



UNIVERSITY OF LEEDS

PhD Thesis

**Development of a synthetic solar irradiance
generator that produces time series with high
temporal and spatial resolutions using readily
available mean hourly observations**

Author:

James Matthew BRIGHT

Supervisors:

Dr Rolf CROOK

Prof Peter G. TAYLOR



Submitted in accordance with the requirements for the degree of PhD in Low
Carbon Technologies

School of Chemical and Process Engineering
Doctoral Training Centre in Low Carbon Technologies

June 2017

Author James M. Bright
pm08jmb@leeds.ac.uk
jamielbright1@gmail.com
School of Chemical and Process Engineering
University of Leeds, UK

Supervisor Dr Rolf Crook
r.crook@leeds.ac.uk
School of Chemical and Process Engineering
University of Leeds, UK

Supervisor Professor Peter Taylor
p.g.taylor@leeds.ac.uk
School of Chemical and Process Engineering
University of Leeds, UK

External Examiner Professor Ralph Gottschalg
r.gottschalg@lboro.ac.uk
School of Mechanical, Electrical and Manufacturing Engineering
University of Loughborough, UK

Internal Examiner Dr Benjamin Chong
b.chong@leeds.ac.uk
School of Electronic and Electrical Engineering
University of Leeds, UK

Internal Chair Professor Tim Cockerill
t.cockerill@leeds.ac.uk
School of Chemical and Process Engineering
University of Leeds, UK

“You can only do your best.”

Brian Sykes (1933-2014)

Declaration of authorship

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated within the text. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2017 The University of Leeds and James Matthew Bright

Signed:

Date:

Acknowledgements

First and foremost I would like to thank my supervisors Dr Rolf Crook and Prof Peter Taylor. Regular meetings and discussions have proven fun, productive, creative and insightful; the shape and nature of this thesis is a direct consequence of such gatherings. Our unspoken agreement to be honest and open has been a crucial element of progress. Having such a free and genuine environment to develop ideas and resolve issues has been a fundamental factor for the completion of this work.

My thanks go to my examiners Prof Ralph Gottschalg and Dr Benjamin Chong. It is far from a simple task to examine a PhD thesis; I very much appreciate the time taken to read, consider and critique my work.

The support structure of the Doctoral Training Centre (DTC) has been incredibly valuable to me, enabling me to explore every interest, both professionally and personally. The DTC has encouraged and afforded me wonderful placement and training opportunities and I appreciate the help offered by James McKay and Emily Bryan-Kinns for facilitating them.

My sincere thanks to Dr Christopher Smith who worked with me on publications; he is always forthcoming and friendly with advice and ideas. Extended also to Oytun Babacan and Professor Jan Kleissl, who showed me that long-distance collaborations over many time zones can be a pleasant and enjoyable experience.

The DTC students and colleagues have become my close friends and allies over the last 4 years. They helped create a comfortable and productive office environment with which to share the PhD experience. I am indebted to them all for this.

My girlfriend, Anna Flanagan, has supported me immeasurably; the final stages of the PhD would have been difficult without her. Lastly, I thank my family, my close friends and the Flanagan's for their support. In particular, my Grandad, Brian Sykes, who was a tremendous influence on my life and will continue to be despite his sad departure during my studies.

Abstract

Photovoltaics (PV) have seen rapid global penetration into the low voltage (LV) electricity distribution grid year-on-year. The result of high PV penetration levels is grid impacts of voltage fluctuations, harmonic distortions and reverse flow among others. Research that attempts to quantify the maximum allowable PV penetration into the LV grid before experiencing detrimental impacts is an important. The most commonly reported barrier to enabling grid impact analysis is the lacking availability of high-resolution and geographically flexible solar irradiance data. As an alternative, synthetically generated solar irradiance data can be used.

There is a distinct lack of synthetic solar irradiance generators that can derive high resolution and statistically accurate solar irradiance data using only readily available inputs. This thesis presents the development of two synthetic generators: the Solar Irradiance Generator (SIG), and the Spatially Decorrelating Solar Irradiance Generator (SDSIG). The SIG proves the concept that synthetic minutely irradiance time series can be generated using readily available mean hourly observations of total cloud amount, atmospheric pressure, wind speed and cloud base height. The SDSIG presents the first ever methodology to synthetically generate unique and spatially decorrelating minutely irradiance time series for any number of uniquely orientated and tilted houses inside a spatial domain using the same inputs as the SIG. The SDSIG employs (1) Markov chains, to derive stochastic weather variable time series, (2) synthetic representations of clouds in the sky, using a novel method called cloud fields, (3) globally flexible irradiance estimation models, and (4) distributions of clear-sky irradiance by total cloud amount, to create the irradiance time series.

The SDSIG outputs are temporally validated using metrics of ramp rates, variability indices and irradiance magnitude frequencies against real world observations at two UK sites and two USA sites, representing three distinct climates. Daily 2-sample Kolmogorov-Smirnov tests of each metric passed a minimum of 95.34% of the time with a 99% confidence limit. The lowest CDF correlation coefficient between modelled and observed data for all metrics and sites was $R = 0.908$; the mean was $R = 0.987$. The SDSIG outputs are spatially validated at Oahu, HI USA, showing $R = 0.955$, RMSE=0.01 and MAPE=0.865% when comparing modelled and observed spatial correlation versus site separation. The SDSIG outputs are applied to a grid impacts power flow model of an LV grid with increasing PV penetration to test the over voltage metric of daily on-load tap changer (OLTC) operations. Using *correlating* irradiance time series at each house in the LV grid overestimates OLTC operations in every instance of PV penetration when compared to using spatially *decorrelating* irradiance time series from the SDSIG.

List of journal and conference papers

Journal papers

Bright, J. M., Babacan, O., Kleissl, J., Taylor, P. G., and Crook, R. 2017. A synthetic, spatially decorrelating solar irradiance generator and application to a LV grid model with high PV penetration. Accepted with corrections and resubmitted to Solar Energy.

Smith, C. J., Bright, J. M. and Crook, R. 2017. Cloud cover effect of clear-sky index distributions and differences between human and automatic cloud observations. Solar Energy 144, 10-21.

Bright, J. M., Taylor, P. G. and Crook, R. 2017. Methodology to stochastically generate synthetic 1-minute irradiance time series derived from mean hourly weather observational data. ISES Solar World Congress 2015 (SWC2015) Special Issue. Accepted for publication in Solar Energy (expected publication date 2 years after conference in October 2017).

Bright, J. M., Smith, C. J., Taylor, P. G. and Crook, R. 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. Solar Energy 115, 229-242.

Conference papers

Bright, J. M., Taylor, P. G. and Crook, R. Methodology to stochastically generate spatially relevant 1-minute resolution irradiance time series from mean hourly weather data. 5th Solar Integration Workshop. International Workshop on Integration of Solar Power into Power Systems. Brussels, Belgium. Oct 19–20, 2015. In proceedings, pages 180–186.

Bright, J. M., Taylor, P. G. and Crook, R. A spatio-temporal methodology to stochastically generate synthetic 1-min irradiance profiles derived from mean hourly weather observational data. SolarPACES. Concentrating Solar Power and Chemical Energy Systems. Cape Town, South Africa. Oct 13–16, 2015. In proceedings, pages 251–252.

Bright, J. M., Smith, C. J., Taylor, P. G. and Crook, R. Methodology to stochastically generate synthetic 10 minute irradiance profiles derived from mean hourly weather observational data. EU PVSEC, 31st European Photovoltaic Solar Energy Conference and Exhibition. Hamburg, Germany. Sept 14–18, 2015. 5BV.1.7.

Bright, J. M., Smith, C. J., Taylor, P. G. and Crook, R. Stochastic generation of minutely irradiance time series derived from hourly weather observation data. PVSAT. Photovoltaic Science Applications and Technology. Conference C97 of the Solar Energy Society. University of Leeds, Leeds, UK. Apr 15–17, 2015. In proceedings, pages 1–4.

Contents

Declaration of authorship	iv
Acknowledgements	v
Abstract	vi
List of journal and conference papers	vii
Contents	viii
List of figures	xiii
List of tables	xix
Abbreviations	xxi
Nomenclature	xxv
1 Introduction to solar resource modelling	1
1.1 Renewable energy and climate change	1
1.2 Solar irradiance data availability	3
1.3 Solar resource modelling	6
1.4 Overview of the research	7
1.4.1 The research aim	8
1.4.2 The research problem	8
1.4.3 Research objectives	9
1.4.4 Thesis structure	9
2 Grid impacts literature review	11
2.1 The electricity network	11
2.1.1 Classification of the electricity grid	12
2.1.2 The research problem posed by grid impacts	13
2.2 Perception and identification of grid impacts	13
2.3 The grid impacts from a distribution network operator’s perspective	15
2.3.1 Voltage fluctuation	15
2.3.2 Harmonic distortion	21
2.3.3 Unintentional islanding	24
2.3.4 Reverse flow	25

2.3.5	Hot-spots	26
2.4	Review of the response time of power systems	28
2.5	Future considerations for grid impact analysis	30
3	Solar Irradiance Generator (SIG) development	33
3.1	Key concepts and definitions	34
3.1.1	Types and definitions of solar radiation	34
3.1.2	Clear-sky and clearness indices	35
3.1.3	Markov chains	36
3.1.4	Meteorology and atmosphere	38
3.1.5	Solar intermittency, variability and fluctuations	41
3.2	Literature review of temporal solar irradiance generators	44
3.2.1	Summary of identified research opportunities	50
3.3	Methodology of the Solar Irradiance Generator	52
3.3.1	Input data requirements	54
3.3.1.1	Input observation data	55
3.3.1.2	User defined variables	56
3.3.2	Cloud sample production	58
3.3.2.1	Producing a one-dimensional-matrix distance series of cloud cover	61
3.3.2.2	Producing a one-dimensional-matrix time series of cloud cover	62
3.3.2.3	Random extraction of a cloud sample	62
3.3.3	The Markov chain process and Markov transition matrices	62
3.3.3.1	Mathematical background of Markov chains	63
3.3.3.2	Constructing the Markov transition matrices	64
3.3.3.3	Implementing a Markov chain	67
3.3.4	Stochastic generation of weather variable time series	68
3.3.4.1	Initialising the weather variable generator	69
3.3.4.2	The weather variable generator process	69
3.3.5	Sun-Earth geometry	71
3.3.5.1	Background to calculating solar angles	71
3.3.5.2	Blanco-Muriel PSA algorithm methodology	75
3.3.6	Generation of clear-sky indices	77
3.3.6.1	Relationship of clear-sky index to total cloud amount	77
3.3.6.2	Clear-sky index with a 1 minute resolution	79
3.3.6.3	Extended periods clear and overcast conditions	82
3.3.6.4	Introduction of cloud edge enhancement events	83
3.3.7	Irradiance calculations	84
3.3.7.1	Obtaining the theoretical clear sky irradiance	85
3.3.7.2	Translating irradiance onto an arbitrary plane	87
3.3.8	Computational demand	89
3.4	Validation of the Solar Irradiance Generator	89
3.4.1	Hourly validation for Leeds, UK	89
3.4.2	Minutely validation for Cambourne, UK	91
3.5	Discussion of results	97
3.5.1	Diurnally weighted okta Markov transition matrices	98

3.5.2	Pressure weighted okta Markov transition matrices	99
3.5.3	Seasonally weighted okta Markov transition matrices	101
3.5.4	Discussion of the synthetic irradiance	102
3.6	Chapter summary	105
3.6.1	Key points, applicability and potential development	105
4	Spatially Decorrelating Solar Irradiance Generator (SDSIG) development	109
4.1	Literature review of spatio-temporal solar irradiance generators	110
4.1.1	Summary of identified research opportunities	118
4.2	Methodology of the Spatially Decorrelating Solar Irradiance Generator	118
4.2.1	Data requirements	120
4.2.2	Cloud field production	121
4.2.2.1	Methods of synthesising a cloud field	121
4.2.2.2	Developed approach to synthesising cloud fields	125
4.2.3	Simulating cloud motion	129
4.2.3.1	Weather variable generation and cloud field selection	129
4.2.3.2	Cloud field motion across a spatial domain	130
4.2.4	Generation of clear-sky indices	134
4.2.4.1	Clear-sky index distributions by both okta and elevation angle	134
4.2.4.2	Implementation of clear-sky index distributions	137
4.2.4.3	Inclusion of cloud edge enhancement events	138
4.2.4.4	Extended periods of clear and overcast sky	142
4.2.5	Spatially Decorrelating Solar Irradiance Generator outputs and computational performance	142
4.2.6	Computational demand	143
4.3	Validation of the Spatially Decorrelating Solar Irradiance Generator	143
4.3.1	Temporal validations for the UK and USA	144
4.3.2	Spatial validation for Oahu, Hawaii	149
4.4	Discussion of the Spatially Decorrelating Solar Irradiance Generator	152
4.4.1	Illustration of the instantaneous spatial correlation	152
4.4.2	Cloud fields	158
4.4.3	Cloud motion	165
4.4.4	Clear-sky indices	167
4.4.4.1	Clear-sky indices fluctuations	167
4.4.4.2	Differences between climate regions	168
4.4.4.3	Cloud edge enhancement events	169
4.4.5	Simple case study demonstrating the use of the Spatially Decorrelating Solar Irradiance Generator	170
4.5	Application of the Spatially Decorrelating Solar Irradiance Generator in a grid impact assessment case study	172
4.5.1	Overview of the grid impact study	172
4.5.2	Test circuit information and data sources	173
4.5.3	On-line tap changer operations grid impact metric	175
4.5.4	Results of grid impact assessment	175
4.6	Chapter summary	176

5	Conclusions and future work	179
5.1	Research summary	179
5.2	Contributions to knowledge	182
5.3	Limitations and ideas for future work	185
5.3.1	General future enablers to advance the impact of theoretical solar resource modelling	185
5.3.2	Specific opportunities to develop the Spatially Decorrelating Solar Irradiance Generator	186
A	Solar Irradiance Generator scripts	191
A.1	Cloud sample production	191
A.2	User defined variables	194
A.3	Data import	195
A.4	Preliminary Markov chain framework	195
A.5	Markov chain creation	196
A.6	Preliminary weather variable generation framework	202
A.7	Stochastic generation of weather variable and sun obscured time series	204
A.8	Global horizontal clear-sky irradiance calculations	208
A.9	Generation of minutely clear-sky index time series	210
A.10	Calculation of tilted irradiance by component	214
B	Spatially Decorrelating Solar Irradiance Generator scripts	215
B.1	Cloud field production	215
B.2	User defined variables	219
B.3	Matrices rotational function	221
B.4	Data import	222
B.5	Cloud motion and clear-sky index implementation	222
B.6	Clear-sky modification and incident irradiance calculations	227
B.7	Clear-sky distributions by okta and solar elevation angle	233
	Bibliography	237

List of Figures

1.1	Locations that have 1 min solar radiation data as part of the Base-line Surface Radiation Network as maintained by the World Radiation Monitoring Centre (WRMC-BSRN, 2014). Red dots indicate an expired monitoring location, green dots indicate active stations.	4
1.2	Distribution of UK Met Office synoptic observation sites (MIDAS, 2015)	4
2.1	The end-user voltage for a year with increasing installation capacity sizes at each property connected to the LV network (Widen et al., 2010)	17
2.2	The relationship between the voltage-imbalanced factor (VUF) and PV power output (Wong et al., 2014)	19
2.3	Total harmonic distortion relative % and PV generation factor over a winter day in Croatia (Fekete et al., 2012)	22
2.4	Power response time of different power equipment (plot produced by Pai (2016)	29
3.1	Example of a three-state Markov chain with the associated transition probabilities from each of the three states.	37
3.2	The annual variation in extraterrestrial radiation (PVPerformance Modeling Collaborative, 2016)	39
3.3	Solar radiation arriving on the ground under cloudy skies, inclusive of the many types of scattering and reflection losses in the Earth’s atmosphere (Iqbal, 2012).	40
3.4	Global horizontal incident irradiance in Leeds, UK on 25/06/2015 between the hours of 10:00 and 14:00. The data was logged using a horizontally-mounted silicon photodiode (BPW20) in short-circuit current mode with a linear current response converted to voltage with a transconductance amplifier, logged at 10 Hz using a 10-bit DAC. The data has been averaged across timescales of, from top to bottom: 1 second, 10 seconds, 1 minute, 10 minutes and 1 hour.	43
3.5	Demonstration of the sun obscured methodology derived by Morf (1998). Top) the stochastic insolation function over time, where 1 and 0 indicate clear and clouded periods, respectively. Bottom) the corresponding irradiance output as guided by the stochastic insolation function.	51
3.6	Flow chart summarising processes within the SIG. Green parallelograms represent an input into the model, blue rectangles represent a process, and the yellow parallelogram at the bottom indicates the synthetic irradiance output times series.	54

3.7	An example output of the SIG for a typical day in June for Cambourne, UK, that visually demonstrates the fundamental working of the SIG. Plot a) shows the mean okta number stochastically derived at each hour. Plot b) shows the corresponding cloud cover time series in a minutely resolution, where black is indicative of cloud present and white indicating periods of clear sky. Plot c) shows the resultant tilted irradiance, although this simulation was performed on a plane horizontally aligned for Cambourne, UK.	55
3.8	Colour plot illustrating the MTM for the mean annual transition probability of okta number from the current hour (y-axis) to the next hour (x-axis). This MTM was made using data taken from the MIDAS (2015) data sets for Cambourne, UK.	66
3.9	Demonstration of the stochastic Markov chain process from the current hour, $M_{t-1} = 1$, to the next hour, $M_t = 3$. In the bottom left, the cumulative first order MTM for an arbitrary variable containing 3 states is shown, \mathbf{P}_C^1 . As the current hour is at state 1 ($M_{t-1} = 1$), the appropriate row of $\mathbf{P}_{C(S 1)}^1$ is selected. The stochastic process is looped through, as indicated by orange arrows until M_t is found to be at $s = 3$. This process is repeated for $t(1, 2, \dots, T)$. However, M_t becomes M_{t-1} at each time step, t	68
3.10	Examples of the stochastically generated time series of (top) the cloud amount C in tenths, (middle) the cloud height z_{ref} in m, and (bottom) the cloud speed u_{ref} in ms^{-1} . The resolution is 1-hour and spans for 15 full days modelled for 1st-15th January 2012	72
3.11	Simple diagram of a sky dome indicating measurements of the solar zenith angle, θ_z , measured between the local vertical and the direction of the solar beam; and the solar azimuth angle, A_z although denoted τ in this thesis, measured from the south meridian to the direction of the sun. All these angles are in relation to a point P, on the Earth's surface (image reproduced with permission from ITACA (2015)).	73
3.12	Top) bi-modal histogram of all hourly k_c observations taken from 63 UK weather stations between 2010-2013. Bottom) histograms of k_c separated by N , with analytical distributions fitted and shown in red. Plots produced by Dr Chris Smith.	80
3.13	Global map of the Linke turbidity, T_L , to 5 minutes of arc angle for January 2002 taken from SoDa (2016). $T_L = 1$ indicates a perfectly clear and dry atmosphere. T_L increases as the presence of atmospheric aerosol and water vapour increases.	87
3.14	Comparison between modelled (mod) and observed (obs) mean diurnal irradiance profiles separated by season and the annual mean diurnal irradiance profile. The means are made from 10-years of modelled output and 10-years of validation data for Leeds, UK.	90
3.15	Comparison scatter plot for each 10-year mean hour of modelled output data against the corresponding hour of observation data. The yearly mean is also included. An $x = y$ line is included for reference.	91
3.16	Comparison of the mean hourly irradiance frequency cumulative probability distribution profiles for 10 years of modelled and observational data from Leeds, UK.	92

3.17	Examples of daily global horizontal irradiance time series is categorised using the daily mean variability index (VI). In red is the global horizontal clear sky irradiance and in blue is the global horizontal irradiance. The image was produced by Stein et al. (2012).	93
3.18	Cumulative probability distribution plots of the (a) daily variability index, (b) the minutely irradiance ramp-up and ramp-down occurrences, and the (c) minutely global horizontal irradiance frequencies for both the modelled data and the observed data from Cambourne, UK.	95
3.19	Colour plot indicating the deviation of the diurnally subjective okta MTM minus the annual yearly mean okta MTM. White space indicates a $\leq \pm 0.01$ change in transition probability.	99
3.20	Colour plot indicating the difference between the above average pressure MTM (AA) minus the below-average pressure MTM (BA). White space indicates a $\leq \pm 0.01$ change in transition probability.	100
3.21	Colour plot indicating the deviation of the individual season MTMs from the yearly mean okta MTM. The colour indicates the difference in transition probability, whilst white space indicates a $\leq \pm 0.015$ change in transition probability.	101
3.22	Histogram of the mean hourly observed values of k_c (bar) compared to the values of k_c from the SIG (line).	103
3.23	The diurnal mean hourly irradiance root mean square error (Wm^{-2}), averaged across 10 simulations. The RMSE is displayed diurnally and separated by season, the yearly diurnal RMSE is also shown.	104
3.24	Three days of example irradiance profile outputs from the SIG (top), and from observational data (bottom). Note this is neither a predictive method nor a forecasting model and so the synthetic profiles are not intended to match or recreate observational data. The purpose of this figure is to demonstrate the SIG's output and to show real 1-minute resolution observational data irradiance profiles, the days 176 to 178 were selected purposely as they presented similar corresponding characteristics.	106
4.1	Spatial cross-correlation structure (cross-correlation as function of inter-station distance) of hourly irradiance data gained from an independent long term (10 years) data set of 6 sites in Germany from Beyer et al. (1991). The line represents a fit to these data using an exponential decay by Wiemken et al. (2001)	113
4.2	Irradiance fluctuation correlation coefficients between geographical dispersed solar measurement devices for two different days of 2nd and 7th October 1996 in north of Tokyo, Japan. The 7th October data is fitted with a linear trend line. Plot produced by Otani et al. (1997)	113
4.3	Flow chart summarising the SDSIG 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 SIG methodology to include a spatial dimension.	119
4.4	Example visualisations of cloud fields for coverage values $C = 1$ to 10. The X dimension is equal to $3600u_{ref}$ and the cloud speed of these cloud fields is $u_{ref} = 1 \text{ ms}^{-1}$. These plots are merely visualisations of what are in fact a $3 \times n_c$ 2DM of $x_c-y_x-r_c$	124

4.5	Example of the utilisation of equation 4.2. A is the cloud field domain defined by its four points. c_1 and c_2 combined create c_i and are shown in blue defined by their respective (x_c, y_c) centre points and radii r_c . The spatial resolution x_{res} is the distance between sample points. The colours of the sample points (orange) indicates a point that is within both A and c_i , (green) a point found only within A and (grey) any point outside of A . The resultant $C_{\text{frac}} = 0.367$ and $C = 4$ tenths	127
4.6	Cloud speed distribution for Tucson, AZ USA, produce by Augsburg and Favrat (2013)	130
4.7	Visual demonstration of the cloud motion and sun obscured methodologies The dashed box represents a cloud field full of clouds with centre points (x_c, y_c) and radii, r_c , that is travelling at speed u_{ref} across the spatial domain in solid blue containing the houses positioned at $((x_h, y_h))$ and indicated by crosses. The spatial domain has dimensions XY while the cloud field has dimensions X_cY . 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 field moves at each time step, t , in the direction indicated by the arrow marked u_{ref}	131
4.8	1) Histograms of the clear-sky index distributions by okta for all solar elevation angles. 2) Histograms of the clear-sky index distributions by solar elevation angle bands for all okta. Both plots were produced by and published in Smith et al. (2017)	136
4.9	A sample clear-sky irradiance time series (cyan) alongside observed global horizontal irradiance from WRMC-BSRN (2014) in Cambourne (red) and the percentage difference between the two (blue). The y axis is both irradiance (Wm^{-2}) and percentage difference (%).	139
4.10	The chance of a cloud edge enhancement event occurring at different (top) okta values and (bottom) zenith angles, all separated by varying cut-off margins.	140
4.11	The cumulative probability function of cloud edge enhancement fractional magnitude (top) separated by each okta value and (bottom) separated by bands of solar zenith angle.	140
4.12	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 calendar 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.	145
4.13	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.	146

4.14	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.	150
4.15	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).	151
4.16	Spatial correlation between centre point to all other locations 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.	153
4.17	Line plot of the correlation from the reference 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 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 correlation between the individual X and Y directions for both the control and prevailing scenarios.	154
4.18	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.	155
4.19	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 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.	156
4.20	Top left) histogram of the fractional difference between the actual coverage value experienced at a house in the centre of the spatial domain and the hourly coverage value of the whole sky. Top right) CDF of the histogram in top left. All below) histograms of the actual experienced coverage, C_h in tenths for each target C value indicated in bold above each plot.	161
4.21	Analysis of the $C = 5$ cloud field taken from figure 4.4. Two potential paths that a house may take through the cloud field are indicated. Clouded sections of the path are shaded yellow and clear sections are shaded red. The actual coverage value experienced on both paths are shown below the cloud field.	162

4.22	Convergence of the coverage value over time. Each line represents an hour of simulation where the experienced coverage develops over time, the solution is calculated as a moving average. The solutions are separated by the target coverage values for the whole cloud field. These plots complement the method shown in figure 4.21.	163
4.23	The estimated coverage value versus the target coverage value for increasing spatial grid resolution, res . A line along $x = y$ is provided for reference.	164
4.24	Histogram of cloud radii as drawn 10^8 times from the Wood and Field (2011) distributions and adjusted to apply limits. All radii above a limit are combined in the final bin.	165
4.25	Demonstration of how a house, when initially located at $(0, Y)$, can be positioned outside of the spatial domain when rotated by the cloud motion direction, θ . The method of determining the maximum distance outside of the domain, q , is detailed.	167
4.26	A theoretical set-up used to demonstrate the functionality of the SDSIG. The numbers indicate 5 different houses located in Cambourne, UK, for which spatio-temporally decorrelating irradiance time-series are generated.	170
4.27	Synthetic decorrelating irradiance profiles for 5 houses on the 10th Jan from SDSIG located in Cambourne, UK.	172
4.28	On-line diagram of the 8500-node feeder circuit used in the power flow simulations (IEEE, 2013).	174
4.29	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 represents the power flow scenario using the single correlated time series data across the distribution system and green represents the scenario using the spatially decorrelated and temporally unique synthetic GHI time series from the SDSIG. Blue 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. This plot was produced by Oytun Babacan.	177

List of Tables

3.1	Conversion of the total cloud amount, N , in okta (reported as eighths) to the cloud coverage fraction, C (reported as tenths). $N = 9$ represents sky that is obscured due to fog/haze/other meteorological phenomena (NOAA, 2016b)	41
3.2	Relationships between meteorological variables and radiation. Key: SD = sunshine duration, CF = cloud fraction, CT = cloud type, OCF = opaque cloud fraction, DTR = diurnal temperature range, PR = precipitation. Clearness index, kt, and clear-sky index, kc, were defined in section 3.1. Table produced by Dr C. J. Smith.	44
3.3	Indication of how the MTMs for each weather variable are separated accounting for seasonal, above and below pressure and diurnal differences.	64
3.4	Solar parameters calculated by different authors. This table was reproduced from Blanco-Muriel et al. (2001) and updated by the author. δ is the declination angle, ra is the right ascension, EoT is the equation of time or equivalent, θ_a is the azimuth angle, and θ_z is the zenith or elevation angle.	74
3.5	Distributions of k_c by N derived from histograms shown in figure 3.12. The normal distributions are parameterised by (μ, σ) where μ is the mean and σ is the standard deviation. The Weibull and Gamma ($\Gamma(a, b)$) distributions are parameterised by (a, b) where a is a shape parameter and b is a scale parameter.	79
3.6	The number of days (out of 365) that pass the 2-sample Kolmogorov-Smirnov test on 7 of the same calendar day from successive years of both modelled and observational CDFs containing the minutely variability indices, minutely ramp rate occurrences, and the minutely irradiance frequency, tested at increasing confidence limits.	97
4.1	Station metadata from the NREL Oahu Radiation Monitoring Station (Sengupta and Andreas, 2010) showing the height above sea level (z_l), latitude (Θ) and longitude (Φ) for each ID in figure 4.14.	149
4.2	Physical parameters as user defined inputs into the SDSIG for the 5 houses in the case study as shown in figure 4.26. The letter C denotes the centre of the spatial domain at (750,750).	171
B.1	k_c distribution parameters for each N at bands of θ_e	234
B.2	k_c distributions continued	235

Abbreviations

1DM	One-dimensional matrix
2DM	Two-dimensional matrix
AA	Above average [pressure]
AEMO	Australian Energy Market's Operator
BA	Below average [pressure]
BADC	British Atmospheric Data Centre
BEIS	Department of Business Energy, and Industrial Strategy
BSRN	Baseline Surface Radiation Network
CDF	Cumulative probability density function
CEE	Cloud edge enhancement
CSI	Clear sky irradiance
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DECC	Department of Energy and Climate Change
DG	Distributed generation
DNI	Direct normal/beam irradiance
DNO	Distribution network operator
DSTATCOM	Distribution static compensator
ECMWF	European Centre for Medium Range Weather Forecasts
EPRI	Electrical Power Research Institute
EV	Electric vehicles
FiT	Feed-in tariff
GHI	Global horizontal irradiance
GIS	Geographic information system
IEA	International Energy Agency
IF	Irradiance magnitude frequency [metric]

IPPC	Intergovernmental Panel on Climate Change
KCI	Clear-sky index [metric]
K-S	Kolmogorov-Smirnov [test]
LCOE	Levelised cost of energy
LiDAR	Light detection and ranging
LV	Low voltage [electricity grid]
MAPE	Mean absolute percentage error
MIDAS	Met Office integrated data archive system
MODIS	Moderate resolution imaging spectroradiometer
MTM	Markov transition matrix
MV	Medium voltage [electricity grid]
NOAA	National Oceanic and Atmospheric Administration
NREL	National Renewable Energy Laboratory
OLTC	On-load tap changer
PDF	Probability density function
PSA	Platforma Solar de Almeria
PV	Photovoltaic
PVP	Photovoltaic penetration
QCLCD	Quality controlled local climatological data
R	Correlation coefficient
RET	Renewable energy targets
RMSE	Root mean square error
ROC	Renewables obligation certificate
RR	Ramp rates [metric]
SDSIG	Spatially Decorrelating Solar Irradiance Generator
SIG	Solar Irradiance Generator
SPA	Solar position algorithm
STC	Small scale technology certificate
TcHD	Total current harmonic distortion
THD	Total harmonic distortion
TMY	Typical meteorological year
TvHD	Total voltage harmonic distortion
UCSD	University of California, San Diego

UKMO	United Kingdom Meteorological Office
UTC	Coordinated universal time
V2G	Vehicle-to-grid
VI	Variability index [metric]
WRMC	World Radiation Monitoring Center

For Brian.

Chapter 1

Introduction to solar resource modelling

1.1 Renewable energy and climate change

Renewable energy technologies and global climate change are two topics that intertwine and must be discussed synonymously. Climate change mitigation is one of the primary drivers to the development of renewable energy alternatives, the other being fuel and energy crises ([Ellis, 2014](#)) that have seen uncertain global fossil fuel costs and augmentation of political instability in securing combustion fossil fuels ([de Vos and Sawin, 2012](#)).

The Earth's climate has undergone a rapid and unprecedented global warming trend in the last few decades ([Pittock, 2013](#)); an abundance of scientific publications agree that this climate irregularity is a direct consequence of the anthropogenic release of greenhouse gasses ([Houghton, 1996](#); [USA DoE, 2000](#); [O'Hare et al., 2005](#); [Team, 2008](#)). Of the various significant contributors to global greenhouse gas emissions, fossil fuelled energy production is the principal contributor ([Höök and Tang, 2013](#)). Unfortunately for the mitigation of climate change, the global economy and particularly the economies of industrialised countries are largely dependent on fossil fuels for energy generation ([Jacquard, 2006](#)). Such a dependency exacerbates the problem of greenhouse gas induced climate change. Whilst the impacts related to global climate change are difficult to predict, the Intergovernmental Panel on Climate Change (IPCC) reports a wide range of potentially devastating impacts: rising sea levels leading to coastal flooding, severe disease transmission, an increased occurrence of extreme weather events, loss of ecosystems, among others ([IPCC, 2014](#)). For the mitigation of climate change, it is clear that

the immediate transition away from fossil-fuel-dependent energy generation towards a high efficiency and renewable based system is imperative (Jaccard, 2006).

Growth in the energy industry is particularly prevalent as renewable energy technologies can meet much of the growth in energy demand if given enough financial support (Johansson and Burnham, 1993; Solangi et al., 2011; Fokaides and Kyliyi, 2014). There are numerous types of renewable technologies accessible from a diverse range of resources, such as solar, biomass, wind, wave and tidal, all of them effectively inexhaustible (Tiwari and Mishra, 2011). It is due to this variety that new barriers and issues are presenting themselves, although there are clear net benefits such as emission reduction, cost savings, well balanced economic structure, energy security, and remote electricity access (Jenkins, 2013). There are also indirect benefits, such as improved air quality. Globally, around 6.5 million premature deaths can be attributed to air pollution each year, of which energy production is a significant man-made air quality reduction culprit. This is a larger number of deaths than the combined total attributed to HIV/AIDS, tuberculosis and road injuries (IEA, 2016).

PV technologies have shown remarkable progress in life cycle environmental performances. All PV technologies generate far less life-cycle air emissions per GWh than conventional fossil fuel electricity generation technologies (Fthenakis et al., 2008). The efficiency of solar cells is expected to increase with further research and with a higher emphasis on the use of recycled materials. These changes will bring further environmental benefits, making PV a highly desirable electricity generation technology (Sherwani et al., 2010).

Globally, there has been a rapid increase in solar PV uptake. The International Energy Agency (IEA) recorded that solar PV was the fastest-growing renewable power technology worldwide from 2000–2011 (IEA, 2014), with Germany and Italy accounting for over half the installed capacity at the time, followed by Japan, Spain, USA and China. By 2016, China held the leading market share. The installed global capacity has shown rapid PV deployment annually, with 50 GW installed in 2015 alone, predominantly from China, Japan and USA (REN, 2016). In the UK, there was an 87% growth in generation from 2014 to 2015 from 4,040 GWh to 7,561 GWh, with an installed capacity of 8.915 GWp (start of 2016); this made the UK the 4th largest installer in 2015 (DECC, 2016).

With such rapid growth and projected increases in PV deployment, issues arise when integrating solar PV into the electricity distribution grid. These issues depend on the degree of variability in the power generation and on the flexibility of the power system (Widen, 2015). Overcoming the effects that resource variability has on PV across every temporal scale is pivotal for the uninterrupted and most economic development and

penetration of PV technologies into the grid (Perez and Fthenakis, 2015) so that solar renewable energy can contribute to the reduction of anthropogenic green house gasses from the energy sector.

In order to address the potential issues that PV presents to the electricity grid, it is important to be able to theoretically model and explore them. Copious attempts have been made to assess and quantify the impacts that face the electricity grid, these are well explored and discussed in chapter 2. The complexities of modelling are discussed in chapters 3 and 4. Fundamentally, however, modelling the solar resource comes down to a simple factor — data availability.

1.2 Solar irradiance data availability

Access to data is the most significant component to understanding the solar resource and facilitating PV variability and electricity grid interaction studies. The ideal solar irradiance data availability would be a highly geographically dispersed and well maintained solar observation database.

There are two main types of solar irradiance data: ground based observations and satellite derived irradiances. Satellite derived solar irradiance temporal and spatial resolutions are poor. The most state-of-the-art satellites can provide 1 km² at nadir, however this is limited to one field of view centred over Asia and Australia and the spatial resolution decreases with distance from the equator. The highest temporal resolution available from satellite imagery with no model assumptions is 10-min from the same Himawari-8 satellite from the Japanese Meteorological Agency. As will be demonstrated later, 10-min is an inadequate temporal resolution to capture rapid changes in solar resource availability. The second type of solar irradiance data is ground based observations. These are recorded at very high temporal resolutions, from hundredths of a second to 1-min, and are much more suited to capturing solar radiation fluctuations. The issue, however, is that they are not well geographically dispersed. In order to observe irradiance fluctuations, ground based irradiance monitoring is the only real option and so the inherent disadvantages are discussed.

Using real radiation observation data places too much reliance on data availability (Fernández-Peruchena and Gastón, 2016); these datasets are often plagued with gaps, inaccurate time-stamps or inconsistent measurement techniques (Kumar et al., 2013). Whilst methods exist to fill these gaps so as not to over or underestimate the realistic variability of irradiance (Moreno-Tejera et al., 2016; Polo et al., 2011; Larraneta et al., 2015), datasets are geographically sparse. The majority of solar irradiance datasets are

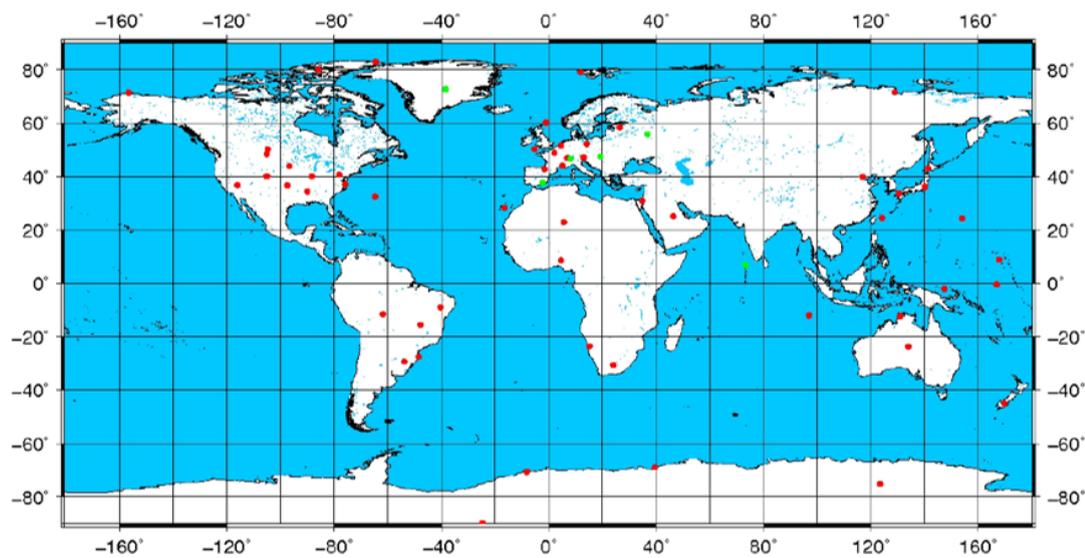


FIGURE 1.1: Locations that have 1 min solar radiation data as part of the Baseline Surface Radiation Network as maintained by the World Radiation Monitoring Centre (WRMC-BSRN, 2014). Red dots indicate an expired monitoring location, green dots indicate active stations.



FIGURE 1.2: Distribution of UK Met Office synoptic observation sites (MIDAS, 2015)

from model derived solar resource assessments. Satellite images are frequently used to refine results from ground based observations, however, it is common for datasets to be purely a construct of model derived values (Vignola et al., 2013).

The World Radiation Monitoring Centre (WRMC) maintain the central archive of the Baseline Surface Radiation Network (BSRN) that offers high resolution, readily available and quality assured datasets (WRMC-BSRN, 2014). 1-min resolution irradiance data is available for many locations around the world and are shown in figure 1.1. This is an example of an excellent source of irradiance data. The problem with them, however, is geographic density. There are only two existing sites in the UK, Cambourne and Lerwick (station number 50 and 51 respectively). The intention of the BSRN is to represent the climatic regions, so whilst it is a useful source of irradiance data, the variation of irradiance across the UK cannot be represented by these two time series. Cambourne is located in the south east and Lerwick on the islands off the northern coast of Scotland. They represents the two extremes of the country. Furthermore, weather pattern distribution is not uniform for the whole country. Total cloud amount varies vastly with both longitude and latitude (Smith et al., 2017) implying that these BSRN datasets cannot be used for other regions around the country.

Other high resolution data sets exist throughout the UK, although as the measurement technology is not consistent, nor are the maintenance practices and format, comparing multiple irradiance datasets can lead to compounded errors. Furthermore, they are usually the result of private endeavours by research facilities and industry. The spatial density of meteorological stations equipped for high resolution solar radiation monitoring is far less than required (Calinoiu et al., 2014). Where the geographic distribution is acceptable (figure 1.2), the time resolution is often lower than is desirable, such as the radiation observations from UK Met Office that are a 1-hour resolution as standard (MIDAS, 2015) among other meteorological offices around the world. This limited data availability provides an opportunity for synthetic irradiance time series modelling that can utilise the good geographic distribution of standard meteorology measurements.

Vignola et al. (2013) state that irradiance measurement devices such as pyranometers, pyrhemometers and cavity radiometers, have accuracies from ± 3 to 7% irradiance magnitudes at 95% confidence. However, all suffer from errors due to maintenance, calibration and spectral response degradation. Daily maintenance is a requirement to keep this equipment in quality assured operation. Therefore, costs are high and so access to high quality data is possible, yet uncommon. As it stands, datasets are never ideal and it is the view of the financing industry that the material miscalculation of

the solar resource is one of the biggest risks of a solar project (Vignola et al., 2013); reliable irradiance data are vital.

As will be demonstrated throughout this thesis, the access to high spatial and temporal resolution irradiance data can vastly improve the understanding of spatial correlation and magnitudes of solar fluctuations, which is useful for a variety of reasons as is discussed in section 1.4. Ground observation data at such a high spatial and temporal resolution is extremely scarce. The only known publicly available dataset is used by Hinkelman (2013) and is maintained by the National Renewable Energy Laboratory (NREL) (Sengupta and Andreas, 2010). The issue is that this dataset is only relevant to the island of Oahu, Hawaii USA. The only other way to use this type of data is through solar resource modelling.

1.3 Solar resource modelling

Solar modelling, solar resource modelling, irradiance modelling and similar terms are used to describe mathematical methods of determining the power availability from the sun over a desired time period, in both the past and future.

Solar resource modelling is needed for a multitude of reasons such as forecasting the power availability for PV farms to compete in generation markets, suitability assessments of a location for a PV farm in the planning phase, estimating the lighting demand of a property, in agriculture when estimating crop yield, in public health care research e.g. UV exposure (Calinoiu et al., 2014), and importantly in estimating the grid impacts from increasing penetrations of PV into the electricity network.

There are three distinct methods of solar resource modelling. The first approach is forecasting where the future irradiance is estimated using the current conditions as framework. This is typically achieved using satellites as frame works over time windows of 10 minutes up to typically 7 days (Reikard, 2009; Chow et al., 2011; Inman et al., 2013; Yang et al., 2014). The second approach is to produce historic, bankable irradiance datasets (Vignola et al., 2013). Historic irradiance modelling, or past-casting, is a method to determine actual irradiance values that occurred in the past from the previous minute to as far back as desired. historic measurements. The third type of solar resource modelling is synthetic irradiance modelling, this term covers any attempt to synthetically generate irradiance time series. The purpose of this field is to generate statistically accurate time-series for use when real datasets are unavailable, to fill in gaps in larger datasets, or to increase the temporal resolution of a dataset. Synthetic modelling is of interest in this thesis and will be discussed in the following section.

The key benefits of deriving synthetic irradiance time series is similar to that of historic irradiance modelling: uninterrupted time series data for the entire duration of the synthetic generation, potential for geographic flexibility, time savings offered by modelling data as opposed to waiting to recording new data, removal of inconsistent measurement technique drawbacks such as spectral response issues, fouling and maintenance; cost reduction because measuring or purchasing irradiance data can be expensive, and potential removal of the data availability dependency.

1.4 Overview of the research

This section discusses the overall research aim, which is then discussed in its associated research problems and the objectives required to meet them. This section closes with a discussion of the structure of this thesis.

This work strives to offer a real option for generating realistic solar irradiance data for use in solar related studies. To date, the only real options available to one requiring an irradiance input is to sacrifice either spatial resolution or temporal resolution. The methodology presented in this thesis presents a real alternative that can achieve both very high spatial resolution and sufficient temporal resolution for a multitude of uses.

This type of work synthetic irradiance of particular interest to solar PV systems engineers and power flow engineers. Firstly, for PV systems modelling, the temporal resolution of study is typically a direct consequence of the input solar data resolution, should the model wish to consider solar variability. Therefore, the study is limited to locations from known data sets as discussed in the previous section. This thesis presents a downscaling methodology whereby readily available low resolution inputs can be manipulated to offer high resolution solar irradiance outputs. For them, they can then operate inverter models, storage control strategies etc. at a high temporal resolution. For power flow engineers considering the electricity distribution network, to date, there are very few options to obtain individual, correlating irradiance profiles for a multitude of properties on an electrical distribution grid with which to analyse different electrical impacts from solar PV. As will be discussed, the most common solution to this problem is to use a single time series of irradiance and apply it to all properties on a grid indicating perfectly synchronised ramping events, and therefore will overestimate the impact of a ramp event. With the methodology presented in this thesis, it will be possible for distribution grid modelling to have a high spatial and temporal resolution data input for their simulations facilitating more accurate and insightful results.

The major benefits of the presented methodology to others is the capability of individually assigning irradiance time series to as many locations within a spatial domain as required, and they will all correlate appropriately with each other. This is the first of its kind to offer such capability with the only real alternative solution is to produce an aggregated or up-scaled power output across the whole spatial domain. Furthermore, this methodology is produced from readily available mean hourly data and captures the ramp, magnitude and variability statistics of actual solar irradiance.

The remainder of this section will outline the steps taken to achieve the intended research.

1.4.1 The research aim

The aim of this research is to develop a novel modelling methodology that produces synthetic irradiance time series that vary on both a temporal and spatial dimension suitable for application in multivariate grid impact analysis, and that are also derived from readily available hourly observation data.

1.4.2 The research problem

The research aim is fundamentally derived from a need for high resolution solar irradiance data that is well geographically dispersed that facilitates multivariate grid impact analysis. To appropriately address this need, a thorough understanding of grid impacts is required. This leads to the first research problem:

1) What are the key distribution network impacts associated with increasing penetration of intermittent solar PV technology into the grid, the identification of which will help guide the requirement criterion for synthetic irradiance time series?

Upon identifying a target grid impact and the nature of studying it, the type and need for synthetic irradiance will be identified. This leads to the technical research problem:

2) Can synthetic irradiance time series be generated with significant statistical accuracy using readily available, well geographically dispersed, mean hourly meteorological observations as an input, facilitating access to more appropriate data for temporal grid and solar systems study?

Upon conceptually proving that synthetic irradiance is producible, it is imperative to introduce a spatial dimension for larger scale assessments of grid impacts such as those identified for problem 1. This leads to the final research problem:

3) *Is it possible to produce statistically accurate, synthetic irradiance time series that vary on both a temporal and spatial dimension, facilitating multi-variate grid impact analysis?*

1.4.3 Research objectives

In order to address the research problems and overall aim, the following research objectives are devised and are defined as follows:

1. Establish the impacts to the distribution network associated with the distributed generation of solar PVs. Identify the critical impact, if any, and the methods of quantifying it.
2. Develop a synthetic and stochastic methodology that produces a temporally relevant irradiance time series from readily available mean hourly weather observations. Complete with validation against real 1-min irradiance time series observations.
3. Further development of the methodology to incorporate a spatial dimension, facilitating the production of multiple, spatially decorrelating irradiance time series that validate against equivalent observation data.
4. Demonstrate the applicability of the spatially decorrelating and temporally accurate irradiance time series in application to a grid impacts methodology and explore one of the perceived impacts.

1.4.4 Thesis structure

This thesis is in 5 chapters. Each chapter will be discussed in turn and detail how it helps address the research problem and where the objectives will be carried out.

1) *Introduction and background* — this chapter brings the reader from the widest context of solar energy and climate change and directs them through the overarching concepts and workings of solar modelling, before presenting the research aim, problems and objectives.

2) *Grid impacts literature review* — an in depth review to the different grid impacts and the current research attempts to analyse them. The lessons learned from the studies are presented as research opportunities that facilitate the fundamental inclusions for the development of a synthetic solar irradiance generator model.

3) *Solar Irradiance Generator (SIG) development* — an in depth review of the current state of solar irradiance generation and lessons learned from literature. Development and validation of a temporal solar irradiance generation methodology.

4) *Spatially Decorrelating Solar Irradiance Generator (SDSIG) development* — further advancement of the SIG to include a spatial dimension through study of alternative, or lack there of, methods and literature. Development and validation of the spatial methodology before application in a grid impact assessment.

5) *Conclusion and future work* — lastly, the work will be summarised and the research aim revisited. Future research opportunities are suggested and discussed with select ideas outlined in detail.

Chapter 2

Grid impacts literature review

The aim of this chapter is to identify the key distribution network impacts associated with increasing penetration of solar PV into the electricity grid. This is achieved through ascertaining the current understanding of potential grid impacts by exploring each in isolation.

This chapter is separated into three sections. The first section discusses the different perceptions and identified grid impacts. The second section is separated into each of the grid impacts in turn. They are introduced, their importance explained and discussed using state of the art literature that examines each one. Finally, the findings and identified research opportunities are summarised.

2.1 The electricity network

Built, developed and redeveloped over many years, electricity networks are unique by country, and often unique within parts of a country ([Berry et al., 2013](#)). This means that grids are incredibly diverse and so there is no consensus example of an electricity network. Technological advances and deployment are tailored to the geographic region where the construction occurs, be it a long feeder traversing swathes of countryside to reach a village, or a short but complex inner-city district feeder that provides electricity for many times the population of a rural feeder. It is this diversity in society that results in electricity grids of incredible density with short average distances between nodes, or quite the opposite in sparse, rural settings ([Schneider et al., 2008](#)). By the very nature of these distinctive grids, the challenge of grid integration of PV becomes equally as variable. Analysis is required on a case by case basis, or at least using a classification system.

The electricity network in the UK, and similarly around the world, comprises three intuitively named sections: generation, transmission and distribution. They are defined by [Richardson \(2016\)](#) as follows.

1. **Generation:** The generation section of the grid comprises all the large scale power stations of the whole country. Electricity generated is fed into the transmission network at entry points across the country so that it can be delivered to consumers.
2. **Transmission:** The transmission network is a series of cables (96% overhead, 4% underground in the UK) that traverse the country linking together the generation networks to distribution networks. The cables in England and Wales typically operate at 400 kV and 275 kV, respectively. Scotland operates at a lower transmission; down to 132 kV.
3. **Distribution:** The distribution network covers all sections of the grid from exit points on the transmission network to the home, typically from 33 kV for industries, and 11 kV for towns and lighter industry, and finally down to 230 V at domestic usage. DG is connected to the distribution network, of which the LV 11 kV grid is considered in this thesis.

2.1.1 Classification of the electricity grid

In order to carry out research on the electricity grid, it is important to be able to physically define what the grid is. Classification of the electricity grid within a country is infrequently performed. However, it was a priority for national bodies in both the USA and Australia. Both governments sanctioned a national feeder taxonomy, which is a study of electricity network schematics that leads to the classification of the voltage power line systems found within the electricity grid. These taxonomies considered the electrical load upon the feeder, physical properties of the cabling, voltage characteristics and network topology to ascertain distinct classifications across the whole country ([Berry et al., 2013](#)).

The Australian taxonomy identified 15 classifications while the USA identified 24 ([Schneider et al., 2008](#)). The different States of the USA were found to have individualised classifications for urban and rural feeders. They were summarised by the density, industrial weight within the system, and whether a feeder location is urban, suburban or rural. The most important lesson to take from these taxonomies is that there is no simple approach or consensus that could cover all types of electricity networks with a single classification. A taxonomy style approach allows for general testing

of grid integration by type and by country. As most countries have not undertaken a national feeder taxonomy, grid impact analysis would have to be considered on a case by case basis as opposed to testing by classification.

[Eichman et al. \(2013\)](#) state that there is a maximum achievable renewable energy penetration limit that satisfies both electricity demand and system reliability; the topology also plays a decisive role ([Papaioannou and Purvins, 2014](#); [Nguyen et al., 2016](#)). This limit will vary depending on the feeder classification studied ([Berry et al., 2013](#)).

2.1.2 The research problem posed by grid impacts

With widespread implementation of PV technologies across the world, there will be considerable transformation in residential electricity demand. High uptake scenarios of residential PV can have detrimental impacts on the electricity distribution network due to newly installed DG. Whilst load is predictable and well-modelled, PV output is intermittent and its inclusion within impact analysis is far less prevalent in literature. The issue of PV intermittency is a substantial unknown, such that some DNOs have halted the deployment of further intermittent renewable technologies until further information is available ([Engerer and Mills, 2014](#)). It is possible that limiting PV penetration is unnecessary on account of overly conservative estimates as derived from the current understanding of PV-DG impacts. This is clear identification of an area for urgent research to prevent the slowing down of currently rapid renewable generation technology uptake. It is important to identify and quantify what are the key distribution network impacts associated with increasing penetration of intermittent solar PV technology into the grid.

2.2 Perception and identification of grid impacts

To assess the grid impacts, it is important to understand who is affected. The consumer typically has very little consideration or understanding of the complexities caused by the adoption of PV-DG within the low voltage (LV) electricity distribution network; they are more financially driven ([Mathiesen et al., 2011](#)). It is assumed that consumers adopt these policies with only monetary and sociological factors in mind. Some countries, such as Germany, have challenged this passive mindset and introduced new regulations to ensure that LV-DG customers provide ancillary services. It is already mandatory in Italy for installations above 6kW ([Caldon et al., 2014](#)).

It is the author's belief that the most important perception is that of the DNOs. The reasoning is threefold. Firstly, DNOs hold the power to place hard limits on PV uptake; this is demonstrated in Australia with Ergon Energy placing a 3.5 kWp limit per system (Parkinson, 2015), down from a previous limit of 5kWp. Secondly, the greatest economic impact is to the DNO as it is their infrastructure that must cope with electrical impacts from PV-DG. Policy makers are also considered to be of high importance. However, it is assumed that they are largely advised by the DNOs. Thirdly, the research was found to be heavily in favour of the DNOs perspective.

The DNO's concern for PV-DG likely stems from experience with wind generation, where a known issue requiring mitigation is "ramping", which is defined as the significant increase or decrease in electricity generation over a small time frame (Hoff and Perez, 2010). Almost all justification is targeted directly at the DNO in economic or technological feasibility (Wagner et al., 2015). The introduction of load changing technologies to the LV distribution network causes a variety of impacts. The aim of this section is to address them from the available literature.

Without new methods of coping with increased PV-DG, and because DNOs are not able to dictate the type, location or size of PV installations (Walbank and Goodhand, 2014), the management method from DNOs is to reactively replace and upgrade equipment such as transformers. In Germany, an additional 380,000km accumulated length of new cabling (a 24% increase in length from 2012 levels) is expected by 2020 at a projected cost of 20bn Euros; all of which are in areas where the LV-DG exceeds the local demand (Nykamp et al., 2012). Favourable incentives are required to allow for greater uptake in more controllable environments (Goli and Shireen, 2014). Zhao et al. (2010) summarised the range of issues caused by PV-DG, and Nykamp et al. (2012) split them into power control considerations and asset stresses. Both can be managed through demand and generation balancing or asset reinforcement (Mokhtari et al., 2014), although these strategies are often inconvenient and difficult to implement and predict. A widely adapted approach to conservative grid impact mitigation is PV curtailment during periods where the generation is higher than is manageable. This is detrimental to the overall *green* effectiveness of the system as well as reducing financial return (Wong et al., 2014). Poor mitigation management can result in negligible benefits from LV-DG (Borges, 2012). It is evident that DNOs are currently forced to adapt reactively to increased penetration of PV-DG.

The grid impacts from PV-DG are at their worst during periods of distinctly large imbalances of generation and load. Balancing the two is essential to minimising grid impacts (Nykamp et al., 2012). The provision of ancillary services is a response taken for mitigating grid impacts from PV-DG in both Germany and Italy. Ancillary services

are defined as the necessary services to maintain electrical power transmission within certain limits. They consist of: reactive power and voltage control, loss compensation, system protection, scheduling and dispatch, load following, and energy imbalance (FERC, 2013). Technologies, such as inverters, are required to modulate the reactive power exchanged with the grid and this continues to fuel innovation within inverter controls (Caldon et al., 2014). The interconnection equipment of PV-DG is the most significant technical requirement affecting the development of PV-DG projects (Caldon et al., 2014).

The issues of PV-DG from literature are summarised well by Passey et al. (2011) and Du et al. (2013) as:

- Voltage fluctuations
- Total harmonic distortion
- Unintentional islanding
- Reverse voltage flow

A further impact was identified through discussion with the DNO that operates locally to the University of Leeds called Northern Powergrid. Walbank and Goodhand (2014) stated that they would wish to identify the particular areas most likely to be affected by PV-DG. This impact has been called:

- Hot-spots

This following section is separated into the above bullet-pointed impacts.

2.3 The grid impacts from a distribution network operator's perspective

This section is separated into the six identified grid impacts. Each impact is introduced and described before a critical review of the current research is made.

2.3.1 Voltage fluctuation

Voltage fluctuations are defined as deviations away from the nominal line voltage. They are caused by both load and generation variation (Nguyen et al., 2016), such as PV

ramping caused by solar intermittency. These voltage fluctuations can cause a wide variety of impacts including premature ageing of infrastructure, equipment failure or overheating (ElNozahy et al., 2016), and the triggering of automated line equipment installed on the distribution feeders such as on load tap changers (OLTC) (Yan et al., 2014; Nguyen et al., 2016). OLTCs will be detailed in section 2.3.4. Voltage fluctuations are normal and occur at every node in the LV grid. However, their frequency can be increased to unsustainable levels with the introduction of new load and generation (Deilami et al., 2011; ElNozahy and Salama, 2014a).

DNOs face the difficult challenge of maintaining power quality to every customer; they are legally obliged to do so. Typically, voltage deviations are permissible at plus or minus 10% of the nominal voltage worldwide. For all EU member states, the EN50160 legal standard exists to maintain the delivered voltage to within 10% of its nominal value for 95% of the time; some EU countries apply stricter limits (Zhao et al., 2010; Widen et al., 2010; Deilami et al., 2011; Nykamp et al., 2012; Fekete et al., 2012). Australian administers stricter limits of maintaining nominal voltage within limits of +10 to -6% for 98% of the time (Vallee et al., 2013). Adhering to these obligations is an extremely high priority for DNOs as considerable fines are issued if they are not met (Thomson, 2000; Walbank and Goodhand, 2014). This economic concern must be minimised before the DNOs will allow for a greater roll-out of PV-DG. Therefore, transformer and feeder overload from voltage fluctuations must be understood and avoided (Clement-Nyns et al., 2010).

Research into PV induced voltage fluctuations is extensive. The most recent and extensive review of methods of voltage fluctuation control is by Shivashankar et al. (2016). They conclude that PV induced voltage fluctuation can be effectively minimised through the use of storage technologies and so increasing PV penetration should not be limited. Whilst this conclusion is accurate, it assumes the adoption of storage technologies. They are becoming more popular, particularly with the advancement of mainstream commercial options such as the Tesla Powerwall (Tesla, 2016). It is also statistically likely that PV adopters will adopt other renewable based technology (Balta-Ozkan et al., 2015). However, there is no guarantee to the DNO that residences with PV will also adopt storage technology. A hypothesis is proposed that understanding the baseline impact of voltage fluctuations is critical in order to appropriately assess its extent. A baseline scenario is one that does not consider the impact of mitigation technologies that are not ubiquitous in the industry. This does not mean that research concerning technological solutions is unnecessary; the development of interconnection equipment and inverter developments are highly important (Hernandez et al., 2012) and ultimately contribute to the wider story.

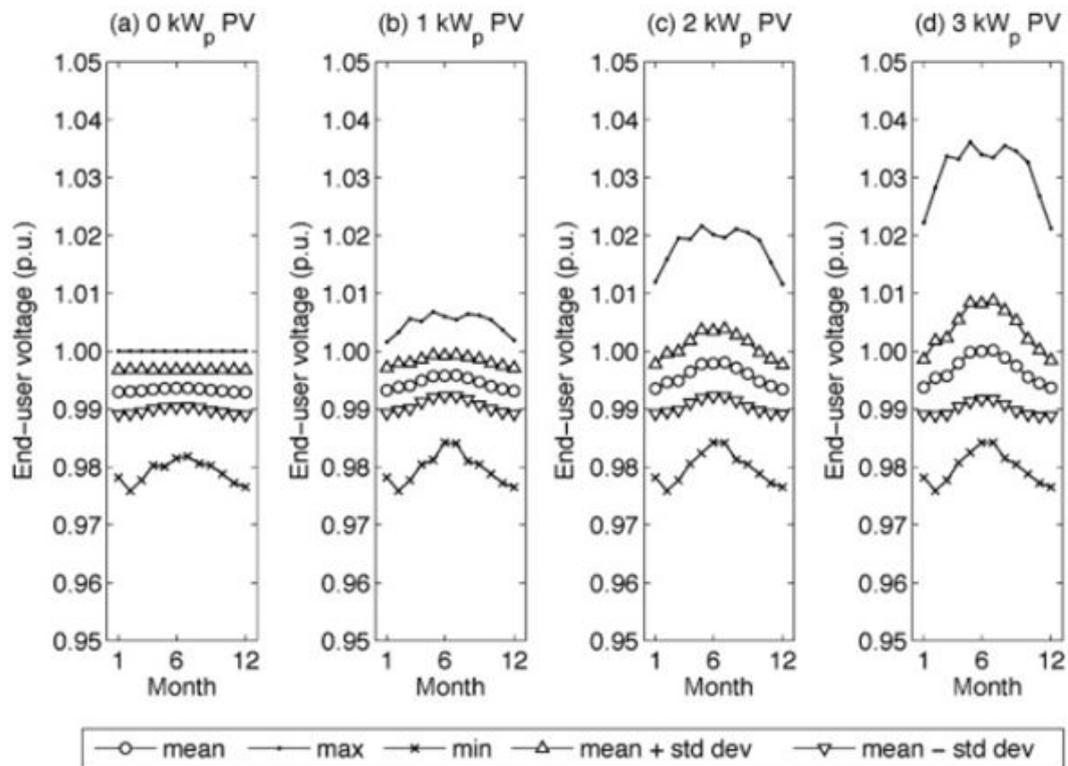


FIGURE 2.1: The end-user voltage for a year with increasing installation capacity sizes at each property connected to the LV network (Widen et al., 2010)

Some methodologies look at the maximum installable capacity of PV into a section of the grid. Paatero and Lund (2007) present a non-technological management strategy whereby planning a range of panel orientations can offset the generational peak output and can extend PV power output duration. They state that there is reportedly no issue with every property installing a 1kWp system (considered at two latitudes of 60N and 39N), however, a 2 kWp on every property leads to multiple over-voltage situations. A 1 kWp level is re-confirmed in a later study by Widen et al. (2010), this is because electricity generated from installations with a capacity greater than 1 kWp is generally exported, as 1 kWp tends to be within household demand. It was also found that the smaller the LV network, the greater potential of the installed capacity per household. This is demonstrated in figure 2.1, which shows how increasing the installed capacity in the network affects the end-user. However, this methodology is required on a higher resolution to appropriately capture intermittency impacts. It and many others like it suffer from using mean hourly input irradiance data.

Voltage impacts caused by solar intermittency are considered in a more recent methodology by Wong et al. (2014) that examined the voltage on a grid connected 7.2 kWp PV system upon a radial LV feeder in Malaysia. Solar intermittency is a particularly important issue in Malaysia as it receives next-to-no clear sky days, defined as daily

mean N of 0 okta, due to its climate and geography. The Malaysian government aims to be one of the largest producers of solar power in the world and so has a great interest in understanding grid impacts associated with PV-DG. Figure 2.2 demonstrates the output from the study. It demonstrates the direct response of voltage imbalance with the power output. It was also shown that short term voltage flicker, found by integrating the voltage fluctuations over a 10 min period, was always outside statutory limitation. Long term voltage flicker (as short term flicker, except integrated over 2 hours) regularly breached the statutory limit. Perhaps the most advantageous part of the study by Wong et al. (2014) is that they determined these values empirically on a physical PV installation without the deployment of additional mitigation technology. They established the baseline impact from a singular large PV system. It was found that for an output greater than 4 kW, the voltage at point of connection is highly likely to violate the Malaysian statutory limit of 252 V from a nominal value of 241–242 V. The study does not offer insight into more typically sized installations, between 1–5 kWp, or for multiple installations on the same section of grid. It is, however, important work that uses high resolution data and begins to address the baseline impact.

The current response to prevent voltage fluctuations in many countries is PV curtailment. This includes Malaysia (Salim et al., 2015) and would be applied to the instances of high voltage fluctuations shown in figure 2.2. Solutions under development to get around voltage fluctuation issues include employment of super-capacitors, static synchronous compensators with integrated storage, or OLTCs with installed reactive power compensators (Passey et al., 2011). Australian Standard AS4777.2 requires that inverters operate at close to unity power factor (i.e. inject only real power into the grid) unless they have been specifically approved by electricity utilities to control power factor or voltage at the point of connection (Passey et al., 2011; Salim et al., 2015). Specific technological solutions to address voltage fluctuation issues include the use of reactive power modulation. Cowley and Ekwue (2015) found that using the power factor led to voltage stability when reactive power was used in support of PV systems. This resulted in benefits for the consumer as more real power is output from the system.

Using real power to control voltage is more effective than using reactive power as the distribution network has a greater resistance than it does reactance. Methodologies for modelling proposed technological solutions are detailed by Paatero and Lund (2007) and Yan et al. (2014). Custom power devices are modelled to explore their effectiveness of controlling voltage fluctuation. Dynamic voltage restorers and distribution static compensators (DSTATCOM) were proven to compensate reactive power, mitigate harmonics and reduce voltage fluctuations. The DSTATCOM was shown to be the better methodology and would be positioned at 2/3rd the way from the transformer to the

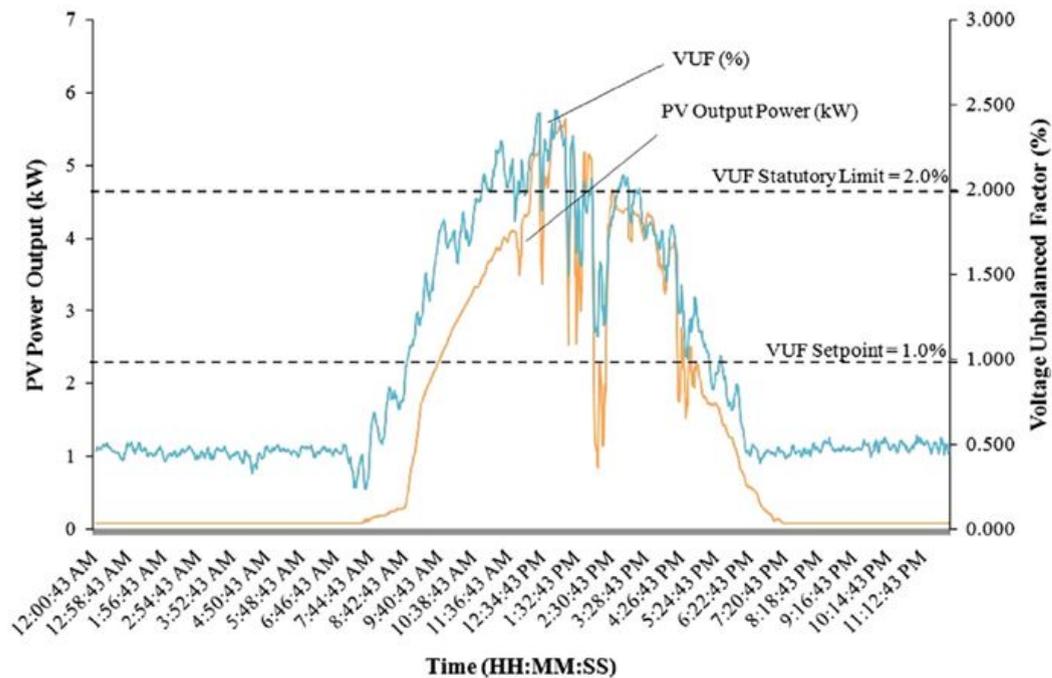


FIGURE 2.2: The relationship between the voltage-imbalanced factor (VUF) and PV power output (Wong et al., 2014)

end of feeder. The DSTATCOM was also proven to reduce the number of OLTC operations, thus extending infrastructure lifetime which would aid in balancing cost. The prevailing conclusion is that technological solutions do exist to mitigate voltage fluctuations, however, consumers are under no obligation to deploy them.

A Swedish based impact assessment model for PV-DG using three grid topologies is presented by Widen et al. (2010), and further progressed by Lingfors and Widén (2016). The three grids are modelled with an increasing number of nodes and cabling length. A stochastic PV and demand model is used to produce a power flow simulation with only solar irradiance as an input. This methodology is simple and versatile in identifying some of the grid impact issues, including reverse flow, voltage fluctuations and load matching. Widen et al. (2010) claim that they produce “detailed and realistic data on voltage fluctuations”, however, only an hourly irradiance input was used. The reason given was due to the computational demand, which is not insignificant when modelling multiple nodes on a grid. Furthermore, voltage fluctuations are claimed to be insensitive to time averaging (Widén et al., 2010).

It is the view of the author that the justification by Widén et al. (2010) stating “statistical investigation of voltage variations in the presence of PV-DG does not require higher resolution than 1 hour”, is too strong a conclusion and would require the amendment of a clause stating “in the presence of high penetration of PV-DG”. Whilst it is accepted

that taking aggregations of demand and voltage has less impact from time averaging, the effect from a single property is still subject to real-time variability. Their analysis shows that for each of the 13 nodes analysed, the 10 minute maximum demand was always greater than the 1 hour time averaged equivalents, and minimum demand always less than the equivalents. When analysing the irradiance, the 2 min maximum power was 2.2% greater than the 1 hour average. An assumption of the research by [Widén et al. \(2010\)](#) is that PV-DG will be installed simultaneously, and that when there is spatial decorrelation and a well distributed uptake of PV-DG in the grid, there is a lesser impact from using lower resolution data. Practically, installations are not likely to happen simultaneously. Therefore, it is not an option to ignore scenarios where a single or few PV installations exist on the LV grid if the apparent impacts from PV-DG be quantified. The need for high resolution data is well demonstrated by [Yan et al. \(2014\)](#) where voltage violations at the end user are observed at 30s intervals.

[Yan et al. \(2014\)](#) detail the analysis of OLTC transformers and voltage regulation under increasing PV penetration using real, high resolution generation and demand data. OLTC operations that maintain voltage within permissible limits are shown to increase with PV penetration. The distance from the OLTC is also shown to significantly impact the effectiveness of OLTCs. When the OLTC is far from the PV, OLTC sensitivity is not enough to detect far away voltage violations and so tap changes do not occur even though end user violations do. This presents further issues from PV as the conventional grid set-up may not be adequate to handle voltage fluctuations. They present a beneficial design of using a small distribution static compensator which compensates the on-line voltage with reactive power. The analysis approach by [Yan et al. \(2014\)](#) is considered one of the best for analysing voltage fluctuations. The use of high resolution real world data for a year of PV power output and residential load consumption allows the production of informative results.

Not all impacts from PV are considered negative. [Mouheb et al. \(2012\)](#) present a study on a lengthy, remote, and presumably weak feeder on a LV network in Wilaya de Chlef, Algeria. The location studied is significantly far from an electricity generation source and so electricity supply fails during high summer tension. PV generation matches the demand profiles of Wilaya de Chlef well as demand is proportional to ambient temperature in Algeria. PV successfully prevented significant voltage drop outside of permissible limits along the length of the feeder. Intermittency could not be considered in this study due to insufficient temporal resolution, however, the methodology is relevant when addressing the lengthy rural feeders, such as those found in the Australian electricity network feeder taxonomy and how PV can be used to support the grid as opposed to damaging it.

Most methodologies reviewed are only concerned with their perceived version of extreme scenarios. This is typically the highest solar PV output coinciding with lowest load. This is quite often in the form of smoothed, summertime clear sky profiles of irradiance that do not consider diurnal variation, solar intermittency issues and the frequency of fluctuations. This type of approach is therefore deemed insufficient to address the issue of voltage fluctuation. High temporal resolution data that can encompass solar intermittency is a requirement to truly capture the impacts from voltage fluctuations for all levels of PV penetration. There was a distinct lack of quantifying the impact. Many papers identified the occurrence of voltage fluctuations but failed to statistically summarise or convert to a form of metric that would be useful in comparative studies. The majority of research concerning voltage fluctuation focusses on technological solutions of the issue. Many control systems and inverter strategies are proposed that can adequately respond to fluctuations, however, their ubiquity in installed or new inverters is not commented upon. Therefore, it is the author's belief that technological solutions, unless uploadable to existing technology, do not satisfy the concern of DNOs for existing installations. Studies on voltage fluctuation require a metric in order to characterise and compare it from location to location, perhaps using overuse of existing grid equipment in the form of OLTC operations would be a useful metric.

2.3.2 Harmonic distortion

Grid connected power sources act as non-linear loads whereby the impedance changes with applied voltage drawing distorted current and voltage waveforms that contain harmonics (Fekete et al., 2012). Harmonics of a waveform are components whose frequencies are multiples of the fundamental waveform. Harmonics occur when a load is connected that draws from voltage that differs from the fundamental wavelength (Wakileh, 2001).

Harmonics have a significant impact upon the operational efficiency and the reliability of a power system, the loads and the protective relaying (Du et al., 2013). Furthermore, harmonics can have knock on effects that cause disruptions in other electrical equipment such as telephone transmission interference and cable TV (Dartawan et al., 2012) as well as degradation of conductor and insulation material (APT, 2011). A further concern is overheating of equipment on account of altered current, which can lead to failure or premature ageing (Dartawan et al., 2012; Wakileh, 2001).

The root mean square of all harmonics against the fundamental harmonic is known as the total harmonic distortion (THD) (APT, 2011). There are no globally established

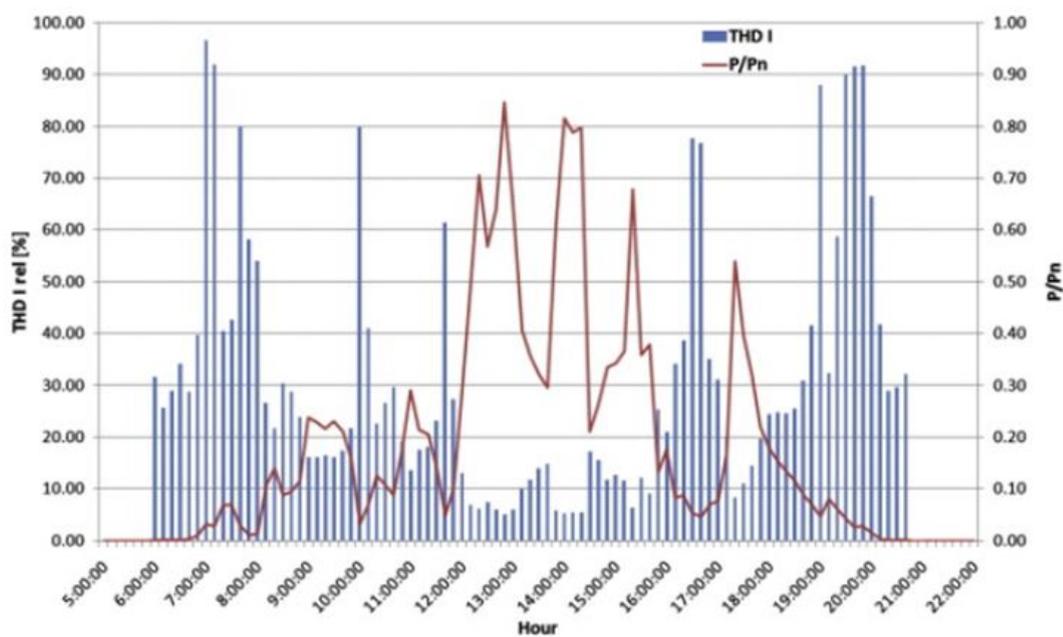


FIGURE 2.3: Total harmonic distortion relative % and PV generation factor over a winter day in Croatia (Fekete et al., 2012)

limits or regulations for THD control, although EU members use the EN50160 that stipulates harmonics up to the 40th harmonic must remain within 8% of their nominal value. However, this can be adapted by country. For example, Croatia set the limit at 2.5% (Fekete et al., 2012). There exist two industry standards by the IEEE: 519-1992, specifying the limits on amount of harmonics allowed within the power system; and 1547-2003, focussing on the interconnection of renewable resource which includes inverter connected PV (Dartawan et al., 2012). Existing inverter standards in Australia (AS4777.2) for small PV installations require $< 5\%$ THD on the current injected into LV grids with tighter stipulation on specific harmonics (Passey et al., 2011). With increasing growth and penetration of PV installations, inherent harmonic distortion caused by the inverter connection to the grid is becoming a concern (Du et al., 2013). Unfortunately, harmonics are difficult to model due to the requirement of sub-second data to truly analyse the impact. It is an important issue to explore (Hernandez et al., 2011; Zhao et al., 2010).

THD 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), further demonstrating that using clear sky irradiance profiles is insufficient for impact modelling as the clear sky profile will always provide the maximum power output, which ignores the lowest rated power. THD fluctuations with PV rated power is illustrated well in figure 2.3, which shows an intermittent generation profile and how the harmonics change through the day. The plot indicates that the THD relative percentage is reasonably insensitive to

intermittency so long as the rated power stays above 25%. The drop from 80% of the rated capacity to 30% between 13:00 and 14:00 causes a <10% increase in THD. Comparing this to a similar drop in output between 16:00 and 17:00 to below the 25% of the rated power, the relative THD level shows a 50% increase. However, this is thought to be an issue in calculation. [Chicco et al. \(2009\)](#) showed that the absolute harmonic changes under low power operations were small while the THD at corresponding times were high. The implication here is that THD is an inadequate metric for describing the actual impact of harmonics on the grid. A suggested alternative would be to limit the absolute harmonics that occur on the system.

