Appendix IV

Spatial and Temporal Patterns of Solar Absorption by Clouds in Australia as Revealed by Exploratory Factor Analysis

Kevin K. W. Cheung

Department of Environment and Geography, Macquarie University, Sydney, Australia

Suree Chooprateep

Department of Statistics, Chiang Mai University, Chiang Mai, Thailand

Jun Ma

Department of Statistics, Macquarie University, Sydney, Australia

Prepared for Solar Energy

First submission January 2014

First Revision June 2014

Second Revision September 2014

Corresponding author address: Kevin Cheung, Macquarie University, North Ryde, NSW 2109, Australia. Email: kevin.cheung@mq.edu.au

Abstract

Daily solar irradiance data collected from 144 surface stations across Australia over a 23 year period is analyzed to determine patterns in cloud absorption. Two statistical models are developed and fitted to the estimated cloud absorption at each station. The first model, which is equivalent to a non-parametric mean model, is a regression that uses variables of year and pentad (5-day) for each surface station. The second model clusters the surface stations using a factor analysis (FA) on the locational variable, while retaining the year and pentad (or alternatively, month) variables. This identifies seven factors, and a regression model is then developed for each of these, with year and month (or pentad) as the variables.

The first model is able to fit into the radiation observations well especially those in the tropical region in which the seasonal variation is clear. The FA for establishing the second model determines seven dominant factors (with the eighth unimportant one ignored). These seven factors have distinct geographic distributions in each of which a unique factor is the dominant component. Examination of the seasonal cycle and inter annual variation of estimated cloud absorption based on this model indicates three major regional groups in term of the estimated annual cloud cover. The southeast region forms a group with the largest cloud cover percentage, the group consisting of central east, northeast and central west Australia has medium percentage, and that with central, central north and northwest Australia as members has the lowest percentage. These three geographic groups have distinct distribution compared with the climate regimes in Australia, and the relationship with the origins of cloud systems is analyzed. The implications of these statistical models to the short-

and long-term estimation of solar energy availability based on ground observations and their utilities are discussed.

Keywords: Solar irradiance, cloud absorption, regression, factor analysis, rainfall trend, climate variability

1. Introduction

Solar energy is supporting all living organisms on Earth and at the same time the fundamental driving force of the climate system including the atmospheric and oceanic circulation. In this sense, solar radiation is the basis of renewable energy ranging from the direct application of it by photovoltaic power production to wind energy conversion. These techniques of renewable energy are particularly feasible in countries such as Australia with the large continental area to receive solar radiation and the long coast line for installing wind plants. Knowledge on the spatial and temporal variability of solar irradiance is thus critical for designing the appropriate energy-conversion equipment and infrastructure, and generating efficient management plan of the facilities. In particular, the spatial distribution of solar energy determines the cost of implementation of the power facilities and its intermittency in time affects the sustainability and cost-effectiveness of such facilities as well (Rayl et al. 2013).

The intensity of solar radiation, which is represented by the solar constant, is quite stable and has not changed substantially over many years of observations. When entering the atmosphere of the Earth, part of the radiation is absorbed by oxygen, ozone and nitrogen (Peixoto and Oort 1992). Once the solar radiation enters the troposphere, it goes through the absorption by the constituent gases in the atmosphere including water vapour and scattering processes by clouds and aerosols. Due to large variability of the atmospheric processes such as convection, the radiation that eventually reaches the Earth surface (often known as global horizontal irradiance or GHI, which is the sum of the direct normal irradiance, DNI (with a factor of the cosine of the incident angle) and the diffusion horizontal irradiance varies over enormous ranges of spatial and temporal scales.

There are indeed limitations on how well the radiation at different layers of the atmosphere is observed and estimated. Radiation at the high atmosphere is now observed by sensors on board the satellites orbiting the Earth (Cess et al. 1989, 1991). However, bias in the satellite-derived datasets has been identified (e.g., Weymouth et al. 2001; Perez et al. 2002, 2003; Rigollier and Lefevre 2004; Blanksby et al. 2013). While the effects from aerosols are usually not considered in these datasets, Blanksby et al. identified no significant evidence of bias due to the presence of aerosols. The other processes involved in cloud estimation may introduce uncertainties too. For example, in the State University of New York (SUNY) model (Perez et al. 2002) estimate of the cloud index from infrared satellite images are first performed, which is then used in a radiation transmittance function. The SUNY model also accounts for effects of atmospheric turbidity, snow cover, ground specular reflectance characteristics and satellite angle effects for individual image pixels (Nottrott and Kleissl 2010). Similarly, in the Australian Bureau of Meteorology (BoM) the global solar exposure is estimated based on visible images from the geostationary meteorological satellite MTSAT-2 followed by downward radiative fluxes calculations that considered data of hourly cloud albedos (Weymouth et al. 2001). Besides the ground observations of radiation, solar irradiance at the Earth's surface can also be estimated by utilizing numerical weather prediction (NWP)models that either utilize radiative parameterization schemes (e.g., Chou and Lee 1996; Li and Moreau 1996)or perform explicit radiative transfer calculations of all the scattering processes within the atmosphere in some recent models (e.g., Nielsen et al. 2013). The latter especially applies to determination of solar irradiance over ocean surface (e.g., Conant et al. 1997). While this kind of radiative transfer calculation is based on

rigorous physical principles, large uncertainties in simulating the cloud development processes in the atmosphere certainly lower the quality of such estimation (e.g., Ohtake et al. 2013). For example, Sun and Liu (2013) discussed several methods to determine the DNI, which include simple derivation from sunshine duration, radiative transfer calculations with consideration of sky condition, parameterized global and diffuse irradiance and retrieval from back-scattered radiation data of satellite measurements (Masuda et al. 1995; Polo et al. 2013). As expected, there are differential sources of errors for these methods. Structures in the radiative transfer models also have to adapt to different cloud patterns such as those identified in this study for Australia.

With all the aforementioned techniques and methodologies, global datasets of radiation budget and energy balance are available, such as the Earth Radiation Budget Experiment (Barkstrom 1984) and the Global Energy Balance Archive (Gilgen and Ohmura 1999). While these datasets represent quite good estimates of the GHI with large spatial coverage and high enough temporal resolution for regional planning of renewable energy production, the regional details and patterns in the cloud absorption component can only be revealed by the ground station observations. Certainly, the best temporal resolution of station observations of irradiance is about daily such as the dataset applied in this study, which is less than that available from NWP models. However, if the timeframe of applications is from the sub seasonal to more extensive periods, such as examining the variability of solar irradiance in a region for power facility planning, simple statistical models of solar irradiance based on ground observations are able to avoid the uncertainties due to the assumptions inherent in the parameterization schemes of NWP models. The major objective of this study is thus

to analyze the solar absorption by clouds and other components in the atmosphere using the ground-based station observations of shortwave radiation in Australia. Statistical analysis is then performed to analyze the spatial and temporal patterns of radiation absorption within the country. The two statistical models applied in the study are both standard time series models based on statistical regression. With the identified model parameters here, the models can be conveniently executed using the popular statistical software packages. Moreover, the observational dataset from BoM in this study represents a substantial extension to some existing databases available to the energy industry (e.g., AuSES 2006), which is beneficial to regional-scale consideration of the availability of solar energy.

The organization of this paper is as follows. The following section describes the solar irradiance observations used in this study and the statistical models applied to analyze the observations. Section 3 presents the analysis results from these models and discusses the spatial and temporal patterns of solar irradiance and estimated cloud cover. Analysis of the identified regions of estimated cloud absorption in comparison with the Australian climate regimes is given in section 4.Section 5 then summarizes this study with further discussion on the implications of the results obtained here as well as future research.

2. Data and Methods

Statistical models are used to analyse the Australia solar radiation data collected from 144BoM weather stations with installed ground pyranometers spread evenly over Australia. These stations use CM-11 pyranometers manufactured by Kipp & Zonen to measure both the GHI and the diffuse component, and the daily GHI

values are used in this study. Calibrations based on an efficient 'Alternative Method' developed by the BoM of the pyranometers for both clear and cloudy sky conditions are carried out regularly to ensure accurate measurement of irradiance. These data are collected over the years of 1990 to 2012 (total 23 years), and station locations are displayed in Figure 1.

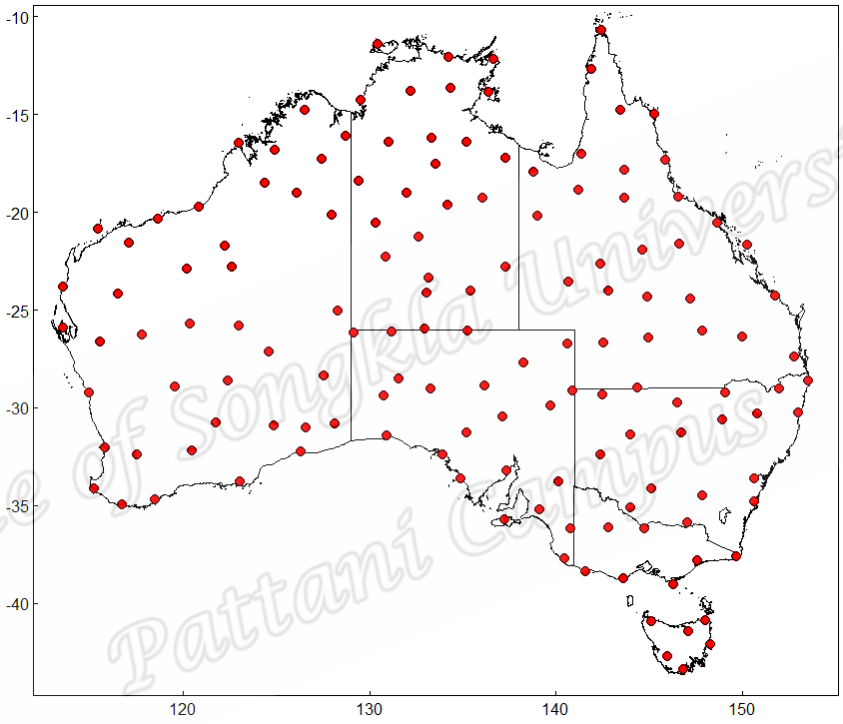

Figure 1 Station locations in Australia where solar radiation data were collected. The state boundaries in Australia are also given.

