Solar output estimation and analysis

Jingjing Yu
U5413000

Supervised by Nicholas Engerer

May 2015

A thesis submitted in part fulfilment of the
degree of Bachelor of Engineering
Department of Engineering
Australian National University

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Jingjing Yu
3 June 2015

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Mr. Nicholas Engerer, for his guidance and support throughout this project. His time and advice helped me overcome the difficulties of the project and gave me much more understanding of solar output estimation modelling. Without his help, this thesis would not have been possible. I am truly grateful and will always appreciate it.

I would also like to thank to my family and friends for their encouragement and support during my studies.

ABSTRACT

This thesis studies the output estimation methods, and investigates the solar forecast in Canberra based on K_{pv} method. Previous works has done the estimates and accuracy analysis using the data of 29 PV sites in Canberra. This thesis did the estimates analysis using the data of more than 300 sites which combined ACT school sites data and data of rooftop systems in Canberra which collected from PVoutput.org. The method of estimating output of one PV site from nearby sites by K_{pv} method was presented. It was demonstrated that the solar output can be predicted accurately with acceptable errors by using K_{pv} method. A comparison has been done between previous work and this project.

CONTENTS

List of Figures

List of Tables

Glossary of Terms

ACT	Australian Capital Territory
E_{bnc}	Beam (direct) solar irradiance
E_{dhc}	Diffuse solar irradiance
E_{ghc}	Global solar irradiance
K_{pv}	Clear sky index for photovoltaic systems
kW	Kilowatt
kWp	Kilowatt-peak rated capacity of a PV system
MAE	Mean absolute error
MBE	Normalised mean bias error
PV	Photovoltaic
RMSE	Normalised root mean squared error

Chapter 1 Introduction

1.1 BACKGROUND

Solar energy has the great potential to be a clean, environment friendly and sustainable source of electricity. Australia has the highest average solar radiation of any continent in the world and has great potential for deploying solar energy. [3] In Australia, the capacity of installed photovoltaics (PV) has been increased from around 50MW to 4 GW over the past five years (2010-2015) [1]. In 2013, 3.4 GW of residential solar capacity was installed in 1.4 million homes [2].

Though the solar energy still relies on the government support in Australia, it is quickly becoming less expensive and more attractive to consumers [2]. The growth of solar energy can be expected to be increased significantly in the further.

However, in Australia, high level of penetration of solar energy may have impacts on stability and the variations of solar power output may cause power quality issue. For large-scale solar energy, many traditional power grid is lack of flexibility to give solar energy to be integrated into the grid. As the solar output power cannot be controlled by the power system, the electricity produced by distributed solar PV generators cannot be absorbed by grids which are designed to supply electricity to consumers [16].

Research from CSIRO and the Australian Solar Institute suggests intermittency of solar systems power outputs may be difficult for electricity markets and regional distributors to integrate [4]. However, accurate forecasts and estimates of solar power output may make it possible to relieve the negative effects of intermittency.
The real time recording of PV site can help to make accurate estimate of PV output, but in Australia, many sites are unmonitored. Therefore, accurate models that estimate the power output of distributed PV systems are needed.

This thesis utilizes the K_{pv} method of Engerer and Mills (2014) [5] to compute an estimation of PV power output and will complete an analysis for the estimates.

1.2 LITERATURE REVIEW

In the literature, there are various instances of solar power output estimation methods, which can be used for generating estimates of power output from non-monitored PV systems.

Lonij 2012[6] has compared and discussed three methods for forecasting PV power output: Numerical weather model, measurements of PV power from a regional network of PV systems, and block motion analysis of ground-based camera images.
The second method, 15-minute interval PV measurements from regional network of PV system, provides better spatial and temporal resolution, PV power output can be inferred from the output of other PV systems much easier. In this method, the estimates were determined by using the clear sky index K based on historical data, the clear-sky index K was defined as

$$K_i = \frac{p_i(t)}{p_{i,clear}(t)} \tag{1.1}$$

where $p_i(t)$ is the normalized power output of site i and $p_{i,clear}(t)$ is the clear sky power output of that site.
The measured output data was from 80 residential rooftop systems in the Tucson. This method is able to provide intra-hour forecasts of quick varying cumulus clouds but requires better cloud velocity measurements.

In Lonij 2012[15], a clear-sky expectation of each system *Ci(t)* was created for analyzing difference between outage, shading, and clouds. It eliminates the effects of clouds and short outage. The expected clear-sky performance was based on the 80th percentile of the historical measurements taken from the same time of day on past 15 days. The *Ci(t)* is represented by

$$Ci(t) = Perc[\{yi(t - n * 1\ Day)\}, 80], with\ n \in \{0 \dots 15\} \tag{1.2}$$

Where *yi(t)* is the yield (kW/kW_{peak}) at time *t* for system *i*, and the Perc({i}, 80) indicates the 80th percentile of a set of numbers {i}, using this instead of an average keeps the outliers from affecting the estimates. Similarly, the global expectation *Cg(t)* and the global expectation for the real time performance *Rg(t)* were defined by:

$Cg(t) = Perc[\{Ci(t)\}, 80]$, $Rg(t) = Perc[\{yi(t)\}, 50]$ respectively.

Finally, the de-rating due to outages, partial shade and clouds were found and compared for analysis.

Golnas2011[7] created another method that estimate performance of PV systems based on measured power output from a subset of total regional PV systems. It also discussed the challenges of using pyranometers, including the cost of accurate instrumentation, optimal placement, maintenance, reliability. A Bird Performance Index (BPI) method was utilized for PV output estimation. The BPI for each day and location is defined as:

$$BPI_i(d) = \frac{E_i(d)}{I_i(d)} \tag{1.3}$$

Where $E_i(d)$ is the ratio of daily total 15-minute energy to potential total energy for each system, $I_i(d)$ is potential total insolation incident on the array, which is integrated from 15-minute clear-sky plane-of-array irradiance estimated by BPI Model, *i* is the system index and *d* is the day.

To obtain the estimate for a given site, each other system's BPI needs to multiply N-1 clear sky isolation values over the same period to obtain *N-1* estimates of local energy and then combine them into a single energy estimates.

There are two dataset in this method: the 15-minute clear sky plane-of-array irradiance values which are estimated by the Bird Clear Sky Model [7], and the 15-minute energy values from 55 systems' generation meter in New Jersey. This method was applied for estimating weekly and monthly energy generation, the result shows that 73%~96% of the weekly estimates was within 5% of measured data and 85% ~ 100% of the monthly estimates between was within 5% of measured data.[7]

Engerer and Mills [5] proposed a method for generating estimates of PV power output using one PV system to estimate the power output at another nearby PV system through their K_{PV} method. They also demonstrated this method to perform better than the Lonij and Golnas approaches.

This method consists of normalising the measured PV power output by a clear-sky power output curve and using the K_{pv} value, at one site to estimate the power output at another site. Here K_{pv} is a clear-sky index for PV. The index is the ratio of measured power output in one site over the simulated theoretical output in that site.

A series of models were used for theoretical clear-sky performance estimates. The ESRA clear-sky model [9] was used to estimate the beam, diffuse and global radiation components and then obtain the clear sky radiation estimates. The Reindl et al formula [10] is used for diffuse components transposition, and the Sandia Performance Model modeled the PV and the inverter performance.

The clear sky index K_{pv} is defined as:

$$K_{pv} = \frac{PV_{meas}}{PV_{clr}} \tag{1.4}$$

where PV_{meas} is the measured PV output of one site and PV_{clr} is the simulated clear sky PV output.

The K_{pv} calculation was demonstrated by the example of calculation of five PV sites in Canberra in September 2011, showing that the effects of orientation, tilt and system size can be removed. The shading effect was also removed from dataset as the shading was known at a particular site during the clearest days in one year.

Since the K_{pv} value is independent, as long as the clear sky curve can be generated for both known PV site and nearby site, the PV output of the nearby site estimation can be calculated by:

$$PV_{est2} = \frac{PV_{meas1}}{PV_{clr1}} * PV_{clr2} = K_{pv1} * PV_{clr2} \tag{1.5}$$

where PV_{meas1} is the measurements in the first PV site, PV_{clr1} is the simulated clear sky output of the this site. PV_{clr2} is the clear sky power output at the nearby site and the accuracy of the calculation is depend on distance, cloud cover, and system specific issues. PV_{est2} is the estimation of the nearby site.
The output data used is collected from January 2011 to December 2012, the results shows that an accurate estimate of PV output was done with RMSE values range from 3.14% to 9.77% and the MAPE values range from 3.51% to 10.8%.

The K_{pv} method was compared with previous PV power output estimation methods: the Bird Performance Index (BPI) [7] and the clearness index K of Lonij [14].
From the definition of BPI, K and K_{pv}, it can be found that the numerators in these three indices are equivalent and the denominators are different. The denominator in BPI does not include performance of PV system, the denominator in K includes the performance of PV system but cloud cover affects it and its calculation requires data for all sites. Cloud cover does not affect the denominator in K_{pv}.
From the results of these three methods for 80 selected clear-sky days in Canberra, the K_{pv} method showed a superior performance in error measurements and provided the most accurate estimates of clear-sky irradiance. At the same time, it can be applied to the systems without historical data while the K method required. Therefore, K_{pv} has a better performance in PV output estimation.

In another paper [8], the K_{pv} method was compared to the pyranometer based method of Engerer 2011 [15]. This paper (Junyan et al 2014) [8] explored both K_{pv} method and pyranometer method for estimating real-time PV energy generation and applied the methods to PV systems in Canberra. The data used for these two estimation methods testing was hourly PV output data of 26 ACT schools and half-hourly-averaged PV output at three CSIRO WERU (weather and energy research unit) monitoring sites during 2012-2013. The weather data was obtained from Australian Bureau of Meteorology Canberra Airport monitoring station (station number 070351).
The pyranometer methods used four pyranometer measurements to generate the PV energy estimates and followed the five-step process, which was discussed in Engerer 2011 [15].

The result shows that the pyranometer method has slightly better RMSE values in the range of 15-20% while the values for K_{pv} method were 15-25%. However, the accuracy of the K_{pv} method improved upon the pyranometer method when the distance between two PV sites less than 5 kilometers.

1.3 MOTIVATION

Given the problems presented by large numbers of unmonitored distributed PV systems, how can we best estimate their power output? Pyranometers provide slightly more accurate estimates, but they are sparsely located. Monitored PV systems meanwhile are abundant and well distributed. If enough of them can used, such that the separation distances are near or below 5km, how well will be able to estimate nearby PV system power output?

Based on these problems and literature review, this study intends to apply the K_{pv} method of Engerer and Mills (2013) [5] to a larger hourly dataset than was previously used, (the previous project used 29 sites [14]). This project uses 327 sites. This method will also be used to do a proper analysis for hourly, daily and monthly estimates of energy generated for all the PV sites.

To complete this analysis, a high quality dataset of PV output from solar energy installations in Canberra will be collated for testing and analysis. The data will contain PV data from as many as 71 ACT school sites data combined with up to 256 PV sites from other sites in Canberra which are available on PVOutput.org. This offers a substantial increase over the 26 used in the previous study.

1.4 RESEARCH QUESTION:

The research questions explored by this project are:.

1. For a larger dataset, how to do the data curation and make a high quality dataset from it, and then make the correct clear sky output estimates for each site?
2. Compare to previous project, by using K_{pv} method, how does the accuracy of hourly PV output estimates of one site changes and how accurate the PV output can be estimated?
3. How does distance affect the accuracy of the PV output estimates?

1.5 SCOPE

This project will use the data from 71 ACT school sites and 150 other ACT sites from 2012 to 2014, which is from the site PVOutput.org in collaboration with installer SolarHub. The main estimate method will be limited to K_{pv} method and the accuracy of this method will be tested, a comparison will be done between it and the accuracy done by previous work. This project will not do research on other models, nor will the the dataset include data of PV sites of other cities.

Chapter 2 Data

There are two types of data used in this project: power output data, and the metadata about the PV systems. Some issues have been found in these data, so each of the data requires quality control to get high quality dataset for analysis in later section. This chapter will contain details of the data and will be divided into two parts: data collection and quality control.

2.1 DATA COLLECTION

Power output data of ACT school sites and PV sites in Canberra are used in this project. The output data and other information about the PV sites need to be collected before doing quality control.

There are three main sources of the data: One is the hourly PV output data of ACT schools obtained from the Water Group website [17]. Another one is the PV output data of other sites in Canberra obtained from PVOutput.org website. And also the weather data obtained from the Australian Government Bureau of Meteorology (BoM). It is used for simulation of PV site output.

