
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Probabilistic evaluation of UK domestic solar photovoltaic systems: An integrated geographical information system PV estimation tool

PLEASE CITE THE PUBLISHED VERSION

<http://dx.doi.org/10.4229/EUPVSEC20162016-6DO.6.6>

PUBLISHER

© WIP Wirtschaft und Infrastruktur

VERSION

VoR (Version of Record)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Leicester, Philip A., Nicholas Doylend, and Paul Rowley. 2019. "Probabilistic Evaluation of UK Domestic Solar Photovoltaic Systems: An Integrated Geographical Information System PV Estimation Tool". figshare. <https://hdl.handle.net/2134/22237>.

PROBABILISTIC EVALUATION OF UK DOMESTIC SOLAR PHOTOVOLTAIC SYSTEMS: AN INTEGRATED GEOGRAPHICAL INFORMATION SYSTEM PV ESTIMATION TOOL

Philip A. Leicester*, Nick Doyle, Paul Rowley

Centre for Renewable Energy Systems Technology (CREST), School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, LE11 3TU, United Kingdom.

* Tel.: +44 (0)1509 635343, Email: p.a.leicester@lboro.ac.uk

ABSTRACT: It is shown how key predictor parameters for the spatial estimation of PV yield, self-consumption and thereby economic and social indicators can be extracted from a GIS system and introduced into a Bayesian Network model. This model endogenises the uncertainties and incorporates spatial variability inherent in these parameters. Empirical monthly and annual yield measurements obtained from over 600 PV installations have been obtained and compared with estimated yields obtained by two key solar tools used for performance estimation in the UK – these are PVGIS and the UK Government’s Standard Assessment Procedure (SAP) for domestic buildings. Mean bias estimates and root mean square error estimations were obtained for each tool and the results used to construct an uncertainty distribution in PV yield prediction given key input parameters such as system rating, orientation and tilt. This uncertainty was used to furnish a probabilistic graphical model with a prior distribution for PV yield estimation. This was integrated into a Geographical Information (GIS) system furnished with roof and building stock parameters including roof attributes obtained from lidar data. Elements held in a vector layer of the GIS system can be selected and the resultant distributions of input parameters automatically fed to the model to yield a posterior distribution of the PV yield. The model is able to propagate the yield uncertainty to other probabilistic models, including ones which predict the internal rate of return and self-consumption. The latter is in turn predicted by empirical marginal distributions of domestic electricity consumption. Thus with a given posterior distributions of PV yield, new posterior distributions for the internal rate of return, self-consumption and carbon emission savings are automatically calculated. By integration with GIS this novel approach allows the spatial analysis of the uncertainty pertaining to representative risk factors for PV adoption in the UK, and facilitate the estimation by installers, investors, and local authorities in a manner which endogenises uncertainty.

Keywords:

1. INTRODUCTION

At a whole system level photovoltaic (PV) technologies are seen as a valuable means to deliver both (i) an effective net energy gain to support multiple consumption patterns of society, and (ii) a decreasing reliance on non-renewable primary energy sources resulting in long term energy sustainability [1]. To support the emergence of a viable innovation system, in which PV can overcome market barriers and contribute to the energy mix, many nations have provided subventions or other market interventions to enable it to compete with incumbent technologies [2]. In the UK the domestic adoption of PV has been rapidly accelerated using the FiT mechanism, introduced in April 2010, which has driven the installation of almost 4GW on UK domestic roofs by April 2016 [3].

The cumulative cost of this subsidy, over the 6 years of FiT operation, commensurate, with a rapid reduction in total system cost, has led to intense political pressure to reduce the real and perceived burden on the electricity bill payer by lowering the FiT rate. The UK’s Department of Energy and Climate Change proposed that a target hurdle rate for investors of 4% was desirable and suggested it should only be realised in the most favourable locations [4]. Following consultation the generation FiT has been significantly reduced, though not as drastically as first proposed [5]. However, adopters and investors are exposed to greater risk of not making an economic return, and a slow-down in the rate of installation in the UK has been observed [3].

