USING A BAYESIAN NETWORK TO EVALUATE THE SOCIAL, ECONOMIC AND ENVIRONMENTAL IMPACTS OF COMMUNITY RENEWABLE ENERGY

Philip A. Leicester[1,2], Chris I. Goodier[1], Paul Rowley[2]

1. School of Civil and Building Engineering, Loughborough University, Epinal Way, Loughborough, LE11 3TU, UK.

2. Centre for Renewable Energy Systems Technology, School of Electronic, Electrical and Systems Engineering, Loughborough University, Epinal Way, Loughborough, LE11 3TU, UK.

ABSTRACT

Social, economic and environmental (**SEE**) impacts resulting from the adoption of solar PV have been modelled at a community scale for the first time using a probabilistic graphical model in the form of a Bayesian Network (**BN**). Model parameters required to conceptualise this multi-disciplinary problem domain are characterised by uncertainty due to stochastic variability, measurement and modelled data errors, or missing or incomplete information. A **BN** conveniently represents the model parameters and the associations between them and endogenises the uncertainty in probability distribution functions or mass functions.

The theory and method of construction of an object-oriented **BN** which encapsulates a number of **SEE** parameters is described. This is used to model small urban areas as potential adopters of solar PV technology. The **BN** has been populated with modelled and empirical quantitative data from a variety of disciplines to create an inter-disciplinary knowledge representation of the problem domain.

The model has been used to explore a number of scenarios whereby ‘observations’ are made on one or more variables of interest thus altering their prior probability distribution. The updated or posterior distributions of all the other variables are then recalculated using inference algorithms. Results are presented which show the utility of this approach in diagnostic and prognostic inference making. For example it is shown that Solar PV can have a small but significant impact on energy poverty.

It is concluded that the adoption of a BN modelling approach that endogenises uncertainty, and reduces investment and policy risks associated with energy technology interventions within communities, can act as a useful due diligence and decision support tool for a number of private, public and community sector stakeholders active in this sector, in particular key decision and policy makers.

Keywords: Solar PV, Bayesian Networks, Community Energy, Uncertainty

INTRODUCTION

Community scale energy efficiency and renewable energy technologies are seen as a valuable means for contributing to a number of energy policy objectives, and are benefitting from a range of financial support mechanisms internationally. These incentives in turn are resulting in rapidly increasing investment in the community scale renewables sector. However, significant uncertainty exists with respect to the potential impacts of community scale renewables in terms of specific policy goals, including actual (as opposed to projected) greenhouse gas reductions, renewable energy generation capacity and socio-economic benefits such as fuel poverty alleviation. This uncertainty represents significant risk for investors and policy makers alike. A multivariate problem domain characterised by uncertainty is ideal for representation by a probabilistic graphical model and more specifically by a Bayesian Network (**BN**) (Pearl, 1990).

The aim of the broader research project to which the work here contributes is to use a **BN** methodology to evaluate the potential **SEE** impacts of low carbon interventions for urban areas in order to endogenise uncertainty in the modelled outputs and thus explore risk. A **BN** as a decision and policy making tool in this arena will be evaluated.

This paper demonstrates how a candidate **BN** was constructed. The two key elements of the method are discussed; firstly the construction of the **BN** to represent the problem domain as a qualitative conceptual model; and secondly the elicitation of quantitative data to define the marginal and conditional probabilities for all the variables in order to construct an accurate knowledge representation. Some preliminary outputs of the model are presented. Finally the efficacy of **BN**s as a tool for decision and policy making pertaining to the deployment of community renewable energy is discussed.

BAYESIAN NETWORK THEORY

A **BN** is encoded over a directed acyclic graph **(DAG)** in which the vertices (nodes) correspond to random variables and directed edges represent direct dependencies between them (see Figure 1). A directed edge from a parent node $\boldsymbol{A}$ to child node $\boldsymbol{B}$ implies that variable $\boldsymbol{B}$ has a causal or influential dependence on variable $\boldsymbol{A}$.

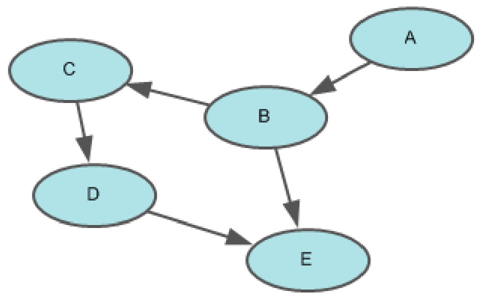

Figure 1. Bayesian Network modelled as a directed acyclic graph

The variable is represented as a *probability mass function* (**pmf**) which gives the probability of each disjoint state. A variable **A**, with n discrete states, a_1, a_2, a_3... a_n, has a **pmf**, $P(\boldsymbol{A})$, represented by a set as in equation 1.