A clear distinction between total current harmonic distortion (TcHD) and the total voltage harmonic distortion (TvHD) exists ([Wakileh, 2001](#)). [Fekete et al. \(2012\)](#) studied these distinctions between voltage and current and found that TvHD was independent of the power output remaining at approximately 3% throughout a day despite solar intermittency. TcHD, on the other hand, was strongly dependent on solar generation conditions. Less research exists exploring this aspect, although [Xavier et al. \(2017\)](#) state that currently installed inverters are able to control TcHD to within acceptable limits. [Bhowmik et al. \(2003\)](#) derived an algorithm to discern the penetration limit of DG on an LV grid using the TcHD as an indication. They tested three arrangements of feeder types at increasing length against three loading patterns. The penetration was defined as the percentage of load that could be generated from PV-DG before encountering 3% TvHD. Shorter feeders could handle more load generated from PV-DG, although larger installed capacity was achievable on longer feeders. The methodology is an excellent tool for DNOs to preliminarily assess potential penetration with regards to harmonics. [Dartawan et al. \(2012\)](#) present a simple PV penetration analysis using grid modelling software. They find that harmonic injection with < 3% TcHD, a penetration level larger than 100% of the LV grids peak load could be achieved suggesting that management of TcHD can facilitate large penetration. The issue is that the TcHD depends on the location of the installations. Furthermore, should TcHD become the limiting factor for increased PV penetration, there are cost effective solutions to reduce TcHD back to permissible levels.

Models proposed to analyse harmonic issues rarely use high frequency data. [Fekete et al. \(2012\)](#) uses 10 minute resolution. Methodologies require improved solar and load data inputs if only to capture solar intermittency. It is suspected by the author that a sub-second analysis of THD is needed to establish what the suitable temporal resolution for modelling harmonics is.

Control strategies to manage THD are well detailed in literature ([Paatero and Lund, 2007](#); [Monfared and Golestan, 2012](#); [Albert, 2016](#); [Jana et al., 2016](#); [Xavier et al.,](#)

2017). Most strategies achieve THD regulation control with a quick dynamic response and many are already installed in new and existing PV-DG inverter technology. Some methodologies (Paatero and Lund, 2007; Guerrero-Rodríguez et al., 2015; Kavali and Mittal, 2016) manage to mitigate harmonics as well as voltage variations. Therefore, THD can be considered as controlled at inverter level for future installations of PV-DG. Looking at TcHD in isolation, the use of high-pass filter and a notch filter in parallel with the grids capacitors is both a cost effective and successful strategy for the DNO (Dartawan et al., 2012).

It is the author's opinion that the increased presence of harmonics on account of PV-DG should not be a priority concern for DNOs, nor is it the main limiting impact. Greater penetration was allowed before harmonic standards were breached than is observed for voltage fluctuations, therefore voltage fluctuations would be the bottle-neck for penetration increase. Furthermore, should the scenario arise where harmonics were the bottle-neck issue inhibiting PV-DG penetration, simple cost effective technical solutions exist to remove it.

2.3.3 Unintentional islanding

Islanding describes the scenario where PV-DG operates independently of the grid, and is no longer controlled at the transformer. Islanding occurs when the grid is subject to failure or routine shut-down but connected PV-DG continues generating power (Pourbabak and Kazemi, 2014).

Islanded operation mode can be beneficial in certain cases, such as back up power reserve and uninterruptable power supply (ElNozahy and Salama, 2014b). Therefore, it is important that it is well managed for the most effective use of renewable energy distributed generation (Palizban et al., 2014). Passey et al. (2011) state that islanding can cause the following problems: safety concerns to technicians, cause the continuation of a fault trapped within the island, reduced power quality, transient over-voltages, inverter damage when reconnected to grid, and damage to connected equipment. Islanding is unintentionally present in almost all countries that deploy DG (Hernandez et al., 2012).

Islanding is a very well researched issue with numerous papers detailing control and detection strategies (Li et al., 2014; Bakhshi and Sadeh, 2016). It is not found to be a function of solar resource intermittency and is controllable using commonly deployed technological solutions found in many currently installed inverters. For this reason it is not considered a high priority grid impact for DNOs, nor is a high temporal resolution irradiance time series imperative. Islanding is not instigated by solar intermittency or

demand peaks but instead grid failure or scheduled shut down. There is the chance that grid impacts associated with PV-DG cause the grid failure, however, this is covered in analysis of voltage fluctuations and rise, increased demand peaks and harmonic distortion.

2.3.4 Reverse flow

One issue identified is voltage rise during high PV generation and low-load situations. [Mokhtari et al. \(2014\)](#) and [Widen et al. \(2010\)](#) state that for conventional operation of electricity networks to continue, the grid requires updating with modernised strategies to manage voltage rise.

Reverse flow is the scenario where the voltage across a transformer flows from the LV grid back to the medium voltage (MV) grid. It is a direct response to voltage rise within the LV grid from installed DG ([Samadi et al., 2014](#)).

The impact of reverse flow within the grid is of concern to DNOs because of uncertainty surrounding potential damage to infrastructure ([Gao, 2013](#)). Old electricity grids are typically designed with intention to flow in one direction, from power plant, to transmission network, to distribution network, to load ([Passey et al., 2011](#)). Older components consist of protection equipment that may be damaged when subject to reverse flow ([Hernandez et al., 2012](#)). However, modern grids are designed to accept reverse flow, although this is typically limited through legislation. Reverse flow across the MV/LV transformer is limited to 60% of the transformer nominal rating in Ontario but more typically limited to 30% around the world ([ElNozahy and Salama, 2014a](#)).

[Haque and Wolfs \(2016\)](#) state that the impact of reverse flow is similar to those of voltage fluctuations. Voltage rise can cause increased usage of network equipment. The number of operations performed by OLTCs can be dramatically increased, which shortens the expected life cycle of these devices. OLTCs are the most widely used voltage control devices on transformers. They are an automated switch device that allows the voltage to be increased or decreased according to predefined limits. OLTCs operate by increasing the length of transformer coil overlap so that a fixed step down can occur, or in reverse flow instances, step up ([Choi and Kim, 2001](#)). As they are automated, increases such as reverse flow and voltage fluctuations can cause overuse ([Gao, 2013](#)). The number of OLTC operations is a metric occasionally used in literature to attempt to quantify grid impacts and can be used to generate baseline comparisons in order to compare the impact of PV-DG ([Lave et al., 2015](#); [Nguyen et al., 2016](#)). OLTCs were shown to respond sub-secondly to voltage fluctuations by [Yan et al. \(2014\)](#). 20% integration of PV-DG causes a four-fold-increase in number of tap changes, which

significantly reduces the lifetime of the equipment. This conclusion was modelled with low resolution data; it is likely that using high resolution data would show further increases in OLTC activity.

A large increase in uptake of PV-DG brings with it substantial amounts of reverse flow (Widen et al., 2010) and can lead to degradation of electricity network stability (Vallee et al., 2013; Vallée et al., 2015). Widen et al. (2010) showed that when all properties within an LV grid have a 1 kWp PV system, reverse flow across the MV/LV transform was observed lasting for 0.1 to 3.5 hours a day. When a 3 kWp system is installed at each property, the daily observed reverse flow was observed lasting for 1.1 to 8.9 hours (Widen et al., 2010). Despite this reverse power flow, Tonkoski and Lopes (2011) state that it can be avoided without conservatively limiting the capacity of PV-DG units through inverter design. This is achieved through smart curtailment whereby power is generated and released into the electricity grid in accordance to the current voltage. In the case that the voltage on the grid is too high, the generated power is curtailed. This is not an ideal solution as it reduces the carbon effectiveness of PV.

There is little research committed to exploring reverse flow in isolation, although some models do allow its evaluation. Without detailed information about the nature of reverse flow upon the transformers, it is difficult to evaluate the extent of impact upon the LV grid. ElNozahy and Salama (2014b) present a methodology that explores how changing the residential load can cause overloading of distribution equipment, infrastructure upgrade requirements, overuse of voltage regulation equipment, end-user power quality degradation, increase in total system power losses, and substation transformer capacity limit breach. However, there are commercial loads present in this methodology and this reduces the potential for reverse flow; typical LV grids do not have commercial loads. Higher resolution data would be required to adopt this model as no solar intermittency was considered through use of 1 hour input data.

The reverse flow of voltage is found to be managed either at the inverter through managed curtailment or at the transformer. The main impact is to DNO operated network equipment such as OLTCs. The frequency of use of OLTCs is identified as a useful metric to benchmark impact of PV-DG in terms of voltage changes. Analysis of reverse flow overlaps with analysis of voltage fluctuations. Any steady state model considering voltage fluctuations could, therefore, also consider reverse flow.

2.3.5 Hot-spots

Hot-spots is not a grid impact as such, it is more the phenomenon whereby PV uptake clustering results in significantly more PV installations on a particular part of the LV

grid; a non-uniform uptake pattern. The clustering effect is well studied in literature and usually attempts to ascertain the driving social, political, environmental and economic factors that most significantly impact PV uptake. The purpose of identifying hot-spots is so that the DNO can know where they should expect to plan upgrades or reinforcement to local LV grids. Predicting where hotspots will occur was a key concern for Northern Powergrid (Walbank and Goodhand, 2014).

The DNOs response to PV-DG penetration is currently reactive because the random location of initial uptake (Balta-Ozkan et al., 2015) of PV-DG makes predicting potential target areas for mitigating future voltage balance incredibly difficult. The clustering effect is best described by Kwan (2012), Snape (2016) and Westacott and Candelise (2016) as aggregation of a technology to a localised point.

Spatial-temporal PV and/or EV uptake models exist in abundance for a whole host of uses. The most notable works are the PV uptake models by Higgins et al. (2013), Snape (2016), Kwan (2012), Cai et al. (2013) and Balta-Ozkan et al. (2015). Economics are well represented by Cai et al. (2013) who claim that the most significant factor for uptake is to consider financial aspects. However, this is in direct contrast to Balta-Ozkan et al. (2015), who find little or no statistical significance between PV uptake and household income. Kwan (2012) found some income weighted in the USA. The suspected cause of the finding by Balta-Ozkan et al. (2015) that income was not statistically significant stems from the opportunities of roof renting and the increasing number of house rentals. Factors included in uptake models are resource availability, government incentives, home value, household income, age, education, ethnicity, area type, political affiliation, among others.

The use of Light Detection and Ranging (LiDAR) and Geographic Information Systems (GIS) is a type of uptake modelling that is receiving increased interest in research. LiDAR data offers a 3D visualisation of an area; some GIS software can calculate the expected solar irradiance based on this height data. A resource methodology is presented in Gooding et al. (2013) that overlays building footprint data to estimate the pitch and aspect of each property. Extending from this methodology, the residential PV capacity of the target area can be established. A GIS approach is presented by Karteris et al. (2013) and can be applied to perform a cost analysis of a policy on PV. Similarly, GIS data and solar radiation maps are used to assess the potential of PV (Bergamasco and Asinari, 2011; Bergamasco, L. and Asinari, P. , 2011; Karteris et al., 2013; Jakubiec and Reinhart, 2013; Lukač et al., 2014). These GIS methodologies are a good way to assess the physical availability, and when combined with uptake models are seemingly effective. This type of methodology is applied by Sun et al. (2013) alongside a UK-based socio-economic uptake model, which considers income,

education, environmental consciousness, building stock and ownership. From this the likely uptake of PV is estimated, the methodology demonstrated the need for more regionally focussed policies for true effectiveness.

The most relevant work to future based modelling efforts have so far come from studies carried out by [Higgins et al. \(2013\)](#) and [Paevere et al. \(2014\)](#). PV uptake modelling are presented by [Higgins et al. \(2014\)](#), taking into consideration previous experience from the other reports. [Snape \(2016\)](#) states that diffusion based methodologies are the most appropriate for uptake modelling. This type of modelling could be used to identify hot-spots.

Once the locations of hot-spots are identified, methodologies outlined in the other grid impact sections can be applied should the DNO provide real grid schematics. The empirical analysis of a real hot-spot feeder would help provide insight on the worst case scenarios for PV integration within the grid and enable appropriate grid impact mitigation. It is worth noting, however, that there are few empirical grid impact studies. The DNO should take a more proactive role in providing empirical analysis opportunities.

2.4 Review of the response time of power systems

It is intuitive that solar radiation will fluctuate on whatever resolution to which it is analysed. There are annual, monthly, daily and sub second fluctuations that can be attributed to the Earth-Sun position, seasons, atmospheric aerosols, weather and cloud. All the attenuating factors combine to make rapidly responding irradiance fluctuations. It is, therefore, important to have an understanding of the resolutions of which to analyse.

As has been demonstrated with this literature review, there is very little that concerns the high temporal resolution study of solar irradiance. There can be two reasons for this. The first is that the data is simply not available with which to do a study. The second is that the power systems analysis being undertaken did not require a high frequency irradiance input as the system does not respond to such changes. It is therefore important to observe the response of different power system controls so that an appropriate temporal resolution can be identified.

[Karimi et al. \(2004\)](#) derive a method to evaluate the rate of response of electrical frequency from various electrical impacts. They find frequency responses in the order of 40 to 200 ms, though the speed of response is almost independent of the magnitude of frequency change. There is a 10 ms delay from a ramping event, and a response time

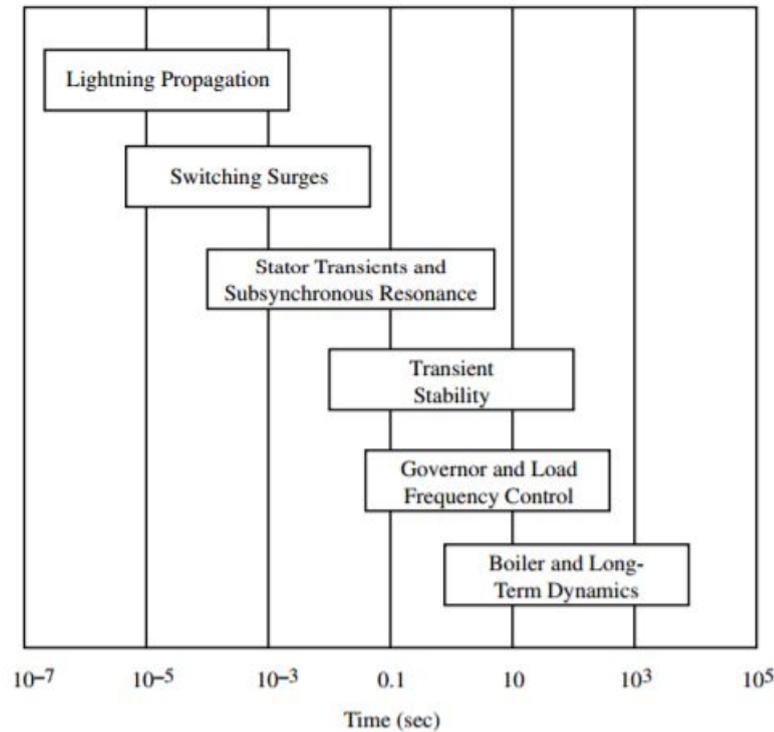


FIGURE 2.4: Power response time of different power equipment (plot produced by [Pai \(2016\)](#))

of 200 ms. Other response times are suggested as 100 ms ([Sánchez et al., 2000](#); [Rao et al., 2000](#); [Lokhande et al., 2017](#)), up to a few seconds ([Patel et al., 2014](#); [Echavarría et al., 2007](#); [Lokhande et al., 2017](#)) with voltage tolerances of $\pm 15\%$ ([Echavarría et al., 2007](#)). Modern OLTC technologies can respond within a single millisecond ([Patel et al., 2014](#)). The dynamic time responses of different power systems is displayed in figure 2.4. There is no question, therefore, that the electrical components operate at a significantly higher resolution than any grid impact study has performed. This is not the only limiting factor, however.

[Li et al. \(2004\)](#) explores the response time of real and reactive power for DG penetrated microgrids. They find responses in power on the same frequency as the voltage, though see a disparity in control reaction. Islanding detection occurred 6-sec after the triggered event, and subsequent reassessment for reconnection of islanding mode after 13-sec. Furthermore, they find that it takes 43-sec for synchronisation algorithms to begin functioning after a triggered DG power fault. This presents a more realistic target with which to produce irradiance, when compared to 50 Hz. Therefore, there is a systematic delay in decision algorithms that are far more detrimental than the near instantaneous response of OLTCs.

[Gao and Redfern \(2010\)](#) states that DNOs strive for faster OLTCs. However, an

intentional time delay is always included to avoid tap changes in response to very short duration voltage changes. This programmed redundancy in the majority of OLTCs mean that intentional response delay is most often set in the range of 30 to 60 seconds. Furthermore, an additional time is included between tap changes in the order of 5- to 10-sec (Zhu et al., 2000), though it is demonstrated that a reduced time delay and inter-operation delay benefits overall voltage quality. Thus, there is a trade-off between voltage quality and number of OLTC operations, as increased operations reduces lifetime of equipment (Zhu et al., 2000).

Irradiance fluctuations have been demonstrated to occur on sub-second temporal scales (Lave et al., 2012; Omran et al., 2011), and so it would be wise to attempt to achieve the same resolution, however, as synthetic irradiance (as will be shown) has not yet been achieved to this degree, there must be a trade-off. The use of OLTCs with their 30 to 60-sec delay response time makes them a suitable candidate to explore the impact upon them. The irradiance, therefore, should be synthetically derived with at least 1-min resolution as a minimum. Fundamentally, the production of spatially correlating solar irradiance at 1-min resolution (as will be demonstrated in chapter 4) has not been achieved before, and therefore serves as a new benchmark tool from which to research from.

2.5 Future considerations for grid impact analysis

This literature review explored applicable research methodologies that analysed grid impact of PV on the LV grid as viewed from the DNO. From this study, the following discrepancies and opportunities present themselves.

There is a distinct lack of real grid network testing. Whilst Mu et al. (2014) and Widen et al. (2010) analysed real networks available through their respective universities, there has been surprisingly little collaboration with DNOs to empirically measure PV within real grids that would help validate work and quantify research models. Research facilities offered by groups, such as CSIRO's Renewable Energy Integration Facility, could be used to validate methodologies. There remains a distinct lack of industrial collaborations between DNOs and research. Until they can be realised, it would be worthwhile to acquire real LV grid topology data so that it can be modelled. Only representative topologies exist that are taken confidentially and typified such as those available from the Electrical Power Research Institute (EPRI, 2008). Through cooperation with DNOs it is possible for them to provide schematics (Walbank and Goodhand, 2014). Data containing transformer details, cable lengths and thicknesses,

installed protective equipment, scheduled upgrading, and accurate geographic location would be the ideal variables.

A study of the literature finds that the majority of research offered introduces new control or detection strategies for future technological improvement. Whilst this is of clear importance, it does not serve for a base level study of the impacts presented. Consensus baseline metrics must be established for grid impact analysis.

The grid impact identified as most concerning is voltage fluctuations. The predominant reason for this is that current inverter technology and cost effective on-line technologies manage harmonic distortion and unintentional islanding. Furthermore, reverse flow and demand peaks are demonstrated in literature to be not a significant concern. Voltage fluctuations, however, require either curtailment or on-line equipment such as the OLTCs in order for it to be managed. Curtailment is an environmentally and fiscally poor option although it is successful in mitigating voltage fluctuations. Overuse of OLTCs is financially detrimental to DNOs as they require replacement sooner than planned, and so a limitation is applied on PV penetration. The use of OLTC operations as a metric would, therefore, be a good indicator of the grid impact of voltage fluctuations caused by PV penetration. In order to establish the baseline grid impact, the use of ancillary services from batteries should not be considered until a later phase of study.

The most limiting factor common to the literature is high temporal resolution data availability. Some work requires significant and extensive data sets, which are usually unavailable or expensive (Gennaro et al., 2014; Higgins et al., 2014; Mu et al., 2014). Data gathering or data synthesis is a fundamental step before any impact analysis can be performed. The literature reviewed found many studies to use an unsuitable temporal resolution data input to appropriately quantify voltage fluctuations. The synthesis of input solar irradiance at an appropriate resolution that facilitates grid impact analysis is, therefore, the basis of this thesis.

Chapter 3

Solar Irradiance Generator (SIG) development

The research aim and objective addressed in this chapter is to ascertain whether it is possible to develop a statistically accurate methodology that can produce synthetic irradiance time series with a 1 minute temporal resolution, using only readily available, well-geographically dispersed, mean hourly meteorological observation data as an input. The model will be referred to as the solar irradiance generator using the acronym SIG.

This chapter will first introduce key concepts and definitions fundamental to understanding this thesis. Secondly, the concept of temporal solar irradiance generation through a review of the most recent literature, highlighting alternatives and opportunities for development will be introduced. Thirdly, the SIG will then be described in overview so that the reader may follow the steps and have a point of reference; each significant step will feature its own section. The SIG is then subject to a temporal validation before discussing the extents of the research. Furthermore, the SIG is provided in its raw script, as produced using [Matlab \(2015\)](#) software. It is appended to this thesis and is separated into sections that correspond well to this chapter, see section [A](#).

The SIG presented here was condensed and published in the Journal of Solar Energy in [Bright et al. \(2015\)](#) and further concept developments published in [Smith et al. \(2017\)](#). Sections of this work were produced in collaboration with Dr Christopher Smith. The author makes clear where Dr Smith carried out independent work or provided data or script, otherwise all work is the author's own.

3.1 Key concepts and definitions

This section will introduce fundamental concepts and definitions required to understand this thesis. There are many more concepts and approaches to be discussed that are not detailed here as their utilisation requires justification from alternatives that may exist. The information in this section is consistent throughout the thesis and so are provided here to enable better understanding of the research.

3.1.1 Types and definitions of solar radiation

The different terminologies for the myriad of ways to describe solar radiation can be confusing. This section will explain the different types of solar radiation that feature in this thesis, as well as the different ways irradiance is expressed.

There are three distinct types of radiation to be understood that feature heavily in this thesis: direct, diffuse and global.

Direct or beam normal irradiance (DNI, G_B) is the irradiance received directly from a 5° field of view concentric around the sun on a surface that is normal to the sun at its position in the sky ([PVPerformance Modeling Collaborative, 2016](#)).

Diffuse horizontal irradiance (DHI, G_D) is the terrestrial irradiance received by a horizontal surface which has been scattered or diffused by the atmosphere. It is the component of global horizontal irradiance which comes from the rest of the sky ([PVPerformance Modeling Collaborative, 2016](#)).

Global Horizontal Irradiance (GHI, G) is defined as the amount of terrestrial irradiance that falls on a surface horizontal to the surface of the Earth. It is the combination of all the incident irradiance upon that plane consisting of both the direct and diffuse components. GHI is the harnessable irradiance by a solar panel (or other) positioned perpendicular to the Earth's surface.

Other definitions are required, however, to express these types of irradiance depending on where they are measured and the nature of incidence.

Extraterrestrial irradiance is the irradiance at the edge of the Earth's atmosphere, one component of which is the horizontal value, G_o , measured perpendicular to the Sun's rays.

The global clear-sky irradiance (CSI, G_{cs}) is a theoretically calculated expression of GHI under conditions where the sky is perfectly clear ([Bird and Hulstrom, 1981b](#)). It represents the global irradiance in the absence of visible clouds ([Reno et al., 2012](#)) and,

therefore, describes the attenuation of G_o through the atmosphere. The distinction between CSI and GHI is that GHI considers all attenuation whereas CSI only concerns atmospheric attenuation and can be considered the maximum available irradiance at ground level. Importantly, the GHI can exceed CSI from either underestimation of CSI or irradiance enhancement events such as additional DNI from reflections off clouds or other surfaces, both of these scenarios are commonplace. CSI has been simply expressed as a function of the zenith angle, although it can encompass many variables such as the atmospheric state such as air pressure, relative humidity, temperature, aerosol content and Rayleigh scattering.

In this thesis, tilted global irradiance, or just tilted irradiance, is defined as the sum of all irradiance incident upon a tilted plane and is denoted G_t . The methods to determine the irradiance on a tilted plane will be discussed in chapter 3.

3.1.2 Clear-sky and clearness indices

The clear-sky index, denoted k_c , is fundamental to research presented in this thesis. It is also important to understand similar parameters such as the clearness or cloudiness index, k_T , so as not to confuse them with k_c .

[Black et al. \(1954\)](#) first posed the concept of normalising radiation measurements to their associated clear-sky potential. Inspired by [Ångström \(1924\)](#), who explored the relationship of actual sunshine duration to potential maximum sunshine duration, and by [Prescott \(1940\)](#), who advanced the concept to explore the duration of sunshine and the theoretical amount of radiation potential if the atmosphere were perfectly transparent ([Brunt, 1939](#)), [Black et al. \(1954\)](#) explored normalising daily values of G by daily values of G_{cs} . [Liu and Jordan \(1960\)](#) provided the first modern interpretation of the concept of “cloudiness index” when normalising the hourly and daily measured global, diffuse and direct irradiance by the extraterrestrial radiation to arrive respectively at indices of k_T , k_d and k_D . The cloudiness index is often termed clearness index ([Duffie and Beckman, 2006](#)) and henceforth will always be denoted as k_T . The reason for this preference is that the k_T implies that it is a better representation of clouds, whereas normalising by the extraterrestrial irradiance actually encompasses all atmospheric losses as well. k_c would be much more deserving of the term “cloudiness index” as the other atmospheric losses and light scattering are inherently contained within G_{cs} . k_T is calculated as

$$k_T = \frac{G}{G_o} \tag{3.1}$$

Bird and Hulstrom (1981b) enabled advancement in irradiance index classification when they produced a methodology to estimate the clear-sky radiation incident upon the Earth's surface, G_{cs} . The derivation of G_{cs} provided a new term that could be normalised against, leading to the birth of the k_c defined as

$$k_c = \frac{G}{G_{cs}} \quad (3.2)$$

G and G_{cs} are both irradiance values that arrive at the surface of the Earth on a horizontal plane, this makes k_c a dimensionless property. G encompasses all losses up until ground measurement, G_{cs} is the theoretical maximum under clear-sky conditions and accounts for atmospheric absorption and scattering losses, but not those as a result of clouds. The implication is that k_c describes the losses or gains that are principally attributed to the presence of clouds, whereas k_T describes all losses from outside of the atmosphere to the ground.

k_c has a significant advantage over k_T as it allows the removal of diurnal and seasonal signals from radiation time series enabling advanced analysis techniques such as wavelets or computing power content fluctuation (Engerer and Mills, 2014).

k_c is equal to 1 when $G = G_{cs}$. k_c is less than 1 when $G_{cs} > G$, the real losses are greater than the theoretical, indicating cloud or heightened atmospheric conditions of scattering, absorption or other. The most common cause of $k_c < 1$ is from clouds. k_c is greater than 1 when $G_{cs} < G$, this can be from either reduced impact of atmospheric considerations or from added reflections, often from the edge of clouds.

3.1.3 Markov chains

To synthetically recreate weather events, Markov chains are employed to capture the transition probabilities of meteorological variables over time. The application and mathematics will be discussed in detail in chapter 3, however, the concept and brief history of Markov chains will be introduced here.

Markov chains are a method of describing changes in events over time. The origin of Markov chains begins with Plato, who speculated that after an uncountable number of years, the universe would return to its pure form. Pure forms are abstract concepts such as a perfectly straight line or a circle existing naturally in the world. In the 16th century, Bernoulli explored the mathematics to try and describe the un-pure forms and patterns of the world by developing upon the theory of expectation. He applied mathematics to accurately estimate the unknown probability of an event as derived

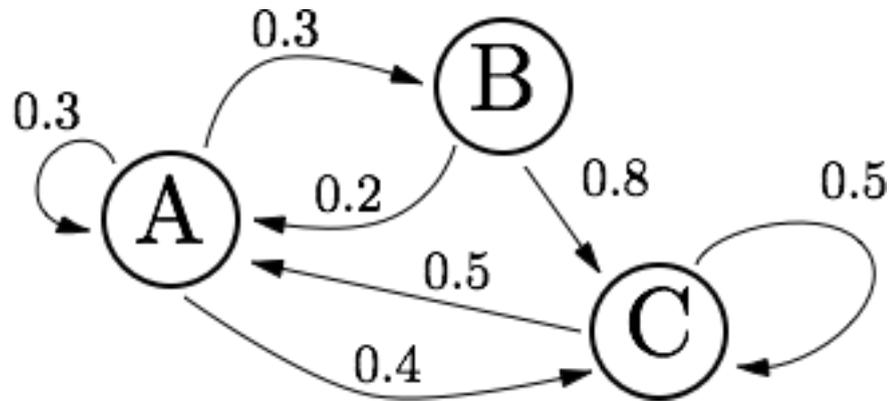


FIGURE 3.1: Example of a three-state Markov chain with the associated transition probabilities from each of the three states.

from the frequency of that event's occurrence within independent trials (Basharin et al., 2004). This is most aptly described with a cup of white and black beads. Selecting a bead from the cup, Bernoulli recorded the outcome before replacing it and repeating the process. He noted that the expected value of white versus black observations will converge on the actual ratio as the number of trials increases, this is now known as the theory of large numbers. With more application and research, probability distributions emerged which suggested that the average fate of events was somehow predetermined, known as the Central limit theorem. Theological opposition began. The idea of predefined statistics governing nature was a dangerous proponent to the religious doctrine of free will and so theologian-turned-mathematician Pavel Nekrasov made a claim that the law of large numbers is only applicable for events with independence, whereby the outcome of previous events does not affect the outcome of future events. Andrey Markov, a Russian mathematician with known public animosity towards Nekrasov, put his expertise to disproving him and extended the law of large numbers in 1906 by applying it to dependent variables (Markov, 1906). By doing so, he developed a new concept known as chain dependence. Chain dependence is a branch of stochastic mathematics that still persists to the present day, although it is popularly known as Markov chains (Seneta, 1966).

The use of Markov chains facilitates a stochastic process in a probabilistic mathematical method whereby transitions from one state to the next are directed by discrete probabilities taken from the statistics of real-world processes. By recording the transitions from one state to the next, the number of transitions can be used to interpret the probability of transition. This can be simply demonstrated by figure 3.1 which shows a three-state Markov chain. Notice that states A and C have a chance of 0.3 and 0.5 respectively to remain at their current state, however, B will always transition to either A or C. The total probability leaving each state will always equal 1 because a new state must always occur and is therefore a certainty.

Markov chains are most effective when used to describe states that have temporal dependencies on previous states. Take a fair coin toss, for example. The probability of any fair coin toss is independent of any toss that preceded it, and so Markov chains would represent the transition of each state to either heads or tails as equal. This is not a good use of Markov chains as there is no dependency on a previous state. However, using a Markovian process, it would be possible to deduce a time dependent statistic, such as the distribution of the number of consecutive heads. Variables that follow a progression are much better described using Markov chains. The first ever use of Markov chains was by Andrey Markov himself. He applied his self-titled mathematics to Alexander S. Pushkin’s poem “Eugeny Onegin” in order to demonstrate dependent trials with what can be regarded as a simple chain. Of 20,000 letters in the poem, Andrey Markov calculated the stationary vowel probability is $P = 0.432$, that the probability of a vowel immediately following another vowel is $P_1 = 0.128$, and the probability of a vowel being followed by a consonant is $P_2 = 0.663$ (Markov, 1906). These fundamentals led Google co-founders, Sergey Brin and Larry Page to develop a page ranking analysis method called PageRank (Page et al., 1999), which is the driving force behind Google’s search engine success. Their method uses Markov chains to assign quality to web pages based on the input search parameters.

Markov models are a very popular method of stochastic data generation as the concept of temporally dependent states can be well applied to many data variables. Weather variables in particular lend themselves well to Markov chains as the weather from one hour to the next depend highly on each other. Markov models that consider weather have been used in many applications, from wind estimates (Masseran, 2015), solar energy estimations (Hocaoglu, 2011; Bhardwaj et al., 2013; Vindel and Polo, 2014b), and in weather time series generation (Yang et al., 2011).

Within this thesis, the states used for Markov chains are mean hourly meteorological observations such as cloud cover. Each state will represent an okta condition and so the Markov process will have nine states, equal to the number of okta conditions shown in table 3.1. By recording the transitions across a fixed time period, transition matrices can be constructed that detail the probability of transition from one state to the next.

3.1.4 Meteorology and atmosphere

The intention of this section is to give an overview of the complexities offered to the irradiance at the Earth’s surface as a result of meteorology and atmosphere.

The concept of the solar constant is important to understanding why meteorological conditions offer the most significant modelling difficulties. The solar constant is a

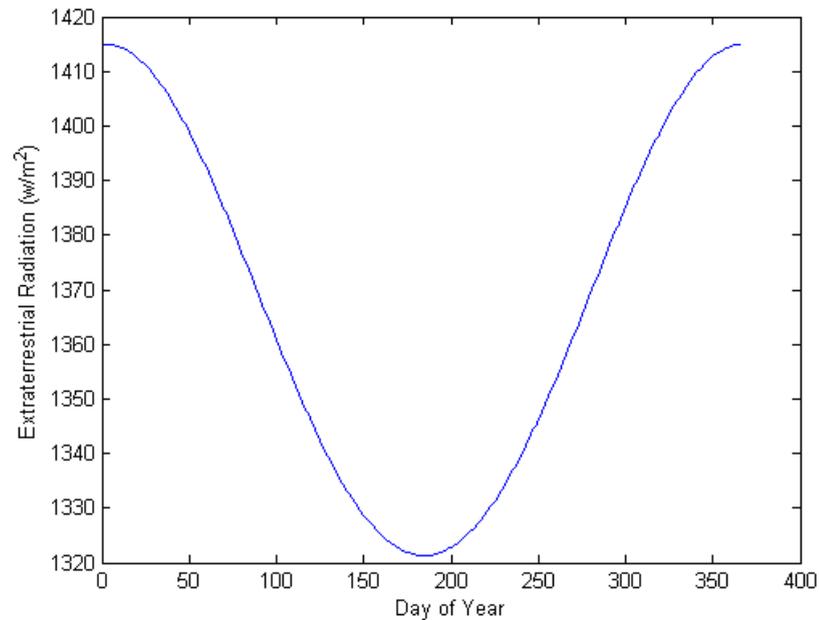


FIGURE 3.2: The annual variation in extraterrestrial radiation (PVPerformance Modelling Collaborative, 2016)

measure of the mean solar electromagnetic radiation incident upon an area of a plane perpendicular to the solar rays over a distance that is approximately the mean distance from the Sun to the Earth (Kopp and Lean, 2011). From maximum to minimum, the solar constant is approximately 1361 to 1362 Wm^{-2} . The significance of this is that the solar flux density is easily calculated without the interference of the Earth's atmosphere (Liu and Jordan, 1960).

The solar irradiance at the edge of the atmosphere is known as the extraterrestrial radiation. It is not as consistent as the solar constant as the Earth's orbit undergoes cyclical changes in distance to the sun. The solar constant is measured at a constant distance of one astronomical unit (AU) and does not account for orbit. The amount of extraterrestrial radiation that is received at the edge of the Earth's atmosphere on a plane perpendicular to the sun is a function of the solar constant and the Earth's position within its orbit, or time of year. Figure 3.2 shows how the extraterrestrial irradiance varies throughout the year.

The extraterrestrial radiation is the last calculation of irradiance before accounting for meteorology, atmosphere and all of the components within it. Iqbal (2012) details many of the factors that must be considered for modelling the solar radiation under clear sky conditions: water vapour, aerosols, optical path length, air mass, ozone, aerosols, scattering of direct solar radiation, Rayleigh scattering, Mie scattering, absorption of radiation by gases, absorbers across the light spectrum, direct spectral irradiance and atmospheric albedo. This is not an exhaustive list, however, it serves to demonstrate

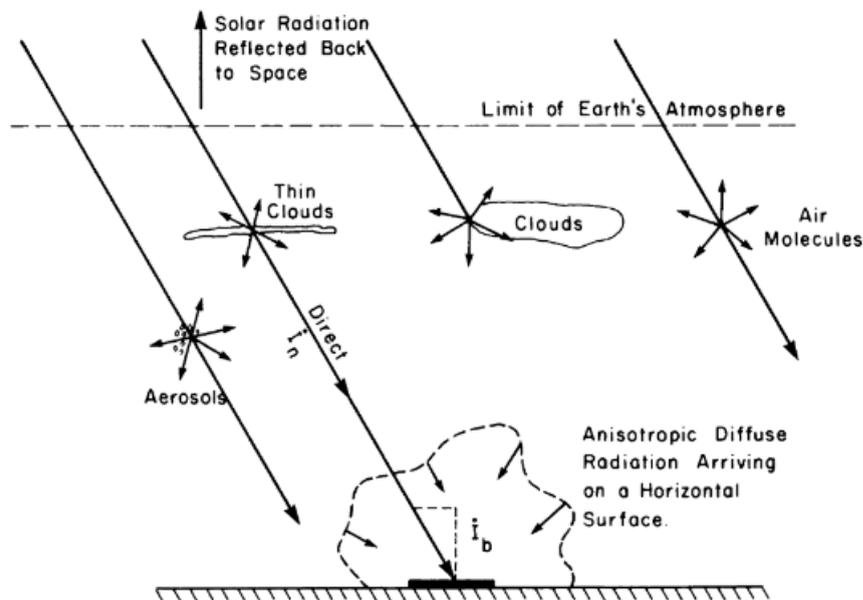


FIGURE 3.3: Solar radiation arriving on the ground under cloudy skies, inclusive of the many types of scattering and reflection losses in the Earth's atmosphere (Iqbal, 2012).

the complexity of the issue. These factors all have well established mathematical methods to incorporate them when calculating the solar radiation at ground level under clear sky conditions. The real problem presents itself when modelling the solar resource in clouded sky situations.

Figure 3.3 depicts the various complications with calculating the incident irradiance in the presence of clouds. Optical losses when travelling through clouds: scattering of thin clouds, cloud edge reflections, aerosol and air molecule scattering, and the diffuse irradiance arriving on a horizontal surface. Section 3.3.6 will cover these factors in detail, however, the concept is introduced here.

Clouds heavily influence the incident radiation received at ground level. When a cloud passes in front of the sun, nearly all of the direct normal irradiance is removed from the available extraterrestrial radiation. Cloud cover is a variable well recorded in the UK and in many parts of the world. Figure 1.2 shows all of the observation stations operated by the UK Met Office. The temperature, pressure, cloud amount, cloud type, rainfall and sunshine hours are routinely measured to a temporal resolution of typically an hour at these stations and at others around the world (Smith et al., 2017). Research has been carried out for nearly a century that has attempted to derive the solar irradiance from these meteorological variables; this is discussed in more detail in Chapter 3. The main variable of interest in this research is the total cloud amount.

TABLE 3.1: Conversion of the total cloud amount, N , in okta (reported as eighths) to the cloud coverage fraction, C (reported as tenths). $N = 9$ represents sky that is obscured due to fog/haze/other meteorological phenomena (NOAA, 2016b)

Okta, N (eighths)	Cloud cover, C (tenths)	Description
0	0	Completely clear sky
1	1 or less, but not zero	Very few clouds
2	2 to 3	Few clouds
3	4	Scattered clouds
4	5	Half of sky covered
5	6	Over half of sky covered
6	7 to 8	Many clouds
7	9 or more, but not 10	Small pockets of clear sky
8	10	Completely overcast
9	10	Obscured by meteorological phenomena

The variable *total cloud amount* is denoted N in this thesis, and is measured with units called *okta*. 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 (UKMO, 2010), it is denoted as C throughout this thesis and the conversion can be seen in table 3.1. The most significant loss of energy to the solar resource is the presence of cloud, the theory on how these losses can be modelled is presented section 3.1.

N is recorded by either a human observer or from a cloud-base recording laser. There is a general disparity between the two with agreements 39% of the time, and 88% within 2 okta (Wauben et al., 2006; Smith et al., 2017). This agreement is attributed to the subjective observer judgement at the end of an hour and the base recorder average over a whole hour (Muneer et al., 1998). Furthermore, directionality of the cloud-base recording laser (vertical) can influence the reading. The angle of the sun away from vertical in the UK is at least 27° and so the automated recording may not be representative. According to Smith et al. (2017), a meteorological convention for recording N is that the total cloud amount is recorded by a human observer and individual layer cloud amounts are recorded by automated equipment.

3.1.5 Solar intermittency, variability and fluctuations

Solar power variability, intermittency and fluctuations are terms that are often used interchangeably in research to have the same meaning. All terms, when applied to

power, describe the inherent nature of the solar energy resource to change in power availability, by which the power provided by the sun is not constant or controllable [Sayeef et al. \(2012\)](#). This thesis, however, considers two categories to the terms that are independent of each other. *Variability* is indicative of a steady change in magnitude progressing over time, such as a lull in the wind speed or the seasonal change in atmospheric water vapour concentrations impacting radiation absorption. *Variability* features within solar resource modelling, however, it does not imply a finality that should be used to describe loss of power from clouds, as is often the case in research. The term *fluctuations*, on the other hand, is much more suggestive of the stopping or starting of something. For example: a flickering light bulb, an engine that is cutting out, or the inhibiting of direct irradiance due to the presence of cloud. *Intermittency* is the preferred term to encompass the idea and concept, or as an adjective. A *fluctuation* is the preferred term to represent the physical change, used as a noun.

Conventional fossil based generation or renewable technologies such as biomass, hydro or nuclear do not suffer from variability and intermittency challenges because the power output requirement from these technologies can be planned through resource management and are therefore dispatchable. With variable resources such as wind, and intermittent resources such as solar power, the resource is available only when the natural conditions allow.

It is well established that the solar resource is highly intermittent. The degree of intermittency, however, is mainly dependent on the temporal resolution on which it is observed ([Lave and Kleissl, 2010](#); [Perez et al., 2011](#)). This is because irradiance fluctuations occur at the sub-second level all the way up to slow, monthly changes and annual variation. [Lave et al. \(2012\)](#) quantify the fluctuations over different temporal frequencies and find distinct wavelets that show an increase in probability of a large ramping event with increasing temporal resolution. Ramping events at a higher temporal resolution are more prevalent and frequent when compared with coarser, lower resolution irradiance data, which nullifies ramping events through averaging. [Figure 3.4](#) demonstrates this using four hours of global horizontal incident irradiance (GHI) observations and temporal resolution ranging from 1-sec to 10-mins measured in Leeds, UK. The GHI is averaged to create coarser, lower resolution irradiance and it can clearly be seen that the intermittency decreases with a lower temporal resolution. Typically, the irradiance curve is curved over 24 hours, however, this plot is a 4-hour segment and the scale on the y-axis is insensitive to real changes, the figure is not intended for extracting irradiance magnitudes, merely to demonstrate the presence of ramps and detail at higher temporal resolutions. The dip at midday is representative of a clouded period.

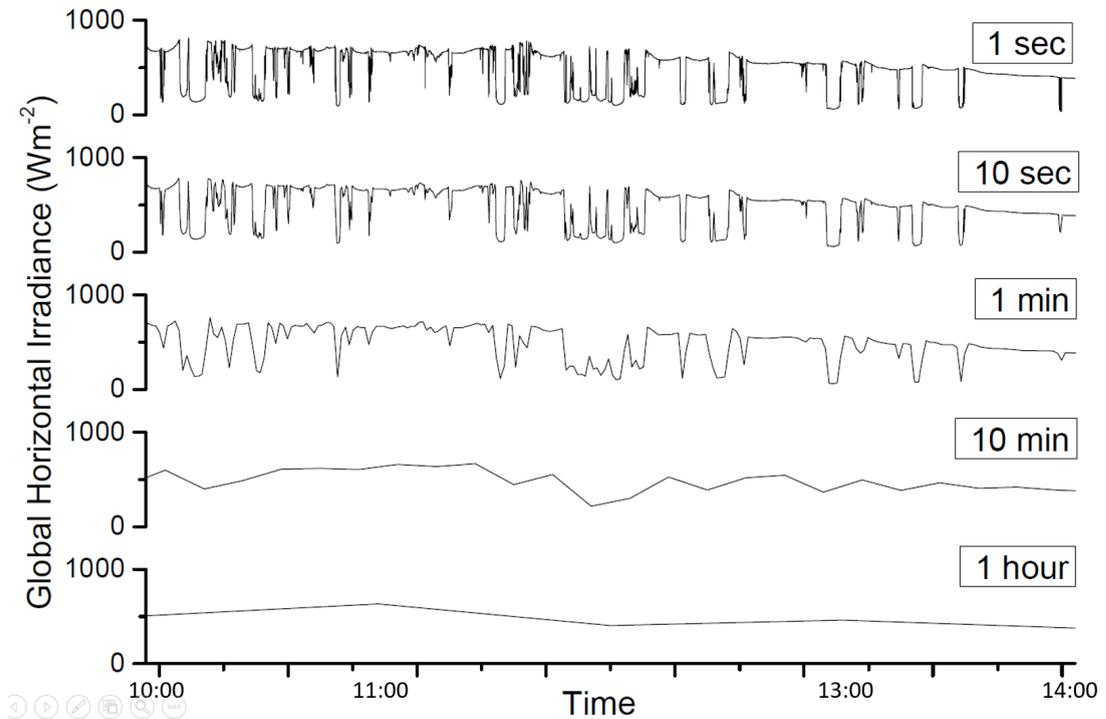


FIGURE 3.4: Global horizontal incident irradiance in Leeds, UK on 25/06/2015 between the hours of 10:00 and 14:00. The data was logged using a horizontally-mounted silicon photodiode (BPW20) in short-circuit current mode with a linear current response converted to voltage with a transconductance amplifier, logged at 10 Hz using a 10-bit DAC. The data has been averaged across timescales of, from top to bottom: 1 second, 10 seconds, 1 minute, 10 minutes and 1 hour.

The power that is available to a power conversion technology, such as a PV inverter, is defined as the incident solar irradiance upon the receiver multiplied by the system’s effective area ($\text{Wm}^{-2} \times \text{m}^2 = \text{W}$) (Kleissl, 2013). Therefore, the power output is directly proportional to the incident irradiance and is subject to its inherent rapid intermittency (Marcos et al., 2011).

Irradiance fluctuations are caused by clouds that pass across the sun, cloud dynamics, atmospheric losses, the transportation of airborne pollutants and more (Suehrcke and McCormick, 1989; Vindel and Polo, 2014a; Calinoiu et al., 2014).

Another example that helps to demonstrate that intermittency is more of an issue than variability is the idea of power inertia between intermittent solar power and variable wind power. Perhaps one of the most significant reasons that intermittency in solar power deliverance is more of an issue than with the other variable technologies is the distinct lack of inertia with which PVs produce energy. Wind power generates power with considerable inertia, meaning that when the resource suddenly drops, the inherent inertia of spinning blades “cushions” the ramp of power loss. When a cloud passes across a PV panel, there is no residual power inertia and so the down ramp of power is instant (Sayeef et al., 2012), it is acknowledged that a large installation spread over a

TABLE 3.2: Relationships between meteorological variables and radiation. Key: SD = sunshine duration, CF = cloud fraction, CT = cloud type, OCF = opaque cloud fraction, DTR = diurnal temperature range, PR = precipitation. Clearness index, kt, and clear-sky index, kc, were defined in section 3.1. Table produced by Dr C. J. Smith.

Study	Derived quantity	Variables
Ångström (1924)	Daily clear-sky index	SD
Prescott (1940)	Monthly clearness index	SD
Wörner (1967)	Hourly clearness index	CF
Kasten and Czeplak (1980)	Hourly clear-sky index	CF, CT
Nielsen et al. (1981)	Hourly net and global radiation	CF
Brinsfield et al. (1984)	Daily global radiation	CF, OCF
Bristow and Campbell (1984)	Daily clearness index	DTR
Hargreaves et al. (1985)	Daily global radiation	DTR
de Jong and Stewart (1993)	Daily clearness index	DTR, PR
Supit and van Kappel (1998)	Daily clearness index	DTR, CF
Muneer et al. (1998)	Hourly beam and diffuse radiation	SD
Matszuko (2012)	10-minutely global radiation	CF, CT

spatial domain would not receive an instantaneous ramp, this is more a function of the spatial dispersion and geographic smoothing, such is the importance of understanding the correlation in the spatial dimension. [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 around the world. They find that there is very little real-world data for use in discovering how solar intermittency affects electricity networks.

It is important that solar intermittency is a well represented component when performing solar resource assessment for integrating PV into the grid. Methods and notable research will be explored in detail in section 3.2.

3.2 Literature review of temporal solar irradiance generators

The generation of synthetic irradiance time series can be broken into two performance categories. The first considers only a temporal dimension (one-dimensional), whilst the second considers temporal and spatial dimensions (two or three-dimensional). Spatio-temporal methodologies are discussed in chapter 4. Whilst all methodologies to generate synthetic irradiance time series include a temporal dimension, the majority do not consider the spatial dimension. This literature review only considers those methodologies.

Synthetic irradiance generation is achieved using assumptions, physics and the inclusion of ground based observations of a variety of meteorological variables. The physics and assumptions are discussed later in this chapter. Table 3.2 lists notable research that have studied the relationship between a meteorological variable and the solar radiation. From as early as 1924, the daily clear-sky index could be determined from the sunshine duration and could then be used to form a time series of mean daily irradiance value for a year. Moving forward from this sees the development of irradiance synthesis opportunities derived from hourly cloud statistics of fraction, type and opacity. The ambient temperature has also been used to determine the daily global irradiance. Hoff and Perez (2010) state that it would be possible to model the variability of a single location based on hourly satellite derived data. This statement is extended within this chapter to be also possible from hourly weather data.

From the grid impact review, a methodology is required that incorporates solar variability and intermittency along the time series at high temporal resolution. Many methodologies exist that do not consider solar variability or intermittency. Badescu et al. (2012) examine the performance of 54 different models that can achieve cloudless sky irradiance time series. These models present the myriad of methodologies that exist with the ability to estimate solar radiation from readily available inputs. They do not, however, produce time series that vary according to state of the sky. In fact all of the models tested failed the stage that explored sensitivity to the state of the sky, for example N . One issue is the large variety of input data that models require. Models exist from a single input of zenith angle, which describes the sun's position in the sky, to nine inputs that include astronomical, geographical, meteorological and turbidity based inputs. The sector of calculating the solar clear-sky radiation is rather saturated and many excellent methodologies exist. However, models examining synthetic production of varying and intermittent solar time series are fewer. These are the models and methodologies of interest.

Polo et al. (2011) produce 10-min resolution synthetic irradiance time series with the inclusion of solar intermittency, derived from hourly mean observations of solar irradiance and adding random fluctuations. This method is suitable for instances with reasonable data availability. Whilst dealing with the issue of adding fluctuations by identifying sky conditions and appropriate k_T distributions, the method fails to offer true synthesis of irradiance without an irradiance input, nor is 10-mins a high enough temporal resolution for the aims of this thesis. The addition of random fluctuations was shown to offer reasonable likeness to real observational data through cumulative probability distribution function (CDF) comparison of the irradiance magnitude of both

modelled and observed data. The daily and monthly means were shown to have excellent reproducibility, indicating that synthesis is viable when increasing the temporal resolution.

Expanding on this work, [Larraneta et al. \(2015\)](#) generate synthetic 10-min resolution DNI data from mean hourly values of sky classification and irradiance using a stochastic driven component and identification of sky classifications as a function of an index expanded from [Skartveit and Olseth \(1992\)](#). Again, this work suffers from insufficient temporal resolution outputs and a requirement of an irradiance time series as an input. The irradiance from [Polo et al. \(2011\)](#) are analysed and separated into four sky classifications, separated by the prevailing k_T values and fluctuations using an assessment method developed by [Perez et al. \(1990\)](#). The use of sky classifications was shown to improve performance. As sky classifications require an irradiance time series to analyse, a similar approach could be employed using N .

An improvement of incorporating solar intermittency and increasing temporal resolution from 1-hour irradiance observation data down to synthetic 1 min irradiance time series is achieved by [Fernández-Peruchena and Gastón \(2016\)](#). The methodology analysed the calculated k_T progression statistics from four 1 min irradiance observations at different latitudes. Daily distributions of k_T were produced and for each hour of observation data, intermittency is applied by implementing a time series of k_T that is synthetically generated as a function of the day and current mean hourly k_T . These stored hourly k_T time series are drawn upon by selecting the current day and the most appropriate hourly k_T time series in accordance to the input hourly GHI irradiance, a 1 minute resolution is therefore achieved. The principal assumption for this methodology to work is that the high-frequency fluctuations caused by clouds are dynamically reproduced from a 2-year long ground measurement of local intermittency, this certainly fails to capture the typical meteorological year (TMY). A TMY in reference to solar irradiance is a time series that captures the statistics of a typical year, such that extreme events do not skew the data. Ten years of time series data is accepted as the standard duration for capturing these annual, seasonal and monthly deviations, and is even discussed by the same authors as a requirement in an earlier paper ([Fernández-Peruchena et al., 2015](#)). The intent of the methodology is that with the statistics of four high frequency GHI time series, many other irradiance time series could be produced. This is known to offer a wide range of mistakes as [Smith et al. \(2017\)](#) demonstrate that the longitude and latitude influence the frequency of cloud fraction present at a site. These cloud fractions have very well defined k_c distributions. Therefore, taking a single location to generate the statistics, and then applying them to a different geographic region can cause inaccuracies. It is important that a methodology

for synthetic irradiance considers the geographic dispersion of meteorology. Furthermore, [Fernández-Peruchena and Gastón \(2016\)](#) require an irradiance input in order to compare magnitudinal similarities with which to fill in the gaps between hours.

There are more methods that attempt to fill the gaps between measurements as opposed to a truly synthetic generation tool with no irradiance input as a guide or foundation. Whilst lessons can be taken from these methodologies, such as the identification of k_T distributions with sky conditions and stochastic drivers of sky classifications, they do not touch upon the aims of this thesis. A common theme of these closely related studies is the identification of sky types or cloud presence. Clouds are one of the largest influences on solar radiation attenuation ([Kasten and Czeplak, 1980](#); [Matszuko, 2012](#)) and so the cloud coverage is a highly useful variable for solar radiation prediction. The analysis of cloud based observations to provide meaningful time series is therefore an interesting and relevant field of study.

To the author's knowledge, the one-to-one relationship between k_c and only N has not been investigated previously. There are cloud fraction relationships that correlate N to a radiation quantity such as [Wörner \(1967\)](#), and employed by [Supit and van Kappel \(1998\)](#), who derive k_t from geographically fitted constants. Later, [Kasten and Czeplak \(1980\)](#) found the relationship between k_c and N using 10 years of data from Hamburg, Germany and was shown to be applicable to the UK by [Muneer et al. \(1998\)](#). Again, this relationship depends on a geographical dependant constant that can be tuned, and so are not one-to-one. [Brinsfield et al. \(1984\)](#) managed to predict solar radiation based on the opaque cloud fraction, an infrequently recorded variable. [Matszuko \(2012\)](#) derived the relationship by N and solar elevation that facilitates the normalising of k_c by $N = 0$. It is pointed out by [Matszuko \(2012\)](#) that cloud opacity has a very diverse range when categorised by N , such that N can equal 8 for 100% sky coverages by clouds, however the type of cloud can cause the opacity to range from 98–100% from cirrus to lows of 7% for nimbostratus. Both cloud types can offer $N = 8$, yet k_c could range from 0.07–0.98. It is for this reason that a one-to-one relationship could be useful. Work presented later in this chapter will detail a methodology from a collaborative piece of research found by the author and [Smith et al. \(2017\)](#) that derives this one-to-one relationship.

An interesting development comes from [Ngoko et al. \(2014\)](#) where 1-min global irradiance time series are synthetically generated from the stochastic progression of k_T . The model described is able to produce truly synthetic time series such that no irradiance data is required as an input once primary analysis of a time series is completed, and so is considered a direct alternative to the SIG. Two high temporal resolution irradiance time series from Tokyo, Japan, are analysed and Markov transition matrices (MTM)

are produced from the progression of calculated k_T , the MTMs are categorised by the daily mean of k_T . Once the MTMs are produced, they can be used stochastically to generate synthetic time series. There are significant drawbacks, however. The most apparent drawback is the validation method compares against the same data analysed for k_T transition probabilities to validate the results. This removes the ability to test whether their k_T transition probabilities can be translated to other regions. What is evident from their validations is that their statistical performance lies very close to the mean of the performances of both validation/input locations. The implication is that the methodology would not be robust for regions that do not have an existing 1-min irradiance input, and therefore it does not satisfy the requirements set out by this thesis of high geographic dispersion. The idea of making MTMs of k_T from larger category bins of the mean daily k_t is one well worth developing and the methodology does demonstrate appropriate recreation of k_T using Markov chains. [Ngoko et al. \(2014\)](#) also demonstrates the suitability of using first order Markov chains when used with irradiance time series generation. This can also be adopted in the SIG.

The sky type is a variable used first by [Perez et al. \(1993\)](#), who derived them using high-quality sky-scan data, and then later by [Torres et al. \(2014\)](#) to categorise the sky and its cloud based conditions into 15 standard categories. These categories are analysed for transition probabilities in order to facilitate the use of first order Markov chains for synthetic and stochastic sky type time series progression. The interesting component of this work is the successful recreation of the sky types validated against two independent locations and the identification of distinct seasonal sky types. Whilst the input data is certainly not easily obtained, as sky type nor historic sky-scan data are often reported, the seasonal dependent MTMs offered an excellent validation. The implication here is that the sky types, when broken by season, are statistically and stochastically reproducible, as well as distinctly different. A seasonal dependency should be explored.

[Pérez-Burgos et al. \(2015\)](#) undertook analysis of k_c (labelled as the cloud modification factor in their study) as a mean hourly diurnal average for each month. This means that the k_c is shown to be statistically predictable by day, and by hour of the day, when taken as a mean. The tool is intended for long term, synthetic time series production and is shown to validate well with long term reproductions. Whilst this research does not touch upon variability, intermittency nor high temporal resolution, it does interestingly show that k_c has a strong diurnal dependency. It is hypothesised within this project that geographically dependent cloud dynamics during different times of the day can be highly influential. The idea of diurnal dependency could be considered in the stochastic production of MTMs.

A highly relevant research paper that satisfies the rationale of this thesis was produced by [Ehnberg and Bollen \(2005\)](#), who also did not wish to use irradiance data as an input into their research recognising the often limited access to it. They instead proposed the use of N , which can be measured without equipment and by eye for intervals of 1 to 3 hours. It is stated that “by assuming a deterministic relationship between cloud coverage and hourly global solar radiation, the need for measurement of the latter disappears.” The model by [Ehnberg and Bollen \(2005\)](#) stochastically and synthetically generates an irradiance time series from a first order Markov chain process with MTMs derived from time series of N . The output is 6 min time series of irradiance with intended application in grid impact analysis. The authors note yearly, seasonal, daily as well as random patterns in the analysed GHI and used standardised clear sky irradiance calculations and coupled them with distributions of GHI at each value of okta, this is further explained by [Gueymard and Wilcox \(2011\)](#) who state that the GHI accumulated during a year is practically independent of GHI of a preceding or subsequent year. The methodology by [Ehnberg and Bollen \(2005\)](#) is the beginning of an excellent line of research. Firstly, it confirms the success of a first order Markov chain approach to modelling N . Secondly, linking the GHI by probability density to each value of N allows for synthetic and stochastic production of irradiance fluctuations. The work is not properly validated, and so without recreating this work, it is difficult to comment on its geographic limitation. The authors do suggest that it is applicable to anywhere within the same climatic region, although this is speculative and presents an opportunity for further research. 6 min resolution is also coarse for certain grid impact analysis.

Moving away from k_c , k_T and GHI based deterministic models, it is possible to use a completely different theory termed a *sun obscured* methodology. A sun obscured based model functions such that, at each time step, it is determined whether there is a cloud blocking out the DNI or not. Previous mentions of models have typically required an irradiance input as guidance, or have added random fluctuations to recreate solar intermittency with no real guide from the most significant attenuator of solar irradiance — clouds. The rationale of this thesis is to be able to produce irradiance time series without the need of a prior irradiance input, similar to the early attempt by [Ehnberg and Bollen \(2005\)](#). [Gafurov et al. \(2015\)](#) reiterates that the need for regional irradiance comparisons within a synthetic irradiance model defeats the purpose of synthetic irradiance modelling, as data must have already existed. [Gueymard and Myers \(2009\)](#) state that relying on irradiance data can be a mistake due to the large systematic errors involved in the recording method. They encourage less dependence on datasets when deriving relationships and statistics. A sun obscured model is a type less explored in the literature and removes dependence on solar irradiance time series as an input.

Morf (1998) was the first to produce a sun obscured type method by stochastically deriving a time series of clear and cloudy periods before using a simple procedure to estimate the irradiance into beam and diffuse subcomponents. Figure 3.5 demonstrates the original sun obscured methodology by Morf (1998). The duration for periods of sun obscured are driven by the negative natural log of a random variate evenly distributed between 0 and 1, R , multiplied by the mean duration of a clouded period, which were determined through analysis of sample observation data; the same method is employed for clear periods. This treatment of clear periods as the same as clouded periods could prove useful for mathematically representing the sky. A first order Markov chain process is employed to drive the MTMs of the clear and cloudy periods. Using conventional clear sky irradiance calculations and separating the irradiance into beam and direct components, a synthetic irradiance time series was generated without the need for irradiance data inputs. This work was extended by Morf (2011) with the conception of quantifying the cloud amount across a spatial domain and analysing the application of a stochastic process using N . The study and validation of the stochastic N time series demonstrated the success of first order Markov chain ability at reproducing N .

A particularly promising finding from the work by Morf (1998) and Morf (2011) was that the steady state probability distribution of N is invariant to the size of the observed area. This opens up the possibility to synthetically map the sky, knowing that over the hour, an element in a spatial domain will converge towards the value of measured N . Neither research papers by Morf (1998, 2011) considered irradiance generation with fluctuations caused by cloud or other variability inducing elements. A further development by Morf (2013) employed a two step process to include irradiance fluctuations. Firstly, the strict on/off shape of the sun obscured methodology (figure 3.5 bottom) was smoothed before feeding the smoothed signal into a window filter, which uses the moving average of the input signal. This approach to fluctuations does not consider observable phenomenon of irradiance fluctuations such as cloud edge enhancements and smoothing as a function of cloud cover, and so there is an opportunity to implement or develop a better tool for modelling the solar variability using a sun obscured methodology.

3.2.1 Summary of identified research opportunities

The literature reviewed here can be summarised into the following research opportunities that facilitate temporal synthetic solar irradiance generation:

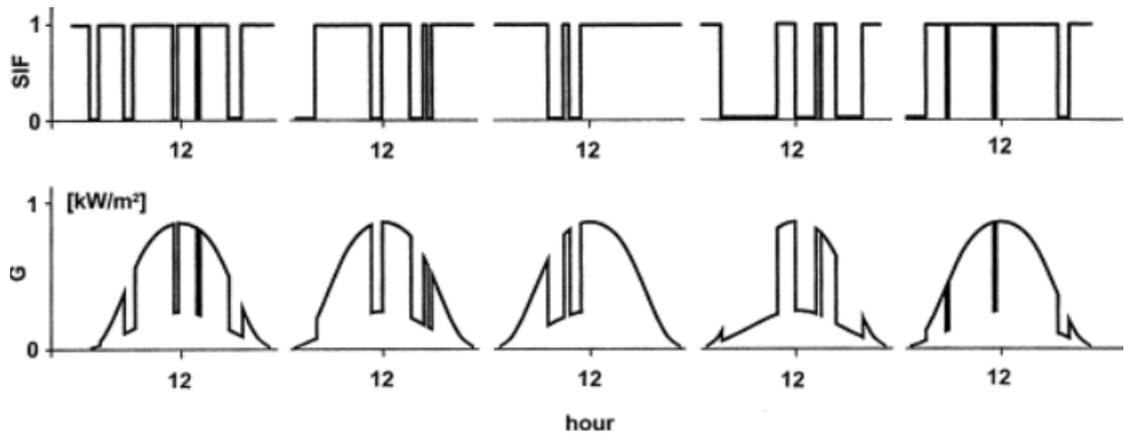


FIGURE 3.5: Demonstration of the sun obscured methodology derived by Morf (1998). Top) the stochastic insolation function over time, where 1 and 0 indicate clear and cloudy periods, respectively. Bottom) the corresponding irradiance output as guided by the stochastic insolation function.

- The application of random number generation to statistically accurate irradiance fluctuations was shown to be a useful method of filling the gaps between hours. Using statistics that do not relate directly to an irradiance time series, such as N , in combination with random number generation could aid in synthetically representing irradiance fluctuations.
- Similarly to the use of sky classifications dictating the distribution of k_T , the use of N in place of sky classifications could be used to determine the distribution of clear sky or clearness indices.
- The capturing of a TMY for well balanced statistical representation of meteorological observation variables is imperative for accurate generation of short and long time scale simulations.
- Stochastically driven models using first order Markov chains are highly appropriate tools for variables such as N , temperature, k_T and k_c . The utilisation of first order Markov chains is proven to be an effective synthetic generation technique.
- Creating density functions of k_T is popular. As discussed in section 3.1, k_c is a more appropriate description of the clouds. There is scope for classifying k_c by different sky classifications or other meteorological variables and producing density functions for stochastic use.
- Sky type that is separated by season can be reproduced synthetically and stochastically using Markov chains. This presents an opportunity for using readily available cloud amount data and performing stochastic analysis for the recreation of seasonally accurate sky conditions. Furthermore, the diurnal predictability of k_c suggests that the inclusion of a diurnal dependency within a model would be a

useful inclusion. As k_c describes solar attenuation from clouds, it suggests that N could also be represented diurnally.

- The development of a model that does not require an irradiance input at point of use so that the methodology can be considered truly synthetic.
- The inclusion of solar variability over yearly, seasonal, monthly and daily time scales could help maintain irradiance fluctuation validity at increasing time scale analysis.
- A sun obscured methodology has demonstrated the most flexibility for synthetic solar irradiance generation.

3.3 Methodology of the Solar Irradiance Generator

The SIG presented in this chapter has been separated into 7 distinct sections. They are listed below with a brief description of what can be expected in each.

- [3.3.1](#) Input data requirements
 - A discussion of the observational data requirements for the SIG and preprocessing for use. A list of the user defined inputs that allow flexible model utilisation.
- [3.3.2](#) Cloud sample production
 - Definition of a cloud sample and explanation of the calculations and processes involved with their production.
- [3.3.3](#) Constructing the Markov transition matrices
 - Description of the fundamental mathematics, method of separating by season, diurnally, and pressure state, creation of the MTMs and their implementation within the SIG.
- [3.3.4](#) Stochastic generation of weather variable time series
 - Detailing of the process of stochastic weather generation to produce time series of N , pressure, cloud height and wind speed. This is sectioned into the preliminary initialisation settings, and by each variable being stochastically generated.
- [3.3.5](#) Sun-Earth geometry

- Method of calculating the solar position relative to a point source on the Earth’s surface. Calculations of solar angles and time required for irradiance calculations.
- 3.3.6 Generation of clear-sky indices
 - Breakdown of the method to generate a time series of the clear-sky index. Detailing the methodology that considers the relationship between N and k_c , adding fluctuations, smoothing during periods of stability, variability and intermittency.
- 3.3.7 Irradiance calculations
 - Review of appropriate irradiance calculations and their implementation within the SIG. Detailing the application of the clear-sky indices, the separation of GHI into its direct and diffuse subcomponents. Translation from GHI onto an arbitrary plane.

For better visualisation of the SIG, a flow chart is produced and shown in figure 3.6. The green parallelograms are indicative of inputs into the SIG. Whilst there appear to be many, only the top left “*Observation Data*” requires a data source. Each of the blue rectangles represents a distinct section of script found within the SIG. These roughly correspond to the sections within this chapter with the exception of the irradiance calculations, this section is separated in the flow diagram into *theoretical clear sky irradiance* process and the *Global horizontal irradiance and tilt onto arbitrary plane* process, although they are discussed together within this chapter. The output irradiance time series are represented by the yellow parallelogram.

All data processing was performed using the commercial software package Matlab r2015a (Matlab, 2015). The raw computer script for the SIG is appended in section A.

An example of the SIG output and a brief representation of the stochastic sun obscured method is demonstrated in figure 3.7. Figure 3.7a shows the stochastically generated N in okta for corresponding hours. N transitioned in 20 of the 24 time steps and so would be considered an unstable day with high scattering of cloud. This is further demonstrated by the figure 3.7b, which presents the one-dimensional sun obscured representation of cloud cover, where black indicates the presence of a cloud, white representing clear sky. The ratio of clear to cloudy within each hour of the simulation is equal to the corresponding okta number in the figure 3.7a. Figure 3.7c shows the resultant tilted irradiance time series for the simulated day typical of June in Leeds, UK. Periods of intermittency are directly proportional to presence of cloud, as is expected in a sun obscured methodology. Furthermore, irradiance variability can be seen

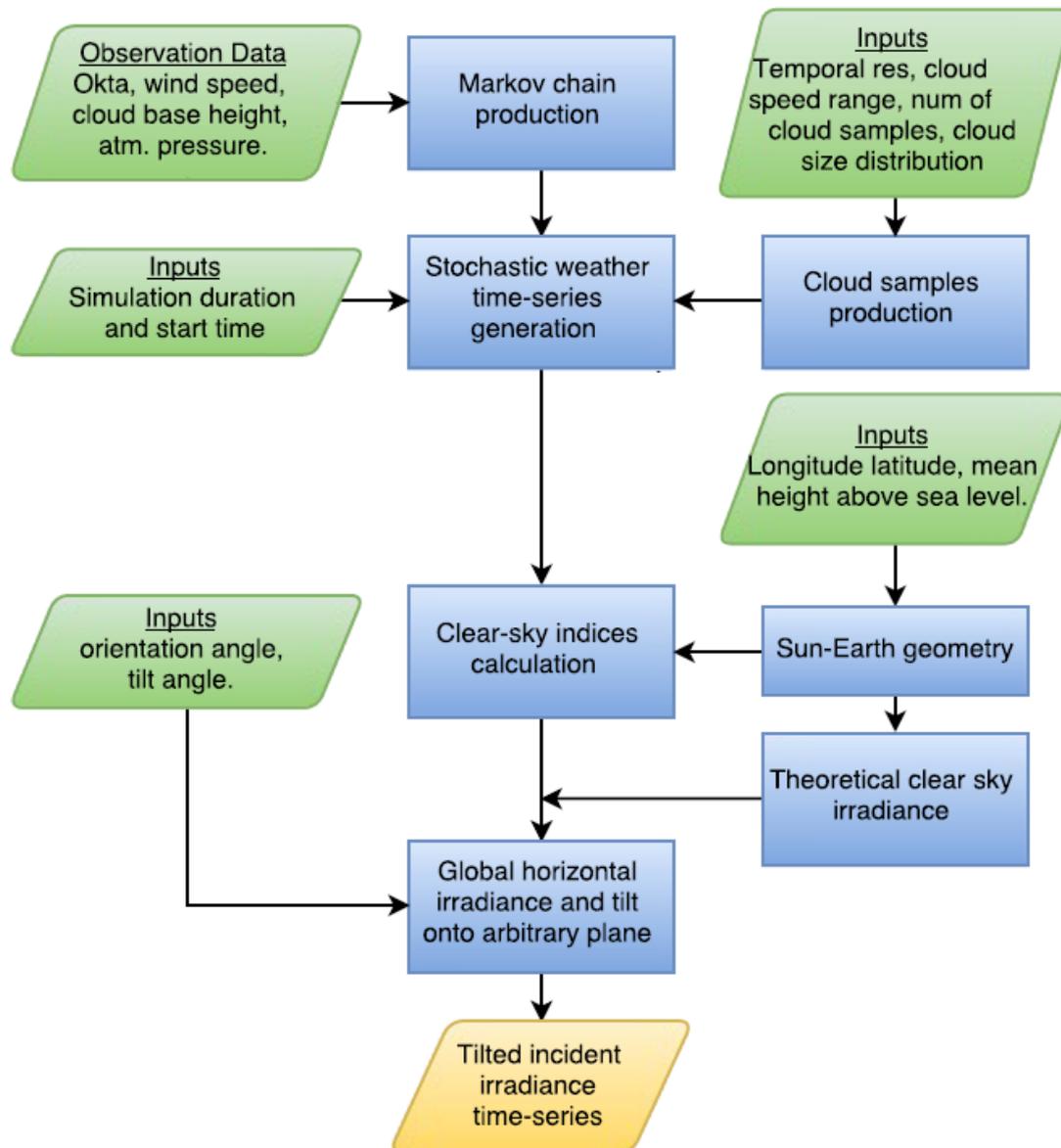


FIGURE 3.6: Flow chart summarising processes within the SIG. Green parallelograms represent an input into the model, blue rectangles represent a process, and the yellow parallelogram at the bottom indicates the synthetic irradiance output times series.

during the intermittent periods, this is most noticeable at intermittent periods at 4pm and 7pm.

3.3.1 Input data requirements

This section is separated into two categories: input observation data and user defined inputs. The input observation data is the raw data used to generate the MTMs, while the user defined inputs are for the SIG simulation parametrisation to enable the most appropriate output irradiance time series.

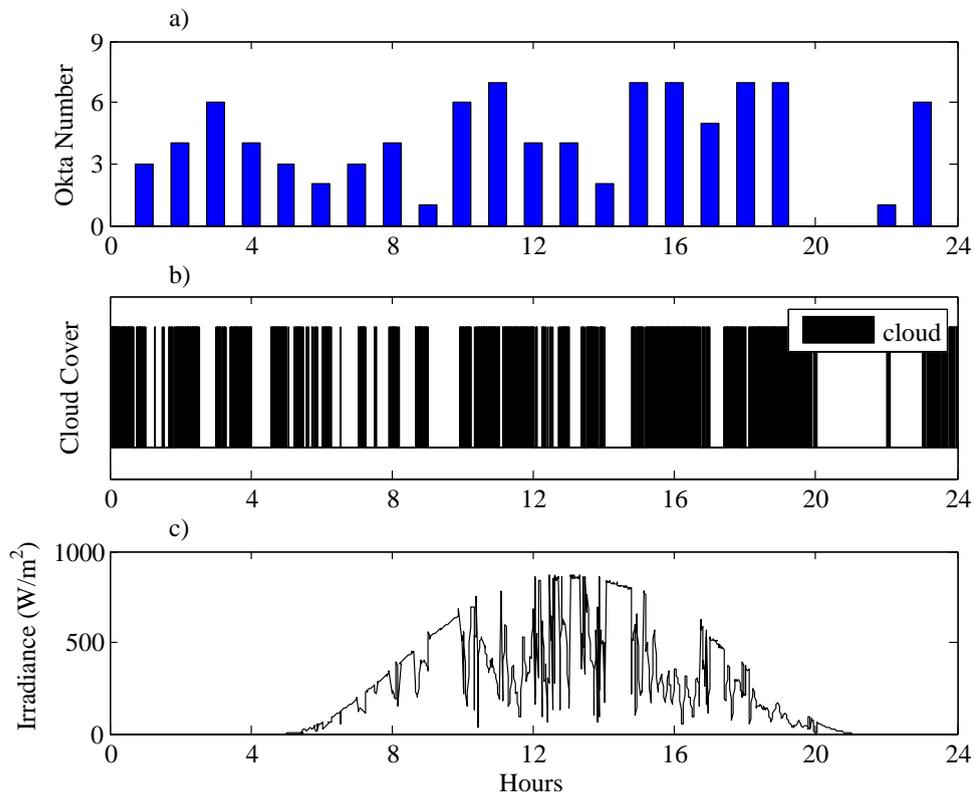


FIGURE 3.7: An example output of the SIG for a typical day in June for Cambourne, UK, that visually demonstrates the fundamental working of the SIG. Plot a) shows the mean okta number stochastically derived at each hour. Plot b) shows the corresponding cloud cover time series in a minutely resolution, where black is indicative of cloud present and white indicating periods of clear sky. Plot c) shows the resultant tilted irradiance, although this simulation was performed on a plane horizontally aligned for Cambourne, UK.

3.3.1.1 Input observation data

The raw observation data inputs required for the SIG are mean hourly weather observations of the following list. Note that the choice of unit is selected as they are the unaltered format of the UKMO:

- mean sea level pressure, p (hPa)
- wind speed, u_{10} (recorded at 10m above surface, knots)
- cloud base height, z (decametres)
- total cloud amount, N (okta)

The listed units are the raw format used in this model. Conversions are used within the SIG from hPa to mbar, knots to ms^{-1} ($0.515 \text{ ms}^{-1}/\text{knot}$ as defined by Met Office), decametres are reduced to metres and okta are converted from N (eighths) to C (tenths).

In order to capture accurately the probability statistics of a TMY, a minimum of 10 years of observational data are recommended (Fernández-Peruchena et al., 2015). It is common that faults and routine maintenance interrupts constant recording of variables; gaps can range from hours to months. For this reason, 12 years of observation data were taken to ensure analysis of at least 10 worth years of weather transitions.

As was discussed in section 1.2 and 3.1.4, the UKMO (2010) maintain long term meteorological data sets of weather observations in the Met Office Integrated Data Archive System (MIDAS) data set (MIDAS, 2015) and is operated by the British Atmospheric Data Centre (BADC, 2013). Within the data sets, the *UK Hourly Weather data* exists that records 104 weather parameters. These data sets are available with excellent geographic dispersion around the UK as was shown in figure 1.2.

The hourly weather data is extracted from the data base all sites by year in text format documents. Each observation station has an individual identification number that can be used to extract the relevant data for the location of study. Data undergoes various quality controls at observation and can result in duplicates of certain hours. Where this occurs, the observation with the most recent time stamp is selected, as is suggested by the Met Office.

The steps for data preparation are as follows:

1. Import data assigning a known value for missing measurements (-9999 was used)
2. Filter required data by using unique station identifier
3. Order chronologically
4. Remove duplicates by selecting the most recent time stamped observation
5. Commit to file for use in the SIG

3.3.1.2 User defined variables

The SIG is built and designed with consideration for potential future development and so has some flexibility with user input. The SIG is robust enough to handle variants of all user defined inputs. The script for the user defined variables is found in appendix section A.2. The user defined inputs are as follows:

- The simulation start day, $d_{t=1}$
 - The day number, where 1 is equal to 1st January up to 366 equalling the 31st December. If the year is not a leap year, 366 will become 1st January of the following year. Numbers outside of these ranges will become the 1st of the same or following year, accordingly.
- The simulation start year, $y_{t=1}$
 - Any year is theoretically achievable, although is limited by the capability of the Sun-Earth geometry methodology selected. Years tested were from 2001-2016.
- The duration of the simulation, $T_{t(1,2,\dots,T)}$
 - The input is required in years. Selecting a single day is possible by inputting 1/365. The minimum selection is one day.
- Longitude, Φ , and latitude, Θ
 - Set the location of the study in degrees. Six significant figures were used in the study. Φ and Θ are provided for each weather station by MIDAS (2015), although can be determined using free online mapping applications.
- Height above sea level, z_l
 - Set z_l for the location of study in metres. Provided at each weather station by MIDAS (2015), although can be determined using free online mapping applications.
- Panel pitch (tilt), β_p
 - The pitch is required in degrees measured from horizontal, 0° , to vertical at 90° . The model will not accept values outside of this range and assign outliers to the minimum or maximum accordingly.
- Panel azimuth (aspect or orientation), β_a
 - The azimuth of the panel is required in degrees measured with 0° as south with angle moving clockwise. Values outside the range of 0° to 359° are assumed to obey circular rotations. E.g. 360° would again represent south and $-90^\circ = 270^\circ$ would represent east.
- Range of wind speeds u_{range}

- The range of wind speeds available within the input data in ms^{-1} . The range could be altered to include higher speeds should a certain site use different input data than [MIDAS \(2015\)](#).
- Number of cloud samples, n_s
 - n_s number of cloud samples that are produced per value of cloud speed, u_{ref} , and per value of C .

