SAMPLE SIZE DETERMINATION OF PHOTOVOLTAIC BY ASSESSING REGIONAL VARIABILITY

Ran LI, University of Bath – UK, rl272@bath.ac.uk
Gavin SHADDICK, University of Bath – UK, G.Shaddick@bath.ac.uk
Haojie YAN, University of Bath – UK, h.yan@bath.ac.uk
Furong LI, University of Bath – UK, f.li@bath.ac.uk

ABSTRACT

In order to assess the effect potential impact of new low-carbon technologies, such as photovoltaic (PV) installations, electric vehicles (EV) and heat pumps (HP) on LV networks it is important to be able quantify the likely outputs from these technologies. However the installation of monitoring equipment is expensive and there may be technical considerations and issues with obtaining customer permission which mean that monitoring large numbers of PV installations may be difficult in practice. This paper proposes a novel statistical approach to determine the effect of sample size in producing an accurate representation of the output of PV installations within South Wales in UK. Probability distributions are used represent the variability in parameters that directly influence PV generation profiles, such as size and orientation, over the area of study. The basic idea is then to use Monte-Carlo simulation techniques to build up representations of the distribution of PV outputs by month within the study area. The potential biases associated with choosing different sample sizes and sampling procedures are then assessed.

INTRODUCTION

There is an increasing interest in low-carbon technologies such as micro-PV, electric vehicles (EV) and heat pumps (HP) at the low voltage (415V) customer side [1]. It needs for greater understanding of the potential impact of such techniques on the LV network and subsequent changes in LV planning. In order to address this need, Western Power Distribution (WPD), the distribution network operator responsible for energy distribution for Southwest England, South Wales and the Midlands in the UK, has initiated a project which aims to identify and quantify the potential impact of new low-carbon technologies. In an additional part of the project, monitoring devices need to be deployed for household PVs in order to assess the impact of these PV at the substation level.
There are currently over 3500 registered PV installations in South Wales, over 1000 of which are associated with substations in the study area. Theoretically, monitoring devices could be installed for every PV installation within the study area, however there are difficulties associated with this including cost, technical issues and obtaining customers' permission. This paper focuses on determining the size of the sample of PV installations that would be required in order to obtain an accurate representation of the distribution of PV outputs over a region.
A number of studies have investigated the relationship between PV input factors and PV output. The output is influenced by a variety of factors including irradiation, azimuth and tilt angle, temperature and system efficiency [2]. Traditionally, methods for assessing required sample sizes are based on assessing the difference between two values which are assumed to follow a given probability distribution, often the Normal [3]. However, in the case considered here this is unlikely to be tenable assumption. Additionally, it is also assumed that the parameters of the distributions are known which ignores the inherent uncertainty and thus underestimates the size of the sample that is required.
In this paper, we propose a more flexible approach to the calculation of sample sizes required to accurately represent PV output over an extended area using Monte Carlo analysis (or error propagation) to produce distributions of output [4]. Within each iteration of the Monte-Carlo simulation, samples from each of these distributions are used to compile a distribution of PV output over the study region. This simulated distribution of PV outputs are then treated as 'real' in the second stage of the process in which the potential effects of different sampling strategies are investigated. At this second stage, a selection of different 'samples' are drawn from the distribution of PV outputs obtained from the first stage and which are now treated as 'real'. The effects of choosing different sampling sizes in terms of obtaining a representative sample are assessed in terms of bias both of mean levels and variability.
The rest of the paper will firstly describe the deterministic relationship between a selection of important input factors and PV output. Next, the rationale for the probability distributions used to represent the variability in the input parameters and the choice of parameters for these distributions are explained. In part IV details of the Monte Carlo method used to simulate the distribution of PV outputs are presented. Part V contains a statistical analysis of the effects of different sample sizes and strategies on the estimation of both mean levels and variability. Finally, we provide a discussion of the findings of this study together with suggestions for future research.