Metadata provides information about the PV sites. The metadata of ACT School sites can be obtained from manuals of each school system. The following table shows the information of PV sites is required for output estimation:

Site ID
Start date
PV model type
Rating of solar panel
Capacity
Orientation
Tilt angle
Number of modules in series

Number of parallel strings per inverter
Inverter model type
Rating of inverters
Longitude
Latitude
Altitude
Postcode
Suburb

Table 2-1 Metadata information

The geography information (longitude, latitude and altitude) was obtained from the web geography calculator by inputting the school location [18]. Compare to previous project, the number of the ACT schools sites was increased from 26 [8] to 71 now. Therefore the metadata of ACT school sites was updated. And all the information is updated to the "sites_info.csv" file. The example of the collected metadata can be seen in Appendix.

The school sites PV output was extracted from the ACT Public Schools Pulse Metering Website of Water Group, which is offering smart meters for water, electricity, solar and gas. Figure 2-1 shows the map of the entire ACT school sites, the accurate location information also can be found in Google Maps. The map provides visualization of sites and the school data used in the project is from all these sites in the map.

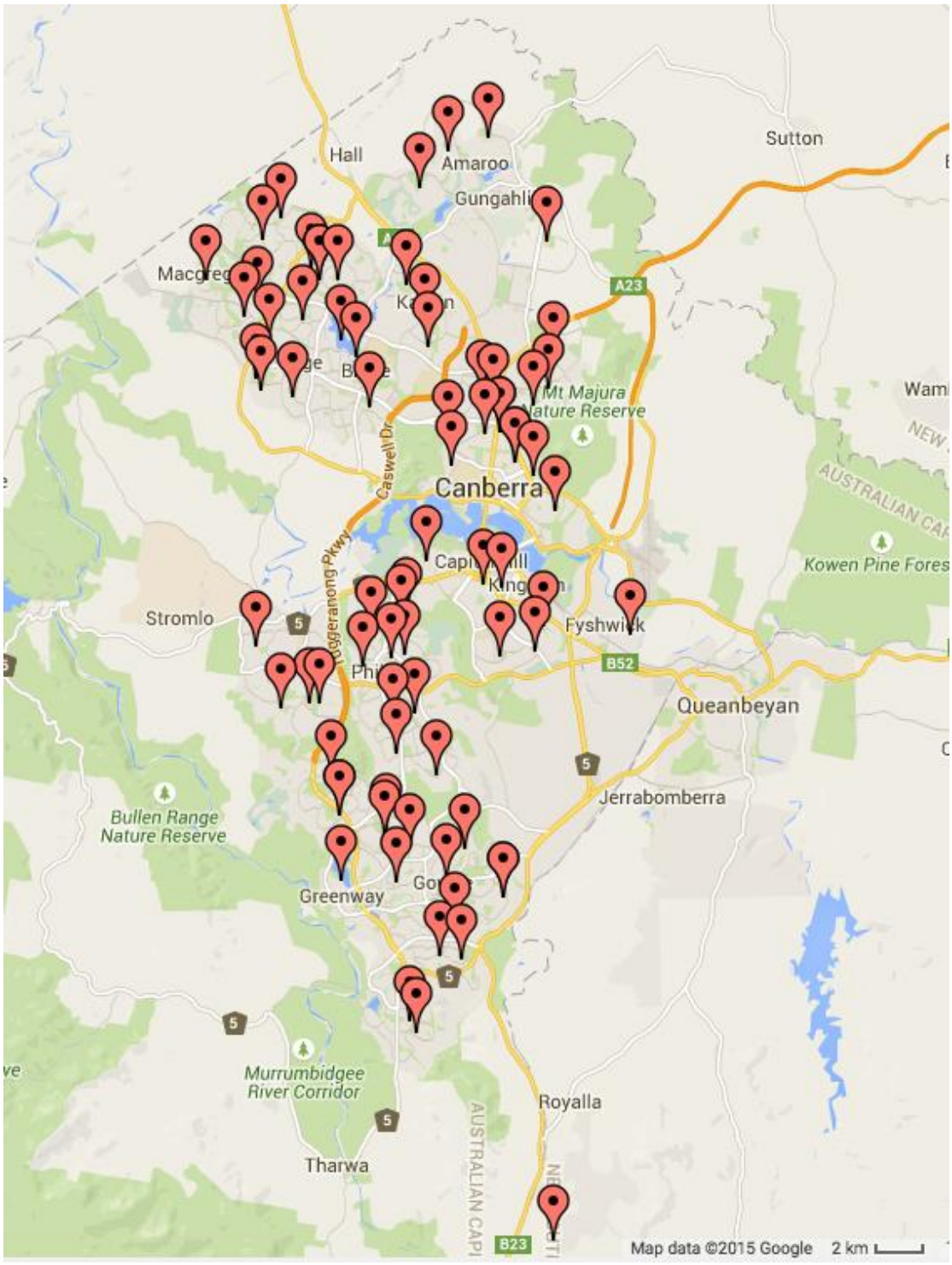


Figure 2-1 ACT School Map [1]

The following two figures are showing the daily PV output by histogram, the daily output can be seen directly from the time series plot in Figure 2-2.

Figure 2-2 shows the daily PV output from Ngunnawal Primary School on 26 Apr.2015. Up to one year of PV data can be extracted from the webpage at a time.

Figure 2-3 is the daily PV output data of site Brand St Hughes in ACT from 26 Nov 2014 to 25 Dec 2014, which was extracted from the website PVoutput.org.

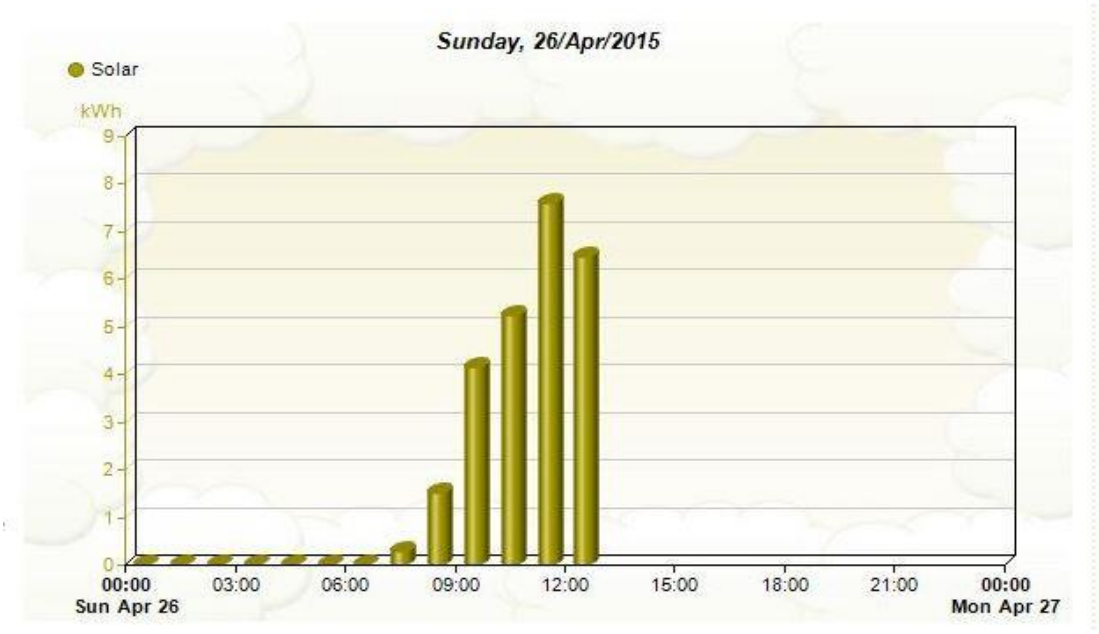


Figure 2-2 Daily PV output of Ngunnawal Primary School on 26 Apr 2015[3]

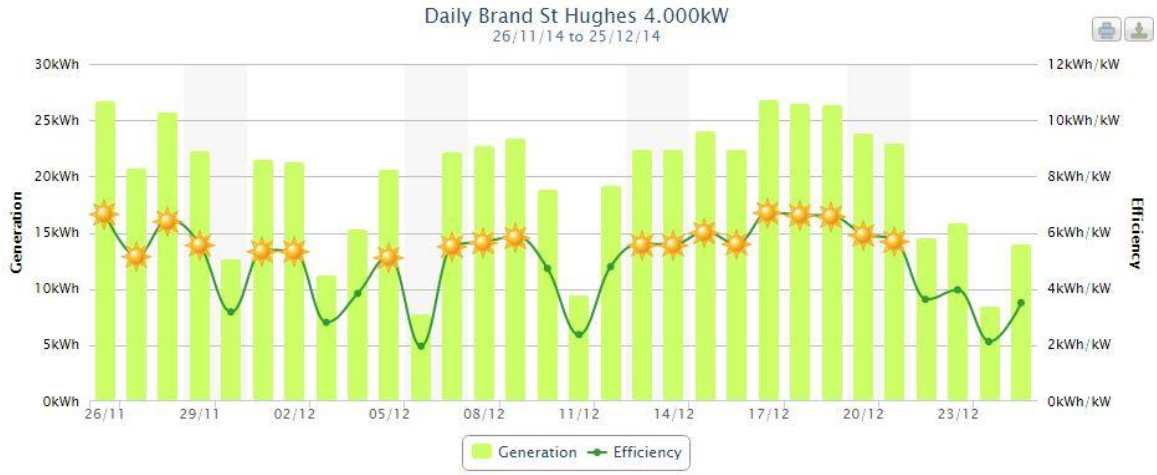


Figure 2-3 Daily PV output of Brand St Hughes from 26 Nov 2014 to 25 Dec 2014[19]

	Solar (kWh)
1/04/2015 0:00	0
1/04/2015 1:00	0
1/04/2015 2:00	0
1/04/2015 3:00	0
1/04/2015 4:00	0
1/04/2015 5:00	0
1/04/2015 6:00	0
1/04/2015 7:00	0.02004
1/04/2015 8:00	0.9519
1/04/2015 9:00	4.36371
1/04/2015 10:00	5.83665
1/04/2015 11:00	3.86772
1/04/2015 12:00	5.55609
1/04/2015 13:00	6.40779
1/04/2015 14:00	5.26551
1/04/2015 15:00	5.29056
1/04/2015 16:00	2.98596
1/04/2015 17:00	1.26753
1/04/2015 18:00	0.18537
1/04/2015 19:00	0
1/04/2015 20:00	0
1/04/2015 21:00	0
1/04/2015 22:00	0
1/04/2015 23:00	0
2/04/2015 0:00	0
2/04/2015 1:00	0
2/04/2015 2:00	0
2/04/2015 3:00	0
2/04/2015 4:00	0
2/04/2015 5:00	0
2/04/2015 6:00	0
2/04/2015 7:00	0.01002
2/04/2015 8:00	1.17234
2/04/2015 9:00	1.68837
2/04/2015 10:00	1.98396
2/04/2015 11:00	3.18135
2/04/2015 12:00	1.81863
2/04/2015 13:00	2.49999
2/04/2015 14:00	3.11622
2/04/2015 15:00	2.43486
2/04/2015 16:00	1.54809
2/04/2015 17:00	0.77655
2/04/2015 18:00	0.23046
2/04/2015 19:00	0
2/04/2015 20:00	0
2/04/2015 21:00	0
2/04/2015 22:00	0
2/04/2015 23:00	0

Table 2-2 An example of raw hourly PV output of Ngunnawal Primary School from 1 Apr 2015 to 2 Apr 2015

The extracted data was saved as csv files which can be found the examples in Table 2-2. Table 2-2 presents the hourly PV output of Ngunnawal Primary School in two days.

2.2 DATA QUALITY CONTROL

After obtaining the data of the PV sites that includes the information of the sites location, power output and the solar panel information, it is necessary to review and check the power output data to quality control the dataset.

By checking the metadata of school sites and the PV output data, it was found that there are some issues in either the metadata or PV output data; therefore, quality control was required for all the data.

Quality control was accomplished by R Studio based on R programming and nickengerer R package. R programming was used for simulation and data analysis in this project. Nickengerer R package is the package for PV data processing which is developed by Engerer [5]. It calculates the sun's position through the solar zenith and azimuth, runs the clear sky radiation models, estimate clear-sky performance of solar energy systems, and most of the project programming is based on R package by using its functions. There are two primary types of input data to this Rpackage: PV data rand solar radiation data (most often from Bureau of Meteorology). These will be referred to as PVO and RAD data respectively. For primary data, there are two supplementary data: Meteorological data and upper atmospheric data [20].