3.3.2 Cloud sample production

A fundamental part of this study is to produce statistically accurate time series using a sun obscured methodology. The sun obscured element is achieved through the modelling of the sky as a series of clouds that will indicate whether or not the sun is covered. To do this, the clouds must be synthetically represented.

A terminology is introduced here for simplicity throughout methodologies. As the SIG is computationally produced, the variables are structured in either one-dimensional or two dimensional computational matrices. These will be denoted as 1DM or 2DM.

A *cloud sample* is a term conceived by the author. It is a Boolean 1DM that represents the sky for one hour. A Boolean 1DM contains only 1's or 0's for use in logical operation. The resolution modelled in the SIG is 1 min and so a cloud sample will have 60 elements within the 1DM, each containing a 1 or 0 representing obscured or not obscured by cloud at each time step for the hour, respectively. An example was demonstrated in figure 3.7b.

An interesting approach comes from [Evans and Wiscombe \(2004\)](#) who stochastically generate “cloud fields” from radar profile statistics. Using these statistics and their novel algorithm, a synthetic and statistically relevant two-dimensional representations of the clouds in the sky can be produced. The key issue is that radar input data is required in order to validate and produce the cloud fields, it is also an additional dimension than is not required for temporal only modelling. Discussion of two dimensional cloud representations is made in section 4.2.2.

To the author's knowledge, there are no other published methods to produce a cloud sample. Other sun obscured methodologies have used statistical probability of a change from an obscured to not obscured moment ([Morf, 2011](#)), however did not need to produce a synthetic representation. A novel approach is presented that generates synthetic 1DM cloud samples.

In the SIG, two variable ranges are previously defined that can be used for cloud sample production: the hourly mean cloud amount in tenths, C ($\frac{0}{10}$ to $\frac{10}{10}$) and the cloud speed, u_{10} (0 to 30 ms^{-1}). A temporal methodology requires only a single passage of beam irradiance from sun to the solar panel, therefore a representation of the whole sky is not required and a one-dimensional approach is used.

The typical method for measuring N is to use a cloud base recorder, which uses a vertical laser pulse to track cloud presence (UKMO, 2010). This measurement device takes readings in a single direction and so the okta number derived is considered an acceptable representation of an hour's ratio between obscured and clear moments directly above measurement, although perhaps not as accurate a measurement of the entire observable sky Smith et al. (2017). The implication for synthetic generation is that a one-dimensional format is well suited from okta measurement techniques and that the final obscured to not obscured ratio within the cloud sample must be directly comparable to the targeted coverage value.

The other variable considered to influence the synthetic generation of cloud samples is the cloud speed, u_{ref} . Solar intermittency was shown to be a function of u_{ref} by Arias-Castro et al. (2014). It is intuitive that the faster a cloud is travelling, the quicker it will pass by a point on the surface and no longer obstruct the direct irradiance. Should u_{ref} be known, the other influencing factor would be the size of the cloud and the number of clouds in the sky.

Three fundamental statements can therefore be used to synthetically generate a cloud sample:

1. Mean hourly N is an appropriate representation of the ratio of obscured to not obscured minutes within a cloud sample.
2. With a fixed u_{ref} over an hour, intermittency becomes a function of the cloud size and number of clouds.
3. Theoretically, a 1DM representation of a period of clear sky “moving” across a point on the surface could be punctuated by a period of cloud defined by u_{ref} and cloud size.

All values of u_{ref} and C can be stochastically selected within the SIG, and so n_s number of cloud samples for all combinations of both u_{ref} and C will be generated. The final components required are the cloud size, x , and number of clouds, n_c . The cloud samples will be generated by following a distribution of x for a single linear dimension in a Boolean 1DM format.

Horizontal x distributions are shown to be well-represented using a single power-law relationship using an exponent of $\beta < 2$ (Wood and Field, 2011; Stull, 1988; Leahy et al., 2012; Pressel and Collins, 2012; Ray and Engelhardt, 1992). The single power-law is shown in equation 3.3.

$$P(x) = \alpha x^{-\beta} \quad (3.3)$$

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

This distribution by Wood and Field (2011) was produced using two separate imaging techniques from the MODIS satellite and from aircraft observations. Their model that proposes equation 3.3 was validated against both sets of data. The use of $\beta < 2$ allows for variation in geography and season, the closer to 2, the wider the range of reproduced cloud lengths making.

Equation 3.3 represents the probability of x without size limits. By rearranging equation 3.3, it is possible to introduce these size limits. This will facilitate its utilisation for pseudo-random number extraction from the single power-law distribution. Equations 3.4 to 3.7 demonstrate how equation 3.3 can be modified for pseudo-random number extraction.

$$x = (\kappa + \lambda R)^{\frac{1}{1-\beta}} \quad (3.4)$$

$$R \sim U(0, 1) \quad (3.5)$$

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

$$\kappa = x_{\max}^{1-\beta} \quad (3.6)$$

and

$$\lambda = x_{\min}^{1-\beta} - \kappa \quad (3.7)$$

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

These equations allow a cloud size to be extracted that follow the distribution outlined by Wood and Field (2011).

The proposed cloud sample methodology is to make a single, very long, 1DM “cloud population” distance series. Cloud population is a term used to describe a very long stretch of cloudy and clear periods that can be used to extract cloud samples from. Firstly, the cloud population is created with of zeros representing a stretch of clear

sky, denoted \mathbf{x} , then populate it with alternating lengths of obscured and not obscured with length pseudo-randomly extracted from the modified [Wood and Field \(2011\)](#) relationship. Each element in the cloud population has a resolution of 10 meters (the smallest resolution defined by x_{\min}). \mathbf{x} is converted into units of time with a resolution of 1-min. This is achieved through the decimation (or resampling) of \mathbf{x} by a sampling rate, ψ , that is a function of u_{ref} . The resultant 1DM cloud population time series contains a binary representation of each minute's coverage condition for an overall maximum duration that is computationally achievable — the 1DM cloud population time series of cloud cover indication is denoted \mathbf{x}_{ψ} . A cloud sample can be randomly extracted from any 60 min stretch within \mathbf{x}_{ψ} , it can then be analysed to determine C . The cloud sample can then be binned by the appropriate u_{ref} and C , for later extraction during the stochastic weather generation phase.

The described methodology for cloud sample production can be broken down into the following steps:

1. Define the length of \mathbf{x} , u_{range} , range of C and n_s
2. Populate \mathbf{x} with cloudy and clear periods using distributions of x
3. Convert the distance series \mathbf{x} to time series \mathbf{x}_{ψ} by resampling by ψ
4. Randomly select a 60 min cloud sample from \mathbf{x}_{ψ}
5. Store the cloud sample in the appropriate bin indexed by u_{ref} and C
6. Repeat process until each bin (all combinations of u_{ref} and C) has n_s samples

3.3.2.1 Producing a one-dimensional-matrix distance series of cloud cover

Computationally looping through \mathbf{x} , a cloud is added with its x determined from equation 3.4. Immediately following the cloud, a gap of clear sky is synthetically added with clear horizontal distance drawn from the same distribution. This assumption is valid as the \mathbf{x} is not intended to be an accurate representation of sky, only a vessel to hold potentially realistic options of which a clouded sky might manifest and is an important tool for sun obscured methodologies ([Morf, 2011](#)). Furthermore, a gap is required, as having two clouds back-to-back would alter the reproduction of the [Wood and Field \(2011\)](#) distribution of x . Once \mathbf{x} is populated, it is ready for resampling by ψ .

3.3.2.2 Producing a one-dimensional-matrix time series of cloud cover

A signal poly-phase filtering technique (implemented using the built in Matlab “re-sample” function (Oppenheim et al., 1999; Matlab, 2012)) is applied that decimates \mathbf{x} by a sampling rate, ψ . The result is the Boolean 1DM time series, \mathbf{x}_ψ . For every u_{ref} , \mathbf{x} is resampled to ensure that the time resolution, t_{res} is kept at 1 min.

The clouds are assumed to be travelling at a constant u_{ref} for each hour of simulation. ψ can therefore be a function of u_{ref} and so the calculation of ψ used in the signal poly-phase filtering technique is given by

$$\psi = \frac{x_{\text{res}}}{u \times t_{\text{res}}} \quad (3.8)$$

Where x_{res} is the spatial resolution of \mathbf{x} . For use in this model $x_{\text{res}} = 10$ m and $t_{\text{res}} = 60$ s, and so equation 3.8 can be expressed as $\psi = (6u)^{-1}$.

3.3.2.3 Random extraction of a cloud sample

Hour long cloud samples can now be randomly extracted from \mathbf{x}_ψ . C of the cloud sample is then determined through equation 3.9.

$$C = \overline{\left(\sum_{i=n_{\text{els}} \times r}^{i+59} \mathbf{x}_\psi \right)} / 60 \quad (3.9)$$

Where the overline represents the mean, n_{els} is the total number of elements in \mathbf{x}_ψ , i is therefore a random start point along the length of \mathbf{x}_ψ .

Cloud samples are continuously extracted from \mathbf{x}_ψ until there is a cloud sample database consisting of n_s samples for each possible u_{ref} at each C .

The script that details the cloud sample production method is shown in section A.1.

3.3.3 The Markov chain process and Markov transition matrices

To implement a stochastic element to the SIG, a Markov chain process is used. Markov models are a very popular method of stochastic data generation and have been used in many applications, from wind estimates (Masseran, 2015), solar energy estimations (Bhardwaj et al., 2013; Vindel and Polo, 2014b; Hocaoglu, 2011), and in weather variable generation (Yang et al., 2011; Ehnberg and Bollen, 2005; Morf, 2011).

The script for construction and implementation of Markov chains can be seen in appendix sections [A.5](#) and [A.6](#).

An MTM contains the probability of transition from any of the states to all other possible states. They are therefore always square, examples are shown later within this section.

3.3.3.1 Mathematical background of Markov chains

A Markovian process is a probabilistic mathematical method whereby transitions from one state to the next are directed by discrete probabilities taken from the statistics of real-world processes; the background was covered in section [3.1.3](#). In the case of the SIG, statistics are developed from real observational transitions of N , u_{10} and z . From these statistics, MTMs can be constructed.

The SIG uses a single order Markov chain process whereby only one previous time-step, $t - 1$, influences the transition of states from $t - 1$ to t . Higher order transitions exist beyond the first order, $t - 1$, through to the n^{th} order, $t - n$, however [Ehnberg and Bollen \(2005\)](#) successfully reproduced N with a first order Markov process and so the same is used for the SIG. [Ngoko et al. \(2014\)](#) describes how a Markov process $(M_t, t = 0, 1, 2, \dots, T)$ that has s allowable states $(1, 2, \dots, s)$ is in state 1 at time t if $M_t = 1$. Note that t is a relative term indicating the current time step. The simulation runs to duration T from $t(1, 2, \dots, T)$ and so $t - 1$ denotes the previous time step regardless of the time step.

In this first order Markov process, allow the previous time step (M_{t-1}) to be in state 1, and the next time step (M_t) to be in state 2. This illustrates that, $M_{t-1} = 1$ has undergone a transition to $M_t = 2$. This transition is denoted $(M_{t-1} = 1 | M_t = 2)$. The chance of this transition occurring is given by discrete probability P_{12} . It is calculated by counting the observable transitions of $(M_{t-1} = 1 | M_t = 2)$ and dividing it by the total number of times that $M_{t-1} = 1$ has undergone any transition to any state $(M_{t-1} = 1 | M_t = s)$. P_{12} can therefore be expressed as

$$P_{12} = \frac{\sum(M_{t-1} = 1 | M_t = 2)}{\sum(M_{t-1} = 1 | M_t = s)} \quad (3.10)$$

These probabilities are stored in a first order MTM, \mathbf{P}^1 , and can be represented as

TABLE 3.3: Indication of how the MTMs for each weather variable are separated accounting for seasonal, above and below average pressure and diurnal differences.

Weather variable	By season	By pressure	Diurnally	Number of <i>MTMs</i>
<i>Cloud amount</i>	✓	✓	✓	12
<i>Cloud height</i>	✓			4
<i>Wind speed</i>	✓			4

$$\mathbf{P}^1 = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1s} \\ P_{21} & P_{22} & \dots & P_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s1} & P_{s2} & \dots & P_{ss} \end{bmatrix} \quad (3.11)$$

Note that the sum of each row is equal to 1 as it contains all the possible state transition outcomes, each row is therefore a probability density function (PDF). Columns cannot be compared in this way as they are independent of each other.

3.3.3.2 Constructing the Markov transition matrices

Constructing an MTM requires the conversion of a variable's range of magnitudes into discrete states. For N , each okta number $(0, 1, \dots, 9)$ is a unique state. Variables of z and u_{10} are rounded to the nearest integer for binning into discrete states.

The number of MTMs must be defined, this is because variables are to be separated by season, diurnally and for above and below average pressure systems. These are indicated in table 3.3. When a variable is separated by season, four MTMs will be produced. N is separated further by above and below average p , doubling the number of MTMs required as each pressure group is then divided by season. An additional four N MTMs are produced for the morning of each season, defined as before 6am. Therefore, there are 20 different MTMs to be produced. These MTMs are filled with the transition probabilities using the previously described mathematical method.

Firstly, the mean of all p observations is calculated, \bar{p} , and used as a marker that defines whether or not an observation is classified as above or below average pressure. High and low pressure systems within the UK offer very diverse and often distinct weather patterns. High pressure systems often bring stability and clear skies, this is due to the sinking air that lowers the tendency for water vapour to form into clouds on account of the temperature and humidity of transported air. In low pressure systems the reverse occurs, resulting in increased cloud formation and often bringing broken and scattered rain clouds and infrequent clear-sky (MetOffice, 2015). The hypothesis is that, by splitting the observations by above and below average pressures, these

pressure dependent weather tendencies may be captured. It is acknowledged that a high or low weather pressure system is relative to the pressure systems and geography that surround them, however, using \bar{p} is a very simple mechanism to achieve consistent differences in the MTM probabilities.

Season indicators are determined. This is done by assigning respective values of 1–4 representing spring through winter. Each season offers unique weather, particularly in temperate climates that have distinct seasons. By separating the weather variables by season, it is hypothesised that the inherent seasonal weather will be statistically captured within each MTM.

The final indicator to be produced is that of a diurnal dependency. There is an increased tendency for higher N states during the morning as the introduction of the sun's heat destabilises the atmosphere, and so the build up to sunrise is appropriately modelled. <6am is considered the cut-off because it is a typical sunrise in the summer for the applied study locations (around $\Theta = 50^\circ$). 5 hours allows 5 okta transitions and is considered an appropriate duration for the slight propensity to shift towards increased okta to manifest.

The steps taken to produce the 20 MTMs are summarised as:

1. Determine the mean sea level pressure, \bar{p} , of all observed p data and assign above or below average pressure indicators to each observation
2. Determine the season using the time stamps and assign seasonal indicators to each observation
3. Determine observations that occur during the morning (between 00:00–06:00) and assign indicators (1–4) to each observation
4. Loop through entire length of observation data of all variables (N , z and u_{ref})
 - Select the appropriate MTM using the three types of indicators (considerations shown in table 3.3)
 - Keep a tally in the MTM at appropriate element (as demonstrated in section 3.3.3.1)
5. Convert tallies into probability density functions, and therefore MTMs, as per equation 3.10

Once these steps are complete, the MTMs holds rows of the PDF of each state transitioning to all other states. An example of a fully populated MTM of variable N is shown in figure 3.8 where the probability of transition is represented as a colour with

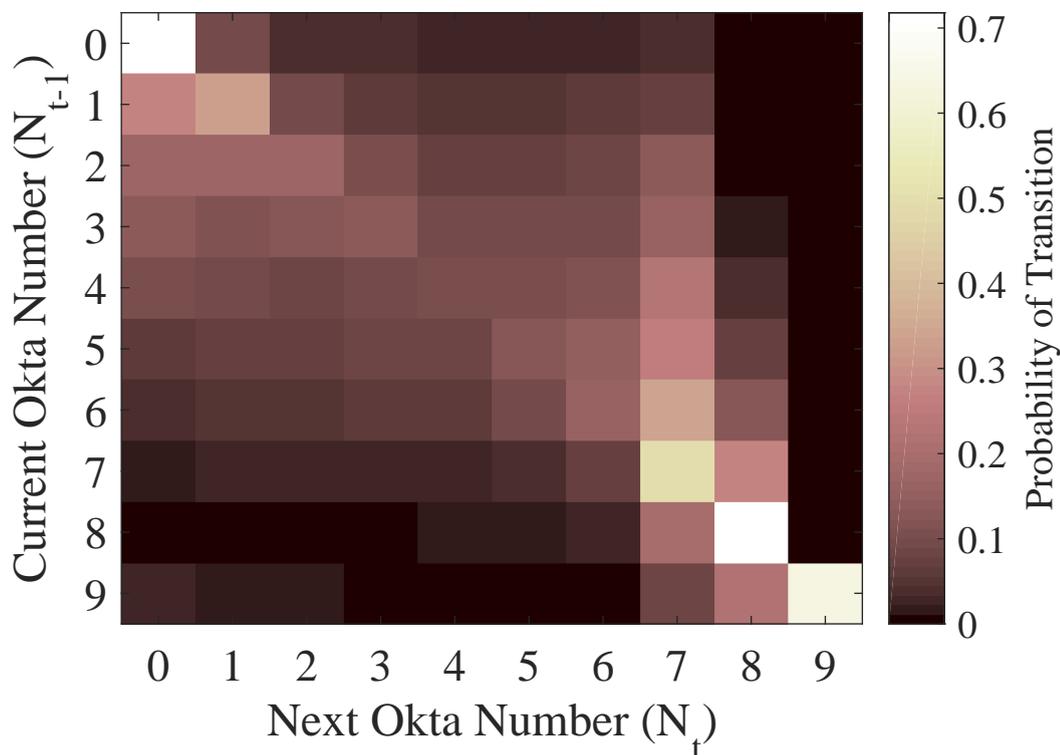


FIGURE 3.8: Colour plot illustrating the MTM for the mean annual transition probability of okta number from the current hour (y-axis) to the next hour (x-axis). This MTM was made using data taken from the [MIDAS \(2015\)](#) data sets for Cambourne, UK.

the key shown on the right. This particular MTM details N that is not separated by season, pressure or time of day. It is therefore the mean transition probability. There is a distinguished pattern in the $y = x$ direction (note the orientation of the y-axis direction, this is the natural orientation for MTMs). The pattern is indicative of a tendency for an okta value to transition to the same value from one hour to the next. Physically this represents N staying the same from one hour to the next. The probability for 0 okta to transition to 0 okta in Cambourne for the whole year is, $P(M_t = 0 | M_{t-1} = 0) = 0.71$.

3.3.3.3 Implementing a Markov chain

For computational use of MTMs within a stochastic process, the row PDFs are converted to cumulative probability functions (CDF), denoted as \mathbf{P}_C^1 , using the following

$$\mathbf{P}_C^1 = \begin{bmatrix} P_{11} & (P_{11} + P_{12}) & \dots & \sum_{n=1}^s P_{1n} \\ P_{21} & (P_{21} + P_{22}) & \dots & \sum_{n=1}^s P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s1} & (P_{s1} + P_{s2}) & \dots & \sum_{n=1}^s P_{sn} \end{bmatrix} \quad (3.12)$$

The implementation of MTMs in a stochastic process is achieved in four steps that are visually demonstrated in figure 3.9. In this plot, the transition from state s at M_{t-1} to a new state at M_t is presented. The green circle indicates the first time step and the yellow circle indicates the second time step. To transition between the two, the process enters the four step process coloured orange. In this example, a simple 3 state variable is used; its \mathbf{P}_C^1 is shown in the bottom left inside the green dashed box. Each row of the MTM represents a unique CDF of a state's transitions. s at M_{t-1} is equal to 1. The first step is to extract the appropriate CDF from the MTM, $P_{C(s|s_{t-1})}^1$, this has been colour coded for ease. The second step is to employ a pseudo-random number generator that drives the stochastic selection.

The variable R is derived from a uniform distribution between 0 and 1, $R \sim U(0, 1)$. R is then queried against each of the probabilities inside of $P_{C(s|s_{t-1})}^1$ using a logical IF statement such that if R is greater than the individual probability, a value of 1 is assigned, else a value of 0 is assigned. The result is a Boolean matrix, \mathbf{R} , that will be used to indicate the next state. It is expressed as

$$\mathbf{R} = R > \mathbf{P}_{C(s_{t-1}|s)}^1 \xrightarrow{\text{if}} \begin{cases} 1 & \text{true} \\ 0 & \text{false} \end{cases} \quad (3.13)$$

The final step to determining the state is by taking a sum of the 1DM Boolean matrix \mathbf{R} . All variables are structured in a 1DM and ordered with predetermined resolution for each element, ω . For example, each element for cloud height represents 100 m and so $\omega = 100$ for z . \mathbf{R} contains the number of states that satisfies equation 3.13, and therefore the state indicated by R can be calculated as

$$s_t = (1 + \sum \mathbf{R}) \times \omega \quad (3.14)$$

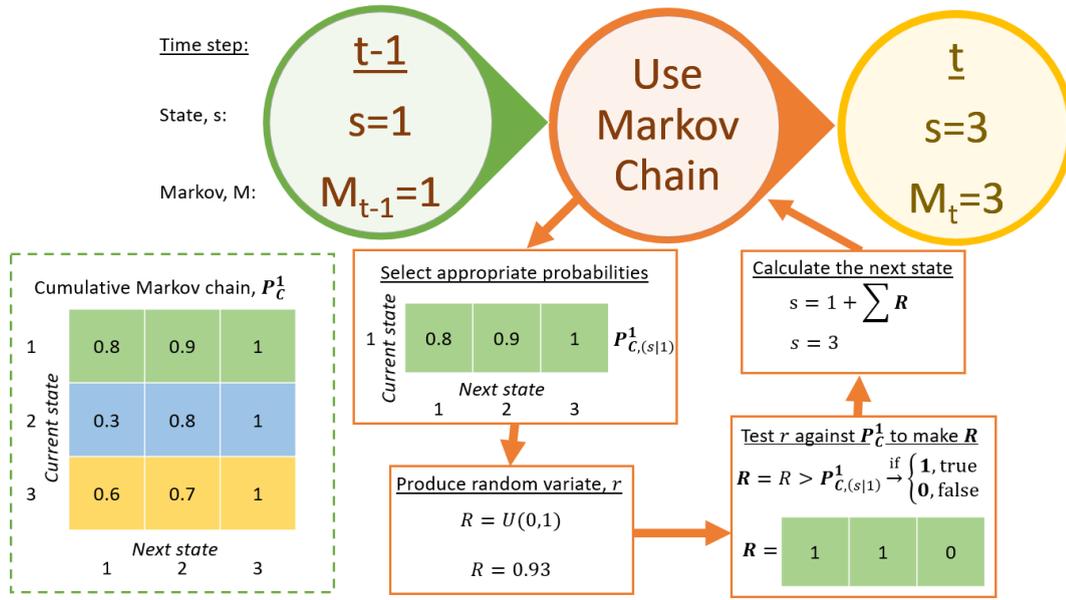


FIGURE 3.9: Demonstration of the stochastic Markov chain process from the current hour, $M_{t-1} = 1$, to the next hour, $M_t = 3$. In the bottom left, the cumulative first order MTM for an arbitrary variable containing 3 states is shown, P_C^1 . As the current hour is at state 1 ($M_{t-1} = 1$), the appropriate row of P_C^1 is selected. The stochastic process is looped through, as indicated by orange arrows until M_t is found to be at $s = 3$. This process is repeated for $t(1, 2, \dots, T)$. However, M_t becomes M_{t-1} at each time step, t .

Once all MTMs are produced, their implementation in stochastic generation of weather time series can occur.

3.3.4 Stochastic generation of weather variable time series

This section will detail the process of obtaining stochastic weather variable time series of z , u_{10} , and N . This part of the model can be thought of as the weather variable generator.

The outcomes of this section are stochastic time series of the weather variables, as well as perfect time series of time stamps that indicate the year, month, day, hour, minute and the season at each time step. Furthermore, through application of the cloud samples, a minutely time series representation of C is produced, which is a binary 1DM containing the state of cloud each minute of simulation denoted \mathbf{C} , this is the only variable that presents intra-hour variability.

3.3.4.1 Initialising the weather variable generator

Before variables can be stochastically generated, an initiation sequence is run that facilitates the weather variable generator. As the SIG allows for user defined inputs, it must be robust and flexible enough to allow for input variation. For example, the SIG must be flexible to leap years, changing starting day, and any simulation duration. This section will discuss the steps to initiate the weather variable generator, the script for this can be referred to in appendix section [A.6](#).

The weather variable generator must be initialised with starting values for each weather variable. [Ehnberg and Bollen \(2005\)](#) state that choosing arbitrary starting conditions of variables has no long term influence on the stochastic process. For this reason, variables are selected randomly from their individual normalised maximum variable ranges using a random variate drawn from a uniform distribution as follows

$$\left. \begin{array}{l} N_{t=1} \\ u_{10,t=1} \\ z_{t=1} \end{array} \right\} = U(1, 0) \times \left\{ \begin{array}{l} N_{\max} \\ u_{10,\max} \\ z_{\max} \end{array} \right. \quad (3.15)$$

z_{\max} is derived from the observable data as the maximum observable, whereas N_{\max} and u_{\max} are previously known as 9 okta and 60 ms^{-1} , respectively. The output of equation [3.15](#) is rounded to the nearest appropriate integer.

In order to select a pressure system duration, P_d , a frequency table for each above and below average system duration is produced. A PDF of P_d is developed for both pressure states. These can then be used in conjunction with a pseudo-random number generator to extract P_d . The current run time of a pressure system is updated with each time step of the simulation. When P_d expires, a new duration is selected. An assumption is that the pressure systems always alternate between above and below average.

3.3.4.2 The weather variable generator process

Using the initialised variables, appropriate MTMs can be selected for each variable as the pressure state, hour of day and season are all now known from initialisation.

The weather variable generator takes each hour of simulation as a time step in chronology. At each hour of simulation, N , u_{10} and z are stochastically selected from the appropriate MTM following the process outlined in section [3.3.3.3](#).

N is converted to C as detailed in table 3.1. The benefit of this conversion is that a wider range of obscured to not obscured fractions can be produced, enabling better total cloud amount flexibility.

An assumption is that clouds travel at the same speed as the wind speed at the height of which they exist. This is intuitive as the wind speeds are what drives cloud motion. The cloud speed at a reference height is denoted u_{ref} . For clouds in the atmospheric boundary layer u_{10} can be extrapolated using the stochastically selected value of z as the target reference height, z_{ref} , as shown in equation 3.16. The atmospheric boundary layer fluctuates in height, however the supporting documentation recommends using an extrapolation for up to 1 km.

To extrapolate u_{10} from 10 m measurement to a height of z_{ref} , the following logarithmic profile extrapolation from Best et al. (2008) is used

$$u_{\text{ref}} = u_{10} \frac{\ln(z_{\text{ref}}/z_{0\text{ref}})}{\ln(10/z_{0\text{ref}})} \quad (3.16)$$

where $z_{0\text{ref}}$ is the roughness length of the location set to 0.14 for rural locations (Best et al., 2008).

Above the atmospheric boundary layer, the geostrophic wind speeds in the free atmosphere are influenced by pressure and thermal gradients (UKMO, 1997). Estimating geostrophic wind speeds is difficult using mean hourly surface observational data, thus presenting one of the limitations of the SIG. Without readily available information of atmospheric pressure and thermal gradients, alternatives must be used. Excellent methods exist to estimate u_{ref} , such as the wavelet variability model (Lave and Kleissl, 2013), however, the data input required is an irradiance profile — an input avoided in the rationale of this methodology. Future development is required to truly address u_{ref} , the validation results however were found to be relatively insensitive to this aspect, and so at this time a guided estimation is utilised. Using the same u_{range} as used in the wavelet variability model (0 to 25 ms^{-1}) and typical geostrophic wind speeds of 3.5 ms^{-1} (Mathiesen et al., 2013), the estimated wind speed at cloud heights above 1 km is determined allowing variation between the suggested range with the mean found at the typical free atmosphere wind speed as

$$u_{\text{ref},(z_{\text{ref}}>1\text{km})} \sim \Gamma(2.69, 2.14) \quad (3.17)$$

Where $\Gamma(\cdot)$ is the gamma distribution function, 2.69 is the Gamma shape parameter of the and 2.14 is the Gamma scale parameter fitted to have mean $\overline{u_{\text{ref}}} = 3.5 \text{ ms}^{-1}$ with a u_{range} of 0 to 25 ms^{-1} (Mathiesen et al., 2013).

Now that u_{ref} and C are defined, a cloud sample can be selected from the cloud sample database. The indexing equation for selecting the appropriate bin can be seen on line 73 in appendix section A.7. The equation is robust for flexibility with range of u_{ref} and n_c , a random variate from $U(1,0)$ is used to select from the bins containing n_s options of cloud sample per u_{ref} per C .

The final step in the weather variable generation is to update the time stamp and re-initialise the variables. The year, month, day and hour must be defined at each time step of simulation. An hour is added at the end of each iteration, all time resolutions are updated against the new hour. The hours within the simulation operate from 1 to 24. These indicate the 1st and 24th hour of the day and so hour 1 represents 00:00–01:00, and hour 24 represents 23:00–00:00. Should the hour move from 24 back to 1, the day number is increased. Should the day number move outside of a month’s range, the month is increased. Should the month move from December to January, the year is increased. The season is defined using day numbers. With the new time stamp, the MTMs can be initialised again for use in the stochastic selection of a variable.

This whole process is repeated for the entire user defined duration of the simulation. The result is a Boolean 1DM time series of cloud cover with a 1-min resolution, \mathbf{C} . \mathbf{C} is the sun obscured part of the SIG and is used in the production of k_c , and ultimately GHI.

An example of stochastic progression of time series can be seen in figure 3.10. It shows the stochastic time series of the three generated variables of C , z and u_{ref} . Notice some correlation between the cloud height and cloud speed, as is consistent with equation 3.16.

3.3.5 Sun-Earth geometry

As each minute within \mathbf{C} is to be aligned with theoretical irradiance, it is critical to know the sun’s position in the sky as viewed from the location on the Earth’s surface, as input from latitude Θ and longitude Φ .

This section discusses the options of calculating the sun’s position relative to a location on the surface of the Earth.

3.3.5.1 Background to calculating solar angles

Before exploring the literature options for solar geometry calculations, some definitions must be defined first. To describe the position of the sun in the sky as it would be

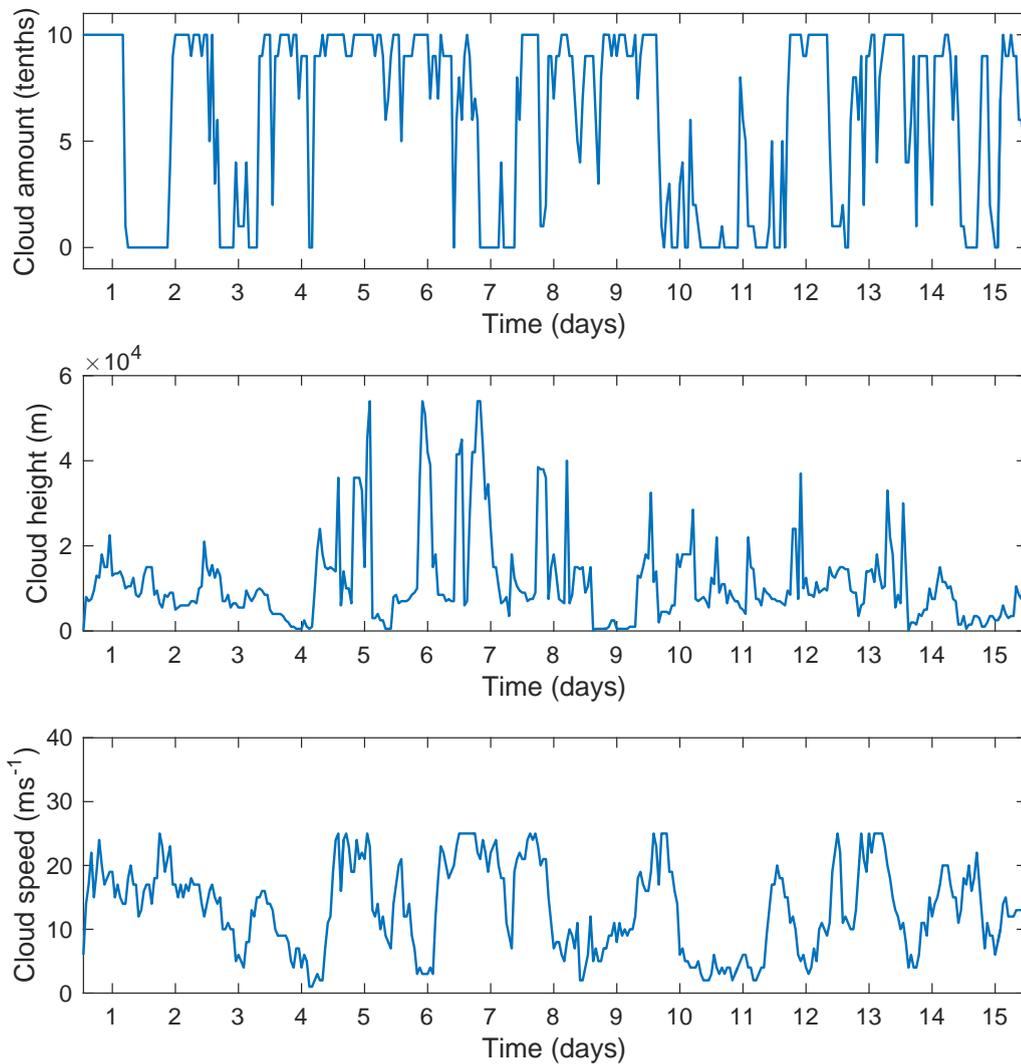


FIGURE 3.10: Examples of the stochastically generated time series of (top) the cloud amount C in tenths, (middle) the cloud height z_{ref} in m, and (bottom) the cloud speed u_{ref} in ms^{-1} . The resolution is 1-hour and spans for 15 full days modelled for 1st-15th January 2012

viewed from a point on the Earth's surface, two angles are required: the solar zenith angle, θ_z , and the solar azimuth angle, τ . Equations for both are described later in this section.

θ_z is the angle that describes the distance of the sun away from the vertical zenith from a point on the Earth's surface. It is dependent on the time of day as well as the solar declination, δ . τ describes the horizontal position of the sun in relation to the southern meridian. Some authors use an alternate definition that describes the τ from the north meridian, this thesis will use the definition where $\tau = 0^\circ$ describes the sun as directly south of a point on the Earth's surface, increasing in a clockwise manner where 90° , 180° and 270° are west, north and east, respectively. Much like θ_z , τ is dependent on time and δ . Both angles are shown in 3.11.

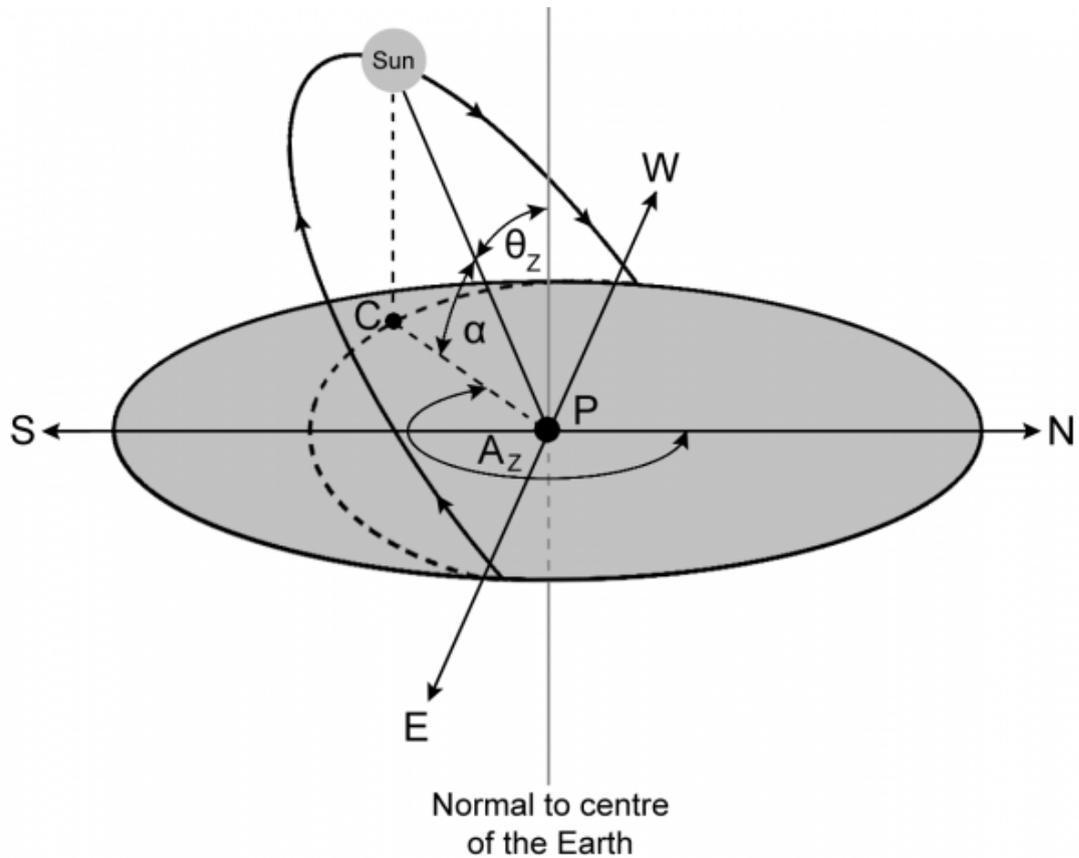


FIGURE 3.11: Simple diagram of a sky dome indicating measurements of the solar zenith angle, θ_z , measured between the local vertical and the direction of the solar beam; and the solar azimuth angle, A_z although denoted τ in this thesis, measured from the south meridian to the direction of the sun. All these angles are in relation to a point P, on the Earth's surface (image reproduced with permission from [ITACA \(2015\)](#)).

There are many methodologies to determine the Sun-Earth geometry. The requirement of the SIG is to produce the geometry to a high accuracy to facilitate high temporal resolution. For this reason, only literature dealing with high accuracy estimates of θ_z and τ are considered. Importance is placed on the accuracy of the model, however there is a trade off with computational complexity and input requirements. The most appropriate model is therefore the greatest accuracy offered within the acceptability of computational power. As this model can be used for concentrating solar power systems as well as PV, a higher accuracy (< 0.005) is desirable. Application in solar concentration and tracking technologies require a higher accuracy than stationary systems, as stationary systems can tolerate errors of a few degrees before significant losses occur ([Grena, 2012](#)). Table 3.4 lists some of the most significant literature that calculates the Sun-Earth geometries. Uncertainties are typically $\leq 0.01^\circ$ in solar zenith and azimuth calculations.

Of these models, [Blanco-Muriel et al. \(2001\)](#) presented the Plataforma Solar de Almeria

TABLE 3.4: Solar parameters calculated by different authors. This table was reproduced from [Blanco-Muriel et al. \(2001\)](#) and updated by the author. δ is the declination angle, ra is the right ascension, EoT is the equation of time or equivalent, θ_a is the azimuth angle, and θ_z is the zenith or elevation angle.

Reference	δ	ra	EoT	θ_a	θ_z
Cooper (1969)	✓				
Spencer (1971)	✓		✓		
Swift (1976)	✓				
Pitman and Vant (1978)	✓		✓		
Walraven (1978)	✓	✓		✓	✓
Lamm (1981)			✓		
Michalsky (1988)	✓	✓	✓	✓	✓
Blanco-Muriel et al. (2001)	✓	✓	✓	✓	✓
Reda and Andreas (2004)	✓	✓	✓	✓	✓
Grena (2008)	✓	✓	✓	✓	✓
Chang (2009)	✓	✓	✓	✓	✓

(PSA) algorithm that is the most often used, reportedly for its simplicity and low error. There are more benefits to this methodology. It has a more efficient and simplified method of calculating the Julian Day from an input of calendar date, has a higher computational efficiency, models earlier than the Blanco-Muriel model are inaccurate for the southern hemispheres, the only input requirements are time and location of which both are input by the user. These reasons make the [Blanco-Muriel et al. \(2001\)](#) model ideal for use in the SIG.

More modern methodologies exist such as the [Reda and Andreas \(2004\)](#) who present the Solar Position Algorithm (SPA). The SPA offers a very low error of $< 0.0003^\circ$ and performs significantly better than the PSA at higher zenith angles, however it requires the local atmospheric pressure and temperature as an input. This is also true of the most accurate method found at time of writing by [Grena \(2008\)](#). Neither the pressure nor temperature are known at each minute, which makes these methodologies unsuitable for use in the SIG.

The PSA by [Blanco-Muriel et al. \(2001\)](#) is not without flaw. The time period of validity is 1995-2015 without further analysis and updating the constants within the method. This still does not present an issue to the proposed methodology as the observation data and data used in validation are within these limits, however, it does present an opportunity for future improvement. Other methods exist to obtain greater flexibility with the time range of such as [Reda and Andreas \(2004\)](#), [Grena \(2008\)](#) and [Grena \(2012\)](#) which operate between 2000 BC to 6000 AD. As mentioned, these models require pressure and temperature, furthermore, they have significantly higher computational demand (orders of 10 times more). For all these reasons, the model

proposed by Blanco-Muriel et al. (2001) is selected for implementation within the SIG as it is the most appropriate for application in grid impact analysis.

3.3.5.2 Blanco-Muriel PSA algorithm methodology

The inputs required in the PSA are time and location. Time can be input in calendar format as five separate variables of year, month and day (y , m and d , respectively) with the Universal Time in hours and minutes (h and min). These are all SIG outputs from section 3.3.4. The location is required in degrees, these are user defined inputs of Θ and Φ , which each MIDAS (2015) station provides to four decimal places.

The following equations detail the PSA model. Some symbols differ from the original. All of the following are calculated using 1DM time series for each variable presented.

The Julian Day, JD , is calculated uniquely by Blanco-Muriel et al. (2001). It is an integer assigned to the whole solar day, beginning with 0 at noon on the 1st January 4713 BC in accordance to the Julian calendar. The PSA expression of JD here allows for the hour to be modelled, and subsequently the minutely fraction of JD through use of h containing minute fractions. JD is calculated as

$$a = \left\lfloor \frac{m - 14}{12} \right\rfloor \quad (3.18)$$

$$JD = \left\lfloor \frac{1461(y + 4800 + a)}{4} \right\rfloor + \left\lfloor \frac{367(m - 2 - 12a)}{12} \right\rfloor - \left\lfloor \frac{3(y + 4900 + \lfloor \frac{a}{100} \rfloor)}{4} \right\rfloor + d - 32075 - 0.5 + \frac{h}{24} \quad (3.19)$$

The use of $\lfloor \rfloor$ is a modulo, floor or integer division function whereby the remainder in the division is disregarded. Note that the final division of equation 3.19 is not an integer division.

In the PSA, JD is normalised to 1st January 2000 and made to begin at midnight as opposed to noon using the following adjustment.

$$JD_n = JD - 2451545.0 \quad (3.20)$$

The sun's ecliptic coordinates of ecliptic longitude Φ_{sun} and the obliquity of the ecliptic ep are derived from JD_n .

$$\Omega = 2.1429 - 0.0010394594 \times JD_n \quad (3.21)$$

$$\overline{\Phi_{\text{sun}}} = 4.8950630 + 0.017202791698 \times JD_n \quad (3.22)$$

Where $\overline{\Phi_{\text{sun}}}$ is the mean longitude of the sun.

$$g = 6.2400600 + 0.0172019699 \times JD_n \quad (3.23)$$

Where g is the mean anomaly of the sun.

$$\Phi_{\text{sun}} = \overline{\Phi_{\text{sun}}} + 0.0334160 \sin(g) + 0.00034894 \sin(2g) - 0.0001134 - 0.0000203 \sin(\Omega) \quad (3.24)$$

$$ep = 0.4090928 - 6.2140e^{-9} \times n + 0.0000396 \cos(\Omega) \quad (3.25)$$

From the ecliptic coordinates, the celestial coordinates of the right ascension, ra , and the solar declination, δ , can be calculated.

$$ra = \tan^{-1} \left[\frac{\cos(ep) \sin(\Phi_{\text{sun}})}{\cos(\Phi_{\text{sun}})} \right] \quad (3.26)$$

care must be taken that ra value is greater than 0. [Blanco-Muriel et al. \(2001\)](#) add a logical query that adds 2π to ra should it fall below 0.

$$\delta = \sin^{-1} [\sin(ep) \sin(\Phi_{\text{sun}})] \quad (3.27)$$

The next step in the PSA is the conversion from celestial coordinates to horizontal coordinates. This requires the time format of Coordinated Universal Time (UTC) whereby hours can be positive and negative either side of UTC with a decimal representation of minutes (e.g. 17.5 would be 17:30). In this time format, the hour angle ω can be calculated. ω is used to describe the solar time as a rotation such that -180° is solar midnight through to 0° at solar noon, approaching 180° as time approaches solar midnight once more.

$$gmst = 6.6974243242 + 0.0657098283 \times JD_n + h \quad (3.28)$$

Where $gmst$ is the Greenwich mean sidereal time.

$$lmst = (gmst \times 15 + \Phi) \frac{\pi}{180} \quad (3.29)$$

Where $lmst$ is the local mean sidereal time.

$$\omega = lmst - ra \quad (3.30)$$

Finally the θ_z and τ can be determined from Θ , ω and δ .

$$\theta_z = \cos^{-1} [\cos(\Theta) \cos(\omega) \cos(\delta) + \sin(\delta) \sin(\Theta)] \quad (3.31)$$

$$\tau = \tan^{-1} \left[\frac{-\sin(\omega)}{\tan(\delta) \cos(\Theta) - \sin(\Theta) \cos(\omega)} \right] \quad (3.32)$$

Blanco-Muriel et al. (2001) apply a correction to θ_z termed the parallax correction, $Plax$

$$Plax = \frac{\bar{r}_E}{u} \sin \theta_z \quad (3.33)$$

Where \bar{r}_E is the Earth's mean radius taken as 6371.01 km and au is the Astronomical Unit taken as 149,597,890 km. Lastly, the parallax correction is applied.

$$\theta_z = \theta_z + Plax \quad (3.34)$$

Further reading and an in depth discussion of the angles and concepts are provided by Kalogirou (2014).

3.3.6 Generation of clear-sky indices

This section will detail how a time series of k_c is created accounting for different okta states, sun obscured moments, random fluctuations and states of cloud stability. This time series of k_c will be denoted \mathbf{k}_c . As discussed in section 3.1.2, the irradiance will be expressed in terms of the clear-sky index k_c , a ratio that parameterises the atmospheric transmission of irradiance. It is found as the ground measured irradiance over the clear sky irradiance, $k_c = G/G_{cs}$. The SIG script regarding this section can be seen in appendix section A.9.

3.3.6.1 Relationship of clear-sky index to total cloud amount

In order to account for the varying optical thicknesses from clouds, k_c is selected from a distribution as opposed to a fixed value. The hypothesis proposed here is that the irradiance presents a statistical distribution with a one-to-one relationship to N , $k_c = f(N)$. In order to test this hypothesis, the author entered into a collaboration with

research partner Dr Christopher J. Smith, who at the time of writing was undertaking a PhD in computational methods for assessment of solar energy potential in present and future climates. The hypothesis proposed here after collaboration was extended and developed upon and published (Smith et al., 2017). The idea was conceptually the authors'. The design of the methodology to analyse $k = f(N)$ was a collaborative effort however the coding to derive the relationship was carried out by Dr Smith. The intention is to develop distributions of $N = 0, 6, 7, 8$, which describe periods of clear, partially cloudy, nearly overcast and overcast skies, respectively. For $N = 1-5$, there are significant areas of clear sky, this presents the challenge to estimating the sun obscured portion of the k_c distribution. For this reason, $N = 6$ is assumed to be representative of the distributions associated with partly clouded sky, and so clouds that exist within $N = 1-6$ will be represented by the optical depths described by the k_c distribution for $N = 6$.

Perhaps one of the most popular methods of calculating G_{cs} is the DISORT radiative transfer model (Stamnes et al., 2000). There were 49 publications mentioning or directly using the DISORT package in 2016 alone. This is unsurprising as it has had commercial development since its creation in 1988, with many journal publications detailing its development. DISORT is one of the standard radiative transfer solvers in the libRadtran package (Mayer and Kylling, 2005). For this ease of application and well received reviews and development from the scientific community, DISORT is the methodology of choice to calculate estimations of G_{cs} for deriving the $k_c = f(N)$ relationship. In the DISORT calculations, the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-interim reanalysis data is accessed (ECMWF, 2014) at a resolution of $1.5^\circ \times 1.5^\circ$ to obtain atmospheric profiles of temperature, O_3 concentration, precipitable water vapour and surface albedo for each month of 2012. Monthly aerosol data was provided by the GLOMAP model at a resolution of $2.8^\circ \times 2.8^\circ$, specifying scattering and absorption coefficients and asymmetry parameters for 6 bands in the shortwave spectrum for sulphate, sea-salt, black carbon and particulate organic matter aerosols in four size modes (Scott et al., 2014). Through use of the DISORT model, G_{cs} is determined which facilitates the derivation of k_c as a function of N .

The following steps were taken to determine the relationship $k_c = f(N)$:

1. Observations of G and N taken from 91 MIDAS (MIDAS, 2015) stations for the year 2012. 111090 observations were used in total.
2. Set 9 okta to overcast hours of 8 okta, $N_{(N=9)} = 8$. Keep only $N = 0, 6, 7, 8$.

Hourly okta, N	k_c distribution for cloud cover
0	$k_c \sim N(0.99, 0.08)$
1–6	$k_c \sim N(0.6784, 0.2046)$
7	$k_c \sim \text{Weibull}(0.5577, 2.4061)$
8	$k_c \sim \Gamma(3.5624, 0.0867)$

TABLE 3.5: Distributions of k_c by N derived from histograms shown in figure 3.12. The normal distributions are parameterised by (μ, σ) where μ is the mean and σ is the standard deviation. The Weibull and Gamma ($\Gamma(a, b)$) distributions are parameterised by (a, b) where a is a shape parameter and b is a scale parameter.

3. Reject hours where θ_z was greater than 80° in order to minimise horizon effects and night time.
4. k_c is calculated from measured G and theoretical calculation of G_{cs} (using the DISTORT package in libRadtran that follows the Blanco-Muriel et al. (2001) algorithm).
5. Group k_c values by associated N .
6. Produce histograms and fit distributions using maximum likelihood estimation ($N = 0$ was fitted visually with a normal distribution).

The final distributions of k_c by N are shown in figure 3.12 and listed in table 3.5. Also shown in figure 3.12 is the bimodal distribution of k_c when not a function of N . Two distinct peaks can be seen that are indicative of clear and cloudy categories of irradiance.

In the $N = 0$ histogram, there are several k_c values that are significantly less than 1, indicating that even when no cloud is registered, the irradiance reported can be much lower than expected. This can be explained by the fact that the cloud sensor is pointing directly upwards whilst the sun is in a different sector of the sky and could possibly be obscured by cloud. This is particularly an issue in the UK where the sun is always at least 27° from zenith. To ensure realistic k_c values when the sun is shining, these low values are rejected and a normal distribution fitted visually to the obvious peak of the histogram, which represents genuinely clear hours.

3.3.6.2 Clear-sky index with a 1 minute resolution

A baseline time series of clear-sky index is made for each day that is applied to moments when the sun *is not* obscured. This sun-not-obscured baseline time series is denoted $\mathbf{k}_{c,\text{clear}}$. Each day of simulation, a single value of k_c is extracted and applied to every minute of that day. Every day is different and, therefore, accounts for solar variability

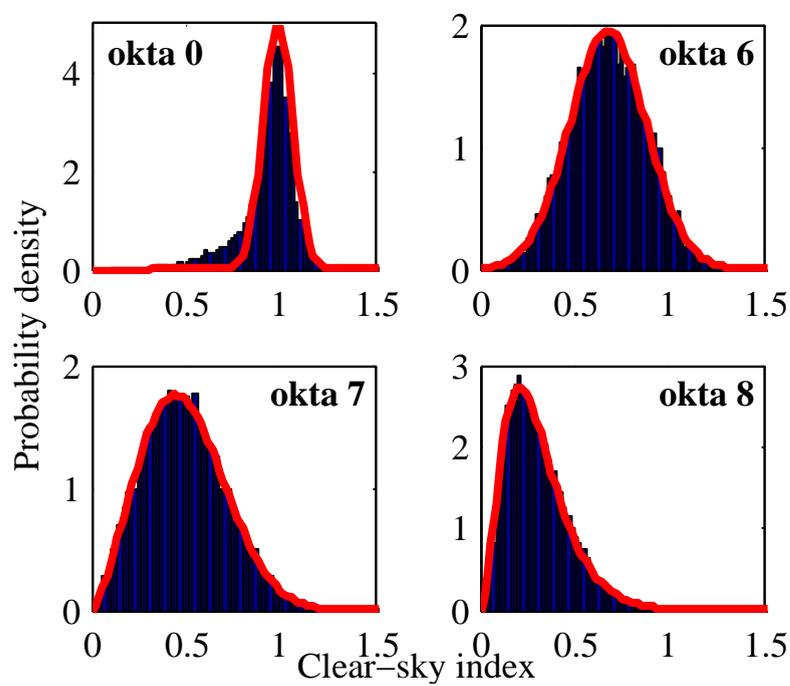
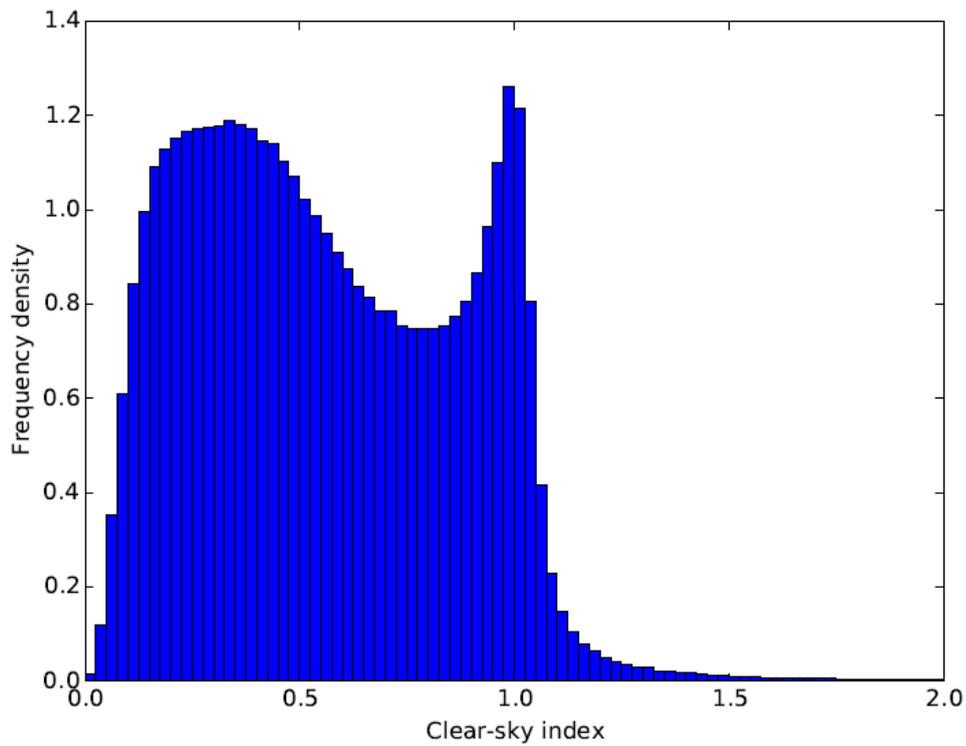


FIGURE 3.12: Top) bi-modal histogram of all hourly k_c observations taken from 63 UK weather stations between 2010-2013. Bottom) histograms of k_c separated by N , with analytical distributions fitted and shown in red. Plots produced by Dr Chris Smith.

between days. The distribution of k_c used to select daily $\mathbf{k}_{c,\text{clear}}$ is dependent on the climate of study. For Cambourne, UK, the sun-not-obscured hours were found to be well represented using a normal distribution $N(0.99, 0.008)$. This distribution is relative to the CSI, which is dependent on the method of calculation, discussed in section 3.3.7.

For clouded hours, the baseline time series of clear-sky index is made for the whole duration of the simulation that is applied to moments when the sun *is* obscured. This time series is denoted $\mathbf{k}_{c,\text{cloud}}$, and is selected from the appropriate distribution of k_c by N .

To move from hourly resolution to minutely representations of both of $\mathbf{k}_{c,\text{cloud}}$ and $\mathbf{k}_{c,\text{clear}}$, a linear interpolation between each value of $k_{c,\text{cloud}}$ and $k_{c,\text{clear}}$ in the time series was performed to obtain baseline minutely time series, denoted $\mathbf{k}_{c,\text{cloud},m}$ and $\mathbf{k}_{c,\text{clear},m}$.

To produce more realistic irradiance profiles that have intra-hour variability and fluctuations during clouds, $\mathbf{k}_{c,\text{cloud},m}$ is allowed to fluctuate between fixed intervals. The SIG allows fluctuations at intervals that are a factor of 60 (as this is the maximum number of fluctuations within the hour). Allowing fluctuations at intervals > 10 minutes did not offer similarities in terms of ramp rate occurrences or the variability indices (described later in section 3.4), whilst fluctuations at < 5 minute intervals saw too much variability. The intention is to capture the gentle rolling of the $\mathbf{k}_{c,\text{cloud},m}$ observed in real irradiance, attributed to changing cloud shape. Every 6 mins, a new k_c value is extracted and linearly interpolated between.

The sun obscured method can now be put in place to produce a single, 1 min resolution time series of the clear-sky index, $\mathbf{k}_{c,m}$. This is achieved using \mathbf{B} , the Boolean 1DM indicating cloud or clear periods every minute. By looping through each time step, t , the appropriate k_c is selected from either $\mathbf{k}_{c,\text{cloud},m}$ or $\mathbf{k}_{c,\text{clear},m}$.

$$\mathbf{B}_t = 1 \xrightarrow{\text{if}} \begin{cases} \text{True} & \mathbf{k}_{c,m}^t = \mathbf{k}_{c,\text{cloud},m}^t \\ \text{False} & \mathbf{k}_{c,m}^t = \mathbf{k}_{c,\text{clear},m}^t \end{cases} \quad (3.35)$$

Where $\mathbf{B}_t = 1$ is a logical query asking whether there is cloud at time step t . $\xrightarrow{\text{if}}$ is the use of a logical IF operator. The result of this equation assigns a k_c value from either $\mathbf{k}_{c,\text{cloud},m}$ or $\mathbf{k}_{c,\text{clear},m}$ into the appropriate t location of $\mathbf{k}_{c,m}$, depending on the outcome of the logical query.

Following this, minutely k_c variations were introduced for both obscured and not obscured minutes by using a Gaussian white noise multiplier as a function of the hourly N .

$$\mathbf{k}_{c,m} = \mathbf{k}_{c,m} f, \quad f \sim N(1, \sigma) \quad (3.36)$$

Where the standard deviation, σ is arbitrarily assigned for obscured minutes as

$$\sigma = 0.01 + 0.003N \quad (3.37)$$

And for not obscured minutes as

$$\sigma = 0.001 + 0.0015N \quad (3.38)$$

Equation (3.37) provides the greatest variation in obscured minutes for higher N while equation (3.38) provides the greatest variation in clear minutes with increasing N .

A further two adjustments are made to ensure that unrealistic values of G are not seen. If a value in $\mathbf{k}_{c,m}$ is less than 0.01, it is set equal to 0.01. At the other end of the scale, there are many situations where $\mathbf{k}_{c,m}$ exceeds a value of 1 in the distribution tails. This is most likely to happen at low solar elevations where horizon and ground effects are more pronounced, and the division of a small G by a very small G_{cs} leads to high values of k_c . The largest values of k_c from the observed irradiances were found to obey the relationship with $R^2 = 0.9931$, fitted using maximum likelihood estimation of θ_z against k_c such that:

$$\begin{aligned} k_{c,\max}(\theta_z) = & 27.21 \exp(-114 \cos \theta_z) \dots \\ & + 1.665 \exp(-4.494 \cos \theta_z) + 1.08 \end{aligned} \quad (3.39)$$

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

3.3.6.3 Extended periods clear and overcast conditions

If the sky is overcast for an extended period, defined as a period of 8 okta lasting longer than 6 hours, a smoothing is made to corresponding moments in $\mathbf{k}_{c,\text{cloud}}$. This is justified because the clear-sky index of overcast skies does not vary extensively (Skartveit and Olseth, 1992; Stein et al., 2012). Firstly, a static number of intervals is

set for simple computational application. 20 intervals will separate the first and last minute of the extended period. The $k_{c,\text{cloud}}$ at each of the 20 intervals is kept, whilst the rest are removed. Between each interval, $k_{c,\text{cloud}}$ values are linearly interpolated using a uniform linear spacing function in Matlab. The Gaussian white noise is re-applied to these periods.

A similar function is applied to extended periods of 0 okta, except the duration is reduced to periods lasting more than 3 hours. Contrastingly to the long period of 8 okta, no intervals are set, therefore, $\mathbf{k}_{c,\text{clear},m}$ is defined as a linear interpolation from the first minute to the final minute of the extended period.

Both $\mathbf{k}_{c,\text{clear},m}$ and $\mathbf{k}_{c,\text{cloud},m}$ are assimilated to form a single IDM of the clear-sky indices \mathbf{k}_c . This is assigned using the Boolean IDM of sun obscured indicators, B , produced in section 3.3.4. The assimilation is achieved as

$$\mathbf{k}_c = \begin{cases} \mathbf{B} = 0 & \xrightarrow{\text{if}} & \mathbf{k}_{c,\text{clear},m} \\ \mathbf{B} = 1 & \xrightarrow{\text{if}} & \mathbf{k}_{c,\text{cloud},m} \end{cases} \quad (3.40)$$

This places the clear-sky indices from the extended clear and overcast periods in the appropriate time step of \mathbf{k}_c .

3.3.6.4 Introduction of cloud edge enhancement events

A final alteration to \mathbf{k}_c is the inclusion of cloud edge enhancements. Cloud edge enhancement, CEE , describes events whereby a point on the Earth's surface receives a larger amount of incident irradiance than is available in the clear-sky irradiance. The events are attributed to irradiance reflecting from the edges of clouds. The typical behaviour of irradiance in the 60 sec leading up to, and after the largest 1 sec ramps is detailed by Lave et al. (2012). This behaviour is applied to periods of transitions from clear to clouded moments, which can be found as transitions from both $1 \leftrightarrow 0$ in \mathbf{B} . The CEE behaviour is normalised to 1 to form a correctional factor for both ramp up events, CEE_{up} , and ramp down events, CEE_{down} , and are expressed as a function of a magnitude, M .

The magnitudes of CEE ramps, M , were analysed using 1 min data from the WRMC-BSRN (2014) against the corresponding mean hourly weather observational data of okta from MIDAS (2015) of the same geographic location. The frequency and magnitude of ramp events attributed to CEE correlated with the N . CEE events were mathematically defined by the author as moments of G that are $> 25\%$ of G_{cs} , which

is calculated using the HELIOSAT method as described by [Hammer et al. \(2003\)](#) and is detailed in section 3.3.7. CDF profiles of the magnitude for each N were made allowing for random variate driven extraction once the appropriate N -guided magnitude distribution is selected. The normalised CEE correction factors are further corrected as a function of M , which is itself a function of N , as detailed in equations 3.41–3.43.

$$M = f(N) \quad (3.41)$$

Once M is determined, the appropriate k_c inside \mathbf{k}_c can be modified according to the CEE moment by using the time index t

$$CEE_{\text{up}} = \mathbf{k}_{c,t}(1 + M) \quad (3.42)$$

$$CEE_{\text{down}} = \mathbf{k}_{c,t}(1 - M) \quad (3.43)$$

The application of the CEE is performed using \mathbf{B} . Iterating through \mathbf{k}_c and using a logical IF statement that queries whether \mathbf{B}_t undergoes a ramp on account of cloud, the appropriate M adjusted CEE correction factor (CEE_{up} or CEE_{down}) is applied to the time before and after the ramp as shown in equations 3.44–3.45.

$$\mathbf{B}_{t-1} = 1 \quad \& \quad \mathbf{B}_t = 0 \quad \xrightarrow{\text{if}} \quad \mathbf{k}_{c,t} = CEE_{\text{up}} \quad (3.44)$$

$$\mathbf{B}_{t-1} = 0 \quad \& \quad \mathbf{B}_t = 1 \quad \xrightarrow{\text{if}} \quad \mathbf{k}_{c,t-1} = CEE_{\text{down}} \quad (3.45)$$

\mathbf{k}_c now contains a time series of the clear-sky index for every minute of simulation. It contains within it daily fluctuations, 6 min variability, cloud edge enhancements, stability in extended overcast and clear periods and 1 min fluctuations guided by cloud amount. The distributions by N contain inherent seasonal, diurnal and pressure system dependent coverage values. It is acknowledged that CEE is more likely to be specific to certain cloud types, namely cumulus, there is no method available to assign a cloud type to the binary representations. As such the chance of occurrence is used to determine if a cloud receives a CEE event correction.

3.3.7 Irradiance calculations

Once \mathbf{k}_c have been generated for the simulation duration, T , the irradiance on an arbitrary plane is calculated that accounts for atmospheric and meteorological attenuation.

The irradiance calculations are in two distinct parts. Firstly, the theoretical CSI and its direct and diffuse components (G_{cs} , $G_{B,cs}$ and $G_{D,cs}$, respectively) based on Sun-Earth geometry are calculated before the atmospheric and meteorological transmission attenuation are accounted for to obtain the simulated GHI, G . Lastly, the actual received direct and diffuse horizontal irradiance (G_B and G_D) are calculated from their clear sky counterparts before being translated onto an arbitrary plane to estimate the tilted irradiance, G_T .

3.3.7.1 Obtaining the theoretical clear sky irradiance

Irradiance models that calculate the CSI at any point on the Earth's surface exist extensively in literature. This thesis does not attempt to create a new version of a CSI model as it is far beyond the scope of this thesis. With the proliferation of CSI models in literature, it can be difficult to select an appropriate method, particularly as many are introduced without appropriate validation (Gueymard, 2012). Reviews of these models have existed as early as 1978 with Atwater and Ball (1978), then later by Bird and Hulstrom (1981a), Carroll (1985), Gueymard (2012) and Badescu et al. (2012).

The DISORT model is complex and, therefore, is not as suitable for deployment in the SIG when there are well-validated, simple alternatives. A model identified as a good methodology with very few inputs in the review of CSI models by Gueymard (2012) is the HELIOSAT method proposed by Hammer et al. (2003). The HELIOSAT method is flexible enough to calculate minutely $G_{B,cs}$ and $G_{D,cs}$ for any location of choice in very little computational time. Therefore, the HELIOSAT methodology is utilised in the SIG and can be seen as model script in appendix section A.8 from line 23 onwards.

The first step is to determine the extraterrestrial irradiance, G_0 , from the solar constant (taken as 1367 Wm^{-2}). As the Earth's orbit around the sun changes throughout the year on account of orbital eccentricity, ϵ , G_0 can be found following equations 3.46 and 3.47 as accounting for ϵ and using the day number, d (Snasel et al., 2011).

$$\epsilon = 1 + 0.03344 \times \cos \left(\frac{2\pi d}{365.25} - 0.048869 \right) \quad (3.46)$$

$$G_0 = 1367 \times \epsilon \quad (3.47)$$

The air mass, m , must be calculated next and is found with an expression introduced by Kasten and Young (1989) shown in equation 3.48. The air mass is the path length that light must travel through the atmosphere before it arrives at the Earth's surface

normalised to the shortest possible path.

$$m = \frac{1 - \frac{z_l}{10000}}{\cos \theta_z + 0.50572(96.07995 - \theta_z)^{-1.6364}} \quad (3.48)$$

Where z_l is the height above sea level of the location being modelled in metres.

The Rayleigh optical thickness, δ_R , is the the optical thickness of a clean and dry atmosphere where only Rayleigh scattering occurs, it is a description of how much light is lost to the scattering of light off air molecules.

$$\delta_R = \begin{cases} (6.6296 + 1.7513m - 0.1202m^2 + 0.0065m^3 - 0.0013m^4)^{-1} & m < 20 \\ (10.4 + 0.718m)^{-1} & m > 20 \end{cases} \quad (3.49)$$

δ_R differs for m above 20 and is the basic parameter in determining the Linke turbidity factor, T_L (Kasten, 1996). T_L is defined as the total number of Rayleigh atmospheres necessary to represent the actual optical thickness of the atmosphere. T_L is described as a very convenient tool of approximation that describes the optical thickness of the atmosphere due to both the absorption and scattering by aerosol particles, and the absorption of radiation by any water vapour present in a clean and dry atmosphere. The larger the value of T_L , the larger the attenuation of radiation within a clear sky atmosphere (SoDa, 2016).

For use in HELIOSAT, T_L is normalised by δ_R and m to avoid daily variations and is denoted $T_{L,2}$.

$$T_{L,2} = \frac{T_L(m)\delta_R}{\delta_{R(2)}} \quad (3.50)$$

To obtain $T_{L,2}$, lookup tables containing $T_{L,2}$ are taken from SoDa (2016); they exist for each month of the year. The lookup tables form a map with a spatial resolution of 5 minutes of arc angle, which approximates to 10 km at mid-latitudes. They were calculated adhering to methods outlined by Diabaté et al. (2003). An example of a T_L look-up table is provided in figure 3.13.

The clear sky direct and diffuse irradiance components ($G_{B,cs}$ and $G_{D,cs}$) are obtained as

$$G_{B,cs} = G_0 e^{(-0.8662 \times T_{L,2} \times \delta_R \times m)} \cos \theta_z \quad (3.51)$$

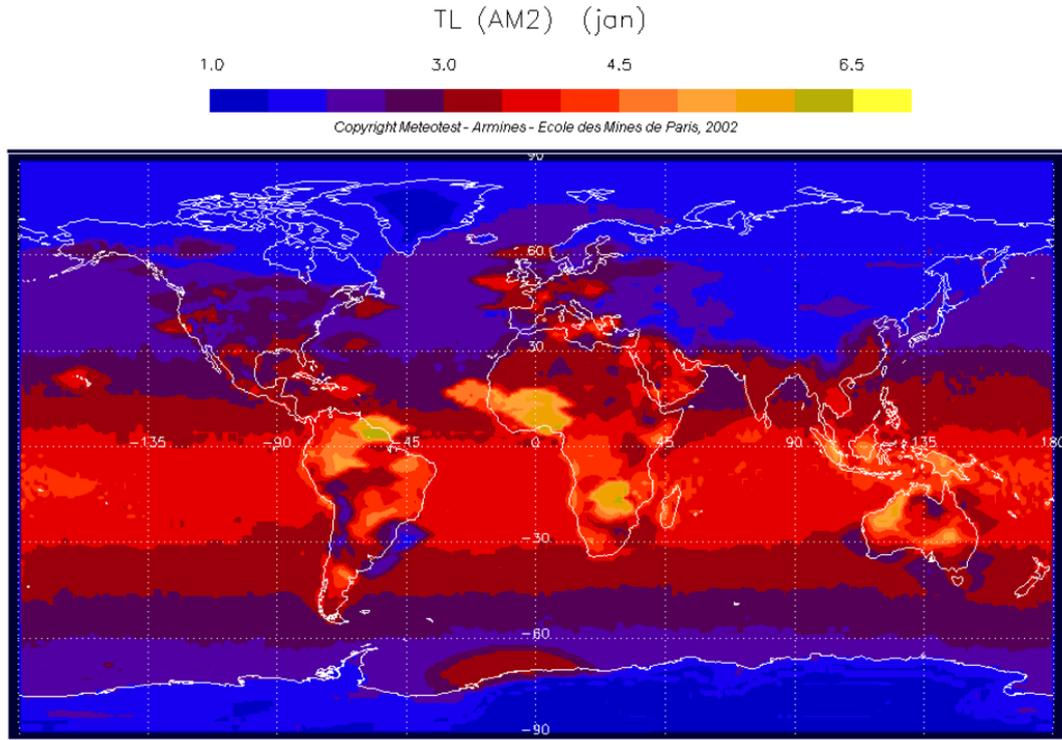


FIGURE 3.13: Global map of the Linke turbidity, T_L , to 5 minutes of arc angle for January 2002 taken from SoDa (2016). $T_L = 1$ indicates a perfectly clear and dry atmosphere. T_L increases as the presence of atmospheric aerosol and water vapour increases.

$$G_{D,cs} = G_0(0.0065 + (-0.045 + 0.0646L_{T,2} \cos \theta_z + (0.014 - 0.0327L_{T,2} \cos \theta_z^2)) \quad (3.52)$$

G_{cs} can then be found as the sum of its components

$$G_{cs} = G_{B,cs} + G_{D,cs} \quad (3.53)$$

The GHI at ground level, G , that incorporates meteorological attenuation of clouds is given by

$$G = k_c G_{cs} \quad (3.54)$$

3.3.7.2 Translating irradiance onto an arbitrary plane

There are numerous options for the translation of irradiance onto an arbitrary plane to estimate G_T . Gueymard and Myers (2009) found in a comprehensive comparison of

ten tilt models that the Klucher (1979) model, a relatively simple model, performed generally better than more complex alternatives. The Klucher (1979) model is an advancement on the Liu and Jordan (1960) model and can be expressed in terms of the three components of diffuse irradiance, G_D : horizontal, circumsolar and isotropic. The horizontal is the increased brightness towards the horizon, the circumsolar describes the forward scattering of radiation from aerosols, and the isotropic component that considers diffuse to be uniform intensity from all directions.

For use of the Klucher (1979) model, the decomposition of G into its G_B and G_D components is necessary in order to calculate G_T . G_B under all sky conditions is shown to be related to both k_c and $G_{B,cs}$ by using an adjustment by Müller and Trentmann (2010) such that

$$G_B = \begin{cases} 0 & k_c < \frac{19}{69}; \\ G_{B,cs}(k_c - 0.38(1 - k_c))^{2.5} & \frac{19}{69} \leq k_c \leq 1; \\ G_{B,cs}k_c & k_c > 1. \end{cases} \quad (3.55)$$

G_D can then be found as the difference from G and G_B .

$$G_D = G - G_B \quad (3.56)$$

G_B and G_D can then be translated onto an arbitrary plane

$$G_T = G_D \left(\frac{1 + \cos \beta_p}{2} \right) \times (1 + F \sin^3 \frac{\beta_p}{2}) \times (1 \dots + F \cos^2 \theta_i \sin^3 \theta_z) + G_B \frac{\cos \theta_i}{\cos \theta_z}, \quad (3.57)$$

where β_p is the arbitrary plane pitch, and F is a modulating factor calculated as

$$F = 1 - (G_D / (G_B + G_D))^2 \quad (3.58)$$

The final calculation required to arrive at G_T using the Klucher model is to derive θ_i . This can be achieved using θ_z , β_p , τ and the azimuth of the arbitrary plane β_a

$$\theta_i = \cos^{-1}(\sin \theta_z \times \sin \beta_p \times \cos(\beta_a - \tau) + \cos \theta_z \times \cos \beta_p) \quad (3.59)$$

3.3.8 Computational demand

The SIG is computationally friendly using the standard university issued desktop operating a dual core with 4 GB of RAM. To produce the 7 years of data in order to validate against the BSRN data sets, the model typically required 15 mins to complete. Reading in the data takes 2 mins (though once the first simulation is run, the data is stored for future). Markov production and preliminary calculations take less than 1 min. Creating the sun obscured time series took around 5 mins. The longest section is the irradiance calculations, which consume the remaining duration (excluding some output plots). This is fundamentally due to the need for storing each of the PSA model variables for each minute of simulation (of which there are over 3,679,200 data points per variable when operating for 7 years). This indicates that there is a larger than proportional response with increasing simulation duration. The model takes roughly 3 mins to produce a full year of irradiance (should the input data be loaded into Matlab).

3.4 Validation of the Solar Irradiance Generator

Two validations of the SIG outputs are carried out. The first using mean hourly GHI observation data from [MIDAS \(2015\)](#) for Leeds, UK. The second using 1-minute GHI observational data from [BADC \(2013\)](#) for Cambourne, UK.

3.4.1 Hourly validation for Leeds, UK

For the application of the model, 12 years of mean hourly weather observations for Leeds Church Fenton (Source ID: 533) were taken from the [MIDAS \(2015\)](#) data set to produce the MTMs. In order to validate the SIG, 12 years of radiation observation data were taken for the same monitoring station for comparison. Missing data points are assumed not to impact the validation significantly. In cases where duplicate measurements exist, the hour with the most recent observation time stamp is selected as is advised by [MIDAS \(2015\)](#).

