

# A synthetic, spatially decorrelating solar irradiance generator and application to a LV grid model with high PV penetration



Jamie M. Bright<sup>a,\*,1</sup>, Oytun Babacan<sup>b</sup>, Jan Kleissl<sup>b</sup>, Peter G. Taylor<sup>a,c,d</sup>, Rolf Crook<sup>a</sup>

<sup>a</sup> School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK

<sup>b</sup> Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA 92093, USA

<sup>c</sup> School of Earth and Environment, Sustainability Research Institute, University of Leeds, Leeds LS2 9JT, UK

<sup>d</sup> Centre for Integrated Energy Research, University of Leeds, Leeds LS2 9JT, UK

## ARTICLE INFO

### Article history:

Received 30 September 2016

Received in revised form 7 February 2017

Accepted 7 March 2017

### Keywords:

Irradiance generation

Spatial correlation

Grid impacts

Distribution grid

## ABSTRACT

Residential photovoltaic (PV) technology is expected to have mass global deployment. With widespread PV in the electricity distribution grids, the variable nature of the solar resource must be understood to facilitate reliable operation. This research demonstrates that synthetic, 1-min resolution irradiance time series that vary on a spatial dimension can be generated based on the following inputs: mean hourly meteorological observations of okta, wind speed, cloud height and atmospheric pressure.

The synthetic time series temporally validate against observed 1-min irradiance data for four locations—Cambourne, UK; Lerwick, UK; San Diego, CA USA; and Oahu, HI USA—when analysing 4 metrics of variability indices, ramp-rate size, irradiance magnitude frequency and clear-sky index frequency. Each metric is calculated for the modelled and observed data at each location and CDF profile correlation compared as well as applying the Kolmogorov-Smirnov (K-S) test with 99% confidence limits. CDF correlation coefficients of each metric are all above  $R \geq 0.908$ , and a minimum of 90.96% of daily irradiance time series passed the K-S test. A spatial validation was performed comparing the model outputs to real observation data. The spatial correlation coefficient regression with site separation was successfully recreated with MAPE = 0.865%, RMSE = 0.01 and  $R = 0.955$ . The spatial instantaneous correlation was shown to behave anisotropically when using fixed cloud direction, with different correlation in along and cross wind directions. Cloud cover states of 40–60% showed the most spatial decorrelation while 0% and 100% had the least.

The model outputs are applied to a distribution grid impact model using the IEEE-8500 node test feeder. PV scenarios of 25%, 50%, and 75% uptake were modelled across a  $1.5 \times 1.5$  km grid. The magnitude and frequency of severe tap changing events are found to be significantly higher when using a single irradiance time series for all PV systems versus individually assigning spatially decorrelating time series.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Irradiance variability and correlation

It is well established that the solar resource is highly variable. The degree of variability is dependent on the time resolution at which it is observed (Lave and Kleissl, 2010; Perez et al., 2011). Irradiance fluctuations occur at the sub-second level all the way

up to slow, monthly changes. The variability significantly increases with increasing time resolution. Copious research demonstrates that PV arrays respond to instantaneous radiation fluctuations caused by passing clouds (Suehrcke and McCormick, 1989), cloud dynamics, atmospheric losses (Calinoiu et al., 2014), and the transportation of airborne pollutants (Vindel and Polo, 2014). This rapid variability can seriously affect power generation from PV (Marcos et al., 2011). Sayeef et al. (2012) performed a thorough exploration of the solar intermittency challenge and described it as a fundamental barrier to the uptake of large-scale solar power in Australia and around the world. They state that there is surprisingly very little real-world data on how solar intermittency affects electricity networks. One limitation on variability studies is data quality. Using real radiation observation data places too much reliance on data availability; these datasets are often plagued with gaps,

\* Corresponding author.

E-mail addresses: [pm08jmb@leeds.ac.uk](mailto:pm08jmb@leeds.ac.uk) (J.M. Bright), [obabacan@ucsd.edu](mailto:obabacan@ucsd.edu) (O. Babacan), [jkleissl@ucsd.edu](mailto:jkleissl@ucsd.edu) (J. Kleissl), [P.G.Taylor@leeds.ac.uk](mailto:P.G.Taylor@leeds.ac.uk) (P.G. Taylor), [R.Crook@leeds.ac.uk](mailto:R.Crook@leeds.ac.uk) (R. Crook).

<sup>1</sup> PhD student with the UK Network of Energy Centres for Doctoral Training at the University of Leeds.

## Nomenclature

### Abbreviations

BADC	British Atmospheric Data Centre
C/PDF	cumulative/probability density function
CEE	cloud edge enhancement
DNO	distribution network operator
GHI	global horizontal irradiance
IF	irradiance frequency metric
K–S	Kolmoorov–Smirnov test
KSI	clear-sky index metric
LV	low voltage
NOAA	National Oceanic and Atmospheric Administration
PV	photovoltaic
QCLCD	quality controlled local clim. data
RR	ramp rate metric
UCSD	University of California, San Diego
VI	variability index metric
WRMC-	World Radiation Monitoring Centre–
BSRN	Baseline Surface Radiation Network

### Variables

$B_i$	Boolean matrix indicating cloud cover
$C$	cloud coverage fraction ( $C/10$ )

$C_t$	target cloud coverage fraction ( $C/10$ )
$d$	distance from property to cloud (m)
$G$	global horizontal irradiance ( $\text{Wm}^{-2}$ )
$G_{cs}$	global clear-sky irradiance ( $\text{Wm}^{-2}$ )
$i$	iterative time-step
$i_a$	iteration value
$k_c$	clear-sky index
$K_c$	vector containing $k_c$ values
$M$	total number of PB systems adopted
$N$	total number of load points available in LV circuit
$N(\mu, \sigma)$	norm-rand distribution (mean, std. dev)
PVP	PV penetration (%)
$r$	radius of cloud (m)
$R$	random variate between 0 and 1
$u$	cloud movement speed ( $\text{ms}^{-1}$ )
$x', y'$	rotated Cartesian coordinate (m)
$x, y, z$	Cartesian coordinates or property (m)
$x_c, y_c$	Cartesian coordinates of cloud centre (m)
$X, Y$	dimensions of spatial domain (m)
$X_C, Y_C$	dimensions of cloud sample (m)
$\theta$	cloud direction angle ( $^\circ$ )

inaccurate time-stamps or inconsistent measurement techniques (Kumar et al., 2013). Whilst methods exist to fill these gaps so as to not over or underestimate the realistic variability of irradiance (Moreno-Tejera et al., 2016), datasets are geographically sparse. There is obvious use for a methodology that could synthetically generate time series for spatio-temporal PV modelling applications.

Much research exists that explores the nature of the variable solar resource including its spatial correlation across varying distances and its time scale characteristics (Widen, 2015; Perez et al., 2011; Gueymard and Wilcox, 2011; Wiemken et al., 2001; Beyer et al., 1991; Gafurov et al., 2015; Perez and Fthenakis, 2015; Hoff and Perez, 2010; Lave et al., 2012). Further research explores the spatial anisotropic nature of solar irradiance as a function of cloud speed, size and motion (Hinkelman, 2013; Arias-Castro et al., 2014; Perez and Fthenakis, 2015). Solar variability cannot be considered as purely isotropic (Hinkelman, 2013; Lave and Kleissl, 2013), as is often assumed in previous research (Otani et al., 1997). The cloud dynamic is the predominant driver of anisotropy. The direction, speed and size of the clouds determine the correlation between sites. Smaller clouds cause a lower correlation between sites, except with sites directly along the direction of cloud motion. Sites in the crosswind direction do not correlate strongly until the cloud size significantly increases (Hinkelman, 2013). For different climatology, such as in the spatio-temporal variability study by Glasbey et al. (2001) where larger synoptic weather systems are dominant, the correlation in the crosswind is greater. The time scale of observations also make a huge impact on the correlation; the shorter the time scale, the faster the decline in correlation (Inman et al., 2013).

The spatial correlation of solar variability due to presence of cloud is highly dependent on the geographic separation between sites. A geographical smoothing effect plays a role in tempering the disruption caused by dramatic ramping events from clouds (Marcos et al., 2012; Lave and Kleissl, 2010; Suehrcke and McCormick, 1989; Otani et al., 1997; Wiemken et al., 2001; Curtright and Apt, 2008; Lave et al., 2012). This smoothing can be exploited to minimise high frequency variability by increasing the geographic dispersion between sites (Lave and Kleissl, 2010;

Perez and Fthenakis, 2015; Arias-Castro et al., 2014; Lave et al., 2015). Geographic dispersion benefits utility-scale PV power plants and residential PV by reducing the effect of ramps due to the reduced correlation between sites (Arias-Castro et al., 2014). Utility scale PV plants can be planned considering geographic dispersion, by contrast, the uptake location of distributed PV generation is largely unplanned as it is determined by the consumer. However, fluctuations can adversely impact both utility and residential scale PV systems (Marcos et al., 2016). Therefore, it is important to have a thorough understanding of the spatial correlation of solar irradiance variability.

When considering a single site, irradiance time series can be synthetically generated (Bright et al., 2015; Larraneta et al., 2015; Morf, 2013; Ngoko et al., 2014). Methodologies that synthetically generate spatially correlating irradiance time series, however, are much less prevalent. The wavelet variability model is a spatio-temporal methodology by Lave et al. (2013) and Lave and Kleissl (2013) that produces spatially decorrelating irradiance time series from a single point sensor irradiance input and a daily correlation scaling coefficient derived as a function of cloud speed. An alternative approach is presented by Marcos et al. (2016) that can simulate spatial decorrelation of irradiance time series from 1 year of 1 s irradiance data from 6 sites dispersed over 1100 km<sup>2</sup>. Both of these models are upscaling techniques, whereby they are not able to simulate distributed irradiance data. Rather, they are only able to consider the correlations of such data in order to derive the aggregated PV power output. The presented methodology in this paper, however, can be used to simulate unique and decorrelating distributed PV power output time series. Gafurov et al. (2015) states that currently, there are no known methods for integration of spatial correlation of solar radiation into synthetic data by using reduced and easily available inputs.

### 1.2. Grid impacts from distributed generation

Power output from PV can contain variations on short time scales due to high frequency irradiance fluctuations caused by passing clouds. These power ramps can cause voltage fluctuations and flicker, voltage rise, reverse power flow and phase imbalance

in distributed PV systems (Du et al., 2013). Voltage fluctuations that can trigger automated line equipment installed on the distribution feeders such as tap changers, potentially causing voltage quality permissibility limit violation (Yan et al., 2014; Nguyen et al., 2016). It is possible to employ responsive control strategies (Mokhtari et al., 2013) with batteries, electric vehicles, or dynamic inverter applications that could match the ramping up or down at the same high frequencies across different phases (Caldon et al., 2014; Segui-Chilet et al., 2007); these services can be costly. It is common to see a single PV generation time series used for many properties within in grid impact related studies (Thomson and Infield, 2007), this can overestimate the number on load tap changer (OLTC) operations by up to 260% from the resultant power ramp rates and magnitudes, ignoring the geographic smoothing (Nguyen et al., 2016; Lave et al., 2015).

The primary grid impact concern of this research is to look at voltage quality within a low voltage (LV) grid network. The Distribution Network Operators (DNOs) provide each house with a valid voltage at all times. When a load is connected to the grid, a voltage drop occurs. Contrastingly, when electricity generation is delivered to the grid, the voltage at that location rises. DNOs are charged with maintaining the power at a voltage within a certain limits with all loads and generations considered (Hernandez et al., 2012). Typically, voltage deviations are permissible at  $\pm 10\%$  of the nominal voltage worldwide. The EN50160 standard exists for all EU members to maintain voltage within 10% of the nominal voltage for 95% of the time; some EU countries apply stricter limits (Widen et al., 2010; Fekete et al., 2012). OLTC equipped transformers are used in traditional distribution systems to control the network voltage magnitude and maintain it within permissible limits (Carvalho et al., 2008).

Perhaps one of the most relevant pieces of research to date comes from Nguyen et al. (2016) who explore the impact of very high penetrations of PV into the distribution network. They conclude that the use of a single irradiance time series for a large area versus multiple time series across that same area results in overestimation of the voltage volatility or deviation, as well as the number of tap operations and power ramp magnitudes. Multiple PV generation inputs are therefore a requirement for greater accuracy in grid modelling. High frequency 30-s irradiance data are generated by Nguyen et al. (2016) using sky imaging to achieve a greater understanding of the voltage flow. Other research misses the impact of variability due to employing 1-h resolution irradiance data (Caldon et al., 2014; Widen et al., 2010; Paatero and Lund, 2007; Mouheb et al., 2012). The impact of irradiance data's temporal resolution on OLTC operations is explored by Lave et al. (2015) and shows that sub 1-min resolution is required to model the distribution grid on account of voltage regulator time constants. The use of a 1-min input overestimates tap operations by  $<10\%$ , which was likely inside modelling error margin, and is suggested to be an acceptable frequency for rough estimations if no higher frequency data are available. The rationale of this research is for cases where high frequency irradiance data is unavailable and a synthetic estimation is required.