Each station supplies daily solar data observed at the station (R_G), and therefore there are 8395 observations for each station over 23 years. In order to maintain equal number of observations for each year, we omit one observation in each leap year, namely the observation on Feb 29. The extraterrestrial irradiation for each station location, assuming that the Earth is a sphere, is calculated by the formula in Klein (1977). A fraction, P_0, of the incoming radiation (R_E) is absorbed by the upper

atmosphere. Thus, we are able to calculate the "observed" daily percentage solar radiation absorbed by clouds (R_C) above the station with the expression $R_C = [1 - R_G / R_E(1-P_0)] \times 100\%$. The statistical models discussed in the following are then fitted to R_C. In this paper we adopt $P_0 = 0.2$. The rational of this particular P_0 value is evidenced in Figure 2 where plots of $R_E(1-P_0)$ and observed maximum solar radiation are provided for eight selected stations.

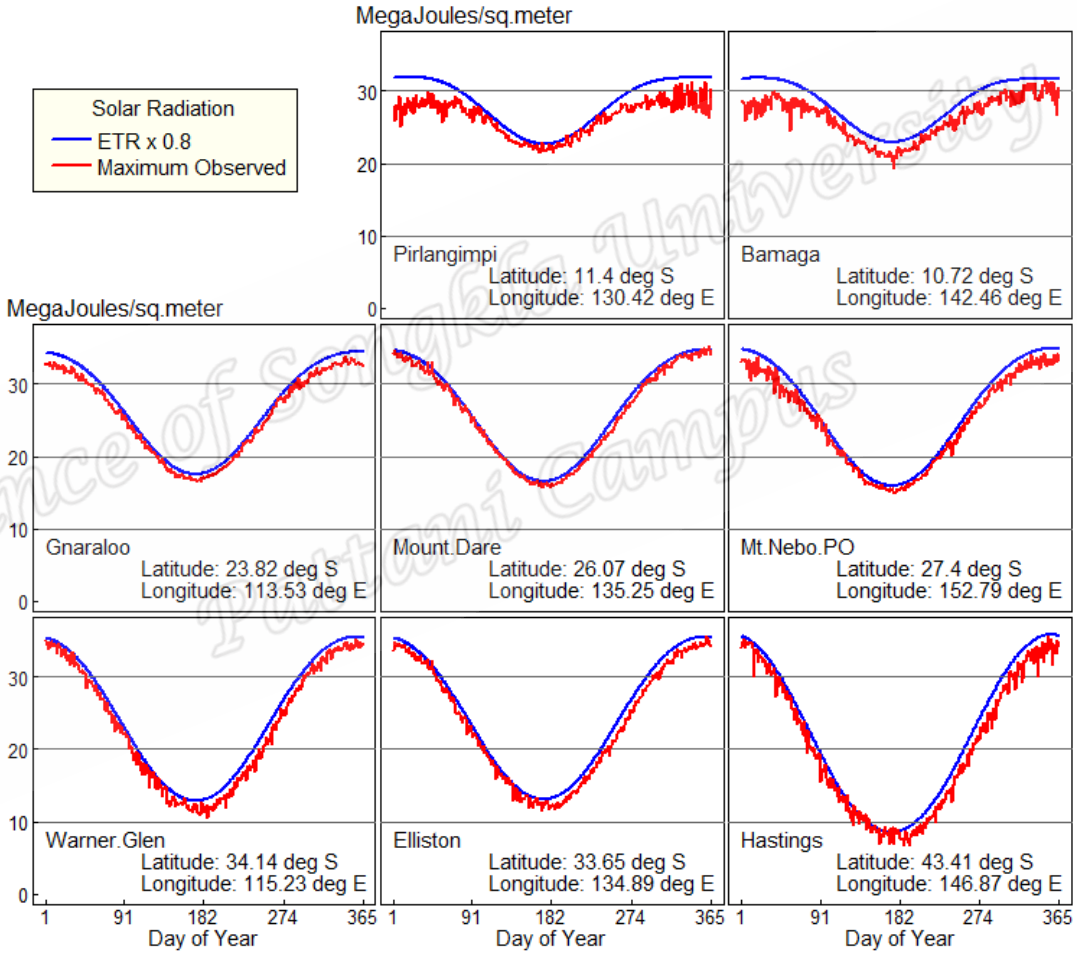

Figure 2 Plots of the observed maximum solar radiation and 0.8 times the extra-terrestrial solar radiation for eight selected stations.

These eight stations evenly cover Australia: two are located in the north, three in the middle and three in the south, all from west to east. It should be noted that there

are more sophisticated methods for estimating upper atmosphere absorption such as taking into account the solar angle, time of year, etc. However, the constant P_0 value adopted here suffices the purpose in this study of developing simple statistical models of solar absorption.

Direct usage of the raw daily percentage radiation absorption data when constructing statistical models is not ideal because they are severely right skewed making the required normal distribution assumption less accurate. One way to partially rectify this problem is to average the data. For each station, we take the five-day average of the daily percentage solar radiation absorption and thus reduce the number of radiation data to 73 in each year. Where data are missing, the average is based on the number present. If all data in a 5-day period are missing, an additive model, described below, is used to interpolate these missing values. These 5-day averages are called the **period** solar radiation absorption data in the sequel. Note that the normality assumption requires that errors be normally distributed and this assumption may be assessed by plotting residuals against normal quantiles after a model is fitted. These plots will be presented and discussed later after the models are fitted.

We use two statistical models to analyze the period radiation absorption data. Although both models involve regressions, they are different. In the first model, the location dependent square root transformed period solar radiation is used as the response variable and year and period factors are the covariates. Here location dependent square root transformation means that it is only applied to those stations with latitude less than 30 degrees. Another issue for this type of data is possible dependence among the response observations, which violates the independence

assumption. We successfully remove dependence by including a lag 1 term (equivalent to 5-day period) of the response variable in the model, evidenced by the corresponding auto-correlation function (ACF) plot displayed in Figure 3.

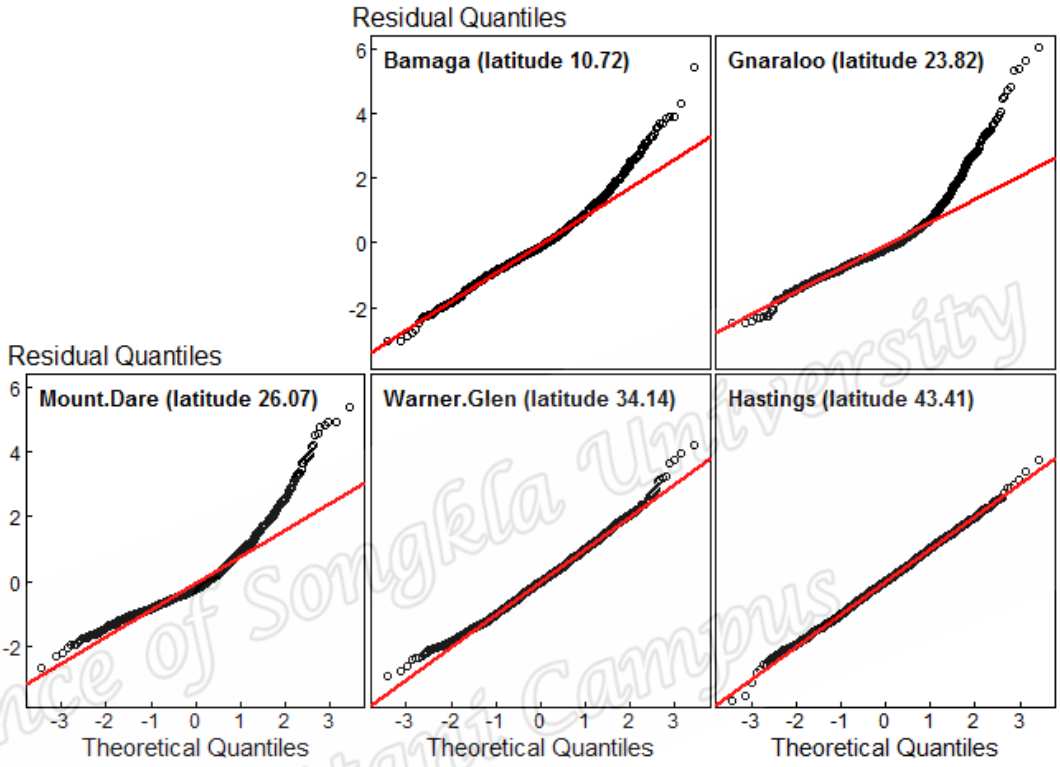

Figure 3 Residual quantile-quantile (Q-Q) plots for five selected stations, where the response of the additive model is five-day average of the daily percentage solar radiation absorption.

The second model first uses factor analysis (Venables and Ripley 2002, chapter 11) to group stations and then adopts the average period radiation absorptions of the stations in each group as the response and the year and period factors as the covariates. Detailed descriptions of these two models are given below.

For station k at year t, let y_{tsk} be the period percentage solar absorption observation for period s. The first regression model considered by this paper is, for

$t = 1, \dots, 23$ and $s = 1, \dots, 73$,

$$g(y_{tsk}) = \alpha_{0k} + \sum_{i=1}^{22} \beta_{ik} x_{ti} + \sum_{j=1}^{72} \gamma_{jk} z_{sj} + g\left(y_{t(s-1)k}\right) + \epsilon_{tsk} \quad (1)$$

Where function $g(\cdot)$ denotes the location dependent square root transformation (details are given later),$\epsilon_{tsk}$are independent error terms following a normal distribution with mean 0 and variance σ^2,x_{ti} are indicator covariates for comparing years$t$and $i$and z_{sj} are indicator covariates for comparing periods$s$ and j. Therefore,x_{ti} takes value 1 if$t = i$ and 0 otherwise and z_{sj} takes value 1if $s = j$ and 0 otherwise.The transformation function $g(\cdot)$ in (1) is defined as: $g(y_{tsk}) = 10 \times \sqrt{y_{tsk}}$ if station k has latitude less than 30 degrees; otherwise $g(y_{tsk}) = y_{tsk}$. Here number 10 is used to multiply square root of y_{tsk}so that the scale of $g(y_{tsk})$ is approximately the same as y_{tsk}, and therefore different plots involving $g(y_{tsk})$ and y_{tsk} are directly comparable. For stations with latitudes larger than 30 degrees, it is found that the data follow normality quite well already and thus no such transformation is necessary. In (1), year1990 and period 1 are used as, respectively, the baseline year and the baseline period. For station k, α_{0k} represents the average period solar radiation absorption corresponding to the baseline year and the baseline period, β_{ik} is the year i effect and γ_{jk} the period j effect. The normal distribution and independence requirements on $\epsilon_{tsk}$are explained by respectively Figures 4 & 5.