Initial data check was done by plotting the curves of theoretical clear sky and measured PV output curve, and comparing them by visual checking. Figure 2-4 is an example of initial checking for output data. The plot presents the theoretical clear sky output and the measured data, which are represented by blue and black curves respectively. From the plot, it was found that there was a time shifting between two curves and the value of the theoretical output value was quite large.

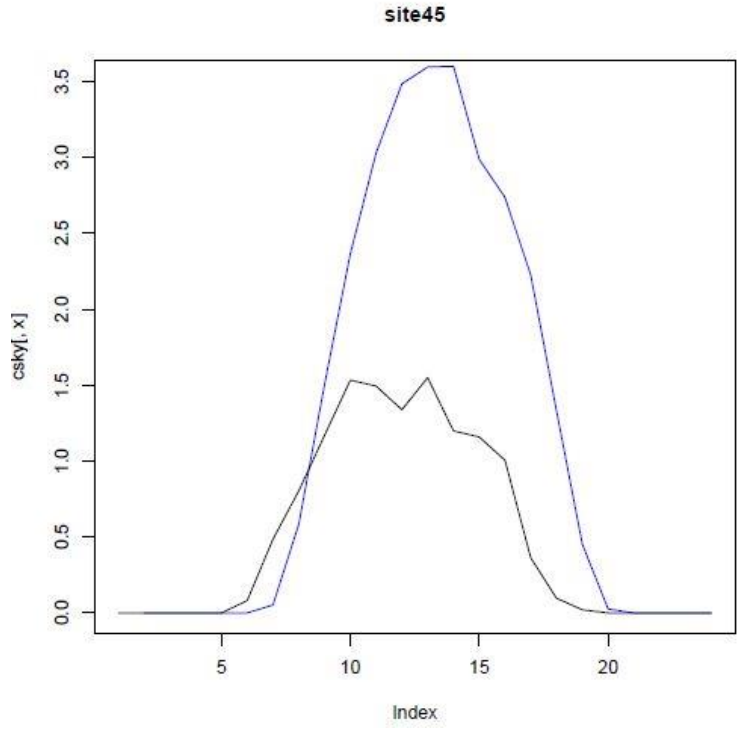


Figure 2-4 Clear sky curve and measured output curve on 15 November 2013(DST) for site45 (site id 1955), x axis is dimensionless time represented by index and y axis is normalised power output kW/kWp

2.2.1 Time stamping issue

By visual checking, firstly the time-stamping issue were identified. In some PV sites, the recorded time was one hour later than the actual time during the Day Light Saving days, this may be caused by the clocks in those sites were not adjusted when it was coming to DST time. In RStudio, the measured PV output values and the calculated theoretical clear sky value of each site was plotted for comparison.

Simulated clear sky model curve and measured PV output curve of each site in different days are plotted in R for comparison.

Figure 2-5 presents an example of simulated clear sky model curve and measured PV output curve for site3 (site id 1851) on 18 February 2014. The blue curve is clear sky curve and the black one represents the measured data. The x-axes represent 24 hours in one day while the y-axes represent the power output (kW/kWp). It was seen obviously from this plot that the black curve was one hour behind the blue one and was not a smooth curve, therefore, the time stamp of this site was one hour late and had the time shifting issue, it was not clear sky at some time of the day as well. Figure 2-5 is the plot of site 3 in a non-DST day, the two curves are almost overlapping, which means the time stamp in this day is correct and it is a clear sky day. The same method has been applied to other sites. The sites which have the same issue were identified and need to do the time stamp correction in quality control.
A list of school names and their site id can be found in the Appendix.

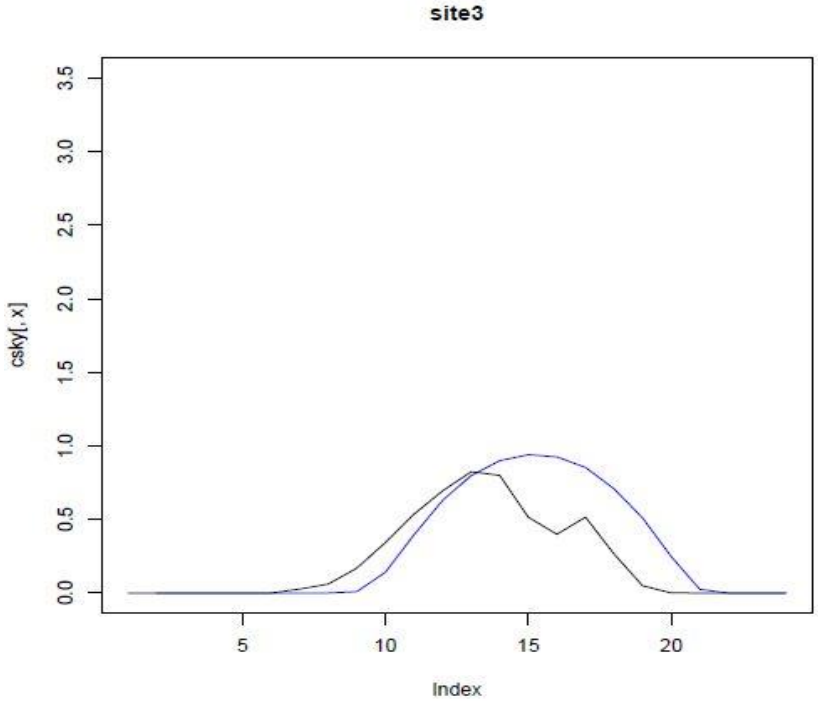


Figure 2-5 Clear sky curve and measured output curve on 18 February 2014(DST), x axis is dimensionless time represented by index and y axis is normalised power output kW/kW$_p$

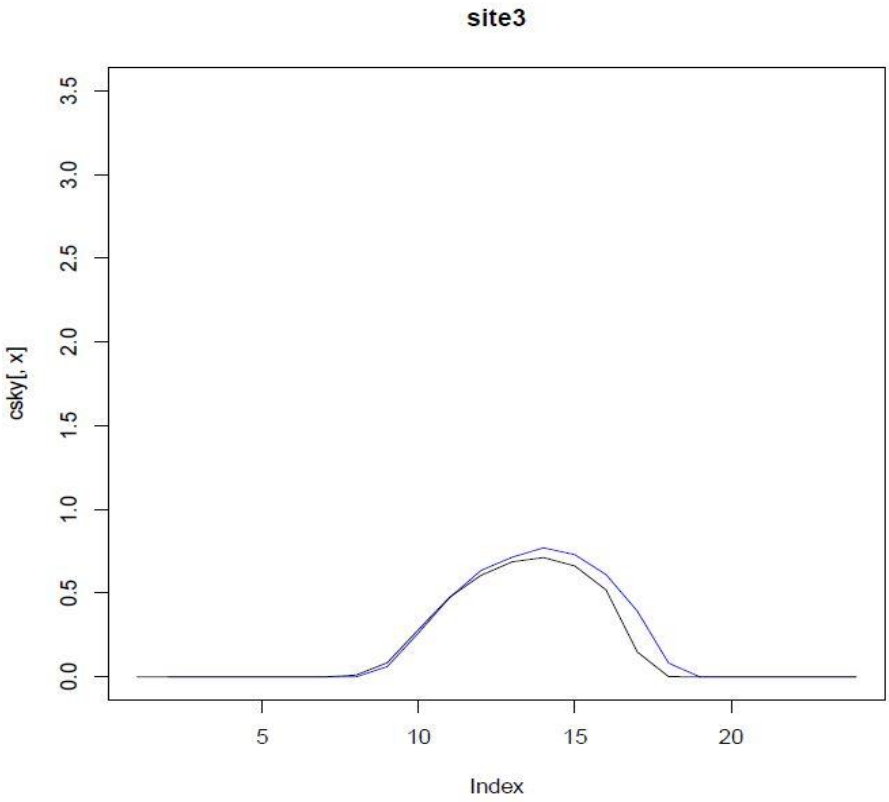


Figure 2-6 Clear sky curve and measured output curve on 22 May 2014(non-DST), x axis is dimensionless time represented by index and y axis is normalised power output kW/kW$_p$

According to the time stamping issues identified, a time stamp correction was applied to the sites with this issue by going through all recorded DST days.

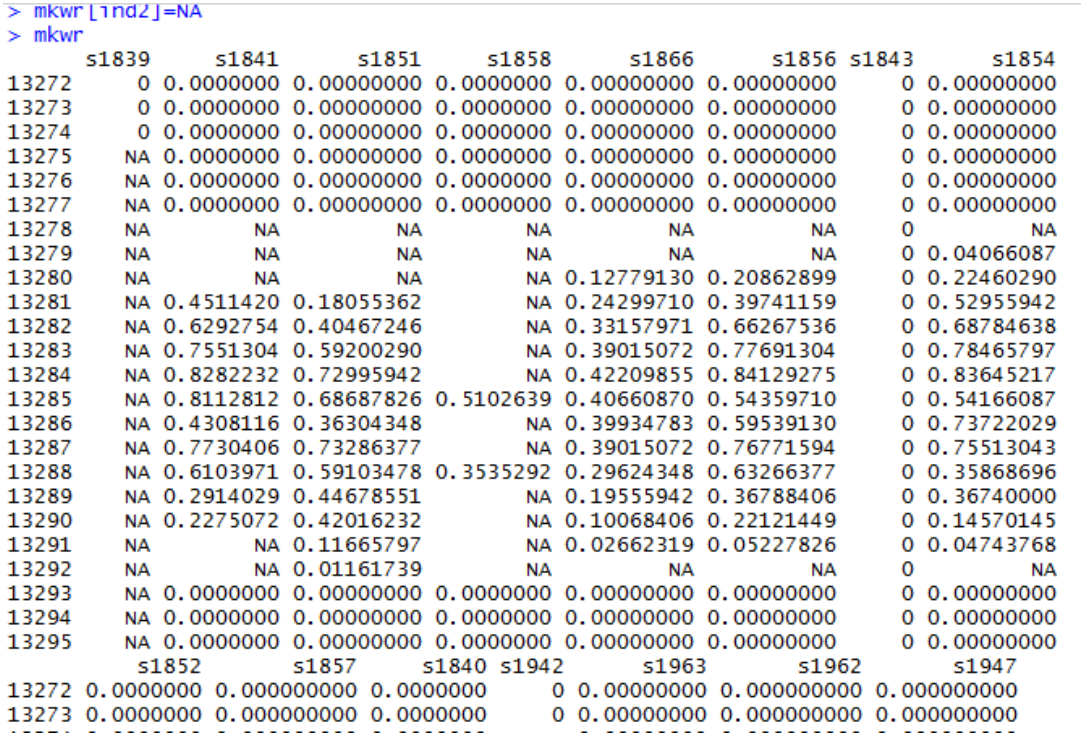
2.2.2 Outlier in PV output data

Secondly, by visual checking for the PV output data, it can be found that some output is physically too large or too small, or no value recorded, so it is necessary to set limits for the output data. Both the physical limits and statistical limits were set.

For clear sky test, the physical lower limit was set to 0.2 kW/kW_p, this was to remove the very small output values, also before sunrise and after sunset, the output values were quite small that can be ignored. The physical upper limit was set to 1.2 kW/kW_p so that if there was measured PV output data of a site quite large and the measured data was more than 20% of the calculated value, the output data would be removed and set to NA.

For statistical test, the standard deviation values were calculated to check whether the measured value was in a reasonable range, if not, it would be removed. Both the statistical upper limit and lower limit were set to 2, this means two standard deviation of output was used, if the output distribution was greater or smaller than 2 standard deviation, it would be regarded as an outlier and be removed.