This has given greater impetus to understanding the uncertainty in the financial return. Research has been carried out to investigate the sensitivity of PV yield to a number of technical system parameters [6]. Taylor et al. have recently calculated a mean yearly integrated performance ratio of 83%, with a standard deviation of 7% (a standard error of 8%) using empirical data for over

7000 mainly domestic systems [7]. This error depends, of course, on many predictor variables, each with their own uncertainties. In particular monthly global horizontal irradiation, estimated by interpolating between ground station measurements, had an RMSE of 4.5%; further errors result from decomposition into direct and diffuse components and transposition to the plane of array [8]. However knowledge of uncertainties in the performance of PV is not enough to understand resultant uncertainties in economic performance. Under the UK’s FiT scheme, value is created for the generation, export and the displacement of imported electricity [9]. Thus Leicester et al. have investigated the uncertainty in self-consumption and characterised a distribution with a mean value of 34% and a standard error of circa 60%, depending on the housing stock studied [10]. This and other parameters have been used to furnish a probabilistic model with the required distributions for a probabilistic discounted cash flow analysis to yield a measure of the uncertainty in a net present value calculation for domestic PV [11]. This research has pioneered the employment of probabilistic graphical models (PGM) in order to endogenise the uncertainties in renewable energy generation and allow uncertain or variable inputs to furnish investors or policy makers with probabilistic outputs [12].

Specifically a PGM approach using Bayesian Networks (BN) has been developed in order to qualitatively define the parameter space, and to quantify the conditional interdependency relationships, between them using a directed acyclic graph and underlying conditional probability tables. In recent years, the BN community has employed graphical information systems (GIS) to furnish BN with the requisite marginal distributions for key parameters [13]. This allows the model to deliver spatially disaggregated probabilistic outputs.

The objective of this paper is to show how GIS/BN integration can be developed using publically available

datasets so as to deliver a spatial probabilistic PV evaluation tool. This can yield spatially disaggregated probabilistic outputs for key indicators allowing community planners to evaluate social, environmental impacts of PV. The paper presents the material as follows. In the Section 2 the use of GIS to furnish BN with probabilistic inputs is presented. In Section 3, GIS data to deliver building stock attributes, and PV yield, household incomes are explored. In Section 4 the use of a GIS interface to select evidence for the probabilistic model is demonstrated. Finally we present a discussion of the utility of this approach, followed by conclusions and suggestions for further work.

2. PROBABILISTIC EVIDENCE FOR BAYESIAN NETWORKS

An entire knowledge domain defined by a set $V = \{V_1, V_2, \dots, V_n\}$ of random variables can be represented by a joint probability distribution (JPD), $P(V_1, V_2, \dots, V_n)$. A jpd with many variables, each with a number of potential values (states) can be mathematically intractable due to an inordinate total number of states. In contrast, a Bayesian network (BN) is a significantly more compact representation of the knowledge domain. A BN is a couple (G, P) , where $G = (N, E)$ is a directed acyclic graph with a set of nodes N , each representing a variable in V , and E is a set of directed edges, which represent conditional dependencies between them. The latter are encapsulated as conditional probability tables (CPT). The joint probability distribution for a BN can be factorised using the chain rule, represented by Equation 1, where π_{V_i} is the set of parent nodes of node V_i . Thus each term in the product is a CPT, or, in the case where the set of parent nodes is empty, its marginal probability distribution. The BN renders the algorithmic calculation of the prior distribution of each variable mathematically tractable. For a brief introduction in to BN theory see reference [11] and references therein.

$$P(V_1, V_2, \dots, V_n) = \prod_{i=1}^n P(V_i | \pi_{V_i}) \quad \text{Equation 1}$$

