Irradiance exponential smoothing using decompositions

Dazhi Yang[a,b,*], Vishal Sharma[a], Zhen Ye[a], Lihong Idris Lim[c]

[a]*Solar Energy Research Institute of Singapore (SERIS), National University of Singapore, 7 Engineering Drive 1, Block E3A, #06-01, 117574, Singapore*
[b]*Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Block E4, #05-45, 117583, Singapore*
[c]*Department of Electronic Systems, University of Glasgow (Singapore), 535 Clementi Road, 599489, Singapore*

Abstract

Time series methods are frequently used in solar irradiance forecasting when two dimensional cloud information provided by satellite or sky camera is unavailable. Furthermore, satellite and sky camera based methods lose resolution at a 1–h time horizon. Exponential smoothing (ETS) has received great attention in the recent years due to the invention of its state space formulation. We explore 1–h ahead ETS forecasting of solar irradiance in this paper. Several knowledge based decompositions are considered to improve the forecast accuracy and computation speed.

Three methods are proposed. The first method considers an additive seasonal–trend decomposition prior to the use of ETS. In such a way, the ETS state space is reduced, thus facilitates online forecasting applications. The second method decomposes the global horizontal irradiance (GHI) time series into a direct component and a diffuse component. These two components are used as forecasting model inputs separately; the results are recombined through the closure equation. We consider the time series of cloud cover index in the third method. ETS is first applied to the cloud cover time series to produce the forecast. The forecast cloud cover is then used to reconstruct GHI through regressions. It is found that the third method performs the best among three methods and all proposed methods outperforms persistence.

Keywords: exponential smoothing, time series, forecast, irradiance

*Corresponding author. Tel.: +65 9159 0888.
Email address: seryangd@nus.edu.sg, yangdazhi.nus@gmail.com (Dazhi Yang)

List of symbols.

T	trend component
S	seasonal component
E	error component
ℓ_t	level term in trend component at time t
b_t	growth term in trend component at time t
ϕ	damping parameter
ϕ_h	$\phi + \phi^2 + \cdots + \phi^h$
y_t	time series $\{y_t\}$ at time t
$\widehat{y}_{t+h\|t}$	h–step–ahead forecast
$\boldsymbol{x}_t$	state vector at time t
ε_t	white noise at time t
$\mu_{t+h\|t}$	conditional expectation of y_{t+h} given $\boldsymbol{x}_t$
$\mathcal{W}$, $\mathcal{R}$, $\mathcal{F}$ and $\mathcal{G}$	functions of the state vectors
α, β, γ and ϕ	ETS model parameters
$\boldsymbol{\theta}$	vector notation for α, β, γ and ϕ
$\mathcal{L}$	likelihood function
G	global horizontal irradiance
$G^{(\mathrm{ext})}$	extraterrestrial horizontal irradiance
I	direct normal irradiance
D	diffuse horizontal irradiance
$\mathcal{CC}$	cloud cover
θ_z	zenith angle
a_0, a_1, a_2 and a_3	cloud cover regression parameters

1. Introduction

The needs for intermittent renewable generation forecasting spawned a large body of literature in the field of energy over the recent years. The utilization of solar energy in particular, requires forecasting at a variety of spatial and temporal scales. The forecast methodologies thus vary for different forecast horizons.

In a spatio–temporal context, the creation, propagation, evolution and distinction of clouds are usually considered, regardless of the choice of forecast horizons and methodologies. In a small spatial scale (a few meters apart), pyranometer networks with designed geometries are used to retrieve

the information on cloud movement speed and direction [1, 2]. Given the geographical proximity, the forecast horizon of this class of methods is ranged from 20 seconds to 3 minutes [3]. On the other extreme of the forecast spatio–temporal scale, satellite images can be used to detect, classify the clouds [4]. However, obtaining the future clouds information does not complete the forecast; ground level irradiance must be derived from the forecast cloud images, and the mapping is not trivial [5].

One way to circumvent the cloud to irradiance conversion is to consider the stochastic nature of the irradiance. Artificial intelligence and statistics are the main tools for such application; they both perform inputs to outputs mapping. Artificial neural networks (ANN) represent a significant bulk in the literature of irradiance forecasting [e.g., 6, 7, 8]. However, the evident disadvantage of ANN owing to its black–box nature poses limitations on the intuitive understanding of the irradiance temporal process. On the other hand, statistical forecasting models emphasize on evolution and dependence; they are described by parameters [9]. We therefore focus on the application of statistical forecasting models in this paper.

Among many famous time series models such as the autoregressive moving average (ARMA) model, exponential smoothing is much overlooked. Although the methods have been around since the 1950s [10], the framework of stochastic formulation, likelihood calculation and model selection was not developed until the publication of two key papers [11, 12]. Since then, exponential smoothing receive attention in many developed areas including, including call centres, power grid, financial markets and inventory control [13, 14, 15, 16, 17]. However, its applications in the field of solar energy are very limited, and most of the works only considered the univariate irradiance time series as the model input [18].

Although exponential smoothing has the multivariate form [19], to realize the vector exponential smoothing, a network of irradiance monitoring stations are required to sample the time series of lattice process [20]. In addition, the selection of relevant spatial neighbors needs attention, as irrelevant information not only increases the model complexity but also introduce additional errors [21]. We therefore focus on the univariate exponential smoothing here, however, the model inputs are not necessarily be the global horizontal irradiance (GHI).

We consider knowledge based decompositions in section 3. In particular, GHI time series $\{G_t\}$, $t \in \{0, 1, \cdots\}$, is decomposed into the diffuse horizontal irradiance (DHI) and the direct normal irradiance (DNI) time series

$\{D_t\}$ and $\{I_t\}$, $t \in \{0, 1, \cdots\}$. Forecasts are produced separably on the two decomposed time series; GHI forecasts are then reconstructed through the closure equation:

$$\widehat{G} = \widehat{I} \cos\theta_z + \widehat{D} \tag{1}$$

where θ_z is the zenith angle, a known quantity; the $\widehat{G}$, $\widehat{I}$ and $\widehat{D}$ denote the forecasts for the respectives. We also propose a more sophisticate decomposition model, namely, GHI as a function of zenith angle and cloud cover ($\mathcal{CC}$):

$$\widehat{G} = f(\theta_z, \widehat{\mathcal{CC}}) \tag{2}$$

where $\widehat{\mathcal{CC}}$ denotes the forecast cloud cover on a discrete scale of 0 to 10 (see below). Before we move on to these, the univariate exponential smoothing is briefly introduced in section 2.