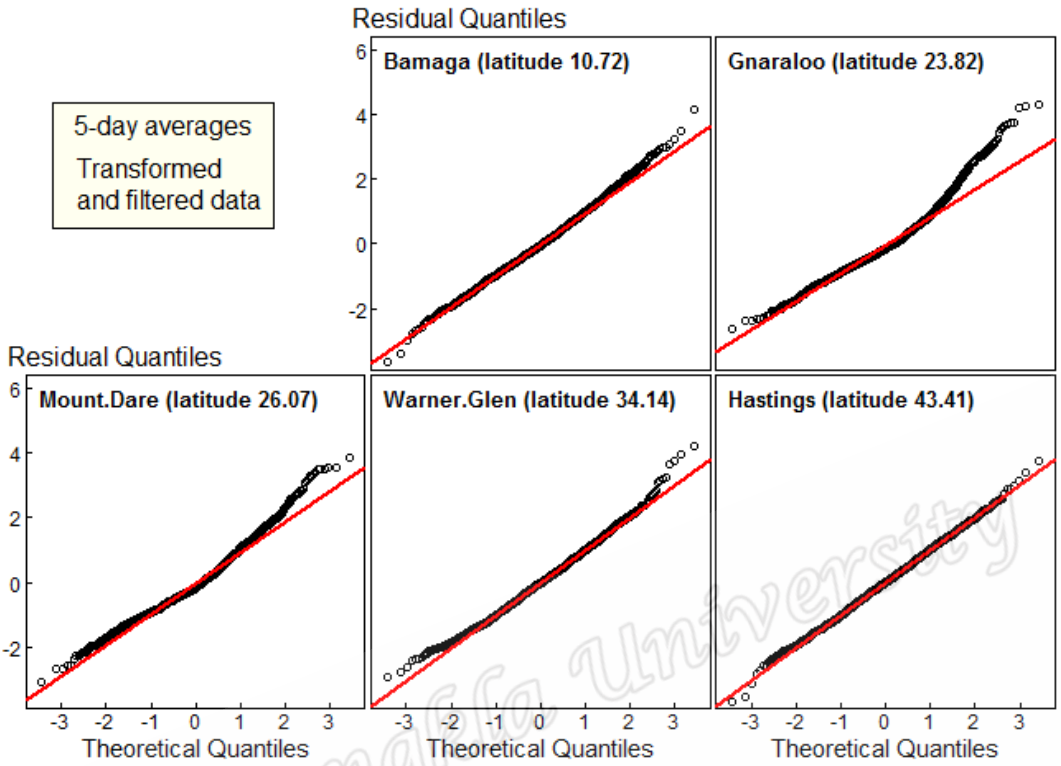

Figure 4 Residual quantile-quantile (Q-Q) plots for the same stations as in Figure 3, where the response is location dependent square root transformed five-day average of the daily percentage solar radiation absorption, and then filtered with a lag 1 term.

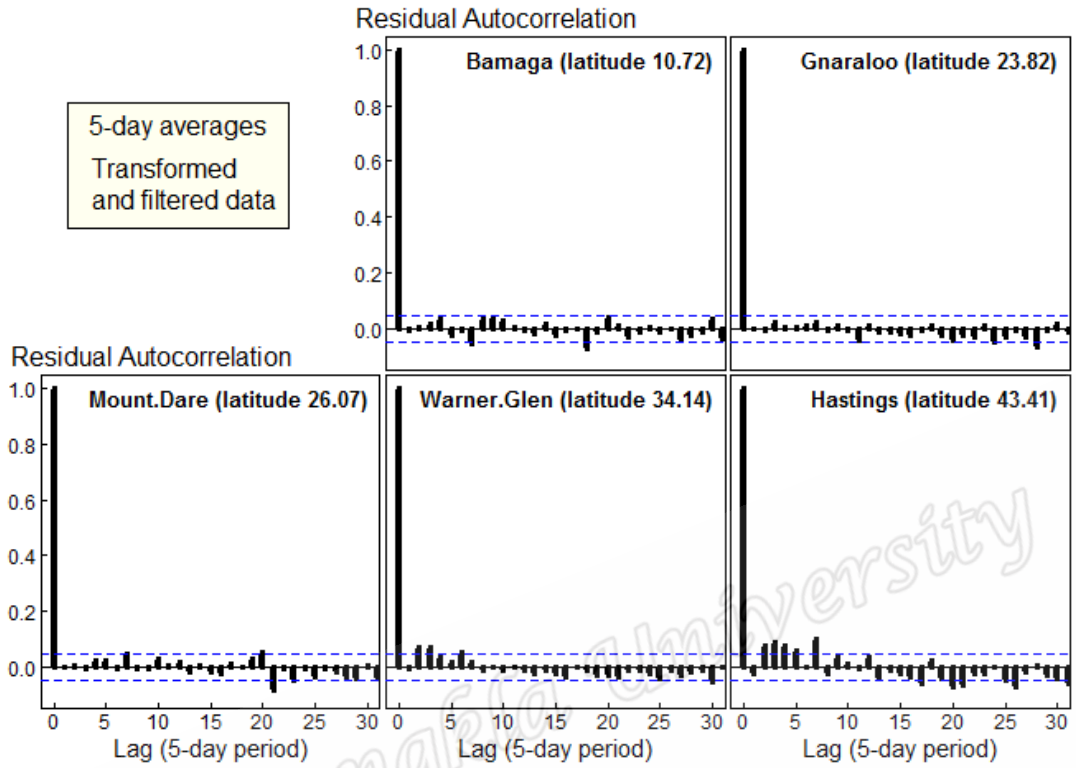

Figure 5 Residual auto-correlation function (ACF) plots for the same stations as in Figure 3. Data are location dependent square root transformed five-day average of the daily percentage solar radiation absorption, and then filtered with a lag 1 term.

In contrast, Figure 3 reveals that for the regression models with untransformed y_{tsk}, some model residuals do not satisfy the normal distribution assumption.

Let y_{sk}be the radiation absorption data for station k at time s. In fact, model (1) can be explained as an approximation to the following non-parametric mean model:

$$E(g(y_{sk}) - g(y_{(s-1)k})) = \alpha_{0k} + f_k(s) \qquad (2)$$

where$f_k(s)$ is the unknown trend function which we wish to estimate (after off-setting by α_{0k}). Several approaches exist to estimate $f_k(s)$, mainly the kernel method and the smoothing spline method (Venables and Ripley, 2002, chapter 8). In fact, model (1) provides an alternative approach to estimate $f_k(s)$, and its idea is based on discretization (or piecewise constant approximation) of $f_k(s)$ where the time interval, in our example, is divided into 1679 (i.e.,23×73) pieces.

The second model is also a regression model, but its response observations are generated from a factor analysis (FA) and the explanatory variables are the year and period factors. More specifically, we first perform a maximum-likelihood FA with 8 factors and "promax" rotation using the correlation matrix calculated from the location dependent square root transformed period radiation absorptions of different stations. We identify 7 significant factors when performing FA, with the details given later. These 7 factors give rise to seven geographical regions of Australia. After conducting FA we then construct the second model where the response observations in our second model are square root transformations of the average period radiation absorptions calculated using the stations within each of the seven regions resulted from the FA; see the model description in equation (3). Note that the average period radiation absorptions for factor 2 are not transformed as they already follow closely to a normal distribution according to the Q-Q plot in Figure 6. Details of FA, its maximum likelihood estimation and factor rotations can be found in chapter 11 of Venables and Ripley (2002). But briefly, FA is a method which creates a new set of **uncorrelated** variables, called the **factors**. These factors are identified using the estimated quantities called factor loadings. Factor loadings can be estimated by either the principal component method or the maximum likelihood method, and then the

estimated loading are usually rotated to achieve better interpretation. “Promax” is one of the “oblique” rotation methods that can “maximally” separate the loadings.

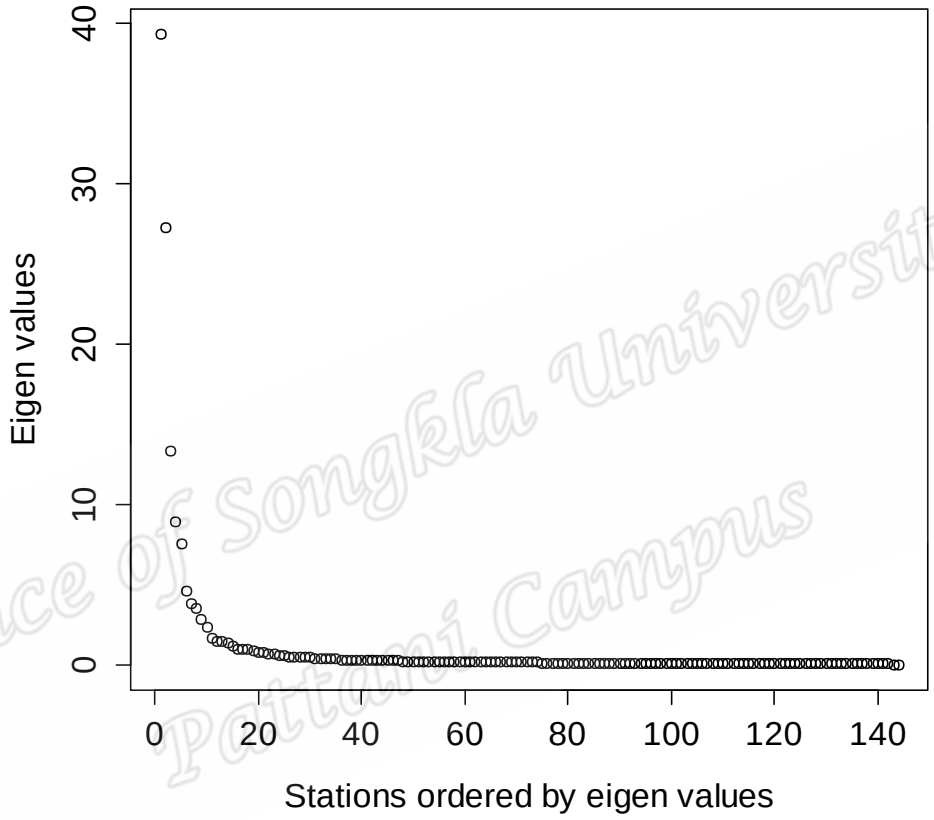

Figure 10 Plot of ordered eigen-values (from largest to smallest) for FA. This plot indicates that no more than 10 factors are required for FA.