### 1.3. Research objectives and paper structure

This paper presents a methodology advanced from previous work by Bright et al. (2015) where a temporal-only, synthetic irradiance generator was presented using readily available mean hourly meteorological observation inputs to produce a 1 min resolution irradiance time series—this previous model is referred to as the Solar Irradiance Generator (SIG) and is spatially correlating as there is no variation in output between two locally situated sites.

The methodology presented in this paper aims to take those same readily available observation data of okta (see Section 2.1), wind speed, cloud height and atmospheric pressure; and produces 1 min resolution irradiance time series that vary on a spatial dimension without the need of input irradiance data; it will be referred to as the Spatially Decorrelating Solar Irradiance Generator (SDSIG), as the correlation between sites falls with distance.

The SDSIG is a spatio-temporal and synthetic time series generation methodology. It outputs spatially decorrelating, 1 min resolution irradiance time series. The SDSIG is coupled to a PV and distribution system simulator employed by the Solar Resource Assessment & Forecasting Laboratory at the University of California, San Diego (UCSD). Not only will this demonstrate the international utilisation potential of the SDSIG, it also allows a geographically flexible tool that facilitates a grid impact analysis to be carried out with the inputs of okta, wind speed, pressure, cloud height, and grid schematics.

The key objectives of this paper are to (1) introduce and detail the SDSIG methodology that generates spatially decorrelating irradiance time series, and (2) demonstrate the SDSIG's application to a distribution grid power flow simulation to assess the associated grid impact of increased OLTC operations as influenced by residential PV penetration. These research aims are achieved through recapitulating the SIG methodology and the introduction of the mechanics behind the SDSIG in Section 2, the SDSIG is temporally validated in Section 3 and the spatial correlation is illustrated in Section 5. The grid impact study is detailed and discussed in Section 6 before the work is concluded and future work ideas listed in Section 7.1.

## 2. Spatio-temporal solar irradiance generation methodology

### 2.1. Summary of previous SIG model

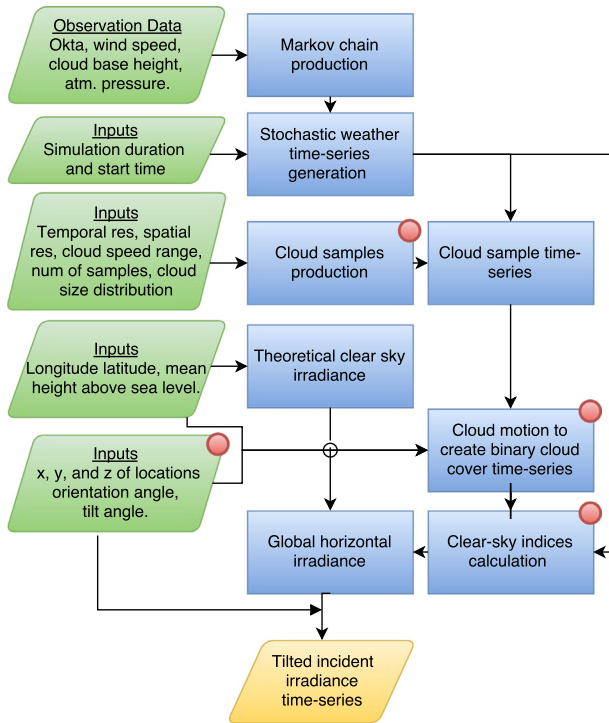
The SIG demonstrated the success of taking mean hourly weather observation data to generate synthetic 1-min temporal resolution irradiance time series. The original methodology has six distinct sections that will be summarised here and in Fig. 1; for a more in-depth understanding of the methodology and equations used, the reader is encouraged to consult the methodology paper (Bright et al., 2015). Firstly, two important definitions must be introduced. An okta is the unit that gives magnitude to the amount of cloud in the observable sky, reported in eighths. 0 okta represents the complete absence of cloud, while 8 okta represent total cloud cover; an additional value of 9 okta represents full coverage due to fog or other meteorological phenomena. Cloud cover fraction is derived from the okta value using the Met Office descriptions and is represented from 0 to 10 out of 10.

#### 1. Cloud samples production

A cloud sample is a binary row vector that represents the cloud cover for a single hour. 1s and 0s signify minutes of sun obscured or sun not-obscured respectively. 500 cloud samples are produced for each value of cloud coverage fraction (0–10 out of 10) at each cloud movement speed ( $1\text{--}30\text{ ms}^{-1}$ ). The cloud sizes are defined following the single power-law relationship of horizontal cloud size distributions defined by Wood and Field (2011). An hour-long row vector is taken at random from a larger row vector before its cloud cover is determined. The sample is stored into the appropriate bins referenced by the coverage and cloud speed.

#### 2. Markov chain production

Using the method of Markov chains, the transition probability matrices of hourly weather variables – okta, wind speed, cloud height and pressure – are created using many years of



**Fig. 1.** Flow chart summarising the methodology. Green parallelograms represent an input, blue rectangles represent a process, and the yellow parallelogram at the bottom indicates the output. A red circle marker to the top right of a box highlights where changes have been made to enhance the methodology to include a spatial dimension. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observation data from the site of interest. The transition of each hour of observation data is analysed and stored in twenty different Markov transition probability matrices (MTM) for the four variables of: okta (under both above and below average pressure conditions and weighted diurnally), wind speed and cloud height are each split four ways by season.

### 3. Stochastic selection of variables

A Markov chain process with the different MTMs is used to stochastically produce a time series of weather variables for every hour within the simulation. Pressure and time of day are used to select the appropriate okta MTM. Cloud height and wind speed is used to estimate the cloud speed. Okta and cloud speed are then used to select the appropriate cloud samples from the bins. A 1-min resolution time series of binary cloud cover is obtained.

### 4. Calculation of clear-sky indices

Observation analysis of the clear-sky index ( $k_c$ ) against okta number found a very strong and well defined set of probability distributions describing the prevalence of  $k_c$  values at each value of okta, this is further explored by Smith et al. (2017) who incorporate the elevation angle making  $k_c$  a function of both okta and elevation angle (see Table A.1 in Appendix A). Using the cloud cover time series and the corresponding okta for each period of coverage, a  $k_c$  is drawn from the appropriate distribution for each moment. Gaussian noise simulating fluctuations and smoothing for long periods of 0 and 8 okta conditions are applied to create irradiance variability.

### 5. Calculation of theoretical clear-sky irradiance

Clear-sky irradiance values and their direct and diffuse components are obtained using the HELIOSAT method (Hammer et al., 2003), which considers the sun-earth geometries of the location selected for simulation. The time series vector of  $k_c$  values are

applied to the corresponding theoretical clear-sky irradiance to obtain the GHI.

### 6. Calculation of incident irradiance

The horizontal direct and diffuse components are calculated using the Müller and Trentmann (2010) method before being translated onto an arbitrary plane of incline – flexible by both orientation and pitch. The incident irradiance is calculated using the Klucher (1979) method.

## 2.2. Spatio-temporal irradiance time series generation

This section will first summarise the methodology improvements and advancements that led to the SDSIG before detailing each change individually. Temporal validations are carried out for three individual locations and a discussion and illustration of the spatial correlation is made.

### 2.2.1. Summary of SDSIG model

The new methodology detailed here is summarised in Fig. 1. Circles in the top right of each component of the flow chart indicate where a development has been made on the previous methodology. Each development will be detailed in this section. A new method of representing the clouds is required to facilitate a spatial dimension. The previous SIG used a one-dimensional matrix approach. Adopting a matrix approach in two dimensions would increase computational demand or reduce spatial accuracy as a trade-off. Instead, a vectorised format is developed to create “cloud samples”, which are 1 h spatial representations of the sky indicating the presence and locations of clouds (Section 2.2.2). Simulating cloud motion using these cloud samples is detailed in Section 2.2.3. The previous SIG methodology had a limited approach to the production of  $k_c$  using extraction of  $k_c$  every 6 min and using only 4 distributions of  $k_c$  by okta.  $k_c$  production within the SDSIG has been redeveloped and improved, as detailed in Section 2.2.4. These three additions and developments facilitate the production of spatially decorrelating irradiance time series.

The circle indicating changes to the inputs are simply the addition of user-defined inputs consisting of individual characteristics for each location where an irradiance time series is desired. These characteristics are the  $x, y$ , and  $z$  coordinates of the location as well as the desired tilt and orientation angles of the plane of each PV system. The simulation domain ranges from 1 to 1500 m for both the  $x$  and  $y$  dimension, this resolution is selected to well represent a typical residential feeder area while still being computationally effective; 1500 m is not a theoretical limit to the methodology. The  $z$  coordinate is the difference in metres away from the central mean height above sea level provided for the baseline irradiance calculations. This methodology has the ability to provide each location with an individual height, panel tilt and orientation, which is rarely employed by current approaches (Engerer and Mills, 2014). The tilt (in degrees for the model input) can be any value, however exceeding  $90^\circ$  will alter the orientation. The panel orientation inputs accept the format  $-180^\circ$  to  $180^\circ$ , where  $-180^\circ = 180^\circ$ , both representing north,  $0^\circ$  is directly south while  $-90^\circ$  is east,  $90^\circ$  is therefore west.

### 2.2.2. Cloud cover sample production

A cloud sample is an array of  $x_c-y_c-r$  values, where  $x_c$  and  $y_c$  are Cartesian coordinates denoting the centres of each cloud and  $r$  are the associated radii. A cloud sample is demonstrated in Fig. 2 as the dashed box containing clouds. The production technique creates 500 different samples for each integer of cloud coverage fraction (0–10 out of 10) and for each value of cloud speed ( $1-30 \text{ ms}^{-1}$ ), these can then be selected as desired once the stochastic weather time series is generated. The coordinates of each cloud’s centre

**Table A.1**

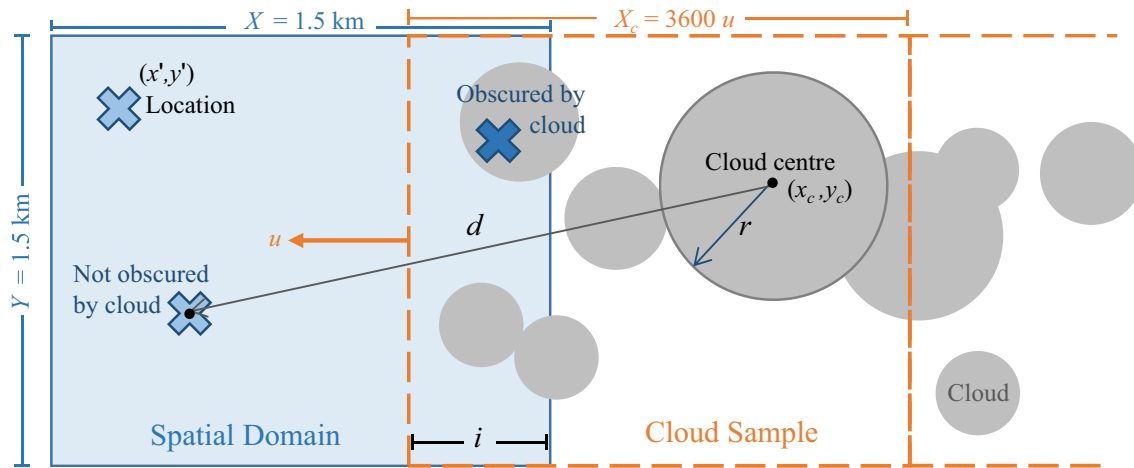
$k_c$  distribution parameters for each okta at different solar elevations. The columns from left to right detail: the type of observation, where auto means that the okta measurement was automated, extrapolated indicates a calculated value; the okta value for the hour of observation; the elevation angle calculated using the Blanco-Muriel et al. (2001) method; the type of PDF distribution referencing Eqs. (A.1) and (A.2); and the scale and shape parameters. where scale, shape1 and shape2 relate to  $a$ ,  $c$  and  $k$  for the Burr distribution and  $a$ ,  $p$  and  $d$  for the generalised Gamma distribution respectively.

