



Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data

J.M. Bright ^{a,*}, C.J. Smith ^{a,1}, P.G. Taylor ^{a,b,c}, R. Crook ^a

^a School of Chemical and Process Engineering, Energy Research Institute, University of Leeds, Leeds LS2 9JT, UK

^b School of Earth and Environment, Sustainability Research Institute, University of Leeds, Leeds LS2 9JT, UK

^c Centre for Integrated Energy Research, University of Leeds, Leeds LS2 9JT, UK

Received 22 September 2014; received in revised form 20 February 2015; accepted 21 February 2015

Communicated by: Associate Editor Jan Kleissl

Abstract

Synthetic minutely irradiance time series are utilised in non-spatial solar energy system research simulations. It is necessary that they accurately capture irradiance fluctuations and variability inherent in the solar resource. This article describes a methodology to generate a synthetic minutely irradiance time series from widely available hourly weather observation data. The weather observation data are used to produce a set of Markov chains taking into account seasonal, diurnal, and pressure influences on transition probabilities of cloud cover. Cloud dynamics are based on a power-law probability distribution, from which cloud length and duration are derived. Atmospheric transmission losses are simulated with minutely variability, using atmospheric profiles from meteorological reanalysis data and cloud attenuation derived real-world observations. Both direct and diffuse irradiance are calculated, from which total irradiance is determined on an arbitrary plane. The method is applied to the city of Leeds, UK, and validated using independent hourly radiation measurements from the same site. Variability and ramp rate are validated using 1-min resolution irradiance data from the town of Cambourne, Cornwall, UK. The hourly irradiance frequency distribution correlates with $R^2 = 0.996$ whilst the mean hourly irradiance correlates with $R^2 = 0.971$, the daily variability indices cumulative probability distribution function (CDF), 1-min irradiance ramp rate CDF and 1-min irradiance frequency CDF are also shown to correlate with $R^2 = 0.9903$, 1.000, and 0.9994 respectively. Kolmogorov–Smirnov tests on 1-min data for each day show that the ramp rate frequency of occurrence is captured with a high significance level of 99.99%, whilst the irradiance frequency distribution and minutely variability indices are captured at significances of 99% and 97.5% respectively. The use of multiple Markov chains and detailed consideration of the atmospheric losses are shown to be essential elements for the generation of realistic minutely irradiance time series over a typical meteorological year. A freely downloadable example of the model is made available and may be configured to the particular requirements of users or incorporated into other models.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Irradiance generation; Resource modelling; Minute resolution; Stochastic modelling; Cloud cover

1. Introduction

Solar irradiance varies on a minutely time scale (Sayeef et al., 2012). The fluctuations are driven by cloud dynamics, atmospheric losses (Calinoiu et al., 2014), and the transport of airborne pollutants (Vindel and Polo, 2014a). Changes in irradiance that occur on the same time

* Corresponding author.

E-mail addresses: pm08jmb@leeds.ac.uk (J.M. Bright), pmcjs@leeds.ac.uk (C.J. Smith), P.G.Taylor@leeds.ac.uk (P.G. Taylor), R.Crook@leeds.ac.uk (R. Crook).

¹ Address: UK Network of Energy Centres for Doctoral Training, University of Leeds, UK.

Nomenclature

Latin alphabet

C	cloud coverage fraction ($C/10$)
C_8	cloud coverage in okta (0–8)
epm	elements per minute in matrix
f	white-noise multiplier for k_c variations
G	irradiance, specified by subscript (W m^{-2})
i	random start point within row vector
k_c	clear-sky index (G/G_{cs})
n	number of elements within a re-sampled cloud length row vector
\mathbf{P}^1	transition probability matrix
$P(x)$	probability of x to occur
r	random variable between 0 and 1
R	resolution of primary \mathbf{x} (100 m/el)
s	number of states in Markov process
t	time-step in Markov process
u	wind speed (ms^{-1})
u_{10}	u measured 10 m above ground (ms^{-1})
x	horizontal cloud length (m)
X	state at point t in Markov process
\mathbf{x}	cloud cover row vector of 1's and 0's
\mathbf{x}_ψ	cloud cover row vector adjusted by ψ
z	cloud height (m)
$z_{0\text{ref}}$	roughness length (m)

Numerical

0	used to represent clear sky minute
1	used to represent a clouded minute
60	used as a conversion for secs to min

Greek alphabet

α	coefficient defined by x_{max}
β	single power law exponent
β	tilt angle from horizontal of inclined plane
δ	coefficient defined by x_{min}
ϵ	minutely fluctuation from k_c
θ_i	solar incidence angle normal to angled plane
θ_z	solar zenith angle
σ	std. dev. around hourly mean of k_c
ψ	sampling rate

Subscript

\mathbf{B}	beam (direct)
\mathbf{B}	denoting a Boolean matrix
clear	clear minute
cloud	cloudy minute
cs	clear sky
C	denoting cumulative probability in matrix
D	diffuse
i	the i th state at time $t - 1$
j	the j th state at time t
m	minutely
max	maximum value
min	minimum value
n	the n th order of a Markov process
P	panel (arbitrary alignment)
ref	value at a reference measurement
s	number of states in Markov process

scale as changes in electricity demand will impact the benefits of storage and self-consumption in a domestic or community PV system (Marcos et al., 2014). Integrated electricity demand, PV supply, and storage simulations must operate on a minutely time scale to capture these effects, and therefore require minutely irradiance time series as an input (Widen et al., 2015; Sayeef et al., 2012; Hummon et al., 2012; Cao and Sirén, 2014). Calibrated minutely irradiance datasets are generally the output of isolated research projects and are often limited in duration, measurement consistency, and location. Hourly weather data, however, is widely collected and made available through national meteorological offices. This hourly data fails to capture the intermittent nature of solar irradiance (Sayeef et al., 2012), therefore some solar irradiance models use hourly weather datasets to artificially generate minutely irradiance time series.

The focus of solar irradiance models can vary from predicting the future irradiance, to providing a general expected irradiance at any location globally. Many examples of these models have been reviewed, analysed and validated in

literature (Gueymard, 2012). The methodology of interest is a sun obscured type approach. This is where the cloud cover is predicted or determined, thereby implying when the solar beam irradiance will be obstructed. A more complex methodology is outlined and developed by Morf (1998, 2011, 2013) where cloud cover is two-dimensionally modelled to replicate sky with certain clouded conditions, whilst a random number generator driven model separates irradiance into its subcomponents of diffuse and beam. Atmospheric transmission losses from extraterrestrial irradiance to global horizontal irradiance is extensively detailed in literature, however its inclusion on a time series irradiance generation model is less so. Simplistically, clouded periods can be subjected to a random variate to represent these losses (Ehnberg and Bollen, 2005), however there is scope for a more sophisticated approach. Geographically dependent monthly clearness index distributions can be used to deterministically select the transmission losses during clouded periods (Morf, 2013). Probabilistic methods are commonly seen to generate irradiance data by stochastically selecting the clearness index

based on real observations (Ngoko et al., 2014), or by determining monthly probability distributions of clearness index (Graham et al., 1988), or stochastically producing cloud cover and applying beam irradiance transmission losses (Aguir et al., 1988). There are numerous methodologies that can be utilised and developed upon.

The methodology presented within this paper is a temporal only sun obscured type model with two distinct sections, the first generates 1-min resolution cloud cover, and the second calculates the irradiance based on whether the ensuing minute is clear or cloudy. A distinct aspect of this methodology is the use of a multitude of Markov chains to stochastically determine future weather condition states of pressure, wind speed, cloud height and okta,² incorporating the weather variation influence of season, the diurnal dependency and variations caused by pressure. Furthermore, the pattern of cloud cover is generated with greater complexity than in previous literature by using a one-dimensional method that considers cloud height, the speed of the cloud, and the statistical distribution of the horizontal cloud length. The irradiance is calculated in three stages, firstly the atmospheric transmission for each minute is determined, secondly the theoretical clear-sky irradiance is calculated based on earth–sun geometry, and thirdly the irradiance is broken down into its different components in order to obtain irradiance on an arbitrary plane. The most distinct element of the irradiance calculations is the production of clear-sky index probability distributions based on the okta number, and the deterministic approach to applying atmospheric losses minute-by-minute with the inclusion of an intra-hour dependency.

The methodology is geographically flexible, so long as the input data exist for the location desired. However, individual simulations at nearby locations would not correlate due to the non-spatial nature of the methodology. The model output is a synthetic global solar irradiance time series upon an arbitrary plane at a 1-min resolution. It is a temporal data series only, and does not include a spatial dimension. Its applications are therefore limited to cases where the spatial element is not integral, such as small scale studies where a single high-resolution irradiance data series input is ideal. The methodology has suggested application in the improved modelling of small-scale PV supply, demand, and storage systems, through the calculation of electricity supply on a time scale that matches the demand flows. Regular demand flows operate with high power across a small duration, meaning that mean hourly averages fail to capture electricity peaks in the supply and demand. To appropriately capture electricity flows within the residence that can accurately calculate efficiencies, self-consumption losses, battery charge and discharge

states for example, an appropriately high-resolution time scale is required (Darcovich et al., 2015; Torriti, 2014). This methodology offers an improved input estimation of the solar irradiance to these types of studies.