Figure 10 exhibits the plot of the ordered eigen-values (largest to smallest) from the correlation matrix of the transformed period percentage radiation absorptions. Since the change point of this plot occurs at around the 10^{th} ordered eigen-value, it suggests no more than 10 factors are necessitated for FA. This together with the fact that the first 8 eigen-values account for 75% of the total variation in the data, we decided to adopt 8 factors in FA initially. From the calculated factor loadings

of these 8 factors, factor 8 contains only two stations where the loadings for factor 8 are dominant. This indicates we would overfit the FA model if factor 8 is included, and therefore, we select 7 factors from our FA. Then, we associate the stations with these 7 factors according to the following procedure: (i) for each factor, we first identify the stations with loadings greater than 0.333, as explicated in Tabachnick and Fiddell (2007) page 646, then (ii) we associate a station with that factor if its factor loading is the largest among all the loadings for 7 factors.

Let u_{tsk} be the average period radiation absorption for FA-based region $k(k = 1, \dots, 7)$ at year t and period s, then our second model in this paper is

$$h(u_{tsk}) = \delta_{0k} + \sum_{i=1}^{22} \eta_{ik} x_{ti} + \sum_{j=1}^{72} \xi_{jk} w_{sj} + h(u_{t(s-1)k}) + \epsilon_{tsk} \quad (3)$$

where x_{ti} are indicator covariates for comparing year t with year i, w_{sj} are indicator covariates for comparing period$s$ with period$j$ and $\epsilon_{tsk}$are independent errors following a normal distribution with mean 0 and variance σ^2. The normality and independence assumptions on ϵ_{tsk} are verified by respectively Figures 6 and 7.

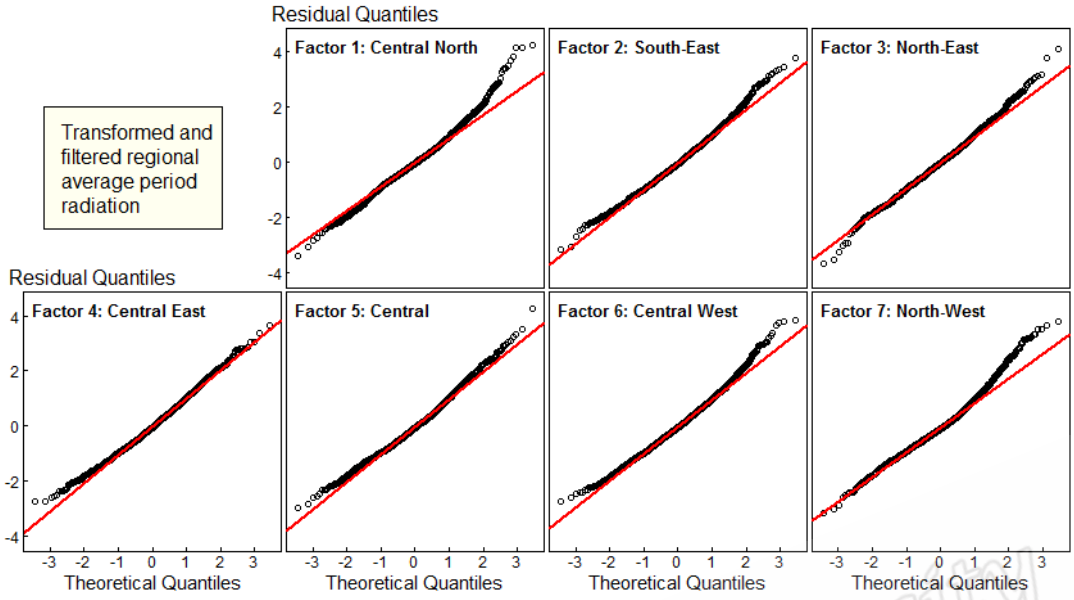

Figure 6 Residual quantile-quantile (Q-Q) plots corresponding to 7 factors, where the response is square root transformed regional average period daily percentage solar radiation absorption, and then filtered with a lag 1 term.

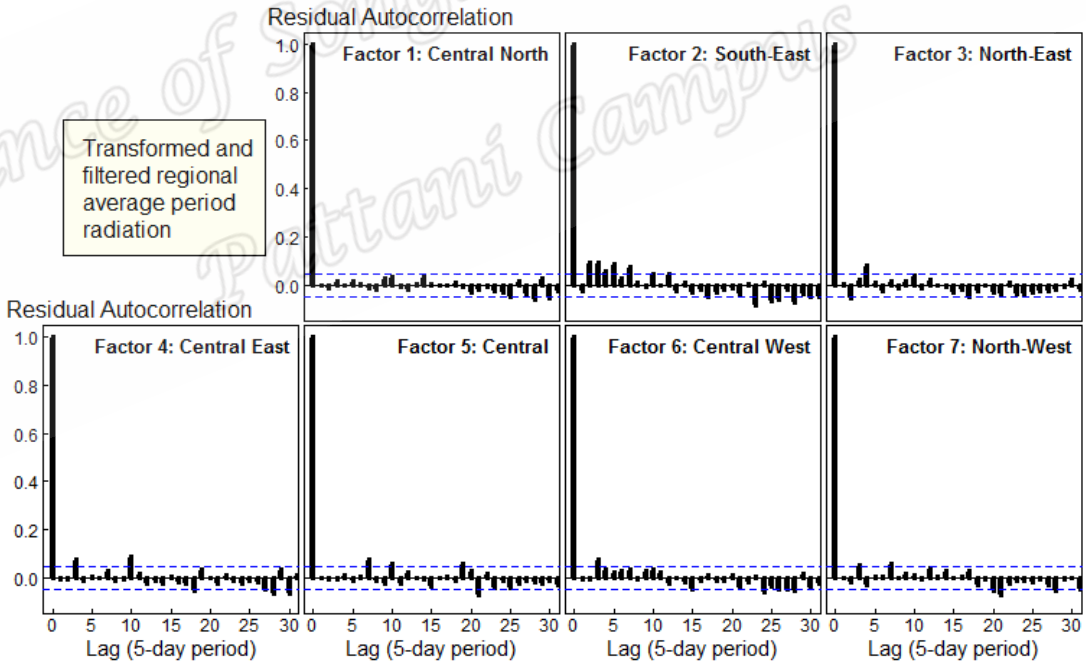

Figure 7 Residual auto-correlation function (ACF) plots corresponding to 7 factors, where the response is square root transformed regional average period daily percentage solar radiation absorption, and then filtered with a lag 1 term.

In (3), function $h(\cdot)$ is defined as $h(u_{tsk}) = 10 \times \sqrt{u_{tsk}}$for $k \neq 2$. By comparing with model (1), model (3) has a coarser discretization grid along the location dimension than model (1) and thus will result in smoother trend estimation.

In order to reveal the month effect we also attempt to replace the period effect in model (3) by month effect variables. This model over-fits the data except the lag term is removed from the model. Thus if we wish to investigate the month effect, such as the plots in Figure 12, we then modify model (3) by including only the year and month effects.

Fitted values $\widehat{g(y_{tsk})}$ and $\widehat{h(u_{tsk})}$ from models (1) and (3) must be backward transformed to give fitted values of $y_{tsk}$and u_{tsk}. Let $\hat{y}$ denote the fitted value for y and $g(y)$ the square root transformation defined above. Then $\hat{y}$ can be accurately approximated from the fitted value $\widehat{g(y)}$by $\hat{y} = \left(\widehat{g(y)}/10\right)^2$.

3. Results

Figure 8 exhibits plots, for five selected stations, of the period solar radiation absorption data (dots) together with the fitted values (red solid curve), where the fitted values are the $\hat{y}_{tsk}$'s given by backward transformation of $\widehat{g(y_{tsk})}$.It is demonstrated that model (1) can estimate the seasonal variation and trend in the time series (i.e., the mean(y_{tsk}) reasonably well, as also demonstrated by the residual Q-Q plots for selected stations in Figure 6,and therefore is useful for prediction of future average radiation absorption values. Closer examination of individual years shows that besides the means and trends, the timing of fluctuations has been quite well modelled (Figure 9).