The crucial concept to grasp for the application of BN to a problem domain is the concept of evidence in order to make prognostic and diagnostic inferences. Evidence means that the model user has new knowledge about the probability distribution applied to one or more variables, and applies these to the model. This results in a new calculation of the JPD and furnishes all the remaining variables with a new posterior distribution. Evidence is of two principal types. Firstly, a variable may be fixed to a specific state i.e. the variable is fixed to one specific value of its potential states. For example, consider a distribution of the discretised states for the PV system rating. The prior distribution of system ratings in a dataset might be as in Table 1. The user could then update the model by applying evidence such that the system rating were set to 4 kWp. This is known as hard evidence since the variable has been fixed to this value. In Table 1 the hard evidence is set to 4kWp. A second type of evidence might be one where the user applies a new distribution which updates the prior distribution. This is called probabilistic evidence since the evidence represents a new local probability distribution applied to the variable in the model [14].

Table 1 Example of a prior distribution of a parameter modified with hard evidence and probabilistic evidence.

Distribution	System Rating (kWp)				
	1	2	3	4	5
Prior Distribution	5	20	40	30	5
Hard evidence	0	0	0	100	0
Probabilistic Evidence	0	15	40	35	10

3. PROBABILISTIC EVIDENCE FURNISHED FROM A GIS SYSTEM

For this work several BN have been developed to understand the relationship between building stock parameters, building energy consumption, PV yield and self-consumption. These have been linked together to deliver an object oriented BN which integrates several knowledge domains to deliver a whole system model. The key elements of this model are reproduced in Figure 1. The individual BN models can be seen in Reference [11].

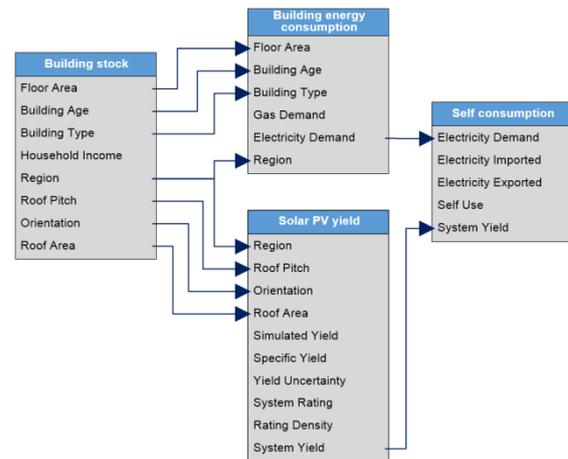


Figure 1. Object oriented BN model for domestic PV

The building stock BN model is key since it delivers probabilistic distributions for its parameters to the building energy consumption and PV yield models. These in turn predict the distribution of self-consumption [10]. The challenge is to extract distributions from GIS vector layers and to furnish the BN model with these distributions as probabilistic evidence. The distributions are determined from the features selected by the GIS user.

There are two approaches which decision and policy makers might employ to achieve this. The first is to select specific geographic areas based on boundary polygons (for example a local authority or community). All the features within the boundary are then used to generate distributions for the required parameters. The second approach is to use the GIS interface to select specific features, for example a row of houses, or those matching particular search criteria, such as all south-facing properties which have a roof size of 25m² or greater. Here we demonstrate the approach using UK census areas known as lower super output areas (LSOA). Figure 2 shows a vector map of one such census area in the South West of England.

The selection of a census area can be used to trigger a software script in order to calculate the distributions required as probabilistic evidence in the BN model. This involves a frequency count of discrete variables, such as building type or, for a continuous variable the allocation to an appropriate interval in a discretised distribution. Some parameters are held in vector layer attribute tables which are linked to the selected features. Others may be calculated using attribute parameters from one or more layers. In the following subsections the parameter sources and processing are described.