A validation of the methodology is carried out using hourly input observation data from Leeds, UK, and with 1-min input observational data from Cambourne, Cornwall, UK. Comparisons with the irradiance output have been made against independent radiation observation data for the same location, and the key features of this methodology are discussed.

2. Methodology

An example of the output can be seen in Fig. 1. Fig. 1a is the mean okta number for the hour, Fig. 1b is the one-dimensional cloud cover where black indicates the presence of a cloud, and Fig. 1c is the corresponding horizontal global irradiance. The simulated day selected was typical of June for Leeds, UK.

The inputs required for this methodology are the hourly mean weather observation measurements of sea level pressure (hPa), 10 m wind speed (knots), cloud base height (decameters), and okta number (okta). In order to capture accurately the probability statistics of a typical meteorological year, a minimum of 10 years of observational data are recommended.

2.1. Cloud size

A two-dimensional approach to producing cloud cover could be used to replicate the measurement type of an okta

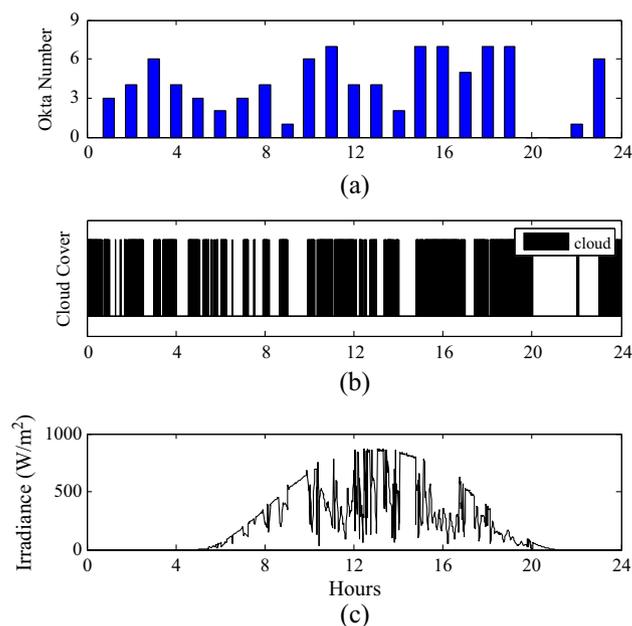


Fig. 1. An example output of a typical day in June for Leeds, UK, that visually demonstrates the fundamental working of the methodology. Plot (a) shows the mean okta number for each hour. Plot (b) shows the cloud cover in a minutely resolution, where black is indicative of cloud present. Plot (c) shows the corresponding global irradiance on a horizontal plane.

² An okta is the representation of cloud amount reported in eighths, where zero okta represents the complete absence of cloud, through to 8 okta representing full cloud cover; an additional value of 9 okta represents full coverage due to fog or other meteorological phenomena (UKMO, 2010).

number (Morf, 2011). The methodology presented, however, requires only a single passage of beam irradiance from sun to the solar panel; a representation of the whole sky is not necessarily required. The typical method for measuring cloud cover is to use a cloud base recorder which uses a vertical laser pulse to track cloud presence (UKMO, 2010). This measurement device takes readings in a single direction, the okta number derived from this type of measurement is considered an acceptable representation of obscured beam irradiance once it is adapted to a synthetic temporal representation. Unlike previous methodologies, the cloud cover will be generated to follow a size distribution for a single linear dimension.

Horizontal cloud size distributions are shown to be well-represented using a single power-law relationship using an exponent of $\beta < 2$ (Wood and Field, 2011; Stull, 1988; Leahy et al., 2012; Pressel and Collins, 2012; Ray and Engelhardt, 1992)

$$P(x) = \alpha x^{-\beta} \quad (1)$$

where x is the horizontal cloud length, β is the exponent taken at 1.66 between 0.1 and 1000 km (Wood and Field, 2011), α is a constant and $P(x)$ is the probability that x will occur.

Eq. (1) represents the probability of x without horizontal cloud length limits. Eq. (2) introduces these limits and to allow for pseudo-random number extraction

$$x = (\alpha + \delta r)^{\frac{1}{1-\beta}} \quad (2)$$

where r is a random variable that is uniformly distributed between 0 and 1, and α and δ are coefficients defined by the upper and lower limits of x as

$$\alpha = x_{\max}^{1-\beta} \quad (3)$$

and

$$\delta = x_{\min}^{1-\beta} - \alpha \quad (4)$$

where x_{\min} is the minimum cloud length and x_{\max} is the maximum cloud length.

These equations allow a 1 metre resolution row vector of cloud cover, \mathbf{x} , to be generated with 1s and 0s representing single metres of cloudy and clear intervals respectively. A signal poly-phase filtering technique (implemented using the built in Matlab 'resample' function (Oppenheim et al., 1999; Matlab, 2012a)) is applied that decimates vector, \mathbf{x} , by a sampling rate, ψ . This converts the cloud length row vector (in m) into a smaller vector that is representative of increasing cloud speeds, and so the resultant vector, \mathbf{x}_{ψ} , is a time series.

The clouds are assumed to be travelling at the concurrent wind speed and cloud height for each hour of simulation. The sampling rate used in the signal poly-phase filtering technique is

$$\psi = \frac{R}{u \times 60} \quad (5)$$

where u is the wind speed (ms^{-1}), 60 is a conversion from seconds to minutes, and R is the resolution of \mathbf{x} (m/element).

\mathbf{x}_{ψ} is the minutely time series cloud cover row vector as a function of the wind speed. Hourly periods are selected at random from \mathbf{x}_{ψ} and the hourly mean coverage value, C , is determined through Eq. (6). C is a fractional representation of the duration of obscured sun, where $\frac{0}{10}$ is clear sky and $\frac{10}{10}$ is fully obscured;

$$C = \frac{\overline{\left(\sum_{i=n \times r}^{i+59} \mathbf{x}_{\psi} \right)}}{60} \quad (6)$$

where the overline represents the mean, n is the total number of elements in \mathbf{x}_{ψ} , and i is a random start point along the length of \mathbf{x}_{ψ} .

Hours are selected from \mathbf{x}_{ψ} until there is a cloud cover database consisting of 1000 examples of an hour of cloud cover for each possible u at each C . During each hour of simulation, an hour of cloud cover is selected from the database using the stochastically determined values of u and C as a reference.

2.2. Markov chains

To implement the stochastic element to the modelling, a Markov process is used. Markov models are a very popular method of stochastic data generation and have been used in many applications, from wind estimates (Masseran, 2015), solar energy estimations (Bhardwaj et al., 2013; Vindel and Polo, 2014b; Hocaoglu, 2011), and in weather generation (Yang et al., 2011).

A Markovian process is a probabilistic mathematical method whereby transitions from one state to the next are directed by discrete probabilities taken from the statistics of real-world processes. In the case of this methodology, statistics are developed from real observational transitions of mean hourly okta, wind speed and cloud height. From these statistics, transition matrices can be constructed.

This methodology uses a single order Markov model whereby only one previous time-step, $t-1$, influences the transition process from $t-1$ to t . Higher order transitions exist, from first order $t-1$, through to the n th order $t-n$. Ngoko et al. (2014) describes how a Markov process ($X_t, t = 0, 1, 2, \dots$) that has s allowable states ($1, 2, \dots, s$), is in state j at time t if $X_t = j$. In this first order Markov process, given that the process is in state i at time $t-1$, the probability that it will transition to state j at time t is given by discrete probability P_{ij} , this is calculated as

$$P_{ij} = P(X_t = j | X_{t-1} = i) \quad (7)$$

These probabilities are stored in a transition matrix \mathbf{P}^1 and can be represented as

$$P^1 = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1s} \\ P_{21} & P_{22} & \dots & P_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s1} & P_{s2} & \dots & P_{ss} \end{bmatrix}$$

Constructing a transition matrix requires the conversion of a variable’s range of magnitudes into discreet states. Fortunately for this study all the variables are extracted in appropriate states, for example each okta number (0, 1, . . . , 9) is its own state. The linear time series of observed states must be converted into a statistical representation of transitions. Firstly, the frequency of transitions f_{s_{t-1},s_t} must be found. This is done by tallying every transition of states found in the observational data in the appropriate f_{s_{t-1},s_t} location. Secondly, the frequencies are converted into probabilities through dividing f_{s_{t-1},s_t} by the total number of occurrences that s_{t-1} transitioned, $f_{s_{t-1}}^*$. In the case where the observed frequency of state i transitioning to state j is f_{ij} , and where f_i^* is the total number times state i has transitioned. The probability of this transition can be expressed as $P_{ij} = f_{ij}/f_i^*$ (8)