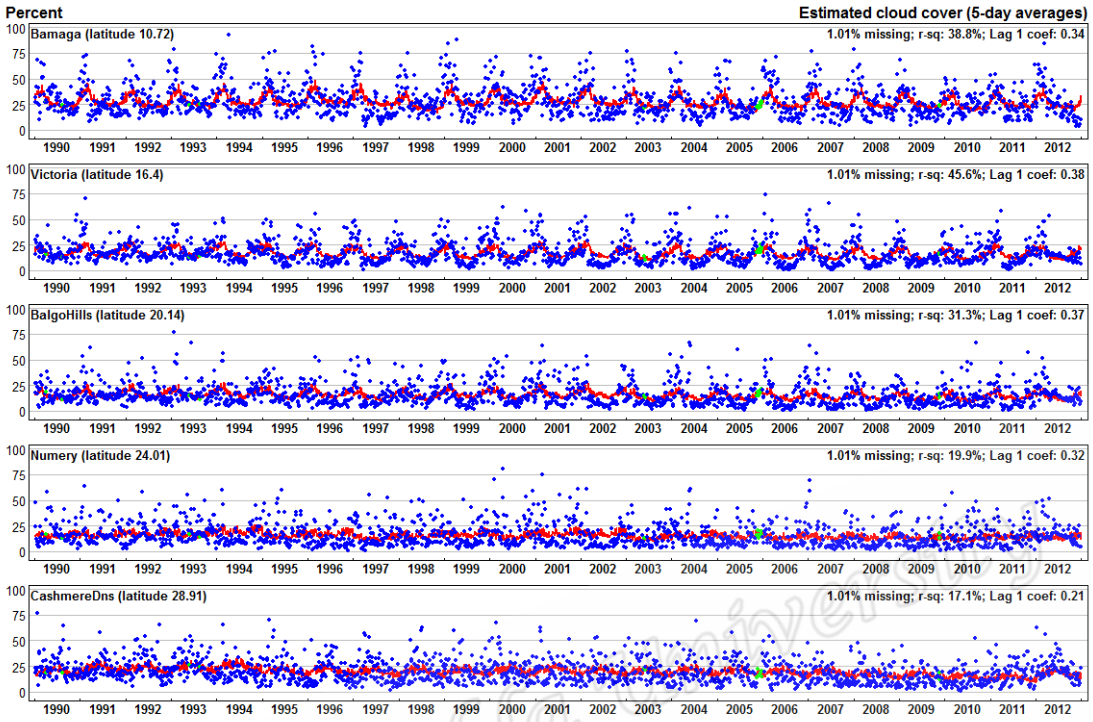

Figure 8 Plots of period solar radiation absorption data (blue dots) and their mean estimation using model (1) (red curves) for 5 stations with latitudes between 10 °S and 30°S. Green dots denote missing values fitted using the model.

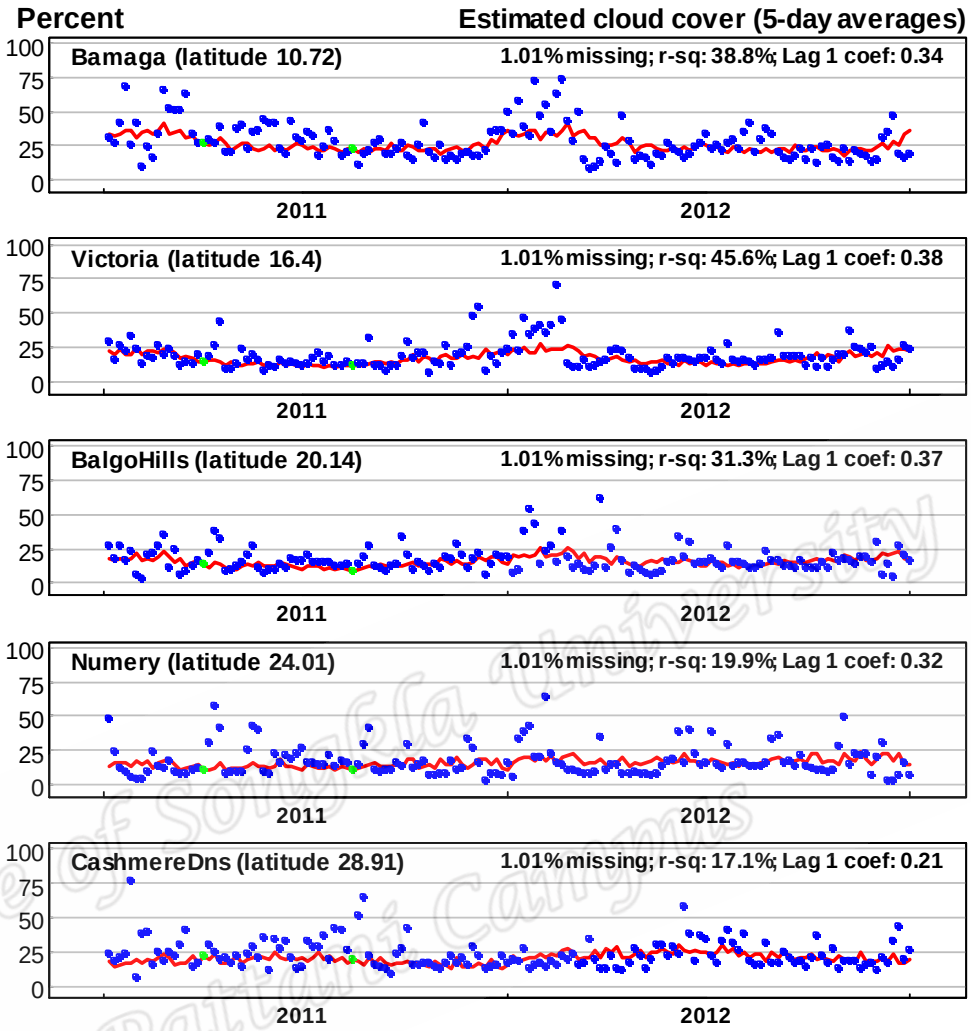

Figure 9 As in Figure 8 except for the years 2011 and 2012 only.

Due to the nature of the model, the magnitudes of fluctuations are much smaller than the observed data; however, this is out of the model's scope to pick up individual fluctuation. The model applies equally well to the other stations. However, there is no consideration of the spatial distribution of radiation absorption in this model.

On the other hand, factor analysis is able to give seven dominant components in solar radiation absorption, which are related to estimating the variation of the percentage period solar radiation energy absorbed by clouds and other bodies in the lower atmosphere. This result divides the stations into 7 regions geographically (Figure 11), in each of which a dominating factor can be identified (i.e., a single large loading is found).

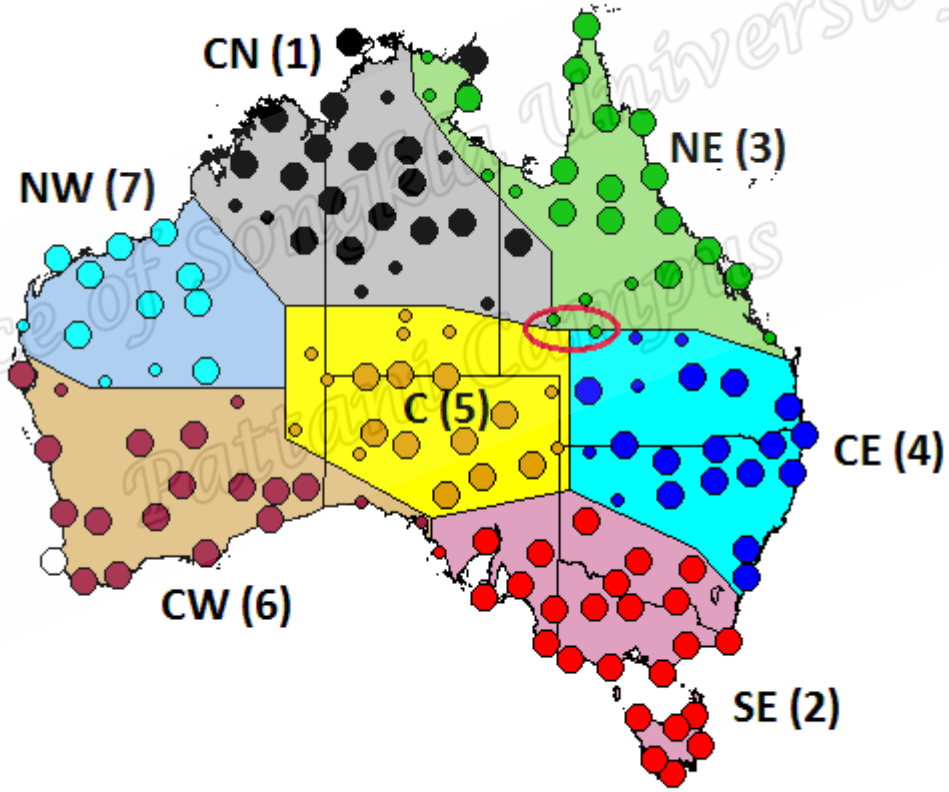

Figure 11 Factor analysis (FA) divides the stations into 7 geographical groups, where larger circles are stations with only a single loading greater than 0.333 and smaller circles are stations with more than one loading greater than 0.333. The number in bracket indicates the dominant factor in a particular region.

The corresponding geographical regions for the 7 factors are Central North (CN, factor 1 is the dominating fact, similarly hereafter), South East (SE, factor 2), North East (NE, factor 3), Central East (CE, factor 4), Central (C, factor 5), Central West (CW, factor 6) and North West (NW, factor 7). While in most of the regions there are a few stations for which mixed factors are necessary with a second factor with loading larger than 0.333 (Table 1), the factor with the largest loading remains the same as those stations with single dominating factor in each of the regions. The model (3) originally gave eight factors but factor 8 only contributes in two stations in Queensland (Springvale and Westerton represented as the circled stations in Figure 11) with smaller loadings than other factors. Therefore, factor 8 is not a dominating factor in any region. Moreover, there is a station (Warner Glen) at South Western Australia (empty circle in Figure 11) with all loadings of factors smaller than 0.333, and thus there is no identified dominating factor for this station.

Table 1 The thirty-four stations in the seven geographic regions with mixed (two) identified dominating factors (in bold).

