week of December 2010) to evaluate the performance of these trend models. After subtracting these trends from the weekly mean GHI, the series of residual of the four trends are displayed in Figure 2. It is clear that unlike the Fourier trend model, the other three models fit poorly at the beginning and the end of the day.

Figure 1: Solar irradiance spectral density of Singapore for 2010 using 5 minutes data

Figure 2: Plot of the four series of trend residual

3.1.2. Comparison with other trend models

In order to test the performance of our Fourier trend model, three other popular trend models (namely a high order polynomials model, a cosine function model and a Gaussian function model) are built for comparison.

The high order polynomial model was first published by Al-Sadah et al. (1990) and claimed to be a good fit to the daily GHI trend:

\[ r_t = a + bt + ct^2 \]  

(2)

where \( a, b \) and \( c \) are the regression parameters.

The autocorrelation function (ACF) of the four series of residuals are shown in Figure 3. Autocor-
relation is the linear dependence between observations as a function of time separation between them. Compared with the other three series of trend residual series, the series of Fourier trend residual has the smallest serial correlation, indicating that it is more stationary than the others.

![Figure 3: Autocorrelation function (ACF) of the four series of trend residual](image)

Table 2: KPSS test for the four trend methods

<table>
<thead>
<tr>
<th>Detrend Method</th>
<th>Critical level</th>
<th>Critical value</th>
<th>KPSS value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier trend</td>
<td>0.01</td>
<td>0.739</td>
<td>0.1211</td>
</tr>
<tr>
<td>Polynomial trend</td>
<td>0.01</td>
<td>0.739</td>
<td>0.6973</td>
</tr>
<tr>
<td>Cosine trend</td>
<td>0.01</td>
<td>0.739</td>
<td>1.0772</td>
</tr>
<tr>
<td>Gaussian trend</td>
<td>0.01</td>
<td>0.739</td>
<td>1.3425</td>
</tr>
</tbody>
</table>

dicating that the two series are not stationary. The residual series of Fourier trend and high order polynomial trend are stationary since their KPSS values are smaller than the critical value. However, the KPSS value of the residual series of Fourier trend is much smaller than that of high order polynomial trend, which proves that the residual series of Fourier trend has much higher probability of obtaining stationarity. It is hence that we use Fourier trend model for our further forecasting.

3.2. Forecast equations

There are fifteen sets of forecast equations in this forecast model. The observed time series is denoted by \( y_1, y_2, \ldots, y_n \). The forecast for \( h \) steps ahead based on all the training data up to time \( t \) is denoted by \( \hat{y}_{t+h} \).

\[
\begin{align*}
\text{Level:} & \quad l_t = \alpha A_t + (1 - \alpha) B_t \\
\text{Growth:} & \quad g_t = \beta C_t + (1 - \beta) D_t \\
\text{Seasonal:} & \quad s_t = \gamma E_t + (1 - \gamma) s_{t-m} \\
\text{Forecast:} & \quad \hat{y}_{t+h} = f(l_t, g_t, s_{t-m+h_m})
\end{align*}
\]

where \( l_t \) is the level component at time \( t \), \( g_t \) is the growth component at time \( t \), \( s_t \) is the seasonal component at time \( t \) and \( h_m = [(h-1) \mod m] + 1 \) where \( m \) denotes the number of seasons in the data. In order to apply this method, the initial states \( l_0, g_0 \) and \( s_{1-m}, \ldots, s_0 \), and the smoothing parameters \( \alpha, \beta \) and \( \gamma \), will be estimated from the training data. The values of \( A_t, B_t, C_t, D_t \) and \( E_t \) vary according to the cell which the method belongs to. Table 3 shows the values of \( A_t, B_t, C_t, D_t \), \( E_t \) and the forecast equation (adapted from Hyndman et al. (2002)).