Figures [3.14](#) and [3.15](#) illustrate a direct comparison between modelled and observational 10-year mean hourly irradiance. The diurnal representation in figure [3.14](#) shows a strong match between observed and modelled GHI for each season and annually. All GHI peaks are well matched, however for spring and winter there is a slight overestimation in midday irradiance. There are many possible causes of this slight overestimation. Firstly, the model can stochastically select better weather conditions than are found in the observation data for midday in spring and winter resulting in

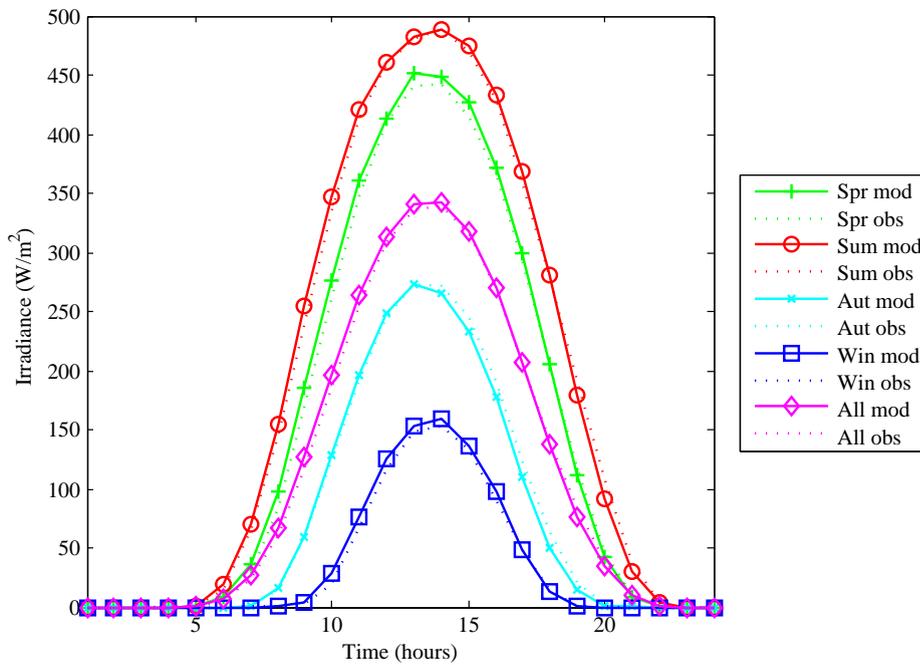


FIGURE 3.14: Comparison between modelled (mod) and observed (obs) mean diurnal irradiance profiles separated by season and the annual mean diurnal irradiance profile. The means are made from 10-years of modelled output and 10-years of validation data for Leeds, UK.

an overestimation. The stochastic selection of weather and $\mathbf{k}_{c,\text{clear}}$ is not expected to provide this overestimation as there were 10 years of spring analysed and so stochasticity provides accurate representation of a TMY. Secondly, the daily background $\mathbf{k}_{c,\text{clear}}$ taken from the $N(0.99, 0.008)$ distribution could randomly select higher values than occurred in observation data. Thirdly, the $T_{L,2}$ selected from the look up tables was taken for a single year were perhaps lower than was apparent in the observation data. Lastly, the final explanation offered is that of the error found in GHI measurement equipment. If the HELIOSAT method is to be trusted, and the distributions of $\mathbf{k}_{c,\text{clear}}$ be accurate, there are well known errors of spectral response and fouling of pyranometers that may place calibration errors in line with those presented here. It is difficult to attribute the overestimation.

The 10-year mean irradiance is indicated by the + scatter point at 112.5 Wm^{-2} observational and 113.5 Wm^{-2} modelled, giving approximately a 0.9% yearly irradiance overestimate for the location of Leeds. Whilst the intent of the model is not to produce hourly mean data in this format, averaging the minutely irradiance generated by the model over hourly time steps shows a strong correlation when compared with the observational data ($R = 0.9715$). This is considered an excellent result when considering the fact that the synthetic data that was not produced to recreate historic irradiance.

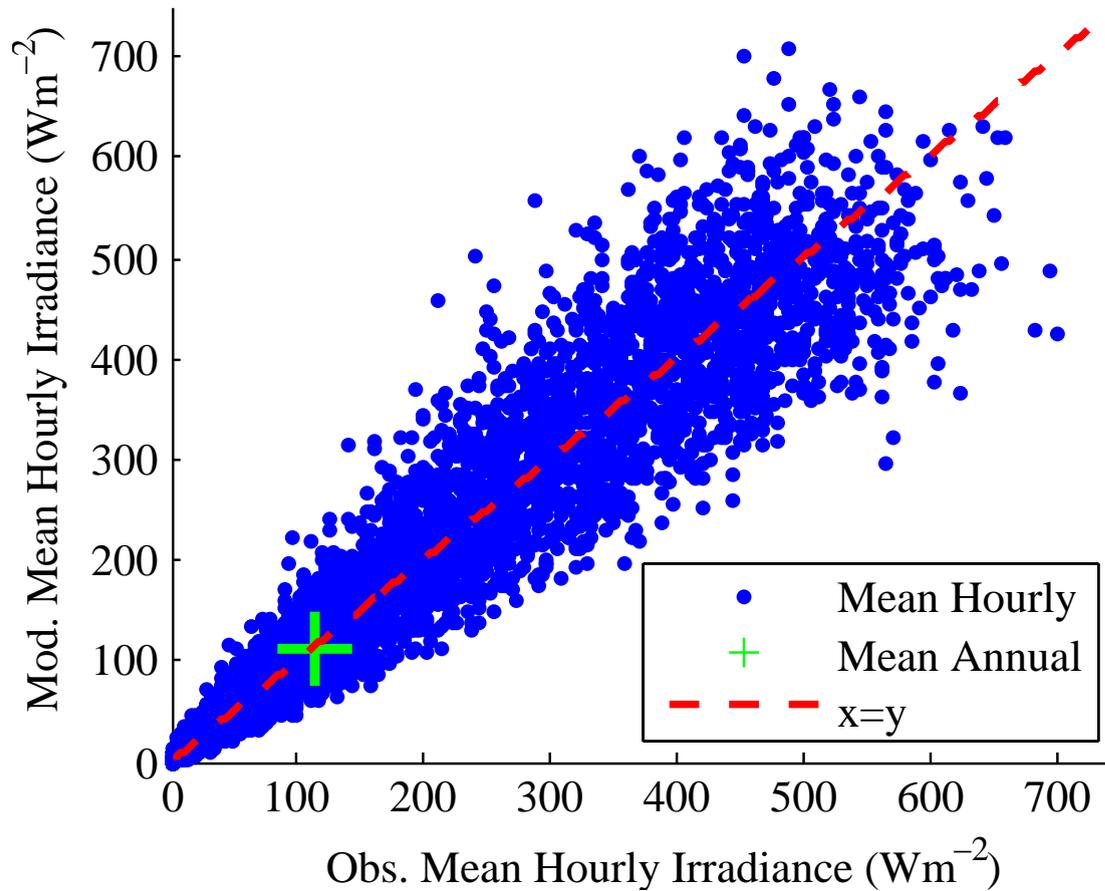


FIGURE 3.15: Comparison scatter plot for each 10-year mean hour of modelled output data against the corresponding hour of observation data. The yearly mean is also included. An $x = y$ line is included for reference.

Figure 3.16 displays mean hourly irradiance CDF distributions of both modelled and observed irradiance data. The correlation between the modelled and observational outputs is $R = 0.9963$, which demonstrates an exceptional fit for magnitudinal representation of the mean hourly irradiance. Whilst it is accepted that the hourly mean values are not suitable for grid impact analysis, what is demonstrated by this is that long term statistics of the synthetic irradiance generator are captured. Figure 3.16 is particularly telling of this as it demonstrates that the SIG captures a TMY.

For a more significant measure of performance, a minutely validation is performed.

3.4.2 Minutely validation for Cambourne, UK

Validation against a minutely irradiance dataset is necessary to confirm that the SIG can successfully generate statistically representative, 1 minute resolution irradiance time series.

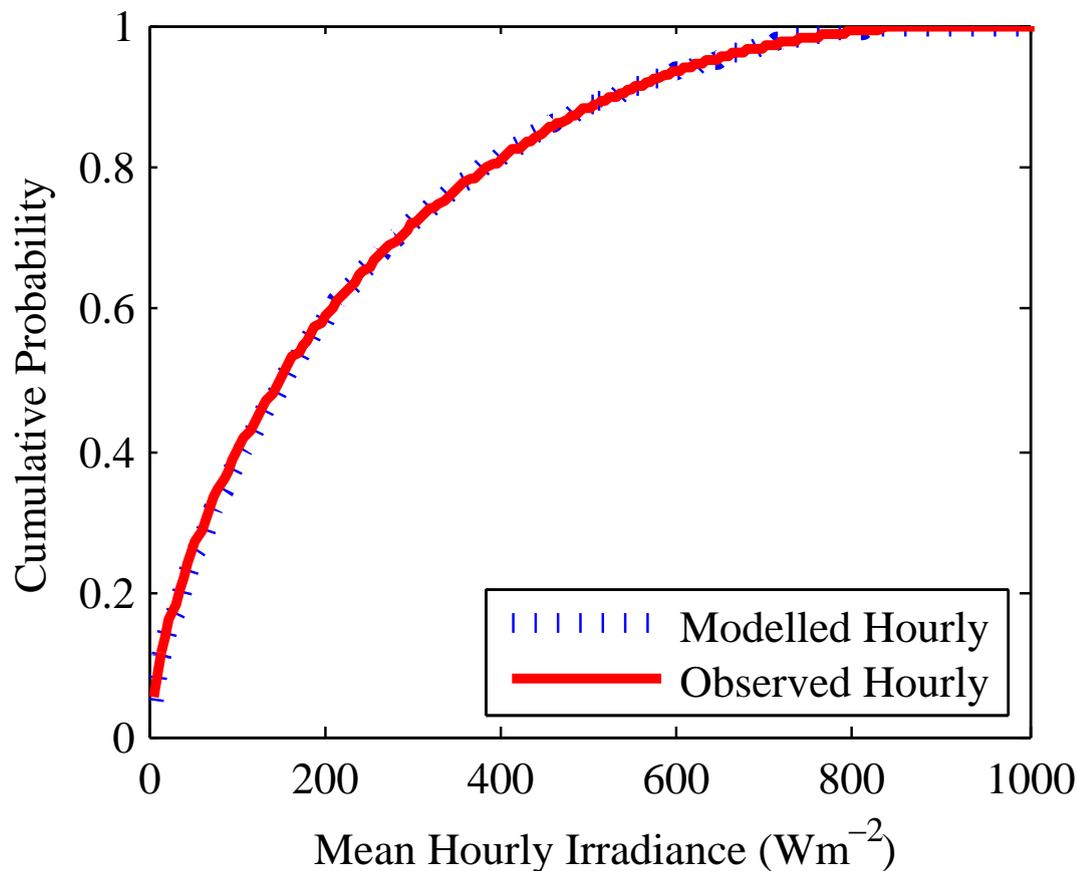


FIGURE 3.16: Comparison of the mean hourly irradiance frequency cumulative probability distribution profiles for 10 years of modelled and observational data from Leeds, UK.

To demonstrate this, 7 years of minutely radiation data was taken from the World Radiation Monitoring Centre - Baseline Surface Radiation Network (WRMC-BSRN) from BSRN station number 50, located in Cambourne, Cornwall, UK ([WRMC-BSRN, 2014](#)). Missing data points were ignored and deemed not to impact the distributions for comparison significantly. 12 years of mean hourly weather observations for the monitoring site at Cambourne (Source ID: 1395) were taken from [MIDAS \(2015\)](#) to produce the appropriate MTMs. The SIG inputs were adjusted for the location of the [MIDAS \(2015\)](#) Cambourne weather station using a latitude of 50.2178, longitude of -5.32656 , and height above sea level of 87m.

Three metrics are used to validate the intermittent nature of the SIG outputs: the variability index (VI), the irradiance magnitude frequency (IF) and the ramp rates (RR). A validation is made through both a comparison of CDFs through correlation checking and by using the 2-sample Kolmogorov-Smirnov (K-S) test for each metric, for each day of the year. They are described here in turn.

The VI is a metric created by [Stein et al. \(2012\)](#) used occasionally in literature to

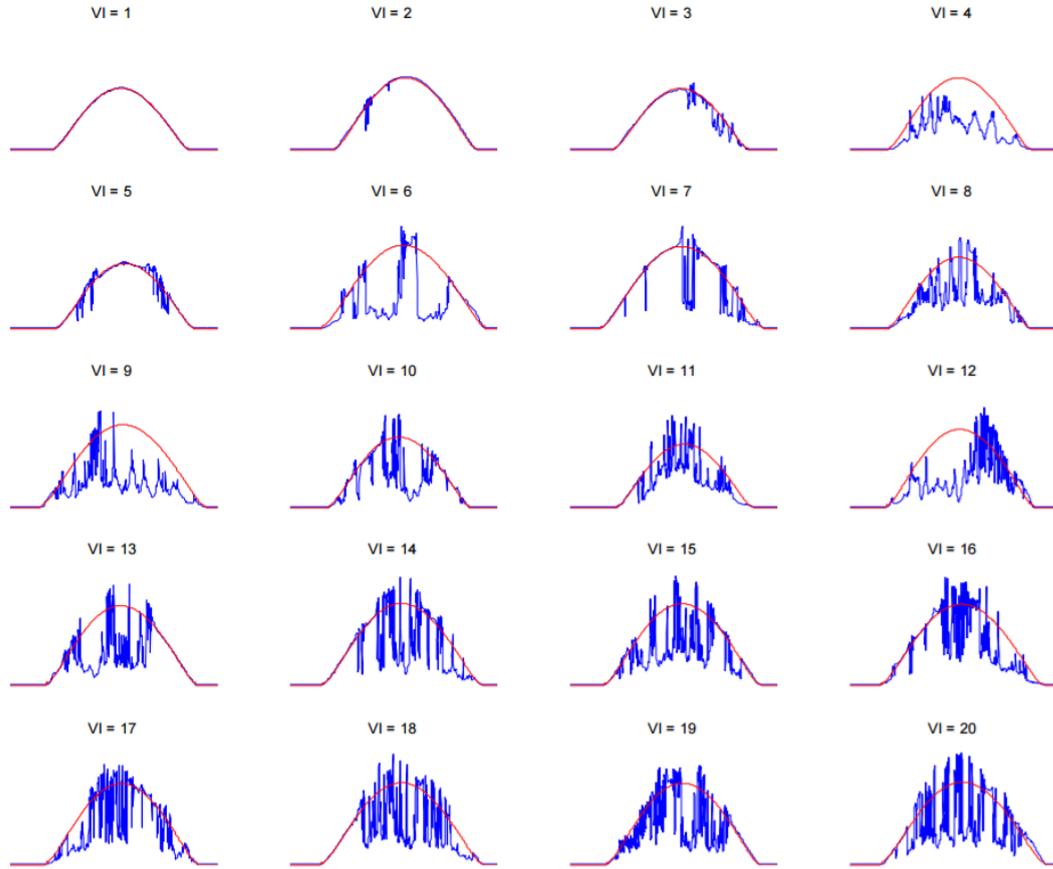


FIGURE 3.17: Examples of daily global horizontal irradiance time series is categorised using the daily mean variability index (VI). In red is the global horizontal clear sky irradiance and in blue is the global horizontal irradiance. The image was produced by [Stein et al. \(2012\)](#).

quantify the amount of solar variability over a set period of time, it is defined as the ratio between the rate of change of GHI and the rate of change of CSI. The more intermittent the day's irradiance, the higher the VI. It is calculated as

$$VI = \frac{\sum_{t=2}^T \sqrt{(G_t - G_{t-1})^2 + \Delta t^2}}{\sum_{t=2}^T \sqrt{(G_{cs,t} - G_{cs,t-1})^2 + \Delta t^2}} \quad (3.60)$$

Conceptually, the VI can be thought of as a comparison of smoothness between G and G_{cs} . Should G be a perfectly clear day and mirror appropriately to G_{cs} , then VI would be equal to or approaching 1. By nature of the equation 3.60 and the smoothness of G_{cs} , VI does not fall below 1. Figure 3.17 shows examples of different values of the VI.

The IF metric is a direct magnitudinal comparison of the full spectrum between observed GHI, and modelled GHI, over a specified time period and binned to the nearest integer.

The RR metric is a comparison of the ramps present in observation and modelled GHI. A ramp was defined earlier as the fractional change in output from one time step to the next. RR is > 1 when G increases between time steps, and is < 1 for decreases in G . It can be expressed as

$$RR = \frac{G_t}{G_{t-1}} \quad (3.61)$$

CDF comparisons are made using the inbuilt Matlab function of the 2-sample correlation coefficient (MathWorks, 2016) and displayed in equation 3.62.

$$R = \frac{\sum_x \sum_y (A_{xy} - \bar{A})(B_{xy} - \bar{B})}{\sqrt{\left(\sum_x \sum_y (A_{xy} - \bar{A})^2\right) \left(\sum_x \sum_y (B_{xy} - \bar{B})^2\right)}} \quad (3.62)$$

Where R is the correlation coefficient, \bar{A} and \bar{B} represent the mean GHI within the respective time series A and B , and A_{xy} and B_{xy} are individual GHI values within the time series A and B . R is equal to 1 when the correlation is perfect, -1 when the correlation is perfectly negative, and 0 when no correlation is present.

The 2-sample K-S test was carried out for each of the three metrics. The K-S test is a non-parametric evaluation of the difference between two CDFs. It tests the absolute difference in probability for all x-axis bins of the CDFs for the two samples, this difference is then checked against the pre-defined confidence limit to determine the success of the test. The K-S test is provided as an inbuilt function in Matlab with both CDFs and the confidence limit as inputs (MathWorks UK, 2016). The test returns two values. The first is a logical indicator if the test has passed or not. A pass is defined as the probability of any y value in the CDF being within the confidence limit. The second returned value from the K-S test is called the asymptotic p-value, defined as the probability of observing a test statistic as extreme as an observed value that would fail the test. The p-value is always between 0 and 1. 0 indicates an impossibility for the point within the CDF to manifest as a value outside of permissible range. 1 indicates a certainty that the value is always outside of the permissible range. To pass the K-S test, the p-value must be below the confidence limit set. Therefore, for a day of modelled GHI to pass against observed GHI, the modelled p-value must be within the confidence limit of the observed p-value.

In order to perform a K-S test, an appropriate sized sample is required. As the intent is to derive the minutely significance of data, a yearly subset would be too large to discern overly meaningful statistics of smaller time scales. As a single day is the smallest simulation duration permissible by the SIG, a day will be used to define the subset size. The results will therefore represent the minutely resolution success over

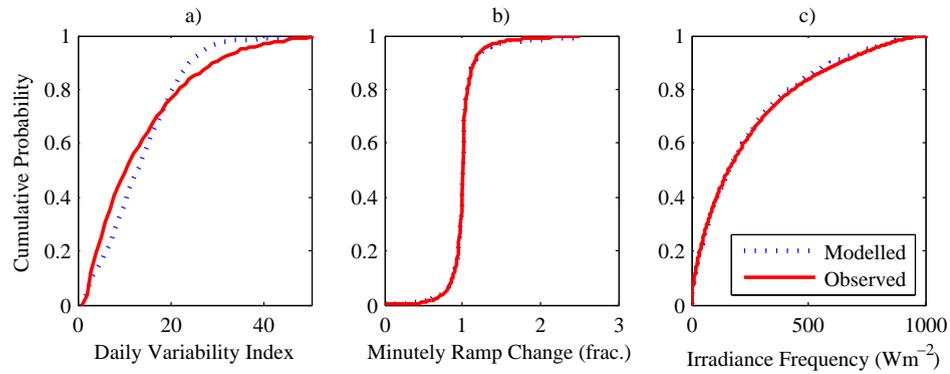


FIGURE 3.18: Cumulative probability distribution plots of the (a) daily variability index, (b) the minutely irradiance ramp-up and ramp-down occurrences, and the (c) minutely global horizontal irradiance frequencies for both the modelled data and the observed data from Cambourne, UK.

the course of a day, it is worth noting that the test is only performed on the portion of the day when irradiance is present between sunrise and sunset, or when $\theta_z < 90$, and so a day's data does not contain the full 1440 minutes of G . As the synthetic data is not a historical representation or reproduction, a single day of modelled data is not expected nor intended to align with observed data. For this reason, each day selected from the subset is combined with all its corresponding days from successive years, and so the GHI data is binned by calendar day. 7 years of observation data for Cambourne is available from the BSRN between 01/01/2007–31/12/2013, and so 7 of the same day from successive years is used for the K-S subset. For example, a single subset contains 7 versions of the 1st of January from 2007 to 2013. 365 subsets will be produced (ignoring 29th February, which only occurs twice in the specified range) and so success is reported in the number of days that passed the K–S test out of 365.

To carry out the K-S test on the VI, the 1 minute values of VI were taken as the subset, this differs to the daily value demonstrated in figure 3.17, however equation 3.60 is not limited by time step and 1 min is suitable. The comparison of the SIG output and the observed data CDFs for VI are shown in figure 3.18a. The two CDFs correlate with a correlation coefficient of $R = 0.9903$. The model generated VIs have an increased frequency of mildly variable days between variability indices 10 to 25, as is indicated by the steeper slope, whilst having a slightly reduced frequency of extremely variable days and extremely stable days.

Table 3.6 indicates that 345 of the 365 days passed the 2 sample K-S test when using a significance level of 99%. Although 20 days seems significant, it equates to 94.52% of days being recreated with a stringent confidence limit. The fact that 5 days failed the test with a 90% confidence suggests an imbalance between the statistical likelihood of the meteorology. The observation data possibly manifests weather that is considered

extremity, or away from the TMY. The weather generator can reproduce these periods, however not on the required day to validate against observation data, and so any attempt to validate against an uncommon day of weather will result in failure of the K-S test. No annual temporal bias was observed from the days that fail the K-S test, such that the failed days were randomly positioned throughout the year. This compliments the point that the synthetic data is not intended to recreate historic data, yet statistically captures the minutely variability in 94.52% of days. There was a correlation to the days that failed that happened to have more cloudier days that is statistically likely at that time of year. The stochastic model captures the TMY through aggregating the discreet Markov transition probabilities from 12 years. Therefore, a single day of coincidental cloudiness would be considered extreme when comparing in this calendar confined manner, and so would regularly fail the K-S test for that day.

The same K-S subset production approach is further applied for the RR and IF metrics. The comparison of the RR CDF, as shown in figure 3.18b, has a correlation coefficient of $R = 1$. Both the 1-minute interval of ramp-up and ramp-down events are captured excellently using the SIG. This is furthermore demonstrated with the results from the day-by-day K-S tests shown in table 3.6. All days tested passed with a significance level of 99% with mean asymptotic p-value of $6.7365e^{-29}$.

Figure 3.18c shows the comparison of the IF CDFs. The correlation coefficient is $R = 0.994$ with a very slight underestimation of the mid-range irradiance occurrences. A possible cause for this is assigning okta values 1–6 with k_c values derived from the distribution of $N = 6$. It is expected that because the difference in $k_c(N)$ distributions shows a shift from very low k_c at $N = 8$ to medium k_c at $N = 6$, that perhaps as $N \rightarrow 1$ the k_c values also approach 1. This implies that by assigning clouded periods during $N = 1$ from the $k_c(N = 6)$ distribution, an underestimation in irradiance would be observed. The K-S test on the IF shows that over 350 days passed the test which is equal to 95.89% at a significance level of 99%.

Figure 3.18 cannot be compared with literature benchmarks, as there are no 1-min competitors when it comes to synthetic solar irradiance generators. There are a few higher resolution synthetic irradiance generators. Figure 6 by Ngoko et al. (2014) displays the reproduction of CDF comparisons of both the clear-sky index and IF from their stochastic model using Markov chains. Their methodology is flawed, such that their validation data feeds the input of the model, and yet their reproduction of data fits neither of the two input high resolution time series; the SIG outperforms this model. There exist comparisons of interpolation methods of increasing temporal resolution of solar irradiance data from figure 8 by Fernández-Peruchena and Gastón (2016) who successfully recreate 1-min IF with high accuracy. Their use of PDF over

K-S test Significance Level	90%	92.5%	95%	97.5%	99%
Variability index	360	358	357	352	345
Ramp rate occurrence	365	365	365	365	365
Irradiance frequency	363	360	359	356	350

TABLE 3.6: The number of days (out of 365) that pass the 2-sample Kolmogorov-Smirnov test on 7 of the same calendar day from successive years of both modelled and observational CDFs containing the minutely variability indices, minutely ramp rate occurrences, and the minutely irradiance frequency, tested at increasing confidence limits.

CDF makes like-for-like comparisons to the SIG difficult, particularly as no empirically derived statistical comparison is made in their work; visually our works are comparable. Perhaps more interesting is the work on VI by the creator of the metric, figure 12 by ? where they demonstrate the 1-min ramp rates for different groupings of cloud type. Whilst the overall CDF is not presented, only a clip of CDF probabilities from 0.95 to 1, this is directly comparable to figure 3.18 as it shows how the fractional ramp change curves from no noticeable ramp to a 70% at similar probability magnitudes. The SIG does not have the capabilities to separate out cloud types to directly compare, however, the similarities are clear. The closest competitor to the SIG is [Ehnberg and Bollen \(2005\)](#) who used a simple Markov model to synthetically simulate irradiance values, however they made no attempt at validation. For these reasons, the SIG validations in figures 3.16 and figure 3.18 can be considered benchmarks.

3.5 Discussion of results

This section will discuss the implementations within the SIG and analyse their significance.

A distinctive element of the SIG is the use of 20 different MTMs in order to capture seasonal, diurnal and pressure based variations. These will be discussed in turn, as will the irradiance and ppotential research opportunities.

Combined, the 20 MTMs successfully replicated the okta frequency distribution of the 12 years of observation data. The mean percentage error between the modelled and observed okta transitions was -0.03% , calculated as the mean of the percentage difference for each okta state. Furthermore, a comparison of CDFs of the modelled and observed okta frequency had a correlation coefficient of $R = 0.9956$. This demonstrates that the overall statistics for Leeds were retained using 20 separate MTMs, whilst also capturing more detailed transition characteristics at certain times of a year.

Each type of okta MTM (diurnal, pressure, and season) is analysed through examining the deviation away from the annual mean okta MTM (shown earlier in this chapter in figure 3.8) to assess the impact each of the different types have on the MTM probabilities.

The mean okta MTM is produced accounting for every transition in the observation data, and is not separated by season, pressure or time of day. It is considered the baseline for this type of methodology, as it has been demonstrated previously by [Ehnberg and Bollen \(2005\)](#).

3.5.1 Diurnally weighted okta Markov transition matrices

The most significant deviation of transition probability away from the mean okta MTM is seen with the morning diurnal dependency MTM. A comparison of the mean of the diurnal MTMs (so that variations caused by pressure and seasonal differences are excluded), minus the mean okta MTM shown in figure 3.19. The differences in probability below ± 0.01 are removed from the plot so that only the most significant deviations from the mean are coloured.

There is a very distinct pattern of the probability for an okta value to remain the same from one state to the next, with a decrease of between 0.01 and 0.25 for 0 to 6 okta. The chance of $N_t = N_{t-1}$, or the okta state remaining the same, is defined by the author as *stability*, as it represents no change in N . *Instability* is defined as the increased chance to move away from the current state, $N_t \neq N_{t-1}$. 7, 8 and 9 okta have an increased probability of remaining the same, and so cloudier weather states will tend to last longer. This can be viewed as a tendency for instability at lower okta states, and the reverse at 7–8 okta.

Significant probability increases for all okta states to transition to a higher okta state are observed. This is indicated by the continuous blue strip seen to the right of $x = y$ (note the $x = y$ direction is from top left to bottom right). The most significant increase is seen for 0 okta, where an increase towards 3 through 6 okta is observed at the expense of stability. Interestingly, there is also a slight increase for 3, 4 and 8 okta to transition down to 0 okta. This diurnal MTM could be described as capturing the tendency for scattered clouds to either clear or thicken.

The physical implication of this is that during the morning period (00:00–05:00am), the current okta state has an increased chance to transition towards an okta state of 6–8 or 0, and so a clouded state is most likely to develop into either fully obscured or complete clear sky by 6am. This increased tendency towards both okta extremes

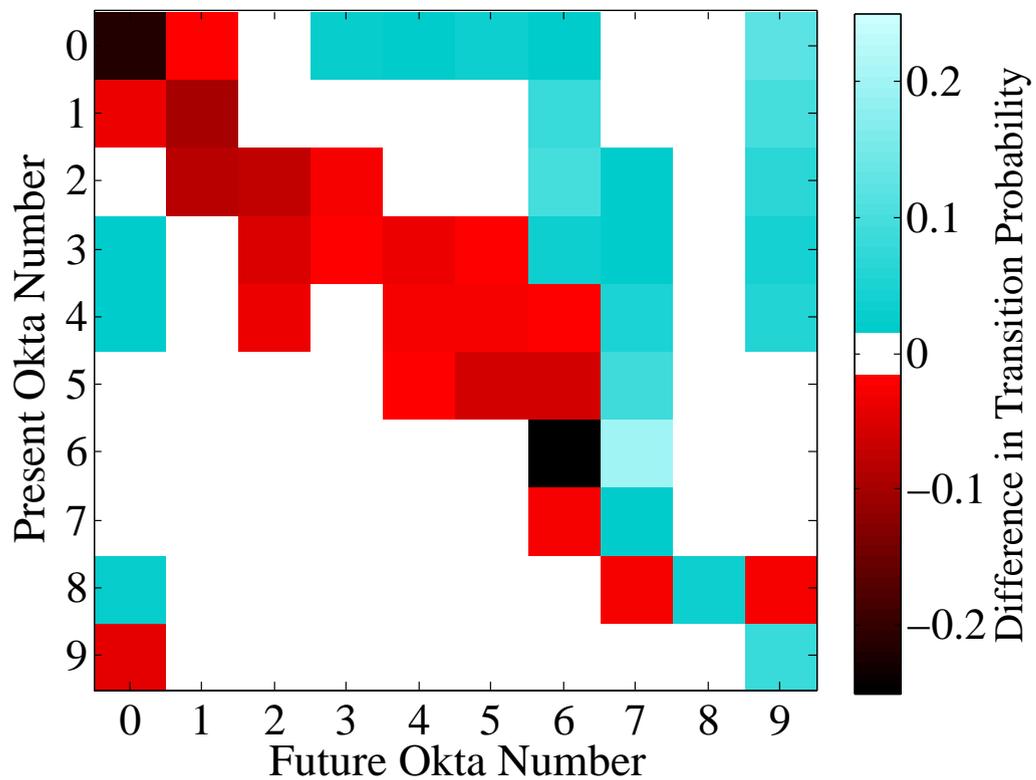


FIGURE 3.19: Colour plot indicating the deviation of the diurnally subjective okta MTM minus the annual yearly mean okta MTM. White space indicates a $\leq \pm 0.01$ change in transition probability.

away from stability during the morning period was described in section 3.3.3 as the unstable atmosphere with the introduction of the sun's heat. This phenomenon cannot be captured using only the mean okta MTM.

3.5.2 Pressure weighted okta Markov transition matrices

The reason for the inclusion of transitions based on pressure is to attempt to capture variations in weather that are caused by high and low pressure systems. The approach is simplified to be either above or below the average pressure.

Figure 3.20 shows a slightly different analysis than was presented in the diurnal comparison for simplicity of direct comparison, and so does not use the mean okta MTM. The mean below average (BA) transition probabilities are subtracted from the mean above average probabilities (AA). This means that blue squares represent occurrences where the BA transition probability is greater than that of AA, implying that *blue transitions are more likely in BA pressure systems*. The green squares represent where

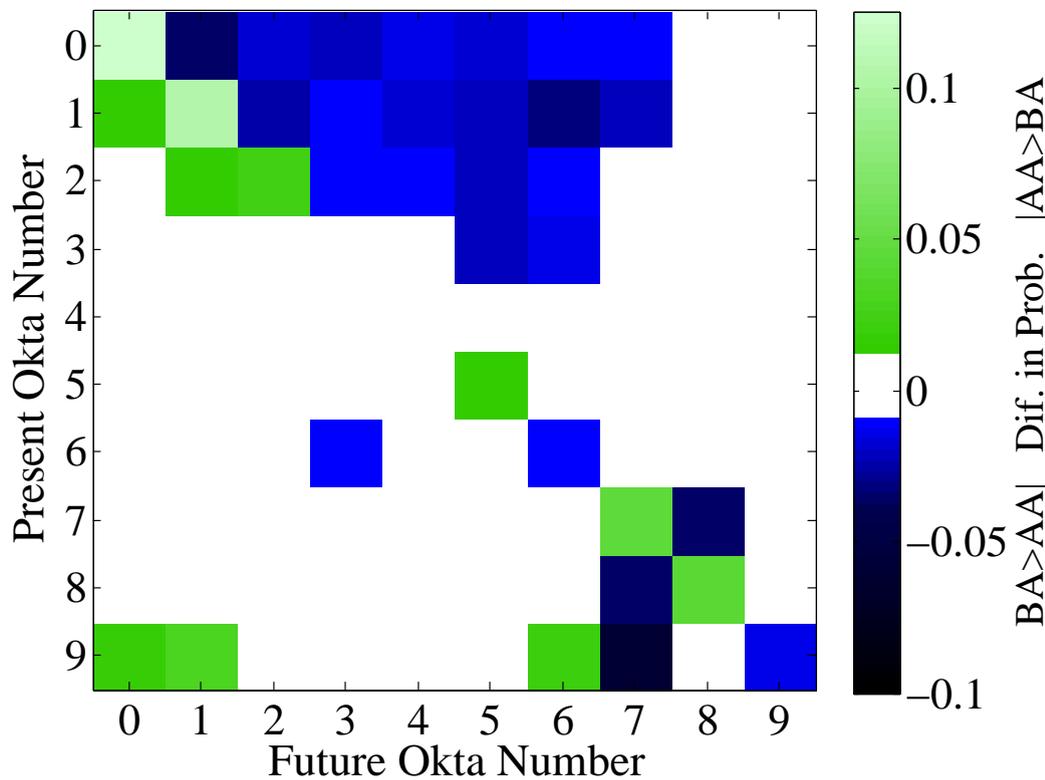


FIGURE 3.20: Colour plot indicating the difference between the above average pressure MTM (AA) minus the below-average pressure MTM (BA). White space indicates a $\leq \pm 0.01$ change in transition probability.

AA transition probability is greater than that of BA, implying that *green transitions are more likely in AA pressure systems.*

The most distinctive observation is that the probabilities for 0–3 okta to transition to a cloudier state are consistently and significantly greater during periods of BA pressure. This supports the rationale that low pressure systems have increased unstable weather conditions and so clear skies are much more likely to develop into cloudier skies. This instability is further demonstrated with 7–8 okta having a higher tendency change between themselves, suggestive of the complete overcast periods to be more likely a function of larger rain clouds than they are stable overcast conditions.

The transitions are more stable during AA pressure as is demonstrated with an increased probability in the $x = y$. 1 to 2 okta under AA pressure conditions have a slight increase in probability to become less cloudy, moving towards clear sky. This is representative of typical high pressure system weather of completely clear. There is also the increased tendency for 7 and 8 okta remaining the same during AA pressure conditions. This could be attributed to the very high cirrus clouds that occur during

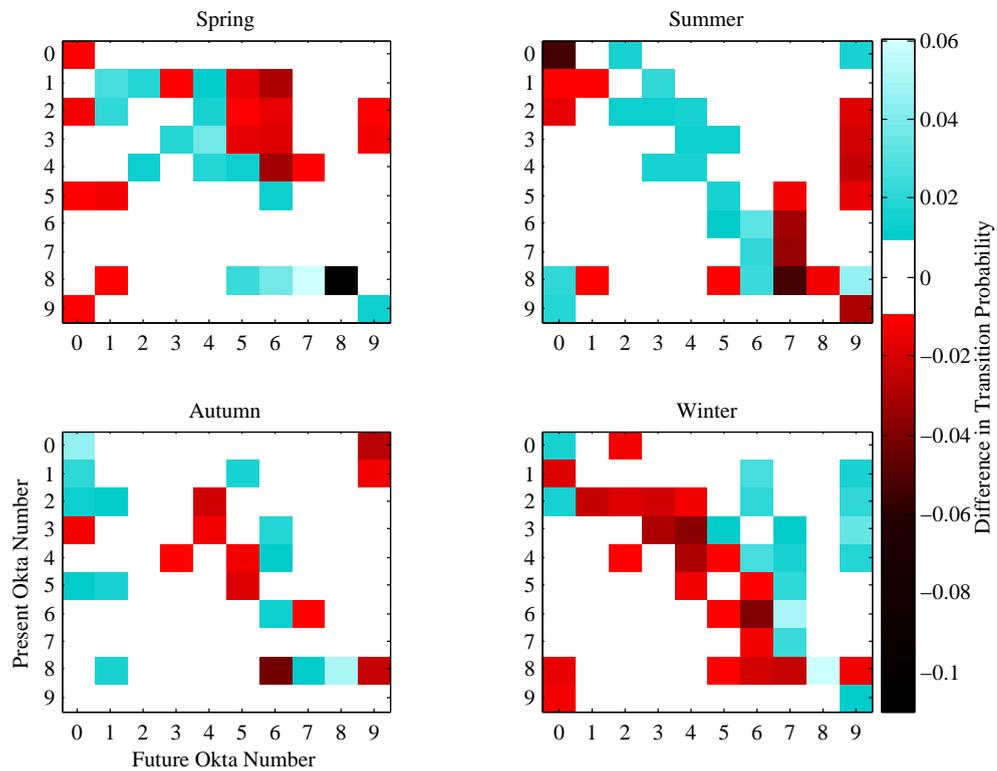


FIGURE 3.21: Colour plot indicating the deviation of the individual season MTMs from the yearly mean okta MTM. The colour indicates the difference in transition probability, whilst white space indicates a $\leq \pm 0.015$ change in transition probability.

high pressure systems that can form fully overcast sky. When these conditions occur, they have an increased stability, whereas overcast conditions dissipate over time.

3.5.3 Seasonally weighted okta Markov transition matrices

The variation in N transition probability due to the different seasons is detailed in figure 3.21, where the deviation from the the mean okta MTM is shown for each season. In order to observe only significant differences, deviations of ± 0.015 are removed. Each season will be discussed in turn.

Autumn variations are the closest to the mean okta MTM. The frequency of occurrence of 2–6 okta are the lowest in observation data and, therefore, differences in this range offer the least impact on the overall SIG output. There is a tendency for states to move to either 0–1 or 6 okta, as is indicated by the blue squares. There is a significant increase in stability for 0 and 8 okta. This physically represents clear and cloudy periods lasting longer, perhaps typical of longer duration weather systems.

Spring transitions are similar to those of autumn in that the most significant deviations occur during the least sensitive okta range. There is, however, a distinct shift to favour stability between medium okta states of 1 and 5. There is a small reduction of ~ 0.11 in probability of remaining clear sky, and a very significant reduction in the probability of remaining fully overcast. This suggests that spring is more stable at states 1–7, yet less stable at 0 and 8 okta. The interpretation is increased amount of scattered cloud and fewer periods of clear sky and over cast. It is worth noting, however, that the transition ($M_{t-1} = 0|M_t = 0$) has a probability of 0.7157 within the mean okta MTM, the reduction seen in spring only lowers the probability to ~ 0.7 , which is still considerable. ($M_{t-1} = 8|M_t = 8$) has a mean probability of 0.7179, which is reduced to ~ 0.605 , presenting a more significant reduction.

Summer and winter contain the most variations away from the mean okta MTM. Summer is the most important month as it is the season receiving the majority of yearly irradiance. Summer sees an increase in the probabilities along $x = y$, suggesting stable meteorology; this is in direct contrast to the pattern observed in winter. Interestingly, winter has an increased tendency to move towards the two extremes of 0 and 8 okta while summer does not. Intuitively, it is expected that summer would contain the most clear sky to clear sky transitions on account of more high pressure systems, however, there is a large reduction in ($M_{t-1} = 0|M_t = 0$) that cannot be directly attributable to probability increases seen towards 2 and 9 okta. This is because all other transitions from 0 okta to all states undergo a slight increase that is below the cut-off limits on the figure, and so are not reported as significant. As discussed, the probability of ($M_{t-1} = 0|M_t = 0$) is still considerable despite the reduction. Winter sees the most intuitive changes. States of 1–7 okta all have an increased tendency towards a cloudier state, as expected from a more adverse weather conditions during the season.

It is worth noting that the differences reported here are specific to the location of Leeds, whilst the differences may be similar in nature, the seasonal, diurnal and pressure differences are specific to the input observational data, such is the motivation of having a geographically flexible model.

3.5.4 Discussion of the synthetic irradiance

Hourly averages of k_c derived from the SIG were compared with k_c values from the MIDAS (2015) data as is shown in figure 3.22. Note how there is a bi-modal set of peaks in the PDF of k_c . The two peaks are attributed to cloudy periods and clear periods.

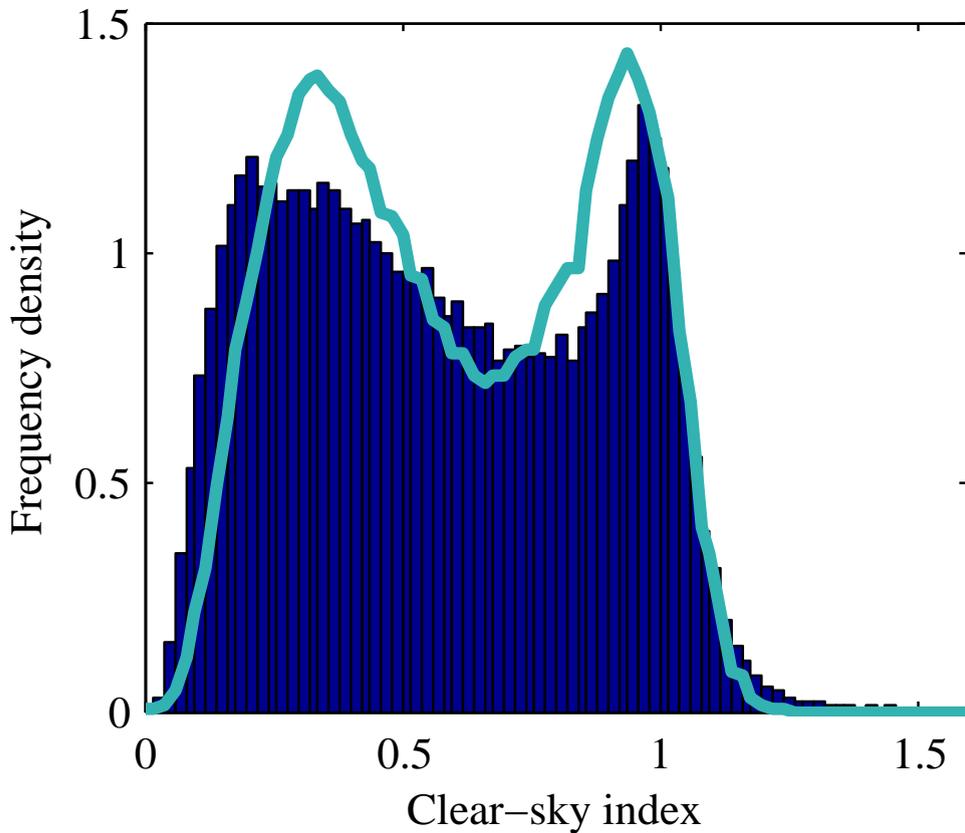


FIGURE 3.22: Histogram of the mean hourly observed values of k_c (bar) compared to the values of k_c from the SIG (line).

It can be seen that the SIG recreates the bimodal structure of the real-world distribution of clear-sky indices. The peaks from the simulated data are of similar height to the MIDAS data, although the intermodal spread is lower and, in particular, there are fewer extreme high or low values in the simulated data. The use of $k_c(N = 6)$ for $N(1, 2, 3, 4, 5, 6)$ is the most likely cause for the misalignment of the peaks. Using this distribution skews the frequency of k_c values towards the $k_c(N = 6)$ distribution, which is a normal distribution of $N(0.6784, 0.2046)$. Appropriately assigning individual distributions to all N states may potentially introduce a more aligned intermodal spread.

Figure 3.23 displays the diurnal root mean square error (RMSE) of the simulated data for each season and annually, averaged across 10 years. The RMSE is calculated as a direct comparison of 10 years of modelled mean hourly irradiance data aligned with 10 years of mean hourly observational data, it is found as

$$RMSE = \sqrt{\frac{\sum_{d=1}^D (\hat{G}_d - G_d)^2}{D}} \quad (3.63)$$

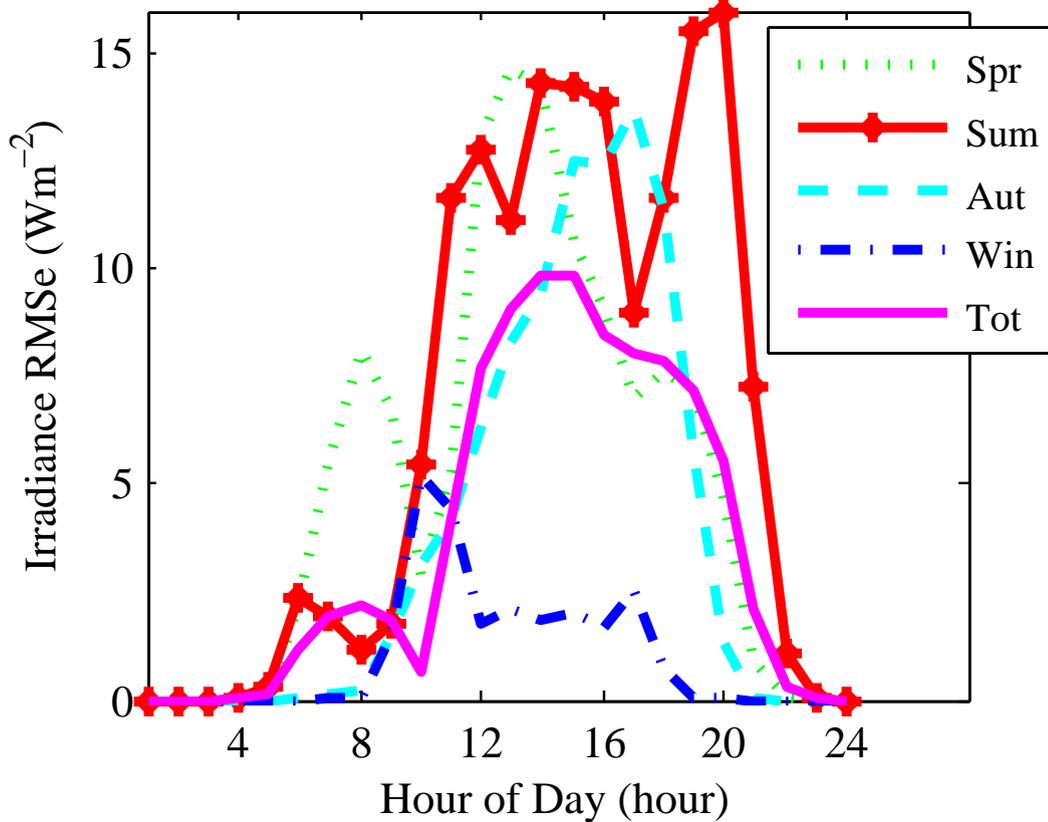


FIGURE 3.23: The diurnal mean hourly irradiance root mean square error (Wm^{-2}), averaged across 10 simulations. The RMSE is displayed diurnally and separated by season, the yearly diurnal RMSE is also shown.

Where \hat{G}_d is the modelled value of the irradiance and G_d is the observed for each of the same hour of the day, for the total number of days that exist within that season, across all 10 years, D . The output represents a scale-dependent aggregate of all the error magnitudes into a single value.

The time of day of most importance is around midday during the summer season, as it has the most clear-sky irradiance and, therefore, the potential for the largest ramp-rates and peak outputs. Typical clear irradiance outputs during these times are conservatively around $900 Wm^{-2}$ at Cambourne. The RMSE in figure 3.23 at midday in summer is at $14 Wm^2$, offering potential mean hourly irradiance error of $\pm 1.5\%$. The majority of hourly RMSE values fall below $15 Wm^2$.

There is also a tendency for increased error at sunset, when compared to sunrise. This could potentially be due to a sun-Earth geometry alignment with the time stamp, or perhaps the time within the minute at which irradiance is recorded.

3.6 Chapter summary

The aim of this chapter was to ascertain whether a synthetic irradiance generation methodology could be developed that can produce a high temporal resolution, statistically accurate irradiance time series using only readily available, geographically dispersed, mean hourly meteorological observations as an input. In order to achieve this, the SIG was developed following these criteria.

This chapter presented the SIG, a methodology for a stochastic, sun obscured type, temporal-only, synthetic irradiance generator with a temporal resolution of 1 minute. One-dimensional cloud samples were generated so that the horizontal cloud length follows a power law relationship. Twenty different MTMs were used in setting the hourly weather variable conditions and were used to determine the 1 minute indication of the state of sun obscured. These MTMs accounted for seasonal, diurnal and pressure variability. Distributions of the clear-sky index by okta value were developed and implemented. Further methods were adopted to generate clear-sky indices that enable 1 minute intermittency and variability. Typical outputs can be seen in figure 3.24.

The method was applied to the city of Leeds, UK, and validated using independent hourly radiation measurements from the same site. Furthermore, a 1 minute resolution validation was carried out using irradiance data from the town of Cambourne, UK. Metrics of variability index, ramp rates magnitudes and irradiance magnitude frequency were shown to be captured excellently with the SIG when tested to 99% confidence limits using the 2 sample Kolmogorov-Smirnov test.

It can be concluded that it is possible to generate 1 minute resolution irradiance time series from readily available 1 hour weather data that capture the statistics of 1 minute irradiance observations. The geographic dispersion and spatial separation will be further investigated in chapter 4.

3.6.1 Key points, applicability and potential development

The use of diurnally representative MTMs was shown to have the most significant deviation from conventional methods of stochastic cloud amount generation. The difference in okta transition probability due to pressure system and season also allows for greater accuracy in capturing the weather conditions of a geographical location.

The distribution of hourly clear-sky indices is recreated, whilst allowing for minutely clear sky index fluctuations. There is potential for extending the research between the clear-sky index and cloud amount.

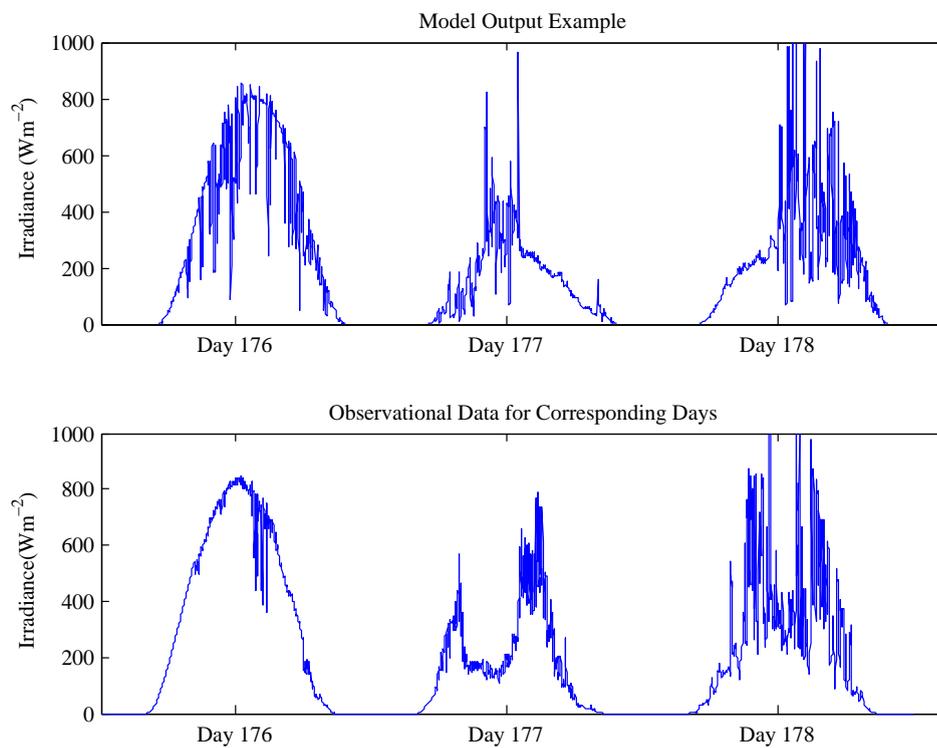


FIGURE 3.24: Three days of example irradiance profile outputs from the SIG (top), and from observational data (bottom). Note this is neither a predictive method nor a forecasting model and so the synthetic profiles are not intended to match or recreate observational data. The purpose of this figure is to demonstrate the SIG's output and to show real 1-minute resolution observational data irradiance profiles, the days 176 to 178 were selected purposely as they presented similar corresponding characteristics.

Individual simulations at nearby locations using the SIG do not correlate due to the non-spatial nature of the methodology. The SIG output is a synthetic global solar irradiance time series upon an arbitrary plane at a 1-minute resolution. It is a temporal data series only, and does not include a spatial dimension. Its applications are therefore limited to cases where the spatial element is not integral, such as small scale studies where a single high-resolution irradiance data series input is ideal. The SIG has suggested application, therefore, in the improved modelling of small-scale PV supply, demand, and storage systems, calculating electricity supply on a time scale that better matches the demand flows. Regular demand flows operate with high power across a small duration, meaning that mean hourly averages fail to capture electricity peaks in the supply and demand. To appropriately capture electricity flows within the residence that can accurately calculate efficiencies, self-consumption losses, battery charge and discharge states, an appropriately high-resolution time scale is required (Torriti, 2014; Darcovich et al., 2015).

The SIG is not theoretically limited to a 1 minute resolution, although for confidence,

higher quality resolution observation data would be required for analysis and validation. This will not be considered within this thesis as 1 minute is already a large gain on the literature, more pressing is that the SIG could be extended through the inclusion of a spatial dimension. A spatial dimension is crucial for aggregated grid impacts analysis and, therefore, forms the basis of the next chapter.

Chapter 4

Spatially Decorrelating Solar Irradiance Generator (SDSIG) development

The research problem addressed within this chapter asks whether or not it is possible to produce statistically accurate, synthetic irradiance time series that vary on both a temporal and *spatial* dimension, and that facilitate multi-variate grid impact analysis.

In order to answer this question, the first objective presented in this chapter is the development from the SIG to become a model that can produce irradiance time series that vary on a spatial dimension whilst continuing to use only readily available, well geographically dispersed, mean hourly meteorological observations as an input. This development is treated as a new model and is referred to as the *Spatially Decorrelating Solar Irradiance Generator* using the acronym SDSIG. The term *decorrelating* is used as it captures the concept that the irradiance at two different locations are different yet related. The alternative is to use the term *correlating*, however, the author feels that this should be used to describe perfectly matching irradiance time series in use at different locations. Therefore, they would perfectly correlate. The second objective presented in this chapter is the demonstration of the SDSIG's applicability in grid impact analysis by exploring the grid impact of voltage fluctuations, using the metric of OLTC overuse, as was proposed in section 2.5.

This chapter will first introduce the concept of spatio-temporal solar resource generation through a review of the most recent literature, highlighting alternatives and opportunities for development. The SDSIG will then be described in overview so that the reader may follow the steps and have a reference; each significant step will feature

its own section. The methodology discusses and summarises the elements of the SIG that are retained. Furthermore, the SDSIG is provided in its raw coded script, as produced using [Matlab \(2015\)](#) software. It is appended to this thesis and is separated into sections that correspond well to this chapter, see section [B](#).

The preliminary development concepts of the SDSIG have been accepted in a special issue of the Journal of Solar Energy for International Solar Energy Society's Solar World Congress 2015 (SWC2015) (?), with expected online availability 2 years after the conference in November 2017. The fully validated and finalised version of the SDSIG presented in this chapter was condensed and is published in the Journal of Solar Energy ([Bright et al., 2017](#)). Research into k_c distributions by N was undertaken to further analyse k_c dependency on both N and θ_e , this work was published in [Smith et al. \(2017\)](#), of which I am a co-author, and will feature in section [4.2.4.1](#).

The application of the SDSIG onto a theoretical schematic of a section of the LV grid was produced in collaboration with Oytun Babacan under the supervision of Prof. Jan Kleissl while undertaking a research scholarship to the University of California, San Diego (UCSD). The collaboration successfully combined the SDSIG with Oytun Babacan's power flow model to analyse the impact of using decorrelating irradiance time series on OLTC operations compared to using correlating irradiance time series. Section [4.5](#) contains collaborative work.

4.1 Literature review of spatio-temporal solar irradiance generators

Chapter [3](#) discussed the one-dimensional, temporal solar irradiance generator methodologies that exist in literature. This chapter will not reproduce the lessons learned from section [3.2](#) and will only focus on the introduction of a spatial element to solar irradiance generation.

An important distinction and definition must be made by the author that limits what is, and isn't, considered an application of spatio-temporal solar irradiance generation. Performing a study using perfectly correlating irradiance across a spatial domain is considered by the author *not* to be a spatio-temporal application. This is because the time series do not consider the *influence* of space between sites, despite the simulation physically including this spatial separation. Therefore, only methodologies that consider the influence of space are considered by the author to be spatio-temporal applications; this opinion is shared by [Harvill \(2010\)](#).

Spatio-temporal solar irradiance generation can be achieved through a variety of methods that use many types of available input data in order to generate irradiance time series at two or more locations. Whilst single irradiance time-series like those from the SIG are useful for studies that consider only a single input, such as the supply and demand electricity flows of a photovoltaic connected house, e.g. the transient PV system model by [Patsalides et al. \(2016\)](#), one-dimensional correlating solar irradiance data are only suited to a small spatially-sized applications. Much of the reviewed spatio-temporal research that explore grid impacts from PV intermittency resort to using correlating irradiance time-series, despite a need for decorrelation ([Widén et al., 2010](#); [Kaplani and Kaplanis, 2012](#)), which was considered in the previous paragraph to not be a spatial application. The justification for resorting to correlating data is stated as lacking data availability or the lack of synthetic, spatially decorrelating solar irradiance generation methodologies to supply the time-series.

There are many types of spatio-temporal solar irradiance generation methodologies. Simple approaches interpolate data between sites, whilst more complex methods extrapolate from a single site. Both these types will be discussed, although it is worth stating that the nature of both extrapolation and interpolation require irradiance observation data inputs and so automatically do not complement the aims of this thesis. Some approaches consider the statistical relationship and influence of spatial separation upon irradiance, whilst others adopt these statistics to synthetically generate irradiance. With all these approaches in mind, this literature review will begin with two suggestions. The first is by [Gafurov et al. \(2015\)](#), who suggest that there are no known methods for the integration of spatial correlation of solar radiation into synthetic data using reduced and easily available inputs. The second by [Munkhammar et al. \(2017\)](#), who second this finding and forward that the ideal situation for grid impact analysis would be to have unique irradiance time series for each PV system connected to the grid, as they differ depending on the distance between sites. The author shares these opinions.

The spatial correlation of solar irradiance is highly dependent on cloud dynamics and the geographic separation between sites ([Lave and Kleissl, 2013](#)). Importantly, the irradiance at each location are inherently decorrelating as no two sites can be the identical at all times, this opens the idea for geographic smoothing, whereby one site could receive maximum irradiance whilst another could potentially receive minimum. This concept is well documented in literature through the study of aggregated PV power output. When irradiance time series across a spatial domain are aggregated, a geographical smoothing effect plays a role in tempering the disruption caused by dramatic ramping events from clouds ([Suehrcke and McCormick, 1989](#); [Otani et al., 1997](#); [Wiemken et al., 2001](#); [Curtright and Apt, 2008](#); [Lave and Kleissl, 2010](#); [Lave](#)

[et al., 2012](#); [Marcos et al., 2012](#)). This smoothing can be exploited to minimise high-frequency-variability by increasing the geographic dispersion between sites ([Lave and Kleissl, 2010](#); [Arias-Castro et al., 2014](#); [Lave et al., 2015](#); [Perez and Fthenakis, 2015](#)), benefiting the DNO in both utility-scale PV power plants and residential PV by reducing the affected voltage from synchronised ramping events ([Arias-Castro et al., 2014](#); [Marcos et al., 2016](#)). Utility scale PV plants can be planned considering geographic dispersion, such as maximising the separation of panels should space availability allow. By contrast, the uptake locations of PV-DG is largely unplanned as it is determined by the consumer. Whilst PV-DG does not typically allow for planned geographic smoothing, it naturally occurs so long as multiple installations occur. Aggregation does not solve voltage problems because single intermittent systems can still cause voltage fluctuations, however, it does reduce transformer tap changer operations. Geographic smoothing does not remove the need for high resolution, decorrelating solar irradiance generation as it is imperative to understanding the effect.

Studying the spatial correlation of irradiance can be difficult as it is limited by the need for high resolution data observations that contain an appropriate spatial separation between measurements. The result is researchers regularly use PV power outputs to determine spatial correlation. [Wiemken et al. \(2001\)](#) perform a study on data collected by [Beyer et al. \(1991\)](#), who gathered hourly solar data across an extensive number of sites in Germany. Their comparisons of averaged sites versus single sites demonstrated that decorrelation exists with an exponential decay relationship with separation, figure 4.1 displays the correlation versus separation plot determined by [Wiemken et al. \(2001\)](#). Much research exists that explores the spatial correlation of irradiance across varying distances and time scales ([Beyer et al., 1991](#); [Wiemken et al., 2001](#); [Hoff and Perez, 2010](#); [Perez et al., 2011](#); [Gueymard and Wilcox, 2011](#); [Lave et al., 2012](#); [Widen, 2015](#); [Gafurov et al., 2015](#); [Perez and Fthenakis, 2015](#)).

[Otani et al. \(1997\)](#) developed a method to calculate cross-correlation between multiple sites within an area of $4 \text{ km} \times 4 \text{ km}$. They found that the correlation between sites has a dependency on the day; the sites did not always correlate. Figure 4.2 shows the findings of [Otani et al. \(1997\)](#) looking at two separate days. The 7th October 1996 saw strong linear regression correlation between sites, whereas there was very little correlation for the 2nd October 1996. [Otani et al. \(1997\)](#) did not make a differentiation between cloud motion direction, nor do they explore the influencing factors behind the weather differences between days. Using the longitude and latitude provided by [Otani et al. \(1997\)](#), and exploring historic weather data from nearby Tokyo, Japan, it was found that both days were cloudy, however, the 7th more so than the 2nd. The 7th also featured a significantly higher wind speed which may have influenced the cloud speed. The 7th featured westerly winds whilst the 2nd featured northerly winds, it

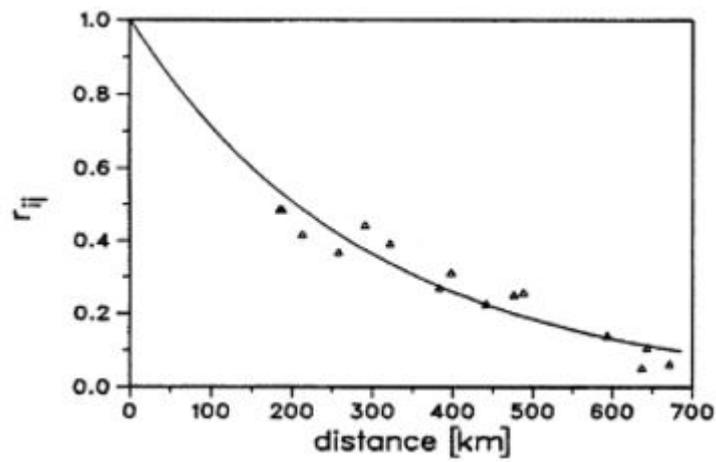


FIGURE 4.1: Spatial cross-correlation structure (cross-correlation as function of inter-station distance) of hourly irradiance data gained from an independent long term (10 years) data set of 6 sites in Germany from [Beyer et al. \(1991\)](#). The line represents a fit to these data using an exponential decay by [Wiemken et al. \(2001\)](#)

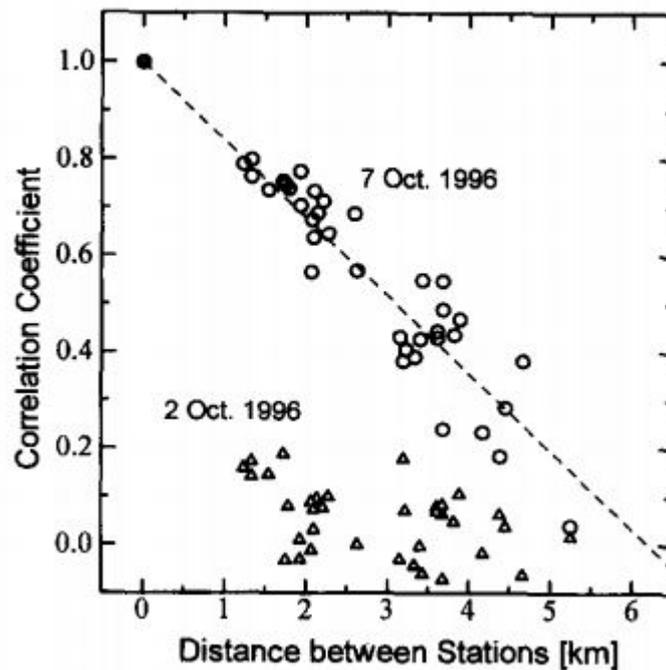


FIGURE 4.2: Irradiance fluctuation correlation coefficients between geographical dispersed solar measurement devices for two different days of 2nd and 7th October 1996 in north of Tokyo, Japan. The 7th October data is fitted with a linear trend line. Plot produced by [Otani et al. \(1997\)](#)

is possible that the connection between wind direction and cloud direction indicates that the 2nd and 7th October had clouds moving roughly perpendicular to each other. [Otani et al. \(1997\)](#) treated all sites as isotropic with expected unity in all directions which may explain why their correlations did not align as expected. [Arias-Castro et al. \(2014\)](#) make the decisive argument that the directionality of cloud motion has significant bearing on the correlation, suggesting that an anisotropic correlation is the only true approach to correctly identifying correlation. Their methodology was tested against real world observations. For clarity, anisotropic models consider the influence of differences in cloud direction, whereas isotropic models do not account for cloud direction or assume a fixed direction.

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. The cloud dynamic and motion is the predominant driver of anisotropy, therefore, cloud direction, speed and size determine the correlation between sites. Smaller clouds cause a lower correlation between sites, except for sites directly in the along wind direction of cloud motion. Sites in the crosswind direction do not correlate strongly until the cloud size significantly increases ([Hinkelman, 2013](#); [Arias-Castro et al., 2014](#)). These small clouds exist most commonly in tropical climates, such as Hawaii, that receives consistent trade cumuli clouds of small radii. [Hinkelman \(2013\)](#) demonstrated that, in Oahu, the larger the separation between sites, the greater the observed ramp decorrelation. For weather systems such as in the United Kingdom, that are dominated by large-scale synoptic weather systems of up to hundreds of kilometres in size, it can be inferred that, there is a greater chance that two locations separated by a large distance will become covered by the same cloud with larger cloud sizes. This was demonstrated in the spatio-temporal variability study by [Glasbey et al. \(2001\)](#) in a location where larger synoptic weather systems are dominant; the correlation in the crosswind was greater. The anisotropic effect is strongest at short time scales ([Perez et al., 2012](#); [Hinkelman, 2013](#); [Widen, 2015](#)). The shorter the time scale, the greater the decorrelation ([Inman et al., 2013](#)). Therefore, cloud size, direction and temporal resolution are important considerations for developing the SDSIG.

Another factor identified as important to the spatial decorrelation of solar irradiance is the topography of the study area. Complex topographies can significantly impact the number of sunshine hours a location can receive. [Olseth et al. \(1995\)](#) studied the impact of geographical positioning to explore the impacts of topography and sun angles at high latitudes. The effect of terrain on available irradiance is described as a microclimatological parameter and varied between 317 to 839 kWhm⁻² across a 10 km²

area. The implication of this is that complex topographies are required to capture this influence, as the aim of this thesis is to develop a methodology to produce spatially decorrelating irradiance time series derived from assumptions and readily available input observation data, incorporating topographical influence is beyond the scope of the SDSIG at this phase.

Extrapolation of irradiance data from a single source is fairly common in literature. [Gueymard and Wilcox \(2011\)](#) state that a solar irradiance measurement station could be used to possibly represent the solar resource at nearby locations, negating the need for additional measurements for around 50 km. [Khalili et al. \(2009\)](#) estimated spatial daily solar irradiance data extrapolating from a point source using autocorrelation, although do not produce a resolution useful enough for grid impact analysis. Almost all extrapolation techniques, whilst using spatial correlation, do not produce time series that are spatially decorrelating to each other. They instead produce a single time series independent of all others generated.

There are two models found in the literature that offer potential alternatives to the SDSIG, however, they require an irradiance input by virtue of an extrapolation technique. [Munkhammar et al. \(2017\)](#) presented a methodology named the Copula model, whereby decorrelating instantaneous k_c are probabilistically generated for any number of locations based on the cross-correlations between k_c values derived from measurements at the same location. [Munkhammar et al. \(2017\)](#) hypothesised that the spatial decorrelation statistics derived at Oahu, HI USA, could be applied globally. The author feels that this hypothesis is highly unlikely due to the lack of considerations in the derivation of the k_c distribution in [Munkhammar et al. \(2017\)](#). Only the hours with $\theta_e > 25^\circ$ are considered, which excludes many periods of lower k_c events ([Smith et al., 2017](#)). Their model lacks the appropriate bimodal relationship of the k_c PDF that they themselves demonstrated to exist. This is where the SDSIG offers an advantage over the very few legitimate alternatives. The k_c distributions by N are able to capture this bimodal shape.

The second alternative methodology is the wavelet variability model by [Lave et al. \(2013\)](#) and [Lave and Kleissl \(2013\)](#). The wavelet variability model is a spatio-temporal methodology 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. The key problem with this method is a requirement of knowing the cloud speed. The authors do propose a method of determining the cloud speed from numerical weather forecasts and have produced daily correlation scaling coefficients for North America. This makes the wavelet variability methodology a partial competitor for spatially decorrelating synthetic irradiance generation for North America.

Interpolation is another method of synthetically producing irradiance time series using station-pair analysis. This type of methodology would be ideal if it could be proven accurate from readily available, high resolution irradiance data sets such as those from the [WRMC-BSRN \(2014\)](#), however, these are typically spaced over hundreds to thousands of kilometres. [Calif and Soubdhan \(2016\)](#) analyse this station pair correlation over 38 km and find a strong influence on the time resolution used in the analysis. Longer time scales demonstrated greater correlation than the 10 min analysis. This evidences the existence of a threshold time scale, below which there is no significant correlation. [Marcos et al. \(2016\)](#) presents an interpolation model that can simulate spatial decorrelation of irradiance time series from a 1 year time series of 1 second irradiance data from 6 sites dispersed over 1100 km². They found that there is a link in correlation between sites, however, it is dependent on the time resolution explored. 6 km separation was significant enough to minimise ramping events from PV plants having an influence on aggregated power flows in real time. This interpolation study does not produce synthetic irradiance time series, it only studies the link between sites.

[Widen \(2015\)](#) use station-pair analysis to integrate the geographic smoothing of distributed irradiance data over a spatial domain of 1 km² and time scales of a few minutes using what is called a virtual network. The model generates irradiance with an estimated aggregate smoothing effect using arbitrary input power output profiles from power plants. This method is useful in power flow models although is not applicable to real scenarios as the model is only simulated on virtual networks, not derived from real world observations.

The most recent developments in interpolation based spatio-temporal irradiance time series modelling comes from [Patrick et al. \(2016\)](#), who present a statistical model of high spatial and temporal resolution at 100 m² and 1 min respectively. Their non-separable spatio-temporal GHI data model outperforms alternative natural neighbour interpolations of the footprint of a PV system.

These methods by [Lave and Kleissl \(2013\)](#), [Calif and Soubdhan \(2016\)](#), [Marcos et al. \(2016\)](#), [Widen \(2015\)](#) and [Patrick et al. \(2016\)](#) require spatially distributed irradiance data and cannot account for variable weather conditions between days. Furthermore, a key distinction between these methodologies and the SDSIG is that they are upscaling techniques. This means that they would only be able to generate either the *aggregated* irradiance of a spatial domain, or a single irradiance time series that would not decorrelate to additional generated time series. They cannot offer individual, decorrelated time series, and are therefore not suitable for multi-variate, house-scale grid impact analysis. They would potentially be more suited to aggregated utility-scale analysis.

[Arias-Castro et al. \(2014\)](#) present the Anisotropic Correlation Model (ACM) that evaluates the spatial decorrelation of irradiance between real world observation sites from Oahu, Hawaii, where small trade cumuli are the norm with typical diameters of 1km or less. This method does not produce irradiance time series and so cannot be considered an alternative to the SDSIG, however, there are crucial assumptions within the ACM that facilitate the move from the SIG to the SDSIG. Firstly, the method of synthetically representing a cloud is novel and proven to be effective, and secondly shows that anisotropy is an essential inclusion in spatial solar modelling. Further discussion regarding the clouds is made in section [4.2.2](#).

[Widen \(2015\)](#) pointed out the complex nature of having to model exact cloud shape, movement and evolution. These elements are fundamental to including a spatial element within the model. To get around this, they make the assumption that clouds move with constant speed across a spatial domain called a virtual network. This concept of a virtual network was designed by [Hoff and Perez \(2010\)](#) and [Perez et al. \(2012\)](#) that allows the exploration of solar correlation with increasing arbitrary distance between station pairs assuming that clouds do not change over that distance and move at constant cloud speed. They find that site pair correlation from clouds predictability decreases as a function of distance. More interestingly, however, they find that the decrease is inversely proportional to the data frequency. These works are the close alternatives to the SDSIG presented in this chapter as they can synthesise irradiance time series with basic assumptions, however, one other target of this thesis is to create statistically accurate irradiance profiles. Statistically accuracy is also site specific such that it captures the local TMY of an area. There are no inputs that guide this weather such as the MTMs in the SIG. The assumptions of cloud direction and speed maintaining a constant between stations is a useful inclusion that can be applied within the SDSIG.

A fitting end to this literature review is to revisit the comment by [Gafurov et al. \(2015\)](#) and [Munkhammar et al. \(2017\)](#) who suggested that currently, there are no known methods for the integration of spatial correlation of solar radiation into synthetic data by using reduced and easily available inputs. A review of the literature finds them to be accurate. Whilst there are some spatio-temporal methodologies that deal with real observational data and the extrapolation or interpolation of measured irradiance to different sites, the authors are unaware of any research that deals with synthetic spatio-temporal irradiance generation derived using readily available data. The research aim of this thesis (section [1.4.1](#)) is the development of a methodology that can facilitate synthetic spatio-temporal irradiance generation using readily available hourly weather observation data. It is therefore a strong indicator of novelty that no true synthetic irradiance generation technique exists as of yet.

4.1.1 Summary of identified research opportunities

There is a clear lack of competitive alternatives to the SDSIG in research. However, the methods employed by extrapolation, interpolation and attempted synthesis techniques offer lessons for consideration on spatial variability and correlation between sites.

Firstly, clouds are difficult and complex to simulate. This is even truer when attempting to use readily available input data. Very few pieces of literature have attempted this, and when they have, simple assumptions were employed. With these simple assumptions, the solar irradiance time series from the models have validated against real observation data and, therefore, it is feasible. Further literature review will take place addressing the ability to synthesize clouds in section 4.2.2.

Secondly, studies considering correlation between station pairs exploring different orientations away from the cloud motion direction. There is clear correlation anisotropy between sites the along and cross wind directions. It is important that the SDSIG enables this behaviour to be captured. This means that the cloud direction must be included within the SDSIG.

Lastly, there are the development opportunities identified from the SIG as outlined in section 3.6.1 and summarised as:

- The SIG could be extended through the inclusion of a spatial dimension.
- The potential for extending the research between the k_c and total cloud amount.
- The methodology could theoretically be applied to produce secondly time series.

4.2 Methodology of the Spatially Decorrelating Solar Irradiance Generator

This section will detail the intricacies associated with advancing the SIG to become the SDSIG. The SDSIG methodology detailed here is summarised in figure 4.3. Red circles in the top right of each component of the flow chart indicate where a development has been made on the previous methodology. Each of these developments will be detailed in their own subsection as listed:

- 4.2.1 Data requirements
 - A discussion and list of the additional user defined data requirements by the SDSIG.

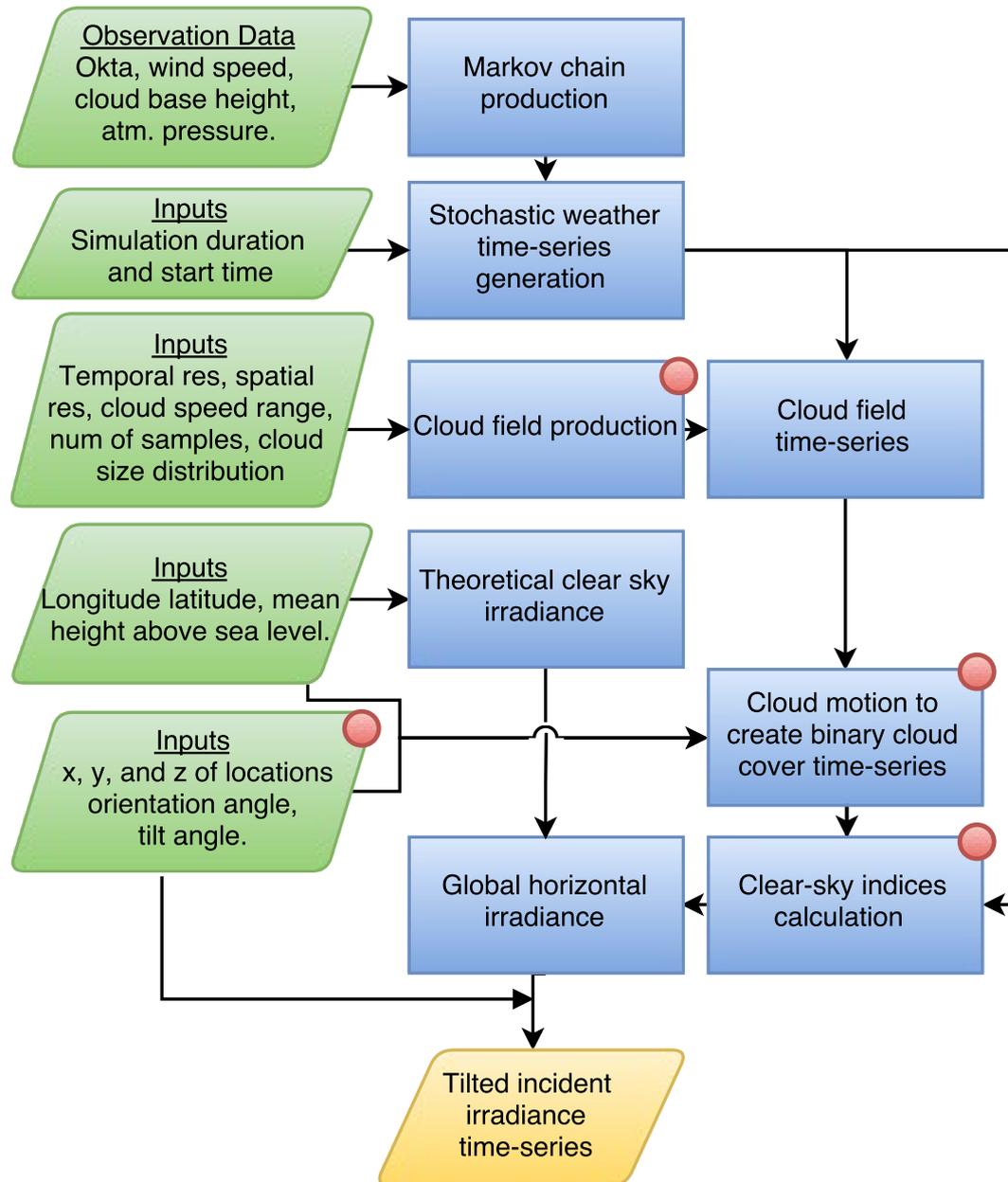


FIGURE 4.3: Flow chart summarising the SDSIG methodology. Green parallelograms represent an input, blue rectangles represent a process, and the yellow parallelogram at the bottom indicates the output. A red circle marker to the top right of a box highlights where changes have been made to enhance the SIG methodology to include a spatial dimension.