Obs. type	Okta	Elevation (°)	Dist. type	Scale	Shape1	Shape2
auto	0	0	burr3	1.004255336	8.052092191	0.334129357
auto	0	10	burr3	1.050077073	30.01220485	0.157472184
auto	0	20	burr3	1.042700528	79.30702738	0.109783689
auto	0	40	burr3	1.041613605	90.44856240	0.114735689
auto	0	50	burr3	1.040348257	106.6777365	0.115167702
extrapolated	0	60	burr3	1.040521909	112.5163000	0.118202507
extrapolated	0	70	burr3	1.039559971	115.5973000	0.120648195
extrapolated	0	80	burr3	1.038727421	114.5183000	0.123093882
auto	1	0	burr3	0.867720611	5.957029683	0.461528066
auto	1	10	burr3	1.034635381	23.30982421	0.135960831
auto	1	20	burr3	1.054028048	45.77649839	0.086326021
auto	1	30	burr3	1.053426748	61.22887273	0.079636671
auto	1	40	burr3	1.048370781	68.68864970	0.080905594
auto	1	50	burr3	1.051091805	83.38138037	0.076478660
extrapolated	1	60	burr3	1.046876702	101.9297000	0.073768447
extrapolated	1	70	burr3	1.045490232	117.3217000	0.070941131
extrapolated	1	80	burr3	1.044103763	132.7137000	0.068113815
auto	2	0	burr3	0.772062039	4.956047790	0.602223618
auto	2	10	burr3	0.997637427	16.12370693	0.171203773
auto	2	20	burr3	1.042715447	33.04399206	0.099150179
auto	2	30	burr3	1.044202341	43.36090288	0.087526150
auto	2	40	burr3	1.039137747	44.65686053	0.094115982
auto	2	50	burr3	1.042453935	59.13488558	0.081633453
extrapolated	2	60	burr3	1.040665084	61.34531164	0.079116354
extrapolated	2	70	burr3	1.040080171	65.44531164	0.074520320
extrapolated	2	80	burr3	1.039495257	67.34531164	0.069924285
auto	3	0	burr3	0.757087430	4.719955631	0.599814980
auto	3	10	burr3	0.962886251	13.26425890	0.200494203
auto	3	20	burr3	1.020610812	24.55724953	0.125177010
auto	3	30	burr3	1.022096040	29.50479982	0.117465722
auto	3	40	burr3	1.016061077	35.57528316	0.108648429
auto	3	50	burr3	1.017760391	40.28732067	0.107029008
extrapolated	3	60	burr3	1.014303520	49.62322278	0.098764717
extrapolated	3	70	burr3	1.012135695	56.75800701	0.092438587
extrapolated	3	80	burr3	1.009967870	63.89279125	0.086112457
auto	4	0	gengamma	0.005480077	0.557400437	7.919701074
auto	4	10	gengamma	0.806798567	3.780633147	3.215699506
auto	4	20	gengamma	0.974592663	7.684412120	3.139798302
auto	4	30	gengamma	0.993042269	10.02624460	3.491953440
auto	4	40	gengamma	0.995177810	11.93575235	3.630871979
auto	4	50	gengamma	0.994027855	14.37955261	4.151900930
extrapolated	4	60	gengamma	1.004320429	17.65246097	4.235087789
extrapolated	4	70	gengamma	1.010364540	20.39297386	4.471435442
extrapolated	4	80	gengamma	1.016408652	23.13348674	4.707783094
auto	5	0	gengamma	0.002024238	0.491298962	8.157538816
auto	5	10	gengamma	0.664559633	2.792447091	3.430206866
auto	5	20	gengamma	0.892248369	5.438275550	3.231241189
auto	5	30	gengamma	0.945393148	7.635869694	3.313058175
auto	5	40	gengamma	0.931245632	8.260396607	3.653753416
auto	5	50	gengamma	0.935768143	9.660919904	3.887662670
extrapolated	5	60	gengamma	0.955266774	12.15815604	3.904411614
extrapolated	5	70	gengamma	0.966907955	13.99957168	4.038153998
extrapolated	5	80	gengamma	0.978549135	15.84098732	4.171896381
auto	6	0	gengamma	0.001677573	0.481989924	8.089286482
auto	6	10	gengamma	0.486147982	2.088140534	3.837157871
auto	6	20	gengamma	0.757850646	3.823759792	3.334064542
auto	6	30	gengamma	0.851623471	5.457121350	3.180219803
auto	6	40	gengamma	0.852367455	5.904250269	3.438878141
auto	6	50	gengamma	0.846455330	6.324495209	3.815566128
extrapolated	6	60	gengamma	0.844980637	8.242714566	3.539666331
extrapolated	6	70	gengamma	0.842396577	9.451120771	3.545829342
extrapolated	6	80	gengamma	0.839812516	10.65952698	3.551992353
auto	7	0	gengamma	0.004537068	0.484621732	4.421152054
auto	7	10	gengamma	0.433486322	1.799709961	2.240029318
auto	7	20	gengamma	0.614610937	2.627897221	2.025277836
auto	7	30	gengamma	0.703741809	3.469979894	2.067728901
auto	7	40	gengamma	0.730381459	4.012040212	2.193316054
auto	7	50	gengamma	0.744532927	4.489305108	2.186545792

(continued on next page)

Table A.1 (continued)

Obs. type	Okta	Elevation (°)	Dist. type	Scale	Shape1	Shape2
extrapolated	7	60	gengamma	0.767009849	5.564174719	2.160900930
extrapolated	7	70	gengamma	0.787405408	6.349960156	2.167008046
extrapolated	7	80	gengamma	0.807800967	7.135745593	2.173115163
auto	8	0	gengamma	0.000167928	0.360116297	5.355314757
auto	8	10	gengamma	0.097204818	1.010969390	3.442829280
auto	8	30	gengamma	0.201529223	1.331344298	2.956426748
auto	8	40	gengamma	0.266130988	1.553540024	2.833357620
auto	8	50	gengamma	0.256395382	1.541268725	2.897819486
extrapolated	8	60	gengamma	0.296218024	1.941970737	2.837260690
extrapolated	8	70	gengamma	0.323651103	2.180435021	2.807957059
extrapolated	8	80	gengamma	0.351084183	2.418899304	2.778653428
auto	9	0	gengamma	0.000342508	0.370450077	4.788361112
auto	9	10	gengamma	0.286555027	1.598964824	2.064839918
auto	9	20	gengamma	0.001227201	0.470980381	5.444603927
auto	9	30	gengamma	0.000303258	0.406115613	5.805599776
auto	9	40	gengamma	0.000559751	0.429564979	5.533348309
auto	9	50	gengamma	0.000815644	0.491761514	7.420808991
extrapolated	9	60	gengamma	0.000815644	0.491761514	7.420808991
extrapolated	9	70	gengamma	0.000815644	0.491761514	7.420808991
extrapolated	9	80	gengamma	0.000815644	0.491761514	7.420808991



**Fig. 2.** Visual demonstration of the methodology. The dashed box represents a cloud sample full of clouds with centre points  $(x_c, y_c)$  and radii,  $r$ , that is travelling at speed  $u$  across the spatial domain in solid blue containing the locations positioned at  $((x', y')$  and indicated by crosses. The spatial domain has dimensions  $XY$  while the cloud sample has dimensions  $X_c Y_c$ . The shade of the crosses indicates overlap by a cloud when darker. The clouds are represented by grey discs. The straight line distance from location centre to cloud centre is denoted  $d$ , and the cloud sample moves at each time step,  $i$ , in the direction indicated by the arrow marked  $u$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

all fall within the sample's spatial domain, which is large enough to cover the simulation area for exactly one hour. The size of the cloud sample is a function of the cloud speed,  $u$ , the temporal resolution and the spatial resolution of the simulation. The size of the area of sky is therefore a function of distance and time, 1.5 km-by-3600 $u$ , where 3600 is the number of seconds in an hour and  $u$  the current cloud speed.

The process to generate a cloud sample is as follows:

1. Set target coverage value,  $C_t$ , within the range  $0-10 \pm 0.5R$ , where  $R$  is a random variate drawn from a uniform distribution between 0 and 1,  $R \sim U(0, 1)$ . The reason for including  $R$  is that the okta observations are rounded quantities. Using  $\pm 0.5R$  creates a continuous variable. E.g. 4 okta = 50% coverage may actually represent the rounding of 54% coverage.  $R$  is redrawn with each  $C_t$ .
2. Randomly add a cloud centre point  $(x_c, y_c)$  into the domain using a homogeneous spatial Poisson point process (Arias-Castro et al., 2014), which selects a point within a mathematical

space using a uniform random variable. The Poisson process is computationally fast however the random placement may not be representative of spatial cloud distributions, furthermore, circular clouds do not represent the geometric diversity of clouds presenting limitations of this method.

3. Extract a cloud radius,  $r$ , from the Wood and Field (2011) horizontal cloud size distribution.
4. Compute the new coverage value,  $C$ . If  $[C] > [C_t]$ , restart the cloud sample.
5. If  $C < C_t$  move to step 2, else move to step 6.
6. Store the cloud sample into a storage bin indexed by coverage and cloud speed.
7. Repeat until all cloud sample bins of each  $C_t$  at each value of  $u$  are full.

### 2.2.3. Cloud movement

The stochastic weather variable generation process creates a time series of okta and cloud speed. The okta is converted to cloud coverage using the recommended conversion provided by UKMO

(2015). Using these two variables, the appropriate storage bin can be selected. A cloud sample is extracted at random from the 500 samples within that bin. The simulation moves the selected cloud sample across the spatial domain at each time-step. In order to simulate the cloud direction, the location's coordinates are rotated about the cloud direction angle,  $\theta$ , using rotational matrices as demonstrated in Eq. (1)

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

where  $x$  and  $y$  denote the initial coordinates of the location within the spatial domain, and  $x'$  and  $y'$  are the rotated coordinates by angle  $\theta$ . The angle by which to rotate is determined using a normally distributed random walk, with standard deviation equal to  $10^\circ$  around the previous time-step's cloud direction.  $10^\circ$  is arbitrarily selected to allow for gradual changes in direction each hour, which are considered to be important as it will affect the correlation between two sites. The alternative is to fix a uniform cloud direction. The impact of the assumption is demonstrated in Section 5 and added to future work. Wind direction at typical measured height of 10 m is not representative of the cloud motion direction, which can have multiple layers travelling in different directions. Methods exist to estimate cloud motion direction that require sophisticated equipment (Wang et al., 2016; Chow et al., 2011), however until a method is available that accurately represents the cloud direction using simple and readily available inputs in order to maintain the rationale of the work, a simple random walk method is applied for simulations where cloud direction changes with weather system (e.g. temperate climates) and a uniform direction can be selected for simulations with prevailing cloud directions (e.g. equatorial).

Whilst the calculations are in a vectorised format, it is simplest to visualise the spatial domain as a static rectangle with the locations of interest marked, this is then gradually overlapped by a second rectangle containing the circular clouds; Fig. 2 visualises this concept. At each time-step of the simulation,  $i$ , the straight line distance from each location's centre situated within the spatial domain,  $(x', y')$ , to the centre of every cloud within the cloud sample,  $(x_c, y_c)$ , is calculated and is denoted as  $d$ . Using  $d$  and each cloud's radius,  $r$ , it is determined if the location is covered by cloud(s) or not. To do this, a logical IF statement is applied such that

$$r > d_i \rightarrow \mathbf{B}_i = 1 \quad \text{else} \quad \mathbf{B}_i = 0 \quad (2)$$

where  $\mathbf{B}_i$  is a Boolean matrix indicating the presence of cloud at the  $i$ th time step;  $\mathbf{B}_i = 1$  signifies cloud whereas no cloud is represented by 0. The symbol  $\rightarrow$  denotes the use of a logical IF statement offering two outcomes. To ensure that clouds are not abruptly cut off in-between cloud fields for situations where cloud centre points are positioned close to the cloud field edge, three sequential cloud samples (past, present and future) are logically queried at each time step, as per Eq. (2). The iterative IF statement is repeated until the current cloud sample has passed outside the spatial domain. At this point the future sample begins to pass directly over the spatial domain and therefore becomes the current cloud sample and a new future cloud sample is selected. This prevents sudden disappearance of cloud once an hour has ended. The outcomes of this step are individual binary time series of cloud cover for each location, which are spatially correlated with each other.

#### 2.2.4. Clear-sky indices

Each cloud contained within a cloud sample is assigned an individual  $k_c$ . The cloud's associated  $k_c$  is applied to the entire area of the disc. The value of each  $k_c$  is derived from a distribution that ranges from potential  $k_c$  typical lows of  $\approx 0.1$  to typical highs of, but not limited to,  $\approx 1.2$ .

Smith et al. (2017) provide the distributions of  $k_c$  corresponding to both okta number and solar elevation in the UK; the distributions are detailed in the Table A.1 in Appendix A. The solar elevation angle dependency of  $k_c$  is hypothesised by Smith et al. (2017) to be a result of larger attenuation within clouds due to the longer solar path at low solar elevations. To extract a  $k_c$  value for each cloud, both the okta number and the solar elevation angle must be known for that time-step of the simulation. The solar elevation angle is calculated at each time-step following the algorithm defined by Blanco-Muriel et al. (2001) and the lowest elevation angle existing within that hour of simulation is taken as the reference with which to select the appropriate  $k_c$  distribution for use. One assumption is that the  $k_c$  distributions for okta hold accuracy when applied globally. A further assumption is that the method of producing  $\mathbf{B}$ , which assumes the sun is directly overhead, captures the coincidence of clouded and clear periods between simulated locations regardless of solar elevation angle, and that the influence of solar geometry is limited to the different  $k_c$  distributions.