Figure 2 Census area LSOA Kerrier 008B in Camborne, Cornwall, England

3.1. DWELLING FLOOR AREA

Building footprints are available in the UK's OS MasterMap® Topography Layer which provides streets and building vector layers [15]. For the correct estimation of energy consumption, using the Building Energy Consumption BN, footprints need to be converted to floor areas. Thus the footprint area is multiplied by the number of floors, the latter estimated using building height data. This has recently been included for many conurbations in the UK MasterMap Building Heights geospatial data [16]. Building footprints in the MasterMap datasets do not always represent single dwelling units. Using the OS AddressBase® File [17], which provides a point vector layer for postal addresses, the number of dwelling units within the building footprint can be estimated using spatial queries. Figure 3 shows an example row of dwellings evaluated using this method. It can be seen that the building footprint on the far left encompasses four address points and the next one encompasses two. The remaining building footprints have only one.

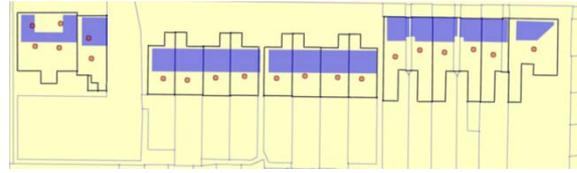


Figure 3 Building footprints with address points

It is important that the floor area is calculated correctly since, as previous work has shown [10], it is an important predictor of building energy demand by a domestic unit.

3.2. BUILDING TYPE AND AGE

Two other attributes, which have been shown to influence domestic energy consumption, are the Building type and the Building Age. These are available in commercially available GIS datasets from the GeoInformation Group [18]. Figure 4 shows the distributions of Building Age determined for four different LSOA census areas in England.

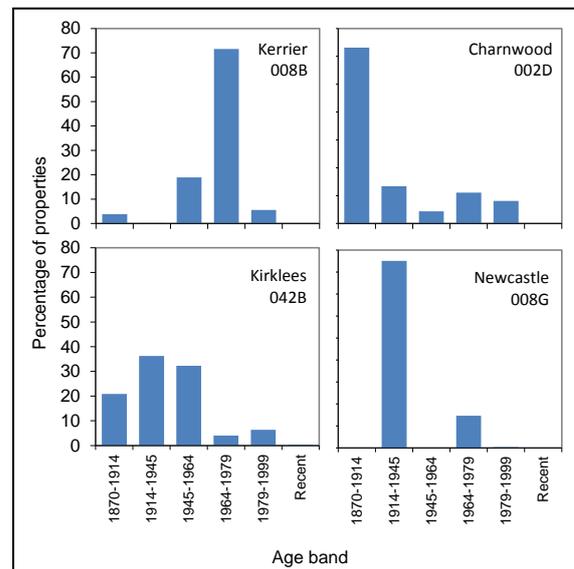


Figure 4 Distributions of Building Age

3.3. ROOF AREA AND ORIENTATION

Publically available lidar data, in combination with building footprints discussed above to serve as cookie-cutters in the GIS data layer, can be used to estimate roof orientations and tilts using methods reported by Palmer et al [19]. Roof areas can also be calculated using lidar data. Whilst this can yield azimuth and tilt distributions for most building stock in the UK, visual inspection does show that there is a high degree of potential error. For example stock which exhibits a high density of complex roof structures common in the UK, such as dormer windows, skylights, hip-roofs and intersecting roofs can overestimate the PV potential of the housing stock. Table 2 shows the percentage of buildings afflicted by problems which are difficult to estimate using automated algorithmic approaches. These were assessed using aerial photography in combination with GIS data. Only two-thirds of dwellings had roofs unaffected by structural constraints or shading. This included significant number of apartment dwellings which had no dedicated roof [12].

Table 2 roof assessments in four census areas (% suitable)

Assessment	Kerrier 008B	Kirklees 042B	Charnwood 002D	Newcastle 008G	TOTAL
Suitable	71.0	66.8	57.2	73.3	66.7
Affected by shading	9.7	11.1	9.4	8.2	9.6
Apartment (No roof)	14.0	11.1	26.4	15.3	16.9
Structural Constraints	5.2	11.0	7.1	3.2	6.8

Nevertheless these techniques enable the creation of GIS datasets of large areas of building stock from which parameter distributions to serve as probabilistic evidence can be determined.