- 4.2.2 Cloud field production
 - A definition of a cloud field. Literature review and details of the methods available to produce a cloud field. Development of new methodology to produce cloud fields of each N and u_{ref} .
- 4.2.3 Cloud field motion across a spatial domain
 - Details and discussion of the methodology developed to move the cloud fields over a spatial domain containing the houses.
- 4.2.4 Generation of clear-sky indices
 - There are new considerations for the generation of \mathbf{k}_c . Cloud edge enhancements, improved accuracy of k_c distributions, and extended stable okta periods are discussed and the methodologies detailed.

4.2.1 Data requirements

The SDSIG still operates on the same input data as detailed in section 3.3.1 as total cloud amount in okta (N), wind speed measured at 10 m above the ground (u_{10}), cloud base height (z_{ref}) and atmospheric pressure (p).

In addition to the user defined variables detailed in section 3.3.1.2, there are additional required data from the user. The new inputs are indicated with the red circle in the top right of the green parallelogram in figure 4.3, furthermore, the script can be seen in section B.2.

The new inputs are individual characteristics for each house within the spatial domain where an irradiance time series is desired. They are described as follows:

- Length and size of the spatial domain, X by Y
 - The spatial domain is selected to be 1500 m in length and square ($X = Y$); however, this is a computational limitation. Larger spatial domains can be produced with this method although they require the reproduction of cloud fields to match this size. Morf (2013) state that 30 km would be suitable for an “under the same sky” approach as N is representative to the horizon.
- The coordinates of each unique location, x_l , y_l , and z_l
 - The SDSIG operates on Cartesian coordinates, and not longitudes or latitudes. The Cartesian coordinates are in metres ranging from 1 m to X . Φ

and Θ are placed centrally within XY such that $(x_\Phi, y_\Theta) = (X/2, Y/2)$. x_l and y_l are therefore in reference to (x_Φ, y_Θ) where $(1, 1) = ((x_\Phi - X/2, y_\Theta - Y/2)$. z_l is the height above sea level at each location in metres.

- Panel pitch (or tilt) and azimuth (or orientation/aspect) of each location, β_p^l and β_a^l , respectively
 - These are used the same as in the SIG, however, the SDSIG can handle unique β_p^l and β_a^l for each house.

The SDSIG has the ability to provide each house with an individual height, panel pitch and azimuth, which is rarely employed by current approaches (Engerer and Mills, 2014).

4.2.2 Cloud field production

The transition from one-dimensional cloud representation, as used in the SIG, to two-dimensional cloud representation requires a new approach to synthetically model the sky and the clouds within it.

The cloud samples introduced in section 3.3.2 were represented using a physical 1DM distance series. Each element within the 1DM physically represented a distance employing binary indicators to indicate the presence of cloud. This process was computationally demanding on one-dimension, to replicate this method with two-dimensions across a spatial domain that is sized XY would increase the computational demand by a factor of XY . This limitation drives the need for a complete redevelopment of the cloud representation methodology, which begins with a return to the literature.

4.2.2.1 Methods of synthesising a cloud field

Morf (2011), who presented a direct alternative to the SIG, made a particularly promising finding that “*The steady state probability distribution of the cloud cover is invariant to the observation area over a wide range*”. This presents the possibility to synthetically map the sky, knowing that over the hour, an element in the spatial domain will converge towards the value of the measured okta. The implication here is twofold such that the sky *can* be synthetically represented and that the passage of cloud *should* cause an element within the domain to converge towards the okta value.

For the selection of the size of a cloud field, there is both a physical and computational limitation. The computational limit is how large a cloud field can be modelled that is within the tolerance of available computational power; larger cloud fields require more

computational power. The computational limit for the equipment used to develop the SDSIG was found to be computationally tolerable around a cloud field size of 1500 m^2 , although this feature is dependent on the accuracy required to assess the C value of a synthesised cloud field, and also the temporal resolution of the simulation. The physical limit is how large a cloud size can be statistically represented by N . The question is therefore, to how far is a value of okta representative of the sky? Morf (2013) suggests this limit is 30 km as is appropriate for horizon distance. This is intuitively reaffirmed as total cloud amount observations are often made by a human observer who assesses the visible sky. Gueymard and Wilcox (2011) suggests that a measurement station could possibly represent the solar resource at nearby locations, suggesting a spatial coarseness of 50 km, or representative to 25 km applicability to each modelled area. These two suggestions find the computational limit to be the bottleneck. There is no further investigation into the representation of okta with distance as the computational limit is well within the physical limit.

Arias-Castro et al. (2014) presented the anisotropic correlation model that produces cloud fields for analysis of the spatial correlation between two sites as a result of simultaneous ramping. A fixed motion vector is applied to a cloud field with arbitrary spatial domain size and fixed cloud radii. The cloud centres are selected using a homogeneous spatial Poisson point process whereby a random location within mathematical space is chosen with uniform distribution. Cloud shape is taken as a fixed circle. To the authors' knowledge, there is no published research on the geometric representation of cloud shape, as was also concluded by Arias and Bae (2016). Their model was validated against spatially dispersed data at Oahu, Hawaii (Hinkelman, 2013), and was found to perform well at short time scales, including 1 minute. From the literature review by Arias-Castro et al. (2014), no methods of modelling the cloud shape or cloud fields were found; there are some newer developments since this publication, however.

Cheng and Yu (2015) use representative sky images of different cloud types to produce a sky field in order to forecast the solar irradiance. The issue with this method is the availability of sky images, correct identification of cloud types, and being able to statistically and stochastically reproduce the presence of different cloud types within a simulation. Furthermore, the use of a sky-image does not consider diffuse and other irradiance calculations. This method could perhaps be modified to fit a more binary pattern to utilise the cloud shape in synthetic modelling. As the SIG and SDSIG do not have the stochastic development of cloud types, this method is not considered further. The use of sky imagers to derive cloud shape and motion are becoming more popular with Tzoumanikas et al. (2016), Dulski et al. (2011), Barbieri et al. (2016) and Wang et al. (2016) all being able to determine the presence of cloud from the imagers. Peng et al. (2016) propose a method of synthesizing cloud motion with stored cloud

templates taken from sky imagers. The need for this additional information presents a barrier as a sky imager is required to produce the database. Otherwise, access to a database is required, without which, this method is not appropriate for the rationale for this thesis; it does pose interesting potential future research potential.

A simpler approach is presented by [Augsburger and Favrat \(2013\)](#) that only considers the worst case scenario to solar concentrating heliostats — cumulus clouds with no solar transmittance. They select a cloud motion vector taken as the mean cloud motion direction and cloud speed at a site in Tucson, Arizona USA. A single cloud, sized larger than the solar receiver field, is passed across the simulated space. For a single and short simulation, this method is computationally simple. However, a physical spatial 2DM approach becomes more computationally demanding with longer passages of time. The most interesting information presented by [Augsburger and Favrat \(2013\)](#) is the distributions of cloud speed for Tucson. The implementation of cloud speed in the SIG was rather arbitrarily defined due to lacking literature discovered on cloud speed. The distribution of cloud speed as $\Gamma(3.1608, 1.46735)$ gives a mean velocity of 4.64 ms^{-1} . Preferably, a historic time series of cloud speeds motion vectors would exist so that the stochastic Markov chain approach could be applied. As this data is not readily available, the use of a distribution is instead employed. This can be implemented in the SDSIG and is presented in the next section.

The addition of satellite derived cloud fields could prove useful for cloud field production. [Luo et al. \(2008\)](#) provide a methodology to extract binary representation of clouds of a predefined type using MODIS satellite images, these images could then be used to synthetically generate realistic cloud fields. For appropriate utilisation of satellite images, there would need to be a relationship derived between N and cloud type, and a relationship between cloud type and k_c . An alternative method by [Dulski et al. \(2011\)](#) uses thermal infra-red analysis to derive temperature statistics of clouds to develop synthetic cloud shapes. They propose an interesting two-dimensional noise production, however, the lacking ability to geometrically represent a cloud is not complementary to a sun obscured type methodology.

In summary, at the time of developing the SDSIG and to the author's knowledge, there were no suitable methods of cloud field production aside from the simple approach by [Arias-Castro et al. \(2014\)](#), which can be developed upon to include variation in cloud size following the [Wood and Field \(2011\)](#) distributions discussed in section 3.3.2.

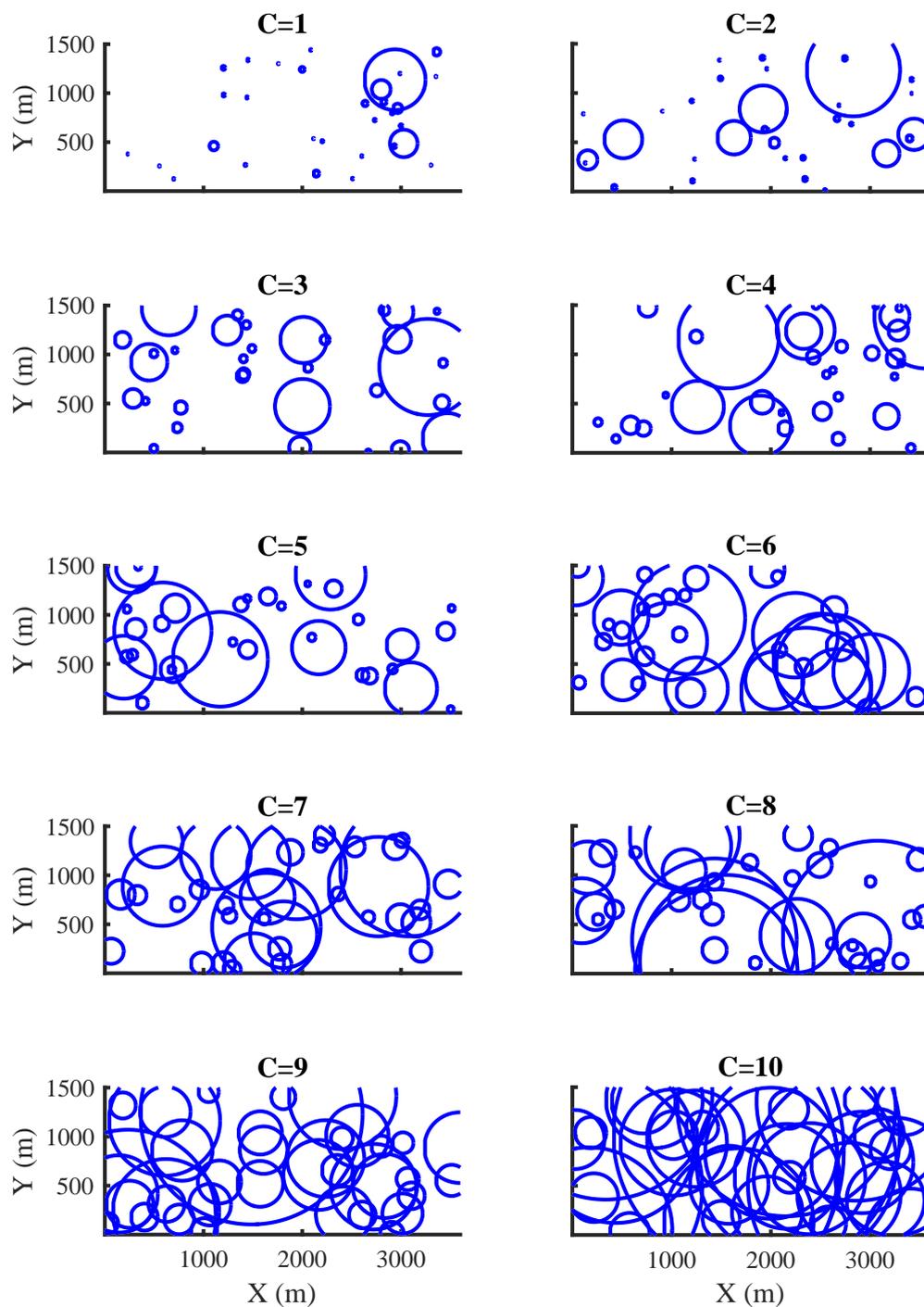


FIGURE 4.4: Example visualisations of cloud fields for coverage values $C = 1$ to 10. The X dimension is equal to $3600u_{\text{ref}}$ and the cloud speed of these cloud fields is $u_{\text{ref}} = 1 \text{ ms}^{-1}$. These plots are merely visualisations of what are in fact a $3 \times n_c$ 2DM of $x_c - y_x - r_c$.

4.2.2.2 Developed approach to synthesising cloud fields

The cloud sample method in the SIG uses a computationally demanding physical representation of distance using a 1DM whereby each element is assigned a representative distance. For progression to a faster and more sophisticated cloud field production, a method of vectorisation is developed whereby the clouds are defined in space with a geometric expression, and are not physically mapped in a 2DM. The computational demand is, therefore, only limited by the number of clouds, n_c , that can be stored in each cloud field. The script for cloud field synthesis is included in section B.1.

As with Arias-Castro et al. (2014), the clouds are assumed to be perfectly circular. This opens the opportunity to represent clouds as a centre point and radii. A centre point of a cloud can be denoted using Cartesian coordinates (x_c, y_c) within defined mathematical space, the radius is denoted r_c . A cloud field is therefore an array with length n_c containing $x_c-y_c-r_c$ values of each cloud; a cloud field is therefore a 2DM that is $3 \times n_c$ elements large. n_c was set to 1200 clouds and so each cloud field contains 3600 elements. Should the SIG cloud sample methodology be used as opposed to this vectorisation method, the spatial granularity would be $\sqrt{3600} = 60$ m. The new vectorised method does not have a spatial granularity limitation and is only limited by the temporal resolution. Overall this is a much more efficient and accurate development.

Figure 4.4 shows actual cloud fields for each coverage value, for clouds moving at 1 ms^{-1} . A simplified cloud field is also seen later in figure 4.5 as black rectangle *A* and the two clouds c_1 and c_2 , or in later in figure 4.7 as the dashed orange box containing clouds.

The proposed production technique creates 500 different cloud fields, n_s , for each integer of cloud coverage fraction ($C = 0$ to 10 out of 10) and for each possible value of cloud speed ($u_{\text{ref}} = 1$ to 30 ms^{-1}), totalling 165,000 different cloud fields to be stored consisting of 594 million elements. Reducing n_s by 1 reduces the number of elements by 1.188 million while reducing n_c by 1 reduces the number of elements by 0.495 million, and so considerable space saving could be made by reducing either. An option could be to link the size of n_s by the prevalence of N at the location of study by using the MTM distributions of N , assigning more cloud fields to oktas with higher prevalence in observations. However, this would unnecessarily require the re-production of cloud fields with each simulation, resulting in extremely poor computational performance..

A cloud field can now be selected as desired once the stochastic weather time series is generated.

The coordinates of each cloud's centre all fall within the cloud field's spatial domain, which is sized to cover the simulation area for exactly one hour, the size of the cloud field is therefore a function of u_{ref} . For computational purposes, cloud fields always travel parallel to the x axis, this allows the y axis of a cloud field to remain independent of u_{ref} . Therefore, the size of a cloud field is $3600u_{\text{ref}}\text{-by-}1.5$ km, where 3600 is the number of seconds in an hour. To produce n_s cloud fields for each u_{ref} at each C , an iterative process is proposed. The process is first described before being summarised.

Firstly, the target C value is set, it is denoted C_t . C_t is the coverage desired for the cloud field being produced and so with each cloud added to a cloud field, the actual C can be checked against C_t to determine satisfaction. It was decided that a tolerance must be permitted due to the nature of recording N . Oktas are reported as discrete integers that have been binned from what is in reality a continuous value, e.g. 3 okta is actually representing 2.5 to 3.49 okta. Including a random variate to C_t would allow for a more representative set of cloud fields per each C . To achieve this, $C_t = C_t \pm 0.5R$ where R is a random variate drawn from a uniform distribution between 0 and 1, $R \sim U(0, 1)$.

The second stage of each iteration is to add a cloud, defined as $x_c\text{-}y_c\text{-}r_c$, to the cloud field. The coordinates of the cloud centre x_c and y_c need to be defined. There is no bias applied to the positioning of clouds into the cloud field as a homogeneous spatial Poisson point process is used, such as was used in (Arias-Castro et al., 2014). A homogeneous spatial Poisson point process randomly selects a point in mathematical space with uniform distribution. This is very simply modelled as $x_c \sim U(0, X)$ and $y_c \sim U(0, Y)$, where $Y = 1500$ m and $X = 3600u_{\text{ref}}$ m. The next step assigns the cloud a radius, r_c . In the SIG, cloud sizes were drawn from the distributions by Wood and Field (2011) and was proved to be an effective tool. The same distributions are used to select r_c as adapted from equation 3.3 to reflect that r is a horizontal cloud length. It is, therefore, inclusive of converting from what is effectively a diameter to become a radius.

$$r = \frac{(\kappa + \lambda R)^{\frac{1}{1-\beta}}}{2} \quad (4.1)$$

Once a cloud is added, the updated C of the cloud field must be determined. The exact geometric solution to C is complex because clouds can overlap. Up to 75% of their area can exist outside of the cloud field environment should it be centred at the very corner of the domain. With $n_c = 1200$, there is a high probability of overlap, as is clearly shown in figure 4.4. Therefore, it is computationally demanding to determine the area obscured from $x_c\text{-}y_c\text{-}r_c$ values. A solution was developed that can be adjusted for different computational availability depending on the accuracy of the required answer. The solution is visually demonstrated in figure 4.5 and is described here.

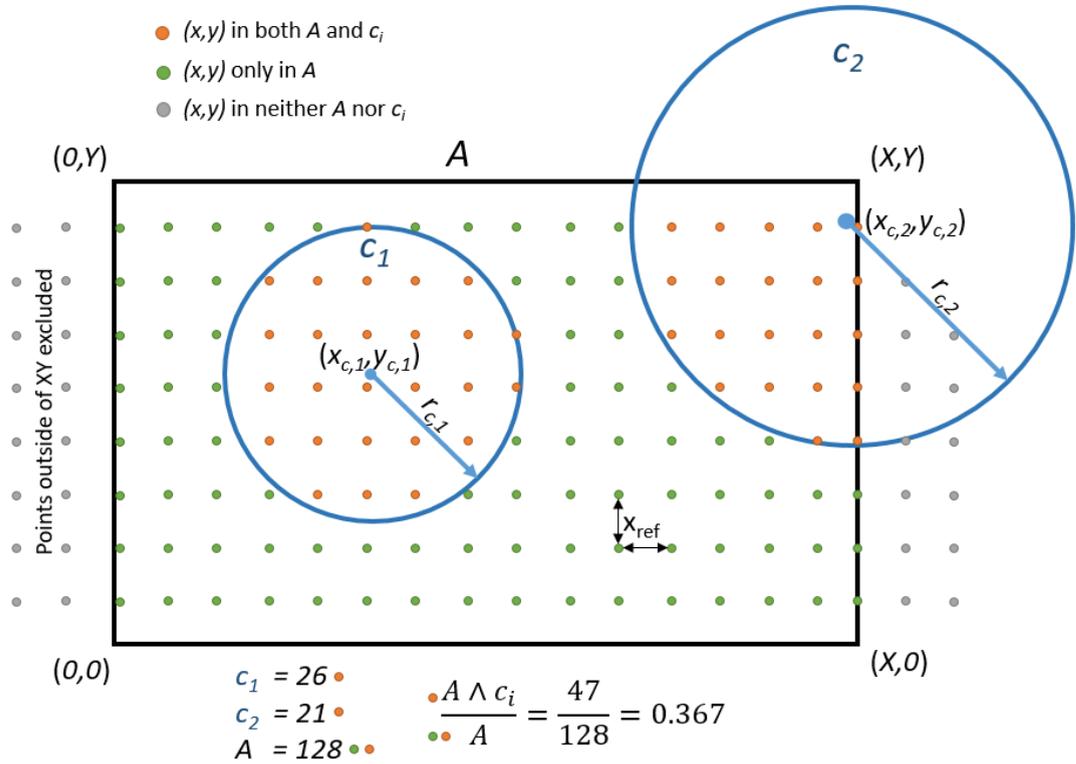


FIGURE 4.5: Example of the utilisation of equation 4.2. A is the cloud field domain defined by its four points. c_1 and c_2 combined create c_i and are shown in blue defined by their respective (x_c, y_c) centre points and radii r_c . The spatial resolution x_{res} is the distance between sample points. The colours of the sample points (orange) indicates a point that is within both A and c_i , (green) a point found only within A and (grey) any point outside of A . The resultant $C_{\text{frac}} = 0.367$ and $C = 4$ tenths

Let the cloud field be rectangle A defined by four coordinates as $(0, 0)$, $(X_c, 0)$, $(0, Y)$ and (X_c, Y) . X_c reflects the different x -axis sizes of the cloud field and the spatial domain such that $X \neq X_c$ for $u_{\text{ref}} > 1 \text{ ms}^{-1}$, whereas the y -axis of both the cloud field and spatial domain are equal regardless of u_{ref} . The space inside A can be expressed using a gridded sample of Cartesian coordinates (x, y) . The rectangle is filled with circles, c_i where $i(1, 2, \dots, n_c)$. The area of each circle can also be expressed using Cartesian coordinates as any coordinate that is within the cloud $x_{c,i}-y_{c,i}$.

Each cloud is defined as any Cartesian coordinate that exists within the circle and can be found using r_i . The list of coordinates within a cloud can be written $c_i = \{(x_{c,i}, y_{c,i}) : (X_c - x_{c,i})^2 - (Y - y_{c,i})^2 \leq r_{c,i}^2\}$. The same view can be taken for the rectangle such that $A = \{(x, y) : 0 \leq x \leq X_c, 0 \leq y \leq Y\}$. The coverage value of the cloud field is therefore the ratio of all Cartesian coordinates that exist in both A and c_i , it can be written using the logical AND wedge symbol as

$$C_{\text{frac}} = \frac{A \wedge c_i}{A} \quad (4.2)$$

Where C_{frac} is the estimated fraction of the coverage value. To convert C_{frac} to C in tenths, a multiplication by 10 and rounding must occur, $C = [10C_{\text{frac}}]$.

The issue with equation 4.2 is that, to be computationally represented, a spatial resolution x_{res} must be used between each Cartesian coordinate considered in A and c_i , this heavily influences the computational demand. The spatial resolution can best be considered a grid of sample points, as is shown in figure 4.5. The equation can be operated with any x_{res} , although for an accurate estimation of C a smaller x_{res} may be required, this is further discussed in section 4.4. The smallest computationally tolerable x_{res} with the equipment available was 50 m, which is equivalent to $31u_{\text{ref}} \times 31$ sample points per cloud field.

The final step of the iteration is to check the calculated C_{frac} value back against the target C_t value. Should C_{frac} be less than C_t , then another cloud (or multiple of clouds for higher u_{ref}) is added into the cloud field and new C is redefined. Once C_{frac} has reached C_t , it is ensured that the same C has been attained by rounding such that if the rounded values of C_t and C are the same, then the cloud field is acceptable. This is expressed in the following equations.

$$C < C_t \xrightarrow{\text{if}} \begin{cases} \text{True} & \text{Add new cloud} \\ \text{False} & \text{Continue to next eq.} \end{cases} \quad (4.3)$$

$$[C] = [C_t] \xrightarrow{\text{if}} \begin{cases} \text{True} & \text{Store cloud field in appropriate bin} \\ \text{False} & \text{Reject cloud field} \end{cases} \quad (4.4)$$

Where $[\cdot]$ is a nearest integer rounding, and $\xrightarrow{\text{if}}$ is a logical IF statement with two outcomes of true or false.

The process to generate a cloud field is summarised as follows:

1. Set C_t within the range $0-10 \pm 0.5R$
2. Add a cloud into the cloud field with random centre point (x_c, y_c)
3. Extract a cloud radius, r_c
4. If the number of clouds exceeds n_c , reset the cloud field
5. Compute the new coverage value, C_{frac} and C

6. If $C < C_t$ move to step 2, else continue
7. If $[C] > [C_t]$ reset the cloud field, else continue
8. Store the cloud field into a storage bin indexed by C and u_{ref}
9. Make new cloud field and repeat until there are n_s cloud fields of each C_t at each value of u_{ref} in every bin

4.2.3 Simulating cloud motion

This section will detail the methodology that “moves” the cloud fields over the spatial domain containing all of the houses. The result is the production of individual binary 1DMs of sun obscured indication (\mathbf{B}) and clear-sky index (\mathbf{k}_c) for each house.

The SIG did not require this step as the cloud sample was already in the form of \mathbf{B} , and k_c was extracted every 6 mins. With the development towards cloud fields in a vectorised format, an entirely new methodology is required.

4.2.3.1 Weather variable generation and cloud field selection

As with the SIG, the same Markov chain production and implementation process and weather variable generation is employed in the SDSIG and can be seen in appendix sections A.4 to A.7. Weather variable generation operates the same as in the SIG with a subtle difference that now, in order to reference a cloud field, an hourly time series of cloud field reference number, n , is also generated. A cloud field is stored as a function of u_{ref} and C with n_s options for each combination of u_{ref} and C . Therefore, n is a random selection from the available options and can be expressed as $n = [n_s R]$ where $R \sim U(0, 1)$. The weather variable generation is the same as the SIG and so all discussion relating to the MTMs and seasonality applies to the SDSIG also.

After weather variable generation, hourly time series of u_{ref} , C and n exist that are applied ubiquitously to the whole spatial domain. The assumption is that all weather variables are applicable to the selected spatial domain size of 1500 m. C is suggested to be well represented to 30 km (Gueymard and Wilcox, 2011; Morf, 2013). u_{10} is highly variable depending on the surrounding topography. As its use is to estimate a cloud speed, that too is assumed to have similar representation over 30 km.

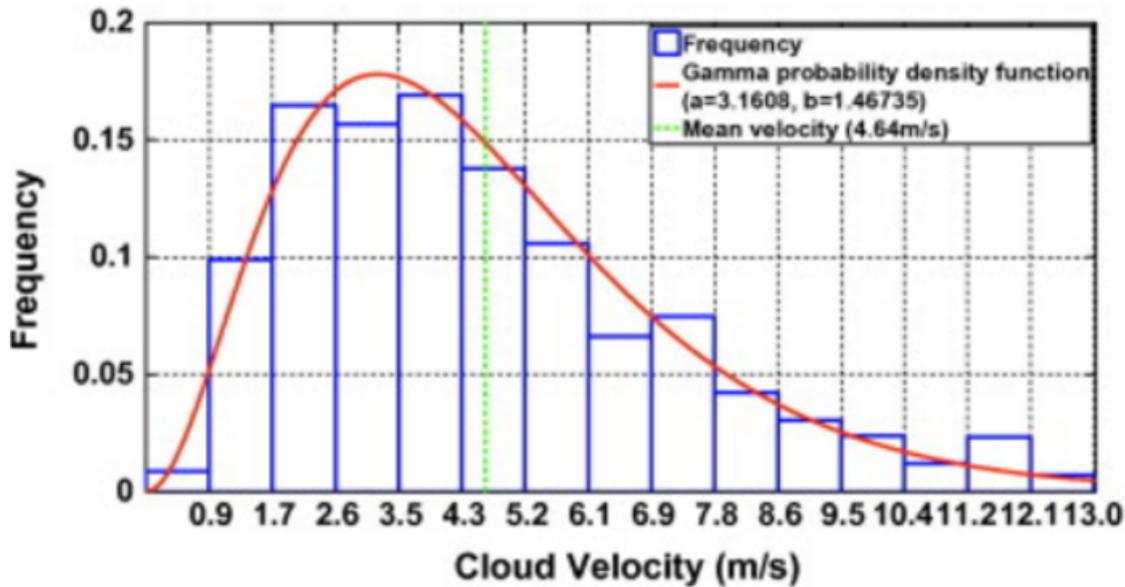


FIGURE 4.6: Cloud speed distribution for Tucson, AZ USA, produce by [Augsburger and Favrat \(2013\)](#)

4.2.3.2 Cloud field motion across a spatial domain

The model script for cloud motion code is intertwined with k_c generation and can be seen in coded format in section B.5 from line 152 onwards. This is because each cloud in the cloud field is assigned a unique k_c value to account for varying cloud thickness.

The best way to visualise cloud motion is to imagine the cloud field moving across the spatial domain from right to left. As the process is vectorised, movement is actually achieved by updating the house-to-cloud distances each time step using u_{ref} and t . The distances facilitate movement of a cloud field, so there is no physical representation, only a constant list of house-to-cloud distances that are updated with each time step. The following methodology description has been accepted for a special issue volume in the Journal of Solar Energy (?). For best understanding of the described methodology, refer to figure 4.7 for visual interpretation.

u_{ref} is extrapolated from u_{10} in both the SIG and SDSIG for clouds below 1 km. For clouds above this limit, the SIG extracted u_{ref} from a gamma distribution of $\Gamma(2.69, 2.14)$. Research for the SDSIG revealed a study by [Augsburger and Favrat \(2013\)](#) that calculated probability distributions for cloud speed and size; they are displayed in figure 4.6. [Augsburger and Favrat \(2013\)](#) find the cloud velocity to be represented by a gamma distribution of $u_{\text{ref}} \sim \Gamma(3.1608, 1.46735)$ for Tucson, Az USA. As the research on this topic is scarce and the SIG assumption was based of a mean and a range from two different studies, using this gamma distribution from [Augsburger and Favrat \(2013\)](#) offers more realism and accuracy to u_{ref} . Whilst Az, USA, is not

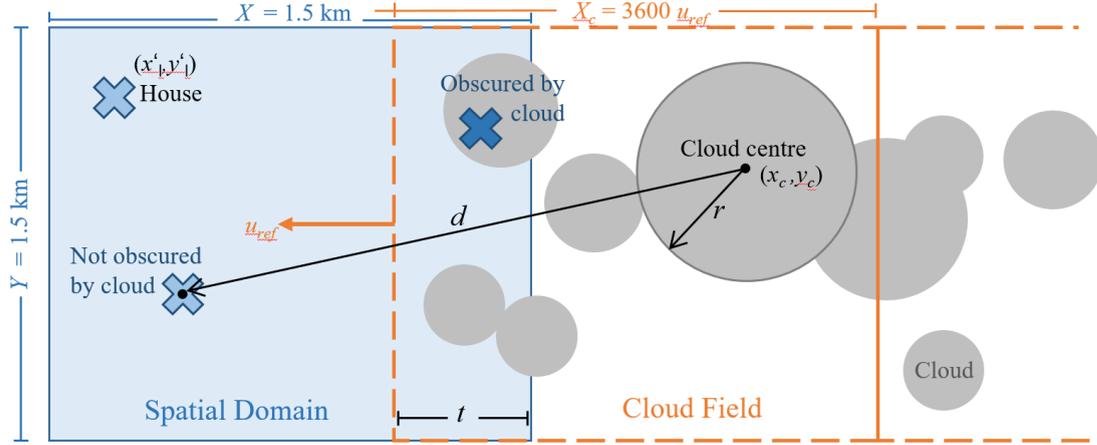


FIGURE 4.7: Visual demonstration of the cloud motion and sun obscured methodologies. The dashed box represents a cloud field full of clouds with centre points (x_c, y_c) and radii, r_c , that is travelling at speed u_{ref} across the spatial domain in solid blue containing the houses positioned at $((x_l, y_l))$ and indicated by crosses. The spatial domain has dimensions XY while the cloud field has dimensions $X_c Y$. 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 field moves at each time step, t , in the direction indicated by the arrow marked u_{ref} .

geographically applicable to the UK, it is the only real indicative distribution of cloud motion vectors published and known to the author. The cloud speed range is within the expected and projected range predicted for the UK, and as the previous distribution was arbitrarily assigned, this is now at least guided by literature. Further research into the cloud motion vectors is assigned to future work. For the most computationally efficient process, the cloud fields are always modelled approaching from the same direction parallel to the x -axis, this allows the y -axis to remain unchanged and halves the number of calculations. In order to facilitate this constant motion direction of the cloud fields whilst also allowing for changing cloud motion direction each hour, the houses are rotated about the cloud direction angle, θ , modifying to rotational matrices demonstrated in equation 4.5 and coded as a Matlab function in section B.3. The modification changes rotation about point $(0, 0)$ to be a function of the centre of the user defined spatial domain, set to $(750, 750)$ in the SDSIG although is flexible to new inputs.

$$\begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_l \\ y_l \end{bmatrix} \quad (4.5)$$

Where x_l and y_l denote the initial coordinates of the house within the spatial domain, and x_l and y_l are the house coordinates rotated by cloud direction angle θ .

The angle by which to rotate is determined using a normally distributed random walk, with standard deviation equal to 10° around the previous time-step's cloud direction, such that $\theta_t \sim N(\theta_{t-1}, 10^\circ)$. 10° is arbitrarily selected to allow for gradual changes in direction each hour. Allowing gradual changes is considered to be important as it will affect the correlation between two sites. The alternative is to fix a uniform cloud direction such as in [Arias-Castro et al. \(2014\)](#). This will result in constant anisotropic correlation over a year, which is not realistic for most geographies. 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. The impact of the assumption is discussed in section 4.4.1 and the development of cloud motion direction distributions added as an interesting field for 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 adopted 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. For example, equatorial and tropical trade wind influenced sites [Hinkelman \(2013\)](#).

At each time step, the distance from each house to the centre of each cloud within the cloud field must be calculated.

Firstly, let the x -axis length of the spatial domain be X and the x -axis length of the cloud field be X_c , the coordinates of a house and each cloud are (x_l, y_l) and (x_c, y_c) respectively.

The distance from a house to the right side edge of the spatial domain is found as

$$\Delta x_l = X - x_l \quad (4.6)$$

While the distance from the cloud centre to the left side edge of the cloud field is equal to x_c .

The overlap of domains is defined as the number of time steps, t , multiplied by the temporal resolution, t_{res} . Therefore, the distance in the x -axis direction from the house to the cloud centre, Δx , is given by

$$\Delta x = \Delta x_l + x_c - (t \times t_{\text{res}}) \quad (4.7)$$

The distance in the y -axis from a house to a cloud is given by

$$\Delta y = y_c - y_l \quad (4.8)$$

Now that the x -axis and y -axis distances from a house to a cloud centre are known, and because it produces a right-angled triangle, the horizontal distance from the house to the cloud centre, d , can therefore be calculated using Pythagoras' theorem as

$$d = \sqrt{\Delta x^2 + \Delta y^2} \quad (4.9)$$

Figure 4.7 visualises the described distances from equations 4.5 to 4.9.

For every time step of the simulation t , using d and r_c it is determined if the house is covered by cloud(s) or not. To do this, a logical IF statement is applied such that

$$r_c > d_t \xrightarrow{\text{if}} \begin{cases} \text{True} & \mathbf{B}_t = 1 \\ \text{False} & \mathbf{B}_t = 0 \end{cases} \quad (4.10)$$

Where \mathbf{B}_t is a Boolean matrix indicating the presence of cloud at the t^{th} time step; $\mathbf{B}_t = 1$ signifies the presence of cloud, whereas no cloud is represented by 0. The symbol $\xrightarrow{\text{if}}$ denotes the use of a logical IF statement offering two outcomes depending on the satisfaction of the logical expression $r_c > d_t$. Equation 4.10 is performed for every house against every cloud and 1DM time series of \mathbf{B} is made for every house.

To ensure that clouds are not abruptly cut off between hours of simulation, in situations where cloud centre points are close to the edge, three sequential cloud fields (*past*, *present* and *future*) are logically queried at each time step, as per equation 4.10. The iterative IF statement is repeated each t until the *current* cloud field has passed outside the spatial domain. At this point, all the sequential cloud fields are updated to reflect that the hour has finished: the *future* sample begins to pass directly over the spatial domain and therefore becomes the *current* cloud field, the *past* cloud field is removed and replaced by the expired *current* cloud field. Lastly, a new *future* cloud field is selected. This prevents sudden disappearance of cloud once an hour is ended, which resulted in consistent clear sky periods each hour.

Each cloud in the cloud field is assigned a unique k_c value so that alongside \mathbf{B} , \mathbf{k}_c can be generated for use in irradiance calculations as in the SIG. This is discussed in the next section.

4.2.4 Generation of clear-sky indices

This section will introduce how \mathbf{k}_c is generated for each house. Recall that, k_c refers to a single value of the clear-sky index while \mathbf{k}_c is a time series of k_c .

The SIG had some limitations. The first limitation was the use of only four different distributions of N , from which to extract a k_c . The limitation is for $N = 1-5$, as they are all extracted from the same distribution of $k_c(N = 6)$; this is now known to depend strongly on θ_e . [Smith et al. \(2017\)](#) show a distinct set of k_c distributions for all N , and also for bands of θ_e . The inclusion of distributions for the ten values of N and for nine 10° bands of θ_e means the SDSIG will now make available 90 different distributions from which to extract $k_c = f(N, \theta_e)$. This is discussed in section [4.2.4.1](#)

Perhaps the most significant limitation to the SIG was the implementation of a resolution by which k_c values were extracted every 6 mins and all minutes in between were linearly interpolated. Furthermore, Gaussian white noise was added to \mathbf{k}_c in an attempt to produce more realistic irradiance profiles. The temporal resolution that the methodology can be applied on is also limited by these assumptions. Theoretically, the SIG is flexible to any time resolution, although the 6 minute and Gaussian fluctuations limit its performance. By developing the SDSIG in a vectorised format and removing the 6 min resolution and Gaussian white noise, the temporal resolution is flexible to temporal resolution. The next step is to generate k_c with the same time flexibility. This is discussed in section [4.2.4.2](#).

Further analysis into what constitutes a CEE event and how it can be incorporated into the SDSIG was performed and discussed in section [4.2.4.3](#).

4.2.4.1 Clear-sky index distributions by both okta and elevation angle

This section considers the author's co-authored publication in the Journal of Solar Energy with Dr Christopher Smith ([Smith et al., 2017](#)). References to [Smith et al. \(2017\)](#) in this subsection indicate work that Dr Christopher Smith did, and not the author of this thesis.

k_c distributions by N and θ_e from UK data are derived by [Smith et al. \(2017\)](#) developing upon those originally published by [Bright et al. \(2015\)](#). The elevation angle dependency

of k_c was hypothesised to be a result of the solar path through the different geometries of clouds, such that low θ_e corresponds to higher attenuation within clouds.

Smith et al. (2017) found that the distribution by N and θ_e can be represented by Burr III distributions for $N(0, 1, 2, 3)$ and by general gamma distributions for $N(4, 5, 6, 7, 8)$.

The PDF of the Burr type III distribution is given by

$$f(x) = \frac{ck}{a} \left(\frac{x}{a}\right)^{-c-1} \left(1 + \left(\frac{x}{a}\right)^{-c}\right)^{-k-1} \quad (4.11)$$

Where a is a positive scale parameter, and both c and k are positive shape parameters.

The PDF of the generalised gamma is given by

$$f(x) = \frac{px^{d-1}\exp(-(x/a)^p)}{a^d\Gamma(d/p)} \quad (4.12)$$

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

Figure 4.8 demonstrates the individual k_c dependences by N and θ_e . A bimodal shape is observable for all $\theta_e > 10^\circ$. For $\theta_e = 0^\circ$, there is very little definition of a k_c peak at 1.

The distributions are limited to elevation angles $\leq 50^\circ$ in Smith et al. (2017) as they were derived for the UK where the sun is always at least 27° from solar zenith. For applicability elsewhere in the world, such as San Diego, CA USA, an extrapolation is required to determine k_c distributions for θ_e up to 80° . This was carried out by extending the best fitting trend of both the shape and scale parameters when plotted against elevation angle. There is only a very small difference observed when extrapolating the distributions from 50° up to 80° . This is perhaps further support of the hypothesis made by Smith et al. (2017) that the dependency was due to the path that the light travels through the cloud, and that clouds are often wider than they are tall for most cloud types. Therefore, the attenuation is higher when the sun is lower in the sky. At $\theta_e > 50^\circ$, the path travelled through the cloud is from above and so share a similar k_c distribution to $\theta_e = 50^\circ$. The use of an extrapolation is acceptable due to the fact that the sensitivity is low.

The distributions from Smith et al. (2017) and extrapolations are included in section B.7 alongside the distribution parameters for each value of N and θ_e .

One assumption with using these distributions is that the k_c distributions by N and θ_e hold accuracy when applied globally. It is thought that annual hourly k_c histograms are different for each geographic location because they experience different quantities

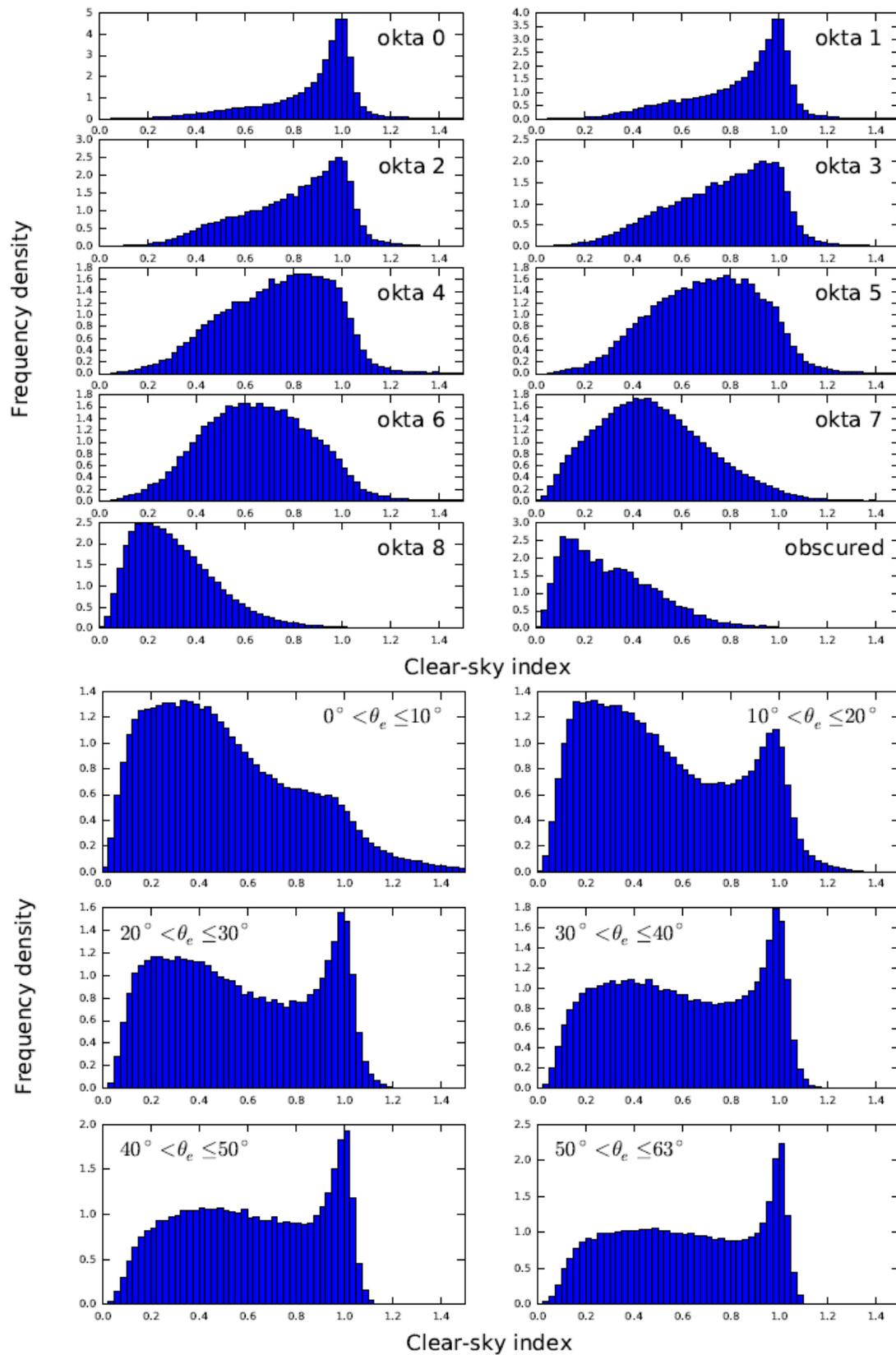


FIGURE 4.8: 1) Histograms of the clear-sky index distributions by okta for all solar elevation angles. 2) Histograms of the clear-sky index distributions by solar elevation angle bands for all okta. Both plots were produced by and published in [Smith et al. \(2017\)](#)

of okta states, not because the k_c distribution by each okta is different around the world. A desert climate would experience greater quantities of 0 okta and fewer 6-8 okta than the north of Scotland would. The annual k_c distribution would therefore be a function of okta frequency. This assumption does have limits as the k_c distribution by N does vary. It is acknowledged that different okta will inherently possess a myriad of different cloud types at different geographies, some types more prevalent than others depending on the climatic region. However, the irradiance magnitude frequencies that are produced using the distributions shown later in this chapter suggest suitability for international use through validation at two additional climates.

The method of producing \mathbf{B} within the SDSIG cannot physically account for the solar elevation. Sun obscured is modelled such that the sun is directly overhead each house within the spatial domain. Therefore, an assumption is made that the SDSIG method captures the influence of solar geometry and a third dimension through using the changes in optical thickness that are limited to the different k_c distributions derived accounting for θ_e .

4.2.4.2 Implementation of clear-sky index distributions

To extract a k_c from the distribution, both the N and θ_e must be known for that time-step of the simulation.

N is stochastically derived in the weather variable generator and θ_e is calculated at each time-step following the algorithm defined by [Blanco-Muriel et al. \(2001\)](#), detailed in section 3.3.5.2. The lowest θ_e that exists within that hour of simulation is taken as the reference with which to select the appropriate k_c distribution for use. Once the appropriate distribution is selected, each cloud contained within a cloud field is assigned an individual k_c extracted from that distribution. The cloud's associated k_c is applied to the entire area of the disc. Should a house be obscured by more than one cloud, a mean of all k_c is taken. When a house is obscured by more than one cloud, it would be intuitive to think that taking a multiple or additive of all obscuring k_c would be the correct solution, however, this would lead to extreme values of k_c that do not appropriately reflect the distributions. Furthermore, the cloud field methodology is not a realistic representation of the sky, therefore, overlapping clouds are not reflective of actual overlapping clouds where the multiple of optical thicknesses would be true. As overlapping discs within the cloud fields do not indicate the presence of more cloud, the mean is taken. To implement the mean, the following logical query is used.

As with equation 4.10 in section that produces \mathbf{B} , at the same time step, a time series of k_c is created and denoted as \mathbf{k}_c for each house as

$$d < r_c \xrightarrow{\text{if}} \mathbf{k}_{c,t} = \frac{1}{n_c} \sum_1^{n_c} k_c \quad (4.13)$$

The final inclusion is to provide k_c for periods of clear sky ($\mathbf{B} = 0$) when a house is not obscured by cloud. This is considered to be the clear sky irradiance, however, it must include inter-day variability similar to those contained in the Linke turbidity. k_c for moments of clear sky within the SIG were drawn from 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 clear sky moments. Using the inbuilt Matlab distribution fitting tool ([Matlab, 2015](#)), the clear-sky peak was found to fit a normal distribution of $N(1.02394, 0.04)$, Oahu, HI USA, was found to fit a Burr distribution of $B(0.9813, 0.72.37, 0.1203)$. This adjustment is accredited to sensor offsets and climatic regional differences, furthermore it is reflective of inaccuracy of the Linke turbidity during the years of validation. Further research would be required to derive daily k_c for periods of clear sky between gaps in the clouds.

4.2.4.3 Inclusion of cloud edge enhancement events

The concept of CEE was introduced in section 3.3.6.4. It is summarised as an event whereby a point on the Earth's surface receives a larger amount of GHI than is available in the theoretical CSI. These events are attributed to additional beam irradiance reflecting off the edge of a cloud or multiple clouds. The effect is examined to a 1-sec resolution by [Lave et al. \(2012\)](#) and simulated with a spatial dimension by [Pecenak et al. \(2016\)](#).

After development of the SIG, analysis was performed to compute the CEE magnitude and probability of CEE occurrence. The magnitudes of the CEE ramps were analysed through comparison of 1-min GHI data from the [WRMC-BSRN \(2014\)](#) against the theoretical CSI data calculated with the PSA from [Hammer et al. \(2003\)](#) and HELIOSAT from [Müller and Trentmann \(2010\)](#).

To identify a CEE event, the fractional difference of the observed GHI from the calculated CSI was calculated and can be described such that a fractional difference of 1 is for $\text{CSI}=\text{GHI}$, < 1 is for $\text{CSI}>\text{GHI}$, and > 1 is for $\text{CSI}<\text{GHI}$. An issue presents itself at this point. Because the method of CSI calculation is synthetic, it does not always produce similar irradiance magnitudes as the clear periods in the observed GHI, and so a fully clear day can exist with a consistently higher irradiance than with using the PSA. This is demonstrated in figure 4.9 where the observed GHI is consistently 16%

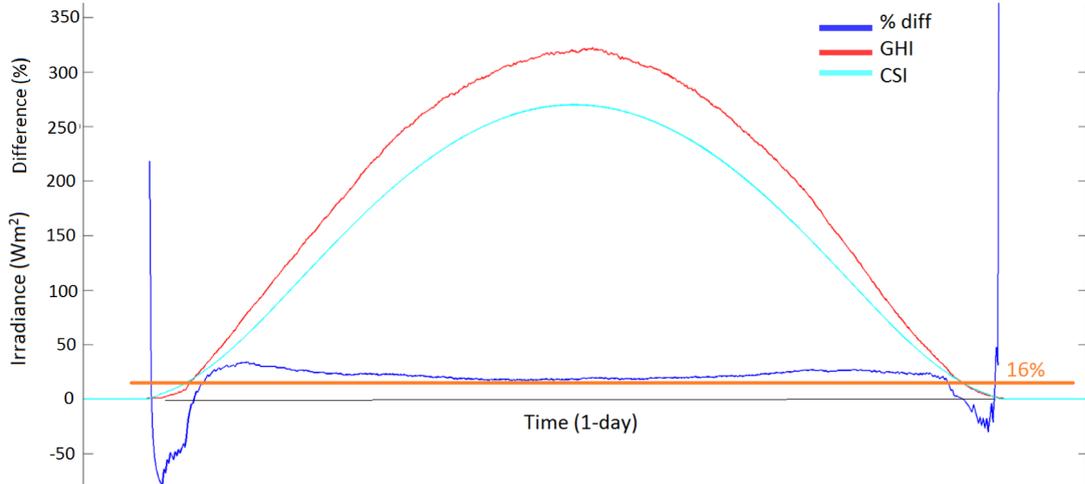


FIGURE 4.9: A sample clear-sky irradiance time series (cyan) alongside observed global horizontal irradiance from WRMC-BSRN (2014) in Cambourne (red) and the percentage difference between the two (blue). The y axis is both irradiance (Wm^{-2}) and percentage difference (%).

greater than the CSI. For this reason, a definition of what constitutes a CEE is made.

For mathematical identification, a CEE event is defined as instances of GHI that are greater than the corresponding CSI by at least a certain margin, this margin is called the *cut-off* and is denoted M_{\min} . Therefore, for a CEE event to be registered, the fractional difference between GHI and CSI must surpass M_{\min} . Four different M_{\min} cut-offs are explored as 0, 0.05, 0.1 and 0.16. $M_{\min} = 0.16$ was selected for use in the SDSIG as it was the largest consistent cut-off observed throughout a day as shown in figure 4.9.

A simple estimation was made to identify the chance of a CEE event occurring. This is defined as the number of events existing above M_{\min} divided by the number of events in total, such that

$$M = \frac{GHI}{CSI} \quad (4.14)$$

$$\frac{\sum (M > (1 + M_{\min}))}{\sum M} \quad (4.15)$$

Where M is the fractional magnitude of the CEE event, the use of \sum in this case is referring to the number of events in each.

To gain more insight into when a CEE event may occur, the chance and magnitude of a CEE event is separated by N , θ_z and M_{\min} , and is shown in figure 4.10 and 4.11. The introduction of M_{\min} shows a stark reduction in probability for lower N . This can

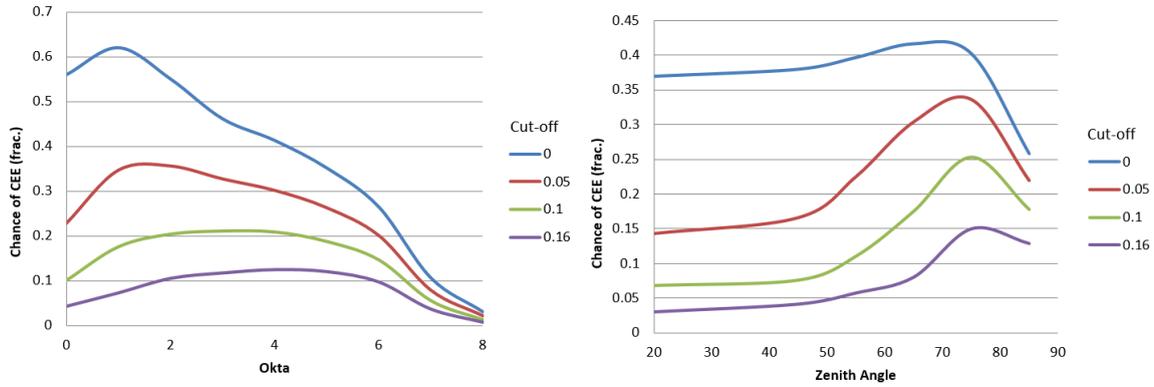


FIGURE 4.10: The chance of a cloud edge enhancement event occurring at different (top) okta values and (bottom) zenith angles, all separated by varying cut-off margins.

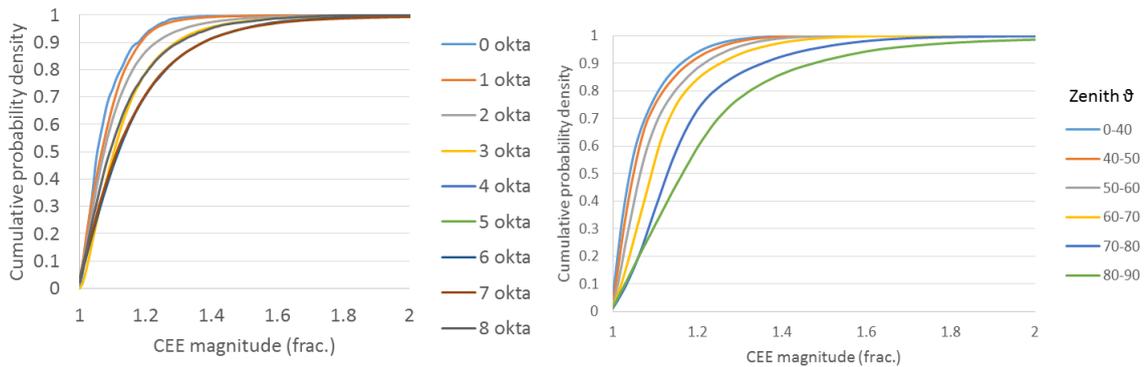


FIGURE 4.11: The cumulative probability function of cloud edge enhancement fractional magnitude (top) separated by each okta value and (bottom) separated by bands of solar zenith angle.

be explained by the ratio of time experienced under clear-sky. Conditions of 7 okta will present very few periods of clear sky with which to undergo a CEE event, whereas conditions of 1 okta had the highest probability. At $M_{\min} = 0.16$, the chance of a CEE event remains between 0.5 and 0.13 for all N . Because there is no DNI in conditions of 8 okta, the chance of CEE is minimal. M_{\min} influences lower θ_z , with the chance of CEE remaining high for $\theta_z > 50$.

This representation of a chance is not truly indicative that, if a cloud is passing across the sun, a CEE will or will not occur. This chance is more indicative of the fraction of the day by N and θ_z that underwent high irradiance moments.

One assumption is that all instances of $M > M_{\min}$ are a result of CEE. Furthermore, the CDF profiles generated are assumed to be accurate to all climates, this assumption is minimised through the inclusion of θ_z , as these are accurate for all climates.

The steepness of CDF profiles separated by both N and θ_z were found to have a proportional relationship with M as is shown in figure 4.11. CEE achieves greater

magnitudes with a larger θ_z (when the sun is lower in the sky). This is unsurprising as the irradiance is low at large θ_z and, therefore, the fractional increase in irradiance can be exaggerated when not considered in absolute terms. A similar pattern is observed with greater magnitudes achieved with greater N . M at large θ_z is less of an issue, in terms of grid impacts as the absolute change in irradiance is lower than during low θ_z and represents only a small ramping event. This discussion is continued in section [4.4.4.3](#).

CDF profiles of M for each value of N and for each band of θ_z were made. A random variate driven extraction is utilised once the appropriate okta-guided magnitude distribution is selected using the appropriate θ_z and N at a time step.

$$M_{\text{CDF}} = f(N, \theta_z) \quad (4.16)$$

$$M = 1 + (M_{\text{CDF}} < R) \quad (4.17)$$

Where M_{CDF} is the CDF profile selected using N and θ_z . R is a random variate from a uniform distribution between 0 and 1. The use of R here allows for random number extraction following the selected M_{CDF} . Adding 1 normalises M to become a correctional factor that can be directly applied to k_c through multiplication.

For application in a 1-min resolution SDSIG simulation, a simple approach is adopted. The chance of a CEE event is extracted using N and θ_z and is logically queried against R . Should a CEE event be required, for the preceding two minutes and subsequent two minutes of the presence of a cloud, a stepwise fractional increase is applied proportional to M . The minutes directly before and after a cloud are allowed an M increase, whilst the minutes surrounding these receive a $\frac{M}{2}$ increase. The reason for this is to allow gradual increase without affecting the overall ramps, similar to observations in [Lave et al. \(2012\)](#).

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}_t undergoes a ramp on account of cloud cover. Should \mathbf{B}_t undergo a ramp, the equivalent time step within k_c is adjusted using the logical if statement as shown below. Equation [4.18](#) represents a ramp down, equation [4.19](#) a ramp up.

$$\begin{aligned} \mathbf{B}_{t-1} = 0 \ \& \ \mathbf{B}_t = 1 & \xrightarrow{\text{if}} \\ \mathbf{k}_c^t = \mathbf{k}_c^t(M) \ \& \ \mathbf{k}_c^{t-1} = \mathbf{k}_c^{t-1}\left(\frac{M}{2}\right) & \end{aligned} \quad (4.18)$$

$$\begin{aligned} \mathbf{B}_{t-1} = 1 \quad \& \quad \mathbf{B}_t = 0 \quad \xrightarrow{\text{if}} \\ \mathbf{k}_c^t = \mathbf{k}_c^t(M) \quad \& \quad \mathbf{k}_c^{t+1} = \mathbf{k}_c^{t+1}\left(\frac{M}{2}\right) \end{aligned} \quad (4.19)$$

4.2.4.4 Extended periods of clear and overcast sky

Extended periods of completely clear or overcast skies do not offer much short-term variability (Skartveit and Olseth, 1992) and so for long, consecutive periods of 0 or 8 okta, a smoothing adjustment is made for periods of 8 okta lasting > 4 hours or for periods of 0 okta lasting > 3 hours. The SIG also had this inclusion, however, a linear interpolation was used that created a poor visual fit.

Within the SDSIG, and observable in code format in section B.6, a random number of intervals is selected from 1 to 5 multiplied by the number of hours in the extended periods of okta 0 or 8. The SIG used 20 intervals regardless of the duration meaning that the minimum change in k_c is every 12 mins. Allowing for this random inclusion of intervals allows the representation of more types of overcast skies. 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, non-linear interpolation technique called a smoothing spline as the spline method has tendency to overshoot and oscillate if the data are not smooth (Moler, 2010). This smoothing is required as the k_c distributions for 8 okta do not take into account the cloud amount duration, this would require a complex stochastic element to the derivation of k_c distributions.

Without the smoothing there are few periods of smooth irradiance during conditions of 8 okta that are regularly seen in real irradiance observations. This would result in an overestimation of the variability index discussed in section 4.3.

4.2.5 Spatially Decorrelating Solar Irradiance Generator outputs and computational performance

The outputs of the SDSIG are multiple irradiance time series for any x - y - z location within the 1.5 km² spatial domain, each with an individual panel pitch and azimuth.

The computational speed of the SDSIG for single day outputs is fast once the initialisation is performed. Once the simulation duration extends to a year, the number of houses in the system becomes crucial to computational speed, impacting it proportionally.

For the grid impact application of the SDSIG in section 4.5, over 2400 locations were required for a whole year. This was very computationally demanding and required over 7 days to complete on a standard desktop computer. It is worth commenting that the intention of this thesis was proof of concept and the development of a mathematical methodology, not for computational perfection. Training and learning to code a computer model were obtained throughout the completion of this thesis, naturally there may exist inexperienced computational execution inefficiencies.

The initialisation of the model requires the production of the cloud field samples and the collation of the input parameters.

4.2.6 Computational demand

The SDSIG is understandably much more computationally demanding than the SIG. Though the SDSIG can operate the same as the SIG, should only a single house be desired for an irradiance time series. The real test of the SDSIG when using a dual core 4 GB RAM standard university machine was to produce the 2438+ (the plus is because some nodes were un-necessary) irradiance profiles for a single year. This operation required 7 days to complete. Reading in the data takes 3 mins and is stored for future operations during the same session. The production of the Markov transition matrices requires a minute. The generation of the stochastic weather variable time series requires less than 1 minute. The issue comes when deriving the clear-sky index time series and subsequently the irradiance for each property within the simulation. The process is iterative, running through an hour at a time before determining the clear-sky index, GHI, DNI, and diffuse time series for each property, then moving to the next hour. The clear sky irradiance is calculated for a year which requires less than a minute. The remaining calculations, demand the whole time. The SDSIG time duration is therefore directly proportional to the product of the number of houses. This methodology, if run through more RAM and processing power would dramatically reduce the requirements of the SDSIG; time requirements would be nullified should access to a supercomputer be available.

4.3 Validation of the Spatially Decorrelating Solar Irradiance Generator

The SIG was validated on both 1 hour and 1 minute resolutions. For the SDSIG, a 1 hour validation is considered not to add value to the methodology and is not included.

The SDSIG was validated to 1 minute in two locations across the UK that each represent a significant weather type of the country. Cambourne, Cornwall, receives some of the highest annual mean solar irradiance in the country and Lerwick, Shetland, receives some of the lowest annual mean solar irradiance. The SDSIG is also demonstrated in San Diego, California USA and Oahu, HI USA. San Diego represents a desert climate while Oahu is a tropical pacific island climate, both are a real test for the SDSIG to demonstrate geographic flexibility.

Validation is difficult on a spatial dimension as there are very few sites world wide with open access to well-maintained, geographically dense, high resolution irradiance time series that also coincide with meteorological weather observations for activation of the SDSIG. Only one site exists as maintained by [Hinkelman \(2013\)](#) that is used with regularity in the research ([Arias-Castro et al., 2014](#)). The data is from Oahu, HI USA.

Presented first in this section is a 1 min temporal validation for Cambourne, Lerwick and San Diego. Secondly, both a temporal and spatial validation for Oahu is made.

4.3.1 Temporal validations for the UK and USA

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 from the SDSIG 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, 2016a](#)) 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

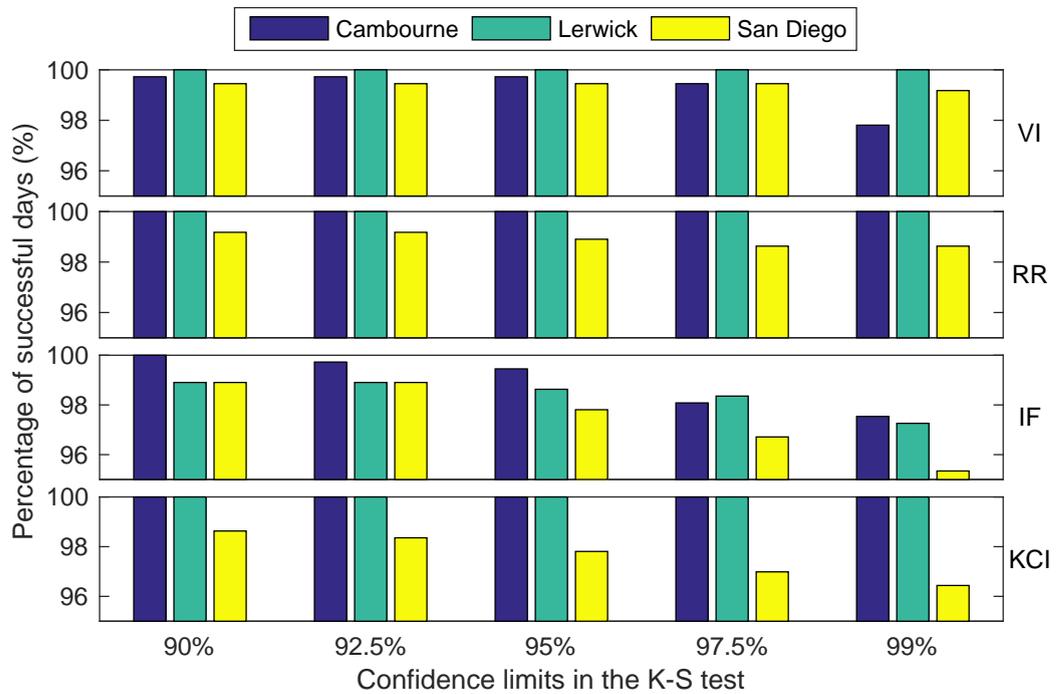


FIGURE 4.12: Results of the K–S test for each of the three locations for the four metrics of variability index (VI), ramp rates (RR), irradiance frequency (IF) and the clear-sky index (KCI). Percentage of successful days that pass the K–S test are shown by location for each metric at increasing K–S test confidence limits. Observation and modelled synthetic data for each calendar 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.

be created. This facilitates the statistical capturing of a TMY. 12 years of data were taken from BADC and 11 years are taken from NOAA.

NOAA data does not report total cloud amount in okta, instead, 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 ± 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.

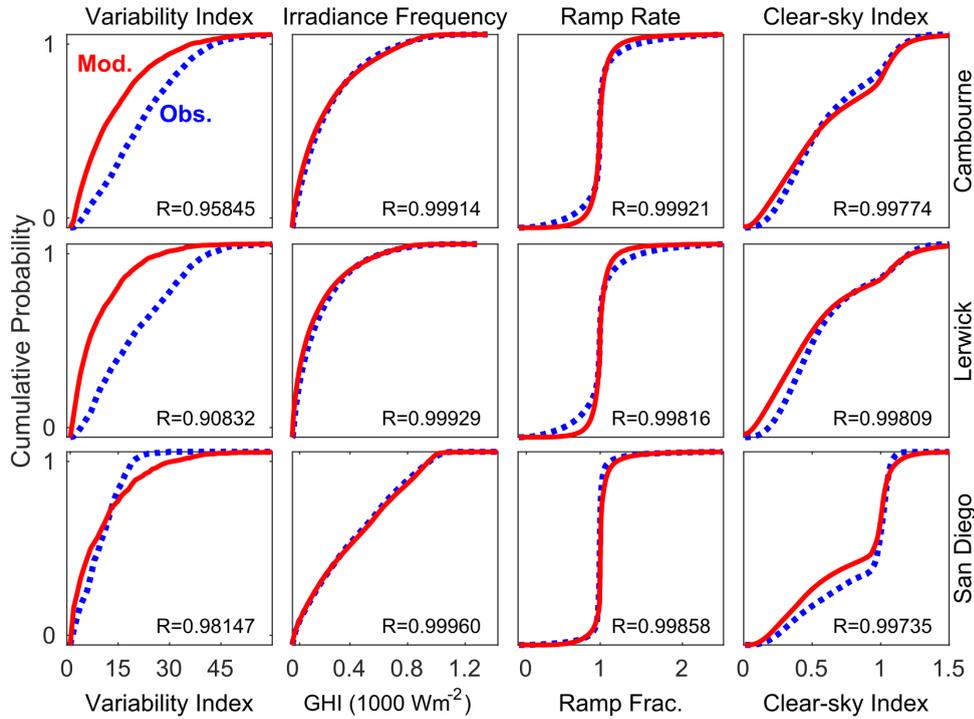


FIGURE 4.13: Comparison of CDF profiles from 1-year of observation data (blue dashed line) and synthetic modelled data (red solid line) of the four metrics, from left column to right: variability index, irradiance frequency, ramp rate occurrence and clear-sky index, at each location, from top row to bottom: Cambourne, Lerwick and San Diego. Each CDF comparison has an individual 2-D correlation coefficient displayed inside the axes.

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-minute values of each metric. Furthermore, the two-sample Kolmogorov–Smirnov (K–S) test was carried out following the same method as outlined in section 3.4. 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 is reportedly acceptable down to sample sizes of 10 to 100. The smallest subset in this validation is 1440, and so the test is suitable. The K–S test results are

displayed in figure 4.12 and comparative CDF profiles of the metrics in figure 4.13 with the correlation coefficient denoted as R and calculated using Matlab's 2-D correlation coefficient function (MathWorks, 2016).

Figure 4.13 shows 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 $R = 0.9991$ 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 successful 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 weather features within both samples of 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, although the long term IF for San Diego are clearly well captured, as is shown by the CDF comparison.

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 minutely VI, as used in the K-S tests, for Lerwick satisfy the K-S test with no annual bias. However, there is clearly a discrepancy between daily mean VI, as shown in the CDF profile. 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. 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. A similar systematic error in representing the daily VI has clearly been carried forward from the SIG to the SDSIG. Whilst it is apparent that the performance of the VI metric worsened from the SIG to the SDSIG, it opens the possibility for further discussion, development and research. It was believed that the added complexity from the former to the latter would facilitate a closer correlation, however this was not the case. It is suggested with firmer hypothesis that a higher order Markov chain is required to capture the daily VI. This does not detract from the excellent validation in all other categories.

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 N and how the distributions of k_c weighted by both N and θ_e accurately create real world distributions of k_c and importantly, a vast improvement on the KCI captured in the SIG. 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.

As discussed with the equivalent plot for the SIG, there is little research that supports these 1-min statistical plots shown in figure 4.13. The IF compared in figure 9 by [Grantham et al. \(2017\)](#) shows excellent reproduction of the DNI and GHI between observed and synthetic data. Their correlations boast R values of 0.984 to 0.999. What is remarkable is that [Grantham et al. \(2017\)](#) present an interpolation model such that the hourly irradiance values are already extremely guided; the SDSIG has no such luxury and instead relies on the stochastic and probabilistic distributions of meteorological variables whilst achieving greater average IF correlation results. [Munkhammar et al. \(2017\)](#), who presented the copula method on virtual spatial networks — a competitor of the SDSIG, did not compare irradiance outputs with data, instead they validated the spatial correlation. Figure 9 by [Larraneta et al. \(2015\)](#) shows CDFs of the IF for 10-min synthetic DNI values separated by k_c . Whilst they do not report the correlation, their M5 model would pass the IF K-S test with a 97.5% confidence with 10-min profiles whereas the SDSIG passes with 1-min resolution at 99% confidence. Figure 6 by [Ngoko et al. \(2014\)](#) displays PDFs of the clearness index demonstrating a strong fit, though without statistically quantifying it. The SDSIG performs in a similar manner, despite not using high frequency irradiance to derive the statistics, one of the key flaws in the [Ngoko et al. \(2014\)](#) methodology.

TABLE 4.1: Station metadata from the NREL Oahu Radiation Monitoring Station (Sengupta and Andreas, 2010) showing the height above sea level (z_l), latitude (Θ) and longitude (Φ) for each ID in figure 4.14.

ID	z_l (m)	Θ ($^\circ$)	Φ ($^\circ$)
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

4.3.2 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 figure 4.14 as maintained by the National Renewable Energy Laboratory (NREL) (Sengupta and Andreas, 2010).

The observation Oahu irradiance time series were averaged from 1 second to 1 minute 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, 2016a) for station location Kalaheo Airport, John Rodgers field, Kapolei, HI USA (ID:22551); the site is at latitude 21.316° , longitude -158.066° , and 10 m above sea level. The user defined input variables that detail each property in the SDSIG are shown in table 4.1. The straight line distance between station pairs are calculated using the haversine formula using Θ and Φ .

The SDSIG is first subject to a temporal validation against one of the Oahu measurement sites selected at random. This is to demonstrate the SDSIG's suitability at creating statistically accurate irradiance time series for Hawaii.

Using the four metrics as before of VI, RR, IF 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 for Hawaii; the SDSIG does

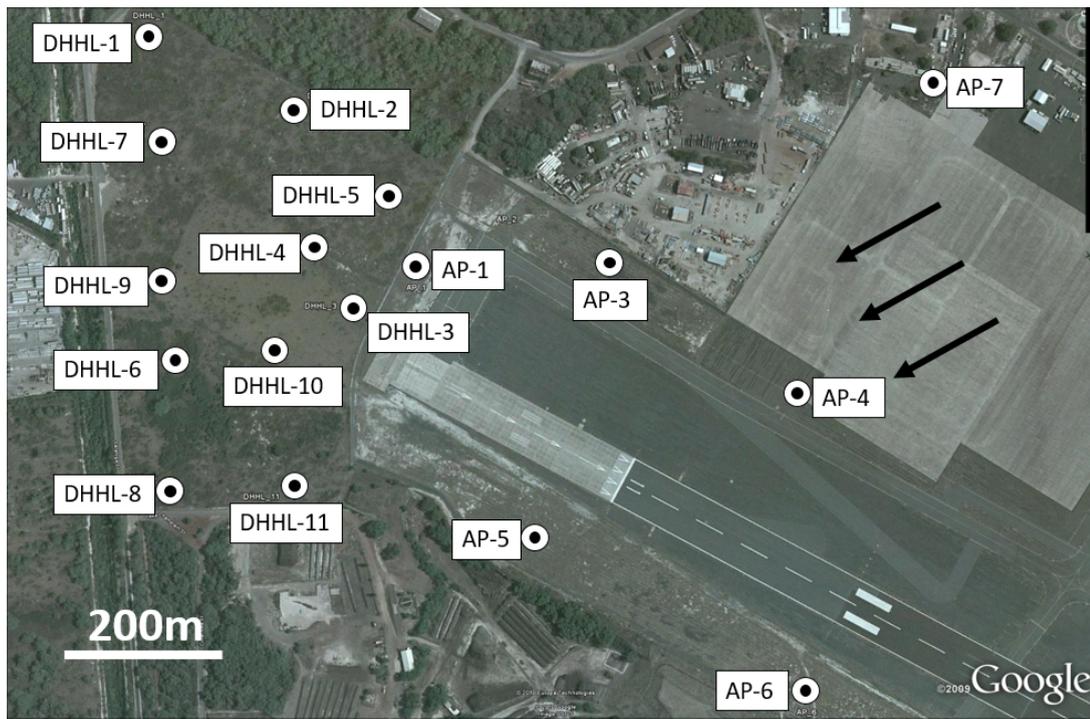


FIGURE 4.14: Geographical layout of the measurement stations at the NREL Oahu Solar Measurement Grid (Sengupta and Andreas, 2010) on Department of Hawaiian Homeland (DHHL) and Kalaeloa Airport (AP) property. The arrows indicate the direction of prevailing wind.

not produce many days of complete clear sky for Hawaii that leads to $VI=1$. 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 interpretation is that the use of a first order Markov chain does not facilitate extremely long periods of clear sky, or 0 okta, for long enough durations. The probability of 0 okta transitioning to 0 okta for Cambourne was shown in figure 3.9 to be 0.71. Therefore, for a whole 24 hours of consecutive 0 okta, as is possible in reality, the chance of this being reproduced within the SDSIG is $0.71^{24} = 0.027\%$, or 1 in 3703 days. This is a potential weakness in the SDSIG.

The K-S test results on the metrics, calculated in the same manner as in section 4.3.1, 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 okta correlations from Smith et al. (2017) are not suited to all geographic and climatic regions. Alternatively, the use of the Burr distribution for all $k_{c,clear}$ allows for a wider range of achievable values. This means that it is possible for days to differ more significantly for the Oahu distribution, resulting in a failure.