It is acknowledged that different okta values will possess a myriad of different cloud types around the world, some types more prevalent than others depending on the climatic region. As is shown later in Section 3, the irradiance magnitude frequencies that are produced using these distributions suggest suitability for international use. The provided distributions are limited to elevation angles  $< 50^\circ$  and so an extrapolation was made to determine  $k_c$  distributions for elevation angles up to  $80^\circ$ . This was carried out by extrapolating the best fitting trend of both the shape and scale parameters when plotted against elevation angle. The distributions use a modified Burr and Gamma distribution, which are detailed in Appendix A alongside the distribution parameters for each okta and elevation angle. The parameters are shown in Table A.1. For overlapping clouds,  $k_c$  is found as the mean of the overlapping clouds'  $k_c$  values. This is because overlapping discs within the cloud samples are not necessarily representing more cloud as the cloud samples are only statistical representations of cloud coverage. Taking the sum of  $k_c$  at this point would lead to extreme values that do not appropriately fit the  $k_c$  distributions derived from observation data.

For each location within the simulation and at each time step,  $i$ , a time series vector of  $k_c$  values is created and denoted as  $k_c^i$ .

A further adjustment to the calculation of  $k_c$  is the inclusion of instances of cloud edge enhancement (CEE). CEE describes an event whereby a point on the Earth's surface receives a larger amount of incident irradiance than is available in the theoretical clear-sky irradiance. These events are attributed to additional irradiance reflecting off the edge of a cloud or multiple clouds. The effect is examined to a 1-s resolution by Lave et al. (2012) and simulated in space by Pecanak et al. (2016). For application in a 1-min resolution simulation, a simple approach is adopted. For the two minutes just before and two minutes just after a cloud, a stepwise fractional increase is applied. The minutes directly before and after a cloud are allowed a  $5\% \times R$  increase, where  $R \sim U(0, 1)$ , and the minutes surrounding these a  $2.5\% \times R$  increase using the same value of  $R$  for each CEE event. This implementation allows for the presence of CEE without potentially overstating its impact by limiting the increase at 5%. Calculating the clear-sky index of 1 min GHI data from BADC (2013), and assuming that  $k_c > 1$  can be attributed to CEE, increases of up to 50% have been observed at high  $\theta_z$ . 5% increases are typical under all okta values and at low  $\theta_z$ . For this reason, 5% increases are selected within the SDSIG.

The application of CEE is performed by stepping through the Boolean matrix of cloud presence,  $\mathbf{B}$ , and using a logical IF statement that queries whether  $\mathbf{B}_i$  undergoes a ramp on account of cloud cover. Should  $\mathbf{B}_i$  undergo a ramp, the equivalent time step within  $k_c$  is adjusted using the logical if statement as shown below

$$\mathbf{B}_{i-1} = 0 \ \& \ \mathbf{B}_i = 1 \longrightarrow k_c^i = k_c^i(1.5R) \ \& \ k_c^{i-1} = k_c^{i-1}(1.25R) \quad (3)$$

$$\mathbf{B}_{i-1} = 1 \ \& \ \mathbf{B}_i = 0 \longrightarrow k_c^i = k_c^i(1.5R) \ \& \ k_c^{i+1} = k_c^{i+1}(1.25R) \quad (4)$$

$k_c$  for moments of 0 okta within the SIG were drawn from the  $k_c$  distribution for an okta value of 0—a normal distribution of  $N(0.99, 0.08)$ . The most significant deviation from using the UK derived  $k_c$  distributions for application in San Diego, CA USA, was visibly fitting a new  $k_c$  distribution for 0 okta moments. Using the inbuilt Matlab distribution fitting tool (Matlab, 2015), the clear-sky peak was found to fit a normal distribution of  $N(1.02394, 0.04)$ . This adjustment is accredited to sensor offsets and climatic regional differences.

Extended periods of completely clear or overcast skies do not show much short-term variability (Skartveit and Olseth, 1992) and so for long, consecutive periods of 0 or 8 okta, a smoothing adjustment is made—periods of 8 okta lasting  $> 4$  h or for periods of 0 okta lasting  $> 3$  h. A random number of intervals is selected from 1 to 5 times the number of hours in the extended periods of okta 0 or 8. The intervals are evenly spaced throughout the duration and  $k_c$  values are drawn from the appropriate  $k_c$  distribution. The progression of  $k_c$  between intervals is filled using an inbuilt Matlab piecewise cubic hermite interpolating polynomial technique (pchip). The pchip is favourable to the next effective interpolation technique called a smoothing spline as the spline method has tendency to overshoot and oscillate if the data are not smooth (Moler, 2010). The reason for including the smoothing is that the cloud fields for 8 okta (overcast) contain many clouds in order to cover all available space. Assigning a  $k_c$  value from the distributions to each of these clouds results in large variability in irradiance. Whilst this is perhaps representative of complex convective cloud systems, it is not true of stable, consistent overcast conditions such as stratus clouds. Without the smoothing there are few periods of smooth irradiance during 8 okta periods which is unlike real irradiance observations and results in an overestimation of the variability index discussed in Section 3.

Multiple irradiance time series can now be produced for any  $x$ - $y$ - $z$  location within the  $1.5 \text{ km} \times 3600u \text{ km}$  spatial domain, each with an individual orientation and tilt.

### 3. Temporal validation

In order to demonstrate the SDSIG's capabilities, three temporal validations were carried out for (1) Cambourne, UK (2) Lerwick, UK, and (3) University of California, San Diego (UCSD), USA. A temporal validation does not consider the spatial correlation. A single synthetic irradiance time series is compared against locally measured GHI. For both UK sites, the GHI data are taken from the World Radiation Monitoring Centre Baseline Surface Radiation Network (WRMC-BSRN) (WRMC-BSRN, 2014) from station numbers 50 for Cambourne and 51 for Lerwick. The GHI data for San Diego is taken from the rooftop of the Engineering Building Unit II at UCSD (Lave et al., 2012). For both UK sites, missing data points were ignored and deemed not to significantly impact the distributions for comparison.

All data processing was performed using the Matlab r2015b (Matlab, 2015). Hourly weather observational data are taken from the British Atmospheric Data Centre's (BADC) Met Office Integrated Data Archive System (MIDAS) (BADC, 2013) for the two UK sites, and from the National Oceanic and Atmospheric Administration's (NOAA) Quality Controlled Local Climatological Data (QCLCD) (NOAA, 2016) for the USA site. As monitoring stations are occasionally taken off-line for repairs or upgrades for months at a time, where possible, more than 10 years of meteorological data are preferable to allow at least 10 years data for each variable that

requires a MTM to be created. This facilitates the statistical capturing of a typical meteorological year. 12 years of data are taken from BADC and 11 years are taken from NOAA.

The NOAA data does not come in okta format, instead the cloud is given a description at three separate levels as: Clear, 0/8; Few,  $>0/8$ –2/8; Scattered, 3/8–4/8; Broken, 5/8–7/8; or Overcast, 8/8; the fractions of 8 indicate the intended cloud cover derived from the descriptions in okta. Where a range of okta can be inferred, equal probabilities for each integer within the range are assigned. As okta values are a discreet value from a continuous measurement, they can realistically be considered  $\pm 0.5$  okta; for the description of Few clouds, the value is allowed to achieve 0 okta due to rounding, despite being  $>0$ .

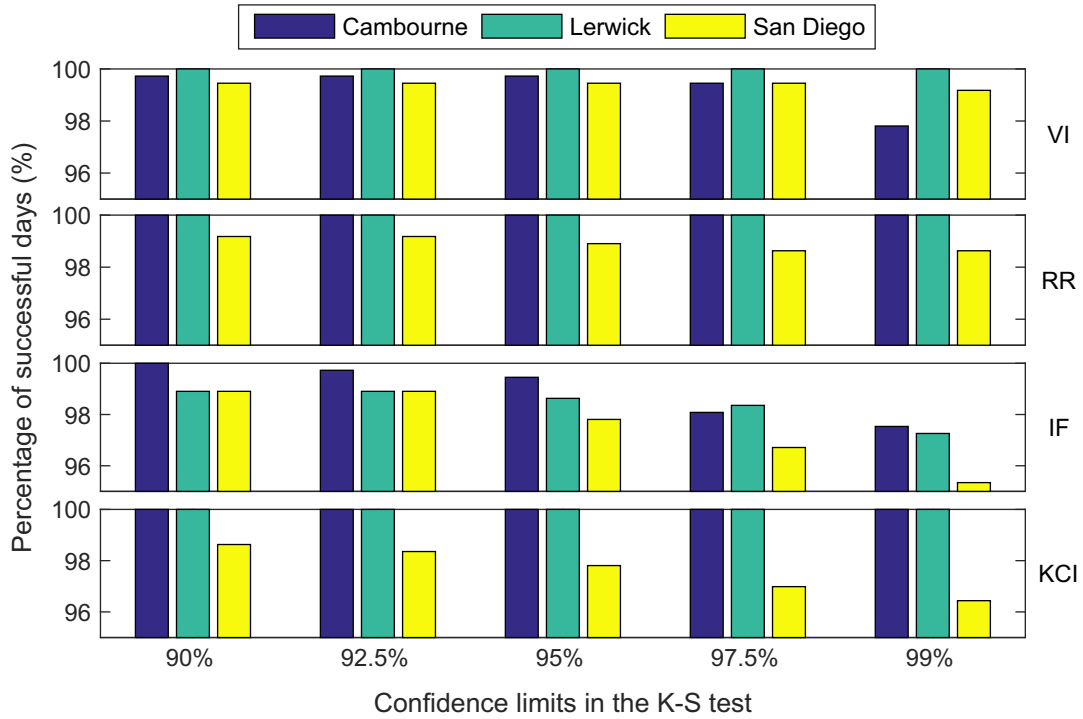
Issues exist with multiple descriptions of cloud layers. An example of a potential code for an hour of cloud cover from NOAA is "FEW BRK OVC", which details the cloud type across three separate layers. In this instance, the description with greatest associated thickness of cloud is taken as the reference—OVC—and so an a value of 8 okta would be assigned.

Four metrics are used to validate the temporal nature of the model's output: the variability index, the irradiance frequencies, the ramp rates, and the 1-min clear sky index. They are denoted as VI, IF, RR and KCI respectively and will be discussed in turn. Comparisons are made between the cumulative distribution function (CDF) profiles of both the synthetic and the observation data made from one year of 1-min values of each metric. Furthermore, the two-sample Kolmogorov–Smirnov (K–S) test was carried out for each metric, for each day of the year; it serves as a goodness of fit test between the simulated CDF profiles alongside the empirical CDFs. The K–S test takes a subset of data and tests the hypothesis that both subsets are from the same sample. In this case, the K–S test is reported as a percentage of days that satisfy the hypothesis at increasing confidence limits—a higher percentage indicates a better performance. The subset of each K–S test consisted of data from seven of the same day from 7 different years for both UK sites. For example, seven modelled samples of the 1st January represent one subset, and is compared against a subset made from seven samples of the corresponding day from observational data. Only two years of observation data are available for the San Diego validation, and so the K–S subset consists of only two days at best. This is not seen as a problem as a smaller subset will be harder to validate against. The K–S test results are displayed in Fig. 3 and comparative CDF profiles of the metrics in Fig. 4 with the correlation coefficient denoted as  $R$  and calculated using Matlab's 2-D correlation coefficient function (MathWorks, 2016).

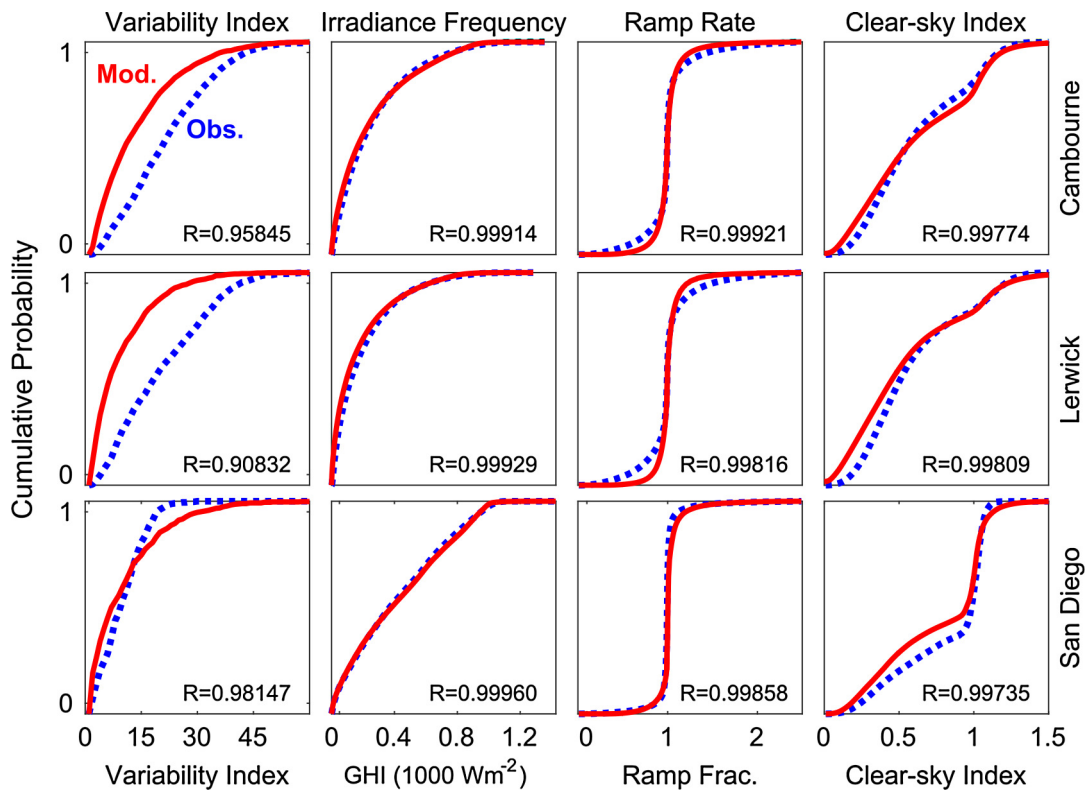
The RR are calculated every minute within the subset as the fractional change in GHI from one time-step to the next. Fig. 4 show that the RR are captured well for all locations with the CDF comparisons correlating at  $R = 0.9982$  to  $0.9992$ , and using the K–S test, 100% of days reject the null hypothesis that the modelled and observed minutely datasets are not from the same dataset with a confidence of 99% for both UK sites, whereas 98.64% of days reject the null hypothesis at 99% confidence for San Diego. The accuracy of capturing the daily RR for each day is of vital importance for suitability in grid impact studies and so this result gives confidence for use of the SDSIG in year long grid impact studies.