Region	Station name	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8
1	Willowra	**0.814**		-0.117		0.250			**0.368**
1	Bulka	**0.796**		-0.131				**0.353**	
1	Nerrima	**0.711**						**0.431**	
1	Cygnet Bay	**0.697**						**0.364**	-0.169
1	Wandie Creek	**0.708**		**0.343**					-0.191
1	Vaughan Springs	**0.641**		-0.179		**0.433**		0.17	0.287
1	Broadmere	**0.647**		**0.461**				-0.108	
1	Marqua	**0.527**		0.138	0.116	0.249		-0.14	**0.476**
2	Elliston		**0.503**			0.227	**0.384**	-0.161	
3	Birricannia			**0.791**	0.22				**0.396**
3	Doomadgee Township	**0.421**		**0.647**					
3	Calvert Hills	**0.529**		**0.564**				-0.124	
3	Tranby	0.159		**0.586**	0.246				**0.485**
3	Bulman AWS	**0.508**		**0.522**					-0.193
3	Captain CreekTM	-0.134	-0.144	**0.490**	**0.475**	-0.104			0.166
3	Maningrida	**0.442**		**0.515**					-0.268
3	Springvale	**0.325**		**0.334**	0.248	0.149		-0.134	**0.461**
3	Westerton	0.150		**0.420**	**0.388**				**0.465**
4	Pingine			0.19	**0.694**				**0.34**
4	Mantuan		-0.106	**0.487**	**0.545**				0.331
4	Tilpa Trevallyn		**0.469**		**0.529**	0.253			
4	Thornleigh			**0.475**	**0.489**				0.424
4	Tibooburra Narriearra		0.277		**0.469**	**0.427**			
5	Pipalyatjara		-0.151			**0.826**		**0.390**	
5	Giles MO		-0.137			**0.745**		**0.476**	

Region	Station name	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8
5	WallaceRH	**0.348**				**0.681**	-0.124	0.113	**0.335**
5	Oak Valley	-0.113				**0.7**	**0.491**		-0.117
5	Cameron Corner Lindon		0.253		**0.349**	**0.546**			0.109
5	Amburla	**0.509**		-0.107		**0.554**	-0.108		**0.338**
5	Numery	**0.419**				**0.538**	-0.107		**0.424**
5	Cordillo Downs	0.142			**0.337**	**0.459**			0.325
5	Ilkurlka		-0.145			**0.601**	**0.528**	0.213	
6	Nullarbor		0.131			**0.362**	**0.672**	-0.133	-0.179
6	Muggon					-0.198	**0.520**	**0.380**	
6	Tjukayirla		-0.179			**0.426**	**0.480**	**0.450**	
6	Smoky Bay		0.353			**0.376**	**0.457**	-0.164	
7	Koonmarra					-0.1	**0.479**	**0.531**	
7	Cunyu					0.142	**0.359**	**0.655**	
7	Gnaraloo		0.161			-0.118	**0.341**	**0.477**	

The seven dominating factors show different variability and trends through the fitted models (Figure 12).

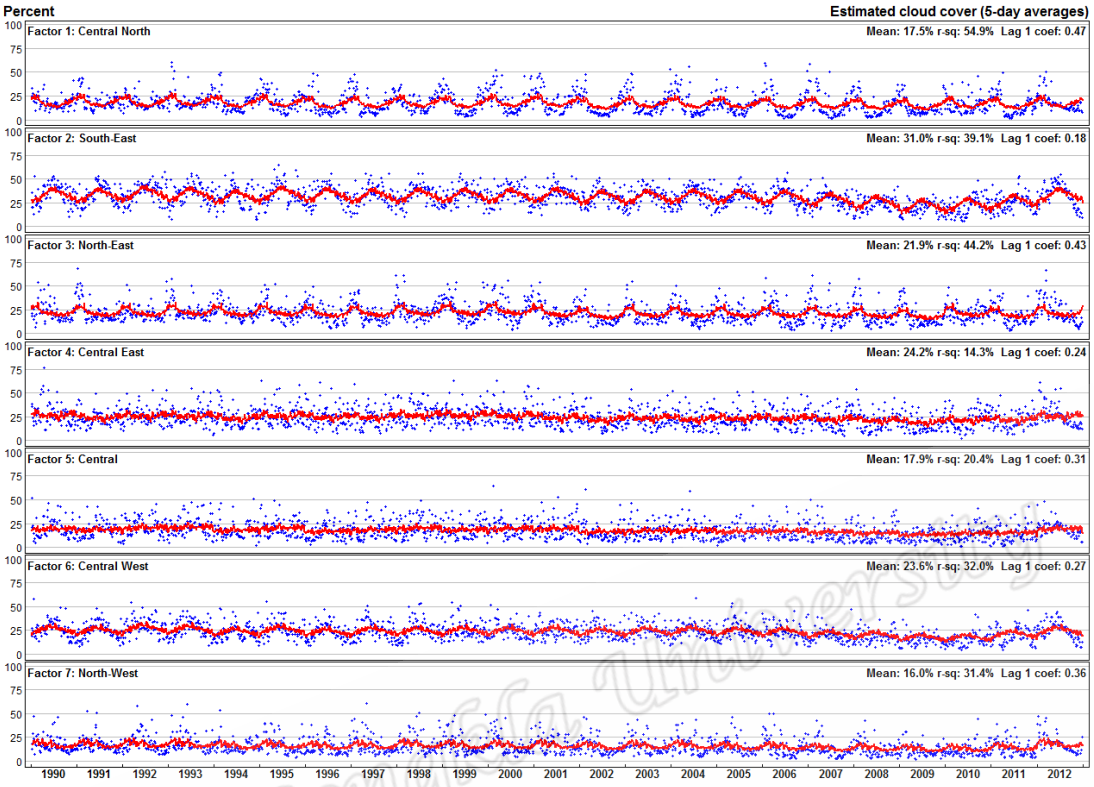

Figure 12 The seven dominating factor in the trend estimation based on model (3) in which they include the year and period (5-day average) effects.

It can be seen that factors 1, 3 and 7, which represent the tropical regions of CN, NE and NW Australia respectively, have distinct seasonal variation. Similarly for factor 2 that represent the SE Australian region with the highest latitudes among all the regions. Comparatively, solar absorption in the regions of CE, C and CW with dominating factors 4, 5 and 6 respectively has large temporal fluctuation in the sub-seasonal timescale, which reflects the fact that a large number of transient weather systems such as fronts and low pressure systems may affect the cloud cover in these regions. It can be seen from Figure 11 that model (3) mainly captures the mean solar absorption as well as its long-term trend in these regions.

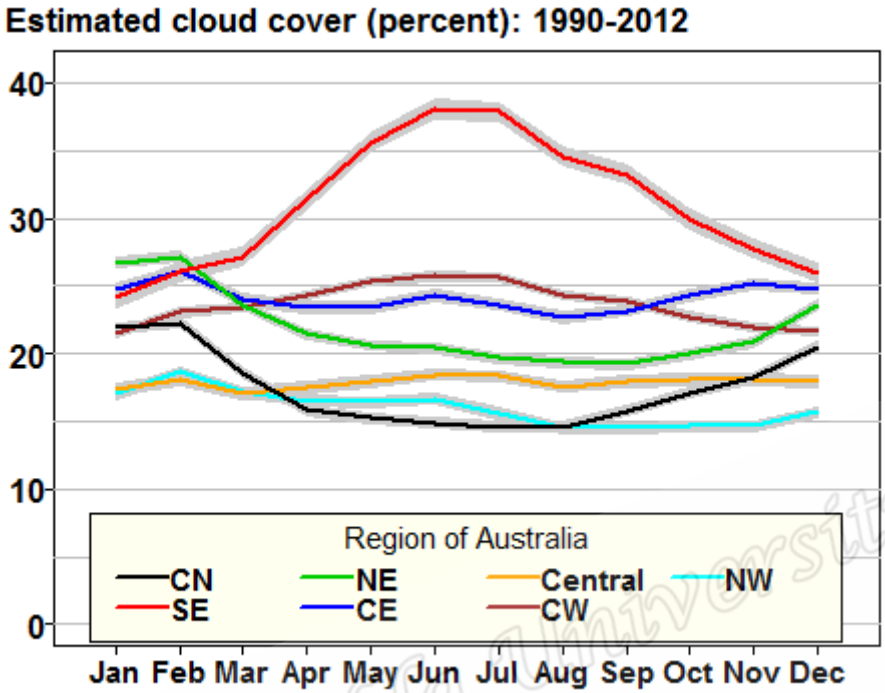

Figure 13 Estimated percentage cloud cover (line) with 68% confidence interval (shaded) in each month of a year.

Figure 13 shows estimates of the seasonal variation within a year of cloud cover for each region, highlighted with colours corresponding to those in Figure 11. Since this plot reflects month effects, it uses the results of model fitting where model (3) is replaced by a model containing the year and month effects only. In the SE regions the percentage increases sharply from January to June, then decreases from July to December. The largest percentage in the CN, CE and NW is in February, with a substantial decline to the minimum in August. However, the percentages in the CE and NW regions increase slightly in June. The largest percentage in the NE region is in February, with a sharp decrease to the minimum in September, whereas the percentage of solar radiation in the Central region was relatively steady, ranging from 17.1% (in March) to 18.5% (in June).Overall, the percentage of cloud cover is

different in each region: the tropical regions have the highest cloud cover in the Austral summer during which the monsoon is active, while the mid latitude region of SE has the highest cloud cover in the Austral winter. In contrast, the other central regions do not have clear seasonal cycles. Inter annual trends in these percentages are plotted similarly in Figure 14.

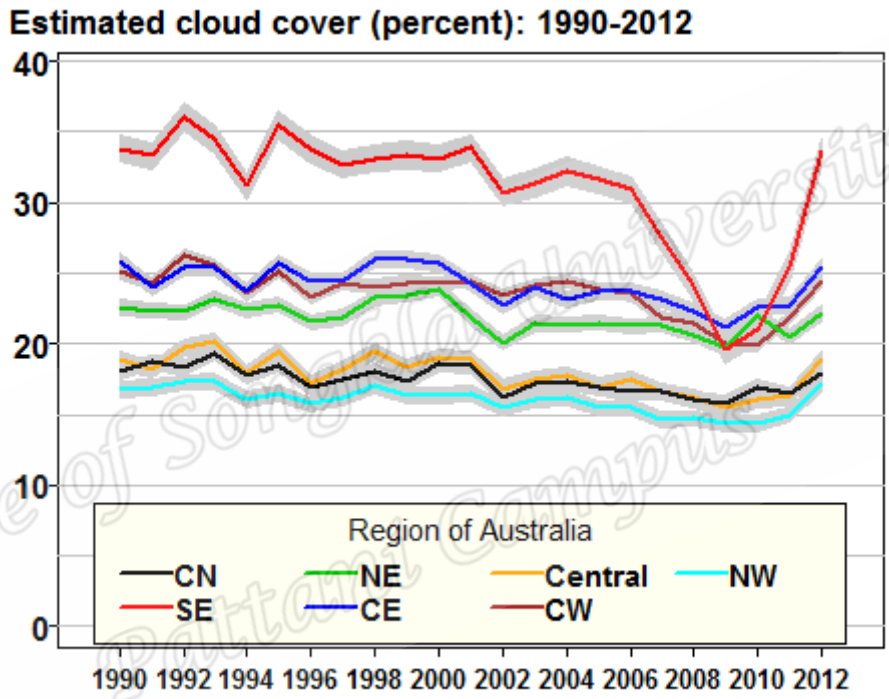

Figure 14 Estimated annual average percentage cloud cover (line) with 68% confidence interval (shaded) during 1990-2012.