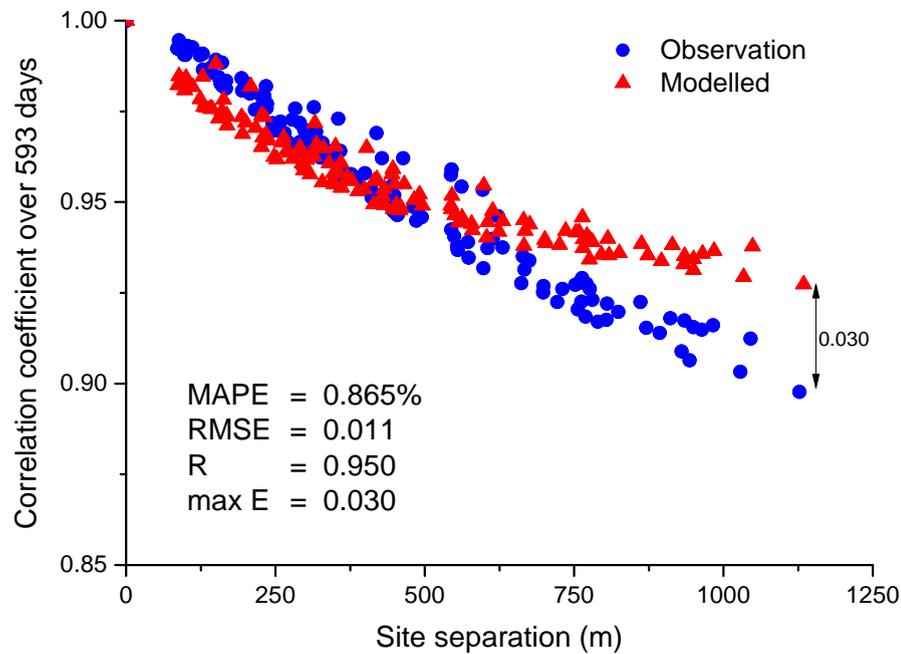


FIGURE 4.15: Correlation coefficients of individual station-pair w min GHI time series for 593 days plotted against the station-pair site separation. Every station-pair combination of the 17 site from the Oahu Solar Measurement Grid is represented as a symbol. Key: mean absolute percentage error (MAPE), root mean square error (RMSE), correlation coefficient (R), and maximum observed error (max E).

The 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) as was discussed in section 3.4. The correlation is plotted against the straight line separation between pairs and shown in 4.15 and displayed in equation 3.62.

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

The SDSIG performed similar 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 $R = 0.9523$. The maximum observed error in correlation was 0.03, or 3%; this is a strong performance. There is an clear disparity between the two curves. The correlation between observed and modelled data is well matched over 500m, however, divergence exists beyond. Both curves behave exponentially with site separation, as was also observed by Wiemken et al. (2001) at similar magnitudes. The cause of this separation is thought to be the cloud direction.

[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° . 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° , 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 [4.3.1](#).

4.4 Discussion of the Spatially Decorrelating Solar Irradiance Generator

This section will present a discussion of the successes and limitations offered with the development of the SDSIG. The spatial correlation is illustrated and discussed. Each methodological update within the SDSIG is discussed in turn. Lastly, a simple case study to demonstrate typical outputs and usage is performed.

4.4.1 Illustration of the instantaneous spatial correlation

Four simulations are run to explore the instantaneous spatial correlation. (1) Running the SDSIG with no influencing or weighting 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 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 correlation is assigned as 1, otherwise an anti-correlation of -1 is assigned for opposing states. Taking the mean of these correlations over a time period determines the overall instantaneous correlation of all locations to the reference point over that time scale. This is achieved by adopting a systematic grid-reference approach and testing each point individually against the reference.

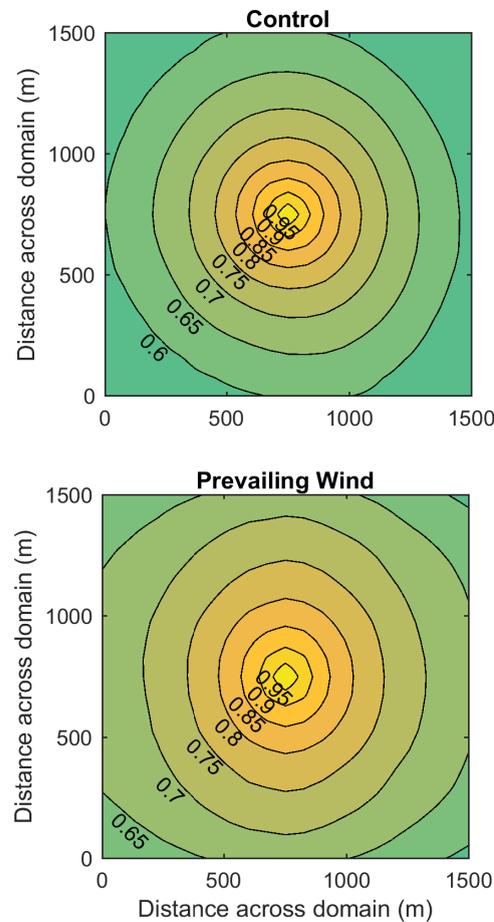


FIGURE 4.16: Spatial correlation between centre point to all other locations within the spatial domain. The top plot is the control simulation where no variables were fixed, whereas the bottom plot has the wind direction fixed from the north (top of plot). The numbers indicate the correlation along the labelled contour.

The instantaneous correlation, as used by [Arias-Castro et al. \(2014\)](#), does not consider time delay of ramping events as with many other literature. Instead, it expresses the probability of two sites to be covered by cloud (not necessarily the same cloud) at any one instant. This type of correlation cannot be compared to the curves shown in the section [4.3.2](#) because that considered the correlation coefficient of extended irradiance time series. The purpose of the instantaneous correlation is for illustration and discussion.

Figure [4.16](#) shows how the spatial correlation manifests across a time period of a year. The reference point is selected as the central most point within the spatial domain at (750,750). The prevailing wind scenario over a year shows less decorrelation with site separation as well as a tendency for anisotropy. The control scenario is more isotropic.

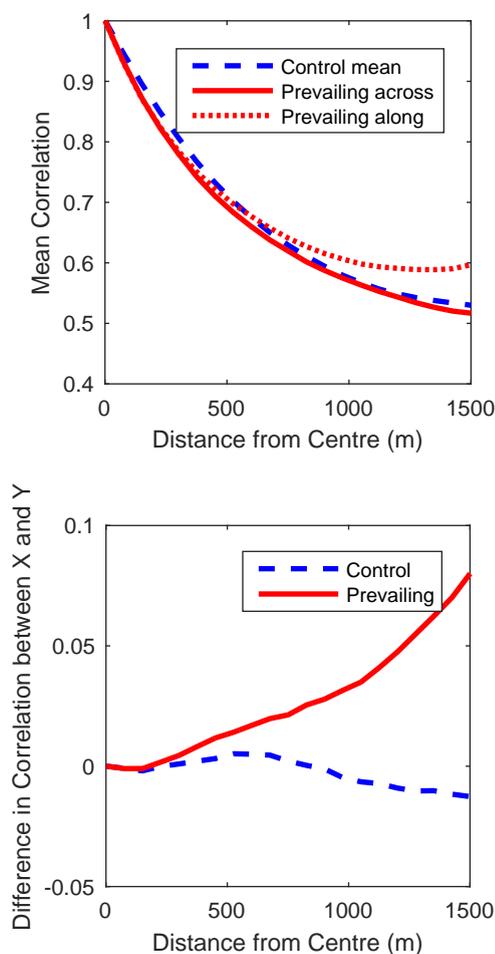


FIGURE 4.17: Line plot of the correlation from the reference point to the edge of the spatial domain for the control scenario (blue dashed line) and the prevailing wind scenario (red solid or dotted line). Top: mean correlation over the domain for the mean of X and Y directions in the control scenario and the individual along and cross wind profiles for the prevailing scenario. Bottom: difference in correlation between the individual X and Y directions for both the control and prevailing scenarios.

Further assessment was carried out by relocating the reference point to the north-western most point within the spatial domain at (1, 1500) so that the decorrelation over a longer distance can be examined. This can be seen in figure 4.17 where the top plot demonstrates how correlation in both the along and cross 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.

Figure 4.16 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

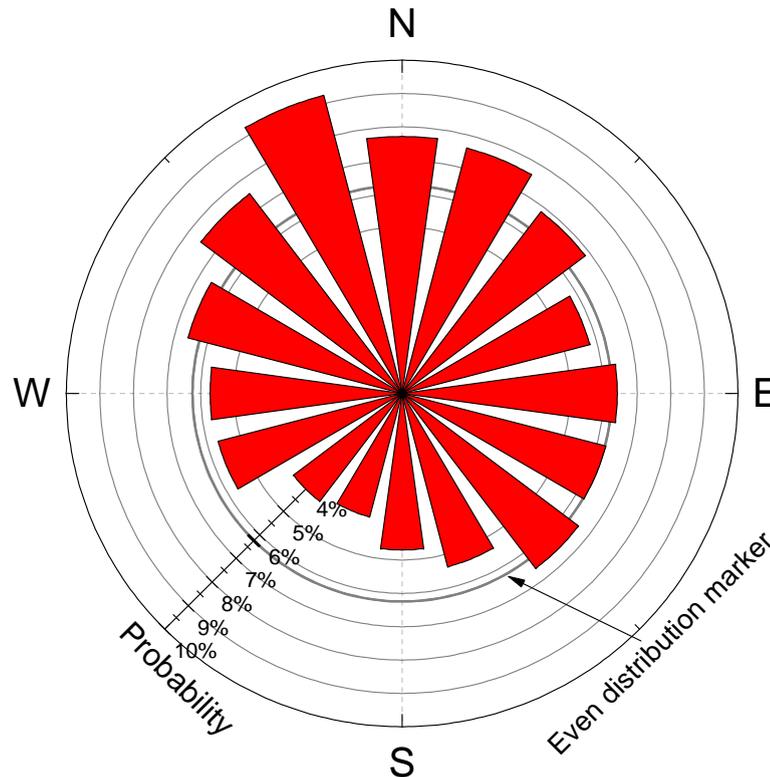


FIGURE 4.18: Wind rose showing the probability density of the hourly cloud motion approach direction over 1-year simulation. Each bin represents 22.5° . The direction indicates where the clouds are travelling from. The grey thick grey even distribution marker is a probability of 6.25% and represents perfectly even distribution.

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 figure 4.17 where the correlation differences in the along and cross directions are plotted against distance from the reference point within the spatial domain. The anisotropy is evidenced by the diverging levels of correlation in the different axial directions. For the control scenario, a steady fluctuation around a 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 4.18 where the control simulation's hourly cloud motion direction for the year is shown on a wind rose. Clouds 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

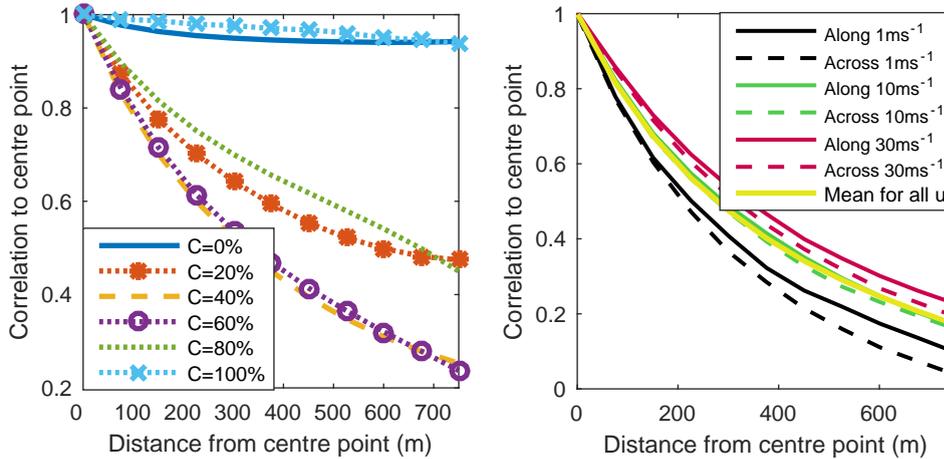


FIGURE 4.19: Line plot of the mean instantaneous correlation from the centre point to the edge of the spatial domain. The left plot shows the results of simulations with fixed cloud coverage percentages and all other variables operate as normal. The right plot shows how the along (solid) and cross (dashed) wind correlation at different cloud speeds, as well as the mean 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.

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 spatial correlation is no longer influenced by the size of cloud and cloud motion direction, the 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.

Figure 4.19 shows how the correlation from the centre point to the edge of the domain is influenced by fixing N or the u_{ref} . Separate simulations were performed for each line plot. The scenario for the left plot of figure 4.19 was the same as the control scenario conditions with exception of the N , which was incrementally increased each simulation and fixed for the entire year. It can be observed that for both fully overcast and fully clear scenarios, the 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 fields 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 correlation for 40 to 60% cloud cover scenarios, which undergo a steep decline in correlation from 1 to 0.25 over 750m. Cloud coverage constants of 20% and 80% share a similar correlation regression to each other, dropping < 0.5 over 750m. 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 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 greater probability of synchronised ramping events (reduced geographic smoothing), which are of high concern to 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 because the frequency of ramps is much lower. The more decorrelated an area the less synchronised the ramps. What is evident from figures 4.16, 4.17 and 4.19 is that to achieve favourable, decorrelated conditions, the separation between locations must be maximised.

The simulation conditions for the right plot of figure 4.19 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 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. Cloud speeds of $u_{\text{ref}} = 1, 10$ and 30 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 ms^{-1} dominate the simulation cloud speed frequency. The instantaneous spatial correlation increases with u_{ref} ; analysis intervals were 5 ms^{-1} . As C , u_{ref} 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_{ref} in both the along and cross wind directions. The cause of this is suspected to be that, 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_{ref} . 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.

4.4.2 Cloud fields

The most distinctive element of the SDSIG is the introduction of a spatial dimension. This is achieved through the development of cloud fields that pass across a 1.5 km^2 spatial domain.

A literature review on synthetically representing a cloud field found very few options of synthesis. For this reason, a new method was devised whereby clouds were assumed to be circular with centre point randomly located inside a rectangular domain with radii extracted from [Wood and Field \(2011\)](#).

The following assumptions were used to generate the cloud fields:

1. Clouds can be represented by circles
2. A two-dimensional approach can be representative of a three-dimensional property
3. Only a single layer of cloud is used
4. Clouds move in a constant direction and speed for the whole hour
5. An okta value is fully representative of the sky for the whole hour and across the spatial domain
6. An estimation of the coverage is sufficient to categorise a cloud field
7. Overlapping clouds should be taken as the mean of clear-sky indices, not the multiple
8. The computational maximum number of clouds is representative of the sky for the hour

It is very difficult to test the validity of some of these assumptions because there are no alternatives to test against. The validation of the SDSIG shows that the VI, IF, RR

and KCI metrics perform excellently with both the K-S test and correlation coefficient comparisons with observation data.

Some of these assumptions are clearly false in reality, such as the first assumption that represents clouds as perfect circles. The need for representing them as a simply-definable for computational use, alongside the lack of literature offering alternatives, necessitates the assumption. There are alternative geometric assumptions that could have been used such as squares or ovals, however, there is as little support for any geographic shape. Circles were used by [Arias-Castro et al. \(2014\)](#), however, no supporting justification was offered. The reason for selecting a circle for the SDSIG is that the area of a circle can be defined using a single parameter, r , and so is the least computationally demanding geometry to produce and use. It is accepted that using a circular cloud sacrifices description of the cloud type in effect. However, the distributions of $k_c(N)$ facilitate the engineering of cloud types. The cloud type is not a variable accessible by the model and such, a circular cloud of very high k_c would be more indicative of high cirrus clouds more-so than thicker cumulus.

The second assumption is that the two-dimensional discs represent a three-dimensional cloud. The third spatial dimension to a cloud is given through the addition of an optical depth using k_c . The information required to generate accurate three-dimensional representations of clouds would move the SDSIG away from the initial rationale of being able to produce irradiance time series from readily available simple inputs. Information on atmospheric pressure and temperature gradients as well as cloud types and weather system progression would be required to capture the clouds. The use of k_c for the third-dimension, at least, has some statistical background. Ultimately, increased complexity to the clouds would need to add value to the validation of the IFs. As the validation demonstrates a strong validation for the IFs, no future work is recommended in the addition of a third-dimension to the cloud fields at this stage.

The third assumption that only a single cloud layer is used can potentially influence the true spatial correlation. This is another assumption that is incorrect in reality. From the beginning of this thesis, the author has witnessed many days where there are three layers all moving seemingly independent of each other. The presence of three cloud layers is a regular occurrence and the variables of motion direction, cloud speed or cloud amount do not necessarily correlate between layers. Some databases report on the cloud amount over three layers of cloud, assigning a total cloud amount to each layer, such as the NOAA QCLCD and UKMO MIDAS data from BADC. In practise, three layers could be simply achievable within the SDSIG by repeating the cloud motion three separate times and collating the resultant \mathbf{B} and \mathbf{k}_c , however, it would take three times the duration to produce with little evidence of added value.

Sites that record N using ceilometers report it for each cloud layer and so there is potential to identify relationships of N between the layers, however, the issue still exists for direction and speed. N is deemed to be representative of the sky regardless of the number of layers. 8 okta represents 100% coverage with any number of cloud layers, and 5 okta is assumed to mean 50% coverage considering every layer. For this reason, the exclusion of multiple cloud layers is assumed to only influence direction and speed. What can be deduced from the illustration of spatial correlation in figure 4.16 is that the prevailing wind scenario demonstrates the anisotropic correlation over 1.5 km² and can be considered the baseline without considering cloud direction across multiple layers. With multiple layers moving in individual directions, the correlation will appear closer to isotropic as in the control scenario. With more cloud directions, the more isotropic the spatial correlation becomes, as was demonstrated in the bottom plot of figure 4.17. For these reasons, the future inclusion of multiple cloud layers is not considered to impact significantly the time series outputs, particularly while there is little or no supporting research to facilitate its inclusion using simple inputs or statistical relationships.

The fifth assumption comes from Morf (2011) who found that a point in the sky will converge on the okta value over time. This finding justified the use of a synthetic method to represent the binary sun obscured nature in a model. To analyse whether the cloud field method performs as suggested by Morf (2011), a 4 year simulation was performed for a house positioned in the very centre of the spatial domain. Every hour, the hourly C value is compared to the actual C experienced at the house, C_h . Figure 4.20 (top left) shows a histogram of the fractional difference between C_h and the cloud field's C . It can be seen that C_h most often matches the cloud field's C over an hour, it resembles a normal distribution with mean around a fractional difference of 1 (indicating no difference). There are some surprising instances where C_h is nearly 3 times greater than the cloud fields C . In order to explain this, further analysis across the coverage range is performed. Figure 4.20 shows histograms of C_h for each different cloud field C indicated in bold above each plot. $C = 10$, or 100% coverage, experiences the greatest accuracy with very few instances of C_h falling below the cloud field's C . Interestingly, it shows how even under predominantly covered skies, a house has the potential to experience much lower coverage. To explain those instances whereby the house can experience C_h that is > 3 times greater than the cloud field's C , we can observe the histograms for $C = 1$ to 3, where it is possible to experience the whole range of possible coverages within the cloud field. For $C = 6$ to 9 there is a high probability for the house experience $C_h = 10$. To understand the wide range of C_h values when targeting a particular C , figure 4.21 shows analysis of two potential paths that a house could take through a cloud field. This idea of a path experiencing a

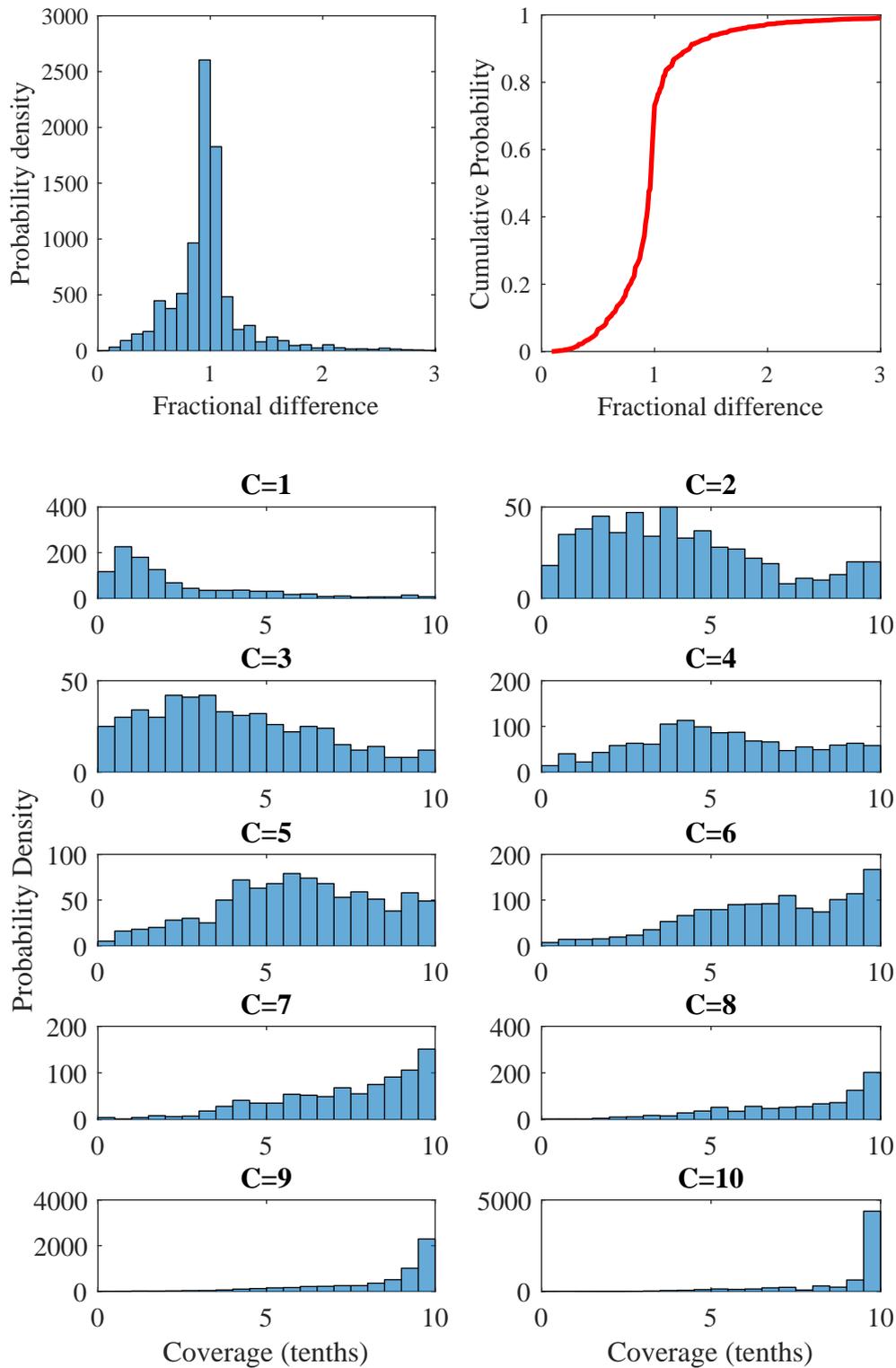


FIGURE 4.20: Top left) histogram of the fractional difference between the actual coverage value experienced at a house in the centre of the spatial domain and the hourly coverage value of the whole sky. Top right) CDF of the histogram in top left. All below) histograms of the actual experienced coverage, C_h in tenths for each target C value indicated in bold above each plot.

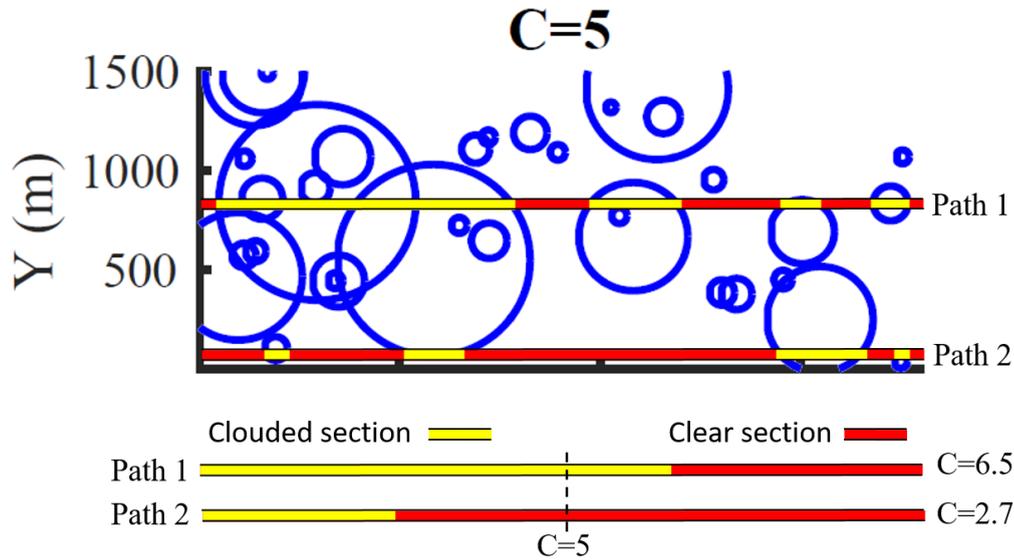


FIGURE 4.21: Analysis of the $C = 5$ cloud field taken from figure 4.4. Two potential paths that a house may take through the cloud field are indicated. Clouded sections of the path are shaded yellow and clear sections are shaded red. The actual coverage value experienced on both paths are shown below the cloud field.

different C to the overall cloud field is also seen in reality with the method of recording N . N can be measured by either an automated ceilometer or a human observer. The human observer is found to be more accurate as they can easily scan the whole sky to estimate N (Smith et al., 2017). The ceilometer measures the path such as in figure 4.21, and will experience greater distribution of errors. The SIG has a more accurate method of synthetically generating C to match observation. Accuracy here is defined as the 1-min time series of C achieving an hourly mean average cloud cover closest to the target okta value. However, the added spatial dimension with the cloud field method allows for realistic variations in C_h .

To assess the convergence over time, figure 4.22 plots the moving average of C_h throughout an hour. As Morf (2011) suggested, a point location within the cloud field converges close to the a solution overtime. However, what is found is that C_h is not always that of the cloud field with the exception of $C = 10$. Solutions typically converge within 20 minutes, although infrequently on the correct C . This means that Morf (2011) is correct in suggesting that the solution will converge over time, although is incorrect in suggesting that it will converge on the correct okta within the hour, particularly for $C = 2$ to $C = 7$. An important consideration is the influence of cloud speed. As the area of the cloud field is directly proportional to the cloud speed ($X = 3600u_{\text{ref}}$), larger u_{ref} produces larger sample size with which to estimate the cloud field's actual C value and is therefore more accurate at estimating. The variability in C_h is not considered to be a negative of the SDSIG, in fact, it is considered by the author to add more realism to the outputs.

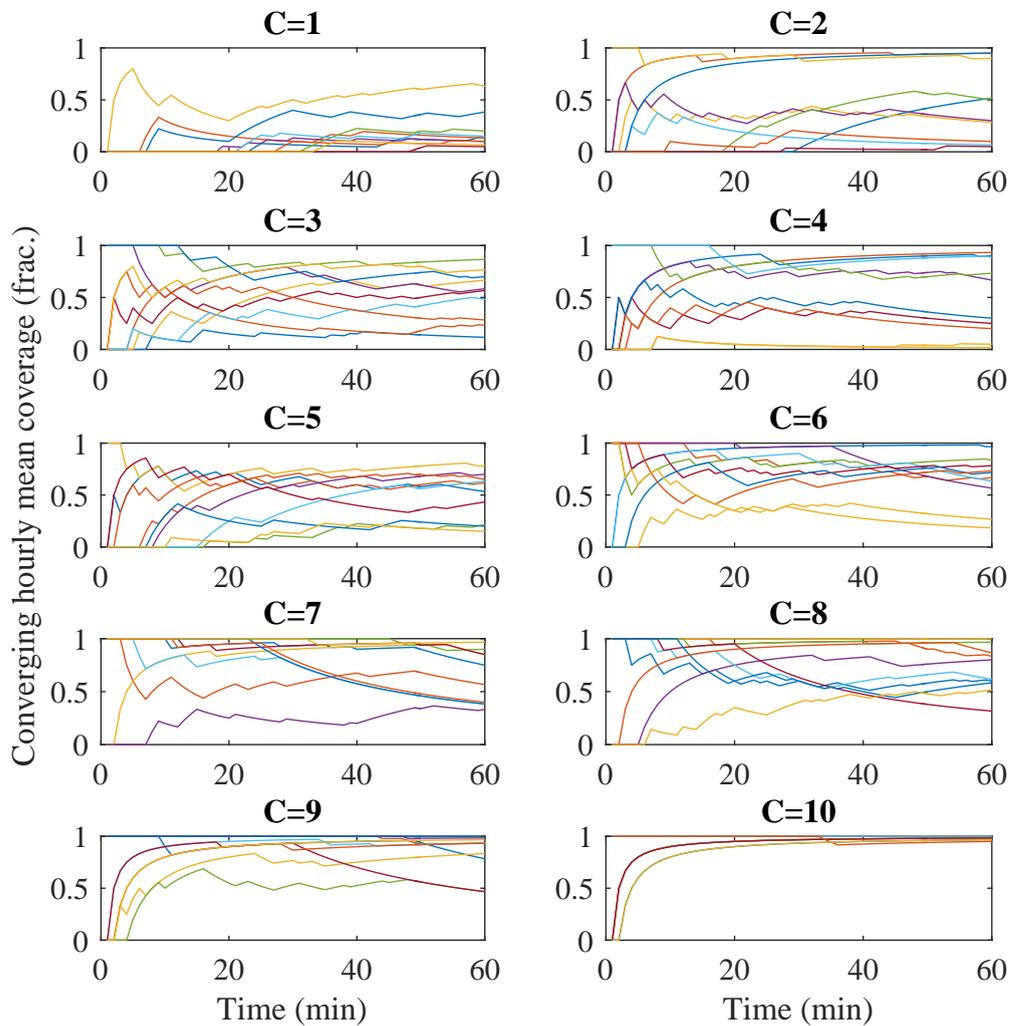


FIGURE 4.22: Convergence of the coverage value over time. Each line represents an hour of simulation where the experienced coverage develops over time, the solution is calculated as a moving average. The solutions are separated by the target coverage values for the whole cloud field. These plots complement the method shown in figure 4.21.

The sixth assumption, that an estimation of the coverage is sufficient to categorise a cloud field, was analysed and presented in figure 4.23. The method presented to estimate C was shown in figure 4.5 whereby a spatial resolution was used to estimate the C of a cloud field. The SDSIG used a 50 m resolution to estimate a cloud field's C . Figure 4.23 shows analysis of the difference in the grid resolution used to estimate C in order to establish the accuracy of this assumption. There are two distinct conclusions to draw from this figure. The first is that there is no significant difference between estimating the coverage using a 1 m grid mesh resolution or using 250 m. The exact solution presented in equation 4.2 can be assumed to be found with a resolution of 1 m, therefore, an estimation using a resolution of up to 250 m is suitable for the model. The cloud fields used in the validation were produced using a resolution of 50 m, this

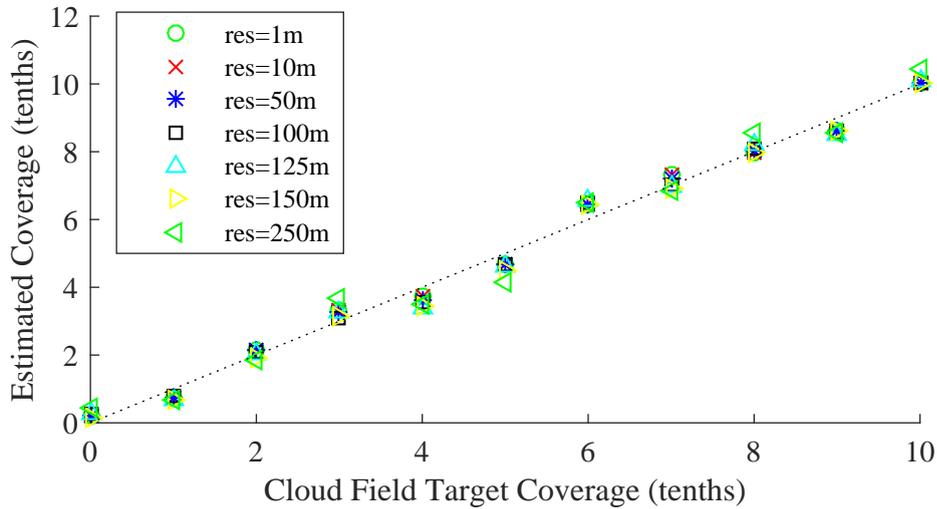


FIGURE 4.23: The estimated coverage value versus the target coverage value for increasing spatial grid resolution, res . A line along $x = y$ is provided for reference.

was at the cost of increased computational demand, however, it assured that accuracy and confidence were maintained. The deviation away from the $x = y$ line is due to the inclusion of $C_t \pm 0.5R$ as discussed in section 4.2.2.2, the important information from the figure is how close together the points are.

The seventh assumption that overlapping clouds should be taken as a mean was justified in the methodology section 4.2.2.2. There is no method of assessing the extent of this assumption without detailed clear-sky indices with associated cloud thicknesses.

The eighth assumption, that the computational maximum number of clouds is representative of the sky for the hour of cloud, poses a problem. Setting a maximum number of clouds when analysing a high cloud speed and high okta value ($u_{ref} = 30$ and $N = 7 = 87.5\%$), the area, A , required to be covered by cloud is $1500 \text{ m} \times 3600 \text{ s} \times 30 \text{ ms}^{-1} = 162,000 \text{ km}^2$. The maximum number of clouds in a cloud field is set to $n_c = 1200$. The area of the cloud is found as πr^2 and so to determine the average \bar{r} required per cloud (assuming no overlap) in order to achieve a cloud field within $\leq n_c$ clouds is found to be 194 m using the following equation.

$$\bar{r} = \sqrt{\frac{AN}{n_c} \frac{1}{\pi}} \quad (4.20)$$

Overlap within the SDSIG is a extremely likely, as was easily demonstrated in figures 4.21 and 4.4, and so realistically $\bar{r} = 194 \text{ m}$ would be too small an estimate. Figure 4.24 shows the distribution of r from the Wood and Field (2011). The mean r produced in 1200 clouds was found to be represented by a gamma distribution of $\Gamma(0.61134, 9.0875)$ where the modal radius was 400 m, suggesting that the choice of

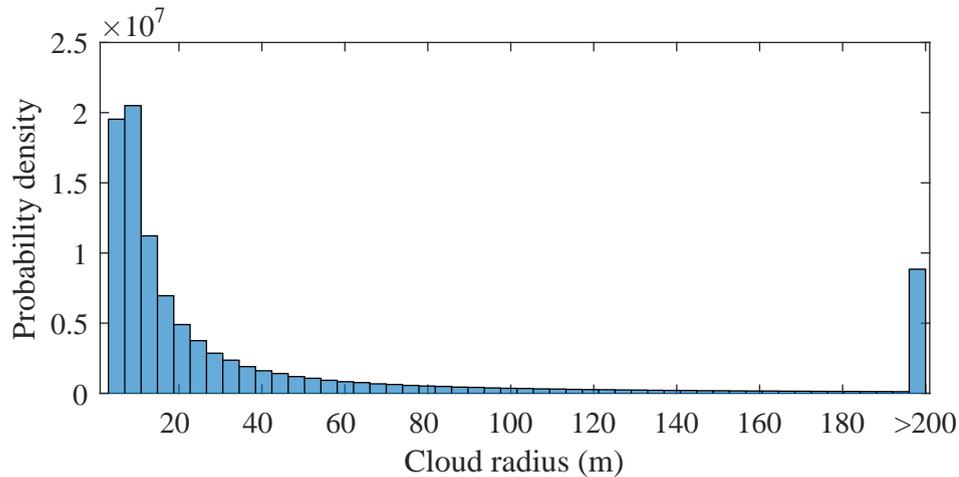


FIGURE 4.24: Histogram of cloud radii as drawn 10^8 times from the [Wood and Field \(2011\)](#) distributions and adjusted to apply limits. All radii above a limit are combined in the final bin.

1200 clouds is satisfactory. This was found using distribution fitting software from [Matlab \(2015\)](#).

The implication of the validation is that the cloud field method, inclusive of all its assumptions, is successful in answering the research question that it is possible to produce statistically accurate, synthetic irradiance time series that vary on both a temporal and spatial domain of the thesis. Furthermore, the cloud field method helps satisfy the research aim to develop a methodology that incorporates a spatial dimension, facilitating the production of multiple, spatially decorrelating irradiance time series. The success of the cloud field is best demonstrated with the RR accuracy, the calculation of the RR is heavily dependent upon the cloud field and so success here demonstrates the suitability of the cloud field methodology in synthetically generating a real world statistic.

4.4.3 Cloud motion

The cloud motion development facilitated the passage of cloud fields over a spatial domain containing the houses. The spatial domain is limited in size by the cloud field size, which is itself only limited by computational storage and power requirements.

The cloud motion section of the model was developed with robustness in mind and so it is designed such that it is flexible to both spatial and temporal resolution changes, meaning that any size cloud field or time step under an hour can be managed. Time steps over an hour can be used, however, it would be found by taking the average of a higher resolution and so there would be no benefit of using the SDSIG. There

are two identified limitations to the cloud motion methodology. The first is potential overestimation of cloud between hours and the second is a limitation to where houses may be positioned within the spatial domain.

Firstly, there is a potential over estimation with clouds centred at the edge of a field being accounted for in adjoining cloud fields as there is no hard severance of cloud between hours. It is possible and computationally simpler to include this severance, however, clouds will simply disappear or appear at every hourly transition, as such can cause synchronised ramping events at the edges of the domain that are to be avoided. The correlations are truer to the circular geometries by allowing each cloud to fully pass across the spatial domain; however, there is the potential for overestimation of cloud with cloud centre $x_c < r$. This could be particularly significant if a cloud of cloud size x_{\max} is situated near the edge of the cloud field. It would influence upon the C value of adjacent cloud fields. If this potential over estimation of cloud cover is significant, it is not evident in the validations shown in figure 4.13. It is perhaps responsible for the wide breadth of C_h in the histograms of figure 4.20.

A further potential overestimation exists depending on the type of okta recording method used for the input data. When a human observer records N , the slightest presence of cloud is not treated as $N = 0$, instead $N = 1$ is assigned. When cloud fields are being made, the target coverage value, C_t , has a $\pm 0.5R$ inclusion. This means that some clouds will be added to the cloud field for $N = 0$ targets. When using ceilometers, however, this inclusion is fair as ceilometers can only track a single path along the sky (Smith et al., 2017). These overestimation potentials are not evident in the validation and not considered to be detrimentally significant.

Secondly, there is a limitation with limiting the spatial domain where the houses can exist. Because the method of simulating cloud direction is to rotate the houses instead of changing the approach direction of the cloud field, houses outside of a circle centred at (750,750) with $r = 750$ m, when rotated, will exist outside of the spatial domain. This is demonstrated in figure 4.25. A hard limit is placed that moves any house that falls outside of the spatial domain back within the limits to the nearest satisfactory (x_l, y_l) . The maximum displacement, denoted as q in figure 4.25, can be found as $q = |0.5(\sqrt{2}X - X)|$; for $X = Y = 750$ and $q = 220$ m. A potential solution, which would redefine the the shape of spatial domain in where houses can exist is changed to be a circle that fits perfectly to the XY domain defined as $(x, y, r) = (0.5X, 0.5Y, 0.5X)$, this would remove any displacement. A better solution would be to use more computationally powerful equipment and extend the study to a larger spatial domain.

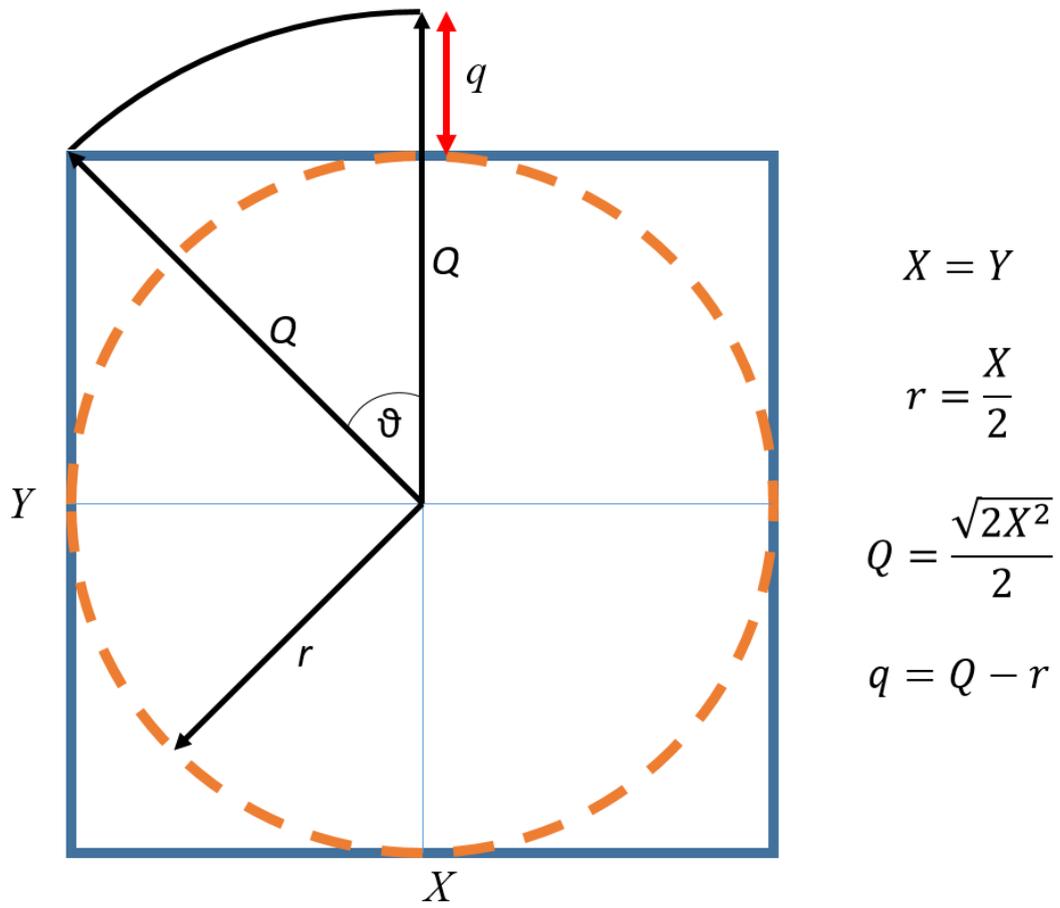


FIGURE 4.25: Demonstration of how a house, when initially located at $(0, Y)$, can be positioned outside of the spatial domain when rotated by the cloud motion direction, θ . The method of determining the maximum distance outside of the domain, q , is detailed.

4.4.4 Clear-sky indices

k_c generation within the SDSIG differs from SIG in three distinct categories: synthetic fluctuations, applicability to different climatic regions, and CEE events. These will be discussed in turn.

4.4.4.1 Clear-sky indices fluctuations

k_c within the SDSIG are *cleaner* in comparison to the SIG, in that fewer adjustments are made. The SIG applied Gaussian noise and 6 min fluctuations. Both these methods are removed as there is not enough research to support their use besides their successful implementation in the validation of [Bright et al. \(2015\)](#).

The fluctuations in k_c are inherent to the cloud field methodology. Each cloud is assigned a k_c from the new $k_c(N, \theta_e)$ distributions. Due to the nature of taking the

mean of overlapping clouds, variability is achieved without the need to add in Gaussian noise or fluctuations.

The only synthetic change to the k_c time series after the cloud motion is that of smoothing events of extended 0 and 8 okta periods. The SIG employed a linear interpolation between points; however, better and more realistic interpolations exist to facilitate smooth and gradual changes in k_c . Using linear interpolations increases the daily variability index and lowers the probability of achieving a VI=1. When using a smoother interpolation, for example a spline or a pchip, these lower VI are improved. The VI was the weakest metric performance from all validation locations and so this section is considered to be highly important. Using the pchip over the linear interpolation does contribute well to smooth irradiance time series, however, it is longer term statistics of higher order Markov chains that may facilitate more realistic daily VI values.

4.4.4.2 Differences between climate regions

The SDSIG was applied in different climatic regions and performed well according to validations. A key assumption to facilitate this was that the distributions of $k_c(N, \theta_e)$ are globally applicable. This assumption requires substantially more research to validate, however, some intuition into the subject can enable discussion.

The different okta conditions are made by weather systems and, therefore, cloud types that are inherent to weather systems. Weather systems are typical to the region that they exist in with the addition of extreme events. Considering the UK, a thin blanket layer of cloud can result in 8 okta, furthermore a thick layer of cloud will also cause 8 okta. Both of these conditions have inherent cloud types that influence the wide distribution of mean hourly k_c . Therefore, for this distribution to be applicable to a different climatic region, a similar construction of cloud types are required to recreate the same 8 okta distribution. So the driving question that influences this assumption is *is the total cloud amount from different regions made up of similar cloud types?* This question has not been answered in this thesis, however, there is indication of this being the case between the UK and San Diego as was suggested by the KCI validation CDFs in figure 4.13.

One assumption involved the derivation of the k_c distribution for clear periods. For the UK, moments of clear sky were found to be satisfied using $k_c \sim N(0.99, 0.08)$, however, this had to be revisited when repositioning the SDSIG to San Diego and Oahu. Analysing the irradiance data from San Diego, clear sky conditions were found to be satisfied using $k_c \sim N(1.02394, 0.04)$ and by $B(0.9813, 0.72.37, 0.1203)$ for Oahu. This is indicative of either the limitation of the clear-sky model by [Hammer et al.](#)

(2003) or inaccuracies in the Linke turbidity. The PSA by Hammer et al. (2003) is supposed to estimate the CSI for a location, however, it underestimates San Diego by 2.4% and overestimates the UK by 1%. Using these normal distributions, the IF validations for all locations had excellent performances in both correlation and K-S test. This is indicative of the need for a better estimation of CSI that is more globally applicable. Failing that, it would be possible to determine the CSI for every BADC site and select the nearest distribution from user defined inputs of Θ and Φ . Further investigation into this clear sky condition dependency on geography would be a useful addition.

4.4.4.3 Cloud edge enhancement events

The inclusion of CEE chance is not effective as there is too much uncertainty surrounding what is and isn't a CEE event. To appropriately determine whether CEE events ($\frac{GHI}{CSI} > M_{\min}$) is attributable to a cloud's edge requires the knowledge of the presence of the cloud, and so this simple estimation method is deemed to be insufficient as it cannot know whether a cloud is present. Intuitively, it can be hypothesised that the irradiance will *always* undergo some form of CEE event, even if the impact is negligible. The chance, determined with the simple estimation, merely describes the fraction of time that GHI is potentially undergoing CEE, it does not describe if a particular cloud edge will cause an enhancement event or not. Without a method to determine which data within the GHI time series are recorded with a cloud near to the sun's location within the sky, it is not sufficient enough to state with certainty that observed moments of high GHI are directly attributable to CEE. The magnitudes, M , derived are representative of real enhancement events and can be assumed to be caused by CEE, for this reason M can still be used to implement a CEE event. The distinct pattern of increasing M with both increasing θ_z and N can also be used. There must be a better method to determine the chance of CEE.

The use of sky-imagers could be utilised to assist in furthering the understanding of CEE statistics such as directly attributable magnitudes and chance of occurrence, however, this was outside the scope of this PhD project. Furthermore, PV inverters are typically oversized to manage CEE events. These modern technologies and grid architecture minimise the impact from CEE at a grid impacts analysis level. Therefore, the only added value to the SDSIG from more accurate CEE events is a better representation of a ramp-up or ramp-down events. As the primary aim of the thesis is to model irradiance, it is deemed by the author a useful line of further research.

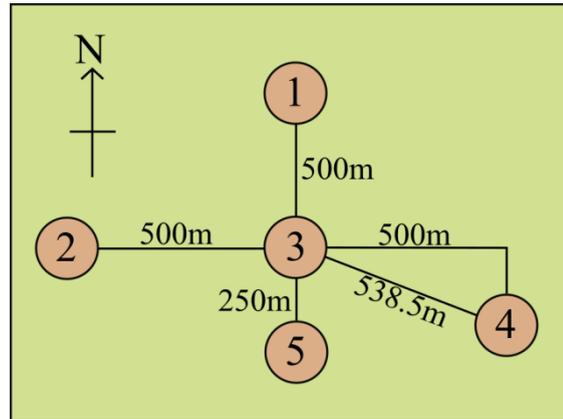


FIGURE 4.26: A theoretical set-up used to demonstrate the functionality of the SDSIG. The numbers indicate 5 different houses located in Cambourne, UK, for which spatio-temporally decorrelating irradiance time-series are generated.

CEE implementation was first developed for the SIG and a similar methodology was applied to the individual \mathbf{B} 1DMs for each house in the SDSIG. A future development is to employ a more sophisticated methodology that can incorporate the spatial influence of CEE. This would be more complementary to the vectorised cloud motion section in the SDSIG. A proposed methodology could be achieved by defining a banded ring around each cloud that is sized as a fraction of r_c . The 1 sec profiles of CEE produced by Lave et al. (2012) could be applied across the partial radius of the banded ring. Within the cloud motion methodology, whenever the time step lands within this banded ring, a CEE event can be applied adjusting by an extracted M from the CDFs made in section 4.2.4.3. This proposed method would be flexible to changes in both the spatial and temporal resolution. Due to time constraints within the PhD, this has not been explored further within this thesis, however, remains as an interesting future work consideration.

4.4.5 Simple case study demonstrating the use of the Spatially Decorrelating Solar Irradiance Generator

This demonstration was published in ? after being selected for publication from the International Solar Energy Society's special issue in the Journal of Solar Energy for the Solar World Congress 2015, Daegu, South Korea.

To illustrate how the SDSIG takes user inputs and delivers spatially decorrelating irradiance time series, the model is applied to a hypothetical configuration for the location of Cambourne, UK. Figure 4.26 illustrates the configuration.

Table 4.2 shows the physical parameter inputs required for each of the houses within the simulation. The x_l and y_l coordinates are noted in reference to the centre of the

TABLE 4.2: Physical parameters as user defined inputs into the SDSIG for the 5 houses in the case study as shown in figure 4.26. The letter C denotes the centre of the spatial domain at (750,750).

	1	2	3	4	5
x_l (m)	C	C-500	C	C+500	C
y_l (m)	C+500	C	C	C-200	C-250
z_l (m)	87	87	87	87	87
β_a^l ($^\circ$)	0	0	0	0	0
β_p^l ($^\circ$)	0	0	0	0	0

spatial domain, C, defined as (750,750). The azimuth angles are defined as -180° to 180° East to West, with South being an angle of 0° , although β_a^l does not influence this simulation as $\beta_p^l = 0$ and so all irradiance values are GHI.

The model is functional for all β_a^l and β_p^l with reference to the input Φ and Θ . The weather station in Cambourne, UK is at $\Phi = -5.32656^\circ$ and $\Theta = 50.2178^\circ$, with an elevation of $z = 87$ m above sea level. The temporal resolution of 1 minute has a spatial granularity defined as the smallest distance between houses where a difference in irradiance can be observed. This is determined using the temporal resolution of 60 s and the typical cloud speed of 5 ms^{-1} giving typical granularity around $300 \text{ m}\cdot\text{min}^{-1}$. The scenario depicted in figure 4.26 shows that the houses are separated by at least 250-1020 m and should, therefore, demonstrate decorrelation. In order to observe spatially decorrelating irradiance time series for houses closer together, a decreased granularity would be required. This is achieved by increasing the temporal resolution.

Figure 4.27 displays a typical output from the model using inputs from table 4.2. The profile is typical of a clear day in mid-January for the location of Cambourne, UK. Typical patterns can be observed associated with the spatial decorrelation. Most notably at around 700 minutes, there is a gradual ramp down in output at house 1 a few minutes before the other 4 houses undergo a similar ramp down, this can be attributed to the wind speed and direction at the time; the clouds were travelling north to south. At 860 minutes, house 2 is the only house to undergo a large ramp down event.

The implication of this is that, on aggregate, this ramping event may not significantly impact the voltage across a transformer. Were each property given correlating irradiance time series, e.g. all equal to house 2, all houses would be subject to the same ramp at 860 min resulting in understating the aggregated irradiance. The SDSIG provides individual irradiance profiles with individual house installation characteristics of β_a^l and β_p^l . It is therefore ready to be applied to a grid impact analysis.

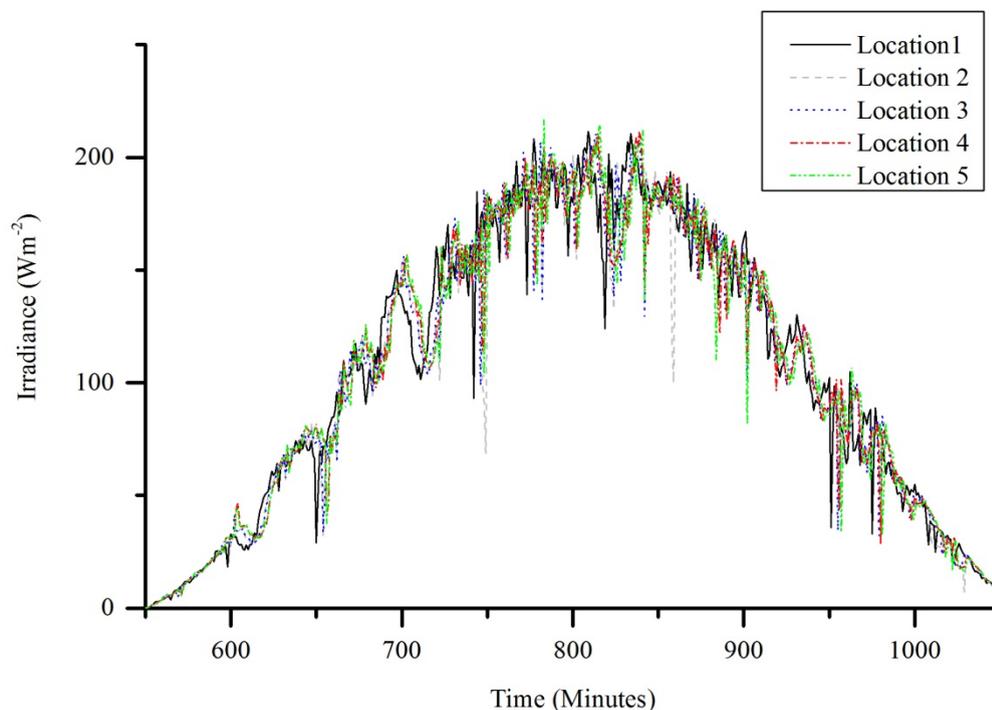


FIGURE 4.27: Synthetic decorrelating irradiance profiles for 5 houses on the 10th Jan from SDSIG located in Cambourne, UK.

4.5 Application of the Spatially Decorrelating Solar Irradiance Generator in a grid impact assessment case study

This section was produced collaboratively with Oytun Babacan and Prof. Jan Kleissl on a research scholarship to UCSD. The power flow model presented was constructed by Oytun Babacan, although assumptions, case study design, interpretation of results and presentation are collaborative. This section is published in the Journal of Solar Energy [Bright et al. \(2017\)](#).

The final research problem of this thesis is to demonstrate the applicability of the SDSIG outputs in application to a multi-variate grid impacts analysis and explore one of the identified grid impacts. To do this, a grid impact assessment was devised.

4.5.1 Overview of the grid impact study

The power flow simulation simulates the electricity flows around a typical LV distribution network feeder spread over 1.5 km^2 with 2438 nodes that each have the potential

for PV installation.

In order to demonstrate the advantages of spatially decorrelating irradiance time series over correlating irradiance time series, 2438 spatially decorrelating irradiance time series are generated from the SDSIG and three scenarios are proposed.

Firstly, a baseline is established by running the power flow model with no PV installations and so irradiance has no generational impact on the electricity grid. Secondly, spatially decorrelating and temporally unique synthetic irradiance time series from the SDSIG are assigned for each available PV system in the distribution system. Lastly, a single correlating irradiance time series, randomly selected from the SDSIG, is applied to for all available PV systems in the distribution grid.

The simulations are performed as individual daily simulations for a complete year (365 days) of data and for varying solar PV penetration levels. The location chosen for the study is UCSD, CA USA. Power flow simulations are conducted using OpenDSS ([EPRI, 2008](#)), an open source electric power distribution system simulator.

4.5.2 Test circuit information 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 in literature. The test case description states that the circuits are realistic ([IEEE, 2013](#)). The IEEE test circuits are anonymised actual circuits from throughout the USA. They are realistic such that they are in fact real, however, the data is protected. The purpose of using an American grid was to boost both the collaborative initiative with UCSD as well as aiding the validation and applicability to the San Diego validation site. A diagram of the test circuit is shown in figure 4.28. 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 possible distance from the substation is approximately 17 km and the circuit has a peak load of approximately 10.7 MW, at which it exhibits approximately a 10% loss.

PV generation installations are 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. Their pitch angles are kept at 20° and their azimuth angles are kept at 0° (due south).

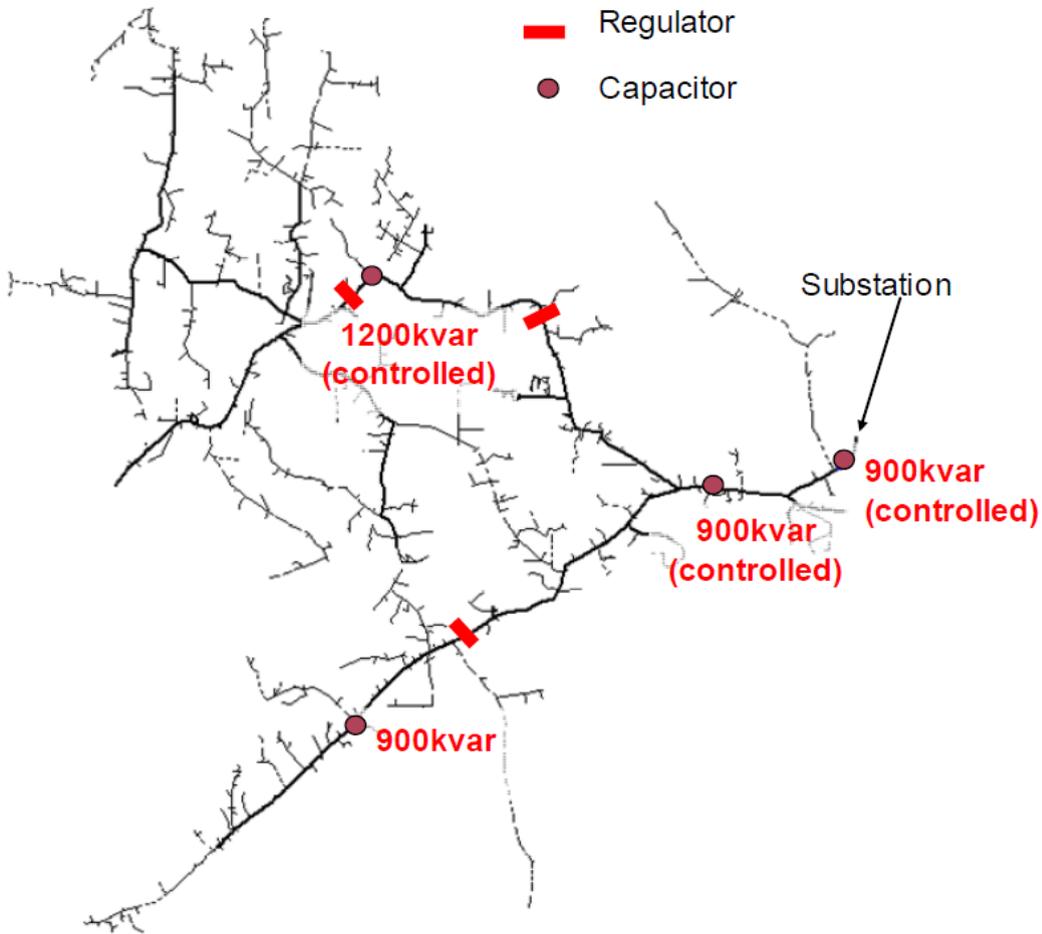


FIGURE 4.28: On-line diagram of the 8500-node feeder circuit used in the power flow simulations (IEEE, 2013).

PV systems are sited at randomly selected load points in the circuit until a desired PV penetration is achieved. As the PV penetration level in the distribution system is increased, already existing PV systems are retained and new systems are added in a similar random fashion until the new desired PV penetration is achieved again. The PV penetration (PVP) definition used in the grid impacts study is:

$$\text{PVP} = \frac{\sum_{m=1}^M P_{PV}^m}{\sum_{n=1}^N P_{load}^n} \quad (4.21)$$

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

system voltage targets during high solar and low demand periods on certain simulation days.

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

4.5.3 On-line tap changer operations grid impact metric

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 house 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.

As was well researched and detailed in chapter 2, there is no consensus metric developed to compare grid impacts. The number of OLTC operations was identified as an appropriate metric to compare grid impact analysis scenarios and for this reason is used to demonstrate the impact of increasing PVP upon the distribution grid. In each power flow scenario, the accumulated depth of required OLTC operations at all voltage regulators are recorded and compared. The depth of OLTC operation is the magnitude (depth) per operation.

4.5.4 Results of grid impact assessment

The resulting daily OLTC operations depth are shown in figure 4.29 for a year of simulations for varying PV penetration levels. The baseline *No PV* case has a median OLTC depth of 187 per day, with 4 outliers and a maximum depth 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 figure 4.29. The results for all PVP levels are as follows:

- 25% PVP simulations
 - *red case*: 18 outliers with a maximum of 477 per day
 - *green case*: 9 outliers with a maximum of 365 per day
- 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
- 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 those of *green cases* with increasing PVP by 10, 49, 93 OLTC operation depth per day, respectively

For both the correlating and decorrelating irradiance time series, OLTC operation depth increases with PVP. However, the magnitude and frequency of severe tap changing events rise significantly faster in the simulations using the correlating time series. In very high PVP cases, such as a distribution system with 75% PVP, using correlating time series results in cases where voltage regulators change taps up to an accumulated depth of 1249 compared to an accumulated depth of 697 when using decorrelating time series, giving an 80% potential overestimation. Such extreme results would result in overly conservative PV impact mitigation measures for the distribution system in question, which would be detrimental to PV uptake.

4.6 Chapter summary

The research problem addressed in this chapter asked if it was possible to produce statistically accurate, synthetic irradiance time series that vary on both a temporal and spatial dimension, facilitating multi-variate grid impact analysis. The first objective of this chapter was to address this question through the development the SDSIG, a spatio-temporal methodology that meets all the criteria demanded in the scope of this thesis. The second objective of this chapter was to demonstrate the applicability of the

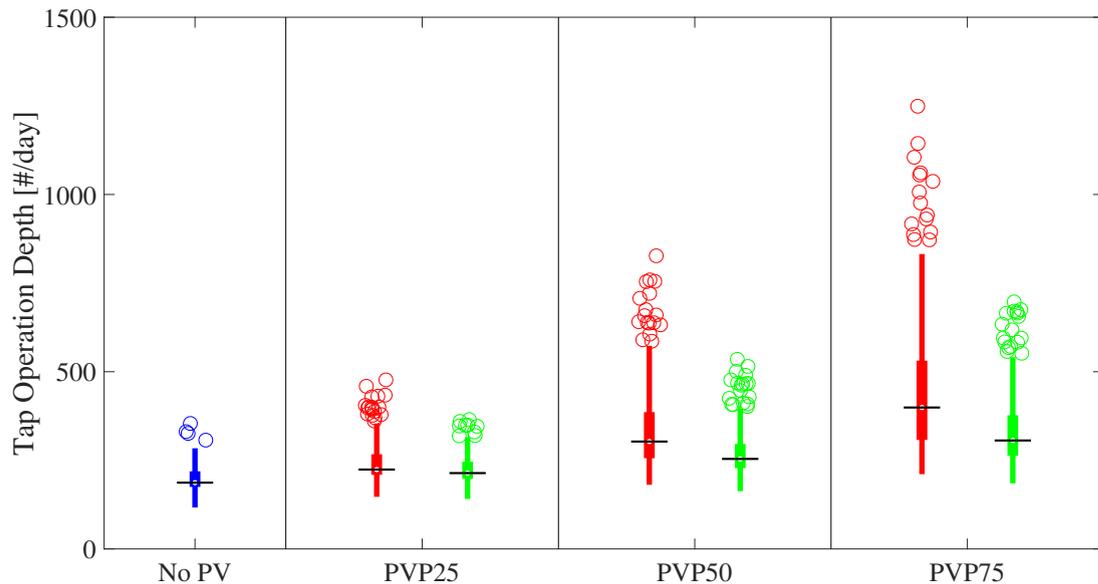


FIGURE 4.29: Number of tap operations over a day for a complete year (365 days) of simulation under varying PV penetration level in the IEEE8500 node distribution system (0%, 25%, 50% and 75%). Red represents the power flow scenario using the single correlated time series data across the distribution system and green represents the scenario using the spatially decorrelated and temporally unique synthetic GHI time series from the SDSIG. Blue 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. This plot was produced by Oytun Babacan.

SDSIG in application to a grid impacts methodology, exploring a grid impact metric of OLTC operations identified in the literature review.

This chapter achieved these objectives through a thorough literature review of the current research identifying the potential fields and opportunities to develop the SDSIG from the SIG. The appropriate developments were made and then output time series validated. The SDSIG was then applied to a grid impacts analysis to assess the advantages of using decorrelating time series over correlating time series by comparing the affect they have on the number of OLTC operations in a power flow simulation.

A temporal validation was carried out using four metrics comparing the correlation between CDF observation data and the synthetic model output data at four sites around the world. The K-S test using increasing confidence limits from 90% to 99% was 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 magnitude frequency and the clear-sky index. Each metric passed the K-S test with 99% confidence limit with minimum success of 90.96% days and average score of 95.99% days. The CDF comparisons for all metrics had a minimum correlation coefficient of $R = 0.9083$ and mean of $R = 0.9872$.

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 same test using SDSIG outputs. The mean absolute percentage error of spatial correlation against site separation was 0.8648%, $RMSE = 0.01$, and the correlation coefficient was $R = 0.9552$.

The SDSIG illustrates 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 uniformly distributed cloud motion direction. With the analysis time scale at 1-year, the minimum 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. However, this is suspected to be a result of the tendency for increased cloud size with higher u_{ref} and C_t in the cloud field production and not due to the actual cloud speed. Furthermore, there is increasing correlation convergence with increasing u_{ref} . This is suspected to be due to clouds passing too quickly to be captured in detail in a 1-min resolution study, resulting in correlation becoming a function of C and not u_{ref} .

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 correlating GHI time series when compared to assigning individually decorrelating GHI time series to each house. Using correlating GHI time series for PV penetration scenarios of 25%, 50% and 75%, an increase in the OLTC operation depth had a median that is higher by 10, 49, and 93 per day with increasing PV penetration, respectively, than scenarios using spatially decorrelating GHI time series. The SDSIG time series would allow the grid operator to determine more realistic PV mitigation estimates and could avoid over-investment in voltage regulation equipment.

Chapter 5

Conclusions and future work

The aim of this research was to develop a novel modelling methodology that produces synthetic irradiance time series that varied on both a temporal and spatial dimension, and that were suitable for application in multivariate grid impact analysis whilst being derived from readily available hourly observation data only.

This research aim was sub-categorised into three research problems. Each problem was examined and addressed in its own chapter. They are concluded in turn within this section. The research problem is introduced, brief summary of key findings given, and a discussion of the implications of the findings made.

5.1 Research summary

1) What are the key distribution network impacts associated with increasing penetration of intermittent solar PV technology into the grid, the identification of which will help guide the requirement criterion for synthetic irradiance time series?

To address this research problem, a thorough literature review was carried out on the identified grid impacts, as well as applicable studies upon each impact. The literature review identified the key impact as voltage fluctuations, harmonic distortions, unintentional islanding, reverse flow, increased demand peaks and hot-spots. The impact of voltage fluctuations was identified as, and considered to be, the most concerning impact as it is most commonly mitigated by degradable devices, overuse of which presents a significant potential expense to the DNO. A review of studies regarding voltage fluctuations found a clear lack of a consensus performance metric in order to establish baselines and comparisons, and so it is suggested and implemented that the overuse of on-load tap changers (OLTCs) be adopted for this purpose. From the literature

attempting to quantify and address grid impacts from solar PV, the key observation was the frequent claim of a lack of readily available irradiance data for use. Too many studies used insufficient resolution irradiance data as an input into a grid impact study, claiming lack of data availability as justification. Using 1-hour resolution irradiance data can underestimate the magnitude and impact of certain grid impacts, particularly voltage fluctuations. This observation was the principal driver for the development of the SIG and SDSIG as there was a clear research gap for the synthetic generation of high resolution input solar irradiance time series that are statistically accurate, reproducible and are globally applicable for use in grid impact analysis.

2) Can synthetic irradiance time series be generated with significant statistical accuracy using readily available, well geographically dispersed, mean hourly meteorological observations as an input, facilitating access to more appropriate data for temporal grid and solar systems study?

This research problem required a proof of concept study to assess whether or not it is possible to make the synthetic irradiance. A review of literature found few options of producing statistically accurate time series from readily available data, most required an irradiance time series as an input from which to interpolate a higher resolution time series. Due to lacking alternatives, a new methodology and model was built and titled the Solar Irradiance Generator (SIG). The SIG presented in this this was the first model to simulate statistically sized clouds alongside clear-sky index distributions by okta in order to stochastically derive weather conditions and ultimately synthesise irradiance. The SIG was the first method to employ multiple MTMs in order to capture the identified seasonal, diurnal and pressure differences in the cloud cover amount. Novelty was demonstrated in the sizing of clouds and the utilisation of cloud samples. Further novelty was demonstrated in the method of producing the clear-sky index by using distributions by okta value and allowing cloud edge enhancement and other fluctuation features to be captured.

The validation of the SIG at the location of Cambourne, UK, demonstrated that it is indeed possible to produce high resolution, synthetic irradiance time series from readily available, mean hourly weather data. The implication of this is that every piece of research reviewed that cited lack of data availability as a limit could be repeated at a temporal resolution down to 1 minute. Applications of the SIG are limited to cases where a spatial element is not integral, such as small scale studies where a single high-resolution irradiance data series are desired. The literature review found numerous studies using inadequate time resolution data where the SIG methodology would have been ideal in supplying data. Example applications of the SIG include PV generation models, PV supply/demand models, PV and battery storage integration

models, inverter models and PV system power control strategies. The validation of the SIG demonstrates that it captures the inherent 1-min qualities of solar irradiance that were identified as crucial in the grid impacts literature review, however, for suitability of grid impact studies, a spatial element is required.

3) Is it possible to produce statistically accurate, synthetic irradiance time series that vary on both a temporal and spatial dimension, facilitating multi-variate grid impact analysis?

The final research problem required further proof of concept as to whether or not it possible to take readily available hourly inputs and produce spatially decorrelating irradiance with a 1 minute resolution. In order to answer this, the Spatially Decorrelating Solar Irradiance Generator (SDSIG) was developed. A study of the literature found two independent journal articles that agree with the author that there is currently no known method of producing synthetic, spatially decorrelating irradiance time series from readily available data. All identified potential alternatives required an irradiance time series with the same temporal resolution as the desired output in order to represent spatial decorrelation. For this reason, the SDSIG is considered the first of its kind and so contains inherent novelty within the methodology. The cloud field production method for representing the sky with increasing amounts of cloud, as well as their deployment within the cloud motion vectorisation, facilitated a brand new method of solar irradiance estimation. Furthermore, the deployment of 90 different distributions of clear-sky index categorised by total cloud amount and solar elevation is certainly a first that facilitated greater geographic flexibility of the SDSIG through incorporation of solar elevation. The additional distributions allowed for the improvement in synthetic reproduction of irradiance magnitudes that are statistically accurate to observational data. The robustness of the SDSIG method can facilitate any time resolution or spatial dimension, although to operate at higher temporal resolutions with confidence requires more quality irradiance time series datasets to validate against.

Temporal validations were carried out at four sites that represent three different climates (temperate, desert-like, and tropical pacific island) in two countries (UK and USA). The success of these validations demonstrates the statistical accuracy of the SDSIG generated solar irradiance time series. However, it was the success of the spatial validation that answer the research problem that it *is* possible to produce synthetic irradiance time series that have a spatial dimension. This is not known to have been achieved in previous literature. The implication is that, with readily available inputs, such as those from the Met Office or NOAA that cover the whole UK and North America, synthetic irradiance time series that vary on a spatial dimension can be generated. Therefore, there is now no need to have to resort to insufficient temporal resolution

irradiance data when analysing grid impacts, unless the input observations are not available. The SDSIG has suggested use in any type of PV-DG grid modelling, solar farm modelling, or situation where the irradiance at two sites separated more than roughly 60 m (at 1 minute resolution) is required. The SDSIG could provide unique, spatially decorrelating irradiance profiles to each point node on a grid case study, each can have distinct orientation and roof tilts. This has never been an option before using synthetic modelling.

The final objective to the research problem was to demonstrate the SDSIG's applicability into multi-variate grid impact analysis. A power flow study demonstrated that the magnitude and frequency of severe tap changes were significantly higher in the simulations for all PV penetrations when using a correlated irradiance time series compared to using decorrelating irradiance time series from the SDSIG. This demonstrates that excluding decorrelation from grid impact studies will likely result in over estimation of the impacts, this could lead to implementing over-conservative mitigation policy or technology to the grid. The SDSIG not only solves the problem of lacking, high resolution irradiance time series data availability, it can facilitate more robust and realistic analysis through allowing decorrelating and individualised house pitch and azimuth variables.

5.2 Contributions to knowledge

The principal contribution to knowledge is the development of a methodology that synthesises irradiance at the highest-temporal resolution downscaling for synthetic modelling available at time-of-writing. The model is certainly a first to the field to provide individualised property options (location, roof tilt, orientation) for any number of locations in a spatial domain. The availability of this well-validated tool to produce spatially decorrelated profiles fills a significant gap where there is only one real set of observation measurements at the site of Oahu, Hawaii. Other measurement sites with spatial separation exist, however are private installations such as those at UCSD. This model can, therefore, be used by researchers to produce irradiance time series that are statistically accurate and spatially relevant without having to limit their geographical region of study. This is a particularly useful contribution of knowledge, and to date (June 2017), there does not exist an alternative methodology to derive such spatial irradiances, and yet research into this field has increased. The author's papers at time of submission have received a combined 24 citations, some of which are attempting to improve upon the validation and spatial correlation component developed in this thesis (widely recognised in the field J. Munkhammar and J. Widén of Uppsala University).

Many of these citations are in relation to grid impacts on the low voltage grid. It is therefore, the authors opinion that the methodology presented here is a significant contribution to knowledge in the field. There are key features of the model that make this true and are subsequently detailed.

The Markovian approach to solar modelling was expanded upon and demonstrably improved the diurnal, seasonal and effective weather system statistics of cloud amount. The most related study in the field uses only a single Markov transition matrix to define the change of total cloud amount. This research expanded upon the single matrix to using 20 separate matrices and showed that ignoring seasonality, diurnality and pressure dependencies will lead to significant biases. Seasonality strongly impacts the transition states by up to 10% difference in probability when compared to the single Matrix. The implementation of 20 matrices is not computationally demanding, in fact, the separation and construction of all Markov transition matrices is the fastest part of the model. This demonstrates that all future studies can easily apply this technique to remove annual, seasonal and diurnal biases from stochastic solar irradiance modelling, presenting a significant improvement to the field. The bias of time-of-day for the site studied was demonstrated to have the largest bias. The time of day is intuitively the most variable as it is the smallest change allowable in the variables. Furthermore, it is intrinsically linked to the solar elevation angle, which was later demonstrated to hold significant relationships to clear-sky index and therefore solar irradiance measured at the ground; thus, corroborating the finding.

The use of geometrically fixed cloud fields is a valid tool to produce accurate binary spatial correlation and highly accurate ramp-rate reproductions across four different climatic regions, marrying the computational burden with empirically derived cloud size distributions. The idea had been proposed once before, however the utilisation with sizing distributions is entirely novel. The spatial validation application to the Hawaii dataset showed strong performance between synthesised cloud cover and reality. Such a simple computational approach opens future avenues of modelling cloud fields of many sizes.

The use of okta to produce distinct clear-sky index distributions is an incredibly effective approach in order to capture diurnal, seasonal trends of realistic irradiance magnitudes. The separation of okta by the solar elevation angle bands allows for more geographic flexibility and reduction in diurnal bias than using just the diurnal Markov transition matrix alone. Solar elevation angle is easier categorised at different latitudes than the hour of day itself as an integer, this allows for greater accuracy in synthesising irradiance magnitudes at the variety of latitudes studied for all times of day. These distributions are being further extended by the research group at Uppsala University

to determine if they are globally applicable. This is a fundamental improvement to understanding how different weather types influence the solar attenuation, particularly useful to the field is the representation of solar irradiance at low sun angles, where the sun is near the horizon, which is often ignored or excluded in solar resource assessment (sometimes ignoring up to 50% of data when solar zenith less than 60° are removed).

A strict relationship was discovered between the cloud fraction and spatial correlation, whereby the spatial correlation is least at 50% cloud cover and increases as one of the states (clear or cloudy) becomes more dominant. This is a useful finding as it is the first study to demonstrate that the spatial correlation of solar irradiance is directly related to the state of sky. All studies prior to this have not isolated a variable to discuss the influence of spatial separation and have instead considered it as aggregated. The interesting use of this relationship is that it local weather statistics, which are readily available from satellite imagery or meteorological agencies, can be used to assess the typical spatial decorrelation at a large scale PV site or distributed PV generation in a grid system depending on prevailing weather and to identify potential days that pose synchronous ramping risks— remembering that lower correlation is beneficial to avoid synchronous ramping. A site such as Hawaii, where the clouds are known to be constant and consistently 50% coverage, it can be assumed that there will be maximum decorrelation between sites. Contrastingly to the finding of cloud fraction, the cloud speed was found to be non-influential on the instantaneous correlation over small distances. Other studies have demonstrated that the cloud speed is fundamental when linking ramp coincidences, however they only consider the time delay between sites, which does not inform power engineers of likely scenarios for any time-instance of a grid. Knowing that the cloud speed is insignificant on the instantaneous spatial decorrelation is useful as it facilitates the option of ignoring the accuracy of this factor, or assume a simple distribution. As this was a large computational expense in producing the cloud fields, this is a useful finding.