The third typical meteorological year (TMY3) data from National Renewable Energy Laboratory is used in this work. The data is freely available online at `http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/tmy3/`. The user manual of the datasets is found at the same website. The TMY3 dataset provides an annual dataset that holds hourly meteorological values that typify conditions at a particular site over a longer period of time, such as 30 years. There are more than one thousand sites in the TMY3 dataset; for demonstration of the proposed forecasting techniques, we choose a class one site, namely, the San Diego Lindbergh Field with USAFN number 722900.

2. Univariate exponential smoothing

2.1. Time series components

Exponential smoothing considers time series as a combination of three components, namely, the trend (T), seasonal (S) and error (E) components. The trend components consists of another combination of a level term (ℓ) and a growth term (b). When we describe the forecast trend T_h over the next h time periods, ℓ and b can be combined in the following 5 ways:

$$\begin{aligned}
\text{None}: &\quad T_h = \ell \\
\text{Additive}: &\quad T_h = \ell + bh \\
\text{Additive damped}: &\quad T_h = \ell + (\phi + \phi^2 + \cdots + \phi^h)b \\
\text{Multiplicative}: &\quad T_h = \ell b^h \\
\text{Multiplicative damped}: &\quad T_h = \ell b^{(\phi+\phi^2+\cdots+\phi^h)}
\end{aligned}$$

where $0 < \phi < 1$ is a damping parameter. Beside the trend component, the seasonal components can be additive ($T+S$), multiplicative ($T \times S$) or none. This gives rise to the 15 combinations of time series components:

Trend	Seasonal		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A_d (Additive damped)	A_d,N	A_d,A	A_d,M
M (Multiplicative)	M,N	M,A	M,M
M_d (Multiplicative damped)	M_d,N	M_d,A	M_d,M

Furthermore, the formulae of these 15 models are shown in Table 1. To understand this seemingly complex table, we consider the state space model.

2.2. State space model

Suppose the errors in an exponential state space model (abbreviated as ETS, which refers to error, trend and seasonal components) can be either additive or multiplicative, we use the triplet (E,T,S) to denote an ETS model. For example, ETS(A,M,N) denotes an ETS model with multiplicative trend and additive errors without seasonal component. Each of the models in Table 1 can be written as a state space model, we use ETS(A,A,N) to illustrate; this model is also known as linear exponential smoothing or Holt's linear method [22].

Let y_t denotes the observation at time t and $\boldsymbol{x}_t$ denotes the state vector contained unobserved components ℓ_t, b_t and s_t, the state space can be written as:

$$y_t = \mathcal{W}(\boldsymbol{x}_{t-1}) + \mathcal{R}(\boldsymbol{x}_{t-1})\varepsilon_t; \quad (3)$$

$$\boldsymbol{x}_t = \mathcal{F}(\boldsymbol{x}_{t-1}) + \mathcal{G}(\boldsymbol{x}_{t-1})\varepsilon_t, \quad (4)$$

where $\mathcal{W}$, $\mathcal{R}$, $\mathcal{F}$ and $\mathcal{G}$ are functions of the state vector; ε_t is a while noise series. Eqns. (3) is the measurement equation and Eqn. (4) is the state equation. The above state space representation is in the innovations form (or the single source of error form), meaning that both the errors in the measurements and systems are represented by ε_t, which allows non–linear formulations [12].

Let $\mu_{t|t-1} = \mathbb{E}(y_t|\boldsymbol{x}_{t-1})$ be the conditional expectation of the current observation and define $\mu_{t|t-1} \equiv \mu_t$. Following the formulae in second row

first column of Table 1, $\mu_t = \widehat{y}_{t|t-1} = \ell_{t-1} + b_{t-1}$ for $h = 1$. Let $\varepsilon_t = y_t - \mu_t$, we write:

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t \tag{5}$$

We can thus write:

$$y_t = \begin{pmatrix} 1 & 1 \end{pmatrix} \boldsymbol{x}_{t-1} + \varepsilon_t; \tag{6}$$

$$\boldsymbol{x}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \boldsymbol{x}_{t-1} + \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \varepsilon_t \tag{7}$$

where $\boldsymbol{x}_t = (\ell_t, b_t)'$, and $\beta = \alpha\beta^*$. These equations give the state space representation of the ETS(A,A,N) model. Complete tables of state space equations for both additive and multiplicative errors using the general state vector

$$\boldsymbol{x}_t = \begin{pmatrix} \ell_t & b_t & s_t & s_{t-1} & \cdots & s_{t-m+1} \end{pmatrix}'$$

can be found in reference [19].

2.3. Parameter estimation and model selection

To automate the model selection from 30 candidates (15 models shown in Table 1 with either additive or multiplicative errors) for online forecasting purpose, each model is fitted using the same set of data. Model parameters $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \phi)'$ and the initial state vector $\boldsymbol{x}_0$ are estimated using the maximum likelihood (ML) estimation through *minimizing*:

$$\mathcal{L}^*(\boldsymbol{\theta}, \boldsymbol{x}_0) = n \log \left(\sum_{t=1}^{n} \varepsilon_t^2 \right) + 2 \sum_{t=1}^{n} \log |\mathcal{R}(\boldsymbol{x}_{t-1})| \tag{8}$$

The reason we use $\mathcal{L}^*$ instead of $\mathcal{L}$ is made apparent in Appendix A. Let $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{x}}_0$ be ML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{x}_0$, then for each candidate model, the Akaike's information criterion (AIC) can be calculated via:

$$\text{AIC} = \mathcal{L}^*(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{x}}_0) + 2q \tag{9}$$

where q is the total number of parameters in the model. The ETS model which *minimizes* the AIC is selected.