$$P(\boldsymbol{A}) \equiv \{p(\boldsymbol{A} = a_i)\ \forall i = 1,2,3 \ldots n\} \qquad \textit{Equation 1}$$

The dependency of a child variable on a parent variable is modelled using a conditional probability table (**cpt**) $P(\boldsymbol{B}|\boldsymbol{A})$, whereby each state of a child variable **B** has a probability conditional on the state of the parent variable **A**. The joint probability distribution (**jpd**), $P(\boldsymbol{A},\boldsymbol{B})$, is calculated using the fundamental rule from the prior **pmf** of the parent and the **cpt** (equation 2).

$$P(\boldsymbol{A},\boldsymbol{B}) = P(\boldsymbol{B}|\boldsymbol{A}) \cdot P(\boldsymbol{A}) \qquad \text{Equation 2}$$

Using the **jpd,** the prior **pmf** of the child variable **B** can be deduced by the process of marginalisation which involves summing the probability of the child variable for each value of the parent:

$$P(\boldsymbol{B}) = \sum_{i=1}^{m} P(\boldsymbol{B}, a_i) \qquad \text{Equation 3}$$

These techniques, which follow from the axioms of probability, can be applied to all the variables in a BN. The fundamental rule yields the chain rule which can be factorised to a more tractable form. Thus the **jpd** for all the variables in the **BN** shown in Figure 1 can be represented using the factorised chain rule as in equation 4.

$$P(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D},\boldsymbol{E}) = P(\boldsymbol{A})P(\boldsymbol{B}|\boldsymbol{A})P(\boldsymbol{C}|\boldsymbol{B})P(\boldsymbol{D}|\boldsymbol{C})P(\boldsymbol{E}|\boldsymbol{B},\boldsymbol{D}) \qquad \text{Equation 4}$$

The **pmf** for each variable can be calculated by a repeated process of marginalisation. Such a calculation produces the prior probability distribution - a measure of the prior uncertainty - of each variable. The utility of a **BN** is realised when one or more variables are fixed to a particular state (hard evidence), or state probabilities are adjusted in the light of new observations (soft evidence). The **jpd** can be updated and the **pmf**s for all other variables recalculated using Bayes Rule and the chain rule. The new posterior distributions allow the **BN** to be used as a decision support tool.

For a large **BN**, with say 20 variables each with 10 discrete states, the number of elements in the **jpd** would number 10^{20}. Its calculation is intractable and computationally NP-hard (Cooper, 1990). Software algorithms enable **pmf**s to be calculated without determining the entire **jpd**, rendering the problem feasible for all **BN** structures (Jensen et al., 1990).

Software packages are available which allow the encoding of a **BN** and the execution of rapid inference calculations. In this work AgenaRisk® was used. This software allows the encoding of hybrid **BN**s which have both discrete and continuous random variables (Neil et al., 2005). Continuous variables, parameterised by a probability distribution function (**pdf**), are automatically discretised (converted to intervals) to yield a **pmf**. AgenaRisk® makes use of dynamic discretisation algorithms to create non-uniform intervals. This ensures that narrower intervals are used in regions of the **pdf** where broad intervals would lead to approximate inference. AgenaRisk® thereby maximises the fidelity of the resultant discretised **pmf** to the **pdf** without a "*heavy cost of computational complexity*" (Neil et al., 2007).

CONSTRUCTING THE BN

The first task is to construct the qualitative component of the **BN**, namely the **DAG**, by deciding upon the key variables of interest and determining the causal influence between them using expert knowledge in a fashion similar to causal mapping (Goodier et al., 2010). For a large interdisciplinary problem domain it is convenient to break the graph into smaller connected sub-domains to create an object-oriented BN (Koller and Pfeffer, 1997; Molina et al., 2010). Each object encapsulates a more granular **BN** in which a number of variables define the object's domain. The prior, or posterior, **pmf** of an object's variables can act as inputs for other objects thus creating a declarative object-oriented application. This approach was adopted here and eight objects which encapsulate a total of 40 variables have been defined, with their relationships shown in Figure 2. These are discussed further below.