Figure 2-7 is an example of removing abnormal values by the physical limits. The data used is the measured data on 10 Jan 2014, the output values which were greater than 20% of the theoretical values, they were removed and set to NA.



```
> mkwr[ind2]=NA
> mkwr
       s1839     s1841      s1851     s1858      s1866      s1856 s1843       s1854
13272      0 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13273      0 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13274      0 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13275     NA 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13276     NA 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13277     NA 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13278     NA        NA         NA        NA         NA         NA     0          NA
13279     NA        NA         NA        NA         NA         NA     0 0.04066087
13280     NA        NA         NA        NA 0.12779130 0.20862899     0 0.22460290
13281     NA 0.4511420 0.18055362        NA 0.24299710 0.39741159     0 0.52955942
13282     NA 0.6292754 0.40467246        NA 0.33157971 0.66267536     0 0.68784638
13283     NA 0.7551304 0.59200290        NA 0.39015072 0.77691304     0 0.78465797
13284     NA 0.8282232 0.72995942        NA 0.42209855 0.84129275     0 0.83645217
13285     NA 0.8112812 0.68687826 0.5102639 0.40660870 0.54359710     0 0.54166087
13286     NA 0.4308116 0.36304348        NA 0.39934783 0.59539130     0 0.73722029
13287     NA 0.7730406 0.73286377        NA 0.39015072 0.76771594     0 0.75513043
13288     NA 0.6103971 0.59103478 0.3535292 0.29624348 0.63266377     0 0.35868696
13289     NA 0.2914029 0.44678551        NA 0.19555942 0.36788406     0 0.36740000
13290     NA 0.2275072 0.42016232        NA 0.10068406 0.22121449     0 0.14570145
13291     NA        NA 0.11665797        NA 0.02662319 0.05227826     0 0.04743768
13292     NA        NA 0.01161739        NA         NA         NA     0          NA
13293     NA 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13294     NA 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
13295     NA 0.0000000 0.00000000 0.0000000 0.00000000 0.00000000     0 0.00000000
        s1852       s1857     s1840 s1942      s1963       s1962       s1947
13272 0.0000000 0.000000000 0.0000000     0 0.00000000 0.000000000 0.000000000
13273 0.0000000 0.000000000 0.0000000     0 0.00000000 0.000000000 0.000000000
```

Figure 2-7

After data quality control finished, the data of ACT school sites was combined with the data of other PV output sites in Canberra, this was also done in the R, and then the assembled data could be used for analysis in the following sections.

2.3 SUMMARY

All the ACT school sites data collected from the school manual was recorded in the "sites_info.csv" file, the file was updated from 26 school sites to 71. The information of school sites was formatted into "all_sites" format in R package for integration.

The quality control correction was applied for ACT schools data based on quality control functions written in R. The quality control did the correction for time shifting issue, set the clear sky limit that if the measured output of one site is greater than 20% of the clear sky output, the value of that site will be removed to limit the max value. In the statistical test, the statistical limit was set, abnormal values were removed to make the data more accurate.

Chapter 3 Method

This chapter will describe the methods that are used in this project. It includes modelling method for clear sky simulation, clear sky index normalization and error analysis method.

The method of estimating theoretical PV output for one PV site is to use the measured PV output from a nearby site and apply the clear sky index K_{PV} [1] to generate the theoretical estimates. Therefore, in order to generate the estimates for a given PV site, theoretical clear sky performance need to be estimated first and then calculate the clear sky index.

3.1 MODELLING

According to Engerer and Mills 2013 [5], two primary steps were used to estimate the clear sky performance. First was to calculate the available transposed clear sky radiation and second was to simulate the PV system based on the available radiation.

The theoretical clear sky PV output was computed with ESRA clear sky beam and diffuse radiation model. The ESRA model was used to estimate beam, diffuse and global radiation components. Then by the Reindl transposition formulae, the diffuse components was transposed to a tilted PV system surface [21]. Then the PV and inverter performance were modelled via Sandia PV [22] and Inverter Performance model [23] by using the available radiation as input. And then integrated the PV power output estimate to energy estimates.

The total amount of clear sky radiation incident upon a tilted surface (E_{gtc}) must be determined to generate the clear sky performance estimates. This was comprised of the tilted beam radiation (E_{btc}), which was based on system geometry calculation and the tilted diffuse radiation (E_{dtc}), which was estimated by Reindl transposition model. Also, the ESRA model computed the clear sky normal beam component (E_{bnc}) and the horizontal diffuse component (E_{dhc}).

The normal beam radiation E_{bnc} calculated by ESRA model is given by:

$$E_{bnc} = E_n * \exp(-8.662 * T_L * AM_r * \delta_R) \tag{3.1}$$

And the horizontal diffuse radiation component (E_{dhc}) is computed by:

$$E_{dhc} = E_h * T_R(T_L) * F_D(\theta_z, AM_r) \tag{3.3}$$

The tilted beam radiation therefore can be computed by:

$$E_{btc} = E_{bnc} * \cos(\theta_\alpha) \tag{3.4}$$

where θ_α is the angle of incidence for tilted surface and determined by its tilt and azimuth orientation by:

$$\cos(\theta_\alpha) = \cos(\alpha_m) * \cos(\theta_z) + \cos(\varphi - \varphi_m) * \sin(\alpha_m) * \sin(\theta_z) \tag{3.5}$$

where φ is the solar azimuth angle and θ_z is the solar zenith angle.

Since only the clear sky condition was considered, the Reindl transposition model was chosen for diffuse radiation transposition. In Gueymard[24], the Reindl transposition model has a quite good performance under clear sky condition and has a simply computation. Therefore, the diffuse radiation available to a tilted surface is given by:

$$E_{dtc} = E_{dhc} * \{\left[(1 - A_i) * \left(1 + \frac{2}{\alpha_m}\right)\right] * \left[1 + f * \sin^3(\frac{\alpha_m}{2})\right] + A_i * R_b\} \tag{3.6}$$

Where A_i is the anisotropy index and R_b is the geometric factor.
Then the total amount of clear sky radiation incident upon a tilted surface (E_{gtc}) was computed by summing up the tilted beam radiation and the tilted diffuse radiation.

The obtained metadata which includes geographic information (longitude, latitude, altitude) of PV systems can be used for solar zenith angle and azimuth angle calculation. Then clear sky radiation was computed with these values.

By using the clear sky radiation as the input to the Sandia Performance Model, the simulated PV power output was generated. The theoretical clear sky output was

transferred to theoretical hourly clear sky energy output via numerical integration. Finally, the clear sky energy output was ready for K_{pv} calculation.

3.2 CLEAR SKY INDEX VALUES

The clear sky index K_{pv} normalizes measured values by that PV system's clear sky output.

In Engerer and Mills 2013 [5], the clear sky index is defined as:

$$K_{pv} = \frac{PV_{meas}}{PV_{clr}} \tag{3.7}$$

where thePV_{meas} is the measured PV output of one site and PV_{clr} is the simulated clear sky PV output.

And for estimating PV output of a nearby PV system, the estimated PV output can be calculated by:

$$PV_{est2} = \frac{PV_{meas1}}{PV_{clr1}} * PV_{clr2} = K_{pv1} * PV_{clr2} \tag{3.8}$$

where PV_{clr2} is the clear sky power output at the nearby site and the accuracy of the calculation is depend on distance, cloud cover, and system specific issues.

The calculation and simulation was accomplished with RStudio. Our obtained meta data and PV output data were used to demonstrate the K_{pv} method. Firstly, the measured PV output, simulated clear sky output and K_{pv} values were calculated and then plotted respectively.

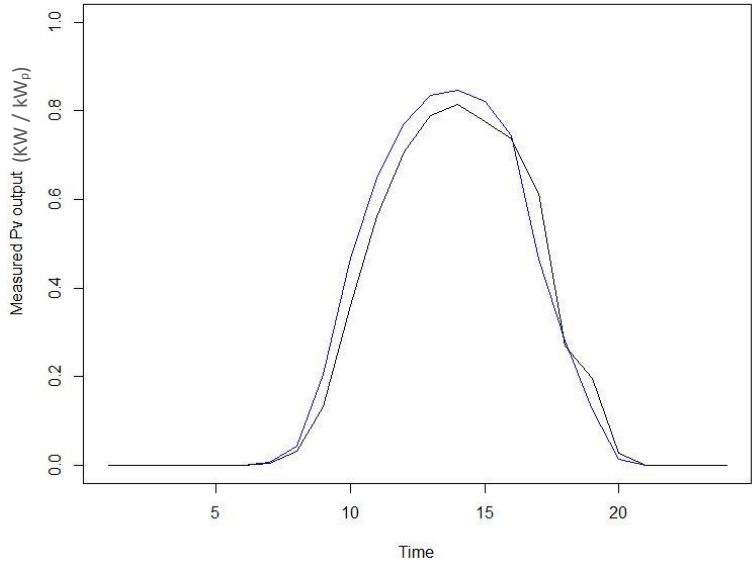


Figure 3-1 Measured clear sky PV output for Belconnen High School (site id1942) and Forrest Primary School (site id1949) on 15 January 2014, with black line and blue line respectively, x axis is local time and y axis is the output (kw/kW$_p$)

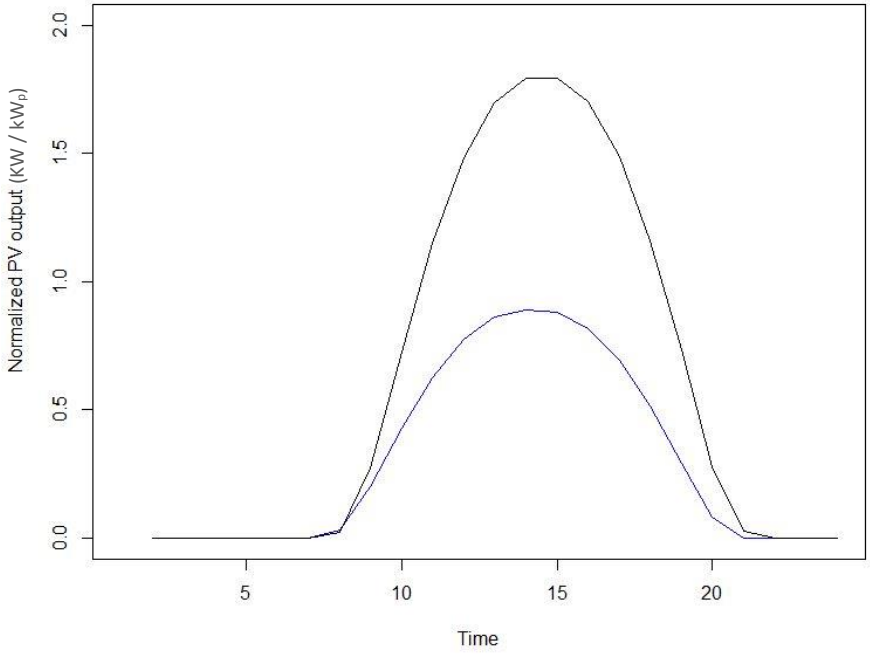


Figure 3-2 Simulated clear sky PV output for Belconnen High School (site id1942) and Forrest Primary School (site id1949) on 15 January 2014, x axis is local time and y axis is the output (kw/kW$_p$)

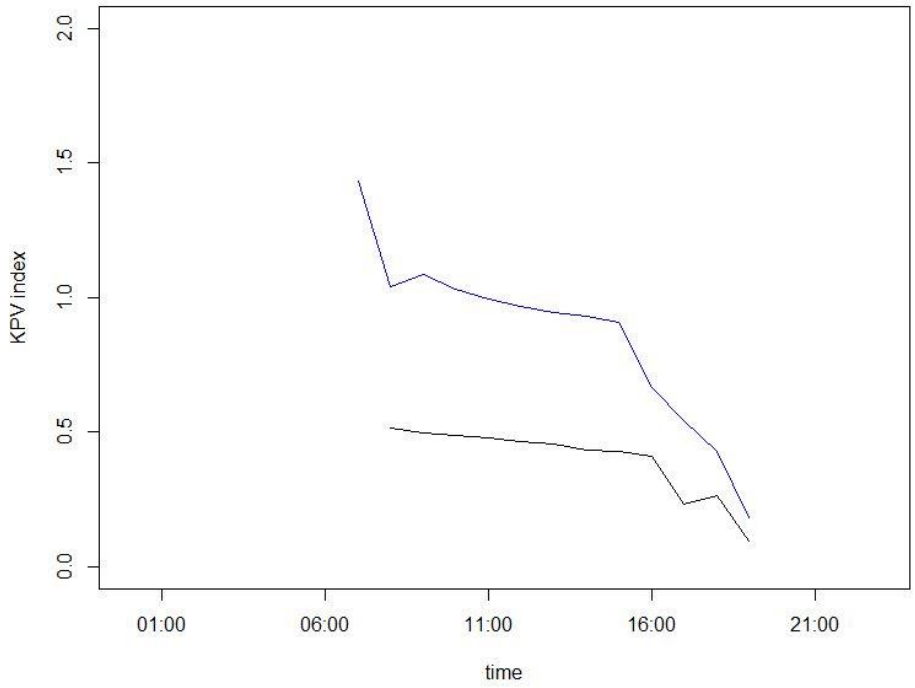


Figure 3-3 K_{pv} calculations for Belconnen High School (site id1942) and Forrest Primary School (site id1949) on 15 January 2014, x axis is local time and y axis is the K_{pv} index

Figure 3-1 shows the measured PV output curve for Belconnen High School and Forrest Primary School in Canberra on 15 January 2014, the black curve is output of Belconnen High School and the blue one is Forrest Primary School. Figure 3-2 is the simulated clear sky output curve for these two sites, the black one is Belconnen High School and the blue one is Forrest Primary School. Figure 3-3 is an example of K_{pv} values which is the ratio of measured output and simulated output of the two sites on this day, the black and blue curves represent Belconnen High School and Forrest Primary School respectively. During clear sky period which is from around 9 a.m. to 4 p.m., the K_{pv} value for the second site is around one, the curves of both sites are smooth and steady.