3. Knowledge based decompositions

If the atmosphere is cloud free, solar irradiance time series on any day would appear bell–shaped owing to the Earth's self rotation. This bell–shaped curve is known as the clear sky model. In irradiance forecasting, clear sky models are often used to detrend the irradiance time series [3]. A physical clear sky model considers atmosphere parameters such as the aerosol optical depth, column ozone and precipitable water [23], which are only measured infrequently and at few locations. Empirical clear sky models are therefore commonly used. However, these models are location dependent [24, 25]. More general detrend method that does not require irradiance modeling should be considered when the forecast algorithm is applied at an arbitrary location.

As a matter of fact, detrending the irradiance time series is not necessary as the exponential smoothing considers the seasonal components. In our case, the seasonal components are diurnal. The seasonal components can be assumed to be additive in irradiance time series, the third column of Table 1 can therefore be relaxed. Even though, fitting an additive seasonal model greatly increases the computation burden in terms of parameter estimation and likelihood calculation. Especially for online forecasting, to produce each point forecast, the algorithm needs to run through the search space of 20 (30 minus the 10 multiplicative error models) models. We consider using a seasonal–trend decomposition procedure based on loess (abbreviated as STL) [26] to further relax the second column of Table 1. The use of STL here is novel, and the decomposition is not considered in references [18, 5].

3.1. STL decomposition

Consider an hourly irradiance time series over a period of one month as shown in the top panel of Figure 1. The GHI time series can be decomposed into the seasonal, trend and remainder components using STL. STL is an iterative filtering procedure. In each iteration of the STL, the moving average smoothing and the locally–weighted regression (LOESS) smoothing [27] are used multiple times. We do not repeat the detailed procedures here; interested readers can refer to the original publications. The purpose of STL is to identify the seasonal component, the residual (trend and remainder) is used as input to the ETS models. Once the forecast is made using the ETS, seasonal component at the forecast time step is added back to the estimate. We refer this procedure as model 1.

3.2. Decomposition using the closure equation

Model 2 is very similar to model 1. Instead of using GHI time series as the input to the STL, DNI and DHI measurements are fed into the STL separately. As a result, two time series of seasonal components can be identified respectively. The forecasts are then made using the two residual series separately. After the seasonal components are added back to the estimates, forecast GHI, $\widehat{G}$, is reconstruct using the forecast DHI and forecast DNI through the closure equation shown in Eqn. (1).

3.3. Decomposition using cloud cover

In the TMY3 dataset, the opaque sky cover measurements are recored on a discrete scale of 0 to 10. We use $\mathcal{CC}$ to denote the data. $\mathcal{CC} = 0$ refers to a clear sky situation; $\mathcal{CC} = 10$ refers to a completely opaque sky. There is no doubt that the irradiance level reaching the Earth's surface is related with the cloud situation, however, the relationship is governed by complex physics. Without detailed measurements of the physical parameters, we seek the simplistic regressive models to link the cloud cover measurements to GHI. Beside the cloud cover, zenith angle is also responsible for the GHI transient. To obtain the function $f(\cdot)$ in Eqn. (2), we examine the scatter plot. Figure 2 shows the scatter plot of the triplet G, $\cos\theta_z$ and $\mathcal{CC}$. Their relationship is immediately seen.

The function form of $f(\cdot)$ can be subjective, despite the physical drawbacks (the choice is not supported by any physical law), we consider the polynomial models as they are mathematically flexible in curve fitting. 11 polynomial models can be fitted using the subsets of data at 11 $\mathcal{CC}$ values. Furthermore, we hypothesize that third order polynomials:

$$\widehat{G} = a_0 + a_1 \cos\theta_z + a_2 \cos^2\theta_z + a_3 \cos^3\theta_z \tag{10}$$

can suffice the fitting. The use of higher order polynomials, i.e., order four or five, may result in statistically insignificant regression coefficients. In linear regression, the p–value tests the null hypothesis that a coefficient is equal to zero. Table 2 shows the regression coefficient and standard errors for higher order polynomial models at the selected $\mathcal{CC}$ values.

It is clear from Table 2 that polynomial models of order four and five produce many insignificant regression coefficients while the number of insignificant coefficients in the third order polynomials is fewer. Although some quadratic models at particular $\mathcal{CC}$ values may better describe the data,

third order polynomial model is selected to keep our choice of model consistent for all $\mathcal{CC}$ values. The fitted coefficients are shown in Table 3; the computation is based on the San Diego TMY3 dataset.

At this stage, $f(\cdot)$ in Eqn. (2) is selected to be the third order polynomials. To produce a GHI forecast using this decomposition, the cloud cover time series is used as the input to the ETS models. The forecast cloud cover $\widehat{\mathcal{CC}}$ is discretized. The GHI forecast can be read from the "look–up table" of the polynomials. For example, if $\widehat{\mathcal{CC}} = 0$ and $\theta_z = 60$, then

$$\widehat{G} = -5.34 + 597.77 \times 0.5 + 1012.16 \times 0.5^2 - 585.59 \times 0.5^3 = 473.4\text{W/m}^2.$$

We refer this method as model 3.