It should be noted that there is an additive damped trend and a multiplicative damped trend in the table. The damped trend acts to dampen the trend with the increasing length of the forecast horizon. The damped trend is applied because it is unrealistic to invariantly use the final estimate of the growth rate at the end of the training data to forecast different horizon. To forecast \( h \) periods ahead, the trend is damped by a factor of \( \phi_h = \phi + \phi^2 + \cdots + \phi^h \).
3.3. State space model

The state space model is a more sophisticated stochastic concept than the exponential smoothing method. The exponential smoothing method can only produce a point forecast but the state space model provides the framework for computing prediction interval and other properties. Every exponential smoothing method corresponds to two possible state space models, one with additive errors and the other with multiplicative errors. Additive errors and multiplicative errors provide different prediction intervals with the same parameter values. Since there are 15 exponential smoothing methods, 30 state space models are deduced accordingly. The state space model for exponential smoothing here is expanded from the general OKS framework (Ord et al., 1997):

\[
y_t = h(x_{t-1}) + k(x_{t-1}) \epsilon_t
\]
\[
x_t = f(x_{t-1}) + g(x_{t-1}) \epsilon_t
\]

where the first equation is called observation equation and the second is called state equation. \(x_t\) is the state vector which in this case defined as \(x_t = (l_t, b_t, s_t, s_{t-1}, \ldots, s_{t-(m-1)})\). We also define \(e_t = k(x_{t-1}) \epsilon_t\) and \(\mu_t = h(x_{t-1})\) so that \(y_t = \mu_t + e_t\). The additive error form can be expressed as \(y_t = \mu_t + \epsilon_t\) where \(\mu_t = y(t-1) + 1\) stands for the one-step point forecast at time \(t - 1\) and \(k(x_{t-1}) = 1\). The multiplicative error form can be expressed as \(y_t = \mu_t(1 + \epsilon_t)\) where \(k(x_{t-1}) = \mu_t\). All of the 15 exponential smoothing methods can be written in the state space model forms which is illustrated in Table 4.

| Table 3: Formulae for recursive calculations and point forecast equations |
|-----------------------------|-----------------------------|-----------------------------|
| N (None) | A (Additive) | M (Multiplicative) |
| \(A_t = y_t\) | \(A_t = y_t - s_{t-m}\) | \(A_t = y_t / s_{t-m}\) |
| \(B_t = l_{t-1}\) | \(B_t = l_{t-1}\) | \(B_t = l_{t-1}\) |
| \(C_t = l_t - l_{t-1}\) | \(C_t = l_t - l_{t-1}\) | \(C_t = l_t - l_{t-1}\) |
| \(D_t = g_{t-1}\) | \(D_t = g_{t-1}\) | \(D_t = g_{t-1}\) |
| \(\hat{y}_{t+h} = l_t + s_{t-m+h_m}\) | \(\hat{y}_{t+h} = l_{t-1} + s_{t-m+h_m}\) | \(\hat{y}_{t+h} = l_{t-1} + s_{t-m+h_m}\) |

3.4. Model selection criterion

Akaike’s Information Criterion (AIC) is used to select the model to be used to perform the point forecasting. The AIC is specified by the following equations:

\[
AIC = L(\hat{\theta}, X_0) + 2P
\]  \hspace{1cm} (11)
\[
L(\theta, X_0) = -\frac{1}{n} \sum_{t=1}^{n} e_t^2 / k^2(x_{t-1})
\]
\[
+ 2 \sum_{t=1}^{n} \log|k(x_{t-1})|\]
\]  \hspace{1cm} (12)

where \(L\) is the maximum likelihood function, \(P\) is the number of parameters in \(\theta, \hat{\theta}\) and \(X_0\) are the estimated of \(\theta\) and \(X_0\) where \(\theta = (\alpha, \beta, \gamma, \phi)\) and initial states \(X_0 = (l_0, g_0, s_0, s_{-1}, \ldots, s_{-m+1})\). The reason...
why the AIC is used instead of mean squared error (MSE) or mean average percentage error (MAPE) is that AIC can select between additive error and multiplicative error models. Since the two kinds of error models produce the same point forecast, MSE or MAPE cannot differentiate between them. However AIC is based on likelihood which meets our requirement. The model which minimizes the AIC will be chosen from among all the 30 models.