Following 3 plots show the examples of PV output for these two sites on a cloudy day. Figure 3-4 and Figure 3-5 are the curves for measured and theoretical solar output respectively on 16 March 2014. Black curve is Belconnen High School while

the blue one is Forrest Primary School. Both measured output curves and K_{pv} curves are not smooth, and K_{pv} curves have a rapid negative slope with redundancy.

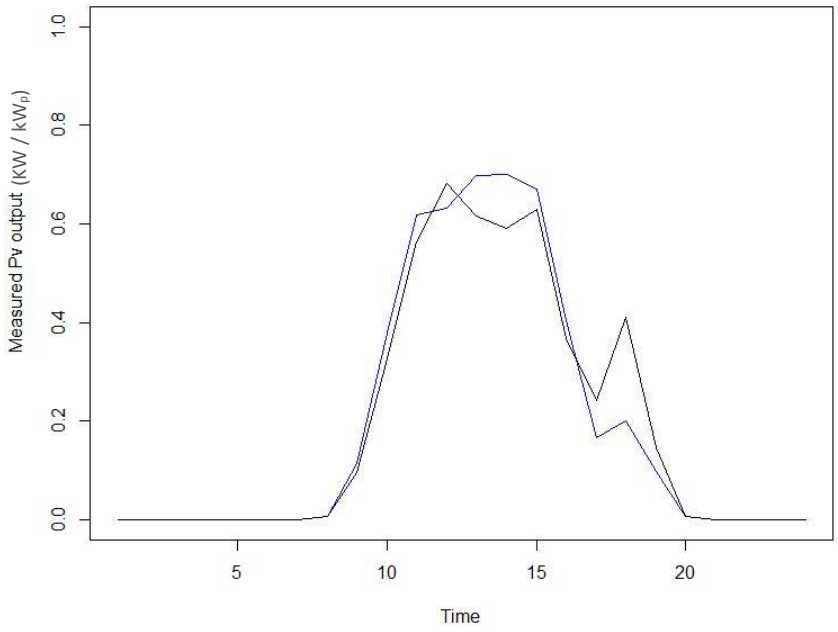


Figure 3-4 Measured clear sky PV output for Belconnen High School (site id1942) and Forrest Primary School (site id1949) on 16 March 2014, x axis is local time and y axis is the output (kw/kW$_p$)

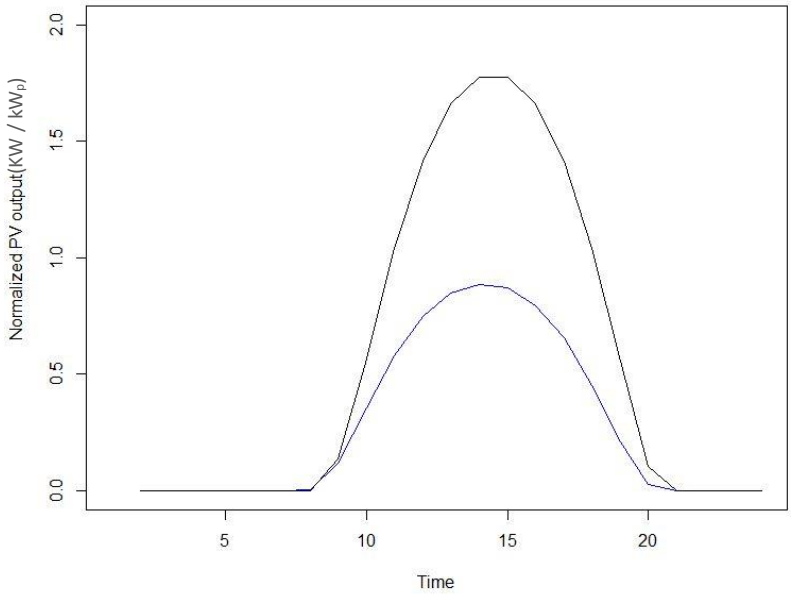


Figure 3-5 Simulated clear sky PV output for Belconnen High School (site id1942) and Forrest Primary School (site id 1949) on 16 March 2014, x axis is local time and y axis is the output (kw/kW$_p$)

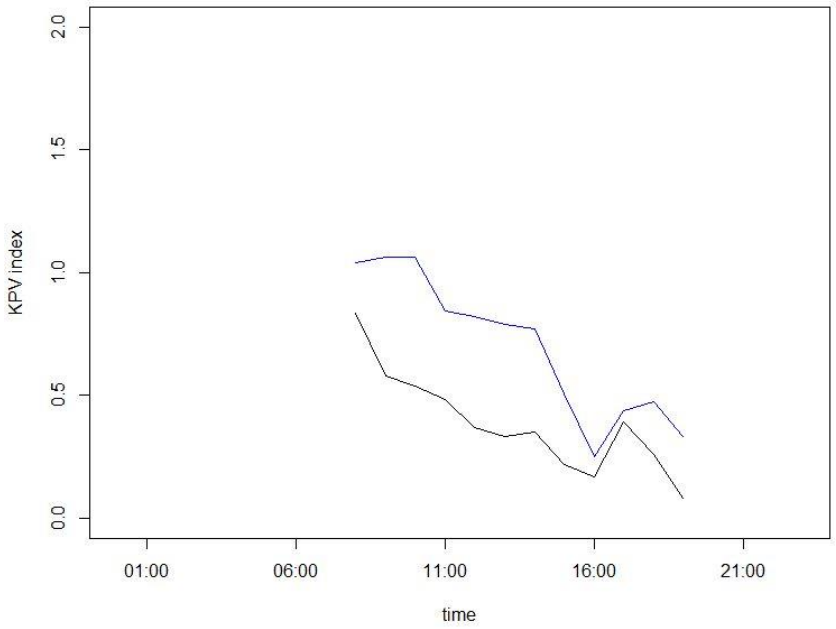


Figure 3-6 K_{pv} Calculations for Belconnen High School (site id1942) and Forrest Primary School (site id1949) on 16 March 2014, x axis is local time and y axis is the K_{pv} index

3.3 ERROR ANALYSIS

By applying the K_{pv} method, the errors of the calculation also need to be determined in order to analyze the accuracy of the method.

Four error metrics are used to do the error analysis for the data as the same in Junyan [14]: the Root Mean Squared Error (RMSE), Mean Bias Error (MBE), the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE).
The RMSE and the MBE values have been the standard for assessing clear sky radiation model.

The normalized RMSE is given by

$$\text{RMSE} = \frac{1}{\overline{O}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{3.9}$$

Where P_i is the estimated PV output and O_i is the observed value, i is the time, n is the number of observed sites, and $\overline{O_\iota}$ is the averaged observed value. The RMSE is used to measure the differences between P_i and O_i, and it sums the errors in predictions. It represents standard deviation of the differences; a large positive RMSE implies a big deviation [25].

The normalized MBE is given by

$$\text{MBE} = \frac{1}{n\overline{O}}\sum_{i=1}^{n}(P_i - O_i) \tag{3.10}$$

The MBE is used to measure the bias of the estimates, it represents the average deviation of the differences between the estimated values and observed values. It is suitable for analyzing long-term performance of models. A positive MBE value indicates the amount of over estimation and a negative MBE value indicates the amount of under estimation in the predicted energy [25].

The normalized MAPE is given by

$$\text{MAPE} = \frac{1}{\bar{n}}\sum_{i=1}^{n}\left|\frac{(P_i - O_i)}{O_i}\right| \tag{3.11}$$

The MAPE is used to measure the overall absolute error of a given model. It measures the difference between the estimated values and observed values over the observed value and it indicates the accuracy of the model as well [25]. But it will cause quite large values if the denominator, which here is the observed value, becomes very small. So for a small observed value, even the error can be large, it does not always imply a low accuracy for the model.

And the MAE is given by:

$$\text{MAE} = \frac{1}{\bar{n}}\sum_{i=1}^{n}|(P_i - O_i)| \tag{3.12}$$

The MAE is used to measure the predicted error. It is the average of the absolute errors as the name states. It is also a common measure for estimates.

Chapter 4 Results

This analysis is aimed at answering the following two research questions:

Research question 2: Compare to previous project, by using K_{pv} method, how does the accuracy of hourly PV output estimates changes and how accurate the PV output can be estimated?

Research question3: How does distance affect the accuracy of the PV output estimates?

From the literature, it is already known that different factors such as distance, sky conditions have different effects to the PV output and therefore affect the model performance.

To identify the effects of these factors in K_{pv} model, and find out how the model performance changes with these factors, the accuracy of estimates will be investigated.

4.1 ESTIMATES ACCURACY

Firstly, the estimates need to be calculated. In order to calculate the estimated clear sky PV output for all the PV sites, the method described in Chapter 3 was used. Estimates of PV output can be generated by K_{pv} method of Engerer and Mills 2013. Large data set and calculations are completed by using the RStudio. The data used for calculation is hourly data.

4.1.1 Yearly analysis

After the estimated PV output was calculated, it was compared to measured data to measure the accuracy and correlation.

A comparison between estimates and measurements was plotted as shown in Figure 4-1. In Figure 4-1, the measured PV output is x-axis and y-axis represents the estimated output. The unit for hourly estimates and measurements are $kW/kW_{p.}$ The data through a whole year which is from July 2013 to July 2014 was plotted. The red line is regarded as a base line which indicates the perfect estimation of accuracy. The closer to this line, the more accurate the data point is. Additionally, if the data point is above the line, it is over estimated; and if the data point is under the line, it is under

estimated. In the plot, the actual data points were replaced by density of points, so it can be seen more clearly that the darker colour area has a larger point density, which means the darker area has more data aggregation.

From the plot, when the measurements and estimates are smaller than 0.2 kW/kW$_p$, there is the highest density region in the graph, which represents that quite a lot of measured and estimated outputs are ranging from 0 to 0.2 kW/kW$_p$. At top right corner of the graph, there is another higher density area, showing lots of PV output estimates and measurements are in the range of 0.7 to 0.9 kW/kW$_p$.

Near the perfect prediction line area, the density is higher than the other area, which means each point are quite close to each other, and the model has a good performance of estimation. The calculated RMSE between estimates and measurements for these data is 0.224 kW/kW$_p$ and 53.9%. The absolute RMSE value 0.224 kW/kW$_p$ is a relatively small value, but 53.9% seems large, and this will be discussed later.

And the MAPE is 64.7%, the MBE is 2.86%. The 2.86% MBE indicates that in this estimation, the amount of over estimation is small, while the 64.7% MAPE indicates the difference between the observed and predicted values. Though the MAPE seems large, the absolute RMSE errors between estimation and measurement is small. When considering the accuracy of the prediction, usually the RMSE values are used. Therefore, the output is estimated accurately.