4. Results and discussion

Composition of the TMY3 seek data that typify the long–term climate behavior, as a result, data for each month are from different years. For example, in the case of the San Diego dataset, January data are from the year 1996 and February data are from 1991. Since the data are discontinuous between months, each month should be evaluated separately. Following the proposed model in section 3.1, below shows the work flow for the empirical study using model 1:

1. Select January data form the TMY3 dataset.
2. One week hourly time series $\{G_\tau\}$, $\tau = \{t - (24 \times 7) + 1, t - (24 \times 7) + 2, \cdots, t-1, t\}$, $t \geq 24 \times 7$, is used as training data for point forecast $\widehat{G}_{t+1}$.
3. STL is applied on $\{G_\tau\}$ to identify the seasonal component, $\{S_\tau\}$, in the training time series. Residual time series $\{R_\tau\}$ is obtained by subtracting the seasonal component from $\{G_\tau\}$.
4. $\{R_\tau\}$ is used as the ETS input. Parameter fitting and model selection lead to $\widehat{R}_{t+1}$, the forecast residual at time $t+1$.
5. $\widehat{G}_{t+1} = S_{t+1-24} + \widehat{R}_{t+1}$, where S_{t+1-24} is the seasonal component at the forecast hour one day ago.
6. Go back to 2 until data points from January are used up, i.e., when $t = 24 \times 31 - 1$. The "-1" is to make sure that all point forecasts can be validated using actual data.
7. Repeat for other months.

The work flow for models 2 and 3 are similar to model 1 with identical length of training data. The results of the empirical studies are shown in Table 4. Three error metrics are used, namely, the mean bias error (MBE), normalize root mean square error (nRMSE) and the expanded uncertainty (U_{95}). The errors are calculated as follows:

$$\text{MBE} = \frac{1}{N}\sum_{n=1}^{N}(\widehat{G}_n - G_n) \tag{11}$$

$$\text{nRMSE} = \frac{\sqrt{\dfrac{1}{N}\displaystyle\sum_{n=1}^{N}(\widehat{G}_n - G_n)^2}}{\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} G_n} \tag{12}$$

$$U_{95} = k \times \frac{\sqrt{\dfrac{1}{N}\displaystyle\sum_{n=1}^{N}(\widehat{G}_n - \text{MBE} - G_n)^2}}{\dfrac{1}{N}\displaystyle\sum_{n=1}^{N} G_n} \tag{13}$$

where k is the coverage factor, equal to 1.96 for a 95% confidence level. All errors are calculated using daylight hours only.

To benchmark our proposed models, persistence forecast results are also shown in Table 4. Two types of persistence models are considered. The first persistence model assumes the forecast GHI is identical to the current observation, i.e., $\widehat{G}_{t+1} = G_t$. The second persistence model considers the extraterrestrial horizontal irradiance $G^{(\text{ext})}$. Reference [3] note that the use of $G^{(\text{ext})}$ can greatly improve the persistence forecast:

$$\widehat{G}_{t+1} = \frac{G_t}{G_t^{(\text{ext})}} \times G_{t+1}^{(\text{ext})} \tag{14}$$

as $G^{(\text{ext})}$ is a known quantity describing the bell–shaped irradiance transient just above the atmosphere. These two persistence models are noted as Pers 1 and Pers 2 in Table 4, respectively.

Table 4 shows that the forecasts made using model 3 are more accurate in general; and all proposed models outperform the persistence models. It is

also found that model 3 performs worse in 1989 October and 1980 November. This is believed due to the degeneracies in the cloud cover measurements. To examine the degeneracies, scatter plots of the measured and modeled GHI are made, as shown in Figure 4. To exclude the forecast error, $\mathcal{CC}$ measurements are used instead of $\widehat{\mathcal{CC}}$ to reconstruct GHI, i.e.,

$$\tilde{G} = f(\theta_z, \mathcal{CC}) \tag{15}$$

where $\tilde{G}$ is the modeled GHI. November (model 2 performs best) and December (model 3 performs best) data are used for the comparison. It is clear from Figure 4 that the November data is more degenerate than the December ones, as the November scatter has a wider spread.

5. Conclusion

Table 1: Formulae for recursive calculations and point forecasts (adopted from [19] without modification). In each case, $\widehat{y}_{t+h|t}$ denotes an h–steps–ahead forecast using information up to time t; ℓ_t denotes the level; b_t denotes the slope; s_t denotes the seasonal components and m denotes the number of seasons in the data. α, β^*, γ and ϕ are coefficients; $\phi_h = \phi + \phi^2 + \cdots \phi^h$ and $h_m^+ = [(h-1) \bmod m] + 1$.

Trend	Seasonal		
	N	A	M
N	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$ $\widehat{y}_{t+h\|t} = \ell_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t / \ell_{t-1}) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$
A	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $\widehat{y}_{t+h\|t} = \ell_t + hb_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$
A_d	$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $\widehat{y}_{t+h\|t} = \ell_t + \phi_h b_t$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$
M	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $\widehat{y}_{t+h\|t} = \ell_t b_t^h$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t / (\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$
M_d	$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $\widehat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$	$\ell_t = \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t / \ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t / (\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$ $\widehat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$

Table 2: Regression coefficients for higher order polynomial models at various cloud cover conditions. Standard error of each coefficient is shown in the brackets. Insignificant coefficients (without stars) have a p–value greater than 0.1.