The IF are calculated by binning each irradiance value of both the synthetic and observation datasets to the nearest integer before creating a frequency table of each daily subset. Binning is necessary as the small subsets do not produce well defined PDFs with irradiance values at 2 decimal places. The IF CDF profiles correlate across all locations with  $R = 0.99914$  to  $0.9996$ . The K–S test result shows that Cambourne, Lerwick and San Diego have 97.53%, 97.26% and 95.34% of days respectively that reject the null hypothesis and pass the K–S test with a confidence limit of 99%. The K–S test for San Diego has the lowest percentage of suc-





**Fig. 3.** Results of the K–S test for each of the three locations for the four metrics of variability index (VI), ramp rates (RR), irradiance frequency (IF) and the clear-sky index (KCI). Percentage of successful days that pass the K–S test are shown by location for each metric at increasing K–S test confidence limits. Observation and modelled synthetic data for each day of the year are subject to the test and the percentage indicates how many of these days passed the K–S test at the indicated confidence limit.



**Fig. 4.** Comparison of CDF profiles from 1-year of observation data (blue dashed line) and synthetic modelled data (red solid line) of the four metrics, from left column to right: variability index, irradiance frequency, ramp rate occurrence and clear-sky index, at each location, from top row to bottom: Cambourne, Lerwick and San Diego. Each CDF comparison has an individual 2-D correlation coefficient displayed inside the axes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cessful days for IF. This is expected to be caused by a low sample size availability for validation. To pass the K–S test using only 2 days of observation data requires similar sky conditions in both the synthetic and observation data, e.g. comparing two cloudy days against two clear days would fail the K–S test. The result for San Diego still offers high confidence in the model's ability to statistically recreate IF for as few as two days of simulation.

The VI can be thought of as the ratio of the “length” of the irradiance line normalised by the “length” of the clear sky irradiance plotted against time (Stein et al., 2012). VI is a metric to describe how variable the irradiance is over a time period. In this case, the VI is calculated every minute and the K–S subset considers all the 1 min VI values for each day of the year. The CDFs of a year of VI correlate with  $R = 0.9585$  for Cambourne and  $0.9815$  for San Diego; K–S tests show 97.81% and 99.18% for both locations respectively with a 99% confidence. Lerwick sees VI correlation at  $R = 0.9083$  yet a K–S result showing 100% of days rejecting the null hypothesis with a 99% confidence. This indicates that whilst the VI may be slightly higher than reality, it is still within the confines of typical daily VI values. The largest discrepancy for Lerwick and Cambourne is for a VI value of 1, which is under-represented in the model. A VI of 1 is seen for a clear or an overcast day. One of the assumptions within the cloud sample production was that rounding the continuous coverage fraction to a discreet okta value of 0 allows for clouds to be present up to <5% coverage; this presence of clouds could prevent VI values equal to 1. The success of the K–S test suggests that the discrepancy in the overall CDF comparison is well distributed across the annual irradiance time series so as not to influence the daily subsets with a high confidence.

The KCI is calculated as  $k_c = G/G_{cs}$  where  $G$  is the global horizontal irradiance and  $G_{cs}$  is the global clear sky irradiance, both in  $Wm^{-2}$ . The KCI CDFs correlate highly for Cambourne, Lerwick and San Diego with  $R = 0.9977, 0.9981$  and  $0.9974$  respectively. This is a demonstration of how using a Markov chain produces a well represented distribution of okta and how the distributions of  $k_c$  weighted by both okta and elevation angle accurately create real world distributions of  $k_c$ . The K–S test results show that both Lerwick and Cambourne have 100% successful days while San Diego sees 96.44%, all with a 99% confidence limit. The lower score for San Diego is suggestive that the  $k_c$  distributions are not exact for the desert climate, however they do offer good accuracy.

#### 4. Spatial validation for Oahu, Hawaii

In order to validate the spatial dimension of the SDSIG, a test was carried out on global horizontal irradiance time series taken from the Oahu Solar Measurement Grid shown in Fig. 5 as maintained by the National Renewable Energy Laboratory (NREL) (Sengupta and Andreas, 2010).

The spatial validation irradiance time series were averaged from 1 s to 1 min in adherence with the SDSIG outputs. The data exists for 593 days from 18th March 2010 to October 31st 2011 inclusive. The input meteorological data for the SDSIG were taken from the QCLCD archive (NOAA, 2016) for station location Kalaeloa Airport, John Rodgers field, Kapolei, HI USA (ID:22551); the site is at latitude  $21.316^\circ$ , longitude  $-158.066^\circ$  and 10 m above sea level. The user defined input variables that detail each property in the SDSIG are shown in Table 1. The straight line distance between station pairs are calculated using the haversine formula using  $\Theta$  and  $\Phi$ .

The SDSIG is first subject to a temporal validation against one of the measurement sites selected at random, this is to demonstrate the SDSIG's suitability for creating statistically accurate irradiance time series for Hawaii. Using the four metrics as before of VI, RR, IF

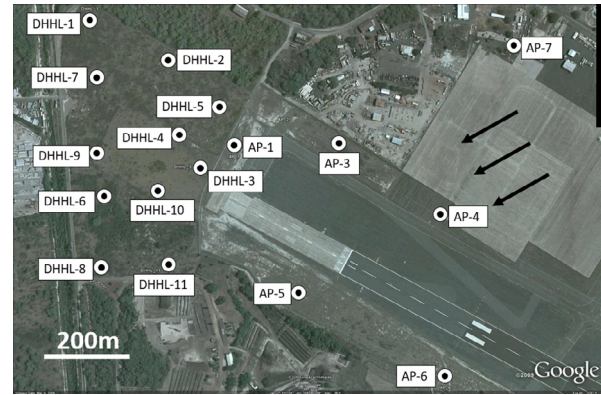


Fig. 5. Geographical layout of the measurement stations at the NREL Oahu Solar Measurement Grid (Sengupta and Andreas, 2010) on Department of Hawaiian Homeland (DHHL) and Kalaeloa Airport (AP) property. The arrows indicate the direction of prevailing wind.

Table 1

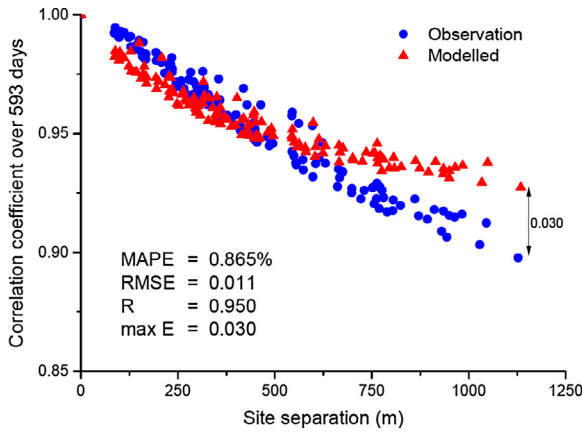
Station metadata from the NREL Oahu Radiation Monitoring Station (Sengupta and Andreas, 2010) showing the height above sea level ( $z_l$ ), latitude ( $\Theta$ ) and longitude ( $\Phi$ ) for each ID in Fig. 5.

ID	$z_l$ (m)	$\Theta$ ( $^\circ$ )	$\Phi$ ( $^\circ$ )
DHHL-1	8	21.31533	158.08700
DHHL-2	10	21.31451	158.08534
DHHL-3	9	21.31236	158.08463
DHHL-4	9	21.31303	158.08505
DHHL-5	10	21.31357	158.08424
DHHL-6	3	21.31179	158.08678
DHHL-7	9	21.31418	158.08685
DHHL-8	3	21.31034	158.08675
DHHL-9	5	21.31268	158.08688
DHHL-10	7	21.31183	158.08554
DHHL-11	7	21.31042	158.08530
AP-1	10	21.31276	158.08389
AP-3	10	21.31281	158.08163
AP-4	9	21.31141	158.07947
AP-5	7	21.30983	158.08249
AP-6	6	21.30812	158.07935
AP-7	11	21.31478	158.07785

and KCI, the correlation coefficient when comparing CDFs from modelled and observed data are  $R = 0.9825, 0.9945, 0.9990$  and  $0.9840$ , respectively. The VI metric correlates the least well for Hawaii; the SDSIG does not produce many days of complete clear sky for Hawaii that leads to  $VI = 0$ . Two possible explanations are presented for this. The first is that the SDSIG outputs are representative of a typical meteorological year and the 593 days of observation data maybe had days of non-typical clear sky stability. The second is that using only a single order Markov chain may not capture longer term weather characteristics that may prevail at the Oahu site. The K–S test results on the metrics, calculated in the same manner as in Section 3, found that 99.73%, 100%, 98.90% and 90.96% days pass the test to a 99% confidence level, respectively. The lower success rate for KCI is perhaps indicative of using okta correlations from Smith et al. (2017), which are not completely accurate for all geographic and climatic regions.

The spatial correlation between every station-pair is calculated over a time scale of 593 days. The correlation between GHI time series of two sites is calculated as the two-sample correlation coefficient (MathWorks, 2016). The correlation is plotted against the straight line separation between pairs and shown in Fig. 6.

One assumption used for the validation was to fix the cloud direction from a north-easterly of  $60^\circ$ , as was used by both Arias-Castro et al. (2014) and Hinkelman (2013).



**Fig. 6.** Correlation coefficients of individual station-pair w min GHI time series for 593 days plotted against the station-pair site separation. Every station-pair combination of the 17 site from the Oahu Solar Measurement Grid is represented as a symbol. Key: mean absolute percentage error (MAPE), root mean square error (RMSE), correlation coefficient ( $R$ ), and maximum observed error (max E).

The SDSIG outputs have a similar spatial correlation to observation data. The mean absolute percentage error for the site-separated, time-averaged correlation is 0.865%. The root mean square error is 0.01. The correlation coefficient between the data is  $R = 0.9523$ .

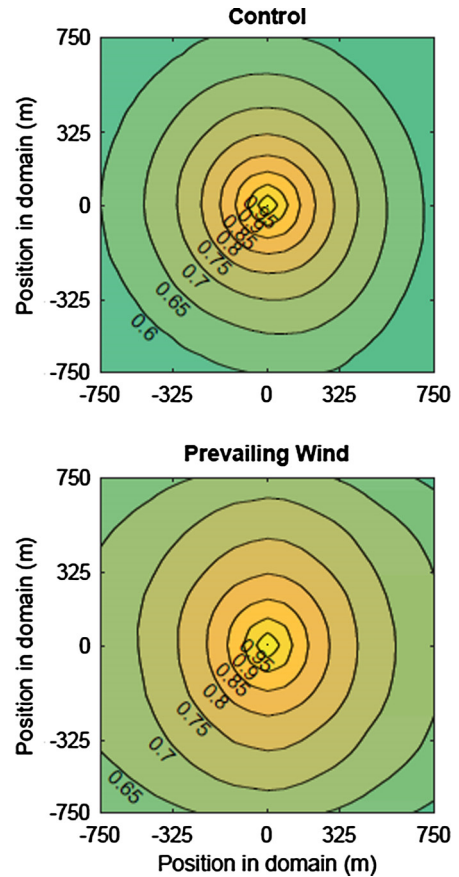
Arias-Castro et al. (2014) studied 13 days of data and found unique correlation patterns in along and cross wind station pairs. For these 13 days, the cloud motion direction was near-constant around  $60^\circ$ . The cloud direction over 593 days is variable and different cloud layers were prevalent in satellite images. One limitation within the SDSIG is that there is no empirically derived method of including cloud motion direction. Using a cloud motion direction of  $60^\circ$ , the correlation can be seen to curve away from the observation data with increasing site separation. This is expected to be caused by fixed along and cross wind station pairs in the SDSIG, yet the along and cross wind pairs change with cloud direction within the observation data. Spatial correlation is sensitive to cloud motion direction and will be discussed further in Section 5.