A pseudo-random number generator is used to determine s_t from s_{t-1} within the simulation. To do this, the cumulative sum of the probabilities P_C^1 is found as

$$P_C^1 = \begin{bmatrix} P_{11} & P_{11} + P_{12} & \dots & \sum_{n=1}^s P_{1n} \\ P_{21} & P_{21} + P_{22} & \dots & \sum_{n=1}^s P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s1} & P_{s1} + P_{s2} & \dots & \sum_{n=1}^s P_{sn} \end{bmatrix}$$

To implement the Markov process in Matlab r2012a, a random variate r , evenly distributed between 0 and 1, is introduced at each transition. P_C^1 is queried against r in a logical statement, resulting in a Boolean matrix

$$P_C^1 < r \rightarrow P_B^1 = 1 \tag{9}$$

where \rightarrow denotes a logical *if* statement, and P_B^1 is the resultant Boolean matrix. For example, if each of the cumulative transition probabilities are less than r , a value of 1 is assigned, else a value of 0 is set. The future state s_t is then given by

$$s_t = 1 + \sum (P_{B,s_{t-1}}^1) \tag{10}$$

where the subscript s_{t-1} indicates the matrix row to sum using the previously determined state. In the subsequent iteration s_t becomes s_{t-1} and the process is repeated.

2.3. Okta, wind speed, pressure and cloud height

Twenty different Markov chains are utilised to stochastically select weather variables for the subsequent hour,

based on the conditions of the current hour. The variables selected through the Markov process are the okta number for each season during both above and below average pressures (8 matrices), wind speed for each season (4 matrices), cloud height for each season (4 matrices), and the diurnal changes for each season (4 matrices). The transition probabilities are calculated through statistical analysis of the transitions between hours within the observation data.

The wind speed and cloud height Markov chains are produced accounting for seasonal variations. A Markov chain is used for each variable representing each of the four seasons, capturing the variability at different times of the year, totalling four chains each. The okta number Markov chains also consider the effect of season, with the inclusion of impacts from pressure and diurnal variation. Eight okta Markov chains are produced that are split by above and below average pressure for each season, and four additional morning okta Markov chains are produced to capture the diurnal variation for okta transitions between 00:00 and 05:00 am for each season. The intent is to capture the variation in transition probability that occurs as a result of weather changes due to the presence of solar energy. 5 am is considered the cut-off because it is a typical sunrise in the summer for the applied study locations. 5 h represents 5 okta transitions and is considered an appropriate duration for the slight propensity to shift towards an increased okta to be represented, Fig. 8 demonstrates the diurnal transition differences. Fig. 2 visually demonstrates the mean okta Markov chain for the entire year, whilst the effect of season can be seen in Fig. 11.

The hourly mean sea level pressure is considered to be either above or below average at any hour, this is to account for weather condition variation between high and low pressure systems. The mean pressure is calculated from the observation data and the frequency distribution of below and above average pressure durations are

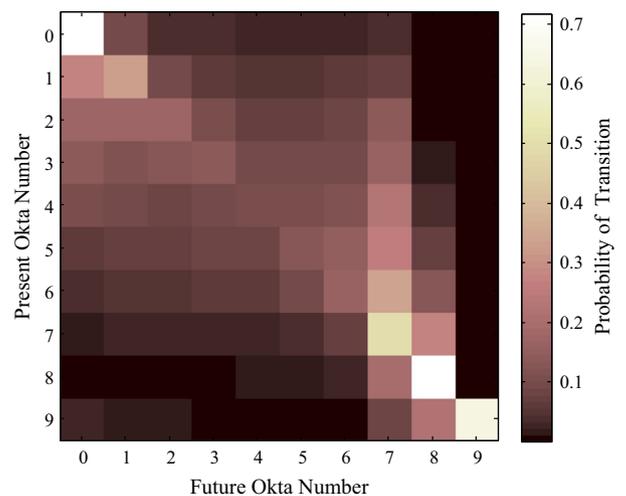


Fig. 2. Colour plot illustrating the Markov chain for the mean annual transition probability of okta number from the present hour to the future hour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

produced. The duration of the pressure system is then selected using a pseudo-random number generator that follows the duration probability distribution. The pressure systems always alternate between periods when they are above and below average.

Wind speed measured by weather stations is typically recorded at 10 m above ground level. For clouds in the atmospheric boundary layer (below 1 km) the wind speed can be extrapolated using the stochastically selected cloud base height as the target reference height, z_{ref} . The extrapolated wind speed is assumed to be the speed at which the cloud travels. To extrapolate the wind speed from the 10 m measurement to a reference cloud height, the following logarithmic profile extrapolation (Best et al., 2008) is used

$$u_{\text{ref}} = u_{10} \frac{\ln(z_{\text{ref}}/z_{0\text{ref}})}{\ln(10/z_{0\text{ref}})} \quad (11)$$

where u_{ref} is the wind speed at the reference height (ms^{-1}), u_{10} is the wind speed at 10 m above ground level (ms^{-1}), z_{ref} is the vertical reference height, and $z_{0\text{ref}}$ is the roughness length of the location (set to 0.14 for rural locations (Best et al., 2008)).

Above 1 km, the geostrophic wind speeds in the free atmosphere are influenced by pressure and thermal gradients (UKMO, 1997). Estimating geostrophic wind speeds is difficult using mean hourly surface observational data; presenting one of the limitations of the current state of the methodology. Without knowledge of pressure and thermal gradients, alternatives must be used. Excellent methods exist such as the wavelet variability model (Lave and Kleissl, 2013), however the data input required is an irradiance profile, an input avoided in the rationale of this methodology once clear sky statistical analysis has been performed. Future development is required to truly address the cloud speed, the validation results however were found to be relatively insensitive to this aspect, and so at this time a guided estimation is utilised. Using the same range as used in the wavelet variability model (0–25 ms) and typical geostrophic wind speeds of 12.5 km h^{-1} (Mathiesen et al., 2013), the estimated wind speed at cloud heights above 1 km is determined allowing variation between the suggested range with the mean found at the typical free atmosphere wind speed as $u_{>1\text{km}} = \text{Gamma}(2.69, 2.14)$.

For each hour of simulation, the wind speed, cloud height and okta are stochastically selected from the appropriate Markov chain. From here, an hour of cloud cover at a minutely resolution can be selected from the cloud cover database, resulting in a temporal Boolean vector representing 1 min cloud cover.

2.4. Irradiance calculation

Once the cloud cover is generated for the entire time series desired, the irradiance is calculated. The irradiance model is in three parts. Firstly, atmospheric transmission for each cloudy and cloud-free minute is determined.

Theoretical clear-sky irradiance based on earth–sun geometry is then calculated and the atmospheric transmission factor applied to obtain simulated global irradiance. Lastly, simulated all-sky direct and diffuse irradiance is calculated and then applied to obtain irradiance on an arbitrary plane.

2.4.1. Generation of clear sky indices

The atmospheric transmission is parameterised in terms of the clear-sky index, $k_c = G/G_{\text{cs}}$, where G_{cs} is a theoretical cloud-free irradiance value. To account for the varying optical thicknesses of clouds, it is appropriate to choose k_c from a distribution rather than to assign a fixed value. To this end, observations of G from the 91 MIDAS³ stations that provided observations of both cloud okta and hourly global irradiance in the year of 2012 were used. Hours in which the mean solar zenith angle was greater than 80° were rejected to minimise the impact of horizon effects and exclude hours during which the sun rose or set.

Values of G_{cs} were calculated by the DISORT radiative transfer solver (Stamnes et al., 2000) in the libRadtran package (Mayer and Kylling, 2005). In the radiative transfer calculation, atmospheric profiles of temperature, O_3 concentration, precipitable water vapour and surface albedo were obtained for each month of 2012 from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-interim reanalysis data (ECMWF, 2014) at a resolution of $1.5^\circ \times 1.5^\circ$. Monthly aerosol data was provided by the GLOMAP model at a resolution of $2.8^\circ \times 2.8^\circ$, specifying scattering and absorption coefficients and asymmetry parameters for 6 bands in the short-wave spectrum for sulphate, sea-salt, black carbon and particulate organic matter aerosols in four size modes (Scott et al., 2014). As each simulated hour is to be aligned with actual data to obtain k_c , it is critical to obtain the solar geometry to high accuracy. The latitude and longitude of each MIDAS station was therefore provided to four decimal places, from which the effective cosine-weighted solar zenith angle for each hour was calculated using the Blanco-Muriel et al. (2001) algorithm (accurate to within 0.01°) included in libRadtran.