4. Experiments, results and discussion

The first set of experiments is based on the solar irradiance time series data from Singapore and the comparison set of experiments is based on the solar irradiance time series data from Colorado. Since it is meaningless to forecast solar irradiance after sunset, we only investigate samples from 07:00 to 19:30. The error type used is the normalized root mean square error (nRMSE) which is described by the following equation.

\[ nRMSE = \frac{\sqrt{E[(X - \hat{X})^2]}}{\sqrt{E[X^2]}} \]  

(13)

All the predicted values are true out-of-sample forecasts, in that they use only data prior to the start of the forecast horizon. The models are estimated over the data prior to the start of the forecast, the points of the next time step is forecast, and the forecast values are compared with the actual ones. The procedure is iterative until forecasts have been run over the daily data set. The SERIS data consists of twelve months from September 2010 to August 2011. Data from September 2010 is used to generate an average daily residual to fit in as the initialization of our prediction model, in order to forecast the time series of the next month, i.e. October 2010. In the forecasting test after October 2012, the data from the previous month is used to form average daily residual initially fitting in the prediction model.

4.1. Forecasting accuracy

To test the accuracy of our exponential smoothing state space (ESSS) model, 5.0 minute average solar irradiance data is used. There are 150 samples from 07:00 to 19:30 for each day. Every 5.0 minute step is predicted and compared with the actual irradiance data to calculate the nRMSE. The test is repeated over the whole month and then the average nRMSE of that month is calculated. For comparison purposes, four other well-established forecasting methods, namely Autoregressive integrated moving average (ARIMA) model, linear exponential smoothing model (LES), simple exponential smoothing (SES) model and random walk (RW), are also tested in the same way. All the test results are listed in Table 5 and Table 6.

Table 5 shows that the ESSS model has generally better prediction accuracy than other time series models in Singapore. Although the SES model performs better in January and March 2011, the ESSS
Table 5: nRMSE (%) comparison among 5 forecasting methods for SERIS data

<table>
<thead>
<tr>
<th>Forecast Period</th>
<th>ESSS</th>
<th>ARIMA</th>
<th>LES</th>
<th>SES</th>
<th>RW</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 October</td>
<td>17.40</td>
<td>20.95</td>
<td>20.47</td>
<td>20.98</td>
<td>23.52</td>
</tr>
<tr>
<td>2010 November</td>
<td>21.59</td>
<td>24.00</td>
<td>23.31</td>
<td>23.69</td>
<td>26.07</td>
</tr>
<tr>
<td>2010 December</td>
<td>27.77</td>
<td>29.77</td>
<td>28.72</td>
<td>29.26</td>
<td>32.20</td>
</tr>
<tr>
<td>2011 January</td>
<td>33.71</td>
<td>33.62</td>
<td>33.89</td>
<td>33.49</td>
<td>37.55</td>
</tr>
<tr>
<td>2011 February</td>
<td>29.15</td>
<td>30.37</td>
<td>30.44</td>
<td>31.00</td>
<td>35.23</td>
</tr>
<tr>
<td>2011 March</td>
<td>33.03</td>
<td>33.60</td>
<td>34.43</td>
<td>32.84</td>
<td>34.85</td>
</tr>
<tr>
<td>2011 April</td>
<td>29.84</td>
<td>31.36</td>
<td>30.55</td>
<td>30.14</td>
<td>35.95</td>
</tr>
<tr>
<td>2011 May</td>
<td>18.55</td>
<td>22.78</td>
<td>22.48</td>
<td>22.37</td>
<td>25.59</td>
</tr>
<tr>
<td>2011 June</td>
<td>19.64</td>
<td>23.44</td>
<td>22.73</td>
<td>22.42</td>
<td>26.00</td>
</tr>
<tr>
<td>2011 July</td>
<td>20.80</td>
<td>23.46</td>
<td>23.05</td>
<td>22.56</td>
<td>27.22</td>
</tr>
</tbody>
</table>