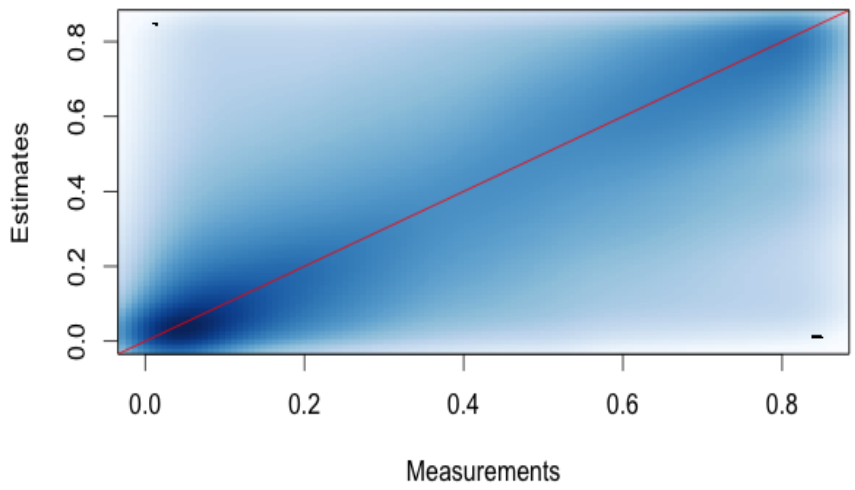


Figure 4-1 Measurements and estimates plot of all the sites from July 2013 to July 2014, the unit for power output is kW/kW$_p$

4.1.2 Monthly analysis

In this section, the monthly analysis will be repeated using hourly PV output data of all sites. Figure 4-2, 4-3, 4-4, 4-5 present the estimates versus the measurements plots for February, May, August, and November in 2013 respectively.

From the plots, it can be seen more clearly that the overall spread of estimate/measurement is along the perfect prediction line, which means the accuracy for monthly data is in an acceptable range. The highest density occurs at the bottom left area, in the range of 0 to 0.1kW/kW$_p$, for normalised measurements and estimates. The plot of February has the clearest view of density distribution along the perfect prediction line, so the accuracy for this month is be better than the other three months. But overall, the difference between each month is not much.

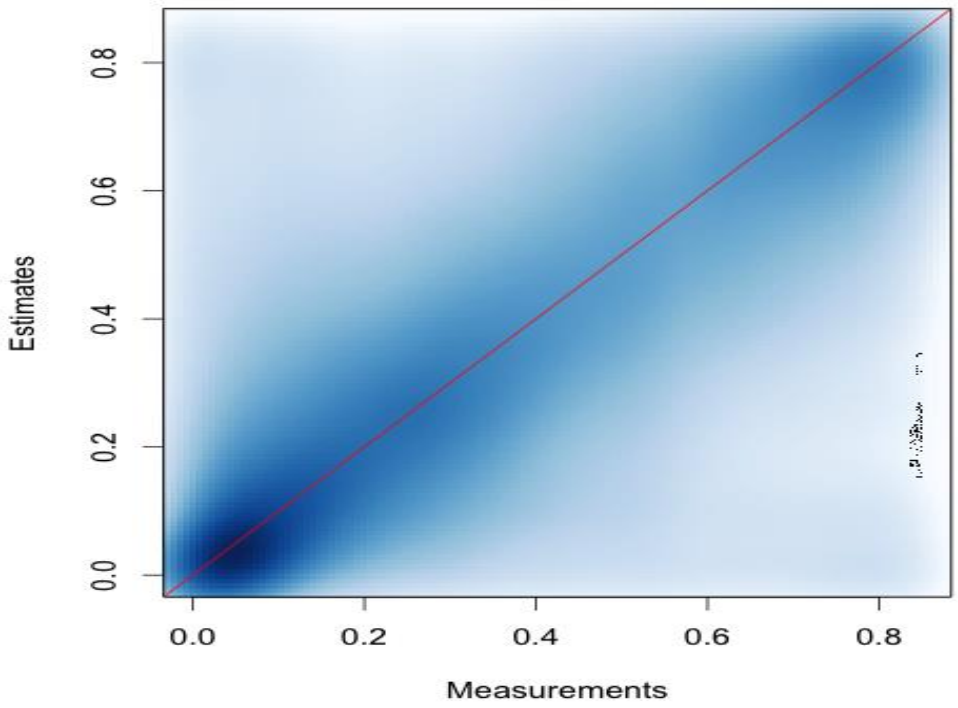


Figure 4-2 Measurements and estimates plot of all the sites in February 2013, the unit for output is kW/kW$_p$

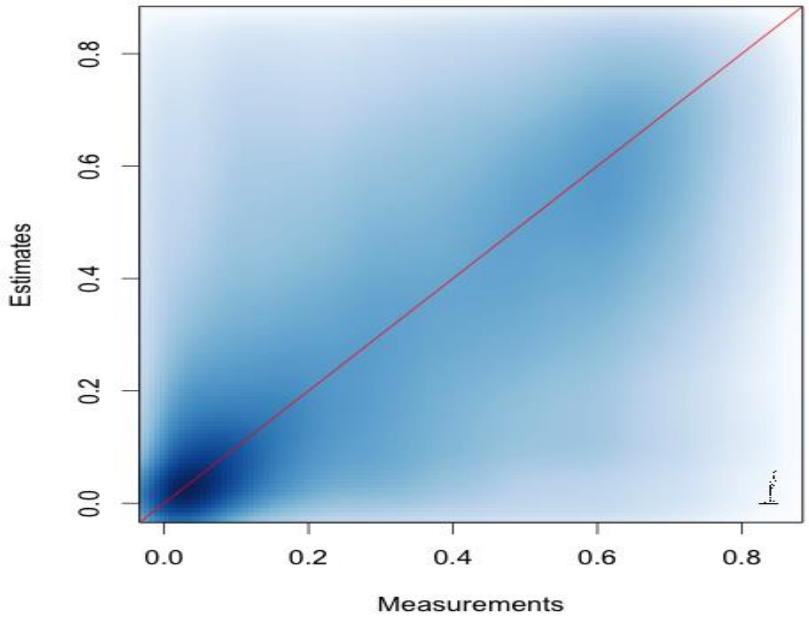


Figure 4-3 Measurements and estimates plot of all the sites in May 2013, the unit for output is kW/kW_p

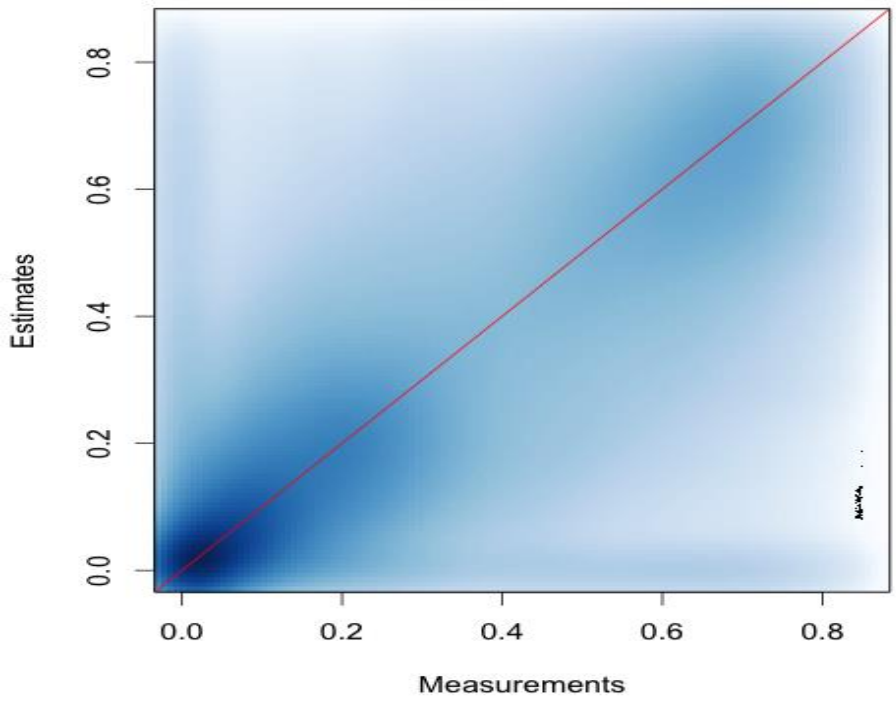


Figure 4-4 Measurements and estimates plot of all the sites in August 2013, the unit for output is kW/kW_p

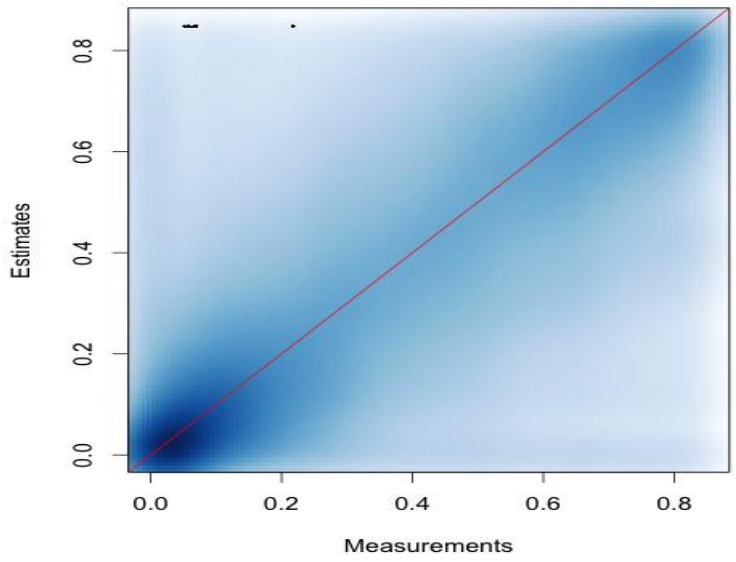


Figure 4-5 Measurements and estimates plot of all the sites in November 2013, the unit for output is kW/kW$_p$

The calculated errors have been shown in table 4-1.

Month	RMSE (kW/kW$_p$)	MBE(%)
February 2013	0.178	1.8
May 2013	0.209	2.525
August 2013	0.220	1.916
November 2013	0.251	3.34

Table 4-1 Summary of hourly accuracy analysis for four months in 2013

From the table, it can be found that for the comparison of these months in different seasons in 2013, the absolute RMSE value of February is the smallest one which is 0.178 kW/kW$_p$. As the same as the accuracy found from the plots, it can be conclude that the accuracy in February is the best. Compare the MBE values of four months; the MBE of February is the smallest which is 1.8%. The range of the MBE is from 1.8% to 3.34%, these are small values meaning the overestimation is not too much

and is in an acceptable range. Given the smallest RMSE and MBE values, February has the best t performance for the model, but there is not much difference between these four months. The RMSE and MBE values for four months are small.

4.1.3 Daily analysis

The daily comparison and calculation will be done to further understand the performance of the model.

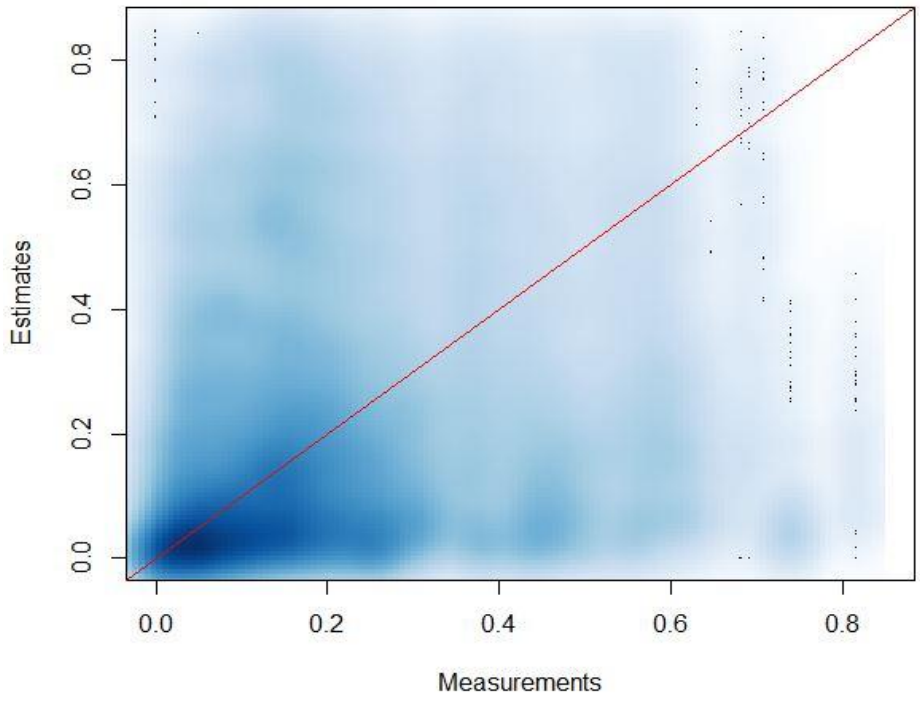


Figure 4-6 Measurements and estimates plot of all the sites from 01 February 2013 to 02 February 2013, the unit for output is kW/kW_p