THE OUTPUT PROFILE OF PHOTOVOLTAIC

The output of a PV is influenced by various factors, including irradiation, azimuth, tilt angle, temperature, system efficiency and PV material [5]. Mathematically, the power of a PV can be by expressed as follows

$$P = \frac{G}{G_0} \times eff_{rel}(G, T_m) \times P_{STC} \quad (1)$$

where G_0 =1000W/m^2. G is the solar irradiance on the PV module, P_{STC} is the nominal peak, and eff_{rel} is the relative module efficiency which is a function of irradiance and module temperature.

The nominal peak power is the power output of the module(s) measured at Standard Test Conditions (STC). The module efficiency measured at Standard Test Conditions is referred to as eff_{nom}. The relationship between P_{STC} and eff_{nom} can be expressed as

$$P_{STC} = A \times eff_{nom} \quad (2)$$

where P_{STC} is the nominal peak power and A is the panel surface area of the PV modules.

The nominal peak power can be written in the following form:

$$P_{STC} = V_{STC} \times I_{STC} \quad (3)$$

where V_{STC} and I_{STC} are the voltage and current of the PV, respectively under the electric load that produces the maximum power at Standard Test Conditions (STC). When the values of temperature and irradiance differ from STC, the maximum current and voltage become I_m and V_m. From (3) and (4), the relative efficiency eff_{rel} can be therefore expressed as [6]:

$$eff_{rel}(G, T_m) = \frac{P \times G_0}{P_{STC} \times G} = \frac{I_m V_m G_0}{I_{STC} V_{STC} G} = (1 + \alpha_i \times (T_m - T_0))$$
$$\times (1 + c_1 \times \ln(\frac{G}{G_0}) + c_2 \times (\ln(\frac{G}{G_0})^2 + \beta_\gamma \times (T_m - T_0)) \quad (4)$$

where the coefficients, α_i, β_γ, c_1 and c_2 are empirical constants. The module temperature has a corresponding relationship with the ambient temperature of the form:

$$T_m = (T_{NOCT} - 20) \times \frac{G}{800} + T_{amb} \quad (5)$$

where T_m is a PV module's temperature, T_{amb} is ambient temperature and T_{NOCT} is nominal operating cell temperature.

THE CHOICE OF PROBABILISTIC DISTRIBUTIONS TO REPRESENT PV INPUT FACTORS

Predicting the power output of a PV using the formula given can be very challenging as some of the input factors may vary significantly under different conditions or in different regions. Normal distributions have been adopted for this purpose in a number of previous studies and [7] gave an example of using probability distributions in calculating PV cost. However the assumption of normality is unlikely to be realistic for all the PV input and in this paper we introduce an improved method to estimate the output of PVs based on a selection of probability for sample size assessments.

Temperature and Solar Irradiation

As can be seen from (1), G_0 is a constant as discussed above. Information on temperature T_m and solar irradiance was obtained from The Photovoltaic Geographic Information System (PVGIS) of European Communities [8]. As the study region, part of South Wales, is a relatively small geographically area, it may be assumed that temperature and solar irradiance are constant at any time point over the study area.

Previous studies have found that solar irradiation and temperature within a small area can be reasonably represented by normal distributions [7, 9]. Here, the irradiation and temperature for each month are assumed to be normally distributed with means and standard deviations based on historical data obtained from PVGIS [10]. Figure1 shows the distribution for average daily irradiation distribution in June.

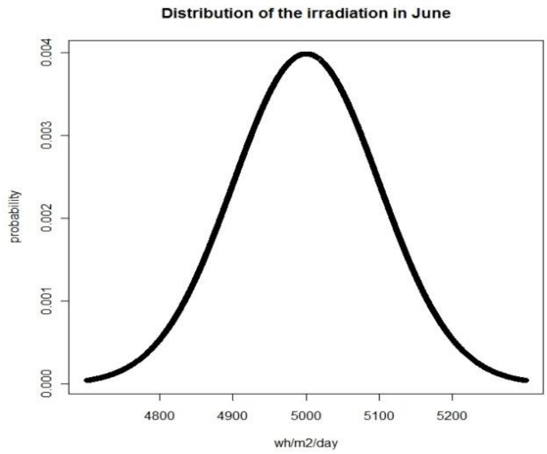

Fig.1. Normal distribution used to represent the solar irradiation in June in South Wales