### 5. Illustration of instantaneous spatial correlation

Four simulations are run to explore the instantaneous spatial correlation. (1) Running the SDSIG with no influencing or weighting of any variable is labelled the “Control” scenario, it represents normal operation of the SDSIG. (2) The “Prevailing” scenario has only the cloud motion direction fixed from the same location (North) to represent constant along and cross wind directions. (3) Fixing only the cloud coverage, and (4) fixing the cloud speed with constant coverage and cloud motion direction.

The instantaneous spatial correlation is calculated by comparing a reference point located within the spatial domain to all other points in the spatial domain at the same time step and querying the sun-obscured state, found in **B**. There only exist two states of cloudiness: obscured by cloud or not obscured by cloud. When a location in the grid shares the same state of cloudiness as the reference state, the instantaneous correlation is assigned as 1, otherwise an anti-correlation of  $-1$  is assigned for opposing states. Taking the mean of these instantaneous correlations over a time period finds the overall instantaneous correlation of all locations to the reference point over that time scale. This is achieved by adopting a systematic grid/mesh approach and testing each point individually against the reference.

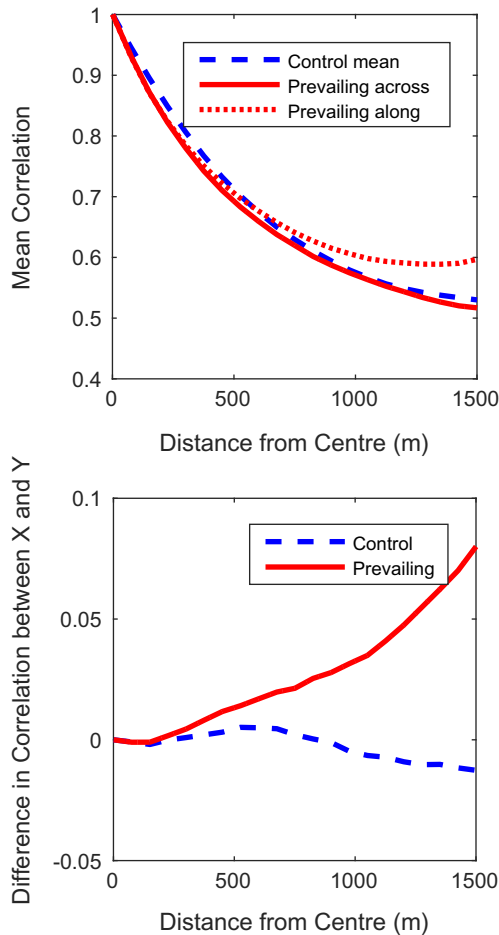
Fig. 7 shows how the instantaneous spatial correlation manifests across a time period of a year. The reference point is selected



**Fig. 7.** Instantaneous spatial correlation from the centre point at (0,0) to all other points within the spatial domain. The top plot is the control simulation where no variables were fixed whereas the bottom plot has the wind direction fixed from the north (top of plot). The numbers indicate the correlation along the labelled contour.

as the central most point within the spatial domain at (0,0). The prevailing wind scenario over a year shows less decorrelation with distance from the reference point as well as a tendency for anisotropy. The control scenario is more isotropic. Further assessment was carried out by relocating the reference point to the north-western most point within the spatial domain at  $(-750, 750)$  so that the decorrelation over a longer distance can be examined. This can be seen in Fig. 8 where the top plot demonstrates how instantaneous correlation in both the  $x$  and  $y$  directions change with distance for both scenarios. Decorrelation is observed for both scenarios, however the prevailing scenario in the along direction undergoes less decorrelation while the control and prevailing along direction follow a similar, more decorrelated trajectory. The control  $x$  and  $y$  directions were very similar and so only the mean is shown.

Fig. 7 shows that a prevailing wind scenario within the model captures the anisotropic tendency, which is also discussed by Arias-Castro et al. (2014) and Lonij et al. (2013), it also demonstrated by Lave and Kleissl (2013) and measured by Hinkelman (2013). This anisotropy is further shown in the bottom plot of Fig. 8 where the instantaneous correlation differences in the  $x$  and  $y$  are plotted against distance from the reference point within the spatial domain. For the control scenario, a steady fluctuation around a instantaneous correlation difference of 0 is observed. With a perfect distribution of cloud movement direction, the control sample becomes more circular. Throughout a year’s simulation, however, the cloud motion direction is not entirely uniform as is shown in Fig. 10 where the control simulation’s hourly cloud motion direction for the year is shown on a wind rose. Clouds



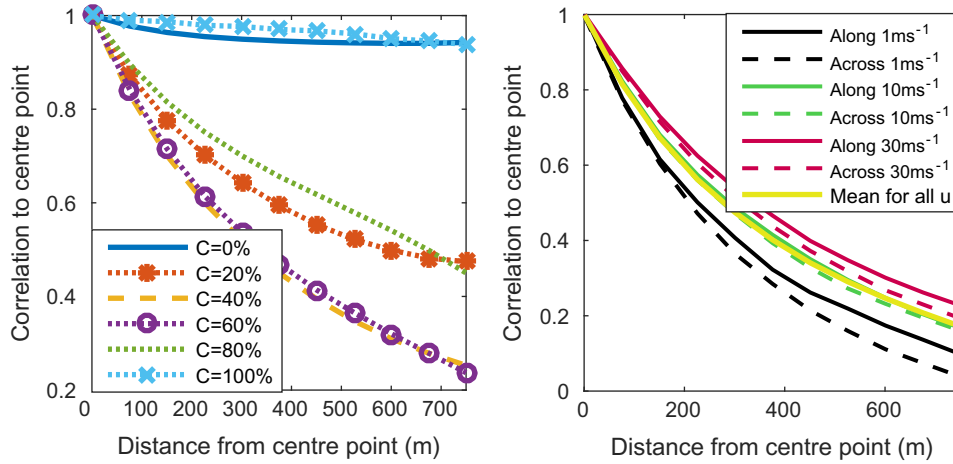
**Fig. 8.** Line plot of the instantaneous correlation from the centre point to the edge of the spatial domain for the control scenario (blue dashed line) and the prevailing wind scenario (red solid or dotted line). Top: mean instantaneous correlation over the domain for the mean of X and Y directions in the control scenario and the individual along and cross wind profiles for the prevailing scenario. Bottom: difference in instantaneous correlation between the individual X and Y directions for both the control and prevailing scenarios. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

approaching from a northerly direction dominated the simulation due to the nature of the random walk method applied. The random walk method does not produce a uniform distribution as it does not facilitate large changes in cloud direction between hours. The output is random and arbitrary as it is not weighted or biased, however it is limited in step size in order to reproduce gradual change. In this particular simulation, by chance, the random walk remained for longer approaching from a northerly direction, because of this a greater instantaneous correlation is observed in the y-axis, which explains the falling trend in correlation difference over distance. The difference in cross wind and along wind correlation for the prevailing wind scenario sees a steady increase with distance as the decorrelation in the along wind direction plateaus at 0.6 while the cross wind continues decorrelating. The expectation is that with increasing distance from the reference point and when the instantaneous spatial correlation is no longer influenced by the size of cloud and cloud motion direction, the instantaneous correlation will become a function of  $C$  for as long as both points share the same  $C$ ; Lonij et al. (2013) also observe correlation change in along and cross wind directions and see a separation up until 10 km. Anisotropy was observed to be more defined over sampling periods of a day to a month, however there was small change when observing 6 month to a year's correlation. Shorter

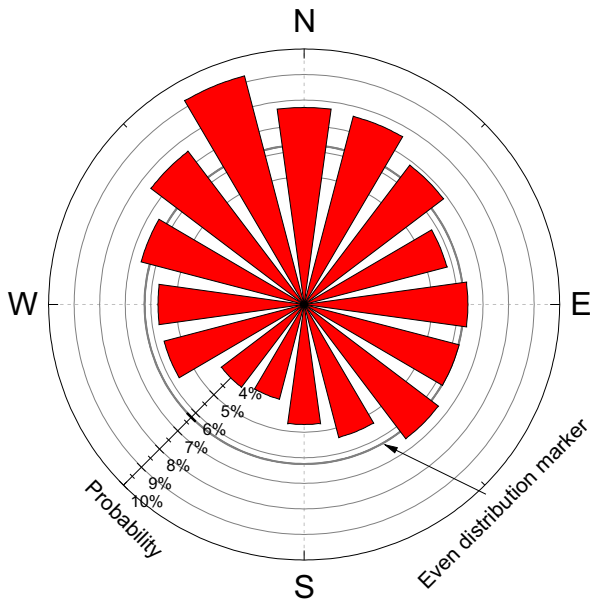
sampling periods are more prone to a prevailing wind scenario and so will favour anisotropy. With a longer sampling period, the SDSIG will offer a more evenly distributed cloud motion direction.

Fig. 9 shows how the instantaneous correlation from the centre point to the edge of the domain is influenced by fixing the cloud coverage or the cloud speed. Separate simulations were performed for each plot. The scenario for the left plot of Fig. 9 was the same as the control scenario conditions with exception of the cloud coverage, which was incrementally increased and fixed for the year. It can be observed that for both fully overcast and fully clear scenarios, the instantaneous correlation is  $> 0.95$  for the entire length of the spatial domain. This is expected as constant obscured or not obscured skies experience very few ramp occurrences. Ramp occurrences do occur for some hours of fully clear or overcast sky as the cloud samples are produced by assigning a discrete value of  $C$  through rounding the continuous data allowing  $C = C_t \pm 0.5$ . More significant is the rapid drop in instantaneous correlation for 40 to 60% cloud cover scenarios, which undergo a steep decline from 1 to 0.25 over 750 m. Cloud coverage constants of 20% and 80% share a similar instantaneous correlation regression to each other, dropping  $< 0.5$  over 750 m. The okta frequency during a year long simulation for all three study sites is dominated by cloud coverage events of 0% and 100% with probability densities of 14% and 21% respectively, where 9% is the evenly distributed probability. This dominance is why the annual instantaneous correlation over the domain only falls to 0.6 and not to the lower values shown by other coverage scenarios. Sites that have a higher correlation will suffer from synchronised ramping events, which are of high concern for DNOs. The frequency of ramp rates is also of importance; the highest correlating state of a coverage equal to 0% does not present as large an issue as the scattered cloud states, as the frequency of ramps is much lower. The more decorrelated an area the less synchronised the ramps. What is evident from Figs. 6–8 is that to achieve favourable, decorrelated conditions, the separation between locations must be maximised.

The simulation conditions for the right plot of Fig. 9 had the cloud coverage set to  $C = 5$  and the cloud direction fixed at approaching from the north as with the prevailing scenarios. The plot shows how fixing the speed of the cloud influences instantaneous correlation over the spatial domain. The cross wind directions in both east and west were similar if not the same and so only the mean of both is shown. The along direction is the mean of the correlations in both the north and south directions; the upwind correlation showed slightly more decorrelation at all cloud speeds for unknown reasons. Cloud speeds of  $u = 1, 10$  and  $30 \text{ ms}^{-1}$  are shown to represent the upper and lower limits of the cloud speed range as well as the mode. In the control scenario, the probability of occurrence for a cloud motion of  $\geq 25 \text{ ms}^{-1}$  is  $\leq 0.0114\%$ , speeds between 1 and  $10 \text{ ms}^{-1}$  dominate the simulation cloud speed frequency. The instantaneous spatial correlation increases with  $u$ ; analysis intervals were  $5 \text{ ms}^{-1}$ . As  $C$ ,  $u$  and cloud direction are fixed for each simulation, the variation in correlation must be independent of these variables. One possibility is the influence of cloud size. By nature of the cloud sample production technique, larger clouds are more prevalent at higher wind speeds as adding a large cloud to the lower coverage values causes  $[C] > [C_t]$  and the cloud sample must be reset. This is intuitive as a large cloud cannot exist in a near cloudless environment. At higher  $C_t$ , as more clouds are required to fill the cloud sample, it causes long overlaps that effectively increase the cloud size. This is further explained by Arias-Castro et al. (2014), who demonstrated that if two locations are covered by the same cloud, the instantaneous correlation will be higher and therefore larger clouds offer greater correlation. There is increasing convergence in the instantaneous correlation with increasing  $u$  in both the along and cross wind directions. The cause of this is suspected to be that,



**Fig. 9.** Line plot of the mean instantaneous correlation from the centre point to the edge of the spatial domain. The left plot shows the results of simulations with fixed cloud coverage percentages and all other variables operate as normal. The right plot shows how the along (solid) and cross (dashed) wind correlation at different cloud speeds, as well as the mean instantaneous correlation of all wind speeds in solid yellow. The simulation fixed the variables cloud coverage to  $C = 5$  and the cloud direction from the north. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 10.** Wind rose showing the probability density of the hourly cloud motion approach direction over 1-year simulation. Each bin represents 22.5°. The direction indicates where the clouds are travelling from. The grey thick grey even distribution marker is a probability of 6.25% and represents perfectly even distribution.

as the cloud speed increases, two locations will not be under the shading of the same cloud for as many time steps as with low cloud speeds. This means that the instantaneous correlation becomes more a function of  $C$  than the direction or  $u$ . This is further explained in a study by Munkhammar et al. (2017) who note that the instantaneous correlations are not dependent on the cloud speed.