	$\mathcal{CC} = 0$	$\mathcal{CC} = 2$	$\mathcal{CC} = 4$	$\mathcal{CC} = 6$	$\mathcal{CC} = 8$	$\mathcal{CC} = 10$
	$\widehat{G} = a_0 + a_1 \cos\theta_z + a_2 \cos^2\theta_z + a_3 \cos^3\theta_z$					
a_0	-5.338^{*}	-1.546	11.248	10.409	10.215	4.268
	(3.092)	(9.000)	(17.525)	(25.004)	(25.631)	(5.735)
a_1	597.771^{***}	504.982^{***}	379.305^{***}	178.540	206.202	214.268^{***}
	(24.994)	(76.770)	(142.157)	(220.199)	(216.746)	(53.276)
a_2	$1,012.155^{***}$	$1,051.806^{***}$	955.749^{***}	$1,388.503^{***}$	$1,128.554^{**}$	274.893^{**}
	(55.471)	(176.310)	(317.074)	(516.697)	(490.858)	(131.938)
a_3	-585.592^{***}	-601.851^{***}	-417.631^{**}	-760.813^{**}	-685.354^{**}	-56.631
	(35.269)	(114.271)	(202.661)	(337.872)	(319.228)	(92.290)
	$\widehat{G} = a_0 + a_1 \cos\theta_z + a_2 \cos^2\theta_z + a_3 \cos^3\theta_z + a_4 \cos^4\theta_z$					
a_0	16.619^{***}	17.885	21.153	13.982	29.888	4.268
	(4.026)	(11.400)	(23.058)	(32.747)	(31.437)	(5.735)
a_1	221.909^{***}	124.451	196.482	106.047	-170.728	214.268^{***}
	(51.644)	(158.712)	(310.653)	(480.573)	(410.881)	(53.276)
a_2	$2,576.822^{***}$	$2,700.700^{***}$	1,725.778	1,709.654	$2,809.690^{*}$	274.893^{**}
	(196.996)	(628.518)	(1,205.579)	(1,960.306)	(1,632.621)	(131.938)
a_3	$-2,914.380^{***}$	$-3,108.543^{***}$	$-1,566.108$	$-1,257.619$	$-3,303.000$	-56.631
	(283.966)	(924.877)	(1,746.456)	(2,944.048)	(2,445.519)	(92.290)
a_4	$1,133.258^{***}$	$1,236.103^{***}$	558.534	248.008	1,316.892	
	(137.163)	(452.655)	(843.594)	(1,459.897)	(1,219.782)	
	$\widehat{G} = a_0 + a_1 \cos\theta_z + a_2 \cos^2\theta_z + a_3 \cos^3\theta_z + a_4 \cos^4\theta_z + a_5 \cos^5\theta_z$					
a_0	20.598^{***}	17.650	17.136	45.254	-12.254	13.545
	(5.282)	(14.925)	(30.497)	(41.884)	(38.241)	(8.468)
a_1	124.138	130.975	300.344	-762.622	1,083.439	3.165
	(98.647)	(311.252)	(601.401)	(871.294)	(775.046)	(174.196)
a_2	$3,201.539^{***}$	2,657.411	1,061.508	7,513.999	$-5,869.688$	1,436.565
	(572.040)	(1,883.632)	(3,505.753)	(5,238.992)	(4,841.553)	(1,132.182)
a_3	$-4,491.491^{***}$	$-2,997.256$	105.799	$-16,374.670$	19,495.460	$-2,487.405$
	(1,385.226)	(4,657.074)	(8,466.085)	(12,993.640)	(12,228.430)	(3,033.878)
a_4	$2,844.085^{*}$	1,114.034	$-1,251.649$	17,076.740	$-24,017.160^{*}$	2,103.519
	(1,477.148)	(5,026.715)	(9,008.046)	(14,165.140)	(13,373.430)	(3,534.125)
a_5	-667.229	47.964	705.210	$-6,708.282$	$10,047.260^{*}$	-621.936
	(573.605)	(1,967.068)	(3,493.865)	(5,616.550)	(5,282.053)	(1,482.504)
Observations	1,483	309	238	134	157	701

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Cloud cover regression coefficients a_0, a_1, a_2 and a_3 for San Diego data, $\mathcal{CC}$ is opaque cloud cover.

$\mathcal{CC}$	a_0	a_1	a_2	a_3
0	−5.34	597.77	1012.16	−585.59
1	−1.78	495.71	1161.59	−707.07
2	−1.55	504.98	1051.81	−601.85
3	0.83	472.14	981.80	−525.60
4	11.25	379.31	955.75	−417.63
5	2.27	358.34	993.68	−493.17
6	10.41	178.54	1388.50	−760.81
7	27.52	13.04	1617.38	−909.58
8	10.21	206.20	1128.55	−685.35
9	11.84	147.15	919.35	−520.80
10	4.27	214.27	274.89	−56.63

Table 4: Error comparison use various proposed forecasting models. Two persistence models (using raw/detrended GHI) are included for benchmarking. Mean bias error (in W/m^2), normalized root mean square error (in %) and the expanded uncertainty at 95% confidence interval (in %) are the error metrics.

Forecast Period	Forecasting models				
	Pers 1	Pers 2	Model 1	Model 2	Model 3
	Mean bias error [W/m^2]				
1996 January	-0.06	-7.86	3.39	3.59	-7.44
1991 February	2.00	-4.63	4.44	5.32	-5.20
1979 March	0.06	-13.56	1.14	0.45	-10.32
1987 April	-3.70	0.86	9.20	7.72	-20.02
1989 May	1.46	-15.55	0.99	0.16	-9.94
1996 June	0.67	-19.61	1.82	3.35	3.39
1989 July	-0.49	-23.16	-0.67	-1.16	-9.34
1978 August	-5.12	-14.20	3.32	-0.01	-4.62
1990 September	4.74	-16.02	5.43	3.35	-11.55
1989 October	-3.06	-15.60	5.04	-0.30	-22.18
1980 November	3.39	1.18	2.10	-1.75	-24.24
1997 December	5.48	-6.12	2.25	-1.28	-12.21
	Normalized root mean square error [%]				
1996 January	42.02	26.40	25.48	25.75	**22.09**
1991 February	37.10	21.88	22.41	20.61	**18.81**
1979 March	34.18	18.31	20.04	18.35	**15.62**
1987 April	30.90	19.31	17.21	15.74	**14.40**
1989 May	36.10	24.42	23.48	23.07	**22.49**
1996 June	35.60	25.08	22.93	23.73	**20.91**
1989 July	32.17	19.64	18.49	18.76	**18.47**
1978 August	31.26	17.10	14.33	14.30	**12.36**
1990 September	35.25	21.17	20.43	20.28	**19.09**
1989 October	37.68	22.51	22.50	**20.71**	22.00
1980 November	36.67	23.54	17.80	**16.65**	20.31
1997 December	37.96	16.87	19.51	16.64	**14.76**
	Expanded uncertainty (U_{95}) [%]				
1991 February	72.70	42.81	43.85	40.30	36.76
1979 March	66.99	35.37	39.28	35.96	30.20
1987 April	60.55	37.84	33.52	30.68	26.95
1989 May	70.75	47.45	46.02	45.22	43.87
1996 June	69.77	48.63	44.95	46.48	40.96
1989 July	63.06	37.59	36.23	36.77	36.03
1978 August	61.24	33.12	28.06	28.03	24.17
1990 September	69.06	40.99	39.97	39.72	37.08
1989 October	73.84	43.45	44.03	40.59	41.47
1980 November	71.85	46.13	34.87	32.61	37.01
1997 December	74.32	32.85	38.21	32.61	27.78