Azimuth Angle

It is noted that the irradiance G absorbed by PV modules in (1) is not only influenced by the solar irradiation but also the tilt angle, azimuth and the material of the PV [11]. The optimum azimuth angle in the UK is south. For north-south facing houses, most customers will install their PVs on south roof in order to obtain the highest power output. Although for the optimum power output PV panels should be installed at this orientation and ,angle factors such as roof types and non-south facing properties, mean the range of orientation is extended from almost 360 degrees for azimuth angle (south=0, west=90, north=180=-180 and east=-90) and from 0 to 45 degrees for tilt.

Based on these factors, the distribution of the azimuth angle is assumed to be normally distributed with a mean angle of 0 degrees and standard deviation of 60 degree.

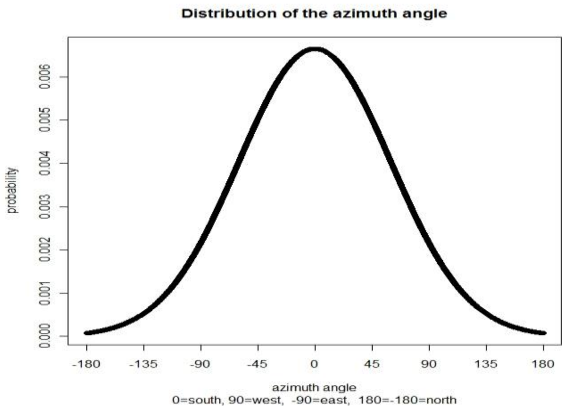

Fig.2. Normal distribution representing the azimuth angles of PVs in South Wales

Tilt Angle

The optimum tilt angle in the UK is 30-40 degrees [11] and the majority of the PVs are put within the range [12]. However in the case of flat roof, due to safety requirements, some PVs may be installed horizontally. In such cases, the solar radiation received may still be almost 90% of that at the optimum angle [12]. Previous research has shown that the received solar radiation starts to fall sharply when the angle is above 50 degree and is substantially reduced when above 70 degrees [12].

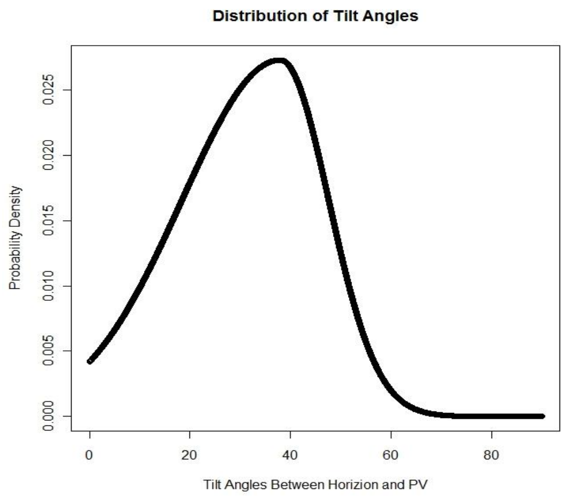

Fig.3. Distribution representing tilt angles of PVs in South Wales

Angles below 30 degrees are more likely to appear than those above 40 degrees due to them being more efficient. Based on this information, the distribution for tilt angles of all PVs in the study region is assumed to be a truncated skewed normal 15] with a mean of 35 degrees, a standard deviation of 15, a skewness parameter of 0.7 and with truncation occurring at 0 degrees seen in Fig.3.

PV Size

The size of a PV is often represented in terms of its total nominal peak power. For example, a 1 Kw-size PV indicates a nominal peak power of 1 Kw and a 10 Kw-size PV has nominal peak power of 10 Kw, therefore higher power is associated with larger PV size.