In all regions these trends oscillate between 14% and 36%, reached minimum values in year 2009, and then grew rapidly to their previous levels by the following year. Maximum values in the NW, CW and SE regions occurred in around 1992, whereas in the CN, Central, NE and CE regions they occurred in 1993, 1993, 2000 and 1998, respectively. In terms of the average percentage of cloud cover, it is interesting to note that the seven regions based on factor analysis can be classified

into three groups. The SE region forms a group by itself with the highest cloud cover percentage usually above 30% throughout the period of analysis. The regions NE, CE and CW form another group with an average cloud cover percentage of about 25%. The remaining regions of Central, CN and NW then form the last group with the lowest percentage of less than 20%. It can be seen that this grouping roughly follows the geographic distribution of these seven regions from the more tropical group to that in the mid latitudes. One exception is the CW region that has its estimate cloud cover close to those in CE and NE, and with similar inter annual variability as well.

Although no overall monotonic increase or decrease over the 23-year period is evident, it can be seen that the estimated cloud cover is particularly low in 2009 especially for the SE region, which represents the minimum from a decreasing trend from the early twenty-first century. Since cloud cover is roughly proportional to actual rainfall, the annual total rainfall recorded by rain gauges is examined. It is found that there was also a nationwide decreasing trend in rainfall up to 2009 (Taschette and England 2009), and that the serious drought in southeast Australia has been well-known (Figure 15)

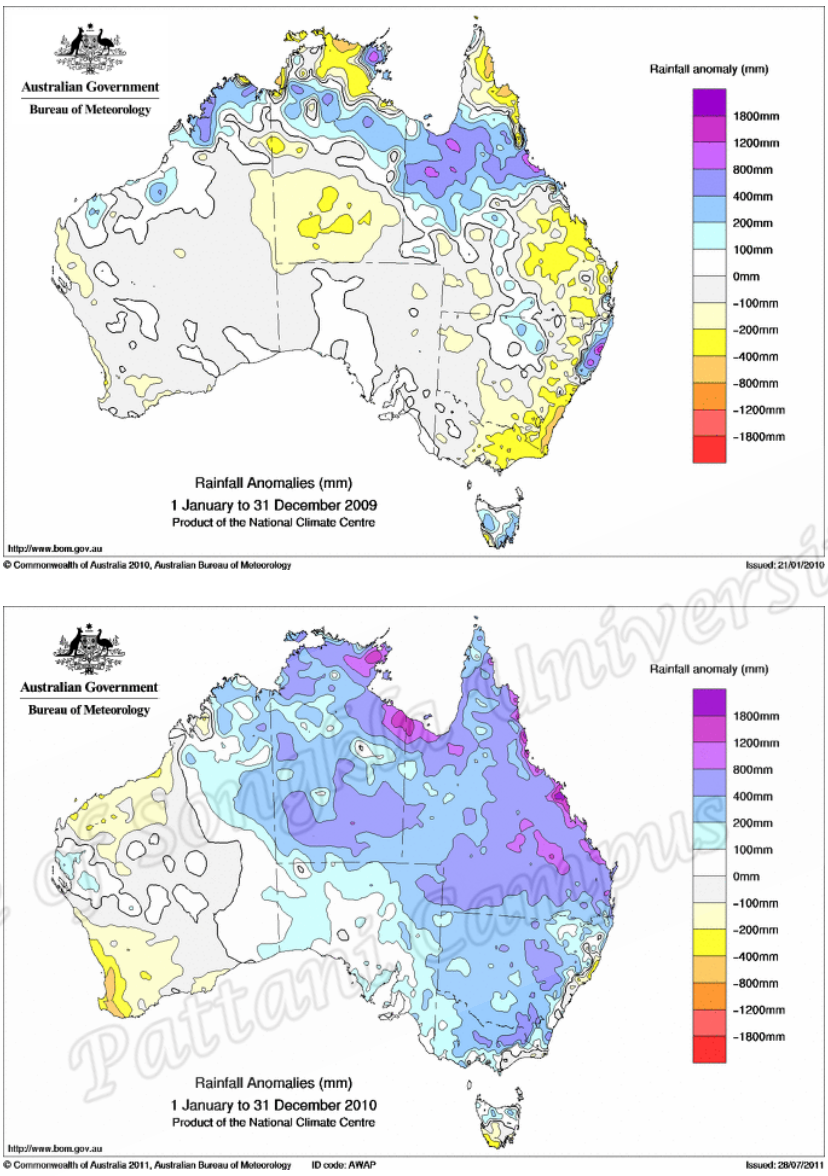

Figure 15 Annual rainfall anomalies (mm) in 2009 (upper panel) and 2010 (lower panel) (Source: Australian Bureau of Meteorology).

Afterward from 2010, the annual rainfall increased again and this trend is consistent with the estimated cloud cover in the last three years of analysis here. The increase of rainfall in this decade is attributed to influences from climate variability such as the El Niño Southern Oscillation (Power et al. 1999).

4. Comparison with the Australian climate regimes

The BoM has classified the climate in Australia into six regimes, which include equatorial, tropical, subtropical, desert, grassland and temperate group (Figure 16).

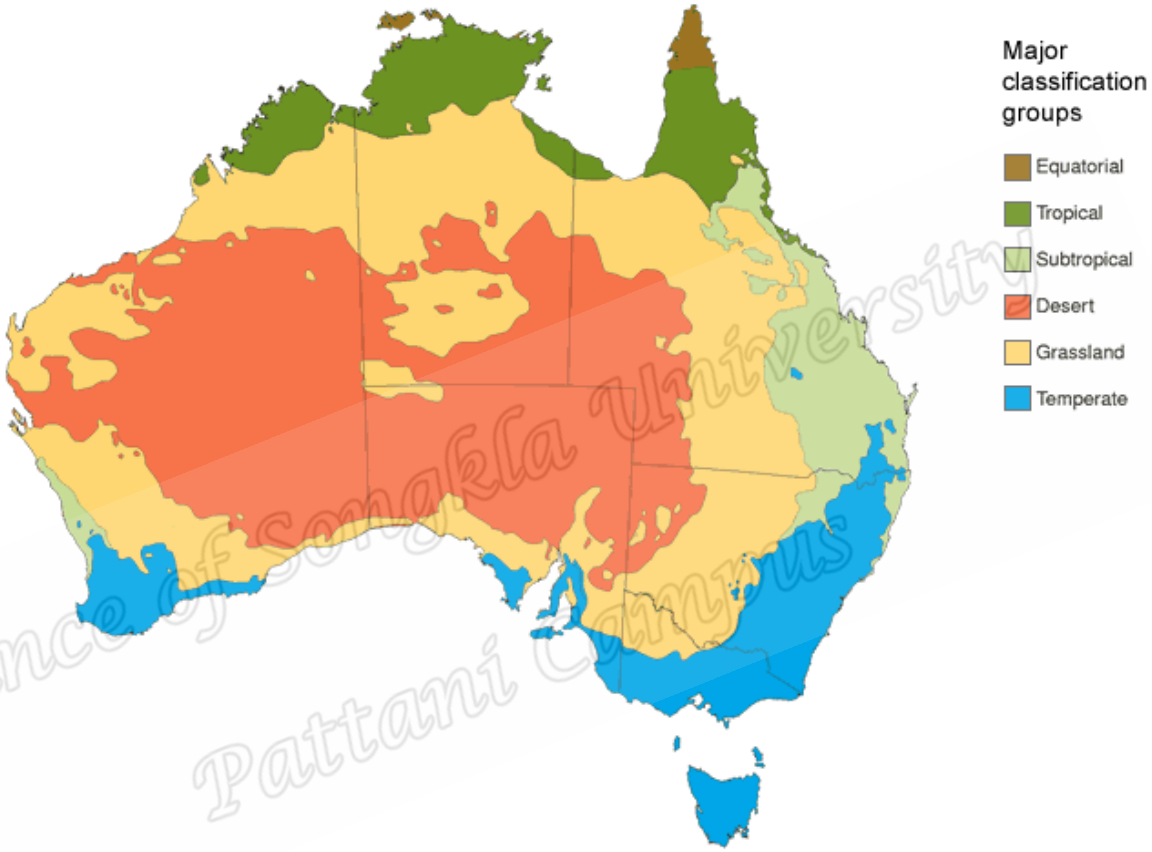

Figure 16 The climate regimes in Australia according to the BoM classification (Source: Australian Bureau of Meteorology).

These regimes were determined based on the modified Köppen scheme (Köppen 1918), which identified climate boundaries with a combination of natural landscape features and aspects of human experience. As shown in Figure 15, the six climate regimes are essentially distributed meridionally with the tropical regime in Northern Australia, grassland regime to the south followed by desert regime in central Australia. Then the temperate regime, which possesses clear seasonal variation,

mainly resides in southeastern Australia, and the subtropical regimes are at the more coastal region of northeastern Australia.

It is interesting to find that the geographic groups as identified by FA do not follow the climate regimes in a straightforward manner, as shown in both the seasonal (Figure 13) and inter annual (Figure 14) variability. Obviously, the SE region stands out as part of the temperate climate regime given its clear seasonal cycle in percentage cloud cover. In Figures 12 and 13, the group of regions with intermediate percentage cloud cover (~20-30%) consists of the NE, CE and CW regions. When comparing with Figure 16, it can be seen that this group crosses from the tropical/subtropical regime of NE to the partly temperate, partly grassland regimes for CE and CW. As discussed in the last section when these FA geographic regions are first identified, the FA factors 4 (associated with CE) and 6 (associated with CW) consist of more sub-seasonal fluctuation in the model and thus these two regions are affected by the transient cloud systems from the west. Although these two regions may not experience as many cloud systems as in the SE, which is much influenced by the polar frontal systems, the CE and CW regions have quite high percentage cloud cover.