Table 6: nRMSE (%) comparison among 5 forecasting methods for Colorado data

<table>
<thead>
<tr>
<th>Forecast Period</th>
<th>ESSS</th>
<th>ARIMA</th>
<th>LES</th>
<th>SES</th>
<th>RW</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 October</td>
<td>16.86</td>
<td>18.86</td>
<td>18.97</td>
<td>19.98</td>
<td>20.22</td>
</tr>
<tr>
<td>2010 November</td>
<td>12.76</td>
<td>13.80</td>
<td>13.85</td>
<td>13.47</td>
<td>15.49</td>
</tr>
<tr>
<td>2010 December</td>
<td>15.37</td>
<td>17.96</td>
<td>17.64</td>
<td>17.00</td>
<td>19.14</td>
</tr>
<tr>
<td>2011 February</td>
<td>9.75</td>
<td>11.04</td>
<td>10.78</td>
<td>10.46</td>
<td>11.38</td>
</tr>
<tr>
<td>2011 March</td>
<td>15.17</td>
<td>17.24</td>
<td>17.50</td>
<td>16.08</td>
<td>17.08</td>
</tr>
<tr>
<td>2011 April</td>
<td>15.12</td>
<td>17.21</td>
<td>17.10</td>
<td>16.98</td>
<td>17.86</td>
</tr>
<tr>
<td>2011 May</td>
<td>18.30</td>
<td>19.22</td>
<td>19.16</td>
<td>19.02</td>
<td>20.22</td>
</tr>
<tr>
<td>2011 June</td>
<td>19.86</td>
<td>20.90</td>
<td>20.91</td>
<td>20.72</td>
<td>21.91</td>
</tr>
<tr>
<td>2011 July</td>
<td>29.04</td>
<td>28.93</td>
<td>29.14</td>
<td>29.02</td>
<td>29.19</td>
</tr>
<tr>
<td>2011 August</td>
<td>24.25</td>
<td>24.40</td>
<td>24.24</td>
<td>24.06</td>
<td>24.00</td>
</tr>
</tbody>
</table>

The ESSS model is behind by less than 0.5%. By comparing with the benchmark RW model, we infer that the solar irradiance variability is high from December 2010 to April 2011, since the RW nRMSEs are all above 30%. The performance of the ESSS model may decrease in these months. In other months, however, the ESSS model outperforms other forecasting models.

Table 6 shows that the ESSS model performs well compared with other time series models in Colorado, US. The solar irradiance variability in this area is generally smaller than that in Singapore due to the huge difference in latitude. The peak variability occurs in summer months like July and August 2011, when the ESSS model accuracy is slightly less accurate than the other models.

In general, the ESSS model has better forecasting accuracy over other time series forecasting models in both Singapore and Colorado, US. The performance of different models can be influenced by the solar irradiance variability in the region.

4.2. Forecasting interval

The state space model of exponential smoothing provides the framework for computing prediction intervals. Since the prediction confidence interval is useful for power station managers and grid operators, the accuracy of the forecasting confidence interval is tested. In this test, accuracy is calculated by the ratio of forecasting points fall within the confidence interval, to the total number of forecasting points. The result is shown in Table 7. The theoretical 95% confidence interval usually gives us forecasting accuracy around 90%, which is an acceptable forecasting range for industrial application.

Figure 4 and Figure 5 illustrate the forecasting interval examples for a day with relatively smooth irradiance and a relatively intermittent day. It is seen that large variations of the solar irradiance would expand the confidence interval which is expected.

Table 7: ESSS 95% confidence interval accuracy (%)

<table>
<thead>
<tr>
<th>Forecast Period</th>
<th>Singapore</th>
<th>Colorado</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 October</td>
<td>92.67</td>
<td>91.45</td>
</tr>
<tr>
<td>2010 November</td>
<td>93.33</td>
<td>94.67</td>
</tr>
<tr>
<td>2010 December</td>
<td>94.00</td>
<td>86.67</td>
</tr>
<tr>
<td>2011 January</td>
<td>91.67</td>
<td>89.00</td>
</tr>
<tr>
<td>2011 February</td>
<td>87.33</td>
<td>88.00</td>
</tr>
<tr>
<td>2011 March</td>
<td>86.67</td>
<td>91.87</td>
</tr>
<tr>
<td>2011 April</td>
<td>91.67</td>
<td>89.80</td>
</tr>
<tr>
<td>2011 May</td>
<td>89.02</td>
<td>96.00</td>
</tr>
<tr>
<td>2011 June</td>
<td>89.72</td>
<td>87.25</td>
</tr>
<tr>
<td>2011 July</td>
<td>91.02</td>
<td>87.53</td>
</tr>
<tr>
<td>2011 August</td>
<td>92.25</td>
<td>91.33</td>
</tr>
</tbody>
</table>

Figure 4: The ESSS forecast plot for relatively smooth day.