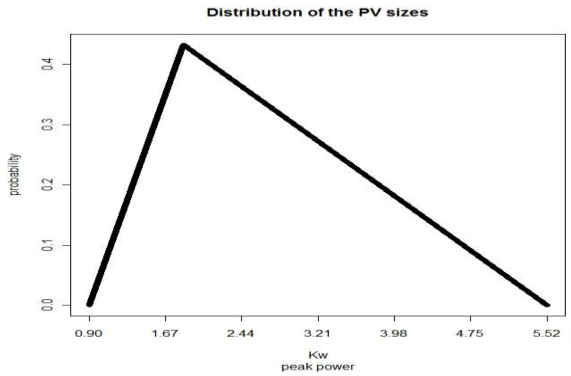

Fig.4. Triangular distribution representing sizes of the PVs in South Wales

Bases on information supplied from WPD relating to a sample of approximately 300 PVs in the study area, a suitable distribution for PV sizes is the triangular distribution. Within this sample PVs, the maximum size is 5.52Kw and minimum size is 0.9Kw with a mode of 1.85. Figure 4 shows the corresponding triangular distribution used to represent PV sizes of all PVs.

Other Factors

There are other factors that are less amenable to being assigned a probability distribution due to difficulties in obtaining enough information on which to basis a sensible choice. Examples include trees shading, PV material and inverter failure. In the absence of additional information, the effect of such factors is represented white noise, in the form of the variance of a normal distribution.

SIMULATION OF THE PV OUTPUT PROFILES BY COMBINING THE DISTRIBUTIONS OF INPUT FACTORS

Monthly Profiles of Photovoltaic outputs

The monthly power output profile is the daily average power output of each month over a year. Although the daily profile is most suitable for electricity network analysis, for simplicity we demonstrate the approach on monthly profile due to constraints in the data required to determine the required distributions at the higher level of temporal resolution. .

The product Pi for average daily power output of a PV in month i with irradiation G_i, efficiency $eff_{rel}(G_i,T_m)_i$ and peak power P_{STCi} is represented by,

$$P_i = \frac{G_i}{G_0} \times eff_{rel}(G_i, T_{mi})_i \times P_{STCi} \tag{6}$$

where i=1, … , 12 represents months of the year.

In order to obtain the distribution of power outputs from PVs of the PVs can be calculated by drawing repeated samples from the distributions defined for factors that feed into the calculations leading to (6), resulting in distributions for power outputs.

Monte Carlo simulation and correlated inputs

Using Monte-Carlo simulation techniques samples are repeatedly taken from the distributions of the input factors used to build up distributions for the power. Here we aim to simulate the output profiles to represent approximately 1000 PVs within the study area.
Profiles were simulated for the outputs of 1000 PV installations based on the input parameters and their associated distributions.
However, there are a number of factors which may have major effects on PVs such as system losses, shading and inverter failure. Based on a study of these factors, a noise term representing 50 W is incorporated into the simulation models, via the standard to represent this additional variability. The value of the noise term will be used as the standard deviation in a normal distribution. Further, in order to investigate the uncertainty of the white noise, a series of different values of the white noise from 1 W to 100 W is tested.
Figure 5 shows the resulting estimated daily output profiles for the 1000 PVs over a year together with the mean over all the profiles.

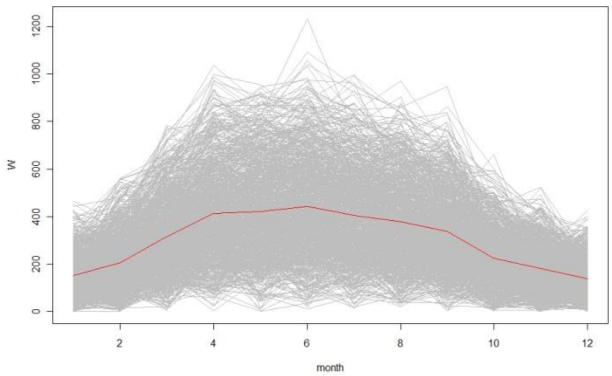

Fig.5. Daily output profiles of 1000 PVs over a year. The red line shows the mean output over all PVs