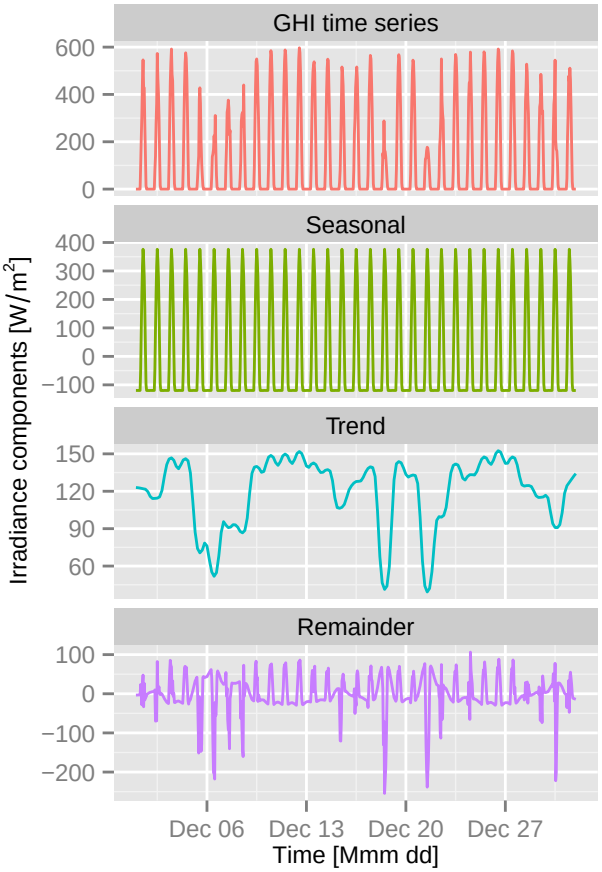

Figure 1: STL decomposition for hourly global horizontal irradiance (GHI) time series. Top panel shows the GHI for San Diego 1997 December. The other three panels show the decomposed seasonal, trend and remainder components respectively.

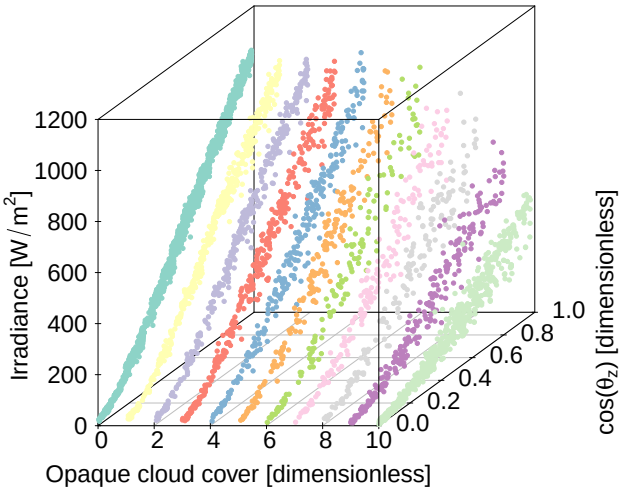

Figure 2: Global horizontal irradiance as functions of cloud cover and cosine of the zenith angle.

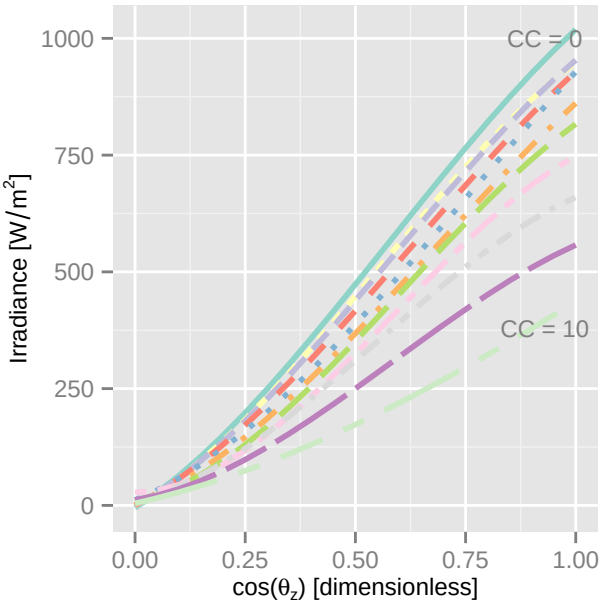

Figure 3: 11 regression lines fitted using the scatters shown in Figure 2, based on 11 cloud cover (a discrete scale of 0 to 10). Third order polynomials are used in the non–linear regressions.