Clear-sky index observations were grouped by corresponding cloud okta for the hour and histograms of the observations created (Fig. 3). There were 111090 observations in total. From this, the distributions from okta 0, 6, 7 and 8 were kept. Periods of okta 9, which represents sun obscured due to fog or other meteorological phenomena, were treated as overcast hours. For hours where okta is in the 1–6 range, the distribution for okta 6 is used, as the aim is to obtain a distribution for cloud optical thickness for partially cloudy hours which differs from cloud optical thickness in overcast or nearly overcast hours, but is relatively unaffected by periods of clear sky. As okta

³ MIDAS: the UK Met Office Integrated Data Archive System who have maintained and updated meteorological observations from land and marine surface stations since 1853 (UKMO, 2014).

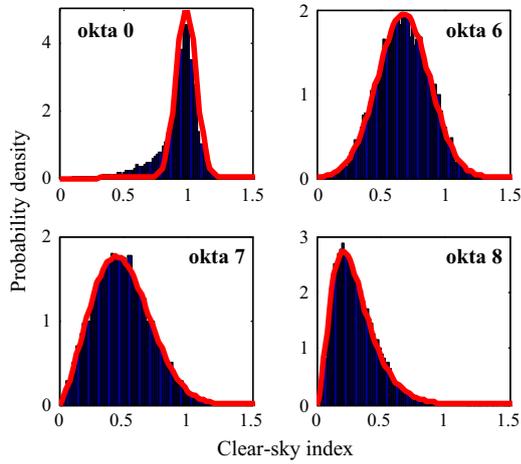


Fig. 3. Distributions of k_c broken down by okta, with analytical distributions fitted that are used in the model.

1–5 h include a significant amount of clear sky, it is difficult to estimate a sun-obscured k_c distribution for these hours. Furthermore, observations of okta 1–5 are comparatively rare. The distributions were fitted using maximum likelihood estimation (Table 1), except in the case of okta 0 where a normal distribution with mean of 0.99 and standard deviation of 0.08 was selected to provide a good visible fit to the peak of the histogram. In the okta 0 histogram, there are several k_c values which are much less than 1, indicating that even when no cloud is registered, the irradiance reported can be much lower than expected. This can be explained by the fact that the cloud sensor is pointing directly upwards whilst the sun is in a different portion of the sky and may be obscured by cloud. This is a particular issue in the UK where the sun is always at least 27° from zenith. To ensure realistic clear-sky values where the sun is shining, these low values have been rejected and a normal distribution fitted visually to the spike of the histogram, which represents genuinely clear hours.

2.4.2. Construction of minutely clear-sky index

For each day, a baseline cloudless clear-sky index, $k_{c,clear}$, is selected from the okta 0 distribution, and applied to all minutes where the sun is not obscured. For cloudy minutes, the baseline cloud-obscured clear-sky index, $k_{c,cloud}$, is selected from the appropriate distribution. If the sky is overcast for an extended period (hourly okta 8 for > 6 hours), the model sets the baseline value of $k_{c,cloud}$ for each interval within the overcast period from the okta

Table 1

Sky states used in the weather generator model and the distribution driving the initial value of k_c for each change of state. The normal distribution is parameterised by (μ, σ) where μ is the mean and σ is the standard deviation. The Weibull and Gamma distributions are parameterised by (a, b) where a is a shape parameter and b is a scale parameter.

Hourly okta	k_c distribution for cloud cover
≤6	$k_c \sim N(0.6784, 0.2046)$
7	$k_c \sim \text{Weibull}(0.5577, 2.4061)$
8	$k_c \sim \text{Gamma}(3.5624, 0.0867)$

8 distribution. This is justified because the clear-sky index of overcast skies does not vary extensively (Skartveit and Olseth, 1992). For partially cloudy hours (okta 1–7), the hourly $k_{c,cloud}$ value is allowed to change each hour and is drawn from the okta 6 distribution if the hourly okta C_8 is 6 or less, and the okta 7 distribution if $C_8 = 7$. If the value of $k_{c,cloud}$ selected from the distribution exceeds 1, a new value from the distribution is drawn.

To move from hourly averages of $k_{c,cloud}$ to minutely instantaneous values, a linear interpolation between two values of $k_{c,cloud}$ was performed to obtain baseline minutely values $k_{c,cloud,m}$. To produce more realistic irradiance profiles, the $k_{c,cloud,m}$ were allowed to fluctuate between intervals of 6 min. The model allows fluctuations at intervals that are a factor of 60. Allowing fluctuations >10 min intervals did not offer similarities in terms of ramp rate occurrences or the variability indices, whilst fluctuations at <5 min intervals saw too much variability. The intention is to capture the gentle rolling of the $k_{c,cloud,m}$ observed in real profiles. Following this, minutely clear-sky index variations were introduced for both cloudy and clear minutes by using a Gaussian white noise multiplier based on the hourly cloud okta:

$$k_{c,m} = f k_{c,m}, \quad f \sim N(1, \sigma) \quad (12)$$

where

$$\sigma = 0.01 + 0.003C_8 \quad (13)$$

for cloudy minutes, and

$$\sigma = 0.001 + 0.0015C_8 \quad (14)$$

for clear minutes. Eq. (13) gives the greatest variation in cloudy minutes for higher okta; Eq. (14) gives the greatest variation in clear minutes with increasing cloud cover.

2.4.3. Generation of irradiance time series

Clear-sky irradiance values, and their direct and diffuse components, were obtained from the HELIOSAT method (Hammer et al., 2003). HELIOSAT is fully described by Hammer et al. (2003) and so is not repeated here. Whilst clear-sky irradiance from the radiative transfer code can be used, the HELIOSAT method is flexible enough to calculate minutely clear-sky direct and diffuse irradiances $G_{B,cs}$ and $G_{D,cs}$ for any location of choice in very little computational time.

The global clear sky horizontal irradiance is the sum of the components

$$G_{cs} = G_{B,cs} + G_{D,cs} \quad (15)$$

and the global *all-sky* irradiance is given by

$$G = k_{c,m} G_{cs}. \quad (16)$$

A further two adjustments are made to ensure that unrealistic values of G are not seen. If the value of $k_{c,m}$ is less than 0.01, it is set equal to 0.01. At the other end of the scale, there are many situations where $k_{c,m}$ exceeds 1 in the distribution tails. This is most likely to happen at low solar elevations where horizon and ground effects are

more pronounced, and the division of a small G by a very small G_{cs} leads to high values of k_c . The largest values of k_c from the observed irradiances were found to obey the relationship given by ($R^2 = 0.9931$)

$$k_{c,\max}(\theta_z) = 27.21 \exp(-114 \cos \theta_z) \dots + 1.665 \exp(-4.494 \cos \theta_z) + 1.08 \quad (17)$$

so that $k_{c,m}$ is set equal to $k_{c,\max}$ if the value drawn from the distribution exceeds this upper threshold.

2.4.4. Decomposition of all-sky irradiance into direct and diffuse components

The decomposition of global irradiance into direct and diffuse components is necessary in order to calculate the irradiance on a plane of arbitrary alignment. Direct irradiance under all sky conditions is related to clear-sky index and clear-sky direct irradiance by an adjustment (Müller and Trentmann, 2010) such that

$$G_B = \begin{cases} 0 & k_c < \frac{19}{69}; \\ G_{B,cs}(k_c - 0.38(1 - k_c))^{2.5} & \frac{19}{69} \leq k_c \leq 1; \\ G_{B,cs}k_c & k_c > 1. \end{cases} \quad (18)$$

Diffuse all-sky irradiance is the difference between the global and direct horizontal values, $G_D = G - G_B$. The direct and diffuse horizontal irradiances are translated to the irradiance incident on a panel of arbitrary inclination and aspect G_p using the Klucher model (Klucher, 1979) which is found to perform well even when compared to more recent models (Gueymard, 2009):

$$G_p = G_D \left(\frac{1 + \cos \beta}{2} \right) \left(1 + F \sin^3 \frac{\beta}{2} \right) (1 \dots + F \cos^2 \theta_i \sin^3 \theta_z) + G_B \frac{\cos \theta_i}{\cos \theta_z}, \quad (19)$$

where θ_i is solar incidence angle normal to the angled plane, and $F = 1 - (G_D / (G_B + G_D))^2$.

3. Validation

All data processing was performed using the commercial software package Matlab r2012a (Matlab, 2012b). Hourly weather observational data are taken from the British Atmospheric Data Centre (BADC) (BADC, 2013) which stores data from MIDAS' land surface monitoring stations, which are well geographically dispersed across the UK. As monitoring stations are occasionally taken off-line for repairs or upgrades for months at a time, 12 years of data sets are used to allow at least 10 years data for each variable.