OBTAINING A REPRESENTATIVE SAMPLE

Programming and simulation

After the profiles of the 1000 PVs have been simulated, this is considered the true population, or 'real data'. We now present a method to assess the efficacy of using samples to make inference about the entire population in a similar fashion as would be carried out in practice. The difference here is that we know the truth, i.e. the features of the entire population we are sampling from, and so can assess the bias associated with different samples. The performance of samples of varying sizes: 2, 10, 100, 250 and 500, is assessed in terms of biases of the sample mean and standard deviation.
To allow for sampling variability, samples of each size are repeated drawn and the bias calculated for each sample. This procedure is repeated 1000 times for each sample size. Fig.6 shows clearly that as the sample size increases, the sample mean approaches to the population mean and that the return in increasing sample size is greatest when increasing very small samples. There are fewer effects when increasing sample size above 100, after which the sample means are very close to the true population mean.

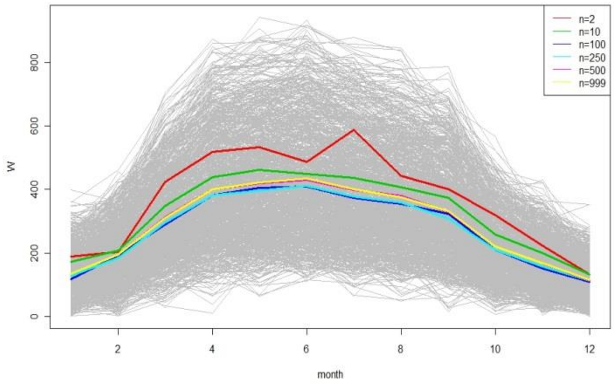

Fig.6. Mean profiles associated with different size samples with additional variation term (white noise) of 50W

The corresponding results when the term representing the additional variability associated with factors such as system losses, shading and inverter failure is increased to 100W can be seen in figure 7.

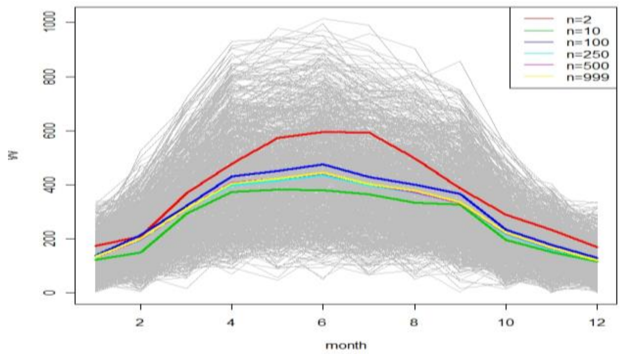

Fig.7. Mean profiles associated with different size samples with additional variation term (white noise) of 100W

Bias of sample mean and standard deviation

In order to assess the bias associated with samples of different sizes, the root mean square error (RMSE) is calculated for both means and standard deviations. For any given sample size, the RMSE is calculated using the mean of the repeated samples and the (known) mean of the population for each month. Formally, the bias is expressed in the following form:

$$Bias_{mean} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{spi} - P_{smi})^2} \tag{7}$$

where n=12 represents the 12 months, P_{spi} is the sample mean of PVs output in month i and P_{smi} is the 'true' mean
The mean values of bias, over the repeated samples, for different sample sizes can be seen in Figs 8. Using an additional variability term from 1W to 100W, it can be

seen that the bias can be as high as 100W when using a sample size of 2, reducing to 50W with a sample size increase to 10 and decreasing markedly after that. With sample sizes of 250 or greater, the bias is lower than 8W.

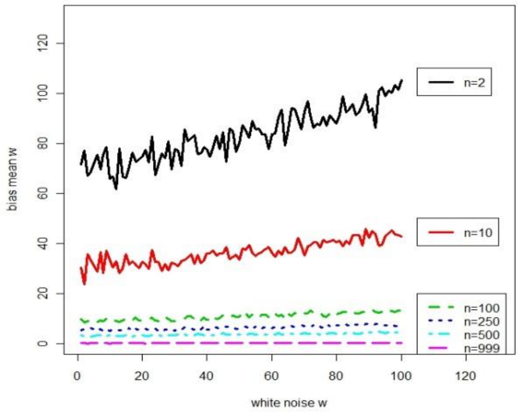

Fig.8. Bias (W) of means under different sample sizes with varying amounts of additional variability (white noise) terms

As observed in Fig 8, the effect of additional variation is more marked with smaller sample sizes and this is again seen in the estimation of biases in which biases increase markedly as the noise term increases but this effect is decreased with large sample sizes.