## 6. Application in a grid impact assessment study

### 6.1. Overview of the power flow study

In order to show the advantages of using synthetic irradiance time series that are unique for each PV system, a comparison study

is conducted by simulating steady-state power flow on distribution systems.

In the first scenario, spatially decorrelating and temporally unique synthetic irradiance time series from the SDSIG are assigned for each available PV system in the distribution system. In the second scenario, a single irradiance time series randomly selected from the SDSIG is used for all available PV systems, and so is a correlating irradiance time series for all sites. The simulations are done as individual daily simulations for a complete year (365 days) of data and for varying solar PV penetration levels. Power flow simulations are conducted using OpenDSS (Institute, 2008), an open source electric power distribution system simulator.

### 6.2. Test circuit and data sources

The IEEE 8500-node test circuit is chosen to build the distribution system as it is publicly available, well-documented and well-tested. This circuit is a radial distribution feeder with multiple feeder regulators and capacitors (Arritt and Dugan, 2010); it resembles a large network with many common power system elements found in a residential distribution feeder. The longest distance from the substation is approximately 17 km and the circuit has a peak load of approximately 10.7 MW.

The PV generation is assumed to be rooftop systems distributed in the test circuit. Each PV system is installed to the secondary side of the service transformer adjacent to the respective load point. Each system is specified to have a capacity equal to the peak demand of that load point and their tilt angles are kept at 20° and their azimuth angles are kept at 180°. Each PV system is coupled with an inverter that has an efficiency of 0.95 and operates at unity power factor. The inverters are sized to match the PV system nameplate capacity at the location. The resulting AC power profiles are directly fed into OpenDSS using its "LoadShape" object.

PV systems are sited among randomly chosen load points of the circuit until a desired PV penetration is reached. As the PV penetration level in the distribution system is increased, already existing PV systems are kept fixed and new systems are added in a similar random fashion until the desired PV penetration is reached again. PV penetration (PVP) definition used in this study is as follows:

$$PVP = \frac{\sum_{m=1}^M P_{PV}^m}{\sum_{n=1}^N P_{load}^n} \tag{5}$$

where  $M$  is the total number of PV systems adopted in the distribution system,  $N$  is the total number of load points in the circuit and  $P_{load}$  and  $P_{PV}$  indicate peak rated power of load points and PV systems, respectively. PVP levels used in this study are 0%, 25%, 50% and 75%. Higher PVP simulations are not considered as voltage regulation using tap changers alone cannot maintain the voltage targets in the distribution system during high solar and low demand periods on certain simulation days.

Simulations are carried out using 1-min resolution demand and solar generation data. Generic demand profiles for residential buildings in San Diego are imported from the dataset provided by Open Energy Information (OpenEI, 2014). The synthetic irradiance time series are generated using the SDSIG. The PV system power outputs are computed using a power conversion model for distributed PV systems presented in Jamaly et al. (2013).

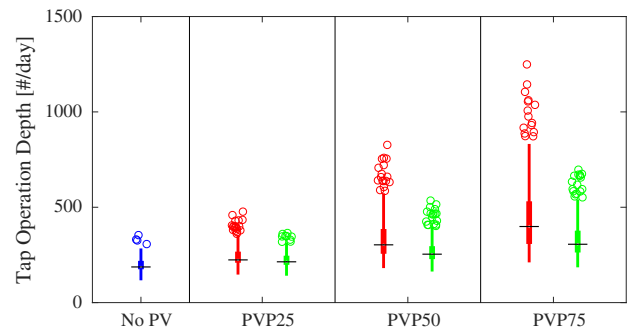
### 6.3. Grid impact metric and results

Voltage in a distribution system fluctuates due to local variations in real power or reactive power injections and/or absorptions due to changes in power supply and demand. In the scenarios considered in this study, voltage fluctuations are either caused by location variations in real power injections by solar PV, or real power absorptions by local demand. The sensitivity of both data sets to the distributed solar PV generation impacts are determined by computing the increase in the number of OLTC operations required to maintain the voltage within pre-defined limits of the test circuit.

The voltage regulation decisions for the simulations are output by the “RegControl” object defined in OpenDSS. The target voltage and bandwidth definitions of the voltage regulators are kept the same as in the original test circuit. To the authors’ knowledge, there is no consensus metric to compare grid impacts. The number of OLTC operations is a widely used Yan et al. (2014), Lave et al. (2015), Nguyen et al. (2016) and relevant metric and therefore used to demonstrate the impact upon the distribution grid. In each power flow scenario, the accumulated depth of required OLTC operations at all voltage regulators are recorded for day-long simulations.

The resulting values are given in Fig. 11 for a year of simulations for varying PV penetration levels. No PV case has a median depth of tap operations of 187 per day, 4 outliers and a maximum of 354 per day. The power flow scenario using the single correlated time series data is named herein as the red case and the scenario using the spatially decorrelated and temporally unique synthetic GHI time series is named as the green case following the colour choice in Fig. 11. The results for all PVP levels are as follows: For 25% PVP simulations, red case has 18 outliers with a maximum of 477 per day, green case has 9 outliers with a maximum of 365 per day; for 50% PVP simulations, red case has 17 outliers with a maximum of 827 per day, green case has 18 outliers with a maximum of 535 per day; for 75% PVP simulations, red case has 15 outliers with a maximum of 1249 per day, green case has 17 outliers with a maximum of 697 per day. Red cases have a median that is higher than green cases with increasing PV penetration by 10, 49 and 93 per day, respectively.

For both the uniform and decorrelated irradiance time series, tap operations increase with PV penetration, as expected. However, the magnitude and frequency of severe tap changing events are rising significantly faster in the simulations using the single GHI time series. In very high PV penetration cases such as a distribution system with 75% PV penetration, using the single GHI time series data set results in cases where voltage regulators change taps up to an accumulated depth of 1249 compared to an accumulated depth of 697 in the simulation using the spatially and temporally unique synthetic GHI time series. Such extreme results would require very



**Fig. 11.** Number of tap operations over a day for a complete year (365 days) of simulation under varying PV penetration level in the IEEE8500 node distribution system (0%, 25%, 50% and 75%). Red colour (left-side of each bracket) represents the power flow scenario using the single correlated time series data across the distribution system and green colour (right-side of each bracket) represents the scenario using the spatially decorrelated and temporally unique synthetic GHI time series. Blue colour represents the No PV case. The black horizontal bar indicates the median. The thick vertical lines show the upper and lower quartiles. Thin vertical lines extend between maximum and minimum values excluding the outliers. The outliers are shown as circles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conservative PV impact mitigation measures for the distribution system in question.

## 7. Conclusions and future work

### 7.1. Conclusions

This paper presented the SDSIG methodology that successfully generates spatially decorrelating irradiance time series from mean hourly weather observation data that validate temporally at four locations across two continents and climates: Lerwick, UK; Cambridge, UK; San Diego, CA USA; and Oahu, HI USA.

A temporal validation was carried out using four metrics comparing the correlation between CDF observation data and the synthetic model output data. The K–S test using increasing confidence limits from 90% to 99% is also carried out on daily subsets containing 1-min resolution irradiance data comparing the goodness of fit of each metric. The metrics are the variability index, ramp rate size, irradiance frequency and the clear sky index. Each metric passed the K–S test with 99% confidence limit with minimum success of 90.96% and average score of 95.99%. The CDF correlations for all metrics correlated with a minimum of  $R = 0.90832$  and mean of  $R = 0.98721$ .

A spatial validation was carried out by calculating the correlation coefficient of 593 day long GHI time series between every station-pair combination of 17 irradiance measurement devices from NREL’s Oahu Solar Measurement Grid. This was compared to the SDSIG outputs subject to the same testing. The mean absolute percentage error was 0.8648% and the correlation coefficient was  $R = 0.95523$ .

The model demonstrates the instantaneous spatial decorrelation of the output time series. The instantaneous correlation is shown to behave anisotropically with a fixed cloud motion direction. Spatial decorrelation is shown to become more isotropic with a more evenly distributed cloud motion direction. With the analysis time scale at 1-year, the minimum instantaneous correlation observed across 1.5 km separation was 0.52 for both prevailing and control scenarios. The most noticeable decorrelation caused by cloud cover conditions is under scenarios of 40% to 60% cloud cover, whereas the least decorrelation is for 0% and 100% cloud cover. Increasing cloud movement speed is shown to increase both the along and across wind instantaneous correlation with distance,

this is suspected to be a result of the tendency for increased cloud size with higher  $u$  and  $C_t$  in the cloud sample production. Furthermore, there is increasing instantaneous correlation convergence with increasing  $u$ . This is suspected to be due to clouds passing too quickly to be captured in detail in a 1-min resolution study, resulting in instantaneous correlation becoming a function of  $C$ .

From the application into the power flow study, it was shown that the magnitude and frequency of severe tap changing events are significantly higher in the simulations using a single GHI time series when compared to assigning individually correlating GHI time series to each location. Using a single GHI time series for PV penetration scenarios of 25%, 50% and 75% showed an increase in the tap operation depth which have a median that is higher by 10, 49, and 93 per day with increasing PV penetration respectively than scenarios using spatially decorrelating and temporally unique GHI time series. The spatially decorrelating and temporally unique synthetic GHI time series would allow the grid operator to determine more realistic PV mitigation measure estimates and would avoid over-investments in voltage regulation equipments.

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., 2016).

## 7.2. Future work

There are certain aspects of the methodology that could be improved with further substantial research, they are detailed here.

Whilst this work has demonstrated geographic flexibility by validating in desert, island and temperate climates, there is more that can be done to allow for geographic flexibility as a function of just longitude or latitude. Analysis of clear-sky index distribution with cloud cover for different climatic regions following the example by Smith et al. (2017) would be of interest as well as the inclusion of long time scale fluctuations such as seasons, as is shown by Perez and Fthenakis (2015).

The CEE could be applied as a function of space as opposed to time. As the temporal resolution is increased, it will become more poignant to explore the CEE spatial influence and include it within the model. Further research into the probability of occurrence and magnitudes of CEE would also better help guide their inclusion in the SDSIG.

Analysis of the minimum cut-off duration for both 0 and 8 okta smoothing periods would benefit this methodology. The use of  $>3$  h for 0 okta within the SDSIG is a conservative estimation and is potentially as low as 30 min, this would reduce the VI in the SDSIG and potentially improve the temporal validation.

Shading due to terrain topography, such as significant blocking mountainous landscapes, would be an interesting inclusion to the SDSIG. Currently, only the topographical height is utilised for the Blanco-Muriel et al. (2001) sun-Earth calculations. Geographic areas of significant height differences would change the irradiance availability at certain times. It is noted that this would only improve accuracy of periods with lower irradiance, however total harmonic distortion is reportedly at its highest during power outputs that are below 20–25% of the rated capacity (Du et al., 2013; Fekete et al., 2012). The EN 50160 states that harmonics up to the 40th harmonic must remain within 8% of their nominal value and so it is possibly a useful and interesting improvement to the SDSIG.

Wind direction at measured height is not representative of the motion of clouds, which can have multiple layers travelling in different directions. To maintain the rationale of the work of using simple and readily available inputs, methods to determine the cloud direction such as Wang et al. (2016, 2011) require sophisticated equipment and so are not suitable in this study. Methods or statistics that could be simply employed to estimate realistic

cloud direction and general cloud motion would be an imperative inclusion. Current work at Leeds is examining this and has preliminarily shown positive correlation between ground based wind speed measurements and the calculated cloud speed based of triangulated pyranometer GHI data.

Perhaps most noticeable with Lerwick, UK, longer term cloud cover patterns are more prevalent. The weather systems that cross Scotland can be many days long, and so using a single order Markov chain process offers a limitation in that, currently, only the cloud cover of a single previous hour is considered within the stochastic element. Increasing the Markov chain power order could potentially capture weather patterns better. Furthermore, it would be interesting to see analysis of the variability index in the same way the okta is studied, this may enable it to be applied using a Markov chain process and may help to identify any statistical significances that may better guide the cloud cover time series.

Transition to 1-s resolution is possible with this methodology.

## Acknowledgements

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). Thanks are offered to NREL for providing free access to the Oahu Solar Measurement Grid.

## Appendix A. Clear-sky distributions

Smith et al. (2017) details the distribution probability density functions (PDFs) for both the Burr and Gamma distributions (see Table A.1).

The PDF of the Burr (type III) distribution

$$f(x) = \frac{ck}{a} \left(\frac{x}{a}\right)^{-c-1} \left(1 + \left(\frac{x}{a}\right)^{-c}\right)^{-k-1} \quad (\text{A.1})$$

where  $c$  and  $k$  are positive shape parameters and  $a$  is a positive scale parameter.