4.3. Forecasting horizon

In the previous sections, 5.0 minute average solar irradiance data is used to test the performance of
different time series forecasting models. In real applications, different time series forecast horizons must be tested in order to meet various forecasting requirements from the power grid managers. It is also useful to determine the best measurement dataset to be used at a certain forecast horizon. We have made used of the SERIS data ranging from 5.0 minute averages to 60 minute averages, to forecast from 5.0 minutes up to 60 minutes. The nRMSE maps of four different time series models is shown in Figure 6. The nRMSE value is displayed using colors. Clearly, the ESSS model exhibits the best performance when using less than 20 minute average data to predict less than 20 minutes ahead.

5. Conclusions

We apply a novel time series technique to forecast high resolution solar irradiance data using an exponential smoothing state space (ESSS) model. To stationarize the solar irradiance data before applying linear time series models, we propose an original Fourier trend model and compare the performance with other popular trend models using residual analysis and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationary test. The KPSS values of the residual series of cosine trend and Gaussian trend are higher than the critical value, indicating that the two series are not stationary. The residual series of Fourier trend and high order polynomial trend are stationary since their KPSS values are smaller than the critical value. However, the KPSS value of the residual series of Fourier trend is much smaller than that of high order polynomial trend, which proves that the residual series of Fourier trend has much higher probability of obtaining stationarity. It is hence that we use Fourier trend model for our further forecasting. Using the optimized Fourier trend, ESSS model is implemented to forecast the stationary residual series of Singapore and Colorado US. The climate of Colorado is temperate semi-arid which is a good comparison with humid tropical climate of Singapore. To compare the performance with other time series forecasting models, autoregressive integrated moving average (ARIMA), linear exponential smoothing (LES), simple exponential smoothing (SES) and random walk (RW) is also tested using the same data samples. The simulation results show that the ESSS model has generally better performance than other forecasting models. Finally in order to discuss the reliability of the forecasting model in real application, a complementary study on the forecasting 95% confidence interval and forecasting horizon of the ESSS model is proposed. The ESSS model 95% confidence interval gives us forecasting ac-

Figure 5: The ESSS forecast plot for relatively spiky day.

Figure 6: The forecast horizon nRMSE map
accuracy around 90%, which is an acceptable forecasting range for industrial application.

The application can be extended to two dimensional irradiance forecast which we aim to achieve in the future work. 25 monitoring stations are going to be deployed in Singapore. With these 25 monitoring stations, after the forecast for each station using ESSS model, we can generate a forecast irradiance map using the 2D spline interpolated trend surfaces. The 2D spline is an extension of the one dimensional cubic spline interpolation. In the one dimensional cubic spline interpolation, the first and second order derivatives at each point are minimized. Bookstein (1989) states the generalization of this measure for a two-dimensional problem:

\[ J_f = \int \int_{R^2} \left[ \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 \right] \, dx \, dy \]  

(14)

The 2D spline is proportional to the bending energy of an idealized thin plate of infinite extent.

This two dimensional irradiance map (Figure 7) shows the big picture of the solar irradiance situation around Singapore, which is essential to assess solar potential around Singapore.

Figure 7: Sample 2D forecast irradiance map using 10 monitoring stations

Acknowledgment

This work is supported by project “Optimization of photovoltaic electricity generation in tropical power grids through irradiance forecasting and system monitoring”. The project is funded by the National Research Foundation of Singapore under the Clean Energy Research Programme Grant No. NRF2010EWT-CERP001-030.

References

Mellit, A., Pavan, A.M., 2010a. A 24-h forecast of solar irradiance using artificial neural network: Application for per-