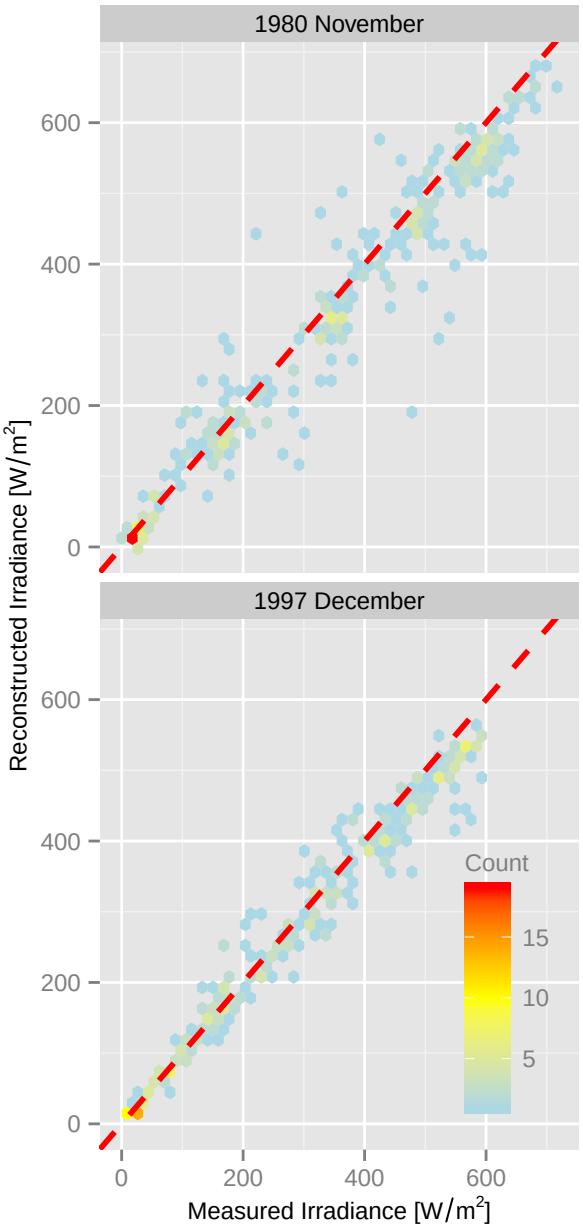

Figure 4: Cloud cover data accuracy in 1980 November and 1997 December. $\mathcal{CC}$ measurements are used to reconstruct GHI via Eqn. (15). Hexagon binning algorithm is used for scatter display. The dotted lines are the identity lines.

Appendix A. Maximum likelihood estimation

The likelihood function of the series vector $\boldsymbol{y}$ is a function of the model parameters $\boldsymbol{\theta}$ and the initial state vector $\boldsymbol{x}_0$. Recall the definition of innovation, $\varepsilon_t = y_t - \mu_t$, representing the new and unpredictable component in the series. The likelihood function also depends on the innovation variance σ^2. It can be shown that the joint density of the series is the weighted product of the densities of the individual innovations:

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x}_0, \sigma^2) = \prod_{t=1}^{n} \frac{\mathbb{P}(\varepsilon_t)}{|\mathcal{R}(\boldsymbol{x}_{t-1})|}, \tag{A.1}$$

with $\mathbb{P}$ denotes probability. Furthermore, when the innovations are Gaussian:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{x}_0, \sigma^2|\boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} \left| \prod_{t=1}^{n} \mathcal{R}(\boldsymbol{x}_{t-1}) \right|^{-1} \exp\left(-\frac{1}{2} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{\sigma^2} \right), \tag{A.2}$$

and the log likelihood is

$$\log \mathcal{L} = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{t=1}^{n} \log |\mathcal{R}(\boldsymbol{x}_{t-1})| - \frac{1}{2} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{\sigma^2}. \tag{A.3}$$

The ML estimate of σ^2 is thus given by:

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2, \tag{A.4}$$

which can be found by setting the derivative of Eqn.(A.3) w.r.t. σ^2 to zero. Therefore, the concentrated likelihood is

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{x}_0|\boldsymbol{y}) = \left[2\pi \exp(1) \widehat{\sigma}^2 \right]^{-n/2} \left| \prod_{t=1}^{n} \mathcal{R}(\boldsymbol{x}_{t-1}) \right|^{-1} \tag{A.5}$$

by plugging Eqn. (A.4) into Eqn (A.2). The two times of negative log–likelihood is:

$$\begin{aligned} -2 \log \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{x}_0|\boldsymbol{y}) &= \text{const.} + n \log \left(\sum_{t=1}^{n} \varepsilon_t^2 \right) + 2 \sum_{t=1}^{n} \log |\mathcal{R}(\boldsymbol{x}_{t-1})| \\ &= \text{const.} + \mathcal{L}^*(\boldsymbol{\theta}, \boldsymbol{x}_0) \end{aligned} \tag{A.6}$$

Hence the ML estimates of the parameters are obtained through minimizing $\mathcal{L}^*$.

[1] J. Bosch, J. Kleissl, Cloud motion vectors from a network of ground sensors in a solar power plant, Solar Energy 95 (0) (2013) 13 – 20. doi:http://dx.doi.org/10.1016/j.solener.2013.05.027.
URL http://www.sciencedirect.com/science/article/pii/S0038092X13002193

[2] J. Bosch, Y. Zheng, J. Kleissl, Deriving cloud velocity from an array of solar radiation measurements, Solar Energy 87 (0) (2013) 196 – 203. doi:http://dx.doi.org/10.1016/j.solener.2012.10.020.
URL http://www.sciencedirect.com/science/article/pii/S0038092X12003854

[3] R. H. Inman, H. T. Pedro, C. F. Coimbra, Solar forecasting methods for renewable energy integration, Progress in Energy and Combustion Science 39 (6) (2013) 535 – 576. doi:http://dx.doi.org/10.1016/j.pecs.2013.06.002.
URL http://www.sciencedirect.com/science/article/pii/S0360128513000294

[4] H. Escrig, F. Batlles, J. Alonso, F. Baena, J. Bosch, I. Salbidegoitia, J. Burgaleta, Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast, Energy 55 (0) (2013) 853 – 859. doi:http://dx.doi.org/10.1016/j.energy.2013.01.054.
URL http://www.sciencedirect.com/science/article/pii/S0360544213000856

[5] Z. Dong, D. Yang, T. Reindl, W. M. Walsh, Satellite image analysis and a hybrid esss/ann model to forecast solar irradiance in the tropics, Energy Conversion and Management 79 (0) (2014) 66 – 73. doi:http://dx.doi.org/10.1016/j.enconman.2013.11.043.
URL http://www.sciencedirect.com/science/article/pii/S0196890413007644

[6] J. Cao, S. Cao, Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis, Energy 31 (15) (2006) 3435 – 3445, {ECOS} 2004 - 17th International Conference on Efficiency, Costs, Optimization, Simulation, and Environmental Impact of Energy on Process Systems 17th International Conference on Efficiency, Costs, Optimization, Simulation, and Environmental Impact of Energy on Process Systems.