This research was the first to demonstrate the impact of using correlating instead of decorrelating irradiance on a modelled electricity grid. The application to an IEEE test feeder showed that the increase in on-load tap changer operations was significantly impacted when using synchronous ramping (same irradiance profile for all properties in a spatial domain) when compared to using decorrelating irradiance profiles (unique irradiance profiles individualised to each property). The importance of this is that any study using correlating irradiance — of which there are many, if not nearly all — have over estimated the impact of “synchronous” ramping events, as the events are not truly synchronous. The overestimation of impact from solar PV is damaging to the opinion, desirability and attractiveness of the technology. This is likely the exact opposite intention held by the researchers. The presented research allows a fairer test

solution for the impact of solar PV at aggregated 1-min time resolution, which was identified to be a good starting time resolution in from a study of the grid impacts and power systems response. Ideally this time resolution could be reduced, however, 1-min downscaling from 1-hour presents the new benchmark in the field.

In summary, being able to operate now with statistically accurate spatial decorrelation and at a high temporal resolution is a significant advantage to the field of solar modelling and grid integration studies. The developed model in this thesis provides this capability. Naturally, there are some limitations to the work and some areas for improvement. These are discussed in the following section.

5.3 Limitations and ideas for future work

There are general considerations and specific ideas for advancement of both the field of synthetic solar resource assessment and of the SDSIG itself. This section will be split between those general comments regarding the solar resource assessment field, and those that are specified to certain aspects of the SDSIG.

5.3.1 General future enablers to advance the impact of theoretical solar resource modelling

An impression from the grid impacts literature review was that more industrial collaborations are required in order for progress in maximising PV penetration into the LV grid. Without real grid schematics or empirical testing on networks to appropriately validate grid impact analysis and power flow tools, it is difficult to state with empirical certainty how conservative an approach is, or isn't, when considering voltage fluctuations. This link to industry could be benefited through the establishment of grid impact baseline consensus metrics with which to benchmark studies and methodologies, such as the over use of on-load tap changers; the empirical analysis of these grid impact metrics would help connect theoretical conclusions to the real world, such as those of the power flow application of the SDSIG.

Access to well-distributed, high resolution solar irradiance data is imperative to advance the field of spatially decorrelating irradiance models. To date, the only real place to access trusted and well-maintained data with a spatial separation is from Oahu, HI USA. Whilst this is an excellent resource, it limits all studies to an area with a prevailing cloud direction, a predominant cloud type and a tropical island climate. There is a need to establish a network of irradiance monitoring sites with good spatial separation

that are structured with appropriate station-pairing in order to determine the cloud direction. This would advance the knowledge of different climatic zones and provide a better platform for those studying solar resource modelling with opportunities for validation and statistical analysis of the spatial correlation of solar irradiance.

5.3.2 Specific opportunities to develop the Spatially Decorrelating Solar Irradiance Generator

In its current format presented in this thesis, the SDSIG operates within a 1.5km^2 spatial domain as was found to be computationally tolerable for producing > 2400 year-long irradiance time series for the grid impact study, however, the methodology is robust enough to explore much larger areas and could deliver greater understanding of spatial correlation at areas around the world. Furthermore, the spatial granularity is found to be 60m as is the smallest distance travelled by a cloud travelling 1 ms^{-1} for 60 s. This does not mean that there will be no differences between two sites located closer together than the granularity, as the clouds can be centred anywhere within the cloud field, it means that the minimum a cloud can move with each 1 min time step is 60 m. Transition to a 1-sec resolution is theoretically plausible with the SDSIG and would reduce the granularity to 1m.

Analysis of the minimum cut-off duration for both 0 and 8 okta smoothing periods would benefit this methodology. The use of > 3 hours for 0 okta within the SDSIG is a conservative estimation and is intuitively as low as 30 mins. This may reduce the variability index (VI) in the SDSIG, improving the temporal validation metric performance. The VI is the metric with the weakest performance within both the SIG and SDSIG. The implication of this is that daily solar variability is not being statistically reproduced as well as it could be. The author believes this is a result of using only a single order Markov chain dependency, considering the weather variable transitions from only one hour previous. The best demonstration of this is that the probability for complete clear sky to remain stable between hours in Leeds is 0.71. For a full day of complete clear sky, which is reasonably well experienced in winter and summer anti-cyclone weather systems each year, the chance is $0.71^{24} \equiv \frac{1}{3703}$ days, or approximately 1 day in 10 years. With hindsight, it is a reasonably obvious limitation, however, literature used similar assumptions and validations were not too negatively impacted using a single order. After the above future work of expanding the spatial and temporal resolutions, the progression to a higher order Markov chain, and a study to analyse the most appropriate order, is a recommended piece of future research.

The cloud fields are a highly novel aspect of this thesis, as such there is plenty of scope for development. The SDSIG at present does not allow for the clouds passing across the domain to develop in the dynamic way with progression of time like clouds do in reality. Cloud edges in the cloud field are clearly defined such that the direct irradiance is subject to a binary on-or-off. Ideally the model could take a more dynamic approach and allow the clouds to develop over time with knowledge of weather systems beyond what is simply employed at current using high and low pressure environments. The Markov chains perform well enough to allow variation in the systems that occur, however, to capture a cloud system of hundreds of kilometres long, the stochastic weather would have to produce multiple events of $N = 8$. It would be interesting to explore the concept that, instead of selecting the cloudiness for the hour ahead, but to select the weather system for the next 3 days to 3 weeks to have dynamic cloud transitions starting in the cirrus region before gradually moving into a fully developed cumulonimbus cloud. One more development to the cloud field implementation would be to achieve more complex three-dimensional shapes through a change in clear-sky irradiance across the domain of the circular cloud. The clouds could be assumed to be spherical through applying different attenuations across the diameter of the cloud. There is little scientific evidence that justifies this; however, it would be a very interesting piece of research.

Whilst the model in its current state satisfies validations for the geographic regions attempted (San Diego, Oahu, Cambourne and Lerwick), this is not enough to suggest the SDSIG is globally applicable, despite hypothesising that this may be the case. More locations must be validated against before global accuracy can be claimed. One of the limits that will likely hold back the global applicability of the SDSIG is the use of the relationships of clear-sky index to okta and solar elevation. The pattern caused by solar elevation is expected to be globally applicable as the nature of this correction is to do with the geometric path that light takes through the clouds, not necessarily to do with the cloud type. The okta number, however, is expected to be a function of cloud type. 8 okta in the UK may not share the same clear-sky index distribution as 8 okta from each of the climatic region. For true geographic flexibility, the clear-sky index distributions would need to be examined globally and compared for similarity or distinct and definable differences. An alternative is to research a possible link between the clear-sky index and the cloud type, and then to further analyse the distribution of cloud type with total cloud amount around the world. This would be an incredibly large undertaking. An estimation of the added value gained of this line of research would be needed as recommended starting point. A better understanding of the global validity of clear-sky index distributions could facilitate the SDSIG to perform globally.

The cloud edge enhancement (CEE) application within the SDSIG could potentially be much more sophisticated whilst being computationally less demanding. CEE could be applied as a function of space, as opposed to its current implementation of time. As the temporal resolution of the SDSIG is increased, it will become more poignant to explore the spatial influence of CEE and include it within the SDSIG. Further research into the probability of occurrence and magnitudes of CEE would also help guide implementation into the SDSIG. The principal concept of spatially applied CEE has been theorised by the author such that a banded ring is made around each cloud as a function of that cloud's radius. The ramping pattern discovered by [Lave et al. \(2012\)](#) can be applied along the band's width. The cloud field is then moved across the spatial domain as usual. Should the house be positioned inside one of the bands and not also inside a cloud, a CEE correctional factor from Lave married with a magnitude M from the analysis in section 4.2.4.3 can be applied to the clear-sky index for that house. Intuitively, the CEE is a spatial variable similar to the circumsolar of the sun, and its temporal nature is only due to the speed of the cloud which would lend itself well to the CEE correctional factor by [Lave et al. \(2012\)](#). The added value should be explored first as, with intended application to PV power models, the typical oversize of inverters may nullify the benefits of, and need for, CEE accuracy.

Shading due to terrain topology such as blocking mountainous landscapes, would be an interesting inclusion to the SDSIG. Currently, only the topographical height is utilised for the PSA 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 to 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 thesis of using simple and readily available inputs, methods to determine the cloud direction such as [Wang et al. \(2016\)](#); [Chow et al. \(2011\)](#) require sophisticated equipment and so are not suitable for this study. Methods or statistics that could be simply employed to estimate realistic cloud direction and general cloud motion would be an interesting inclusion. Recent work at the University of Leeds examined this and has preliminarily shown positive correlation between ground based wind speed measurements and the estimated cloud speed using triangulated pyranometer GHI data. At current, the author feels that the inclusion of more accurate cloud direction

is the most limiting assumption to the SDSIG and is the most pressing piece of future work.

Appendix A

Solar Irradiance Generator scripts

The scripts presented in this appendix are all produced using Matlab 2015b software. The colour code presented is as follows:

- *Black* — executable script
- *Green* — non-executable comments
- *Blue* — Matlab function or operator
- *Red* — deletion of a variable

A.1 Cloud sample production

```
1 %% Cloud Sample Production
2
3 %% Variables
4 % User defined
5 cloud_vector_length=10^8; %(decameters) lenght of large cloud sample
6 num_of_options=1000;%number of sample hours per wind speed per coverage
7 u_range=30;%all possible wind speeds (m/s).
8 coverage_range=9;% C=(1:9)/10. 0/10 and 10/10 are both all zeros or ones.
9
10 % Pre allocate arrays
11 cloud_sample=zeros(1,cloud_vector_length); %pre-allocate memory. construct array
    of variable length
12 combined_record=zeros(cloud_vector_length,1);
13
14 % Preliminary definitions
15 marker=1; %indexing initiation within the sample
16 % Power Law Exponent ///  $p(x)=Cx^{-B}$  /// incorporated xmin and xmax
```

```

17 % (Wood & Field, 2011, Journal of Climate, Volume 24, p4800).
18 % powerlaw becomes  $x=(\alpha+\beta*r)^{1/(1-B)}$ .
19 % where  $r=0$  gives  $x_{max}$ .  $rand=1$  gives  $x_{min}$ .
20 B=1.66; % as recommended
21 x_min=10;%minimum cloud length (decameters).
22 x_max=150000;%maximum cloud length (decameters).
23 alpha=x_max^(1-B);
24 beta=x_min^(1-B) - alpha;
25
26 %% Produce Large Cloud Vector
27 while marker<length(cloud_sample);
28     %ADD CLOUD
29     % produce random value for use in power law
30     cloud_rand=rand; %N[0,1]
31     % apply single power law to produce a cloud length
32     cloud_length=floor((alpha+beta*cloud_rand)^(1/(1-B)));
33     % keep a log of the length for later plots (to prove power law)
34     combined_record(marker,1)=cloud_length;
35     % if the cloud length goes beyond vector length, limit it.
36     if cloud_length+marker>length(cloud_sample);
37         cloud_length=length(cloud_sample)-marker;
38     end
39     % put the cloud in the array at the correct point
40     cloud_sample(1,marker:marker+cloud_length)=1;
41     % update the current point along the sample
42     marker=marker+cloud_length;
43
44     %ADD CLEAR
45     clear_rand=rand; %REPEAT for a period of clear sky.
46     clear_length=floor((alpha+beta*clear_rand)^(1/(1-B)));
47     combined_record(marker,1)=clear_length;
48     if clear_length+marker>length(cloud_sample);
49         clear_length=length(cloud_sample)-marker;
50     end
51     cloud_sample(1,marker:marker+clear_length)=0;
52     marker=marker+clear_length;
53 end
54
55 % create frequency and probability stats using tabulate function
56 combined_record=tabulate(combined_record);
57 % find the number of clouds produced over the sample range.
58 num_of_clouds=sum(combined_record(2:length(combined_record),2))/2;
59
60 %% Reshape cloud population, Randomly Sample cloud population, Allocate cloud
    sample
61 %make coverage arrays
62 for i=1:coverage_range
63     bin=zeros(num_of_options*u_range,60);
64     assignin('base', ['coverage_bin_', num2str(i)], bin);
65 end; clear bin
66
67 %loop whole range of u and create samples.
68 for u=1:u_range %cycle through all the different windspeeds
69     %resample the cloud cover
70     epm=u/(1/6);% resample rate to convert the cloud_sample

```

```

71     cloud_resampled=resample(cloud_sample,1,epm); % resample(data,P,Q) resamples
        data such that the data is interpolated by a factor P(1) and then decimated
        by a factor Q(epm)
72     %remove noise from resampled data to maintain binary
73     cloud_resampled(cloud_resampled<0.5)=0;
74     cloud_resampled(cloud_resampled>=0.5)=1;%
75
76     %pre allocate space
77     entry_count=zeros(1,9); %log of cloud sample = a coverage value
78     tally=zeros(1,9); %to check if correct amount of options produced
79
80     %start selecting cloud samples from the resampled vector
81     while sum(tally)~=(num_of_options*coverage_range);
82         %random location from all but final hour of cloud cover
83         r=rand*(1-(61/length(cloud_resampled)));
84         %extract a cloud sample
85         sample_hour=cloud_resampled(ceil(r*length(cloud_resampled)):59+ceil(r*
        length(cloud_resampled)));
86         %determine coverage
87         coverage=round(sum(sample_hour)/6);
88         %assign cloud sample to appropriate bin
89         switch coverage
90             case {0,10} %skip if 0 or 10
91             case 1
92                 %keep tally of entries into each case
93                 entry_count(1,1)=entry_count(1,1)+1;
94                 %fill bin if there is space
95                 if entry_count(1,1)<=num_of_options;
96                     %update bin space
97                     tally(1,1)=tally(1,1)+1;
98                     %place the hour sample in correct space
99                     coverage_bin_1(u*num_of_options-(num_of_options-tally(1,1))
        ,:)=sample_hour;
100                end
101                %repeat for all coverage values
102                case 2
103                    entry_count(1,2)=entry_count(1,2)+1;
104                    if entry_count(1,2)<=num_of_options;tally(1,2)=tally(1,2)+1;
        coverage_bin_2(u*num_of_options-(num_of_options-tally(1,2)),:)=sample_hour;
        end;
105                case 3
106                    entry_count(1,3)=entry_count(1,3)+1;
107                    if entry_count(1,3)<=num_of_options;tally(1,3)=tally(1,3)+1;
        coverage_bin_3(u*num_of_options-(num_of_options-tally(1,3)),:)=sample_hour;
        end;
108                case 4
109                    entry_count(1,4)=entry_count(1,4)+1;
110                    if entry_count(1,4)<=num_of_options;tally(1,4)=tally(1,4)+1;
        coverage_bin_4(u*num_of_options-(num_of_options-tally(1,4)),:)=sample_hour;
        end;
111                case 5
112                    entry_count(1,5)=entry_count(1,5)+1;
113                    if entry_count(1,5)<=num_of_options;tally(1,5)=tally(1,5)+1;
        coverage_bin_5(u*num_of_options-(num_of_options-tally(1,5)),:)=sample_hour;
        end;

```

```

114         case 6
115             entry_count(1,6)=entry_count(1,6)+1;
116             if entry_count(1,6)<=num_of_options;tally(1,6)=tally(1,6)+1;
coverage_bin_6(u*num_of_options-(num_of_options-tally(1,6)),:)=sample_hour;
end;
117         case 7
118             entry_count(1,7)=entry_count(1,7)+1;
119             if entry_count(1,7)<=num_of_options;tally(1,7)=tally(1,7)+1;
coverage_bin_7(u*num_of_options-(num_of_options-tally(1,7)),:)=sample_hour;
end;
120         case 8
121             entry_count(1,8)=entry_count(1,8)+1;
122             if entry_count(1,8)<=num_of_options;tally(1,8)=tally(1,8)+1;
coverage_bin_8(u*num_of_options-(num_of_options-tally(1,8)),:)=sample_hour;
end;
123         case 9
124             entry_count(1,9)=entry_count(1,9)+1;
125             if entry_count(1,9)<=num_of_options;tally(1,9)=tally(1,9)+1;
coverage_bin_9(u*num_of_options-(num_of_options-tally(1,9)),:)=sample_hour;
end;
126     end
127
128
129 end
130 end
131
132 %Combine all bins into single array
133 sun_obscurd_options=[coverage_bin_1;coverage_bin_2;coverage_bin_3;
coverage_bin_4;coverage_bin_5;coverage_bin_6;coverage_bin_7;coverage_bin_8;
coverage_bin_9];
134 %Save the array to file
135 csvwrite('Cloud_Samples',sun_obscurd);

```

A.2 User defined variables

```

1 %% User Defined Variables
2 start_day=1; % choose the day number from which to start from
3 start_year=2001; % choose the year from which to start from
4 num_of_years=7; % set the duration from the model to run
5 leap_years=(start_year:start_year+num_of_years); %make array of years
6 leap_years=floor(floor(leap_years./4)./(leap_years./4)); %determine if leap year
7 num_of_days=365*numel(leap_years==0)+366*numel(leap_years==1); %create number of
days in accordance to leap years
8 latitude=50.2178; % set the latitude. Currently Cambourne, UK
9 longitude=-5.32656; % set the longitude. Currently Cambourne, UK
10 height_above_sea_level=87; %meters above sea level. Currently Cambourne, UK
11 panel_pitch=0; %(degs.)panel angle measured from the horizontal
12 panel_azimuth=0; %(degs.) panel orientation(+180=North, 0=south,-90=west, 90=
east)
13 u_range=60;%the range of windspeeds. standard=60. Defined from Met Office.
Adaptable only with new cloud sample technique

```

```

14 num_of_options=1000;% the number of options per windspeed per coverage. standard
    =1000. Adaptable only with new cloud sample technique

```

A.3 Data import

```

1 %% Read In Data
2 % Select appropriate location raw data
3 pressure=csvread('1395pressure2001to2012.csv');
4 data1=csvread('1395data2001to2012.csv'); %Note not all these variables are
    required: 1) hour% 2) cloud total (okta)% 3) Low cloud type% 4) Medium Cloud
    Type% 5) high cloud type% 6) cloud base height 7) pressure (msl)% 8) 1 -
    cloud amount% 9) 1 - cloud type% 10) 1 - cloud height (decameters)% 11) 2
    "" 12) 2 ""% 13) 2 ""% 14) 3 ""% 15) 3 ""% 16) 3 ""% 17)
    Air temp% 18) year% 19) hour of day 20) day number% insert 6 blank rows to
    fill with markov case, season, hp/lp systems x4% 27) wind direction% 28)
    wind speed
5 data = [data1(:,1:20) zeros(length(data1),6) data1(:,20+1:end)]; %make space for
    season markers, markov chain markers etc. 6 columns.
6 sun_obscured_options = csvread('Sun_Obscured_Options_Generic.csv'); %Read in the
    sun obscured options, each line is an array full of clouded hours in
    minutes(:,60).

```

A.4 Preliminary Markov chain framework

```

1 %% Preliminary Preparation for Markov Chain production
2
3 % Define the seasons.
4 season=zeros(length(data),1); %pre allocate array
5 for i=1:length(data)%For each hour measurement. NB: Spring=Mar to May. Summer=
    Jun to Aug. Autumn=Sept to Nov. Winter=Dec to Feb
6     if isinteger(data(i,18)/4)== 1;%this is a method to check whether the year
    is a leap year. then updates the day at which the seasons start and end.
7         springstart=61;springend=152;
8         summerstart=153;summerend=244;
9         autumnstart=245;autumnend=335;
10        winterstart=and(1,336);winterend=and(60,366);
11    else %account for leap year
12        springstart=60;springend=151;
13        summerstart=152;summerend=243;
14        autumnstart=244;autumnend=334;
15        winterstart=and(1,335);winterend=and(59,365);
16    end
17
18    %query the day number, then assign the appropriate season number
19    if data(i,20)>= springstart && data(i,20)<=springend;
20        season(i,1)=1;
21    elseif data(i,20)>=summerstart && data(i,20)<=summerend;
22        season(i,1)=2;
23    elseif data(i,20)>= autumnstart && data(i,20)<=autumnend;

```

```

24     season(i,1)=3;
25     else season(i,1)=4; %else winter
26     end
27 end
28 data(:,21)=season;%column 21 = season {1,2,3,4}
29
30 % Assign a case number in order to populate the correct Markov Table.
31 pressure_avg=mean(pressure(:,2)); % Calculate the average pressure
32 markovcase=zeros(length(data),1); % pre-allocate memory
33
34 for i=1:length(data);%rows:21 = season. 7=mean sea level pressure
35     % Markov chains for each season and pressure state
36     if data(i,21)==1 && data(i,7)<pressure_avg;
37         markovcase(i,1)=1; %spring low pressure
38     elseif data(i,21)==1 && data(i,7)>pressure_avg;
39         markovcase(i,1)=2; %spring high pressure
40     elseif data(i,21)==2 && data(i,7)<pressure_avg;
41         markovcase(i,1)=3; %summer low pressure
42     elseif data(i,21)==2 && data(i,7)>pressure_avg;
43         markovcase(i,1)=4; %summer high pressure
44     elseif data(i,21)==3 && data(i,7)<pressure_avg;
45         markovcase(i,1)=5; %autumn low pressure
46     elseif data(i,21)==3 && data(i,7)>pressure_avg;
47         markovcase(i,1)=6; %autumn high pressure
48     elseif data(i,21)==4 && data(i,7)<pressure_avg;
49         markovcase(i,1)=7; %winter low pressure
50     elseif data(i,21)==4 && data(i,7)>pressure_avg;
51         markovcase(i,1)=8;
52     end
53 end
54 data(:,22)=markovcase;%populate the markov case into the data file
55
56 % Pre-allocate memory for the 8 different markov transition matrices (MTM)
57 springlp=zeros(10,10); springhp=zeros(10,10);
58 summerlp=zeros(10,10); summerhp=zeros(10,10);
59 autumnlp=zeros(10,10); autumnhp=zeros(10,10);
60 winterlp=zeros(10,10); winterhp=zeros(10,10);
61
62 % Assign okta of (0/8) a value of 10 for use in columns for indexing ease
63 for i=1:length(data); % column meaning = 2)hourly okta. 8,11,14)okta layer
64     if data(i,2)==0; data(i,2)=10; end %if the okta reading is 0...
65     if data(i,8)==0; data(i,8)=10; end % ...assign a value of 10
66     if data(i,11)==0; data(i,11)=10; end
67     if data(i,14)==0; data(i,14)=10; end
68 end

```

A.5 Markov chain creation

```

1 %% Create Markov Chains
2
3 % Populate the Markov chains with real transitions from observation data
4 for i=2:length(data); % index key: i=next, i-1=now.

```

```

5     markov=data(i,22); %column 22 = markov case
6     %if a data point follows another (and is not a missing value of -9999)
7     if data(i-1,2)~-9999 && data(i,2)~-9999;
8         switch markov; %switch to the appropriate MTM
9             %tally/populate the markov chain in appropriate place
10            case 1; %spring low pressure
11                springlp(data(i-1,2),data(i,2))= springlp(data(i-1,2),data(i,2))+1;
12            case 2; %spring high presure
13                springhp(data(i-1,2),data(i,2))=springhp(data(i-1,2),data(i,2))+1;
14            case 3; %summer low pressure
15                summerlp(data(i-1,2),data(i,2))=summerlp(data(i-1,2),data(i,2))+1;
16            case 4; %summer high pressure
17                summerhp(data(i-1,2),data(i,2))=summerhp(data(i-1,2),data(i,2))+1;
18            case 5; %autumn low pressure
19                autumnlp(data(i-1,2),data(i,2))=autumnlp(data(i-1,2),data(i,2))+1;
20            case 6; %autumn high pressure
21                autumnhp(data(i-1,2),data(i,2))=autumnhp(data(i-1,2),data(i,2))+1;
22            case 7; %winter low pressure
23                winterlp(data(i-1,2),data(i,2))=winterlp(data(i-1,2),data(i,2))+1;
24            case 8; %winter high pressure
25                winterhp(data(i-1,2),data(i,2))=winterhp(data(i-1,2),data(i,2))+1;
26        end
27    end
28 end
29
30 % Calculate pressure system duration
31 %columns = 23)high pressure indicator, 24)low pressure indicator
32 for i=1:length(data); %for the length of the pressure data..
33     if data(i,7)>pressure_avg; data(i,23)=1; %if pressure is above average
34     else data(i,24)=1; %assign a value of 1.
35     end
36 end
37
38 % calculate duration of the high pressure system, account for missing data.
39 for i=2:length(data)-1;
40     if data(i:i-5,7)==-9999;%if there are 5 missing data plots,
41     %end the duration count
42     elseif data(i,7)==-9999; %assume pressure is maintained
43         data(i,25)=data(i-1,25); data(i+1,25)=data(i-1,25);
44
45     %if pressure is maintained make cumulative tally of duration
46     elseif data(i,23)==1 && data(i+1,23)==1;
47         data(i+1,25)=data(i,25)+data(i+1,23);
48     end
49 end
50
51 %remove all but the final duration count, leaving only actual duration
52 for i=1:length(data)-1;
53     if data(i+1,25)~=0;
54         data(i,25)=0; end
55 end
56
57 %repeat whole process for above for low pressure
58 for i=2:length(data)-1;
59     if data(i:i-5,7)==-9999; %using column 26 to do this.

```

```

60     elseif data(i,7)==-9999; data(i,26)=data(i-1,26); data(i+1,26)=data(i-1,26);
61     elseif data(i,24)==1 && data(i+1,24)==1; data(i+1,26)=data(i,26)+data(i
    +1,24);
62     end
63 end
64 for i=1:length(data)-1;
65     if data(i+1,26)~=0; data(i,26)=0; end
66 end
67
68 %combine results and index by season
69 season_indicators=data(:,21); %season indicators
70 highpressure=data(:,25); %lp durations
71 lowpressure=data(:,26); %hp durations
72
73 %tabulate the findings into the pressure system and season
74 %"tabulate" produces a frequency dist. table for each discreet value
75 springhpsys=tabulate(highpressure(season_indicators==1));
76 summerhpsys=tabulate(highpressure(season_indicators==2));
77 autumnhpsys=tabulate(highpressure(season_indicators==3));
78 winterhpsys=tabulate(highpressure(season_indicators==4));
79 springlpsys=tabulate(lowpressure(season_indicators==1));
80 summerlpsys=tabulate(lowpressure(season_indicators==2));
81 autumnlpsys=tabulate(lowpressure(season_indicators==3));
82 winterlpsys=tabulate(lowpressure(season_indicators==4));
83
84 %make a diurnal markov chain for the morning (between 1-5am)
85 ammarkov=zeros(length(data),7); %pre allocate
86 ammarkov(:,1)=data(:,19); %hour of day
87 ammarkov(:,2)=data(:,21); %season
88 ammarkov(:,3)=data(:,2); %okta
89 %separate out the okta number by season
90 for i=1:length(ammarkov);
91     if ammarkov(i,1)<6; %for hours 0--5.
92         if ammarkov(i,2)==1; %if the season is spring...
93             ammarkov(i,4)=ammarkov(i,3);%...populate appropriate column.
94             elseif ammarkov(i,2)==2; ammarkov(i,5)=ammarkov(i,3);%summer
95             elseif ammarkov(i,2)==3; ammarkov(i,6)=ammarkov(i,3);%autumn
96             else ammarkov(i,7)=ammarkov(i,3);%winter
97         end
98     end
99 end
100
101 %pre-allocate memory for the morning markov chains
102 morningspring=zeros(10,10);
103 morningsummer=zeros(10,10);
104 morningautumn=zeros(10,10);
105 morningwinter=zeros(10,10);
106
107 %usage is (now,next) i=next, i-1=now.
108 for i=1:length(ammarkov); %change -9999 values to 0
109     for j=4:7;
110         if ammarkov(i,j)==-9999; ammarkov(i,j)=0; end
111     end
112 end
113

```

```

114 %Populate Markov chains with transition frequency
115 %...okta transition from the current state(i) to future state (i+1)
116 for i=2:length(ammarkov);%spring
117     if ammarkov(i-1,4)~=0 && ammarkov(i,4)~=0; %if correct season...
118         %... populate the Markov chain
119 morningspring(ammarkov(i-1,4),ammarkov(i,4))=morningspring(ammarkov(i-1,4),
        ammarkov(i,4))+1;     end
120 end
121 for i=2:length(ammarkov);%repeat for summer...
122     if ammarkov(i-1,5)~=0 && ammarkov(i,5)~=0;
123 morningsummer(ammarkov(i-1,5),ammarkov(i,5))=morningsummer(ammarkov(i-1,5),
        ammarkov(i,5))+1;     end
124 end
125 for i=2:length(ammarkov);%...autumn
126     if ammarkov(i-1,6)~=0 && ammarkov(i,6)~=0;
127 morningautumn(ammarkov(i-1,6),ammarkov(i,6))=morningautumn(ammarkov(i-1,6),
        ammarkov(i,6))+1;     end
128 end
129 for i=2:length(ammarkov);%...winter
130     if ammarkov(i-1,7)~=0 && ammarkov(i,7)~=0;
131 morningwinter(ammarkov(i-1,7),ammarkov(i,7))=morningwinter(ammarkov(i-1,7),
        ammarkov(i,7))+1;     end
132 end
133
134 %Create Markov transition matrices for wind speed
135 windspeedupdate=data(:,28); %separate out the windspeeds
136 windspeedupdate(windspeedupdate==0)=1; %windspeed of 0 knotts is now = 1 knott
        for indexing purposes
137 data(:,28)=windspeedupdate; %replace the column with updated values
138 wind_max=max(data(:,28)); % determine the maximum measured windspeed
139
140 % pre allocate memory for wind markovs using the maximum wind speed for each
        season
141 wind_spring=zeros(wind_max,wind_max);
142 wind_summer=zeros(wind_max,wind_max);
143 wind_autumn=zeros(wind_max,wind_max);
144 wind_winter=zeros(wind_max,wind_max);
145
146 %populate the Markov cains with tranasition frequency
147 for i=2:length(data); % i=next, i-1=now.
148     seasonmarker=data(i,21); %season
149     if data(i-1,28)~-9999 && data(i,28)~-9999; %if a data point follows
        another (and is not a missing value of -9999) (a genuine transition)
150         switch seasonmarker;
151             case 1; %spring
152 wind_spring(data(i-1,28),data(i,28))= wind_spring(data(i-1,28),data(i,28))+1; %
        populate the markov chain tally in appropriate place
153             case 2; %summer
154 wind_summer(data(i-1,28),data(i,28))= wind_summer(data(i-1,28),data(i,28))+1;
155             case 3; %autumn
156 wind_autumn(data(i-1,28),data(i,28))= wind_autumn(data(i-1,28),data(i,28))+1;
157             case 4; %winter
158 wind_winter(data(i-1,28),data(i,28))= wind_winter(data(i-1,28),data(i,28))+1;
159         end
160     end

```

```

161 end
162
163 % Make the cloud height Markov chains
164 %take the cloud heights of the 3 levels measured
165 cloudheightupdate1=[data(:,10),data(:,13),data(:,16)];
166 %make -9999 values a NaN to use nanmean function
167 cloudheightupdate1(cloudheightupdate1==-9999)=0/0;
168 %mean cloud height to nearest 5
169 cloudheightupdate=round(round(nanmean(cloudheightupdate1,2))./5).*5;
170 cloudheightupdate(isnan(cloudheightupdate)==1)=-9999;%revert away from NaN
171 cloudheightupdate(cloudheightupdate==0)=1; % for indexing. returned to 0 later.
172 %update the cloud height values
173 data(:,6)=cloudheightupdate;
174 cl_h_max=ceil(max(data(:,6))); %max cloud height
175
176 %pre allocate Markov chains for processing speed
177 cloudheight_spring=zeros(cl_h_max,cl_h_max);
178 cloudheight_summer=zeros(cl_h_max,cl_h_max);
179 cloudheight_autumn=zeros(cl_h_max,cl_h_max);
180 cloudheight_winter=zeros(cl_h_max,cl_h_max);
181
182 %Populate the Markov chains with the transition frequency
183 for i=2:length(data); % i=next, i-1=now.
184     seasonmarker=data(i,21); %season
185     if data(i-1,6)~-9999 && data(i,6)~-9999; %if a genuine transition
186         switch seasonmarker; %season indicator (1--4)
187             case 1; %spring
188                 cloudheight_spring(data(i-1,6),data(i,6))= cloudheight_spring(data(i-1,6),data(i
189                     ,6))+1; %populate the markov chain in appropriate place
190             case 2; %summer
191                 cloudheight_summer(data(i-1,6),data(i,6))= cloudheight_summer(data(i-1,6),data(i
192                     ,6))+1;
193             case 3; %autumn
194                 cloudheight_autumn(data(i-1,6),data(i,6))= cloudheight_autumn(data(i-1,6),data(i
195                     ,6))+1;
196             case 4; %winter
197                 cloudheight_winter(data(i-1,6),data(i,6))= cloudheight_winter(data(i-1,6),data(i
198                     ,6))+1;
199         end
200     end
201 end
202
203 cloudheight_spring(1,1)=0; %replace ch=0 with value of 0, not 1
204 cloudheight_summer(1,1)=0;
205 cloudheight_autumn(1,1)=0;
206 cloudheight_winter(1,1)=0;
207
208 %Analyse the Pressure
209 Pmsl=pressure(:,2); %extract the pressure values, note that they are recorded
210     and normalised to mean sea level pressure
211 Pmsl=round(Pmsl); %give each value a unique integer for referencing
212 Pmax=ceil(max(Pmsl)); %find the max pressure
213 PAAMarkov=zeros(Pmax,Pmax); %preallocate Markov chains for above..
214 PBAMarkov=zeros(Pmax,Pmax); % ...and below mean pressure
215
216 %populate the appropriate pressure in the Markov chains

```

```

211 for i=1:length(Pmsl)-1; %for the length of the pressure readings
212   if Pmsl(i)<=ceil(pressure_avg) && Pmsl(i+1)<=ceil(pressure_avg) %if two
       consecutive measurements are below average then keep a transition tally
213   PBAmarkov(ceil(Pmsl(i)),ceil(Pmsl(i+1)))=PBAmarkov(ceil(Pmsl(i)),ceil(Pmsl(i+1))
       )+1;
214   end
215
216   if Pmsl(i)>ceil(pressure_avg) && Pmsl(i+1)>ceil(pressure_avg)
217   PAAmarkov(ceil(Pmsl(i)),ceil(Pmsl(i+1)))=PAAmarkov(ceil(Pmsl(i)),ceil(Pmsl(i+1))
       )+1;
218   end
219 end
220
221 % Produce the CDF of each pressure and season for each okta/cloudheight/
       windspeed/pressure/pressure duration Markov chain.
222
223 % seasonal high pressure
224 springhp_prob=springhp./(sum(springhp,2)*ones(1,10));%Calculate the PDF
225 cum_springhp_prob=cumsum(springhp_prob,2); %convert into CDF
226 % this process is repeated for all markov chains
227 summerhp_prob=summerhp./(sum(summerhp,2)*ones(1,10));
228 cum_summerhp_prob=cumsum(summerhp_prob,2);
229 autumnhp_prob=autumnhp./(sum(autumnhp,2)*ones(1,10));
230 cum_autumnhp_prob=cumsum(autumnhp_prob,2);
231 winterhp_prob=winterhp./(sum(winterhp,2)*ones(1,10));
232 cum_winterhp_prob=cumsum(winterhp_prob,2);
233
234 % seasonal low pressure
235 springlp_prob=springlp./(sum(springlp,2)*ones(1,10));
236 cum_springlp_prob=cumsum(springlp_prob,2);
237 summerlp_prob=summerlp./(sum(summerlp,2)*ones(1,10));
238 cum_summerlp_prob=cumsum(summerlp_prob,2);
239 autumnlp_prob=autumnlp./(sum(autumnlp,2)*ones(1,10));
240 cum_autumnlp_prob=cumsum(autumnlp_prob,2);
241 winterlp_prob=winterlp./(sum(winterlp,2)*ones(1,10));
242 cum_winterlp_prob=cumsum(winterlp_prob,2);
243
244 % low pressure system duration
245 springlpsys(:,[2,3])=[];   springlpsys_mean=mean(springlpsys);
246 springlpsys_std=std(springlpsys); %find standard deviation for pressure
247 summerlpsys(:,[2,3])=[];   summerlpsys_mean=mean(summerlpsys);
248 summerlpsys_std=std(summerlpsys);
249 autumnlpsys(:,[2,3])=[];   autumnlpsys_mean=mean(autumnlpsys);
250 autumnlpsys_std=std(autumnlpsys);
251 winterlpsys(:,[2,3])=[];   winterlpsys_mean=mean(winterlpsys);
252 winterlpsys_std=std(winterlpsys);
253
254 % high pressure system duration
255 springhpsys(:,[2,3])=[];   springhpsys_mean=mean(springhpsys);
256 springhpsys_std=std(springhpsys);
257 summerhpsys(:,[2,3])=[];   summerhpsys_mean=mean(summerhpsys);
258 summerhpsys_std=std(summerhpsys);
259 autumnhpsys(:,[2,3])=[];   autumnhpsys_mean=mean(autumnhpsys);
260 autumnhpsys_std=std(autumnhpsys);
261 winterhpsys(:,[2,3])=[];   winterhpsys_mean=mean(winterhpsys);

```

```

262 winterhpsys_std=std(winterhpsys);
263
264 % produce morning markov chains for each season
265 morningspring_prob=morningspring./(sum(morningspring,2)*ones(1,10));
266 cum_morningspring_prob=cumsum(morningspring_prob,2);
267 morningsummer_prob=morningsummer./(sum(morningsummer,2)*ones(1,10));
268 cum_morningsummer_prob=cumsum(morningsummer_prob,2);
269 morningautumn_prob=morningautumn./(sum(morningautumn,2)*ones(1,10));
270 cum_morningautumn_prob=cumsum(morningautumn_prob,2);
271 morningwinter_prob=morningwinter./(sum(morningwinter,2)*ones(1,10));
272 cum_morningwinter_prob=cumsum(morningwinter_prob,2);
273
274 % produce wind speed markov chains for each season
275 wind_spring_prob=wind_spring./(sum(wind_spring,2)*ones(1,length(wind_spring)));
276 cum_wind_spring_prob=cumsum(wind_spring_prob,2);
277 wind_summer_prob=wind_summer./(sum(wind_summer,2)*ones(1,length(wind_summer)));
278 cum_wind_summer_prob=cumsum(wind_summer_prob,2);
279 wind_autumn_prob=wind_autumn./(sum(wind_autumn,2)*ones(1,length(wind_autumn)));
280 cum_wind_autumn_prob=cumsum(wind_autumn_prob,2);
281 wind_winter_prob=wind_winter./(sum(wind_winter,2)*ones(1,length(wind_winter)));
282 cum_wind_winter_prob=cumsum(wind_winter_prob,2);
283
284 % produce cloud height markov chains for each season
285 cloudheight_spring_prob=cloudheight_spring./(sum(cloudheight_spring,2)*ones(1,
length(cloudheight_spring)));
286 cum_cloudheight_spring_prob=cumsum(cloudheight_spring_prob,2);
287 cloudheight_summer_prob=cloudheight_summer./(sum(cloudheight_summer,2)*ones(1,
length(cloudheight_summer)));
288 cum_cloudheight_summer_prob=cumsum(cloudheight_summer_prob,2);
289 cloudheight_autumn_prob=cloudheight_autumn./(sum(cloudheight_autumn,2)*ones(1,
length(cloudheight_autumn)));
290 cum_cloudheight_autumn_prob=cumsum(cloudheight_autumn_prob,2);
291 cloudheight_winter_prob=cloudheight_winter./(sum(cloudheight_winter,2)*ones(1,
length(cloudheight_winter)));
292 cum_cloudheight_winter_prob=cumsum(cloudheight_winter_prob,2);
293
294 %pressure
295 PAAMarkov_prob=PAAMarkov./(sum(PAAMarkov,2)*ones(1,Pmax));
296 cum_PAAMarkov_prob=cumsum(PAAMarkov_prob,2);
297 PBAMarkov_prob=PBAMarkov./(sum(PBAMarkov,2)*ones(1,Pmax));
298 cum_PBAMarkov_prob=cumsum(PBAMarkov_prob,2);

```

A.6 Preliminary weather variable generation framework

```

1 %% Preliminary Stochastic Weather Generation Setup
2 %define the seasons start and end day
3 springstart=60;springend=151;
4 summerstart=152;summerend=243;
5 autumnstart=244;autumnend=334;
6 winterstart=and(1,335);winterend=and(59,365);
7
8 % Randomly set beginning pressure system with 50:50 chance

```

```

9  if rand<0.5; start_pressure_sys=1; %1=above average pressure
10 else start_pressure_sys=0;      % 0= below average pressure.
11 end
12
13 % determine start season for the specified start point
14 if start_day>= springstart && start_day<=springend;      start_season=1;
15 elseif start_day>=summerstart && start_day<=summerend;  start_season=2;
16 elseif start_day>= autumnstart && start_day<=autumnend; start_season=3;
17 else                                                    start_season=4;
18 end
19
20 %determine the first morning okta Markov chain required
21 switch start_season; %use the start season to select appropriate MTM
22     case 1;          weather_start=cum_morningspring_prob;
23     case 2;          weather_start=cum_morningsummer_prob;
24     case 3;          weather_start=cum_morningautumn_prob;
25     case 4;          weather_start=cum_morningwinter_prob;
26 end
27
28 hours=ceil(num_of_days*24); %total number of hours in the simulation
29 day_number=start_day; %set the day number at start of simulation
30 year=start_year; %set the year at start of simulation
31 pressure_sys=start_pressure_sys; %set the pressure at start of simulation
32 current_season=start_season; %set the season
33 hour_number=1; %set the hour number
34
35 %Determine the starting pressure system duration
36 %Normal dist with observed means and stddev
37 if pressure_sys==0; %if it is low pressure
38     switch start_season; % find the current seasons pressure system
39         case 1;
40             sys_duration=floor(normrnd(springlpsys_mean, springlpsys_std));
41         case 2;
42             sys_duration=floor(normrnd(summerlpsys_mean, summerlpsys_std));
43         case 3;
44             sys_duration=floor(normrnd(autumnlpsys_mean, autumnlpsys_std));
45         case 4;
46             sys_duration=floor(normrnd(winterlpsys_mean, winterlpsys_std));
47     end
48 else % else if it is high pressure
49     switch start_season; % find the current seasons pressure system
50         case 1;
51             sys_duration=floor(normrnd(springhpsys_mean, springhpsys_std));
52         case 2;
53             sys_duration=floor(normrnd(summerhpsys_mean, summerhpsys_std));
54         case 3;
55             sys_duration=floor(normrnd(autumnhpsys_mean, autumnhpsys_std));
56         case 4;
57             sys_duration=floor(normrnd(winterhpsys_mean, winterhpsys_std));
58     end
59 end
60
61 % Randomly set all variables at the start of simulation from N[0,1]
62 current_cloud_height=ceil(rand*250);
63 current_wind_speed =ceil(rand*wind_max);

```

```

64 current_weather_okta =ceil(rand*10);
65 current_pressure=round(pressure_avg); %take the mean for start P
66
67 %select appropriate above or below pressure Markov chain
68 if pressure_sys==0; %if pressure system is below
69     Pressure_markov=cum_PBAmarkov_prob;
70 else Pressure_markov=cum_PAAmarkov_prob;
71 end
72
73 %Determine starting Wind speed and cloud height using season Markov chains
74 switch start_season;
75     case 1;
76         wind_start=cum_wind_spring_prob;
77         cloud_height_start=cum_cloudheight_spring_prob;
78     case 2;
79         wind_start=cum_wind_summer_prob;
80         cloud_height_start=cum_cloudheight_summer_prob;
81     case 3;
82         wind_start=cum_wind_autumn_prob;
83         cloud_height_start=cum_cloudheight_autumn_prob;
84     case 4;
85         wind_start=cum_wind_winter_prob;
86         cloud_height_start=cum_cloudheight_winter_prob;
87 end
88
89 % Stochastically select first of weather variables
90 future_weather_okta=1+sum(weather_start(current_weather_okta,:)<rand);
91 future_wind_speed=1+sum(wind_start(current_wind_speed,:)<rand);
92 future_cloud_height=1+sum(cloud_height_start(current_cloud_height,:)<rand);
93 future_pressure=1+nansum(Pressure_markov(current_pressure,:)<rand);
94
95 % Pre-allocate memory for speed by making the arrays
96 weather_record = zeros(hours,10);
97 sun_obsured = zeros(hours*60,1);
98 sun_obs_record=zeros(hours*60,3);
99 u_ref=zeros(hours*60,1);
100 hour_x=zeros(hours*60,1);

```

A.7 Stochastic generation of weather variable and sun obscured time series

```

1 %% Stochastic Generation of Weather Variable Time Series
2 for hour=1:hours; %loop through every hour of the desired simulation time
3
4     % Stochastic selection of weather variables
5     current_weather_okta=future_weather_okta;
6     current_wind_speed=future_wind_speed;
7     current_cloud_height=future_cloud_height;
8     current_pressure=future_pressure;
9
10    % Determine the current month

```

```

11     month = ceil(day_number/30.501);
12
13
14     % Convert Okta number into its proportionate sky coverage (out of 10)
15     % WMO2700 code provides conversions of okta in 8ths to 10ths.
16     switch current_weather_okta;
17         case 1; coverage=1;                %<=1/10 but not zero
18             okta_minutely(hour*60-59:hour*60)=1;
19         case 2; coverage=2+floor(2*rand);  %2/10 - 3/10
20             okta_minutely(hour*60-59:hour*60)=2;
21         case 3; coverage=4;                %4/10
22             okta_minutely(hour*60-59:hour*60)=3;
23         case 4; coverage=5;                %5/10
24             okta_minutely(hour*60-59:hour*60)=4;
25         case 5; coverage= 6;                %6/10
26             okta_minutely(hour*60-59:hour*60)=5;
27         case 6; coverage=7+floor(2*rand);  %7/10 or 8/10
28             okta_minutely(hour*60-59:hour*60)=6;
29         case 7; coverage=9;                %9/10+ but not 10/10
30             okta_minutely(hour*60-59:hour*60)=7;
31         case 8; coverage=10;                %10/10
32             okta_minutely(hour*60-59:hour*60)=8;
33         case 9; coverage=10;                %meteorological phenomena
34             okta_minutely(hour*60-59:hour*60)=9;
35         case 10; coverage=0;                %0/10
36             okta_minutely(hour*60-59:hour*60)=0;
37     end
38
39     % Calculate the windspeed at the cloud height (met office):
40     % Use log scale for <1km height. Use gamrnd for >1km.
41     if current_cloud_height<100; %height is in decameters so <1km
42         u_ref=ceil(current_wind_speed*0.515*((log(current_cloud_height*10/0.14))
43         /(log(10/0.14))))); %perform the met office interpolation %note the 0.515
44         conversion of knots to m/s. 0.14 is for a rural setting, this could perhaps
45         be updated based on desired location.
46     else % else use the gamrnd function above 1km.
47         %gives mean of 3.49m/s with range 0-25 and 0.5% above 25.
48         u_ref=round(gamrnd(2.7,2.144));
49     end
50
51     % if the windspeed is outside of limits, set to closest min or max
52     if u_ref<1;         u_ref=1;         end
53     if u_ref>60;        u_ref=60;        end
54
55     % Keep a record of weather variable time series
56     weather_record(hour,1)=hour_number;
57     weather_record(hour,2)=day_number;
58     weather_record(hour,3) = year;
59     weather_record(hour,4)=current_season;
60     weather_record(hour,5)=current_weather_okta;
61     weather_record(hour,6) = pressure_sys;
62     weather_record(hour,7)=sys_duration;
63     weather_record(hour,10) = u_ref;
64     weather_record(hour,8) = coverage;
65     weather_record(hour,9) = current_cloud_height;
66

```

```

63 % Create a minutely time series of variables
64 month_record(hour*60-59:hour*60,1)= month;
65 season_record(hour*60-59:hour*60,1)=current_season;
66 pressure_record(hour*60-59:hour*60,1)=current_pressure;
67 coverage_vector(hour*60-59:hour*60,1)=coverage;
68
69 % Select cloud sample from appropriate bin to creat 1-min time series
70 switch coverage
71     case {1,2,3,4,5,6,7,8,9} %for coverage of 1-9.
72         random=rand; % N[0,1]
73         sun_obsured(hour*60-59:hour*60,1)=sun_obsured_options(((u_range*
num_of_options*(coverage-1))+ceil(num_of_options*u_ref-(random*
num_of_options))),:); %note this is a complicated indexing that uses the
coverage and the windspeed to randomly select 1 of the 1000 options for that
particular C and u.
74     case 10 %for fully overcast, the coverage for the hour is total, and so
all 60 elements recieve a value of 1.
75         sun_obsured(hour*60-59:hour*60,1) = 1; %sun_obsured is the primary
array to come out of this long for loop.
76     end
77
78 % Produce minutley hour and day fraction. e.g. minute 1 of hour 1 would be
1.00. minute 60 of hour 1 would be 1.983. This is for use later in the
irradiance calculations
79 hour_run=hour_number+(1/60):(1/60):hour_number+1;
80 hour_x(hour*60-59:hour*60,1)=hour_run';
81 day_run=linspace(day_number+hour_number*(1/60),day_number+(hour_number+1)
*(1/60),60);
82 day(hour*60-59:hour*60,1)=day_run';
83
84 % Determine if current year is leap year and add the additional day
85 if floor(year/4)/(year/4)==1; %if the year is a multiple of 4
86     days_in_year=366; %increase the days of the year to 366.
87 else days_in_year=365; % else set it at 365.
88 end
89
90 % Determine current season
91 if day_number>= springstart && day_number<=springend;
92     current_season=1;
93 elseif day_number>=summerstart && day_number<=summerend;
94     current_season=2;
95 elseif day_number>= autumnstart && day_number<=autumnend;
96     current_season=3;
97 else
98     current_season=4;
99 end
100
101 % Stochastically set length of next pressure system if it has ended
102 %ASSUMPTION: pressure system always switches from low to high
103 if sys_duration==0; %if ended..
104     if pressure_sys==0; %if currently below average...
105         pressure_sys=1;%... switch to an above average.
106         switch current_season;
107             case 1;
108                 sys_duration=round(normrnd(springhpsys_mean , springhpsys_std));

```

```

109         case 2;
110             sys_duration=round(normrnd(summerhpsys_mean , summerhpsys_std));
111         case 3;
112             sys_duration=round(normrnd(autumnhpsys_mean , autumnhpsys_std));
113         case 4;
114             sys_duration=round(normrnd(winterhpsys_mean , winterhpsys_std));
115     end
116 else pressure_sys=0; %...else switch to a below average
117     switch current_season;
118         case 1;
119             sys_duration=round(normrnd(springlpsys_mean , springlpsys_std));
120         case 2;
121             sys_duration=round(normrnd(summerlpsys_mean , summerlpsys_std));
122         case 3;
123             sys_duration=round(normrnd(autumnlpsys_mean , autumnlpsys_std));
124         case 4;
125             sys_duration=round(normrnd(winterlpsys_mean , winterlpsys_std));
126     end
127 end
128 % if pressure system hasn't finished yet, reduce the count by an hour.
129 else sys_duration=sys_duration-1;
130 end
131
132 % Select correct Markov chains
133 switch current_season;
134     case 1 % spring
135         if pressure_sys==0; future_weather_markov=cum_springlp_prob;
136         else future_weather_markov=cum_springhp_prob;
137         end
138     case 2 % summer
139         if pressure_sys==0; future_weather_markov=cum_summerlp_prob;
140         else future_weather_markov=cum_summerhp_prob;
141         end
142     case 3 % autumn
143         if pressure_sys==0; future_weather_markov=cum_autumnlp_prob;
144         else future_weather_markov=cum_autumnhp_prob;
145         end
146     case 4 % winter
147         if pressure_sys==0; future_weather_markov=cum_winterlp_prob;
148         else future_weather_markov=cum_winterhp_prob;
149         end
150 end
151
152 % Select the morning Markov chain when hour number is 1-5am
153 if hour_number<6;
154     switch current_season;
155         case 1 ;
156             future_weather_markov=cum_morningspring_prob;
157         case 2 ;
158             future_weather_markov=cum_morningsummer_prob;
159         case 3 ;
160             future_weather_markov=cum_morningautumn_prob;
161         case 4 ;
162             future_weather_markov=cum_morningwinter_prob;
163     end

```

```

164     end
165
166     % Select the appropriate Markov chain for wind speed and cloud height
167     switch current_season;
168         case 1;
169             wind_start=cum_wind_spring_prob;
170             cloud_height_start=cum_cloudheight_spring_prob;
171         case 2;
172             wind_start=cum_wind_summer_prob;
173             cloud_height_start=cum_cloudheight_summer_prob;
174         case 3;
175             wind_start=cum_wind_autumn_prob;
176             cloud_height_start=cum_cloudheight_autumn_prob;
177         case 4;
178             wind_start=cum_wind_winter_prob;
179             cloud_height_start=cum_cloudheight_winter_prob;
180     end
181
182     % Select appropriate pressure Markov chain
183     if pressure_sys==0;
184         Pressure_markov=cum_PBAmarkov_prob;
185     else Pressure_markov=cum_PAAmarkov_prob;
186     end
187
188     % Stochastically determine the future weather states
189     future_cloud_height=1+sum(cloud_height_start(current_cloud_height,:)<rand);
190     future_weather_okta=1+sum(future_weather_markov(current_weather_okta,:)<rand
191 );
192     future_wind_speed=1+sum(wind_start(current_wind_speed,:)<rand);
193     future_pressure=1+nansum(Pressure_markov(current_pressure,:)<rand);
194
195     % Update the timing structures.
196     hour_number=hour_number+1; %increase the hour_number by one
197     if hour_number==25;
198         hour_number=1;% if this hour number exceeds 25, reset it
199         day_number=day_number+1; %increase day number
200         %has a year ended?
201         if day_number==days_in_year %if yes then reset timings
202             day_number=1;
203             hour_number=1;
204             year=year+1;
205         end
206     end
207 end %and repeat until the simulation is complete

```

A.8 Global horizontal clear-sky irradiance calculations

```

1  %% Irradiance Calculations (by Dr Chris Smith and Jamie Bright)
2
3  % Calculate solar angles and timings - using Muriel-Blanco algorithm
4  %Blanco-Muriel et al., 2001, Solar Energy 70(5) 431-441
5  julian_day=day+2455927.5-1/48;

```

```

6 n=julian_day-2451545; %Normalise to 01/01/2000 00:00
7 Omega=2.1429-0.0010394594*n;
8 L=4.8950630+0.017202791698*n;% mean longitude
9 g=6.2400600+0.0172019699*n;% mean anomaly
10 l=L+0.03341607*sin(g)+0.00034894*sin(2*g)-0.0001134-0.0000203*sin(Omega); %
    ecliptic longitude
11 ep=0.4090928-6.2140e-9*n+0.0000396*cos(Omega); % obliquity of the ecliptic
12 ra=mod(atan2(cos(ep).*sin(l),cos(l)),2*pi);%right ascension
13 delta=asin(sin(ep).*sin(l));%declination
14 gmst=6.6974243242+0.0657098283*n+mod(hour_x'-2,24);%GMT
15 lmst=(gmst*15+longitude)*pi/180;%local mean sidereal time
16 hour_angle=lmst-ra;
17 theta_z=acos(cosd(latitude)*cos(hour_angle).*cos(delta)+sin(delta)*sind(latitude
    ));%Solar zenith angle
18 Parallax=6371.01/149597890*sin(theta_z); %Parallax correctional factor
19 zenith_angle=(theta_z+Parallax)*180/pi;%adjustment to zenith due to parallax.
20 elevation=90-zenith_angle;%elevation angle, i.e. complement of zenith
21 azimuth=atan(-sin(hour_angle)./(tan(delta).*cosd(latitude)-sind(latitude).*cos(
    hour_angle)));%azimuth angle of sun
22 incident_angle=real(acosd(sind(zenith_angle).*sind(panel_pitch).*cosd(
    panel_azimuth-azimuth)+cosd(zenith_angle).*cosd(panel_pitch))');%solar
    incident angle taking into account panel tilt and azimuth
23
24 % top of atmosphere normal irradiance
25 eccentricity=1+0.03344*cos(2*pi*day/365.25-0.048869);
26 solar_constant=1367*eccentricity;clear eccentricity
27
28 % Kasten airmass formula as a function of zenith angle
29 airmass=zeros(1,length(month_record));
30 airmass(zenith_angle<=90)=(1-height_above_sea_level/10000)*((cosd(zenith_angle(
    zenith_angle<=90))+0.50572*(96.07995-zenith_angle(zenith_angle<=90))
    .^(-1.6364)).^(-1));
31 airmass(zenith_angle>90)=Inf;
32
33 %CalculateRayleighoptical depth as a function of airmass
34 Rayleigh=zeros(length(solar_constant),1);
35 Rayleigh(airmass<20)=1./(6.6296+1.7513*airmass(airmass<20)-0.1202*airmass(
    airmass<20).^2+0.0065*airmass(airmass<20).^3-0.00013*airmass(airmass<20).^4)
    );
36 Rayleigh(airmass>=20)=1./(10.4+0.718*airmass(airmass>=20));
37
38 % Retrieve Linke Turbidity for nearest 1/12 degree grid square and month
39 % get values from TIF file of Linke turbidities
40 lat_index=round((-latitude+1/24+90)*4320/360); %round the latitude
41 lon_index=round((longitude+1/24+180)*4320/360); %round the longitude
42 LinkeTurbidity2=zeros(12,1);%pre allocate array with space for each month
43 month_strings={'January','February','March','April','May','June','July','August'
    ,'September','October','November','December'};
44 for i=1:12
45     %Open Linke turbidity for January
46     tiffData=Tiff(['supportingfiles/linke/',month_strings{i},'.tif'],'r');
47     loadup = single(tiffData.read())/20; % Linke Turbidity = greyscale_value/20
48     LinkeTurbidity2(i)=loadup(lat_index,lon_index); %extract the appropriate
    linke turbidity value for that latitude and longitude and place in array
49     clear loadup %clear the loadup data to save space

```

```

50     tiffData.close(); %close the tiff data
51 end
52 clear tiffData %clear all the tiff data.
53
54 % Calculate direct and diffuse clear sky irradiance
55 diffuse_horizontal_cs=solar_constant.*(0.0065+(-0.045+0.0646.*LinkeTurbidity2(
    month_record)).*cosd(zenith_angle'))+(0.014-0.0327.*LinkeTurbidity2(
    month_record)).*cosd(zenith_angle').^2);
56 diffuse_horizontal_cs(diffuse_horizontal_cs<0)=0;
57 direct_horizontal_cs=solar_constant.*exp(-0.8662.*LinkeTurbidity2(month_record)
    .*Rayleigh.*airmass').*cosd(zenith_angle');
58 direct_horizontal_cs(direct_horizontal_cs==0)=0;
59 direct_horizontal_cs(isnan(direct_horizontal_cs))=0;
60 global_horizontal_cs=direct_horizontal_cs+diffuse_horizontal_cs;

```

A.9 Generation of minutely clear-sky index time series

```

1 %% Generate the Minutely Clear-Sky Index
2
3 % Select clear and cloudy properties for each day from appropriate distribution
4 okta_hourly=weather_record(:,5);%hourly okta time series
5
6 hours = numel(okta_hourly); %total number of hours
7 days = hours/24; %number of days
8
9 % Apply variability to clouded moments using 6 min cycles.
10 resolution=6; %must be a factor of 60 1,2,3,4,5,6,10,12,15,20,30,60.
11 shift_factor=60/resolution; %used in indexing.
12
13 % Pre allocate memory
14 obscured_min=zeros(hours*60,1);
15 not_obscured_min=zeros(hours*60,1);
16 obscured_factored=zeros(hours*shift_factor,1);
17 kcMinutely=zeros(numel(sun_obscured),1);
18
19 % Determine Okta value every 6 mins.
20 for i=1:length(okta_hourly) % loop through each hour
21     okta_factored(i*shift_factor-(shift_factor-1):i*shift_factor)=okta_hourly(i);
22 end
23
24 % Pick clear-sky index from the corresponding okta weighted distribution
25 obscured_factored(okta_factored<=6) = normrnd(0.6784, 0.2046, numel(
    okta_factored(okta_factored<=6)),1); %for okta of <=6, choose a kc from the
    normal distribution of mean 0.6784 and stddev 0.2046
26 obscured_factored(okta_factored==7) = wblrnd(0.557736, 2.40609, numel(
    okta_factored(okta_factored==7)),1); %as above but for okta 7 and using
    weibul distribution
27 obscured_factored(okta_factored>=8) = gamrnd(3.5624, 0.08668, numel(
    okta_factored(okta_factored>=8)),1); %as above but for okta 8 and using
    gamma distribution
28
29 % Ensure sun_obscured hourly does not exceed 1

```

```

30 while numel(obsured_factored(obsured_factored>1))>0 % limited obsured kc
    value to 1. if it is, re select from the okta 8 distribution.
31     obsured_factored(obsured_factored>1)=gamrnd(3.5624, 0.08668, numel(
        obsured_factored(obsured_factored>1)), 1); %re select using okta 8
        distribution
32 end
33
34 % Interpolate the k_c values to 1-min resolution from the 6 min
35 for i=1:length(obsured_factored)-1;
36     obsured_min(i*resolution-(resolution-1):i*resolution)=linspace(
        obsured_factored(i),obsured_factored(i+1),resolution);
37 end
38
39 % Pick clearsky minutes from a normal distribution. One k_c each day (
    variability of atmospheric turbidity, etc)
40 not_obsured=normrnd(0.99,0.08,days,1);
41
42 % Combine clear-sky and covered k_c time series based on sun obsured [0,1]
43 kcMinutely(sun_obsured==1)=obsured_min(sun_obsured==1);
44 kcMinutely(sun_obsured==0)=not_obsured(ceil(find(sun_obsured==0)/1440));
45
46 % Add gaussian white noise as a function of cloud cover for all conditions
47 kcMinutely(sun_obsured==1)=kcMinutely(sun_obsured==1).*normrnd(1,0.01+0.003*
    okta_minutely(sun_obsured==1)');
48 kcMinutely(sun_obsured==0)=kcMinutely(sun_obsured==0).*normrnd(1,0.001+0.0015*
    okta_minutely(sun_obsured==0)');
49
50 % For long periods of okta 0, apply a smoothing.
51 % pre allocate memory
52 Period_of_0k0=zeros(length(okta_minutely),1);
53 Ok8_ind=zeros(length(okta_minutely),1);
54
55 % Find beginning and end of each okta 0 period
56 for i=2:length(okta_minutely)-1;
57     % indicate the start and end of okta 0 period with a 1 and 2 among 0s
58     if okta_minutely(i)==0 && okta_minutely(i+1)==0 && okta_minutely(i-1)~=0;
59         Period_of_0k0(i)=1;
60     end
61     if okta_minutely(i)==0 && okta_minutely(i-1)==0 && okta_minutely(i+1)~=0;
62         Period_of_0k0(i)=2;
63     end
64 end
65
66 % Determine the length of the period
67 Ok0_duration=1; %Set tally varibale at 1
68 ok0_cutoff=3.5; %set the cutoff period in hours
69 for i=1:length(Period_of_0k0)-1 %loop through every minute
70     %Once identified the start, begin the tally
71     if Period_of_0k0(i)==1
72         if i+Ok0_duration==numel(Period_of_0k0);break; end
73         while Period_of_0k0(i+Ok0_duration)~=2;
74             %update tally until end of okta period is found
75             Ok0_duration=Ok0_duration+1;
76
77             % model saftey protocols

```

```

78         if i==numel(Period_of_0k0); break; end
79         if i+0k0_duration==numel(Period_of_0k0);break; end
80     end
81 end
82
83 %Check the duration against the desired cutoff period
84 if 0k0_duration>=ok0_cutoff*60;
85     %Apply smoothing from normal distribution around 1
86     kcMinutely(i:i+0k0_duration)= normrnd(1,0.0015,0k0_duration+1,1);
87 end
88
89 % reset the duration tally variable
90 0k0_duration=1;
91 end
92
93 % For long periods of 0k8 apply smoothing. Repeat of above
94 for i=2:length(okta_minutely)-1;
95     if okta_minutely(i)==8 && okta_minutely(i+1)==8 && okta_minutely(i-1)~=8;
96         0k8_ind(i)=1;
97     end
98     if okta_minutely(i)==8 && okta_minutely(i-1)==8 && okta_minutely(i+1)~=8;
99         0k8_ind(i)=2;
100    end
101 end
102
103 0k8_duration=1;
104 ok8_cutoff=5;
105 intervals=ok8_cutoff*4; %20 intervals across any long period of 0k8.
106
107 for i=1:length(0k8_ind)-1
108     if 0k8_ind(i)==1
109         if i+0k8_duration==numel(0k8_ind);break; end
110         while 0k8_ind(i+0k8_duration)~=2;
111             0k8_duration=0k8_duration+1;
112             if i==numel(0k8_ind); break; end
113             if i+0k8_duration==numel(0k8_ind);break; end
114         end
115     end
116
117     if 0k8_duration>=ok8_cutoff*60;
118
119         % Make blank array of each hour within extended okta 8 period
120         els=zeros(1+intervals,1);
121         for j=2:length(els);
122             %evenly split the period by the desired number of intervals
123             els(j)=ceil(0k8_duration*((j-1)/intervals));
124         end
125
126         % Update the kcminutely in the appropriate place with linearly spaced kc
127         % values with small flux.
128         for j=1:length(els)-1;
129             kcMinutely(i+els(j):i+els(j+1))=linspace(kcMinutely(i+els(j)),
130 kcMinutely(i+els(j))*normrnd(1,0.1),els(j+1)-els(j)+1);

```

```

131     % Add structured fluctuations throughout the smoothed data
132     fluxes=roundn(rand(intervals,1),-1);
133     gap=ceil(Ok8_duration/intervals);
134     flux_min=zeros(length(Ok8_duration),1);
135     for k=1:length(fluxes);
136         flux_min(k*gap-(gap-1):k*gap,1)=fluxes(k,1);
137     end
138     %similar to a gaussian noise, however allows for variation in std dev
139     for k=1:length(flux_min);
140         if flux_min(k,1)<0.4
141             flux_min(k,2)=1+abs(normrnd(0,0.005));
142         elseif flux_min(k,1)<0.7
143             flux_min(k,2)=1+abs(normrnd(0,0.03));
144         else flux_min(k,2)=1+abs(normrnd(0,0.05));
145         end
146     end
147     % Apply fluctuations back into the clear-sky index time series
148     kcMinutely(i:i+Ok8_duration)=kcMinutely(i:i+Ok8_duration).*flux_min(:,2);
149
150     end
151     Ok8_duration=1;
152 end
153
154 %Clear sky index upper limit maximum by zenith angle
155 for i=1:length(kcMinutely); %check each k_c
156     %determine the theoretical maximum kc value based on the zenith at time
157     kcmax = 27.21*exp(-114*cosd(zenith_angle(i))) + 1.665*exp(-4.494*cosd(
158         zenith_angle(i))) + 1.08;
159
160     % Reassign random value from ok 8 distribution if too large
161     if kcMinutely(i)>kcmax
162         kcMinutely(i)=wblrnd(0.3, 1.7);
163     end
164
165     % Assign minimum value if below theoretically possible
166     if kcMinutely(i)<0.01;
167         kcMinutely(i)=0.01;
168     end
169 end
170
171 % Add irradiance peaks at moment of cloud shift. Cloud edge enhancement
172
173 %assign a chance of CEE
174 chance=0.40;
175
176 %apply irradiance peaks if chance is satisfied
177 for i=3:length(kcMinutely);
178     a=rand; % select a random value to test against the chance variable
179     %at period the minute before/after a cloud, apply a CEE by norm dist.
180     if sun_obscurd(i-1)-sun_obscurd(i)==1;
181         if a>chance;
182             kcMinutely(i)=kcMinutely(i)*normrnd(1.05,0.01,1);
183         end
184     elseif sun_obscurd(i-1)-sun_obscurd(i)==-1;

```

```

185     if a>chance
186         kcMinutely(i-1)=kcMinutely(i-1)*normrnd(1.05,0.01,1);
187     end
188 end
189 % add slighter enhancement in minutes either side of above
190 if sun_obsured(i-2)-sun_obsured(i-1)==1;
191     if a>chance;
192         kcMinutely(i)=kcMinutely(i)*normrnd(1.025,0.01,1);
193     end
194 elseif sun_obsured(i-2)-sun_obsured(i-1)==-1;
195     if a>chance;
196         kcMinutely(i)=kcMinutely(i)*normrnd(1.025,0.01,1);
197     end
198 end
199 end

```

A.10 Calculation of tilted irradiance by component

```

1 %% Calculate GHI, DNI and DHI, and translate onto arbitrary plane
2
3 %Apply clear-sky index to global clear sky irradiance
4 global_horizontal = kcMinutely .* global_horizontal_cs;
5
6 %Calculate DNI according to Muller and Trentmann (2010).
7 direct_horizontal=zeros(numel(sun_obsured),1);
8 direct_horizontal(kcMinutely<1&kcMinutely>19/69)=direct_horizontal_cs(kcMinutely
    <1&kcMinutely>19/69).*(kcMinutely(kcMinutely<1&kcMinutely>19/69)-0.38*(1-
    kcMinutely(kcMinutely<1&kcMinutely>19/69))).^(2.5);
9 direct_horizontal(kcMinutely>=1)=direct_horizontal_cs(kcMinutely>=1);
10 direct_horizontal(direct_horizontal<0)=0;
11 global_horizontal(global_horizontal<0)=0;
12 diffuse_horizontal=global_horizontal-direct_horizontal;
13
14 % Translate onto arbitrary plane using Klucher model
15 F=1-(diffuse_horizontal./global_horizontal).^2; % modulating factor
16 isotropic=(1+cosd(panel_pitch))/2; % isotropic component
17 horizontal=(1+F.*(sind(panel_pitch)/2).^3); % horizon brightening term
18 circumsol=(1+F.*(cosd(incident_angle)).^2.*(sind(zenith_angle')).^3);%
    circumsolar diffuse irradiance
19 panel_irradiance=diffuse_horizontal.*isotropic.*horizontal.*circumsol+
    direct_horizontal./cosd(zenith_angle').*cosd(incident_angle);
20
21 %take account of night
22 panel_irradiance = (panel_irradiance > 0).*panel_irradiance;

```

Appendix B

Spatially Decorrelating Solar Irradiance Generator scripts

B.1 Cloud field production

```
1 %% Vectorised Method of producing cloud fields of different coverage, C, at
   different cloud speeds, u.
2 clearvars
3
4 %% input variables
5 hour=3600; %hour in seconds
6 temporal_res=3600; %temporal resolution in seconds
7 spatial_res=1500; %spatial resolution in meters
8 num_of_cloud_field=500; %number of options of each u, at each C.
9 c_range=11; % C_r=(0:10)/10. coverage 0:10.
10 u_range=30; % cloud speed range 1:30 m/s
11 max_num_of_clouds=1200; %maximum number of clouds. This is used within the
   output arrays.
12 domain_y_max=spatial_res; %maximum size within the y direction. This is a
   function of the spatial resolution.
13
14 B=1.66; % Power Law Exponent ///  $p(x)=Cx^{-B}$  /// (Wood & Field, 2011, Journal of
   Climate, Volume 24, p4800).
15 cloud_x_min=1;%minimum cloud length (decameters). Power law applicable to
   0.1-1500km (Wood & Field,2011)
16 cloud_x_max=300;%maximum cloud length (decameters). (decameters selected as 10m
   =1 element resolution within vector) %set to 3000m as any cloud larger than
   this would fill the domain.
17
18 %% make the blank arrays to pre-allocate memory.
19 cloud_field0=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field); %make output
   arrays (filled with 3(xyr) x max_num_of_clouds options)
20 cloud_field1=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
21 cloud_field2=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
22 cloud_field3=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
23 cloud_field4=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
```

```

24 cloud_field5=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
25 cloud_field6=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
26 cloud_field7=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
27 cloud_field8=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
28 cloud_field9=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
29 cloud_field10=zeros(max_num_of_clouds*u_range,3*num_of_cloud_field);
30 clouds_store=zeros(u_range*c_range*num_of_cloud_field,1);
31 u_store=zeros(u_range*c_range*num_of_cloud_field,1);
32 C_store=zeros(u_range*c_range*num_of_cloud_field,1);
33
34 %Make the cloud field
35 error_flag=0; % the error flag is used to break the process if cloud number
    exceeded
36 rejections=0; %counts the rejected cloud fields for each u and C to help
    streamlining
37 for u=1:u_range; % loop through each cloud speed
38     domain_x_max=u*3600;
39     uTic=tic;
40     disp(['u = ',num2str(u)]);
41     for C=0:c_range
42         cloud_field=0; %a marker of the current field to populate
43         tic % start a timer to track for user
44         while cloud_field < num_of_cloud_field
45             coverage = -1; %mean coverage of current field. Set below
    permissible to begin
46             clouds = 0; % number of clouds within the field
47             output_x=zeros(max_num_of_clouds,1);%preallocate x,y,r arrays
48             output_y=zeros(max_num_of_clouds,1);
49             output_r=zeros(max_num_of_clouds,1);
50
51             while coverage < C % while the fields coverage is below the target
52
53                 %Add clouds cloud
54                 clouds_per_time=30+round(u/3)+C*u; %add more than 1 each
    iteration to save time
55                 clouds=clouds+1;
56                 num_of_clouds=clouds_per_time*clouds; %report current number of
    clodus
57                 if num_of_clouds>max_num_of_clouds;error_flag=1; end %use error
    flag if larger than permissible
58                 if error_flag==1;
59                     disp('ERROR: max number of clouds exceeded');
60                     break %break the process
61                 end
62
63                 for i=num_of_clouds-(clouds_per_time-1):num_of_clouds;
64                     output_x(i,1)=rand*domain_x_max; %place the centrepoint
    within the x domain
65                     output_y(i,1)=rand*domain_y_max; %place the centerpoint
    within the y domain
66                     output_r(i,1)=10*0.5*((alpha+beta*rand)^(1/(1-B))); %radius
    according to power law. in meters.
67                 end
68
69                 %Analyse Coverage Value

```

```

70         %the coverage will be calculated by regarding the domain area to
        be a set of xy coordinates. The exact area covered within the domain can be
        calculated, however it requires complex surface integrations, and so a
        simple counting squares approach will suffice.
71
72         xref=0;%the first x coordinate to check
73         yref=0;%the first y coordinate to check
74         xres=50;%is u*3600 m long. take a reading every "xres" metres
        along the x-axis of the domain. minimum 3600 long, maximum 108000 long.
        Accuracy and speed lives or dies here.
75         yres=50;%is 1500 m long. will check every "yres" metres along
        the y-axis of the domain.
76         covered_tally=0; %reset the tally of covered points
77         total_xys=(domain_x_max/xres)*(domain_y_max/yres); %total number
        of points within the domain. resembles an area. X/res * Y/res
78
79         while xref<domain_x_max && yref<domain_y_max %once xref and y
        ref == the limits of the domain, every point has been checked.
80
81         distances_to_cloud_centres=sqrt((output_x(1:num_of_clouds)-
        xref).^2+(output_y(1:num_of_clouds)-yref).^2); %find the distances to the
        centre point of each circle using pythagoras to find the hypotenuse of a
        right angled triangle with sides of dx and dy
82
83         if max(distances_to_cloud_centres<output_r(1:num_of_clouds))
        ==1; % d<r returns a binary array with 1 if the distance to cloud centre is
        less than r, and 0 if not. therefore the max of this array will return 1 or
        0. If the answer is 1 then the (xref,yref) coordinate is within a cloud's
        radius, and therefore covered.
84         covered_tally=covered_tally+1; %if it is covered, add a
        point to the tally
85         end
86
87         yref=yref+yres; %add an increment to the x axis for the next
        coordinate to check.
88         if yref==domain_y_max; %if the coordinate is at the end of
        the row...
89         xref=xref+xres; %.. then increase the column yref by the
        one increment==yres...
90         yref=0; %... and reset the xref back to 0. The while
        loop will end once xref and yref are at their max.
91         end
92         end
93
94         coverage=round(10*covered_tally/total_xys); %(the number of (
        xref,yref) coordinates covered, divided by the total number of (xref,yref)
        coordinates within the domain = the coverage fraction*10 and rounded=integer
        of coverage.
95
96         %if the sample is larger than C, then reset the sample and start
        again
97         if coverage>C;
98             clouds = 0;
99             coverage=-1;
100            output_x=zeros(max_num_of_clouds,1);

```

```

101         output_y=zeros(max_num_of_clouds,1);
102         output_r=zeros(max_num_of_clouds,1);
103         rejections=rejections+1;
104 %         disp(['    total rejections=',num2str(rejections),'| C=',
num2str(C),' u=',num2str(u)]);
105         end
106
107     end
108
109     cloud_field=cloud_field+1;
110     clouds_store((u-1)*c_range*num_of_cloud_field+C*num_of_cloud_field+
cloud_field,1)=num_of_clouds;
111     C_store((u-1)*c_range*num_of_cloud_field+C*num_of_cloud_field+
cloud_field,1)=coverage;
112     u_store((u-1)*c_range*num_of_cloud_field+C*num_of_cloud_field+
cloud_field,1)=u;
113
114
115     %% Allocate into bins arranged by coverage.
116     switch C
117     case 0
118         cloud_field0(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
119     case 1
120         cloud_field1(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
121     case 2
122         cloud_field2(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
123     case 3
124         cloud_field3(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
125     case 4
126         cloud_field4(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
127     case 5
128         cloud_field5(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
129     case 6
130         cloud_field6(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
131     case 7
132         cloud_field7(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
133     case 8

```

```

134         cloud_field8(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
135         case 9
136             cloud_field9(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
137         case 10
138             cloud_field10(max_num_of_clouds*u-(max_num_of_clouds-1):
max_num_of_clouds*u,cloud_field*3-2:cloud_field*3)=[output_x,output_y,
output_r]; %cloud_fieldx(x:x+500,y:y+3)=xyr.
139         end
140
141         if error_flag==1;break ;end %continue breakdown if error occurs
142     end
143     timer=toc;
144     disp([' C=',num2str(C), ' time=',num2str(timer/60),'mins'])
145
146     if error_flag==1;break ;end
147 end
148 toc(uTic)
149 if error_flag==1;break ;end
150 end
151
152 disp([' total rejections=',num2str(rejections)]);
153 %Summarise the results
154 u_mean=zeros(length(u_store)/num_of_cloud_field,1);
155 C_mean=zeros(length(u_store)/num_of_cloud_field,1);
156 n_mean=zeros(length(u_store)/num_of_cloud_field,1);
157
158 for i=1:length(u_mean)
159     %take means for summary
160     u_mean(i)=mean(u_store((i-1)*num_of_cloud_field+1:i*num_of_cloud_field));
161     C_mean(i)=mean(C_store((i-1)*num_of_cloud_field+1:i*num_of_cloud_field));
162     n_mean(i)=mean(clouds_store((i-1)*num_of_cloud_field+1:i*num_of_cloud_field)
);
163 end
164
165 %write to files out from matrix "cloud_field"
166 disp('Writing Files')
167 for i=0:c_range-1
168     string=['cloud_field',num2str(i)];
169     X=eval(string);
170     dlmwrite(['V2_clouds_vectors_sample',num2str(i),'__temporal_res_',num2str(
temporal_res),'__spatial_res_',num2str(spatial_res),'__',num2str(
num_of_cloud_field),'_num_of_cloud_field','__',num2str(max_num_of_clouds),'
max_num_of_clouds'],X)
171 end

```

B.2 User defined variables

```
1 %% Preamble
```

```

2 tic %set a timer
3 read_file_check=exist('read_files','var'); %check to see if data is already
    loaded (time saving). Note that this will fail if running two different
    versions simultaneously and "clear all" in the command window will be
    required.
4 data_exist_check=exist('cov1','var'); %check if the data is truly loaded (if
    first run fails (i.e. typo in run_validation, then this will not load
    without this line). This is a double check
5 if read_file_check==0; %if it isn't already loaded...
6     read_files='y'; %read in the files this simulation.
7 elseif data_exist_check==0; %if it isn't already loaded...
8     read_files='y';
9 else read_files='n'; %if they aren't loaded. load them
10 end
11 %clear all other variables so new simulation can be performed
12 clearvars -except pressure data cov* read_files tstart %keep continuous files
13
14 %% User Defined Variables
15 % User defined variables to set
16 start_day=1; % choose the day number from which to start from
17 start_year=2001; % choose the year from which to start from. Not really
    important as leap years not currently accounted for.
18 num_of_years=7; % set the duration from the model to run (1/365.25 will select 1
    day etc.)
19 num_of_days=365*6+366; %could account for leap years here, as per example.
20
21 %Unique to the SDSIG
22 time_zone_difference=0; %GMT=0. California is 8 hours behind.
23 latitude_central=50.2178; % set the latitude. Cambourne in this example.
24 longitude_central=-5.32656; % set the longitude of the desired location.
25 height_above_sea_level_central=87; %meters above sea level.
26 u_range=30; % the range of cloud speeds in m/s. Currently set as standardised
    from 'Cloud_Sampling_Technique.m'.
27 temporal_res=60; %number of time steps to be produced per hour. e.g. 1=1-hour.
    60=1-min. 3600=1-sec
28 spatial_res=1500;% size of the spatial domain in metres. New cloud fields would
    need to be produced if increased. Currently a computational limit, not a
    theoretical one.
29 max_num_of_clouds=1200; %the max number of clouds allowable in a cloud field.
    Computational limit
30 switch temporal_res
31     case 60
32 num_of_samples=500;% the number of options per windspeed per coverage.
33     case 3600
34         num_of_samples=200; %fewer for 1 sec as computationally more demanding
35 end
36
37 %% define the characteristics of the properties within the spatial domain
38
39 num_of_houses=50; %referred to as houses, but only indicates a location
40 house_info=zeros(num_of_houses,5); %define houses into x,y,h,azimuth,paneltilt
41 for i = 1:num_of_houses;
42     house_info(i,1)=1+spatial_res*rand; % assign random x axis
43     house_info(i,2)=1+spatial_res*rand;% assign random y axis

```

```

44     house_info(i,3)=height_above_sea_level_central*normrnd(1,0.1); %assign
        random hasl
45     house_info(i,4)=0; %azimuth, between East-West (-180,180)
46     house_info(i,5)=0; %assign panel tilt
47
48 end
49
50 [house_info(:,1),house_info(:,2)]=matrot(90,house_info(:,1),house_info(:,2),
        spatial_res); %as the cloud fields will computationally pproach from the
        right (arbitrarily selected), and 0deg is a northerly cloud direction,
        rotate the XY 90, so that north now points to the direction of the
        approaching cloud tile. This essentially normalises the cloud angle.