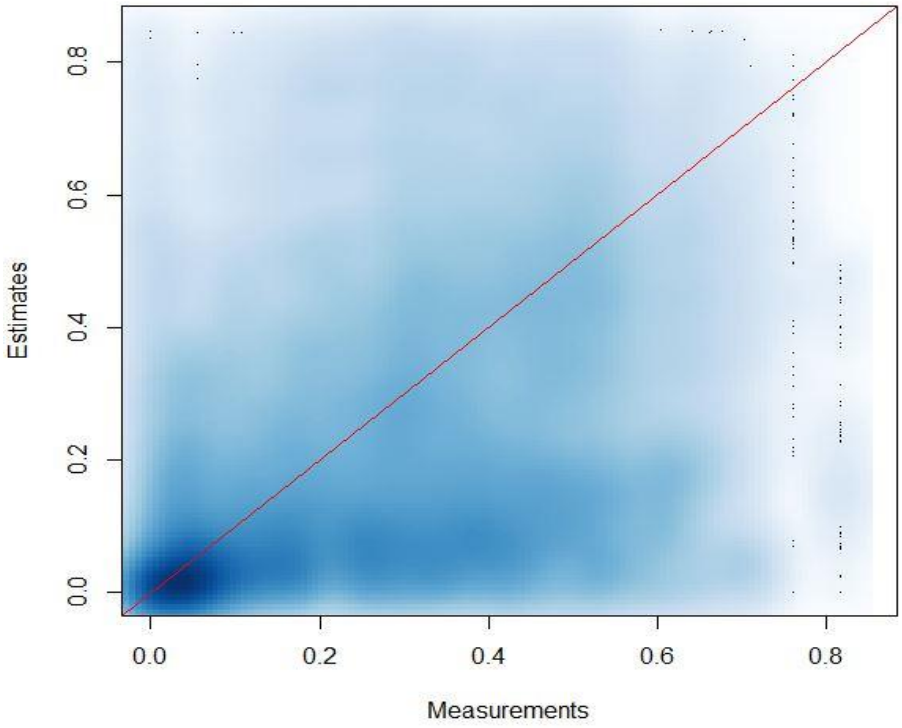


Figure 4-7 Measurements and estimates plot of all the sites from 01 May 2013 to 02 May 2013, the unit for output is kW/kW_p

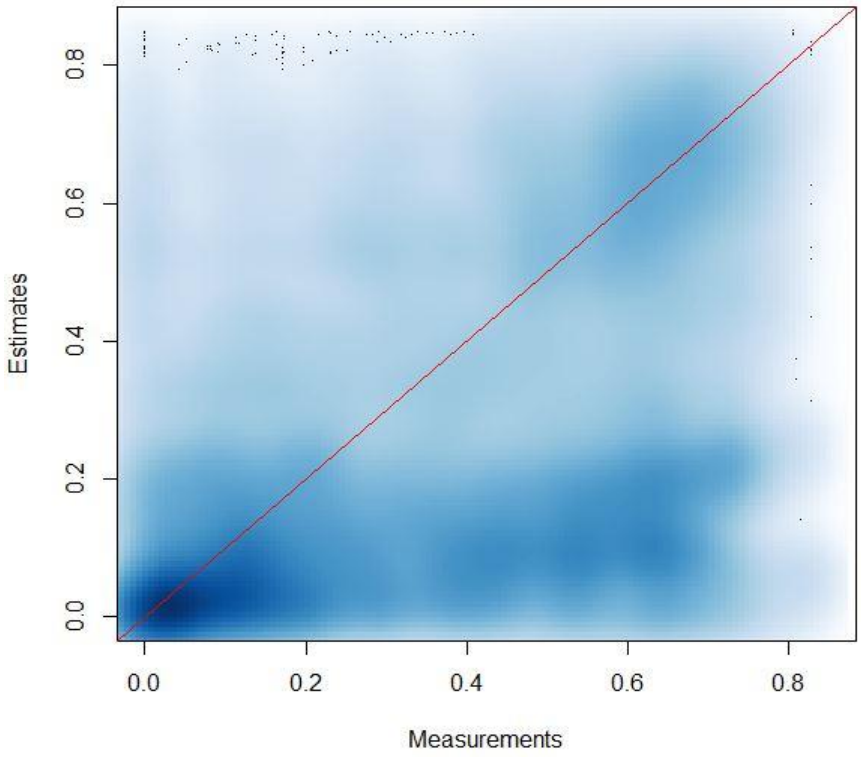


Figure 4-8 Measurements and estimates plot of all the sites from 01 August 2013 to 02 August 2013, the unit for output is kW/kW_p

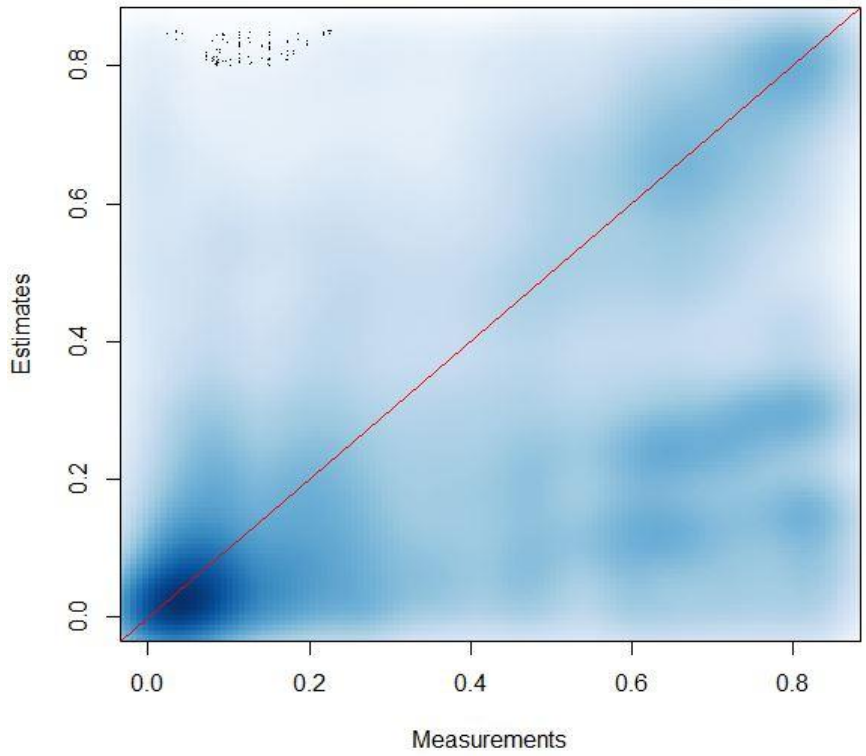


Figure 4-9 Measurements and estimates plot of all the sites from 01 November 2013 to 02 November 2013, the unit for output is kW/kW_p

Figure 4-6, 4-7, 4-8, 4-9 present plots for one day in February, May, August, and November of 2013 respectively. Given the plots above, it can be seen that there is no much difference between each day. Compare to the yearly and monthly plots, the distribution of density is more scattered.

According to the plot, in the area of bottom left, the density is higher, this means that many measured and estimated values are in the range of 0 to 0.1 kW/kW_p. The estimated data in this range is more accurate as it is close to the red line. It also can be found that under estimation points are almost the same as over estimation, since the density distributed almost the same over/under the red line.
The accuracy was found intuitively from the plot as well. It can be concluded that the smaller the measured/estimated value is, the more accurate.

By taking calculation of errors, the accuracy can be understood more clearly. And the RMSE and MBE values are calculated in the following table:

Day	RMSE (kW/kW_p)	MBE (%)
01 February 2013	0.178	1.8
01 May 2013	0.209	2.525
01 August 2013	0.220	1.916
01 November 2013	0.251	3.34

Table 4-2 Summary of hourly accuracy analysis for four different days in 2013

The calculated RMSE between estimates and measurements for February is 0.258 kW/kW_p and 253%. The absolute RMSE value 0.258 kW/kW_p is similar to the RMSE value calculated using one year data, but 253% is quite large and is 5 times the yearly value.

For the MAPE and the MBE value, they are 133.6% and 7.11% respectively. The MAPE value is greater than that of yearly value, and the MBE of this day is close to the yearly value.

For remaining three days, the results show no much difference among them. Therefore, the accuracy for days in different seasons is almost the same.

Overall, the RMSE and MBE results are in the reasonable range as expected, and the output can be estimated accurately.

4.2 THE EFFECT OF DISTANCE TO ACCURACY

To investigate of the effects of different factors to model accuracy, we focused on the factor of distance.

Figure 4-10 provides the plot of RMSE values for changing distances for hourly estimates of all sites through one year.

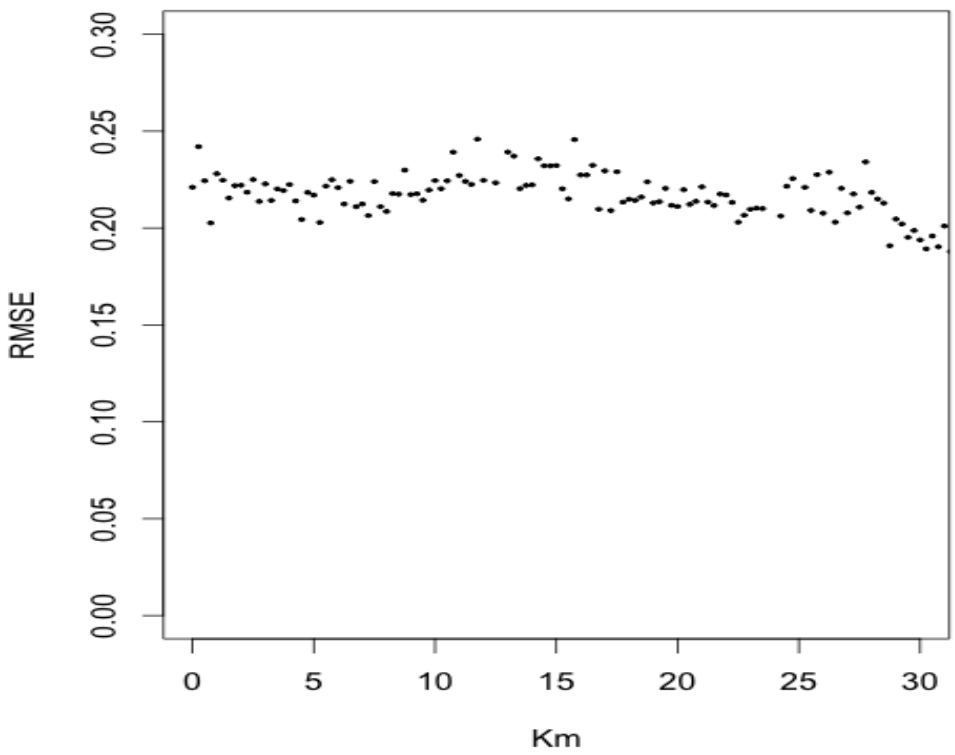


Figure 4-10 Plot of RMSE values changing with distances for all the sites from July 2013 to July 2014

From literature, the RMSE should have a better value when the distance is smaller than 5 kilometres, however, from the generated yearly RMSE plot, the errors keeps in the range of 20% to 25% for all distance. It shows the correlation between the RMSE values and distances is not quite significant.

Chapter 5 Discussion

From the generated results in chapter 4, both the expected and unexpected data has been come out.

A comparison between the results in this project and previous work will be discussed. The effects of different factors to the estimation accuracy will be discussed as well. And this will answer our research question3:

How does distance affect the accuracy of the PV output estimates?

5.1 Comparison with previous K_{PV} method based project for PV output estimation

This section will also answer the research question2.

From the results of Tan et. al. 2014 [8] , the RMSE of K_{pv} method is for WERU sites are 18.8%, 19. 8% and 20.7% respectively, and the MBE values are 6.5%, -6.5% and 5.3% respectively.

A summary of comparison is shown in following table:

	MBE(%)	RMSE(%)
WERU site641	12.8	17.0
WERU site642	-4.4	15.4
WERU site643	-5.9	19.5
All sites in this project	2.86	53.9

Table 5-1 comparison of errors for K_{pv} between results of Tan et. al. 2014 and this project using hourly estimates.