3.1. Hourly validation in Leeds, UK

For the application of the model, 12 years of hourly weather observations for the monitoring site at Leeds Church Fenton (Source ID: 533) were taken from the

BADC MIDAS data sets to produce the Markov chains. In order to validate the model, 12 years of radiation observation data were taken for the same monitoring station for comparison. Missing data points are assumed to insignificantly impact the validation. In cases where duplicate hourly measurements exist, the hour that has undergone quality control by MIDAS is selected.

Figs. 4 and 5 illustrate direct comparisons between modelled and observational mean hourly irradiance, averaged across 10 years. The 10 year mean annual irradiance is also indicated by the + scatter point at 112.5 W m^{-2} observational and 113.5 W m^{-2} modelled, giving approximately a 0.9% yearly irradiance overestimate for the location of Leeds. Whilst the intent of the model is not to produce hourly mean data in this format, averaging the minutely irradiance generated by the model over hourly timesteps shows a strong correlation when compared with the observational data ($R^2 = 0.9715$).

Fig. 6 displays the mean 10-year frequency distribution of the mean hourly irradiance, normalised to a single year. The correlation between the modelled and observational outputs is $R^2 = 0.9963$.

3.2. Minutely validation for Cambourne, UK

Validation against a minutely irradiance dataset is necessary to confirm that the model can achieve statistically representative minutely resolution. To demonstrate this, 7 years of minutely radiation data was taken from the World Radiation Monitoring Centre – Baseline Surface Radiation Network (WRMC-BSRN) from BSRN station number 50, located in Cambourne, Cornwall, UK (WRMC-BSRN, 2014). Missing data points were ignored and deemed to not significantly impact the distributions for comparison. The model inputs were adjusted for the location of the Cambourne weather station using a latitude of 50.2178, longitude of -5.32656 , and

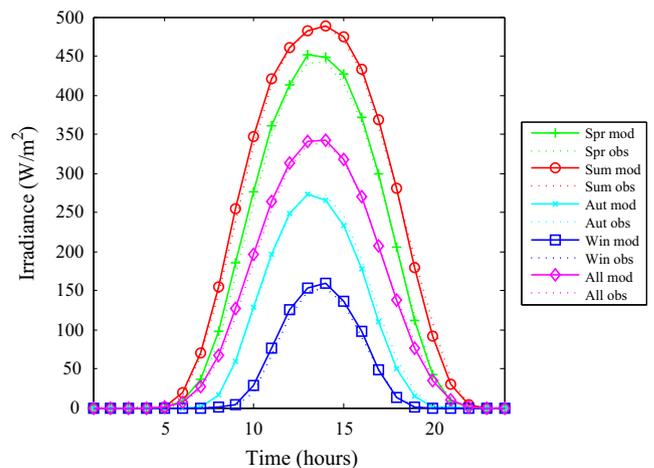


Fig. 4. Comparison of the seasonal mean diurnal irradiance profiles and the annual mean diurnal irradiance profile between 10-years of modelled output and 10-years of validation data from Leeds, UK.

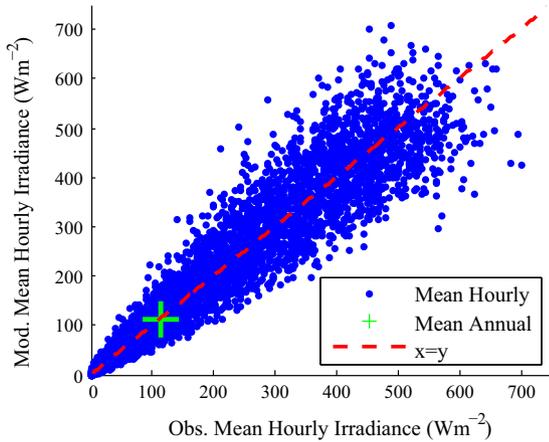


Fig. 5. Comparison scatter plot of each hour of the year, averaged over 10 years, for the model output against the observational data. The yearly mean is also included. An $x = y$ line is included for reference.

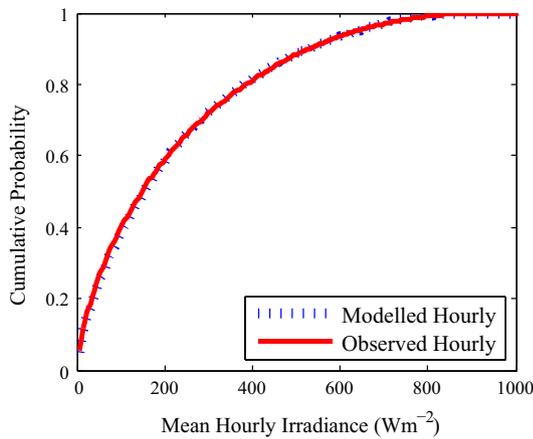


Fig. 6. Comparison of the mean hourly irradiance frequency distribution across 10 years of modelled and observational data from Leeds, UK.

height above sea level of 87 m. 12 years of hourly weather observations for the monitoring site at Cambourne (Source ID: 1395) were taken from the BADC MIDAS data sets to produce the appropriate Markov chains.

Three metrics are used to validate the intermittent nature of the model output, the variability index CDF, the irradiance CDF and the ramp rate CDF. The variability index is the ratio between the rate of change of the measured irradiance and the rate of change of the reference clear sky irradiance (Stein et al., 2012). The more intermittent the day’s irradiance, the higher the variability index. Furthermore, the 2-sample Kolmogorov–Smirnov (K–S) test was carried out for each metric, for each day of the year. The subset of each K–S test consisted of 7 of the same day’s minutely data. For example, 7 modelled samples of the 1st-January represents one subset and is compared against 7 samples of the same day from the observational data.

The comparison of the CDFs of the mean daily variability indices for the model output and the measured validation data is shown in Fig. 7a. The two curves correlate well at $R^2 = 0.9903$. The model generated variability indices

have a slightly increased frequency of mildly variable days between variability indices 10–25 as is indicated by the steeper slope, whilst having a slightly reduced frequency of extremely variable days. Table 2 indicates that 94.38% of days analysed with the 2 sample K–S test reject the null hypothesis that the minutely variability indices of each day of the year observed and modelled cdfs are not from the same sample, when using a significance level of 99%. No annual temporal bias was observed from the days that accept the null hypothesis. The comparison of the observational ramp rate occurrence CDF and that of the model have an excellent correlation, as shown in Fig. 7b with $R^2 = 1$. Both the 1-min interval of ramp-up and ramp-down events are captured excellently using this methodology. This is furthermore demonstrated with the results from the day-by-day K–S tests shown in Table 2. All days tested reject the null hypothesis across all tested significance levels with mean asymptotic p -value of $6.7365e^{-29}$. Fig. 7c shows the comparison of the CDFs of the irradiance occurrence frequency between the 1-min generated data and the 1-min observational data across 7 years. The probability distribution functions correlate well at $R^2 = 0.994$ with a very slight underestimation of the mid-range irradiance occurrences. A possible cause for this is combining the clear sky indices of okta values 1–6 with the same distribution. The day-by-day irradiance frequency distribution compare well with observational data as seen in Table 2, which shows that over 95% of days tested reject the null hypothesis at a significance level of 99%.

4. Application and discussion

Irradiance is generated over a 10 year simulation using Leeds, UK, input observational data to complete the methodology, the same simulation was used in the validation. A distinctive element of this methodology is the use of 20 different Markov chains in order to capture seasonal, diurnal and pressure based variations.

Combined, the Markov chains successfully replicated the okta frequency distribution of the 12 years of observation data. The mean percentage error between the modelled and observed okta transitions is -0.03% , which demonstrates that the overall statistics for Leeds are retained even with 20 separate Markov chains in use, whilst capturing more detailed transition characteristics at certain times of a year. The annual correlation of modelled and observational okta frequency distribution is $R^2 = 0.9956$, showing a highly accurate reproduction of the location’s okta statistics.

Each type of okta Markov chain (diurnal, pressure, and season) is analysed through examining the variation away from the annual mean okta Markov chain (Fig. 2) to assess the impact each of the different types have on the transition probability. The mean okta Markov chain is produced with every transition in the observational data, and not separated for season, pressure and season.

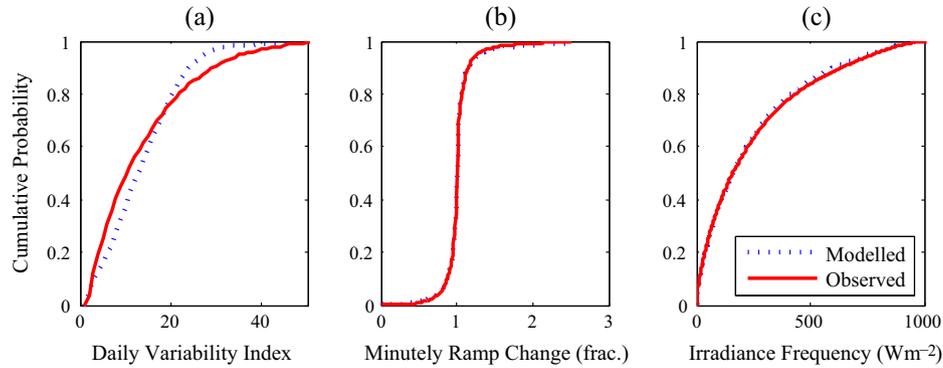


Fig. 7. Cumulative probability distribution plots of the (a) daily variability index, (b) the minutely irradiance ramp-up and ramp-down occurrences, and the (c) minutely global horizontal irradiance frequencies for both the modelled data and the measured data from Cambourne, UK.