3.4.SOCIO-ECONOMICS

A key objective was to be able to evaluate the socio-economic impacts of PV. For example, there is a need to understand the impact of low carbon technologies on fuel spending and the overall household income. To this end household income distributions for census areas were calculated using the iterative proportional fitting technique (IPF) [20]. This technique uses a reference dataset of anonymised census data with income linked to other social parameters to calculate the distribution of income in a target dataset for which only the social parameters are known. This showed that, typically, equalised household incomes are variable, exhibiting a coefficient of variation of over 50%. This delivers a marginal distribution of income for the census area (Figure 5). This data can then be further linked to building stock using the IPF technique. As a reference dataset the UK’s English Housing Survey [21] was used to estimate income for the housing stock in candidate census areas.

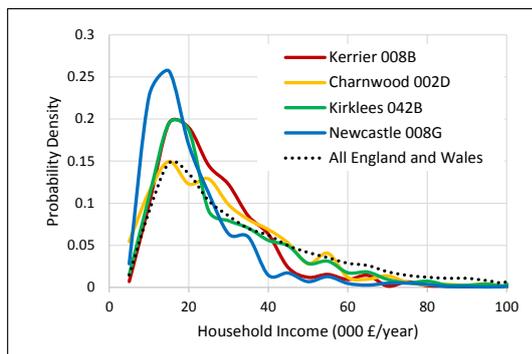


Figure 5 Household income distributions for census areas

Earlier studies have shown that domestic energy consumption, both electricity and gas, are directly and indirectly influenced by household income; the indirect influence is via the dwelling floor area [22] [23]. The result of the IPF technique applied to the building stock data is shown in Figure 6. This shows that the latter relationship has been effectively modelled in the simulated dataset. This probabilistic interpolation of income into the building stock model and the

relationships modelled between building parameters, building energy consumption, and self-consumption of PV generated electricity ensure that the GIS layers deliver useful probabilistic evidence to the BN model in Figure 1.

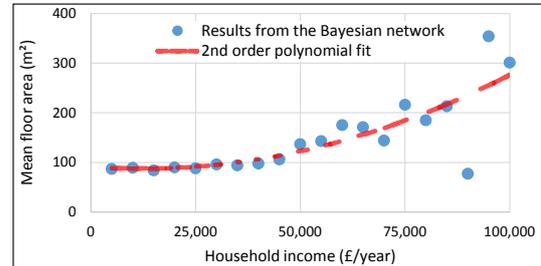


Figure 6 Result of modelled relationship between floor area and household income using iterative proportional fitting

3.5.PV YIELD

Whilst above the characterisation of building stock and occupancy factors which can then introduce probabilistic distributions in to the Building Stock BN model, it is also pertinent to note the sources of data used to estimate the PV Yield. Two PV yield estimation tools have been evaluated for the degree of uncertainty they introduce in to the model, PVGIS [24] and the UK SAP. Estimates from these tools were compared to the empirical annual yields for over 600 UK installations obtained from the Sheffield Microgeneration Database [25]. Figure 7 shows the result for PVGIS using the CMSAF database. Mean bias estimates and root mean square error estimations were obtained for each tool and the results used to construct an uncertainty distribution in PV yield prediction given key input parameters such as system rating, orientation and tilt. This analysis furnishes the model with an uncertainty distribution for the yield estimation by PVGIS or the SAP. In the application developed in this work, PVGIS was used to furnish the model with estimations.

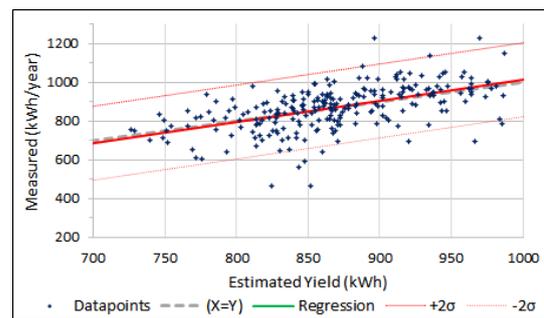


Figure 7 Measured versus PVGIS/CMSAF estimated yield for 600 UK systems

In this section a number of parameters have been presented which can be extracted from GIS layers in order to then be summarised as probability distributions. These are summarised in Table 3.