CONCLUSIONS

This paper proposes method for estimating the distribution of PV outputs over a region. Relationships between in out factors and PV outputs are expanded to incorporate both the inherent variability in values of the input factors but also additional levels of uncertainty. This is performed using Monte-Carlo techniques which enable representative distributions of PV outputs to be generated over the region of interest. We then used these representative distributions to perform a study to assess how accurately data from samples of PVs, rather than monitoring the entire population, would be in assessing output profiles.
Under the assumptions presented in the paper, it is observed that there is a marked relationship between accuracy and sample size when dealing with samples of less than 100 but that after 100 there is a diminishing return in terms of estimating mean and variability. Further development of this approach will focus on generalizing this method to enhance the applicability:

i. Networks analysis is usually based on half-hourly daily profiles. Future work will apply this approach into half-hourly daily profiles.
ii. At present, the distributions of input factors are based on a number of assumptions derived from previous studies and the literature. In the future, as data is collected from the monitors placed as part of the project, these assumption will be tested and updated based on real data

REFERENCES

[1] "2050 Pathways Analysis," DECC, LondonJuly,2010 July,2010.
[2] T. Huld, M. Šúri, and E. D. Dunlop, "Geographical variation of the conversion efficiency of crystalline silicon photovoltaic modules in Europe," Progress in Photovoltaics: Research and Applications, vol. 16, pp. 595-607, 2008.
[3] D. Chih-Shun, W. Qing, H. Cheng-Ta, and M. Pedram, "Stratified random sampling for power estimation," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, vol. 17, pp. 465-471, 1998.
[4] G. C. Christian P. Robert, Monte Carlo Statistical Methods, 2nd ed. New York, NY: Springer-Verlag, 2004.
[5] T. Huld, M. Šúri, and E. D. Dunlop, "Geographical variation of the conversion efficiency of crystalline silicon photovoltaic modules in Europe," Progress in Photovoltaics: Research and Applications, vol. 16, pp. 595-607, 2008.
[6] T. A. Huld, M. Suri, R. P. Kenny, and E. D. Dunlop, "Estimating PV performance over large geographical regions," in Photovoltaic Specialists Conference, 2005. Conference Record of the Thirty-first IEEE, 2005, pp. 1679-1682.
[7] S. B. Darling, F. You, T. Veselka, and A. Velosa, "Assumptions and the levelized cost of energy for photovoltaics," Energy & Environmental Science, vol. 4, pp. 3133-3139, 2011.
[8] Š. M. Huld T.A., Dunlop E.D., Micale F, "Estimating average daytime and daily temperature profiles within Europe," Environmental Modelling & Software, vol. 21, 12, 1650-166,, 2006.
[9] R. D. Harmel, C. W. Richardson, C. L. Hanson, and G. L. Johnson, "Evaluating the Adequacy of Simulating Maximum and Minimum Daily Air Temperature with the Normal Distribution," Journal of Applied Meteorology, vol. 41, pp. 744-753, 2002/07/01 2002.
[10] M. Šúri, T. A. Huld, and E. D. Dunlop, "PV-GIS: a web-based solar radiation database for the calculation of PV potential in Europe," International Journal of Sustainable Energy, vol. 24, pp. 55-67, 2005/06/01 2005.
[11] K. L. Christian N Jardine, "Photovoltaics in the UK," Environmental Change Institute, University of Oxford.
[12] C. T. K. Tovey "Performance of two photovoltaic arrays in the UK," Proceedings of the ICE - Energy, vol. Volume 161, pp. pages 11 –21 01 February 2008 2008.