The generalised gamma is a superset of several common distributions used in mathematics and engineering, and includes the gamma, exponential, Weibull, chi-squared, normal and lognormal distributions as special or limiting cases. The PDF is given by

$$f(x) = \frac{px^{d-1} \exp(-(x/a)^p)}{a^d \Gamma(d/p)} \quad (\text{A.2})$$

where  $a$  is the scale parameter,  $p$  and  $d$  are shape parameters.  $\Gamma(\cdot)$  is the gamma function that generalises factorials to all real numbers.

## References

- Arias-Castro, E., Kleissl, J., Lave, M., 2014. A Poisson model for anisotropic solar ramp rate correlations. *Sol. Energy* 101, 192–202.
- Arritt, R.F., Dugan, R.C., April 2010. The IEEE 8500-node test feeder. In: IEEE PES T D 2010, pp. 1–6.
- BADC, 2013. British Atmospheric Data Centre – National Centre for Atmospheric Science – Natural Environment Research Council. <<http://badc.nerc.ac.uk>>.
- Beyer, H.G., Luther, J., Schumacher-Gröhn, J., 1991. Comparison of some standard pv simulation programs with the simulation system insel. In: Conference Proceedings, 175–180 Solar World Congress. Proc. of the Biennial Congress of the ISES, Denver, Colorado, USA, 1923 August 1991. Pergamon Press, Oxford, New York.
- Blanco-Muriel, M., Alarcon-Padilla, D., Lopez-Moratalla, T., Lara-Coira, M., 2001. Computing the solar vector. *Sol. Energy* 70 (5), 431–441.
- Bright, J., Smith, C., Taylor, P., Crook, R., 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. *Sol. Energy* 115, 229–242.
- Bright, J.M., Smith, C.J., Taylor, P.G., Crook, R., 2016. The Bright Solar Resource Model. <<http://jamiembright.github.io/BrightSolarModel/>>.
- Caldon, R., Coppo, M., Turri, R., 2014. Distributed voltage control strategy for LV networks with inverter-interfaced generators. *Electr. Power Syst. Res.* 107, 85–92.

- Calinoui, 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.
- Carvalho, P.M.S., Correia, P.F., Ferreira, L.A.F., 2008. Distributed reactive power generation control for voltage rise mitigation in distribution networks. *IEEE Trans. Power Syst.* 23 (2), 766–772.
- Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol. Energy* 85 (11), 2881–2893.
- Curtright, A.E., Apt, J., 2008. The character of power output from utility-scale photovoltaic systems. *Prog. Photovolt.: Res. Appl.* 16 (3), 241–247.
- Du, Y., Lu, D.D.-C., James, G., Cornforth, D.J., 2013. Modeling and analysis of current harmonic distortion from grid connected PV inverters under different operating conditions. *Sol. Energy* 94 (0), 182–194.
- Engerer, N., Mills, F., 2014. Kpv: a clear-sky index for photovoltaics. *Sol. Energy* 105, 679–693.
- Fekete, K., Klaić, Z., Majdandžić, L., 2012. Expansion of the residential photovoltaic systems and its harmonic impact on the distribution grid. *Renew. Energy* 43 (0), 140–148.
- Gafurov, T., Usaola, J., Prodanovic, M., 2015. Incorporating spatial correlation into stochastic generation of solar radiation data. *Sol. Energy* 115, 74–84.
- Glasbey, C., Graham, R., Hunter, A., 2001. Spatio-temporal variability of solar energy across a region: a statistical modelling approach. *Sol. Energy* 70 (4), 373–381.
- Gueymard, C.A., Wilcox, S.M., 2011. Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data. *Sol. Energy* 85 (5), 1068–1084.
- Hammer, A., Heinemann, D., Hoyer, C., Kuhlmann, R., Lorenz, E., Müller, R., Beyer, H., 2003. Solar energy assessment using remote sensing technologies. *Remote Sensing Environ.* 86, 423–432.
- Hernandez, J., la Cruz, J.D., Ogayar, B., 2012. Electrical protection for the grid-interconnection of photovoltaic-distributed generation. *Electr. Power Syst. Res.* 89 (0), 85–99.
- Hinkelman, L.M., 2013. Differences between along-wind and cross-wind solar irradiance variability on small spatial scales. *Sol. Energy* 88, 192–203.
- Hoff, T.E., Perez, R., 2010. Quantifying PV power output variability. *Sol. Energy* 84 (10), 1782–1793.
- Inman, R.H., Pedro, H.T., Coimbra, C.F., 2013. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* 39 (6), 535–576.
- Institute, E.P.R., 2008. Open distribution system simulator (OpenDSS) version 7.6.4.36. Tech. rep., EPRI. <<http://smartgrid.epri.com/SimulationTool.aspx>>.
- Jamaly, M., Bosch, J.L., Kleissl, J., March 2013. A power conversion model for distributed PV systems in California using solaranywhere irradiation. Tech. rep., UCSD.
- Klucher, T., 1979. Evaluation of models to predict insolation on tilted surfaces. *Sol. Energy* 23, 111–114.
- Kumar, A., Gomathinayagam, S., Giridhar, G., Mitra, I., Vashistha, R., Meyer, R., Schwandt, M., Chhatbar, K., 2013. Field experiences with the operation of solar radiation resource assessment stations in India. *Energy Proc.* 49, 2351–2361. cited By 4.
- Larraneta, M., Moreno-Tejera, S., Silva-Perez, M., Lillo-Bravo, I., 2015. An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series. *Sol. Energy* 122, 517–528.
- Lave, M., Kleissl, J., 2010. Solar variability of four sites across the state of Colorado. *Renew. Energy* 35 (12), 2867–2873.
- 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 Sol. Energy 3.
- Lave, M., Kleissl, J., Stein, J.S., 2013. A wavelet-based variability model (wvm) for solar PV power plants. *IEEE Trans. Sustain. Energy* 4 (2), 501–509.
- Lave, M., Reno, M.J., Broderick, R.J., 2015. Characterizing local high-frequency solar variability and its impact to distribution studies. *Sol. Energy* 118, 327–337.
- Lonij, V.P., Brooks, A.E., Cronin, A.D., Leuthold, M., Koch, K., 2013. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Sol. Energy* 97, 58–66.
- Marcos, J., Marroyo, L., Lorenzo, E., Alvira, D., Izco, E., 2011. Power output fluctuations in large scale PV plants: one year observations with one second resolution and a derived analytic model. *Prog. Photovolt.: Res. Appl.* 19 (2), 218–227.
- Marcos, J., Marroyo, L., Lorenzo, E., García, M., 2012. Smoothing of PV power fluctuations by geographical dispersion. *Prog. Photovolt.: Res. Appl.* 20 (2), 226–237.
- Marcos, J., de la Parra, Í., García, M., Marroyo, L., 2016. Simulating the variability of dispersed large PV plants. *Prog. Photovolt.: Res. Appl.* 24 (5), 680–691.
- MathWorks, 2016. 2-D correlation coefficients – Matlab corr2 (accessed on: 04/04/2016).
- Matlab, 2015. version 8.6.0.267246 R2015b. The MathWorks Inc., Natick, Massachusetts.
- Mokhtari, G., Nourbakhsh, G., Zare, F., Ghosh, A., 2013. Overvoltage prevention in LV smart grid using customer resources coordination. *Energy Build.* 61, 387–395.
- Moler, C., 2010. Numerical Computing with MATLAB: Revised Reprint. Society for Industrial and Applied Mathematics.
- Moreno-Tejera, S., Silva-Perez, M., Lillo-Bravo, I., Ramirez-Santigosa, L., 2016. Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. *Sol. Energy* 132, 430–441.
- Morf, H., 2013. A stochastic solar irradiance model adjusted on the Angstrom–Prescott regression. *Sol. Energy* 87 (0), 1–21.
- Mouheeb, M., Hamidat, A., Loukarfi, L., 2012. Impact of PV compensation in improving the voltage drop in electrical networks LV. *Energy Proc.* 18 (0), 751–761. terragreen 2012: Clean Energy Solutions for Sustainable Environmen (CESSE).
- Müller, R., Trentmann, J., 2010. Algorithm theoretical baseline document direct irradiance at surface. Tech. rep., EUMETSAT Satellite Application Facility on Climate Monitoring.
- Munkhammar, J., Widn, J., Hinkelman, L.M., 2017. A copula method for simulating correlated instantaneous solar irradiance in spatial networks. *Sol. Energy* 143, 10–21.
- 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.
- Nguyen, A., Velay, M., Schoene, J., Zheglov, V., Kurtz, B., Murray, K., Torre, B., Kleissl, J., 2016. High PV penetration impacts on five local distribution networks using high resolution solar resource assessment with sky imager and quasi-steady state distribution system simulations. *Sol. Energy* 132, 221–235.
- NOAA, 2016. Quality controlled local climatological data (QCLCD). <<http://www.ncdc.noaa.gov/qclcd/>> (accessed on: 31/01/2016).
- OpenEI, 2014. Commercial and residential hourly load profiles for all TMY3 locations in the United States. Tech. rep., Office of Energy Efficiency Renewable Energy. <<http://en.openei.org>>.
- Otani, K., Minowa, J., Kurokawa, K., 1997. Study on areal solar irradiance for analyzing areally-totalized PV systems. *Sol. Energy Mater. Sol. Cells* 47 (14), 281–288.
- Paatero, J.V., Lund, P.D., 2007. Effects of large-scale photovoltaic power integration on electricity distribution networks. *Renew. Energy* 32 (2), 216–234.
- Pecenak, Z.K., Mejia, F.A., Kurtz, B., Evan, A., Kleissl, J., 2016. Simulating irradiance enhancement dependence on cloud optical depth and solar zenith angle. *Sol. Energy* 136, 675–681.
- Perez, M.J., Fthenakis, V.M., 2015. On the spatial decorrelation of stochastic solar resource variability at long timescales. *Sol. Energy* 117, 46–58.
- Perez, R., Kivalov, S., Schlemmer, J., Jr., K.H., Ho, T., 2011. Parameterization of site-specific short-term irradiance variability. *Sol. Energy* 85 (7), 1343–1353.
- Sayeeff, S., Heslop, S., Cornforth, D., Moore, T., Percy, S., Ward, J.K., Berry, A., Rowe, D., 2012. Solar intermittency: Australia's clean energy challenge? characterising the effect of high penetration solar intermittency on Australian electricity networks. Tech. rep., CSIRO.
- Segui-Chilet, S., Gimeno-Sales, F., Orts, S., Garcera, G., Figueres, E., Alcaniz, M., Masot, R., 2007. Approach to unbalance power active compensation under linear load unbalances and fundamental voltage asymmetries. *Int. J. Electr. Power Energy Syst.* 29 (7), 526–539.
- Sengupta, M., Andreas, A., 2010. Oahu solar measurement grid (1-year archive): 1-second solar irradiance; Oahu, Hawaii (data); NREL report no. DA-5500-56506. Tech. rep., NREL.
- Skartveit, A., Olseth, J., 1992. The probability density and autocorrelation of short-term global and beam irradiance. *Sol. Energy* 49 (6), 477–487.
- Smith, C.J., Bright, J.M., Crook, R., 2017. Cloud cover effect of clear-sky index distributions and differences between human and automatic cloud observations. *Sol. Energy* 144, 10–21.
- 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.
- Suehrcke, H., McCormick, P., 1989. Solar radiation utilizability. *Sol. Energy* 43 (6), 339–345.
- Thomson, M., Infield, D., 2007. Impact of widespread photovoltaics generation on distribution systems. *IET Renew. Power Gener.* 1 (1), 33–40. cited By 149.
- UKMO, 2015. UK Met Office Integrated Data Archive System (MIDAS) land and marine surface stations data (1853-current). NCAS British Atmospheric Data Centre. URL <http://badc.nerc.ac.uk>.
- Vindel, J., Polo, J., 2014. Intermittency and variability of daily solar irradiation. *Atmos. Res.* 143 (0), 313–327.
- Wang, G., Kurtz, B., Kleissl, J., 2016. Cloud base height from sky imager and cloud speed sensor. *Sol. Energy* 131, 208–221.
- Widen, J., 2015. A model of spatially integrated solar irradiance variability based on logarithmic station-pair correlations. *Sol. Energy* 122, 1409–1424.
- Widen, J., Wckelgrd, E., Paatero, J., Lund, P., 2010. Impacts of distributed photovoltaics on network voltages: Stochastic simulations of three Swedish low-voltage distribution grids. *Electr. Power Syst. Res.* 80 (12), 1562–1571.
- Wiemken, E., Beyer, H., Heydenreich, W., Kiefer, K., 2001. Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany. *Sol. Energy* 70 (6), 513–518. grid-Connected PV.
- Wood, R., Field, P.R., 2011. The distribution of cloud horizontal sizes. *J. Climate* 24, 4800–4816.
- WRMC-BSRN, 2014. World radiation monitoring center – baseline surface radiation networkPANGAEA – data publisher for earth and environmental science. <<http://bsrn.awi.de/>>.
- Yan, R., Marais, B., Saha, T.K., 2014. Impacts of residential photovoltaic power fluctuation on on-load tap changer operation and a solution using DSTATCOM. *Electr. Power Syst. Res.* 111 (0), 185–193.