Table 3 Summary of parameters obtained from GIS vector layers and source data

Parameter	GIS Source	Comments
Floor area	Mastermap building footprints	Calculated using address point data and building heights
Building Type	Geoinformation group	Commercially available data; can be ignored if total energy not required.
Building Age		
Roof Slope	Lidar data	Calculated and added to attribute table
Roof areas for PV		
Orientation		
Household income	Census data	Iterative proportional fitting to produce simulated data added to attribute table
PV Yield	PVGIS or SAP	Uncertainty must be incorporated using empirical data.

4. APPLICATION OF PROBABILISTIC EVIDENCE TO THE BAYESIAN NETWORK

When selecting GIS features, either by selecting a polygon, or a large number of features using QGIS, it is a simple task to iterate through the selection set to obtain attribute values with which to construct probability distributions for the required parameters in the BN model. This was implemented as a QGIS plugin using a routine written in R [26]. This calculated the required distributions using an R-GIS implementation which were then provided to an RNetica software layer [27]. The latter uses the Netica-C API to establish the distributions in the target Netica BN model.

The application of probabilistic evidence is not straightforward in Netica, or indeed, in most commercial BN software [14], since it requires the application of evidence using a dummy node which is a child node of the one to which the evidence is being applied [28]. It is even more complicated if the new evidence is a set of multiple probabilistic evidences to be applied to a number of target nodes – this is predominantly the case when deriving a set of new distributions from a GIS system. In theory is a straightforward application of Bayes' rule to each node in turn. However the process is noncommutative since the order in which multiple CPTs are adjusted to set target nodes to their desired distribution is significant. In other words, the application of Bayes' rule to set the desired distribution to the second node influences the first and so on. In order to solve this, an iterative proportional fitting procedure as proposed by Pan et al [29] has been applied using a software algorithm written in R. This cycles round each target node, adjusting the dummy nodes' CPTs in turn. It was found that the CPTs rapidly converge such that the correct distribution representing the probabilistic evidence is applied to all the target nodes.

This algorithm, once all the probabilistic evidence from a GIS selection has been applied to the BN model, returns the new posterior distributions to the GIS interface for the user to evaluate.

5. DISCUSSION

In this section some key results obtainable from a GIS/BN integrated model are presented. These results are the probability distributions for any variable in the integrated BN model. As such the results are not unlike those we have previously published [10] [11]. The significance of this approach is discussed with regard to spatial decision making for the deployment of low carbon technologies. The further development of this approach is discussed in the context of the low carbon transition, new value propositions for decentralised energy business models and the management of risk.

Figure 8 shows the model's posterior distributions when a specific census area is selected.

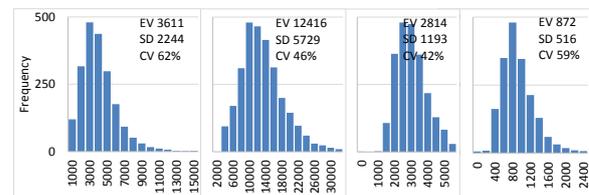


Figure 8 Posterior distributions of key parameters following selection of a census area: (from left to right) Electricity consumption, Gas consumption, PV yield, Self-consumption

As reported in [11] these output distributions can be used to furnish probabilistic discounted cash flow analysis and energy affordability indicators such as for example the percentage of household income spent on fuel following the installation of PV. The fact that these distributions can be obtained for any selection of properties in a GIS model equipped with appropriate vector layers and attribute data provides a powerful evaluation tool for decision and policy makers.

One of the benefits of this technique is that it provides a means of accessing the power of probabilistic graphical models while retaining a GIS interface that may be more familiar or more intuitive to stakeholders in the decision-making process. This paper is therefore presents the development of an integrated GIS/BN model as a logical development of a probabilistic analysis of a low carbon intervention such as PV.