Table 2

The percentage of days that reject the null hypothesis when performing the 2-sample Kolmogorov–Smirnov test on 7 of the same day of modelled and observational CDF profiles of the minutely variability indices, minutely ramp rate occurrences, and the minutely irradiance frequency, tested at increasing confidence limits.

K–S test significance level	90%	92.5%	95%	97.5%	99%
Variability index	98.63%	98.15%	97.81%	96.51%	94.38%
Ramp rate occurrence	100%	100%	100%	100%	100%
Irradiance frequency	99.11%	98.70%	98.29%	97.40%	95.89%

4.1. Diurnally weighted okta Markov chains

The most significant deviation of transition probabilities away from the mean is seen with the morning diurnal dependency type Markov chain. A comparison of the mean of the diurnal Markov chains (so that variations caused by pressure and season are excluded) minus the mean okta Markov chain is shown in Fig. 8. It illustrates the most significant deviations from the mean, as the differences in probability below ± 0.01 are removed from the plot. There is a very distinct pattern for the probability of an okta number to remain the same from one state to the next,

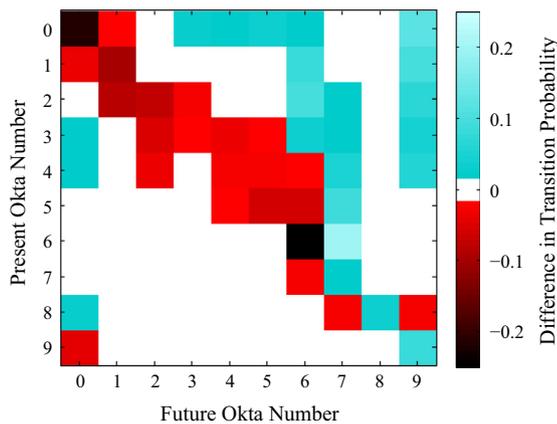


Fig. 8. Colour plot indicating the deviation of the diurnally subjective okta Markov chain minus the annual yearly mean okta Markov chain. White space indicates a $\leq \pm 0.01$ change in transition probability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with a decrease of between 0.01 and 0.25 for okta 0–6. Okta 7, 8 and 9 however have an increased probability of remaining the same meaning that cloudier weather states are longer and more prevalent. Significant probability increases for an okta state transitioning to a higher okta number are seen across all okta states. There is also a slight increase for okta 3, 4 and 8 to transition to okta 0. The physical implication of this is that during the morning period (00:00–05:00 am), the current okta value has an increased tendency to transition towards an okta value of 6–8 or 0, and so cloud state is most likely to develop into either fully obscured or complete clear sky. This increased tendency towards both okta extremes during the morning

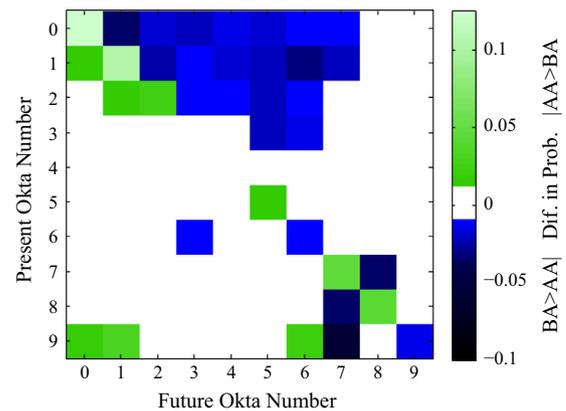


Fig. 9. Colour plot indicating the difference between the above average pressure Markov chain (AA) minus the below-average pressure Markov chain (BA). White space indicates a $\leq \pm 0.01$ change in transition probability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

period is not captured when using a single Markov chain for all times of the year.

4.2. Pressure weighted okta Markov chains

The reason for the inclusion of transitions based on pressure is to attempt to capture variations in weather that are caused by high and low pressure systems. The approach

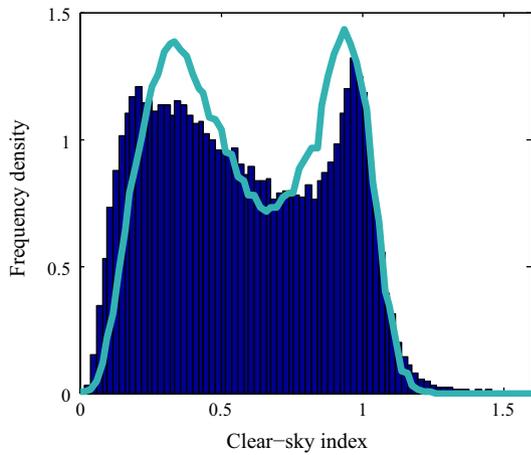


Fig. 10. Histogram of the mean hourly observed values of k_c (bar) compared to the values of k_c from the model (line).

is simplified to be either above or below the average pressure. Fig. 9 shows a slightly different analysis to the diurnal comparison for simplicity. The mean below average (BA) transition probabilities are subtracted from the mean above average probabilities (AA). The darker shade (blue) represents occurrences where the BA transition probability is greater than that of AA, for that particular transition only. The lighter shade (green) represents where the AA transition probability is greater than that of BA. The most significant key point to note is that the probabilities for okta 0–3 to transition to a cloudier state are consistently and significantly greater during periods of BA pressure. The transitions are more stable during AA pressure, with an increased tendency to remain at the same okta state. For okta 1–2 during AA pressure, there is a slight increase in probability to become less cloudy. Okta 7–9 during BA pressure has a higher tendency, +0.1, to remain cloudy.

4.3. Seasonally weighted okta Markov chains

The variation due to the different seasons is detailed in Fig. 11, where the deviation from the mean okta Markov is shown for each of the four seasons. Again, in order to observe the most significant of differences, deviations of ± 0.015 are removed. Autumn variations are the closest to the mean. The frequency of occurrence of okta 2–6 are

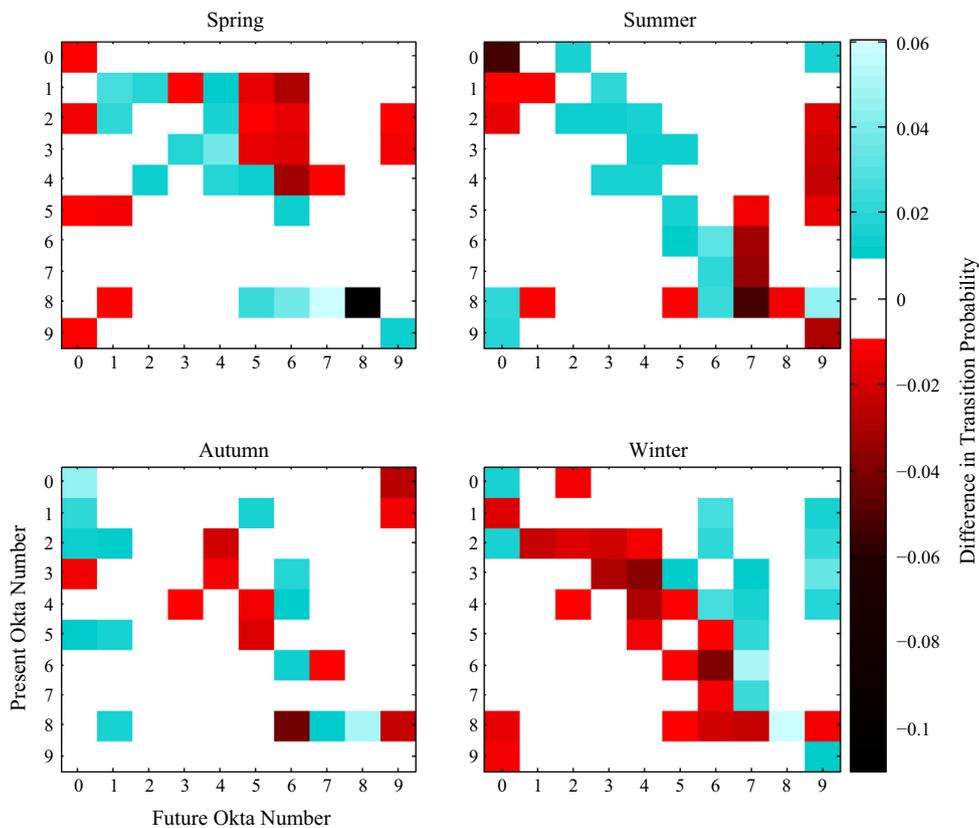


Fig. 11. Colour plot indicating the deviation of the individual season Markov chains from the yearly mean okta Markov chain. The colour indicates the difference in transition probability; white space indicates a $\leq \pm 0.015$ change in transition probability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the lowest and so differences in this region offer the least impact on the overall model. The tendency for okta 1 to remain the same is significantly larger for autumn. Spring transitions are similar to those of autumn, in that the most significant deviations occur during the least sensitive okta range. There is however a distinct shift to favour lower okta values. Summer and winter contain the most significant variations from the mean. This is most important in summer as this is the period receiving the majority of the year's irradiance. Summer sees an increase in the probabilities along the diagonal ($x = y$), hinting towards more stable transitions, directly contrasting the pattern observed in winter. Interestingly, winter has an increased tendency to move towards the two extremes of clear sky and fully obscured. It is worth noting that the differences reported here are specific to the location of Leeds, whilst the differences may be similar in pattern, the seasonal, diurnal and pressure differences are specific to the input observational data.