On the other hand, the NE region locates in the tropical climate regime and thus the origin of clouds is different from that in CE and CW. The NE region also brings interesting contrast to the other tropical to grassland regions of CN and NW, which together with the central desert region C form the group with the lowest percentage cloud coverage in Figures 13 and 14. This contrast between the eastern and western side of the tropical regime is likely due to the climatological cloud systems affecting them. Both the distribution of the annual average rainfall and that for the summer (which is mainly attributed to the summer monsoon, not shown) indicate that more

rainfall is obtained at the eastern tropical region than the western tropical. This may be due to the onset process of the Australian summer monsoon (Hung and Yanai 2004) through which the cloud systems in the inter tropical convergence zone affect the NE region more. The NE region is also more prone to severe weather systems such as tropical cyclones with a lot of convective clouds embedded in them, thus explaining the higher percentage cloud cover in NE compared with that in CN and NW.

5. Conclusion and discussion

This study analyzes the daily solar irradiance data during 1990-2012 collected from 144 surface stations across Australia. Two statistical models are developed to analyze the 5-daily averaged estimated cloud absorption based on the solar irradiance data in each of the stations. The first is a regression model that has variables of year and pentad factors, and is equivalent to a non-parametric mean model (equations 1 and 2). The second model first performs a FA to obtain the year and month explanatory variables, followed by another regression model (equation 3).

The first model is able to fit the radiation observations well in each of the stations especially those in the tropical region (lower latitudes). That is, the mean value in each of the 5-day periods over the 23 years is well estimated by this model. While the first model effectively fits a separate regression to each station, the second model uses FA to establish clusters of stations with similar characteristics and then a regression is fitted to each of these. The FA, using month and year (or pentad and year) explanatory variables, estimated eight dominant components; however, one of these was found to have little significance. The remaining factors are found to

represent distinct geographic regions (i.e., each of these regions has a different monthly or annual pattern of cloud cover).Namely, in each of the identified geographic regions (CN, NE, CE, SE, CW, NW and C) a unique factor is the dominant component. With the monthly explanatory variable, the model is able to capture the seasonal variation of solar radiation in the tropical regions (CN, NE and NW) as well as SE well. In the other central regions of Australia, the monthly fluctuation of radiation is large. The FA-based model is able to estimate the mean seasonal variation for these regions quite well, and the yearly explanatory variable also estimates the long-term trend well. Examination of the seasonal cycle and inter annual variation of solar radiation based on this model indicates three major regional groups in term of the estimated annual cloud cover, and these three regional groups are distinct from the standard climate regimes in Australia. The SE region that locates in the temperate climate regime forms a group with the largest annual cloud cover percentage. The group consisting of NE, CE and CW has medium percentage, but the origin of cloud systems mainly comes from the mid latitude west erlies for CE and CW while that for NE is mainly tropical. Lastly the group with C (essentially with desert climate regime), CN and NW as members has the lowest percentage. While the CN and NW regions belong to the tropical and grassland climate regimes, the difference in cloud cover in these two regions compared with the other tropical region of NE is noted.

As mentioned in the introduction, the ability to estimate and perform short-term (e.g., daily) prediction of solar irradiance is critical to the establishment and operation of solar energy facilities. Although the existing satellite-based solar irradiance datasets and those from NWP models have better spatial coverage and high

temporal resolution, both the mean model and the FA-based regression model we developed in this study would serve this purpose well for providing the baseline climatology of solar radiation. On one hand, the uncertainties due to various kinds of radiation parameterization in NWP models, which also applied to the satellite-based estimates due to the built-in transmittance calculation, have been avoided in our models based on ground observations only; on the other hand the standard time series regression models developed here can be conveniently reproduced and utilized for realistic applications.

In particular, the spatial pattern that consists of seven identified regions, each of which having its unique dominating factor, is of much value for planning the solar energy resources in various regions within Australia, as well as facility inter-site analysis. Moreover, long-term planning of the applications of solar energy needs forecasts of solar irradiance up to the intra seasonal, seasonal time scale or even longer. For this purpose, both statistical models and dynamical NWP models are often applied. The regional pattern of solar radiation and the associated temporal variability within each region obtained by exploratory factor analysis in this study can be the basis for evaluating such statistical and NWP models. The points of evaluating statistical models include whether the models consider the spatial variability of cloud cover well such that surface irradiance can be accurately estimated. For dynamical weather models, the critical evaluations may include whether the parameterization schemes, radiation physics and cloud-radiation interactions within the models are able to generate the regional pattern of irradiance and cloud absorption.

Nevertheless, the decreasing trend of solar irradiance and the estimated cloud cover in the Australian region starting from the early twenty-first century and

reaching a minimum in 2009 indicates that the availability of solar radiation is much influenced by climate variability and change. Under such influences, the major factors affecting surface solar irradiance such as cloud cover, cloud types and water vapour will be modified spatially and temporally. A worthwhile extension of this study is thus to apply the framework of exploratory factor analysis and regression model to distinct climate periods (or climatic phases as are usually discussed in climate science literature) and investigate how the regional patterns of solar radiation vary. With such extension, the present framework of statistical analysis will benefit even the long-term variability of solar energy and its practical applications. In particular, the fact that there exist different cloud pattern regimes as shown in our FA analysis has implications to the structure of radiative transfer models.

Acknowledgements

The authors are grateful to Prof. Emeritus Don McNeil, Department of Statistics, Macquarie University, Australia, for his guidance and helpful suggestions. The comments and suggestions from the anonymous reviewers have greatly improved the manuscript and are much appreciated.

References

AuSES, 2006. *Australian Solar Radiation Data Handbook (ASRDH)*, fourth edition. Australian Solar Energy Society (AuSES), 80 pp.

Barkstrom, B. R., 1984. The Earth Radiation Budget Experiment (ERBE). Bulletin of the American Meteorological Society 65, 1170–1185.

Blanksby, C., Bennett, D., Langford, S., 2013. Improvement to an existing satellite data set in support of an Australia solar atlas. Solar Energy 98, 111–124.

Cess, R. D., Vulis, I. L., 1989. Inferring surface solar absorption from broadband satellite measurements. Journal of Climate 2, 974–985.

Cess, R. D., Dutton, E. G., DeLuisi, J. J., Jiang, F., 1991. Determining surface solar absorption from broadband satellite measurements for clear skies: Comparison with surface measurements. Journal of Climate 4, 236–247.

Chou, M. D., Lee, K.-T., 1996. Parameterizations for the absorption of solar radiation by water vapour and ozone. Journal of the Atmospheric Sciences 53, 1203–1208.

Conant, W. C., Ramanathan, V., Valero, F. P. J., 1997. An examination of the clear-sky solar absorption over the central equatorial Pacific: Observations versus models. Journal of Climate 10, 1874–1884.

Klein, S. A., 1977. Calculation of monthly average insolation on tilted surfaces. Solar Energy 19, 325–329.

Köppen, W. P. 1918. Klassification der KlimatenachTemperatur, Niederschlag und Jahreslauf. Petermanns Geog. Mitt. 64, 193–203; 243–248.

Gilgen, H., Ohmura, A., 1999. The Global Energy Balance Archive. Bulletin of the American Meteorological Society 80, 831–850.

Hung C.-W., Yanai, M., 2004. Factors contributing to the onset of the Australian summer monsoon. Quarterly Journal of the Royal Meteorological Society 130, 739–758.

Li, Z., Moreau, L., 1996. Alteration of atmospheric solar absorption by clouds: Simulation and observation. Journal of Applied Climatology 35, 653–670.

Liu, B. Y. H., Jordan, R. C., 1960. The inter relationship and characteristic distribution of direct, diffuse and total solar radiation. Solar Energy 4, 1–20.

Masuda, K., Leighton, H. G., Li, Z., 1995. A new parameterization for the determination of solar flux absorbed at the surface from satellite measurements. Journal of Climate 8, 1615–1629.

Nielsen, K. P., Gleeson, E., Rontu, L., 2013. Radiation sensitivity tests of the HARMONIE 37h1 NWP model. Geoscientific Model Development Discussions 6, 6775–6834.

Nottrott, A., Kleissl, J., 2010. Validation of the NSRDB-SUNY global horizontal irradiance in California. Solar Energy 84, 1816–1827.

Ohtake, H., Shimose, K., Fonseca Jr., J. G. S., Takashima, T., Oozeki, T., Yamada, Y., 2013. Accuracy of the solar irradiance forecasts of the Japan Meteorological Agency mesoscale model for the Kanto region, Japan. Solar Energy 98, 138–152.

Peixoto, J. P., Oort, A. H., 1992. Physics of Climate. American Institute of Physics, 520 pp.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F., 2002. A new operational satellite-to-irradiance model. Solar Energy 73, 307–317.

Perez, R., Seals, R., Ineichen, P., Stewart, R., Menicuccic, D., 2006. A new simplified version of the Perez diffuse irradiance model for tilted surfaces. Solar Energy 39, 221–231.

Polo, J., Vindel, J. M., Martín, 2013. Angular dependence of the albedo estimated in models for solar radiation derived from geostationary satellites. Solar Energy 93, 256–266.

Power, S., Casey, T., Folland, C., Colman, A, Mehta, V., 1999. Inter-decadal modulation of the impact of ENSO on Australia. Climate Dynamics 15, 319–324.

Rayl, J., Young, G. S., Brownson, J. R. S., 2013. Irradiance co-spectrum analysis: Tools for decision support and technology planning. Solar Energy 95, 364–375.

Rigollier, C., Lefe`vre, M., 2004. The method heliosat-2 for deriving shortwave solar radiation from satellite images. Solar Energy 77, 159–169.

Sun, Z., Liu, A., 2013. Fast scheme for estimation of instantaneous direct solar irradiance at the Earth's surface. Solar Energy 98, 125–137.

Taschetto, A. S. and England, M. H., 2009. An analysis of the late twentieth century trends in Australian rainfall. International Journal of Climatology 29, 791–807.

Tabachnick, B. G. and Fidell, L. S., 2007. Using Multivariate Statistics,5th edition, Allyn& Bacon, Boston, MA.

Venables W.N., Ripley B.D., 2002.Modern Applied Statistics with S, 4th Edition, Springer, Queensland.

Weymouth, G., Le Marshall, J. F., 2001. Estimation of daily surface solar exposure using GMS-5 stretched-VISSR observations: The system and basic results, Australian Meteorological Magazine, 50, 263–278.