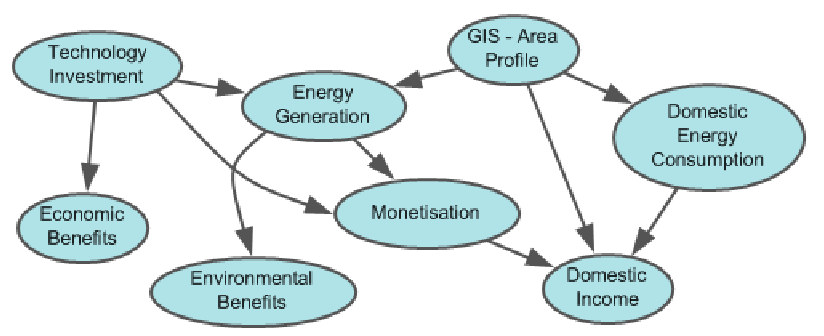

Figure 2. Object-oriented BN for community renewables

The second task is to populate each node with quantitative probabilistic data. For each variable a node probability table (**npt**) needs to be entered. For variables with parents, the data must convey conditional probabilities for each parent state. Preparing data for entry into the **npt** requires processing and conditioning. Occasionally it is convenient to parameterise a dataset into a continuous probability distribution using multivariate fitting algorithms.

In the following section the resulting **BN**'s objects are described. Some key data sources used to populate the **npt**s of variables are described and posterior **pmf**s resulting from observations are demonstrated.

RESULTS AND DISCUSSION

1. **GIS Area Profile** - In this work the geographic unit of analysis was the Lower Super-Output Area (**LSOA**), which is derived from socially homogeneous UK census output areas (Martin et al., 2001) and comprises, on average, 600 dwellings. Variables from several sources have been combined to create a Graphical Information system (**GIS**). Thus far, 4 **LSOA**s have been included (Figure 3). Geo-location, size, aspect and pitch of roofs have been derived using Light Detection and Ranging (LiDAR). Building type, height, footprint and age have been obtained from a commercial database. A **pmf** of the domestic household income has been estimated using spatial micro-simulation modelling. This has been estimated using census and family resource survey data (Anderson, 2013) and fitted with a β-distribution for ease of entry into AgenaRisk®. Figure 4 shows the **pmf** and fitted **pdf** for household income in LSOA E01025703 (Loughborough).

Figure 3. LSOA locations in England

2. **Technology Investment** - This object models the probabilistic relationships between technology costs, loan finance and discount rates to give net present value distributions.

3. **Energy Generation** - This object takes as a key input the solar potential of the **LSOA** from the **GIS** object and uses PVGIS CM-SAF model (Huld et al., 2012) to calculate the solar potential **pmf** of the **LSOA.** The results for LSOA E01018870 (Camborne) are shown in Figure 5.

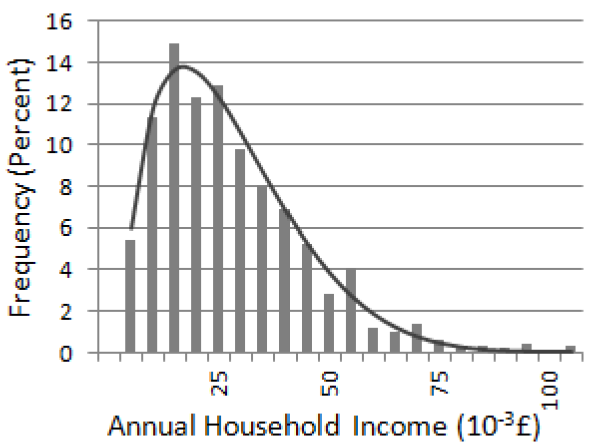

Figure 4. Distribution of annual household income

The deterministic PVGIS Yields are combined with empirical data to estimate system losses in a statistically robust way to introduce a realistic probabilistic measure of solar PV yield for any one property. This is coupled with empirical data for typically deployed UK Solar PV system ratings to calculate probabilistic annual yields.

4. **Monetisation** - The UK FiT scheme pays PV owners a generation and export tariff. This object takes as an input the energy generated and monetises this using current tariffs. A **pmf** of the self-consumption of the PV generated electricity has been estimated using data from Solar PV field trials which enables a **pmf** for export income to be estimated.

5. **Domestic Energy Consumption** - key variables in this object are domestic electricity and gas consumption. Inputs from other objects are the household income and property attributes from the GIS object and building energy performance. Fuel consumptions conditional on the

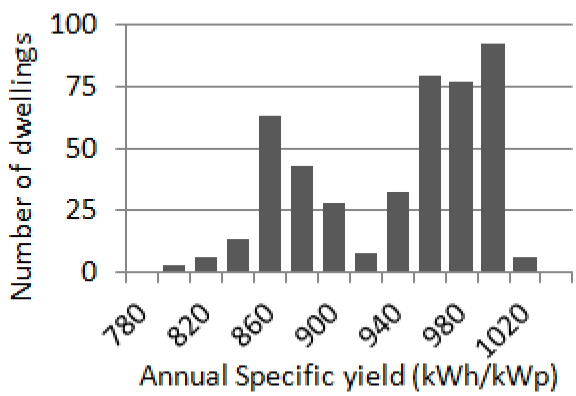

Figure 5. Solar potential for LSOA E01018870,

building performance and household income were derived from literature data (DEFRA, 2005) and deconvolution methods.