The approach is very extendable. It can be easily adapted to yield a number of other impact indicators which will benefit from a spatially disaggregated analysis. For example the evaluation of low voltage grid impacts is a candidate application since there are a number of studies which predict grid impacts such as voltage and thermal issues as a function of the penetration of PV [30]. This requires both an assessment of demand and generation which the model can assess using the selection of building stock.

A third area is the evaluation of community wide cash flows for the evaluation of virtual power flows and the influence of energy storage on the value proposition which can be realised. It is clear that a probabilistic approach is required for an accurate evaluation of risk.

6. CONCLUSION

A Bayesian network which probabilistically estimates generation, demand and direct self-consumption for

domestic PV has been successfully integrated with a GIS system. The latter has been furnished with a number of commercial and publically available datasets which allow the extraction of probability distributions based on a selection of GIS features.

The problem of supplying these multiple probabilistic evidences to the BN has been solved using an iterative algorithm. The model then furnishes a user with posterior distributions of all variables based on the selection of GIS features. This has potential for further development to incorporate other low carbon technologies and evaluate new business models such as virtual power stations and integrate energy storage using a method which endogenises uncertainties and helps evaluate risk.

7. ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom through EPSRC grants EP/K022229/1 (WISE PV - Whole System Impacts and Socio-economics of wide scale PV integration) and EP/K02227X/1 (PV2025 - Potential Costs and Benefits of Photovoltaics for UK-Infrastructure and Society).