4.4. Irradiance

Hourly averages of clear sky indices derived from the model were compared with k_c values from the MIDAS data (Fig. 10). It can be seen that the model recreates bi-modal structure of the real-world distribution of clear-sky indices. The peaks from the simulated data are of similar height to the MIDAS data, although the intermodal spread is lower and in particular there are fewer extreme high or low values in the simulated data.

Fig. 12 displays the diurnal root mean square error (RMSe) performance of the simulated data for each season and annually, averaged across 10 years. The RMSe is calculated as a direct comparison of 10 years of modelled mean hourly irradiance data aligned with 10 years of mean hourly observational data. The time of day of most importance is around midday during the summer season during maximum clear-sky irradiance as this is when the potential for the largest ramp-rates and peak outputs are found.

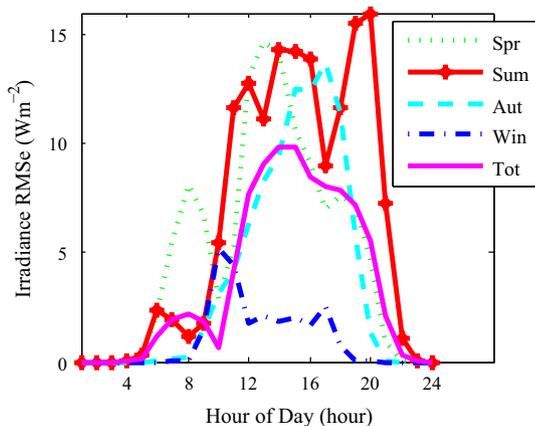


Fig. 12. The diurnal mean hourly irradiance root mean square error (Wm^{-2}), averaged across 10 simulations. The RMSe is displayed diurnally and separated by season, the yearly diurnal RMSe is also shown.

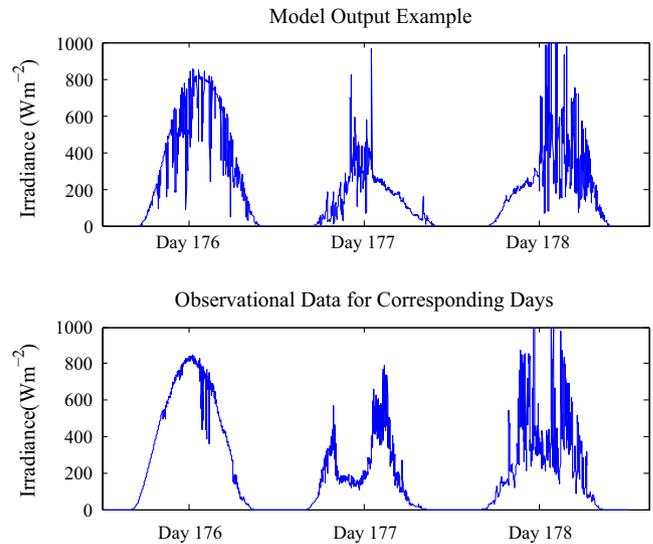


Fig. 13. Three days of example irradiance profile outputs from the model (top), and from observational data (bottom). Note this is neither a predictive method nor a forecasting model and so the synthetic profiles are not intended to match or recreate observational data. The purpose of this figure is to demonstrate the model's output and to show real 1-min resolution observational data irradiance profiles.

Typical un-obscured irradiance outputs during these times are around 900 Wm^{-2} at Cambourne. The RMSe in Fig. 12 at midday in summer is at 14 Wm^{-2} , offering potential mean hourly irradiance error of $\pm 1.5\%$. The vast majority of hourly RMSe values fall below 15 Wm^{-2} . The correlation of 10 years observational and generated mean hourly irradiance profiles is $R^2 = 0.9715$.

5. Conclusion

This paper presented a stochastic sun obscured type synthetic irradiance generation methodology with the intention of taking readily available hourly resolution input data and generating statistically accurate 1-min resolution irradiance profiles. One-dimensional cloud cover has been generated so that the horizontal cloud length follows a power law relationship. Twenty different Markov chains are used in setting the hourly conditions that determine the cloud cover. These Markov chains allow the stochastic selection of wind speed and cloud height whilst accounting for seasonal variability, and the selection of cloud cover accounting for seasonal, diurnal and pressure variability. The use of diurnal Markov chains is shown to have the most significant influence on the output. The variability of okta transition due to pressure system and season also allows for greater accuracy in capturing the weather conditions of a geographical location. The distribution of hourly clear-sky indices are recreated whilst allowing for minutely clear sky index fluctuations.

This methodology is validated on an hourly basis using radiation observations from Leeds Church Fenton, UK;

and on a 1-min basis using radiation observations from Cambourne, Cornwall, UK. The model was shown to capture the irradiance frequency distribution with $R^2 = 0.9963$ and the 10-year mean hourly irradiance with $R^2 = 0.9715$. The 10-year mean diurnal comparison was shown to perform well for each season, the maximum RMSE during summer midday is at 14 W m^{-2} for a single hour. The largest RMSE equates to an error during summer height outputs of $\pm 1.5\%$. The use of clear sky index frequency distributions as a function of the okta number allows the typical cloud characteristics inherent during okta to be produced. The 1-min validation in Cambourne showed the same performance on an hourly resolution as shown for Leeds, however it has been demonstrated that the 1-min irradiance ramp-rates, irradiance frequencies and the variability indices distributions correlated very well at R^2 values of 1.000, 0.9994, and 0.9903 respectively. Day by day 2-sample K–S tests were carried out on the modelled and observed CDFs of the variability index, ramp rate occurrences, and the irradiance frequency; the percentage of days that reject the null hypothesis at significance levels of 99% are 94.38%, 100%, and 95.89% respectively. Whilst not a predictive model, visual inspection of both the 1-min resolution data displayed in Fig. 13 shows how the observational and modelled irradiance profiles have very similar characteristics during overcast and intermittent moments.

This model can be used to generate statistically accurate 1-min time series irradiance data from readily available 1-h meteorological input data. The generated irradiance can be utilised as an input into other applications such as temporal intermittency studies, small scale grid impacts modelling associated with residential PV, and residential PV integrated battery storage modelling. Furthermore, this methodology is not limited to a minutely resolution and can theoretically be applied to produce secondly time series, however, in order to have confidence in the accuracy of the output, higher quality resolution observation data is required. This methodology could be extended by the inclusion of a spatial dimension using geographic smoothing techniques such as those demonstrated in Lave et al. (2012), as well as exploring potential improved accuracy through the use of higher order Markov models.

The solar resource model has other possible applications and a freely downloadable example of the model will be provided in the hope that researchers will adopt and adapt it for their own purposes (Bright et al., 2015).

Acknowledgements

Thanks are given to Dave Allen for his support and advice, and to the editor and anonymous reviewers for their thorough and helpful critiques.

This work was financially supported by the Engineering and Physical Sciences Research Council through the University of Leeds Centre for Doctoral Training in Low Carbon Technologies (Grant No.: EP/G036608/1).