According to this table, it can be found that the accuracy of K_{pv} has been reduced, the RMSE value is increased from around 15.4% ~19.5% to around 53.9%. The accuracy

changes much may due to the increase of the number of tested sites. If the sites did not record the data properly, the total amount of error will increase. Additionally, the sun angles affected the results as well. The sun angles in the morning and evening can bring more errors in the calculation. This requires more data quality control with filtering the sun angles need to be done. There are some other possible reasons for the large RMSE value occurring:

1. If the estimates are calculated under clear sky condition, and the measured data is under non-clear sky condition such like cloudy days, the results can be quite different for different sky conditions and therefore the error between estimates and measurements is large.
2. Some sites may record the output data incorrectly, and the wrong data cannot be qualified in quality control process.
3. During the clear sky estimates calculation, there are some unavoidable errors be generated.

The MBE in this project is better than the results in previous project, which is 2.86%, and it indicates the overestimation for predicted PV output is very small. This is the value we are expecting.

5.2 THE RELATIONSHIP BETWEEN ACCURACY AND DISTANCE

Based on the result in chapter4, the relationship between RMSE value and distance is hard to find. From the RMSE/distance plot (Figure 4- 10), the RMSE value is in the same range with changing of distance, the correlation of them is not obvious.

Following table shows a comparison between results of Tan et. al. 2014 [8] and the results of this project:

Distance	RMSE(%)
Tight cluster	13.6
Loose cluster	28.5
Random	23.4
All sites in this project	20-25

Table 5-2 distance/RMSE comparison of errors for K_{pv} between result of Tan et. al. 2014 and this project using hourly estimates

For the table 5-2, the tight cluster is the cluster of distances between two sites which are in the range of 0.3 to 1.1km, while the loose cluster represents the group of distances between two sites which are in the range of 32.8 to 41.1km.

From the findings in Tan et. al. 2014 [8] using data from 2013 to 2013, there is relationship between hourly estimates accuracy and the distance of two PV sites. When the distance is below approximate 5km, the RMSE values become smaller obviously, which are in the range of 10% to 20% approximately, the accuracy of estimation is better. When the distance is greater than 5km, the accuracy remains almost the same; the RMSE is in the range of 15% to 35% approximately.
However, from the results of this project using data from 2013 to 2014, when the distance between two sites is less than 5km, the RMSE value is the range of 20% to 25%; when the distance increased to greater than 5km, the RMSE values are still the same, in the range of 20% to 25%.

This result is different from the results in Tan et. al. 2014[8] and the correlation cannot be found. There are some possible reasons:
First, in this project, more than 300 sites' data has been used for estimation, much more than the data used in Tan et. al. 2014 [8]. The error occurs more easily for larger data set.
Second, the quality control has been done for the data of ACT school sites. However, quality control has not been applied to the data of PV sites in Canberra properly. This may cause the error as well.
Third, in quality control process, the sun angle needs to be taken into consideration. Sun angles in the morning and evening may have effects on estimate calculation and accuracy, so some of the data may need to be filtered by sun angles.
This requires further reinvestigation for the angles and quality control, which is beyond the scope of this thesis.

5.3 OTHER FACTORS AFFECT THE ACCURACY

5.3.1 Sky conditions

From the literature, the absolute error changes with the K_{pv} value changing, and it indicates that different K_{pv} values have different effects on the accuracy.

According to the K_{pv} values calculated in previous chapter, in some days, the K_{pv} values of the day changed rapidly due different sky condition. For example, in chapter3, the K_{pv} curves for Belconnen High School (site id 1942) and Forrest Primary School (site id 21) on 16 March 2014 (Figure 3-6) are like broken lines and have a trend of fluctuating down. This day is a partial cloudy day with K_{pv} value changes as well as the accuracy. Therefore, it can be assumed that different sky conditions will have different effects on the accuracy as well.

5.3.2 Solar zenith angles

In the test of relationship between distance and accuracy, the sun angle is considered as an important factor of affecting the estimate accuracy. The accuracy and errors may vary through the day. The sun angles in the morning and evening, before sunrise and sunset, makes the PV output be very small. Therefore, the uncertainty in measurement may increase and the accuracy of PV output estimation will reduced.

Chapter 6 Conclusions and Further Work

6.1 CONCLUSION

This thesis investigated the solar estimation in Canberra based on K_{pv} method.
A literature review has been done which studied different methods on solar estimation.
In data curation process, quality control has been applied for the data of 327 PV sites in Canberra. Two main issues in quality control process were solved: First is time-stamping issue in the output data, where the recorded time is one hour later than the local time due to some errors during measurement. Second is the abnormal output values, the output values which are too large or too small, or out of the deviation limit of normal distribution, were removed by physical and statistical limit respectively.

In the output estimation process, the theoretical clear sky output was simulated and the output of one PV site was estimated from a nearby site by using the K_{pv} method.

The result and analysis part presented the estimation accuracy, which has an absolute RMSE value of 0.224 kW/kW_p and a MBE value of 2.86%. Both of the values were small and they have demonstrated that the solar output can be predicted accurately with acceptable errors by using K_{pv} method. In addition, a comparison has been done between previous work and this project.

6.2 FURTHER WORK

Though in the data curation process, the time shifting problem and abnormal output data problem have been solved, more quality control still need to be done for the data. The solar zenith angle is assumed to have negative effect on estimate accuracy, filtering the PV output data by the angles is required for possible further works. Additionally, further research on identification and removal of shading may also be required. Therefore, higher quality dataset can be made, and more accurate estimate results can be generated.

From the result of relationship between distance and accuracy, it was found that this result was different from the findings of Tan et. al. 2014 [8]. This requires reinvestigation on the relationship of distance with errors.

Other factors which can affect the estimation accuracy can also be investigated in further work if possible, and a more accurate and convenient way may be found to do PV output estimation.

Appendix

sid	sb	sid	sb
1839	BRADDON	**1946**	Ainslie
1841	ARANDA	**1930**	Palmerston
1851	HOLT	**1954**	Kambah
1858	WANNIASSA	**1948**	Forrest
1866	FARRER	**1953**	Torrens
1856	Gilmore	**1955**	Wanniassa
1843	Hawker	**1867**	DEAK
1854	FYSHWICK	**1847**	AMAR
1852	MELBA	**1862**	WARA
1857	Richardson	**1187**	OCON
1840	O'CONNOR	**1183**	CAMP
1942	BELCONNEN	**1202**	PHIL
1963	CALWELL	**1189**	CHAP
1962	CALWELL	**1199**	COND
1947	CAMPBELL	**1844**	DICK
1958	Chisholm	**1191**	DUFF
1936	CHARNWOOD	**1185**	BELC
1952	Curtin	**1200**	LYNE
1956	Fadden	**1845**	LYNE
1941	FLOREY	**1863**	LYON
1949	FORREST	**1195**	WATS
1935	FRASER	**1865**	MAWS
1933	GIRALANG	**1864**	PEAR
1931	Nicholls	**1860**	KAMB
1957	GOWRIE	**1870**	NARR
1943	HAWKER	**1869**	REDH
1944	KALEEN	**1849**	SCUL
1960	Greenway	**1861**	WARA
1964	Conder	**1868**	DEAK
1939	LATHAM	**1192**	WANN
1940	MACGREGOR	**1842**	WEET
1951	Garran	**1197**	YARR
1945	KALEEN		
1852	MELBA		
1934	EVATT		
1961	Monash		
1937	MELBA		
1950	NARRABUNDAH		
1927	Ngunnawal		

Table A-1List of ACT School sites

sid	st	ar	az	tl	nm	ms	mp	mr	mt	mm	sh	ni	ii	ir	it	im	pc	sb	div	lat	lon	alt	pr	int
1839	NA	10350	26	20	45	15	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2612	BRADDON	NA	-35.2734	149.1395	574	1	3600
1841	NA	10350	0	30	45	15	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2614	ARANDA	NA	-35.2554	149.0796	622	1	3600
1851	NA	10350	43	25	45	15	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2615	HOLT	NA	-35.2249	149.0285	585	1	3600
1858	NA	20010	-9	32	87	14.5	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2903	WANNIASSA	NA	-35.403	149.0963	608	1	3600
1866	NA	10350	0	30	45	15	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2607	FARRER	NA	-35.3782	149.1068	655	1	3600
1856	NA	10350	0	20	45	15	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2905	Gilmore	NA	-35.4191	149.1349	658	1	3600
1843	NA	10350	-21	25	45	15	3	230	SR-P660230	52	1	1	1	100	SMC10000TL	53	2905	Hawker	NA	-35.2498	149.0348	642	1	3600
sid	st	ar	az	tl	nm	ms	mp	mr	mt	mm	sh	ni	ii	ir	it	im	pc	sb	div	lat	lon	alt	pr	int

Table A-2Collected metadata

Bibliography

[1] I. MacGill, Photovoltaics in Australia: Time for a Rethink, 2015, IEEE Power & Energy Magazine, v. 13 no. 2, pp. 96

[2] Australia: SOLAR energy becomes a success in AUSTRALIA MENA Report, 2014

[3] M.Yu, A.Halog 2015, Solar Photovoltaic Development in Australia—A Life Cycle Sustainability Assessment Study, Environment and Energy: the Industrial Ecology perspective

[4] Sayeef.S, Heslop.S, Cornforth.D, Moore.T, Percy.S, Ward.J, Berry. A, Rowe.D, 2012, Solar intermittency: Australia's clean energy challengek, CSIRO

[5] Engerer, N. A., Mills, F. P., 2013. Kpv: A Clear-Sky Index for Photovoltaics

[6] Vincent P. A. Lonij, Vijai Thottathil Jayadevan, Adria E. Brooks, Jeffrey J. Rodriguez, Kevin Koch,Michael Leuthold, and Alexander D. Cronin, 2012, Forecasts of PV Power Output Using Power Measurements of 80 Residential PV Installs

[7] Golnas, A., Bryan, J., Wimbrow, R., Hansen, C., Voss, S., Clemente, S., 2011. Performance Assessment Without Pyranometers: Predicting Energy Output Based on Historical Correlation.

[8] J. Tan, N. A. Engerer, F. P. Mills, Estimating Hourly Energy Generation of Distributed Photovoltaic Arrays: a Comparison of Two Methods

[9] Rigollier, C., Bauer, O., Wald, L., 2000. On the Clear Sky Model of the ESRA European Solar Radiation Atlas -With Respect to the Heliostat Method. Solar Energy 68 (1), 33–48.

[10] Reindl, D., Beckman, W., Duffie, J., Jan. 1990. Evaluation of Hourly Tilted Surface Radiation Models. Solar Energy 45 (1), 9–17

[11] Gueymard, C., Mar. 2009. Direct and Indirect Uncertainties in the Prediction of Tilted Irradiance for Solar Engineering Applications. Solar Energy 83 (3), 432–444.

[12] King, D., Boyson, W., Kratochvill, J., 2004. Photovoltaic Array Performance Model. Tech. Rep. December, Sandia National Laboratories.

[13] King, D., Gonzalez, S., Galbraith, G., Boyson, W., 2007. Performance Model for Grid-Connected Photovoltaic Inverters.

[14] J. Tan, 2013. An Investigation of Methods for Estimating Distributed Solar Energy Generation, Australian National University

[15] N. A. Engerer, 2011. Simulating Photovoltaic Array Performance using Radiation Observations from the Oklahoma Mesonet, Norman, Oklahoma

[16] A. Zahedi, 2010. Australian renewable energy progress, Renewable and Sustainable Energy Reviews Volume 14, Issue 8, October 2010, Pages 2208–2213

[17]ACT School Pulse Metering, Water group, online available at http://www.outpostcentral.com/remote/(S(botl3thgajjmasuip5wtvzys))/cws/actschools/map.aspx

[18] Mapcoordinates.net, online available at http://www.mapcoordinates.net/en

[19] PVoutput.org, online available at:
http://pvoutput.org/list.jsp?p=4&id=10315&sid=10659&gs=0&v=0& o=date&d=desc

[20] Engerer, N. A., R package manual

[21] Reindl, D., Beckman, W., Duffie, J., Jan. 1990. Evaluation of Hourly Tilted Surface Radiation Models. Solar Energy 45 (1), 9–17

[22] King, D., Boyson, W., Kratochvill, J., 2004. Photovoltaic Array Performance Model. Tech. Rep. December, Sandia National Laboratories.

[23] King, D., Gonzalez, S., Galbraith, G., Boyson, W., 2007. Performance Model for Grid-Connected Photovoltaic Inverters.

[24] Badescu, Computing the Global and Diffuse Solar Hourly Irradiation on Clear Sky.

[25] T. Khatib, A. Mohamed and K. Sopian, “A review of solar energy modelling techniques,” Renewable and Sustainable Energy Reviews, vol. 16, no. 5, pp. 2864-2869, June 2012.