doi:http://dx.doi.org/10.1016/j.energy.2006.04.001.
URL http://www.sciencedirect.com/science/article/pii/S0360544206001009

[7] A. Linares-Rodriguez, J. A. Ruiz-Arias, D. Pozo-Vazquez, J. Tovar-Pescador, An artificial neural network ensemble model for estimating global solar radiation from meteosat satellite images, Energy 61 (0) (2013) 636 – 645. doi:http://dx.doi.org/10.1016/j.energy.2013.09.008.
URL http://www.sciencedirect.com/science/article/pii/S0360544213007597

[8] C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation, Energy 36 (1) (2011) 348 – 359. doi:http://dx.doi.org/10.1016/j.energy.2010.10.032.
URL http://www.sciencedirect.com/science/article/pii/S0360544210005955

[9] N. Cressie, C. Wikle, Statistics for Spatio-Temporal Data, Wiley Series in Probability and Statistics, Wiley, 2011.
URL http://books.google.com.sg/books?id=-kOC6D0DiNYC

[10] C. C. Holt, Forecasting trends and seasonals by exponentially weighted averages, Office of Naval Research (ONR) Memorandum Number 52, Tech. rep., Carnegie Institute of Technology (1957).

[11] J. K. Ord, A. B. Koehler, R. D. Snyder, Estimation and prediction for a class of dynamic nonlinear statistical models, Journal of the American Statistical Association 92 (440) (1997) pp. 1621–1629.
URL http://www.jstor.org/stable/2965433

[12] R. J. Hyndman, A. B. Koehler, R. D. Snyder, S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting 18 (3) (2002) 439 – 454. doi:http://dx.doi.org/10.1016/S0169-2070(01)00110-8.
URL http://www.sciencedirect.com/science/article/pii/S0169207001001108

[13] J. W. Taylor, R. D. Snyder, Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing,

Omega 40 (6) (2012) 748 – 757, special Issue on Forecasting in Management Science. doi:http://dx.doi.org/10.1016/j.omega.2010.03.004.
URL http://www.sciencedirect.com/science/article/pii/S0305048310000307

[14] J. W. Taylor, Short-term load forecasting with exponentially weighted methods, Power Systems, IEEE Transactions on 27 (1) (2012) 458–464. doi:10.1109/TPWRS.2011.2161780.

[15] J. W. Taylor, Density forecasting of intraday call center arrivals using models based on exponential smoothing, Management Science 58 (3) (2012) 534–549. doi:10.1287/mnsc.1110.1434.

[16] J. W. Taylor, Triple seasonal methods for short-term electricity demand forecasting, European Journal of Operational Research 204 (1) (2010) 139 – 152. doi:http://dx.doi.org/10.1016/j.ejor.2009.10.003.
URL http://www.sciencedirect.com/science/article/pii/S037722170900705X

[17] J. W. Taylor, A comparison of univariate time series methods for forecasting intraday arrivals at a call center, Management Science 54 (2) (2008) 253–265. doi:10.1287/mnsc.1070.0786.

[18] Z. Dong, D. Yang, T. Reindl, W. M. Walsh, Short-term solar irradiance forecasting using exponential smoothing state space model, Energy 55 (0) (2013) 1104 – 1113. doi:http://dx.doi.org/10.1016/j.energy.2013.04.027.
URL http://www.sciencedirect.com/science/article/pii/S0360544213003381

[19] R. J. Hyndman, A. B. Koehler, J. K. Ord, R. D. Snyder, Forecasting with Exponential Smoothing, Springer, Deblik, Berlin, Germany, 2008.

[20] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen, W. M. Walsh, Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging, Renewable Energy 60 (0) (2013) 235 – 245. doi:http://dx.doi.org/10.1016/j.renene.2013.05.030.
URL http://www.sciencedirect.com/science/article/pii/S0960148113002759

[21] D. Yang, Z. Dong, T. Reindl, P. Jirutitijaroen, W. M. Walsh, Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage, Solar Energy 103 (0) (2014) 550 – 562. doi:http://dx.doi.org/10.1016/j.solener.2014.01.024.
URL http://www.sciencedirect.com/science/article/pii/S0038092X14000425

[22] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting 20 (1) (2004) 5 – 10. doi:http://dx.doi.org/10.1016/j.ijforecast.2003.09.015.
URL http://www.sciencedirect.com/science/article/pii/S0169207003001134

[23] M. Lefevre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, J.-J. Morcrette, Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmospheric Measurement Techniques 6 (9) (2013) 2403–2418. doi:10.5194/amt-6-2403-2013.
URL http://www.atmos-meas-tech.net/6/2403/2013/

[24] Y. Dazhi, P. Jirutitijaroen, W. M. Walsh, The estimation of clear sky global horizontal irradiance at the equator, Energy Procedia 25 (0) (2012) 141 – 148, PV Asia Pacific Conference 2011. doi:http://dx.doi.org/10.1016/j.egypro.2012.07.019.
URL http://www.sciencedirect.com/science/article/pii/S1876610212011812

[25] D. Yang, W. M. Walsh, P. Jirutitijaroen, Estimation and applications of clear sky global horizontal irradiance at the equator, Journal of Solar Energy Engineering 136 (3). doi:http://dx.doi.org/10.1115/1.4027263.

[26] R. B. Cleveland, W. S. Cleveland, J. E. McRae, I. Terpenning, STL: A seasonal-trend decomposition procedure based on loess, Journal of Official Statistics 6 (1) (1990) 3–73.

[27] W. S. Cleveland, S. J. Devlin, E. Grosse, Regression by local fitting: Methods, properties, and computational algorithms, Journal of Econometrics 37 (1) (1988) 87 – 114. doi:http://dx.doi.org/10.1016/0304-4076(88)90077-2.

URL http://www.sciencedirect.com/science/article/pii/0304407688900772