References

- Aguiar, R., Collares-Pereira, M., Conde, J., 1988. Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. *Sol. Energy* 40 (3), 269–279.
- BADC, 2013. British Atmospheric Data Centre – National Centre for Atmospheric Science – Natural Environment Research Council. <<http://badc.nerc.ac.uk>>.
- Best, M., Brown, A., Clark, P., Hollis, D., Middleton, D., Rooney, G., Thomson, D., Wilson, C., 2008. Small-scale Wind Energy – Technical Report. Tech. Rep., UK Met Office.
- Bhardwaj, S., Sharma, V., Srivastava, S., Sastry, O., Bandyopadhyay, B., Chandel, S., Gupta, J., 2013. Estimation of solar radiation using a combination of Hidden Markov Model and generalized Fuzzy model. *Sol. Energy* 93 (0), 43–54.
- Blanco-Muriel, M., Alarcón-Padilla, D., López-Moratalla, T., Lara-Coira, M., 2001. Computing the solar vector. *Sol. Energy* 70 (5), 431–441.
- Bright, J.M., Smith, C.J., Taylor, P.G., Crook, R., 2015. The Bright Solar Resource Model. <<http://jamiembright.github.io/BrightSolarModel/>>.
- Calinoiu, D.-G., Stefu, N., Paulescu, M., Trif-Tordai, G., Mares, O., Paulescu, E., Boata, R., Pop, N., Pacurar, A., 2014. Evaluation of errors made in solar irradiance estimation due to averaging the Angstrom turbidity coefficient. *Atmos. Res.* 150 (0), 69–78.
- Cao, S., Sirén, K., 2014. Impact of simulation time-resolution on the matching of PV production and household electric demand. *Appl. Energy* 128 (0), 192–208.
- Darcovich, K., Kenney, B., MacNeil, D., Armstrong, M., 2015. Control strategies and cycling demands for Li-ion storage batteries in residential micro-cogeneration systems. *Appl. Energy* 141 (0), 32–41.
- ECMWF, 2014. European Centre for Medium-Range Weather Forecasts – ERA-Interim. Tech. Rep., ECMWF. <<http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>>.
- Ehnberg, J., Bollen, M., 2005. Simulation of global solar radiation based on cloud observations. *Sol. Energy* 78, 157–162.
- Graham, V., Hollands, K., Unny, T., 1988. A time series model for Kt with application to global synthetic weather generation. *Sol. Energy* 40 (2), 83–92.
- Gueymard, C., 2009. Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications. *Sol. Energy* 83, 432–444.
- Gueymard, C.A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. *Sol. Energy* 86 (8), 2145–2169.
- Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R., Beyer, H., 2003. Solar energy assessment using remote sensing technologies. *Rem. Sens. Environ.* 86, 423–432.
- Hocaglu, F.O., 2011. Stochastic approach for daily solar radiation modeling. *Sol. Energy* 85 (2), 278–287.
- Hummon, M., Ibanez, E., Brinkman, G., Lew, D., 2012. Sub-hour solar data for power system modeling from static spatial variability analysis. In: 2nd Annual Solar Integration Workshop, Lisbon, NREL/CP-6A20-56204.
- Klucher, T., 1979. Evaluation of models to predict insolation on tilted surfaces. *Sol. Energy* 23, 111–114.
- Lave, M., Kleissl, J., 2013. Cloud speed impact on solar variability scaling – application to the wavelet variability model. *Sol. Energy* 91 (0), 11–21.
- Lave, M., Kleissl, J., Arias-Castro, E., 2012. High-frequency irradiance fluctuations and geographic smoothing. *Sol. Energy* 86 (8), 2190–2199, progress in Solar Energy 3.
- Leahy, L.V., Wood, R., Charlson, R.J., Hostetler, C.A., Rogers, R.R., Vaughan, M.A., Winker, D.M., 2012. On the nature and extent of optically thin marine low clouds. *J. Geophys. Res.: Atmos.* 117 (D22).
- Marcos, J., Storkel, O., Marroyo, L., Garcia, M., Lorenzo, E., 2014. Storage requirements for PV power ramp-rate control. *Sol. Energy* 99 (0), 28–35.

- Masseran, N., 2015. Markov Chain model for the stochastic behaviors of wind-direction data. *Energy Convers. Manage.* 92 (0), 266–274.
- Mathiesen, P., Collier, C., Kleissl, J., 2013. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. *Sol. Energy* 92, 47–61.
- Matlab, 2012a. R2014a Documentation, Signal Processing, Multirate Signal Processing, Resample. <<http://uk.mathworks.com/help/signal/ug/resampling.html>>.
- Matlab, 2012b. Version 7.14.0.739 R2012a. The MathWorks Inc., Natick, Massachusetts.
- Mayer, B., Kylling, A., 2005. Technical note: the libRadtran software package for radiative transfer calculations – description and examples of use. *Atmos. Chem. Phys.* 5, 1855–1877.
- Morf, H., 1998. The stochastic two-state solar irradiance model STSIM. *Sol. Energy* 62 (2), 101–112.
- Morf, H., 2011. The stochastic two-state cloud cover model STSCCM. *Sol. Energy* 85 (5), 985–999.
- Morf, H., 2013. A stochastic solar irradiance model adjusted on the Angstrom–Prescott regression. *Sol. Energy* 87 (0), 1–21.
- Müller, R., Trentmann, J., 2010. Algorithm Theoretical Baseline Document Direct Irradiance at Surface. Tech. Rep., EUMETSAT Satellite Application Facility on Climate Monitoring.
- Ngoko, B., Sugihara, H., Funaki, T., 2014. Synthetic generation of high temporal resolution solar radiation data using Markov models. *Sol. Energy* 103 (0), 160–170.
- Oppenheim, A.V., Schaffer, R.W., Buck, J.R., 1999. *Discrete-time Signal Processing*, second ed. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Pressel, K.G., Collins, W.D., 2012. First-order structure function analysis of statistical scale invariance in the AIRS-observed water vapor field. *Climate* 25, 5538–5555.
- Ray, J., Engelhardt, F., 1992. *Produced Water: Technological, Environmental Issues and Solutions*; [Proceedings of the 1992 International Produced Water Symposium, Held February 4–7, 1992, in San Diego, California]. Environmental Science Research. Plenum Press.
- Sayeef, S., Heslop, S., Cornforth, D., Moore, T., Percy, S., Ward, J.K., Berry, A., Rowe, D., 2012. *Solar Intermittency: Australia's Clean Energy Challenge*. Tech. Rep., CSIRO.
- Scott, C.E., Rap, A., Spracklen, D.V., Forster, P.M., Carslaw, K.S., Mann, G.W., Pringle, K.J., Kivekäs, N., Kulmala, M., Lihavainen, H., Tunved, P., 2014. The direct and indirect radiative effects of biogenic secondary organic aerosol. *Atmos. Chem. Phys.* 14 (1), 447–470, <<http://www.atmos-chem-phys.net/14/447/2014/>>.
- Skartveit, A., Olseth, J., 1992. The probability density and autocorrelation of short-term global and beam irradiance. *Sol. Energy* 49 (6), 477–487.
- Stammes, K., Tsay, S.-C., Wiscombe, W., Laszlo, I., 2000. DISORT, A General-purpose Fortran Program for Discrete-Ordinate-Method Radiative Transfer in Scattering and Emitting Layered Media: Documentation of Methodology. Tech. Rep., Dept. of Physics and Engineering Physics, Stevens Institute of Technology, Hoboken, NJ, USA.
- Stein, J.S., Hansen, C.W., Reno, M.J., 2012. The variability index: a new and novel metric for quantifying irradiance and PV output variability. In: *World Renewable Energy Forum*, Denver, CO.
- Stull, R., 1988. *An Introduction to Boundary Layer Meteorology*. Atmospheric and Oceanographic Sciences Library. Springer.
- Torriti, J., 2014. A review of time use models of residential electricity demand. *Renew Sustain Energy Rev* 37 (0), 265–272.
- UKMO, 1997. UK Meteorological Office – Source Book to the Forecasters Reference Book. Meteorological Office.
- UKMO, 2010. UK Met Office – Observations: National Meteorological Library and Archive Fact Sheet 17 – Weather Observations Over land. Tech. Rep., UK Met Office.
- UKMO, 2014. UK Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre. <<http://badc.nerc.ac.uk>>.
- Vindel, J., Polo, J., 2014a. Intermittency and variability of daily solar irradiation. *Atmos. Res.* 143 (0), 313–327.
- Vindel, J., Polo, J., 2014b. Markov processes and Zipf's law in daily solar irradiation at earth's surface. *J. Atmos. Sol.-Terrestrial Phys.* 107 (0), 42–47.
- Widen, J., Carpmann, N., Castellucci, V., Lingfors, D., Olauson, J., Remouit, F., Bergkvist, M., Grabbe, M., Waters, R., 2015. Variability assessment and forecasting of renewables: a review for solar, wind, wave and tidal resources. *Renew. Sustain. Energy Rev.* 44 (0), 356–375.
- Wood, R., Field, P.R., 2011. The distribution of cloud horizontal sizes. *J. Clim.* 24, 4800–4816.
- WRMC-BSRN, 2014. World Radiation Monitoring Center – Baseline Surface Radiation Network PANGEA – Data Publisher for Earth and Environmental Science. <<http://bsrn.awi.de/>>.
- Yang, H., Li, Y., Lu, L., Qi, R., 2011. First order multivariate Markov chain model for generating annual weather data for Hong Kong. *Energy Build.* 43 (9), 2371–2377.