```

B.3 Matrices rotational function

```

1  %% The Matrix Rotation Function
2  % function that circularly rotates xy values within a square matrix
3  % clockwise about its centre point by defined angle.
4  % Created by : Jamie Brght
5
6  % [x2 y2] = matrot(theta,x1,y1,m,n)
7  % Where:
8  % theta    is the angle (degs) to rotate coordinates by
9  % x1       is the list of x coordinates (n index)
10 % y1       is the list of corresponding y coordinates (m index)
11 % m        is the number of rows in the matrix to rotate
12 % n        is the number rof columns in the matrix to rotate
13 % x2       is the rotated n-index of the corresponding x1 input
14 % y2       is the rotated m-index of the corresponding y1 input
15
16 %% The function
17 function [x2, y2] = matrot(theta,x1,y1,mat_size)
18 %convert theta for clockwise rotation, clockwise seems more natural
19 theta = -theta;
20
21 %determine centre point from input m and n values
22 C=round(mat_size/2);
23
24 %convert to cartesian coordinates using new centrepoint
25 dx1=x1-C;
26 dy1=y1-C;
27
28 %calculate the x' and y' values using the rotational matrices
29 dx2=dx1.*cos(degtorad(theta)) - dy1.*sin(degtorad(theta));
30 dy2=dx1.*sin(degtorad(theta)) + dy1.*cos(degtorad(theta));
31
32 x2=dx2+C;
33 y2=dy2+C;
34
35 %if the rotation takes the point outside of the defined m-by-n matrix, set it at
        the nearest point.
36 x2(x2<1)=1;

```

```

37 x2(x2>mat_size)=mat_size;
38 y2(y2<1)=1;
39 y2(y2>mat_size)=mat_size;
40
41 end

```

B.4 Data import

```

1 %% Read In Data
2
3 %%When using NOAA data. Extract in the same format as BADC data.
4 ND=csvread('supportingfiles/NOAA_data.csv');%extract the data
5 ND(isnan(ND)==1)=-9999;%replace NaN values with the -9999 format
6 % The required format: hourrun/okta/blanks 3-4-5/CBH/P/blanks8
   -9-10-11-12-13-14-15-16/blankl/air temp/year/hour of day/day number/6 blanks
   /wind direction/wind speed.
7 data=[ND(:,5),ND(:,11),zeros(length(ND),3),ND(:,10),ND(:,9),zeros(length(ND)
   ,10),ND(:,1),ND(:,6)+1,ND(:,3),zeros(length(ND),7),round(ND(:,8))];
8     pressure=ND(:,9);
9
10 %load up the cloud fields.
11 %Cloud fields are stored according to the temporal res, spatial res, number of
   fields per u per C, and number of clouds. Using the user defined inputs,
   select the cloud fields.
12 for i=0:10
13     name=['V2_clouds_vectors_sample',num2str(i),'__temporal_res_',num2str(
   temporal_res),'__spatial_res_',num2str(spatial_res),'__',num2str(
   num_of_samples),'__num_of_samples__',num2str(max_num_of_clouds),'
   max_num_of_clouds'];
14     x=dlmread(fullfile('supportingfiles',name));
15     assignin('base',['cov',num2str(i)],x);
16 end

```

B.5 Cloud motion and clear-sky index implementation

```

1 %% Cloud motion
2 for Hour =1:hours %loop every hour in the simulation
3     okta=okta_hourly(Hour); %extract the okta for that hour
4     elev_hour=round(min(elevation(Hour*temporal_res-(temporal_res-1):Hour*
   temporal_res))/10)*10; %determine the lowest elevation within that hour (
   from the 1-min-res vector) rounded to nearest 10.
5
6     if Hour>1; For all hours after the very first in the simulation..
7         xyr=xyr2; %the current cloud field is set the previous hour's future
   cloud field, and so only need to load new xyr2
8     else %for the first hour, create the cloud field.
9         %extract the weather data to select appropriate cloud field.
10        u_ref=hourly_u(Hour); %cloudspeed
11        c_ref=hourly_c(Hour); %coverage

```

```

12     n_ref=hourly_n(Hour); %random variate for the hour
13     dir_ref=hourly_dir(Hour); %cloud direction
14
15     %Extract a cloud field
16     switch c_ref
17         case 0
18             xyr=cov0(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
19         case 1
20             xyr=cov1(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
21         case 2
22             xyr=cov2(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
23         case 3
24             xyr=cov3(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
25         case 4
26             xyr=cov4(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
27         case 5
28             xyr=cov5(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
29         case 6
30             xyr=cov6(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
31         case 7
32             xyr=cov7(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
33         case 8
34             xyr=cov8(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
35         case 9
36             xyr=cov9(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
37         case 10
38             xyr=cov10(u_ref*max_num_of_clouds-(max_num_of_clouds-1):u_ref*
max_num_of_clouds,3*n_ref-2:3*n_ref);
39     end
40 end
41
42 %ignore empty cloud slots (there can be 0:max_num_of_clouds in each tile),
extract only present clouds for computational efficiency and accuracy
43 r1=xyr(:,3); r1=r1(r1>0);
44 x1=xyr(:,1); x1=x1(r1>0);
45 y1=xyr(:,2); y1=y1(r1>0);
46 kc1 = zeros(1,numel(r1>0));
47
48 %% Assign kc values to each cloud
49 if elev_hour<0 %if the elevation angle is below 0, the sun is set and it is
unimportant, so skip and save time
50 else
51     CompoundConditionInd= (coeff.elevmin==elev_hour) & (coeff.okta1==okta);
%find the appropriate row reference for the distribution parameters within
the coefficients.csv file using compound logical statement

```

```

52     switch okta
53         case {0,1,2,3} %okta 0:3 all use a BurrIII distribution
54
55             %extract the shape and scale parameters from coefficients.csv
56             using the indicator CompoundConditionalInd
57             alpha_ScaleParameter=coeff.scale(CompoundConditionInd==1);
58             c_ShapeParameter=coeff.shape1(CompoundConditionInd==1);
59             k_ShapeParameter=coeff.shape2(CompoundConditionInd==1);
60
61             %Create the BurrIII distribution PDF using the paramaters above
62             burrIIIPDF=((c_ShapeParameter.*k_ShapeParameter)./
63             alpha_ScaleParameter).*(distribution_range./alpha_ScaleParameter).^(-
64             c_ShapeParameter-1).*+(distribution_range./alpha_ScaleParameter).^(-
65             c_ShapeParameter)).^(-k_ShapeParameter-1);
66             burrIIICDF=cumsum(burrIIIPDF)./100; %make CDF.
67
68             %assign each cloud a kc value from the new distribution
69             for ii=1:numel(r1>0)
70                 kc1(ii)=sum(burrIIICDF<rand)./100;
71             end
72
73         case {4,5,6,7,8,9} %okta 4:9 use a generalised Gamma function
74             %extract the shape and scale parameters from coefficients.csv
75             using the indicator CompoundConditionalInd
76             a_ScaleParameter=coeff.scale(CompoundConditionInd==1);
77             p_ShapeParameter=coeff.shape1(CompoundConditionInd==1);
78             d_ShapeParameter=coeff.shape2(CompoundConditionInd==1);
79             %create the Genralised Gamma distribution PDF using the above
80             parameters.
81             genGammaPDF=(p_ShapeParameter.*distribution_range.^(-
82             d_ShapeParameter-1).*exp(-(distribution_range./a_ScaleParameter).^(-
83             p_ShapeParameter))./(a_ScaleParameter.^d_ShapeParameter.*gamma(
84             d_ShapeParameter./p_ShapeParameter));
85             genGammaCDF=cumsum(genGammaPDF)./100; %make CDF
86
87             %assign each cloud a kc value from the new distribution
88             for ii=1:numel(r1>0)
89                 kc1(ii)=sum(genGammaCDF<rand)./100;
90             end
91         end
92     end
93
94 %% Now perform the same for the future hour.
95 if Hour<(hours-1) %so long as it is not the end of the simulation...
96
97     u_ref_next=hourly_u(Hour+1);
98     c_ref_next=hourly_c(Hour+1);
99     n_ref_next=hourly_n(Hour+1);
100    dir_ref_next=hourly_dir(Hour+1);
101
102    switch c_ref_next
103        case 0
104            xyr2=cov0(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
105            u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
106        case 1

```

```

97         xyr2=cov1(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
98         case 2
99         xyr2=cov2(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
100        case 3
101        xyr2=cov3(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
102        case 4
103        xyr2=cov4(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
104        case 5
105        xyr2=cov5(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
106        case 6
107        xyr2=cov6(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
108        case 7
109        xyr2=cov7(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
110        case 8
111        xyr2=cov8(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
112        case 9
113        xyr2=cov9(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
114        case 10
115        xyr2=cov10(u_ref_next*max_num_of_clouds-(max_num_of_clouds-1):
u_ref_next*max_num_of_clouds,3*n_ref_next-2:3*n_ref_next);
116        end
117
118        r2=xyr2(:,3); r2=r2(r2>0);
119        x2=xyr2(:,1); x2=x2(r2>0);
120        y2=xyr2(:,2); y2=y2(r2>0);
121        kc2 = zeros(1,numel(r2>0));
122
123        %% Apply KC
124        if elev_hour<0
125            else
126                CompoundConditionInd= (coeff.elevmin==elev_hour) & (coeff.okta1==
okta);
127                switch okta %loop through each moment within the temporary, hourly
okta factored vector
128                    case {0,1,2,3} %okta 0:3 all use a BurrIII distribution
129                        alpha_ScaleParameter=coeff.scale(CompoundConditionInd==1);
130                        c_ShapeParameter=coeff.shape1(CompoundConditionInd==1);
131                        k_ShapeParameter=coeff.shape2(CompoundConditionInd==1);
132                        burrIIIPDF= ((c_ShapeParameter.*k_ShapeParameter)./
alpha_ScaleParameter).*(distribution_range./alpha_ScaleParameter).^(-
c_ShapeParameter-1).*1+(distribution_range./alpha_ScaleParameter).^(-
c_ShapeParameter)).^(-k_ShapeParameter-1);
133                        burrIIICDF=cumsum(burrIIIPDF)./100;
134                        for ii=1:numel(r2>0)
135                            kc2(ii)=sum(burrIIICDF<rand)./100;
136                        end

```

```

137
138         case {4,5,6,7,8,9} %okta 4:9 use a generalised Gamma function
139             a_ScaleParameter=coeff.scale(CompoundConditionInd==1);
140             p_ShapeParameter=coeff.shape1(CompoundConditionInd==1);
141             d_ShapeParameter=coeff.shape2(CompoundConditionInd==1);
142             genGammaPDF=(p_ShapeParameter.*distribution_range.^(
d_ShapeParameter-1).*exp(-(distribution_range./a_ScaleParameter).^(
p_ShapeParameter))./(a_ScaleParameter.^d_ShapeParameter.*gamma(
d_ShapeParameter./p_ShapeParameter));
143             genGammaCDF=cumsum(genGammaPDF)./100;
144             for ii=1:numel(r2>0)
145                 kc2(ii)=sum(genGammaCDF<rand)./100;
146             end
147         end
148     end
149
150 end
151
152 %% Move the clouds for this hour
153     %pre allocate for computational efficiency.
154     separation=zeros(num_of_houses,60); %1 row per location
155     house_coverages=zeros(num_of_houses,60);%60 mins=cols
156     house_kcvalues=zeros(num_of_houses,60);
157
158
159     for cloudmovement=1:temporal_res; %move the clouds looping each time step
160
161         for house=1:num_of_houses %take each house at a time
162             %orientate by the cloud direction
163             [XY_rotated(:,1),XY_rotated(:,2)]=matrot(dir_ref,round(house_info
(:,2)),round(house_info(:,1)),spatial_res);
164             dxd=spatial_res-XY_rotated(house,1); %distance: house x location to
far edge of domain edge
165             dX=cloudmovement*(3600/temporal_res)*u_ref_next; %Distance: house
domain and cloud domain overlap
166             dxec2=dxd-dX+u_ref*3600; %distance: house to far edge of cloud field
167             separation(house,cloudmovement)=XY_rotated(house,1);
168
169             % attach together the cloud fields
170             r_C1=[r1;r2]; %combine the radii from both cloud fields
171             x_C1=[x1-3600*u_ref;x2]; %combine coordinates of clouds adjusting x
by the size of the cloud field domain
172             y_C1=[y1;y2]; %combine y coordinates of two fields
173             kc_C1=[kc1';kc2']; %combine kc values of all.
174
175             clouds=length(r_C1(r_C1>0)); %determine the number of clouds within
the cloud domain
176             if clouds>0 %so long as clouds are present...
177                 dx=dxd+x_C1-dX; % distance along x axis from house to cloud
centre
178                 dy=XY_rotated(house,2)-y_C1; %distance in y direction of house
to cloud
179                 d=sqrt(dx.^2+dy.^2); %direct line from house to cloud
180                 house_coverages(house,cloudmovement)=sum(d<r_C1(r_C1>0)); %
record how many clouds are covering the house

```

```

181         kcs1=kc_C1(d<r_C1); %extract appropriate kc values that cover
the house
182         if isempty(kcs1)==0
183             house_kcvalues(house,cloudmovement)=mean(kcs1); take a mean
of the kc values
184             end
185         end
186     end
187 end
188
189 %% write separation house_kcvalues and house_coverages
190 for house=1:num_of_houses
191     dlmwrite(['F:\file_location\' ,num2str(house),'_separation.mat'],
separation(house,:),'-append');
192     dlmwrite(['F:\file_location\' ,num2str(house),'_house_kcvalues.mat'],
house_kcvalues(house,:),'-append');
193     dlmwrite(['F:\file_location\' ,num2str(house),'_house_coverages.mat'
],house_coverages(house,:),'-append');
194 end
195 end

```

B.6 Clear-sky modification and incident irradiance calculations

```

1 %% Kc production
2 %initialisation
3 not_obscurd_min=zeros(1,hours*temporal_res); %pre allocate array
4 day_floor=floor(day); %gain day number
5 % create a vector containing 1minute resolution cloud speed values
6 for i=1:length(u_hourly)
7     u_minutely(i*temporal_res-(temporal_res-1):i*temporal_res)=u_hourly(i);
8 end
9
10 %kc values for clear sky periods based on Demroes kc_index using distribution
curve fitting.
11 for i=1:floor(length(not_obscurd_min)/1440) % loop through each day
12     not_obscurd_min(i*1440-1439:i*1440)=normrnd(0.99,0.08);%
13 % this distribution is N(0.99,0.08) for Camborne,
14 % tLocationScale(mu=1.11566,sigma=0.111785,nu=3.2205) for Lerwick
15 % Burr dist(alpha=0.981286,c=72.3685,k=0.120297) for Hawaii
16 % and N(1.02394,0.04) for San Diego
17
18 end
19
20 not_obscurd_min(zenith_angle>90)=0; %take account of night
21
22 for house=1:num_of_houses % loop through each houses
23     %% load up each house's data
24     house_kcvalues_temp=dlmread(['F:\file_location\' ,num2str(house),'
_house_kcvalues.mat']');

```

```

25     house_coverages_temp=dlmread(['F:\file_location\',num2str(house),'
    _house_coverages.mat']);
26     sep_temp=dlmread(['F:\file_location\',num2str(house),'_separation.mat']);
27
28     %pre allocate
29     separation=zeros(size(u_minutely));
30     house_coverages=zeros(size(u_minutely));
31     house_kcvalues=zeros(size(u_minutely));
32
33     for i=1:hours %extract the house data to working files
34         separation(i*60-59:i*60)=sep_temp(:,i)';
35         house_coverages(i*60-59:i*60)=house_coverages_temp(:,i)';
36         house_kcvalues(i*60-59:i*60)=house_kcvalues_temp(:,i)';
37     end
38
39     %separation in time from edge of domain at each moment.
40     separation=separation./u_minutely;
41     separation=floor(separation./temporal_res);
42
43     %produce indicators of 8 okta periods
44     Ok8_ind=zeros(length(okta_minutely),1); %pre allocate
45     for i=2:length(okta_minutely)-1; %loop through okta
46         if okta_minutely(i)==8 && okta_minutely(i+1)==8 && okta_minutely(i-1)
    ~=8;
47             Ok8_ind(i)=1; %indicate start of period
48         end
49         if okta_minutely(i)==8 && okta_minutely(i-1)==8 && okta_minutely(i+1)
    ~=8;
50             Ok8_ind(i)=2; %indicate end of period
51         end
52     end % in 00010000200 format now.
53
54     Ok8_duration=1; %initialise the marker
55     ok8_cutoff=4; %set the okta 8 cutoff period
56     a_ScaleParameter=coeff.scale(78); %extract the shape and scale parameters
    from coefficients.csv using the indicator CompoundConditionalInd
57     p_ShapeParameter=coeff.shape1(78); % parameters for okta 8 at 10deg elev
58     d_ShapeParameter=coeff.shape2(78); %else zeros propagate into obscured min
59     %create the Genralised Gamma distribution PDF using the above parameters.
60     genGammaPDF=(p_ShapeParameter.*distribution_range.^(d_ShapeParameter-1).*exp
    (-(distribution_range./a_ScaleParameter).^p_ShapeParameter))./(
    a_ScaleParameter.^d_ShapeParameter.*gamma(d_ShapeParameter./p_ShapeParameter
    ));
61     genGammaCDF=cumsum(genGammaPDF)./100; %find the CDF
62
63     for i=1:length(Ok8_ind)-max(max(separation)) %cycle through the indicators
64         if Ok8_ind(i)==1 %if this is the start of an okta 8 period
65             if i+Ok8_duration==numel(Ok8_ind);break; end %if its the end of the
    array, break the for loop, else errors.
66             while Ok8_ind(i+Ok8_duration)~=2; %while the period of okta 8
    continues
67                 Ok8_duration=Ok8_duration+1; %keep tally of the duration
68                 if i==numel(Ok8_ind); break; end %if its the end of the array,
    break the while

```

```

69         if i+Ok8_duration==numel(Ok8_ind);break; end %if its the end of
the array, break the while
70     end
71     end
72     if Ok8_duration>=ok8_cutoff*temporal_res; %if the duration is at least X
hours or more.
73         intervals=ceil(5*(Ok8_duration/temporal_res)*(rand+0.0001)); %random
number of intervals
74         els=ones(1+intervals,1); %make blank array of each hour plus the
start hour
75         for j=1:length(els); %loop through each hour in els.
76             els(j)=ceil(Ok8_duration*((j-1)/intervals))+i; %find the element
row reference to split the moment into sections
77         end
78         Ok8_kcvalues=zeros(size(els));
79         for ii=1:length(els)
80             Ok8_kcvalues(ii)=sum(genGammaCDF<rand)./100; %assign kc to each
interval
81         end
82         xx=els(1):els(end);
83         house_kcvalues(els(1)+ceil(mean(separation(iii,els(1):els(end))))):
els(end)+ceil(mean(separation(iii,els(1):els(end))))=interp1(els,
Ok8_kcvalues,xx,'pchip');%piecewise cubic hermit interpolation technique
between the kc values.
84     end
85     Ok8_duration=1; %reset the duration and loop again.
86 end
87
88 % pre-allocate irradiance arrays
89 house_panel_irradiance=zeros(numel(house_coverages(1,:)),1);
90 house_GHI=zeros(numel(house_coverages(1,:)),1);
91
92 disp(['          ...for house: ',num2str(house)]) %indicate to the user
the progress
93 panel_tilt=house_info(house,5) ; %extract the individual location geography
and geometry
94 panel_orientation=house_info(house,4);
95 panel_has1=house_info(house,3);
96 incident_angle=real(acosd(sind(zenith_angle).*sind(panel_tilt).*cosd(
panel_orientation-azimuth)+cosd(zenith_angle).*cosd(panel_tilt))); % solar
incident angle taking into account panel tilt and azimuth
97
98 sun_obsured=house_coverages;%
99 sun_obsured(sun_obsured>1)=1; %normalise the number of clouds covered by
to create B 1DM
100 kcMinutely= house_kcvalues;%temporary kc 1DM
101
102 %% for long periods of Okta 0. As for 8
103 Ok0_ind=zeros(length(okta_minutely),1);
104 for i=2:length(okta_minutely)-1;
105     if okta_minutely(i)==0 && okta_minutely(i+1)==0 && okta_minutely(i-1)
~=0;
106         Ok0_ind(i)=1;
107     end

```

```

108     if okta_minutely(i)==0 && okta_minutely(i-1)==0 && okta_minutely(i+1)
~=0;
109         Ok0_ind(i)=2;
110         end
111     end
112     Ok0_duration=1;
113     ok0_cutoff=3;
114     intervals=1;
115     for i=1:length(Ok0_ind)-1
116         if Ok0_ind(i)==1 %if this is the start of an okta 0 period
117             if i+Ok0_duration==numel(Ok0_ind);break; end
118             while Ok0_ind(i+Ok0_duration)~=2; %while the period of okta 0
continues
119                 Ok0_duration=Ok0_duration+1; %keep tally of the duration
120                 if i==numel(Ok0_ind); break; end
121                 if i+Ok0_duration==numel(Ok0_ind);break; end
122             end
123         end
124         if Ok0_duration>=ok0_cutoff*temporal_res;
125             %instead of a straight linear between the start and end period (very
much not the case in real life), there are a fixed amount of intervals
between kc values for the whole duration of okta 0.
126             els=zeros(1+intervals,1);
127             for j=2:length(els);
128                 els(j)=ceil(Ok0_duration*((j-1)/intervals));
129             end
130
131             kc_ok0=normrnd(1.02394,0.04,size(els));
132             kc_ok0(kc_ok0>1.3)=normrnd(1,0.2,size(kc_ok0(kc_ok0>1.3)));
133             kc_ok0(kc_ok0<0.9)=normrnd(1,0.2,size(kc_ok0(kc_ok0<0.9)));
134             Ok0_kcvalues=kc_ok0;
135
136             for j=1:length(els)-1;
137                 intlinspace=linspace(Ok0_kcvalues(j),Ok0_kcvalues(j+1),els(j+1)-
els(j)+1);
138             end
139         end
140         Ok0_duration=1;
141     end
142
143     % populate kc minutely with the clear moments.
144     kcMinutely(sun_obscored==0) = not_obscored_min(sun_obscored==0);
145     %remove impossible extremes to the limit
146     kcMinutely(kcMinutely<0.01)=normrnd(1.02394,0.04,1,numel(kcMinutely(
kcMinutely<0.01)));
147
148     %% Add irradiance peaks at moment of cloud shift, as observed in data. Increased
reflected irradiance
149     %     observed in observational data is a peak in irradiance just before and
after a moment of cloud, this is due to increase reflected beam irradiance.
150     %     to attempt to recreate this, fluxes based on a normrnd distribution
are applied to the minute before and after a cloud, limited to a chance
defined as:
151     chance=0.30;% 30% of the time, this will be applied
152     for i=3*(temporal_res/60):length(kcMinutely); %loop through kcMinutely

```

```

153     chance_test=rand; % select a random value to test against the chance
      variable.
154
155     %END OF CLOUD
156     if sun_obscurd(i-1)-sun_obscurd(i)==1; % sun obscurd is 0001111000
      indicating cloud. if i-1 - i = 1. then i must be the end of a clouded period
      , and so...
157         if temporal_res/60>1 %for 1 sec res
158             increase =linspace(normrnd(1+0.05*chance_test,0.01,1),1,
      temporal_res/60);
159             increase2=linspace(normrnd(1+0.025*chance_test,0.01,1),1,
      temporal_res/60);
160         elseif temporal_res/60==1 %for 1 min res
161             increase=normrnd(1+0.05*chance_test,0.01,1);
162             increase2=normrnd(1+0.025*chance_test,0.01,1);
163         end
164         kcMinutely(i-temporal_res/60+1:i)=kcMinutely(i-(temporal_res/60)+1:i
      ).*increase;%... apply a small increase in kc.
165         kcMinutely(i-temporal_res/60+2:i)=kcMinutely(i-temporal_res/60+2:i)*
      increase2; %...apply a smaller increase to the kc value
166
167     %START OF CLOUD for 1-sec resolution
168     elseif sun_obscurd(i-1)-sun_obscurd(i)==-1; %else if i-1 - i=-1 (
      indicating that i start of a clouded period, and so i-1 is the last period
      before cloud
169         if temporal_res/60>1 %for 1 sec
170             increase =linspace(normrnd(1+0.05*chance_test,0.01,1),1,
      temporal_res/60);
171             increase2=linspace(normrnd(1+0.025*chance_test,0.01,1),1,
      temporal_res/60);
172         elseif temporal_res/60==1 % for1-min
173             increase=normrnd(1+0.05*chance_test,0.01,1);% pick a single
      increase
174             increase2=normrnd(1+0.025*chance_test,0.01,1);
175         end
176         kcMinutely(i-temporal_res/60:i)=kcMinutely(i-temporal_res/60:i)*
      increase; %apply CEE
177         kcMinutely(i-temporal_res/60-1:i)=kcMinutely(i-temporal_res/60-1:i)*
      increase2; %apply CEE
178     end
179 end
180
181
182 %% irradiance calculations following Muller and Trentman
183
184 kcMinutely=kcMinutely';
185 global_horizontal = kcMinutely .* global_horizontal_cs;
186 direct_horizontal = zeros(numel(sun_obscurd),1);
187 direct_horizontal(kcMinutely < 1 & kcMinutely > 19/69) =
      direct_horizontal_cs(kcMinutely < 1 & kcMinutely > 19/69) .* (kcMinutely(
      kcMinutely < 1 & kcMinutely > 19/69) - 0.38*(1 - kcMinutely(kcMinutely < 1 &
      kcMinutely > 19/69))).^(2.5);
188 direct_horizontal(kcMinutely>=1) = direct_horizontal_cs(kcMinutely>=1);
189 direct_horizontal(direct_horizontal<0)=0;
190 global_horizontal(global_horizontal<0)=0;

```

```
191     diffuse_horizontal = global_horizontal - direct_horizontal;
192     diffuse_to_global_ratio=diffuse_horizontal./global_horizontal;
193
194     house_GHI=global_horizontal; % store the GHI for each house
195     % Panel irradiance using Klucher model
196     F=1-(diffuse_horizontal./global_horizontal).^2; % modulating factor
197     isotropic=(1+cosd(panel_tilt))/2; % isotropic component - invariant to
198     direct/global ratio
199     horizontal=(1+F.*(sind(panel_tilt)/2).^3); % horizon brightening term
200     circumsol = (1 + F .* (cosd(incident_angle)).^2 .* (sind(zenith_angle)).^3);
201     % circumsolar diffuse irradiance
202     panel_irradiance = diffuse_horizontal.*isotropic.*horizontal.*circumsol +
203     direct_horizontal./cosd(zenith_angle).*cosd(incident_angle);
204
205 end %repeat all of the above for each property
```

B.7 Clear-sky distributions by okta and solar elevation angle

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 equations [4.11](#) and [4.12](#); 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. All auto observation types are provided by [Smith et al. \(2017\)](#), extrapolations were produced for this thesis by the author.

TABLE B.1: k_c distribution parameters for each N at bands of θ_e .

Obs. Type	N	θ_e ($^\circ$)	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

TABLE B.2: k_c distributions continued

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	0.846455350	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
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

Bibliography

- Albert, A. S., 2016. Development of solar photovoltaic inverter with reduced harmonic distortions suitable for Indian sub-continent. *Renewable and Sustainable Energy Reviews* 56, 694 – 704.
- Ångström, A., 1924. Solar and terrestrial radiation. report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society* 50 (210), 121–126.
- APT, 2011. White papers – total harmonic distortion and effects in electrical power systems. Tech. rep., Associated Power Technologies, online.
URL <http://www.aspowertechnologies.com/resources/white-papers.aspx>
- Arias, M. B., Bae, S., 2016. Electric vehicle charging demand forecasting model based on big data technologies. *Applied Energy* 183, 327 – 339.
- Arias-Castro, E., Kleissl, J., Lave, M., 2014. A Poisson model for anisotropic solar ramp rate correlations. *Solar 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.
- Atwater, M., Ball, J., 1978. A numerical solar radiation model based on standard meteorological observations. *Solar Energy* 21 (3), 163 – 170.
- Augsburger, G., Favrat, D., 2013. Modelling of the receiver transient flux distribution due to cloud passages on a solar tower thermal power plant. *Solar Energy* 87, 42 – 52.
- BADC, 2013. British atmospheric data centre - national centre for atmospheric science - natural environment research council.
URL <http://badc.nerc.ac.uk>
- Badescu, V., Gueymard, C. A., Cheval, S., Oprea, C., Baci, M., Dumitrescu, A., Iacobescu, F., Milos, I., Rada, C., 2012. Computing global and diffuse solar hourly irradiation on clear sky. review and testing of 54 models. *Renewable and Sustainable Energy Reviews* 16 (3), 1636 – 1656.
- Bakhshi, R., Sadeh, J., 2016. Voltage positive feedback based active method for islanding detection of photovoltaic system with string inverter using sliding mode controller. *Solar Energy* 137, 564 – 577.
- Balta-Ozkan, N., Yildirim, J., Connor, P. M., 2015. Regional distribution of photovoltaic deployment in the UK and its determinants: A spatial econometric approach. *Energy Economics* 51, 417 – 429.
- Barbieri, F., Rajakaruna, S., Ghosh, A., 2016. Very short-term photovoltaic power forecasting with cloud modeling: A review. *Renewable and Sustainable Energy Reviews*, –In Press.
- Basharin, G. P., Langville, A. N., Naumov, V. A., 2004. The life and work of A.A.Markov. *Linear algebra and its applications* 386, 3–26.
- Bergamasco, L., Asinari, P., 2011. Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Further improvements by ortho-image analysis and application to Turin (Italy). *Solar Energy* 85 (11), 2741 – 2756.
- Bergamasco, L. and Asinari, P. , 2011. Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Application to Piedmont region (Italy). *Solar Energy* 85, 1041–1055.
- Berry, A., Moore, T., Ward, J., Lindsay, S., Proctor, K., 2013. National feeder taxonomy – describing a representative feeder set for Australian networks. Tech. rep., CSIRO.

- Best, M., Brown, A., Clark, P., Hollis, D., Middleton, D., Rooney, G., Thomson, D., Wilson, C., 2008. Small-scale wind energy - technical report. Tech. rep., UK Met Office.
- Beyer, H., Luther, J., Schumacher-Gröhn, J., 1991. Comparison of some standard PV simulation programs with the simulation system INSEL. Solar World Congress. Proc. of the Biennial Congress of the ISES, Denver, Colorado, USA, 19–23 August 1991, Pergamon Press, Oxford, New York. Conference Proceedings, 175 – 180.
- Bhardwaj, S., Sharma, V., Srivastava, S., Sastry, O., Bandyopadhyay, B., Chandel, S., Gupta, J., 2013. Estimation of solar radiation using a combination of hidden Markov model and generalized fuzzy model. *Solar Energy* 93 (0), 43 – 54.
- Bhowmik, A., Maitra, A., Halpin, S. M., Schatz, J. E., 2003. Determination of allowable penetration levels of distributed generation resources based on harmonic limit considerations. *IEEE Transactions on Power Delivery* 18, 619–624.
- Bird, R., Hulstrom, R., 1981a. Review, evaluation, and improvement of direct irradiance models. *Journal of Solar Energy Engineering, Transactions of the ASME* 103 (3), 182–192.
- Bird, R., Hulstrom, R., 1981b. A simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Tech. rep., Solar Energy Research Institute.
- Black, J. N., Bonython, C. W., Prescott, J. A., 1954. Solar radiation and the duration of sunshine. *Quarterly Journal of the Royal Meteorological Society* 80 (344), 231–235.
- Blanco-Muriel, M., Alarcon-Padilla, D., Lopez-Moratalla, T., Lara-Coira, M., 2001. Computing the solar vector. *Solar Energy* 70 (5), 431–441.
- Borges, C. L. T., 2012. An overview of reliability models and methods for distribution systems with renewable energy distributed generation. *Renewable and Sustainable Energy Reviews* 16 (6), 4008 – 4015.
- Bright, J., Smith, C., Taylor, P., Crook, R., 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. *Solar Energy* 115, 229 – 242.
- Bright, J. M., Babacan, O., Kleissl, J., Taylor, P. G., Crook, R., 2017. A synthetic, spatially decorrelating solar irradiance generator and application to a LV grid model with high PV penetration. *Solar Energy* 147, 83–98.
- Brinsfield, R., Yaramanoglu, M., Wheaton, F., 1984. Ground level solar radiation prediction model including cloud cover effects. *Solar Energy* 33, 493–499.
- Bristow, K. L., Campbell, G. S., 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology* 31, 159–166.
- Brunt, D., 1939. *Physical and Dynamical Meteorology*. Cambridge University Press.
- Cai, D. W., Adlakha, S., Low, S. H., Martini, P. D., Chandy, K. M., 2013. Impact of residential PV adoption on retail electricity rates. *Energy Policy* 62 (0), 830 – 843.
- Caldon, R., Coppo, M., Turri, R., 2014. Distributed voltage control strategy for LV networks with inverter-interfaced generators. *Electric Power Systems Research* 107, 85 – 92.
- Calif, R., Soubdhan, T., 2016. On the use of the coefficient of variation to measure spatial and temporal correlation of global solar radiation. *Renewable Energy* 88, 192 – 199.
- Calinoiu, D.-G., Stefu, N., Paulescu, M., Trif-Tordai, G., Mares, O., Paulescu, E., Boata, R., Pop, N., Pacurar, A., 2014. Evaluation of errors made in solar irradiance estimation due to averaging the Angstrom turbidity coefficient. *Atmospheric Research* 150 (0), 69 – 78.
- Carroll, J., 1985. Global transmissivity and diffuse fraction of solar radiation for clear and cloudy skies as measured and as predicted by bulk transmissivity models. *Solar Energy* 35 (2), 105 – 118.
- Chang, T. P., 2009. The sun's apparent position and the optimal tilt angle of a solar collector in the northern hemisphere. *Solar Energy* 83 (8), 1274 – 1284.
- Cheng, H.-Y., Yu, C.-C., 2015. Multi-model solar irradiance prediction based on automatic cloud classification. *Energy* 91, 579 – 587.
- Chicco, G., Schlabbach, J., Spertino, F., 2009. Experimental assessment of the waveform distortion in grid-connected photovoltaic installations. *Solar Energy* 83 (7), 1026 – 1039.
- Choi, J. H., Kim, J. C., 2001. The online voltage control of ULTC transformer for distribution voltage regulation. *International journal of electrical power and energy systems* 23, 91–98.

- 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. *Solar Energy* 85 (11), 2881 – 2893.
- Clement-Nyons, K., Haesen, E., Driesen, J., Feb 2010. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *Power Systems, IEEE Transactions on* 25 (1), 371–380.
- Cooper, P., 1969. The absorption of radiation in solar stills. *Solar Energy* 12, 333–346.
- Cowley, T. P., Ekwue, A. O. (Eds.), 2015. Impact of power factor variations on the voltage stability analysis of a solar photovoltaic integrated power network. 5th Solar Integration Workshop, 19-20 October 2015, Brussels, Belgium.
- Curtright, A. E., Apt, J., 2008. The character of power output from utility-scale photovoltaic systems. *Progress in Photovoltaics: Research and Applications* 16 (3), 241–247.
- Darcovich, K., Kenney, B., MacNeil, D., Armstrong, M., 2015. Control strategies and cycling demands for Li-ion storage batteries in residential micro-cogeneration systems. *Applied Energy* 141 (0), 32 – 41.
- Dartawan, K., Austraia, R., Hui, L., Suehiro, M. (Eds.), 2012. Harmonic issues that limit solar photovoltaic generation on distribution circuits. SOLAR 2012 Conference of the World Renewable Energy Forum (WREF 2012), May 13-17, 2012, Colorado Convention Center, Denver, Colorado, USA.
- de Jong, R., Stewart, D. W., 1993. Estimating global solar radiation from common meteorological observations in western Canada. *Canadian Journal of Plant Science* 73, 509–518.
- de Vos, R., Sawin, J., 2012. READY: Renewable Energy Action on Deployment: policies for accelerated deployment of renewable energy. READY: Renewable Energy Action on Deployment Presenting : the ACTION Star, Six Policy Ingredients for Accelerated Deployment of Renewable Energy : IEA-RETD. Elsevier Science.
- DECC, April 2016. Energy trends section 6: renewables. National statistics, Department of Energy and Climate Change and Department for Business, Energy and Industrial Strategy. URL <https://www.gov.uk/government/statistics/energy-trends-section-6-renewables>
- Deilami, S., Masoum, A., Moses, P., Masoum, M. A. S., Sept 2011. Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile. *Smart Grid, IEEE Transactions on* 2 (3), 456–467.
- Diabaté, L., Remund, J., Wald, L., 2003. Linke turbidity factors for several sites in Africa. *Solar Energy* 75 (2), 111 – 119.
- 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. *Solar Energy* 94 (0), 182 – 194.
- Duffie, J. A., Beckman, W., 2006. *Solar engineering of thermal processes*. John Wiley & Sons.
- Dulski, R., Sosnowski, T., Polakowski, H., 2011. A method for modelling IR images of sky and clouds. *Infrared Physics and Technology* 54 (2), 53 – 60.
- Echavarría, R., Claudio, A., Cotorogea, M., 2007. Analysis, design, and implementation of a fast on-load tap changing regulator. *IEEE Transactions on Power Electronics* 22 (2), 527–534.
- ECMWF, 2014. European centre for medium-range weather forecasts - ERA-Interim. Tech. rep., ECMWF. URL <http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>
- Ehnberg, J., Bollen, M., 2005. Simulation of global solar radiation based on cloud observations. *Solar Energy* 78, 157–162.
- Eichman, J. D., Mueller, F., Tarroja, B., Schell, L. S., Samuelson, S., 2013. Exploration of the integration of renewable resources into California’s electric system using the holistic grid resource integration and deployment (HiGRID) tool. *Energy* 50, 353 – 363.
- Ellis, R., 2014. *Solar Energy: Solar Solutions*. Lulu.com.
- ElNozahy, M., EL-Shatshat, R., Salama, M., 2016. Single-phasing detection and classification in distribution systems with a high penetration of distributed generation. *Electric Power Systems Research* 131, 41 – 48.
- ElNozahy, M., Salama, M., Jan 2014a. A comprehensive study of the impacts of PHEVs on residential distribution networks. *Sustainable Energy, IEEE Transactions on* 5 (1), 332–342.

- ElNozahy, M., Salama, M., 2014b. Studying the feasibility of charging plug-in hybrid electric vehicles using photovoltaic electricity in residential distribution systems. *Electric Power Systems Research* 110 (0), 133 – 143.
- Engerer, N., Mills, F., 2014. Kpv: A clear-sky index for photovoltaics. *Solar Energy* 105, 679 – 693.
- EPRI, 2008. Open distribution system simulator (OpenDSS) version 7.6.4.36. Tech. rep., Electric Power Research Institute.
URL <http://smartgrid.epri.com/SimulationTool.aspx>
- Evans, K. F., Wiscombe, W. J., 2004. An algorithm for generating stochastic cloud fields from radar profile statistics. *Atmospheric Research* 72 (1–4), 263 – 289.
- Fekete, K., Klaic, Z., Majdandzic, L., 2012. Expansion of the residential photovoltaic systems and its harmonic impact on the distribution grid. *Renewable Energy* 43 (0), 140 – 148.
- FERC, 2013. Federal energy regularoty commission - third-party provision of ancillary services; accounting and financial reporting for new electric storage technologies.
URL <http://www.ferc.gov/whats-new/comm-meet/2013/071813/E-22.pdf>
- Fernández-Peruchena, C. M., Gastón, M., 2016. A simple and efficient procedure for increasing the temporal resolution of global horizontal solar irradiance series. *Renewable Energy* 86, 375 – 383.
- Fernández-Peruchena, C. M., Gastón, M., Sánchez, M., García-Barberena, J., Blanco, M., Bernardos, A., 2015. Mus: A multiscale stochastic model for generating plausible meteorological years designed for multiyear solar energy yield simulations. *Solar Energy* 120, 244 – 256.
- Fokaides, P. A., Kylii, A., 2014. Towards grid parity in insular energy systems: The case of photovoltaics (PV) in Cyprus. *Energy Policy* 65 (0), 223 – 228.
- Fthenakis, V. M., Kim, H. C., Alsema, E., 2008. Emissions from photovoltaic life cycles. *Environmental Science and Technology* 42 (6), 2168–2174.
- Gafurov, T., Usaola, J., Prodanovic, M., 2015. Incorporating spatial correlation into stochastic generation of solar radiation data. *Solar Energy* 115, 74 – 84.
- Gao, C., 2013. Voltage control in distribution networks using on-load tap changer transformers. Ph.D. thesis, The department of electronic and electrical engineering.
- Gao, C., Redfern, M., 2010. A review of voltage control techniques of networks with distributed generations using On-Load Tap Changer transformers. *Universities Power Engineering Conference (UPEC), 2010 45th International*, 3–8.
- Gennaro, M. D., Paffumi, E., Scholz, H., Martini, G., 2014. GIS-driven analysis of e-mobility in urban areas: An evaluation of the impact on the electric energy grid. *Applied Energy* 124 (0), 94 – 116.
- Glasbey, C., Graham, R., Hunter, A., 2001. Spatio-temporal variability of solar energy across a region: a statistical modelling approach. *Solar Energy* 70 (4), 373 – 381.
- Goli, P., Shireen, W., 2014. PV powered smart charging station for PHEVs. *Renewable Energy* 66 (0), 280 – 287.
- Gooding, J., Edwards, H., Giesekam, J., Crook, R., 2013. Solar city indicator: A methodology to predict city level PV installed capacity by combining physical capacity and socio-economic factors. *Solar Energy* 95 (0), 325 – 335.
- Grantham, A., Pudney, P., Ward, L., Belusko, M., Boland, J., 2017. Generating synthetic five-minute solar irradiance values from hourly observations. *Solar Energy* 147, 209 – 221.
URL <http://www.sciencedirect.com/science/article/pii/S0038092X1730186X>
- Grena, R., 2008. An algorithm for the computation of the solar position. *Solar Energy* 82 (5), 462 – 470.
- Grena, R., 2012. Five new algorithms for the computation of sun position from 2010 to 2110. *Solar Energy* 86 (5), 1323 – 1337.
- Guerrero-Rodríguez, N., Rey-Boué, A. B., de Lucas, L. C. H., Martínez-Rodrigo, F., 2015. Control and synchronization algorithms for a grid-connected photovoltaic system under harmonic distortions, frequency variations and unbalances. *Renewable Energy* 80, 380 – 395.
- Gueymard, C. A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of

- 18 broadband radiative models. *Solar Energy* 86 (8), 2145 – 2169.
- Gueymard, C. A., Myers, D. R., 2009. Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy* 83 (2), 171 – 185.
- 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. *Solar Energy* 85 (5), 1068 – 1084.
- Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R., Beyer, H., 2003. Solar energy assessment using remote sensing technologies. *Remote Sensing of Environment* 86, 423–432.
- Haque, M. M., Wolfs, P., 2016. A review of high PV penetrations in LV distribution networks: Present status, impacts and mitigation measures. *Renewable and Sustainable Energy Reviews* 62, 1195 – 1208.
- Hargreaves, G. L., Hargreaves, G. H., Riley, J. P., 1985. Irrigation water requirements for senegal river basin. *Journal of Irrigation and Drainage Engineering* 111 (3), 265–275.
- Harvill, J., 2010. Spatio-temporal processes. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (3), 375–382.
- Hernandez, J., la Cruz, J. D., Ogayar, B., 2012. Electrical protection for the grid-interconnection of photovoltaic-distributed generation. *Electric Power Systems Research* 89 (0), 85 – 99.
- Hernandez, J., Ortega, M., la Cruz, J. D., Vera, D., 2011. Guidelines for the technical assessment of harmonic, flicker and unbalance emission limits for PV-distributed generation. *Electric Power Systems Research* 81 (7), 1247 – 1257.
- Higgins, A., McNamara, C., Foliente, G., 2014. Modelling future uptake of solar photo-voltaics and water heaters under different government incentives. *Technological Forecasting and Social Change* 83 (0), 142 – 155.
- Higgins, A. J., Paevere, P., Mcnamara, C., Garner, S., 2013. Modelling electric vehicle and solar PV uptake for Townsville. Tech. rep., CSIRO.
- Hinkelman, L. M., 2013. Differences between along-wind and cross-wind solar irradiance variability on small spatial scales. *Solar Energy* 88, 192 – 203.
- Höök, M., Tang, X., 2013. Depletion of fossil fuels and anthropogenic climate change—a review. *Energy Policy* 52 (0), 797 – 809.
- Hocaoğlu, F. O., 2011. Stochastic approach for daily solar radiation modeling. *Solar Energy* 85 (2), 278 – 287.
- Hoff, T. E., Perez, R., 2010. Quantifying PV power output variability. *Solar Energy* 84 (10), 1782 – 1793.
- Houghton, J., 1996. *Climate Change 1995: The Science of Climate Change: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*. Assessment report. Cambridge University Press.
- IEA, 2014. *Solar (PV and CSP)*. Tech. rep., International Energy Agency.
- IEA, 2016. *Energy and air pollution: World energy outlook special report*. Tech. rep., International Energy Agency.
- IEEE, 2013. *Distribution test feeders*. Tech. rep., IEEE Power and Energy Society, online.
- Inman, R. H., Pedro, H. T., Coimbra, C. F., 2013. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science* 39 (6), 535 – 576.
- IPCC, 2014. *Climate change 2014: Mitigation of climate change*. Tech. rep., Intergovernmental Panel on Climate Change.
- Iqbal, M., 2012. *An Introduction To Solar Radiation*. Elsevier Science.
- ITACA, 2015. *The Earth and its orbit around the Sun*. Tech. rep., ITACA.
URL <http://www.itacanet.org/>
- Jaccard, M., 2006. *Sustainable Fossil Fuels: The Unusual Suspect in the Quest for Clean and Enduring Energy*. Cambridge University Press.
- Jakubiec, J. A., Reinhart, C. F., 2013. A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly daysim simulations. *Solar Energy* 93 (0), 127 – 143.

- 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.
- Jana, J., Saha, H., Bhattacharya, K. D., 2016. A review of inverter topologies for single-phase grid-connected photovoltaic systems. *Renewable and Sustainable Energy Reviews*, –In press.
- Jenkins, D., 2013. *Renewable Energy Systems: The Earthscan Expert Guide to Renewable Energy Technologies for Home and Business*. Earthscan Expert. Taylor and Francis.
- Johansson, T., Burnham, L., 1993. *Renewable Energy: Sources for Fuels and Electricity*. Island Press.
- Kalogirou, S. A., 2014. *Solar Energy Engineering - Processes and Systems (2nd Edition)*. Elsevier.
- Kaplani, E., Kaplanis, S., 2012. A stochastic simulation model for reliable PV system sizing providing for solar radiation fluctuations. *Applied Energy* 97, 970 – 981, *Energy Solutions for a Sustainable World - Proceedings of the Third International Conference on Applied Energy*, May 16-18, 2011 - Perugia, Italy .
- Karimi, H., Karimi-Ghartemani, M., Iravani, M., 2004. Estimation of Frequency and its Rate of Change for Applications in Power Systems. *IEEE Transactions on Power Delivery* 19 (2), 472–480.
URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1278398&navigation=1>
- Karteris, M., Slini, T., Papadopoulos, A., 2013. Urban solar energy potential in Greece: A statistical calculation model of suitable built roof areas for photovoltaics. *Energy and Buildings* 62 (0), 459 – 468.
- Kasten, F., 1996. The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness. *Solar Energy* 56 (3), 239–244.
- Kasten, F., Czeplak, G., 1980. Solar and terrestrial radiation dependent on the amount and type of cloud. *Solar Energy*, 177–189.
- Kasten, F., Young, A., 1989. Revised optical air mass tables and approximation formula. *Applied Optics* 28 (22), 4735–4738.
- Kavali, J., Mittal, A., 2016. Analysis of various control schemes for minimal total harmonic distortion in cascaded H-bridge multilevel inverter. *Journal of Electrical Systems and Information Technology*, –In press.
- Khalili, M., Brissette, F., Leconte, R., 2009. Stochastic multi-site generation of daily weather data. *Stochastic Environmental Research and Risk Assessment* 23 (6), 837–849.
- Kleissl, J., 2013. *Solar Energy Forecasting and Resource Assessment*. Elsevier.
- Klucher, T., 1979. Evaluation of models to predict insolation on tilted surfaces. *Solar Energy* 23, 111–114.
- Kopp, G., Lean, J. L., 2011. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters* 38 (1).
- 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 Procedia* 49, 2351–2361.
- Kwan, C. L., 2012. Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar PV arrays across the united states. *Energy Policy* 47, 332 – 344.
- Lamm, L., 1981. A new analytic expression for the equation of time. *Solar Energy* 26 (5), 465 –.
- 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. *Solar Energy* 122, 517 – 528.
- Lave, M., Kleissl, J., 2010. Solar variability of four sites across the state of Colorado. *Renewable Energy* 35 (12), 2867 – 2873.
- Lave, M., Kleissl, J., 2013. Cloud speed impact on solar variability scaling – application to the wavelet variability model. *Solar Energy* 91 (0), 11 – 21.
- Lave, M., Kleissl, J., Arias-Castro, E., 2012. High-frequency irradiance fluctuations and geographic smoothing. *Solar Energy* 86 (8), 2190 – 2199, *progress in Solar Energy* 3.

- Lave, M., Kleissl, J., Stein, J. S., April 2013. A wavelet-based variability model (WVM) for solar PV power plants. *IEEE Transactions on Sustainable 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. *Solar Energy* 118, 327 – 337.
- Leahy, L. V., Wood, R., Charlson, R. J., Hostetler, C. A., Rogers, R. R., Vaughan, M. A., Winker, D. M., 2012. On the nature and extent of optically thin marine low clouds. *Journal of Geophysical Research: Atmospheres* 117 (D22).
- Li, C., Cao, C., Cao, Y., Kuang, Y., Zeng, L., Fang, B., 2014. A review of islanding detection methods for microgrid. *Renewable and Sustainable Energy Reviews* 35 (0), 211 – 220.
- Li, Y., Vilathgamuwa, D. M., Loh, P. C., 2004. Design, analysis, and real-time testing of a controller for multibus microgrid system. *IEEE Transactions on Power Electronics* 19 (5), 1195–1204.
- Lingfors, D., Widén, J., 2016. Development and validation of a wide-area model of hourly aggregate solar power generation. *Energy* 102, 559 – 566.
- Liu, B. Y. H., Jordan, R. C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Solar Energy* 4, 1–19.
- Lokhande, N. M., Kulkarni, S. R., Ghone, V. S., Kikale, V. R., 2017. Design of Triac Based on Load Tap Changer for Constant Load Voltage 10 (8), 313–320.
- 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. *Solar Energy* 97, 58 – 66.
- Lukač, N., Seme, S., Žlaus, D., Štumberger, G., Žalik, B., 2014. Buildings roofs photovoltaic potential assessment based on LiDAR (light detection and ranging) data. *Energy* 66 (0), 598 – 609.
- Luo, Y., Trishchenko, A. P., Khlopenkov, K. V., 2008. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sensing of Environment* 112 (12), 4167 – 4185.
- Marcos, J., de la Parra, Í., García, M., Marroyo, L., 2016. Simulating the variability of dispersed large PV plants. *Progress in Photovoltaics: Research and Applications* 24 (5), 680–691.
- 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. *Progress in Photovoltaics: Research and Applications* 19 (2), 218–227.
- Marcos, J., Marroyo, L., Lorenzo, E., García, M., 2012. Smoothing of PV power fluctuations by geographical dispersion. *Progress in Photovoltaics: Research and Applications* 20 (2), 226–237.
- Markov, A., 1906. Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch.* 15, 135–156.
- Masseran, N., 2015. Markov chain model for the stochastic behaviors of wind-direction data. *Energy Conversion and Management* 92 (0), 266 – 274.
- Mathiesen, B. V., Lund, H., Karlsson, K., 2011. 100% Renewable energy systems, climate mitigation and economic growth. The 5th Dubrovnik Conference on Sustainable Development of Energy, Water and Environment Systems, held in Dubrovnik September/October 2009. *Applied Energy* 88 (2), 488 – 501.
- Mathiesen, P., Collier, C., Kleissl, J., 2013. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. *Solar Energy* 92, 47 – 61.
- MathWorks, 2016. 2-D correlation coefficients - Matlab corr2.
- MathWorks UK, 2016. Two-sample Kolmogorov-Smirnov test - Matlab kstest2. Accessed on: 04/04/2016.
- Matlab, 2012. R2014a documentation, signal processing, multirate signal processing, resample. URL <http://uk.mathworks.com/help/signal/ug/resampling.html>
- Matlab, 2015. version 8.6.0.267246 R2015b. The MathWorks Inc., Natick, Massachusetts.
- Matszuko, D., 2012. Influence of the extent and genera of cloud cover on solar radiation intensity. *International Journal of Climatology* 32, 2403–2414.

- Mayer, B., Kylling, A., 2005. Technical note: The libRadtran software package for radiative transfer calculations – description and examples of use. *Atmospheric Chemistry and Physics* 5, 1855–1877.
- MetOffice, 2015. Learning. Tech. rep., UK Met Office.
URL <http://www.metoffice.gov.uk/learning/>
- Michalsky, J. J., 1988. The astronomical Almanac’s algorithm for approximate solar position (1950–2050). *Solar Energy* 40 (3), 227 – 235.
- MIDAS, 2015. UK Met Office Integrated Data Archive System (MIDAS) land and marine surface stations data (1853-current).
URL <http://badc.nerc.ac.uk>
- Mokhtari, G., Nourbakhsh, G., Zare, F., Ghosh, A., 2014. Improving the penetration level of PVs using DC link for residential buildings. *Energy and Buildings* 72 (0), 80 – 86.
- Moler, C., 2010. Numerical Computing with MATLAB: Revised Reprint. Society for Industrial and Applied Mathematics.
- Monfared, M., Golestan, S., 2012. Control strategies for single-phase grid integration of small-scale renewable energy sources: A review. *Renewable and Sustainable Energy Reviews* 16 (7), 4982 – 4993.
- 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. *Solar Energy* 132, 430 – 441.
- Morf, H., 1998. The Stochastic two-state solar irradiance model STSIM. *Solar Energy* 62 (2), 101 – 112.
- Morf, H., 2011. The stochastic two-state cloud cover model STSCCM. *Solar Energy* 85 (5), 985 – 999.
- Morf, H., 2013. A stochastic solar irradiance model adjusted on the Angstrom–Prescott regression. *Solar Energy* 87 (0), 1 – 21.
- Mouheb, M., Hamidat, A., Loukarfi, L., 2012. Impact of PV compensation in improving the voltage drop in electrical networks LV. *Energy Procedia* 18 (0), 751 – 761.
- Mu, Y., Wu, J., Jenkins, N., Jia, H., Wang, C., 2014. A spatial–temporal model for grid impact analysis of plug-in electric vehicles. *Applied Energy* 114 (0), 456 – 465.
- Müller, R., Trentmann, J., 2010. Algorithm theoretical baseline document direct irradiance at surface. Tech. rep., EUMETSAT Satellite Application Facility on Climate Monitoring.
- Muneer, T., Gul, M., Kambezidis, H., 1998. Evaluation of an all-sky meteorological radiation model against long-term measured hourly data. *Energy Conversion and Management* 19, 303–317.
- Munkhammar, J., Widén, J., Hinkelman, L. M., 2017. A copula method for simulating correlated instantaneous solar irradiance in spatial networks. *Solar Energy* 143, 10 – 21.
- Ngoko, B., Sugihara, H., Funaki, T., 2014. Synthetic generation of high temporal resolution solar radiation data using markov models. *Solar 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. *Solar Energy* 132, 221 – 235.
- Nielsen, L. B., Prahm, L. P., Berkowicz, R., Conradsen, K., 1981. Net incoming radiation estimated from hourly global radiation and/or cloud observations. *Journal of Climatology* 1 (3), 255–272.
- NOAA, 2016a. Quality Controlled Local Climatological Data (QCLCD). National Centres for Environmental Information. Accessed on: 31/01/2016.
URL <http://www.ncdc.noaa.gov/qclcd/>
- NOAA, 2016b. WMO 2700 code conversion table. Tech. rep., National Oceanic and Atmospheric Administration.
URL www.nodc.noaa.gov/archive/arc0021/0000907/1.1/data/0-data/HTML/WMO-CODE/WMO2700.HTM
- Nykamp, S., Molderink, A., Hurink, J. L., Smit, G. J. (Eds.), 2012. Statistics for PV, wind

- and biomass generators and their impact on distribution grid planning. The 24th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy, ECOS. Vol. 45.
- O'Hare, G., Sweeney, J., Wilby, R., 2005. *Weather, Climate, and Climate Change: Human Perspectives*. Pearson Prentice Hall.
- Olseth, J. A., Skartveit, A., Zou, H., 1995. Spatially continuous mapping of solar resources in a complex high latitude topography. *Solar Energy* 55 (6), 475 – 485.
- Omran, W. A., Kazerani, M., Salama, M. M. A., 2011. Investigation of methods for reduction of power fluctuations generated from large grid-connected photovoltaic systems. *IEEE Transactions on Energy Conversion* 26 (1), 318–327.
- OpenEI, 2014. Commercial and residential hourly load profiles for all TMY3 locations in the United States. Tech. rep., Office of Energy Efficiency Renewable Energy.
URL <http://en.openei.org>
- Oppenheim, A. V., Schafer, R. W., Buck, J. R., 1999. *Discrete-time Signal Processing* (2nd Ed.). Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Otani, K., Minowa, J., Kurokawa, K., 1997. Study on areal solar irradiance for analyzing areally-totalized PV systems. *Solar Energy Materials and Solar Cells* 47 (1–4), 281 – 288.
- Paatero, J. V., Lund, P. D., 2007. Effects of large-scale photovoltaic power integration on electricity distribution networks. *Renewable Energy* 32 (2), 216 – 234.
- Paevere, P., Higgins, A., Ren, Z., Horn, M., Grozev, G., McNamara, C., 2014. Spatio-temporal modelling of electric vehicle charging demand and impacts on peak household electrical load. *Sustainability Science* 9 (1), 61–76.
- Page, L., Brin, S., Motwani, R., Winograd, T., November 1999. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab.
- Pai, M. A., 2016. *Power system dynamics and stability* / Peter W . (April).
- Palizban, O., Kauhaniemi, K., Guerrero, J. M., 2014. Microgrids in active network management—part I: Hierarchical control, energy storage, virtual power plants, and market participation. *Renewable and Sustainable Energy Reviews* 36, 428–439.
- Papaioannou, I. T., Purvins, A., 2014. A methodology to calculate maximum generation capacity in low voltage distribution feeders. *International Journal of Electrical Power and Energy Systems* 57 (0), 141 – 147.
- Parkinson, G., June 2015. Ergon imposes new limit on rooftop solar system. Tech. rep., Renew Economy.
URL reneweconomy.com.au/ergon-imposes-new-limit-on-rooftop-solar-systems-66926/
- Passey, R., Spooner, T., MacGill, I., Watt, M., Syngellakis, K., 2011. The potential impacts of grid-connected distributed generation and how to address them: A review of technical and non-technical factors. *Energy Policy* 39 (10), 6280 – 6290.
- Patel, N. R., Lokhande, M. M., Jamnani, J. G., 2014. Solid-State On Load Tap-Changer for Transformer Using Microcontroller (January), 101–104.
- Patrick, J. D., Harvill, J. L., Hansen, C. W., 2016. A semiparametric spatio-temporal model for solar irradiance data. *Renewable Energy* 87, Part 1, 15 – 30.
- Patsalides, M., Efthymiou, V., Stavrou, A., Georghiou, G. E., 2016. A generic transient PV system model for power quality studies. *Renewable Energy* 89, 526 – 542.
- 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. *Solar Energy* 136, 675 – 681.
- Peng, Z., Yu, D., Huang, D., Heiser, J., Kalb, P., 2016. A hybrid approach to estimate the complex motions of clouds in sky images. *Solar Energy* 138, 10 – 25.
- Perez, M. J., Fthenakis, V. M., 2015. On the spatial decorrelation of stochastic solar resource variability at long timescales. *Solar Energy* 117, 46 – 58.
- Perez, R., Ineichen, P., Seals, R., Zelenka, A., 1990. Making full use of the clearness index for parameterizing hourly insolation conditions. *Solar Energy* 45 (2), 111 – 114.
- Perez, R., Kivalov, S., Schlemmer, J., Jr., K. H., Hoff, T., 2011. Parameterization of site-specific short-term irradiance variability. *Solar Energy* 85 (7), 1343 – 1353.

- Perez, R., Kivalov, S., Schlemmer, J., Jr., K. H., Hoff, T. E., 2012. Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. *Solar Energy* 86 (8), 2170 – 2176, progress in Solar Energy 3.
- Perez, R., Seals, R., Michalsky, J., 1993. All-weather model for sky luminance distribution—preliminary configuration and validation. *Solar Energy* 50 (3), 235–245.
- Pitman, C. L., Vant, L. L. (Eds.), 1978. Errors in locating the Sun and their effect on solar intensity predictions. In: Meeting of American Section of the International Solar Energy Society, Denver, 28 Aug 1978.
- Pittock, A., 2013. *Climate Change: The Science, Impacts and Solutions*. Taylor and Francis.
- Polo, J., Zarzalejo, L., Marchante, R., Navarro, A., 2011. A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. *Solar Energy* 85 (5), 1164 – 1170.
- Pourbabak, H., Kazemi, A., 2014. A new technique for islanding detection using voltage phase angle of inverter-based DGs. *International Journal of Electrical Power and Energy Systems* 57 (0), 198 – 205.
- Prescott, J. A., 1940. Evaporation from water surface in relation to solar radiation. *Transactions of the Royal Society of South Australia* 64, 114–118.
- Pressel, K. G., Collins, W. D., 2012. First-order structure function analysis of statistical scale invariance in the airs-observed water vapor field. *Climate* 25, 5538–5555.
- Pérez-Burgos, A., Román, R., Bilbao, J., de Miguel, A., Oteiza, P., 2015. Reconstruction of long-term direct solar irradiance data series using a model based on the cloud modification factor. *Renewable Energy* 77, 115 – 124.
- PVPerformance Modeling Collaborative, 2016. An industry and national laboratory collaborative to improve photovoltaic performance modeling.
URL <https://pvpmc.sandia.gov>
- Rao, P., Crow, M., Yang, Z., 2000. STATCOM control for power system voltage control applications. *IEEE Transactions on Power Delivery* 15 (4), 1311–1317.
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=891520>
- Ray, J., Engelhardt, F., 1992. Produced water: technological, environmental issues and solutions; proceedings of the 1992 International Produced Water Symposium, held February 4 - 7, 1992, in SanDiego, California. Environmental science research. Plenum Press.
- Reda, I., Andreas, A., 2004. Solar position algorithm for solar radiation applications. *Solar Energy* 76 (5), 577 – 589.
- Reikard, G., 2009. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy* 83 (3), 342 – 349.
- REN, 2016. Renewables 2016 global status report. Tech. rep., Renewable Energy Policy Network for the 21st Century.
- Reno, M. J., Hansen, C. W., Stein, J. S., March 2012. Global horizontal irradiance clear sky models: Implementation and analysis. Tech. rep., Sandia National Laboratories.
- Richardson, J., 2016. How electricity is made and transmitted. Accessed on: 05/10/2016.
URL <http://www.nationalgrideducation.com/resources/>
- Salim, N. B., Tsuji, T., Oyama, T., Koide, A., Yamashita, S., Uchida, K. (Eds.), 2015. A reactive power control strategy for preventing voltage collapse for Malaysian power network with multiple photovoltaic plants. 5th Solar Integration Workshop, 19-20 October 2015, Brussels, Belgium.
- Samadi, A., Shayesteh, E., Eriksson, R., Rawn, B., Söder, L., 2014. Multi-objective coordinated droop-based voltage regulation in distribution grids with PV systems. *Renewable Energy* 71, 315 – 323.
- Sánchez, V., Echavarría, R., Cotorogea, M., Claudio, A., 2000. Design and implementation of a fast on-load tap changing regulator using soft-switching commutation techniques. *PESC Record - IEEE Annual Power Electronics Specialists Conference* 1 (c), 488–493.
- Sayeef, S., Heslop, S., Cornforth, D., Moore, T., Percy, S., Ward, J. K., Berry, A., Rowe, D., 2012. Solar intermittency: Australia’s clean energy challenge – characterising the effect of high penetration solar intermittency on Australian electricity networks. Tech. rep., CSIRO.

- Schneider, K., Chen, Y., Chassin, D., Pratt, R., Engel, D., , Thompson, S., 2008. Modern grid initiative – distribution taxonomy final report. Tech. rep., PNNL.
- Scott, C. E., Rap, A., Spracklen, D. V., Forster, P. M., Carslaw, K. S., Mann, G. W., Pringle, K. J., Kivekäs, N., Kulmala, M., Lihavainen, H., Tunved, P., 2014. The direct and indirect radiative effects of biogenic secondary organic aerosol. *Atmospheric Chemistry and Physics* 14 (1), 447–470.
- Seneta, E., 1966. Markov and the birth of chain dependence theory. *International Statistical Review* 64, 255–263.
- 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.
- Sherwani, A., Usmani, J., Varun, 2010. Life cycle assessment of solar PV based electricity generation systems: A review. *Renewable and Sustainable Energy Reviews* 14 (1), 540 – 544.
- Shivashankar, S., Mekhilef, S., Mokhlis, H., Karimi, M., 2016. Mitigating methods of power fluctuation of photovoltaic (Pv) sources – a review. *Renewable and Sustainable Energy Reviews* 59, 1170 – 1184.
- Skartveit, A., Olseth, J., 1992. The probability density and autocorrelation of short-term global and beam irradiance. *Solar 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. *Solar Energy* 144, 10 – 21.
- Snape, J. R., 2016. Spatial and temporal characteristics of PV adoption in the UK and their implications for the smart grid. *Energies*, 210.
- Snasel, V., Platos, J., El-Qawasmeh, E., 2011. Digital Information Processing and Communications: International Conference, ICDIPC 2011, Ostrava, Czech Republic, July 7-9, 2011. Proceedings. *Communications in Computer and Information Science*. Springer Berlin Heidelberg.
- SoDa, 2016. SoDa - Linke turbidity factor. Tech. rep., SoDa.
URL www.soda-pro.com
- Solangi, K., Islam, M., Saidur, R., Rahim, N., Fayaz, H., 2011. A review on global solar energy policy. *Renewable and Sustainable Energy Reviews* 15 (4), 2149 – 2163.
- Spencer, J. W., 1971. Fourier series representation of the position of the sun. *Search* 2 (5), 172+.
- Stamnes, K., Tsay, S.-C., Wiscombe, W., Laszlo, I., 2000. DISORT, a general-purpose Fortran program for discrete-ordinate-method radiative transfer in scattering and emitting layered media: Documentation of methodology. Tech. rep., Dept. of Physics and Engineering Physics, Stevens Institute of Technology, Hoboken, NJ, USA.
- Stein, J. S., Hansen, C. W., Reno, M. J., 2012. The variability index: A new and novel metric for quantifying irradiance and PV output variability. In: *World Renewable Energy Forum*, Denver, CO. pp. 1–7.
- Stull, R., 1988. *An Introduction to Boundary Layer Meteorology*. Atmospheric and Oceanographic Sciences Library. Springer.
- Suehrcke, H., McCormick, P., 1989. Solar radiation utilizability. *Solar Energy* 43 (6), 339 – 345.
- Sun, Y. W., Hof, A., Wang, R., Liu, J., jie Lin, Y., wei Yang, D., 2013. GIS-based approach for potential analysis of solar PV generation at the regional scale: A case study of Fujian Province. *Energy Policy* 58 (0), 248 – 259.
- Supit, I., van Kappel, R. R., 1998. A simple method to estimate global radiation. *Solar Energy* 63, 147–160.
- Swift, L.W., J., 1976. Algorithm for solar radiation on mountain slopes. *Water Resources Research* 12 (1), 108–112.
- Team, B., 2008. *Assessment of Climate Change for the Baltic Sea Basin*. Regional Climate Studies. Springer.
- Tesla, 2016. Tesla powerwall.
URL https://www.tesla.com/en_GB/powerwall

- Thomson, M., April 2000. Automatic voltage control relays and embedded generation. I. Power Engineering Journal 14 (2), 71–76.
- Tiwari, G., Mishra, R., 2011. Advanced Renewable Energy Sources. Royal Society of Chemistry.
- Tonkoski, R., Lopes, L. A., 2011. Impact of active power curtailment on overvoltage prevention and energy production of PV inverters connected to low voltage residential feeders. Renewable Energy 36 (12), 3566 – 3574.
- Torres, J., de Blas, M., Torres, L., García, A., de Francisco, A., 2014. Synthetic generation of standard sky types series using Markov transition matrices. Renewable Energy 62, 731 – 736.
- Torrìti, J., 2014. A review of time use models of residential electricity demand. Renewable and Sustainable Energy Reviews 37 (0), 265 – 272.
- Tzoumanikas, P., Nikitidou, E., Bais, A., Kazantzidis, A., 2016. The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system. Renewable Energy 95, 314 – 322.
- UKMO, 1997. UK Meteorological Office - Source Book to the Forecasters Reference Book. Meteorological office.
- UKMO, 2010. Uk met office - observations: National meteorological library and archive fact sheet 17 - weather observations over land. Tech. rep., UK Met Office.
- USA DoE, 2000. Scenarios for a Clean Energy Future Interlaboratory Working Group on Energy-Efficient and Clean Energy Technologies and United States. Department of Energy. Office of Energy Efficiency and Renewable Energy. NREL/TP. Oak Ridge National Laboratory.
- Vallee, F., Klonari, V., Lisiecki, T., Durieux, O., Moiny, F., Lobry, J., 2013. Development of a probabilistic tool using monte carlo simulation and smart meters measurements for the long term analysis of low voltage distribution grids with photovoltaic generation. International Journal of Electrical Power and Energy Systems 53 (0), 468 – 477.
- Vallée, F., Moutier, F., Klonari, V., de Grève, Z., Lobry, J., Lecron, F. (Eds.), 2015. Clustering of photovoltaic generation for the consideration of time changing geographical correlation in probabilistic analysis of low voltage distribution systems. 5th Solar Integration Workshop, 19-20 October 2015, Brussels, Belgium.
- Vignola, F. E., McMahan, A. C., Grover, C. N., 2013. Bankable solar-radiation datasets. Elsevier.
- Vindel, J., Polo, J., 2014a. Intermittency and variability of daily solar irradiation. Atmospheric Research 143 (0), 313 – 327.
- Vindel, J., Polo, J., 2014b. Markov processes and Zipf’s law in daily solar irradiation at Earth’s surface. Journal of Atmospheric and Solar-Terrestrial Physics 107 (0), 42 – 47.
- Wagner, L., Tröster, E., Krämer, J., Betz, B. (Eds.), 2015. Impact of 100% renewables strategy on a distribution grid in Rhineland-Palatinate. 5th Solar Integration Workshop, 19-20 October 2015, Brussels, Belgium.
- Wakileh, G., 2001. Power Systems Harmonics: Fundamentals, Analysis and Filter Design. Engineering online library. Springer.
- Walbank, M., Goodhand, C., 2014. Interview with system planning managers of Northern Powergrid on 06/10/2014. Personal communication.
- Walraven, R., 1978. Calculating the position of the sun. Solar Energy 20 (5), 393 – 397.
- Wang, G., Kurtz, B., Kleissl, J., 2016. Cloud base height from sky imager and cloud speed sensor. Solar Energy 131, 208 – 221.
- Wauben, W., Baltink, H. K., de Haij, M., Maat, N., The, H. (Eds.), 2006. Status, evaluation and new developments of the automated cloud observations in the Netherlands. In: World Meteorological Organization (ed), TECO-2006 – WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation. Geneva, Switzerland.
- Westacott, P., Candelise, C., 2016. A novel geographical information systems framework to characterize photovoltaic deployment in the UK: initial evidence. Energies, 26.
- Widen, J., 2015. A model of spatially integrated solar irradiance variability based on logarithmic station-pair correlations. Solar Energy 122, 1409 – 1424.

- Widén, J., Wäckelgård, E., Paatero, J., Lund, P., 2010. Impacts of different data averaging times on statistical analysis of distributed domestic photovoltaic systems. *Solar Energy* 84 (3), 492 – 500.
- Widen, J., Wäckelgård, E., Paatero, J., Lund, P., 2010. Impacts of distributed photovoltaics on network voltages: Stochastic simulations of three Swedish low-voltage distribution grids. *Electric Power Systems Research* 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. *Solar Energy* 70 (6), 513 – 518.
- Wong, J., Lim, Y. S., Tang, J. H., Morris, E., 2014. Grid-connected photovoltaic system in malaysia: A review on voltage issues. *Renewable and Sustainable Energy Reviews* 29 (0), 535 – 545.
- Wood, R., Field, P. R., 2011. The distribution of cloud horizontal sizes. *Journal of Climate* 24, 4800–4816.
- Wörner, H., 1967. Zur frage der automatisierbarkeit der bewölkungsangaben durch verwendung von strahlungsgrößen. *Abh. Met. Dienst DDR* 82.
- WRMC-BSRN, 2014. World radiation monitoring center - baseline surface radiation network-PANGEA - data publisher for earth and environmental science.
URL <http://bsrn.awi.de/>
- Xavier, L. S., Cupertino, A. F., de Resende, J. T., Mendes, V. F., Pereira, H. A., 2017. Adaptive current control strategy for harmonic compensation in single-phase solar inverters. *Electric Power Systems Research* 142, 84 – 95.
- 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. *Electric Power Systems Research* 111 (0), 185 – 193.
- Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., Kleissl, J., 2014. Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. *Solar Energy* 103, 502 – 524.
- Yang, H., Li, Y., Lu, L., Qi, R., 2011. First order multivariate Markov chain model for generating annual weather data for Hong Kong. *Energy and Buildings* 43 (9), 2371 – 2377.
- Zhao, S., Liu, Y., Li, W., Song, S., Sept 2010. Impacts of grid-connected photovoltaic generation on distribution networks. In: *Electricity Distribution (CICED), 2010 China International Conference on*. pp. 1–4.
- Zhu, T. X., Tso, S. K., Lo, K. L., 2000. An investigation into the OLTC effects on voltage collapse. *IEEE Transactions on Power Systems* 15 (2), 515–521.