8. REFERENCES

- M. Raugi and E. Leccisi, "A comprehensive assessment of the energy performance of the full range of electricity generation technologies deployed in the United Kingdom", *Energy Policy*, vol. 90, pp. 46-59, 2016.
- A. Campoccia, L. Dusonchet, E. Telaretti and G. Zizzo, "An analysis of feed-in tariffs for solar PV in six representative countries of the European Union", *Solar Energy*, vol. 107, pp. 530-542, 2014.
- Department of Energy and Climate Change, "Solar photovoltaics deployment", gov.uk, 2016. [Online]. Available: <https://www.gov.uk/government/statistics/solar-photovoltaics-deployment>. [Accessed: 31- May-2016].
- Department of Energy and Climate Change, "Consultation on a review of the Feed-in Tariffs scheme", 2015. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469476/Consultation_on_a_Review_of_feed-in_tariff_scheme.pdf. [Accessed: 31- August- 2015].
- Department of Energy and Climate Change, "Review of the Feed-in Tariffs Scheme - Government Response", 2015. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/487300/FITs_Review_Govt_response_Final.pdf. [Accessed: 06-January-2016].
- B. Goss, R. Gottschalg and T. Betts, "Uncertainty Analysis of Photovoltaic Energy Yield Prediction", in *8th Photovoltaic Science Application and Technology (PVSAT-8) Conference and Exhibition*, Newcastle, England, 2012, pp. 157-160.
- J. Taylor, J. Leloux, L. Hall, A. Everard, J. Briggs and A. Buckley, "Performance of Distributed PV in the UK: a Statistical Analysis of over 7000 Systems", in *EU PVSEC*, Hamburg, 2015, pp. 2271-2276.
- G. Saluja and T. Muneer, "An Anisotropic Model for Inclined Surface Solar Irradiation", *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 201, no. 1, pp. 11-20, 1987.
- R. Cherrington, V. Goodship, A. Longfield and K. Kirwan, "The feed-in tariff in the UK: A case study focus on domestic photovoltaic systems", *Renewable Energy*, vol. 50, pp. 421-426, 2013.
- P. Leicester, C. Goodier and P. Rowley, "Probabilistic analysis of solar photovoltaic self-consumption using Bayesian network models", *IET Renewable Power Generation*, vol. 10, no. 4, pp. 448-455, 2016.
- P. Leicester, C. Goodier and P. Rowley, "Probabilistic evaluation of solar photovoltaic systems using Bayesian networks: a discounted cash flow assessment", *Prog. Photovolt: Res. Appl.*, p. n/a-n/a, 2016.
- P. Leicester, "The Development Of Object Oriented Bayesian Networks To Evaluate The Social, Economic And Environmental Impacts Of Solar PV", PhD Thesis, Loughborough University, 2015.
- D. Landuyt, S. Broekx, R. D'hondt, G. Engelen, J. Aertsens and P. Goethals, "A review of Bayesian belief networks in ecosystem service modelling", *Environmental Modelling & Software*, vol. 46, pp. 1-11, 2013.
- A. Mrad, V. Delcroix, S. Piechowiak, P. Leicester and M. Abid, "An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence", *Applied Intelligence*, vol. 43, no. 4, pp. 802-824, 2015.
- OS MasterMap Topography Layer [GML geospatial data], Coverage: UK, Updated Jan 2016, Ordnance Survey, GB. Using: EDINA Digimap Ordnance Survey Service, <<http://edina.ac.uk/digimap>>, Downloaded: January 2016
- OS MasterMap Building Heights [CSV geospatial data], Coverage: UK Various, Updated Dec 2014, Ordnance Survey, GB. Using: EDINA Digimap Ordnance Survey Service, <<http://edina.ac.uk/digimap>>, Downloaded: January 2016
- OS AddressBase [CSV Geospatial Data] Coverage: UK, Updated Dec 2013, Available from Ordnance Survey, Adanac Drive, Southampton SO16 0AS
- Geoinformation group datasets
- D. Palmer, I. Cole, R. Gottschalg and T. Betts, "Assessment of potential for photovoltaic roof installations by extraction of roof tilt from light detection and ranging data and aggregation to census geography", *IET Renewable Power Generation*, vol. 10, no. 4, pp. 467-473, 2016.
- B. Anderson, "Estimating small-area income deprivation: an iterative proportional fitting approach". In: Tanton, R., Edwards, K. (Eds.), *Spatial Microsimulation: A Reference Guide for Users, Understanding Population Trends and Processes*, vol. 6. Springer, Netherlands, pp. 49-67, Chapter 4, 2013.
- EHS data reference
- K. Steemers and G. Yun, "Household energy consumption: a study of the role of occupants", *Building Research & Information*, vol. 37, no. 5-6, pp. 625-637, 2009.
- S. Kelly, "Do homes that are more energy efficient consume less energy?: A structural equation model of the English residential sector", *Energy*, vol. 36, no. 9, pp. 5610-5620, 2011.

- 24 ECJRC (European Commission Joint Research Centre), "Photovoltaic Geographical Information System (PVGIS) Version 4", [online computer program] Available from: <http://re.jrc.ec.europa.eu/pvgis/> [Accessed 12-Jan-2016].
- 25 G. Colantuono, A. Everard, L. Hall and A. Buckley, "Monitoring nationwide ensembles of PV generators: Limitations and uncertainties. The case of the UK", *Solar Energy*, vol. 108, pp. 252-263, 2014.
- 26 R Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, [online computer program] Available from <https://www.R-project.org/>.
- 27 R. Almond "RNetica: R interface to Netica(R) Bayesian Network Engine", R package version 0.4-4, 2015, [online computer program] Available from <http://pluto.coe.fsu.edu/RNetica>
- 28 J. Pearl. "Jeffery's Rule, Passage of Experience, and Neo-Bayesianism". In H.E. et al. Kyburg, Jr., editor, *Knowledge Representation and Defeasible Reasoning*, 245-265. Kluwer Academic Publishers. 1990.
- 29 R. Pan, Y. Peng, Z. Ding, "Belief update in Bayesian networks using uncertain evidence. In: *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)*, pp 441-444, 2006.
- 30 A. Navarro and L.F. Ochoa, "Probabilistic impact assessment of low carbon technologies in LV distribution systems," *IEEE Transactions on Power Systems*, vol 31, no 3, p 2192-2203, 2016.