6. **Domestic Income** - The income distribution from the GIS object is combined with cash flows from the monetisation and energy consumption objects. This object provides a **pmf** of the change in household income as a probabilistic economic indicator.

7. **Environmental Benefits** - Using carbon intensity of displaced grid electricity, this object delivers a **pmf** of the carbon emission reduction for the technology.

8. **Economic Benefits** - Cash flows into business from the Technology Investment object can be used to estimate business income, growth and employment creation.

The model allows variables from a range of knowledge domains to be rendered as distributions in the form of **pmf**s. Such a representation of uncertainty is commensurate with Bayesian statistical inference. This is potentially more intuitive for stakeholders since a **pmf** gives an immediate and tangible variability of model parameters as opposed to alternative statistical inference methods using p-values and confidence intervals (Iversen, 1984). There are numerous aspects of this that can be demonstrated using this model across a range of **SEE** indicators e.g. carbon emission savings, domestic household income impacts and contributions to business growth. Figure 6 shows the **pmf**s for income from displaced electricity saving under 3 observations of solar PV system rating.

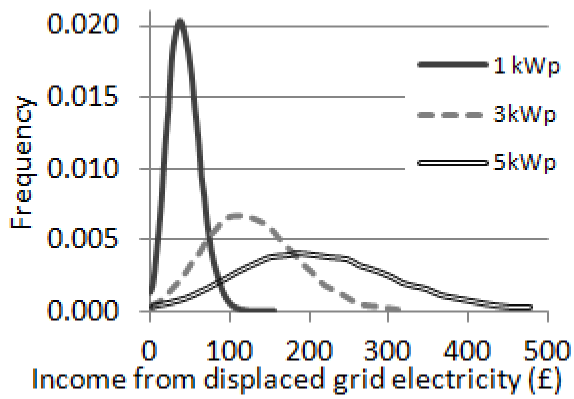

*Figure 6. **pmf**s of saving due to displaced grid electricity in LSOA E0102570, Loughborough.*

kWp	%
0	22.9
1	20.5
3	17.8
5	15.0

Table 1. Percentage households spending over 10% on fuel.

In the UK context one pressing policy concern is fuel (energy) poverty (Boardman, 2012) and questions arise as to whether microgeneration can have an impact (Walker, 2008). Posterior **pmfs** for the percentage of household income spent on fuel have been generated by the **BN** for several PV system rating observations. From these the percentage of households spending over 10% of their income on fuel has been calculated. The results are shown in Table 1. This demonstrates how the **BN** allows scenarios to be tested and recalculates the posterior distributions accordingly.

CONCLUSIONS AND FURTHER WORK

An extensive interdisciplinary knowledge representation for the deployment of solar PV in four communities has been created. The probabilistic assessment of a number of **SEE** parameters has been made, providing a powerful inference making tool to aid multi-criteria decision making (**MCDM)**. This robust whole-system approach facilitates multi-scale (e.g. building, community, or city) analysis whilst managing constraints such as cost or carbon emissions (Rowley et al., 2013). Further development will incorporate probabilistic impact distributions for a range of domestic low carbon technical and behavioural interventions. This can be further enhanced by adding utility and value nodes (Delcroix et al., 2013) for various **SEE** indicators to create a triple bottom line **MCDM** tool to aid deliberative policy making.

ACKNOWLEDGEMENTS

The authors are immensely grateful to James Eddy, Blue Sky Solutions, Coalville; Alastair Buckley and the Sheffield Solar Hub, Sheffield University; Ben Anderson, Essex University. This work is funded by the Loughborough University Centenary Fund.

REFERENCES

Anderson, B., 2013. Estimating small area income deprivation : an iterative proportional fitting approach, in: Tanton, R., Edwards, K. (Eds.), Spatial Microsimulation: A Reference Guide for Users. Springer, pp. 49–67.

Boardman, B., 2012. Synthesis fuel poverty: Lessons learnt, actions needed. Energy Policy 49, 143–148.

Cooper, G.F., 1990. The computational complexity of probabilistic inference using bayesian belief networks. Artificial Intelligence 42, 393–405.

Delcroix, V., Sedki, K., Lepoutre, F.-X., 2013. A Bayesian network for recurrent multi-criteria and multi-attribute decision problems: Choosing a manual wheelchair. Expert Systems with Applications 40, 2541–2551.

Department for Environment and Rural Affairs, Building Research Establishment, The Energy Saving Trust, 2005. Energy Use in Homes - Fuel Consuption.

Goodier, C., Austin, S.A., Soetanto, R., Dainty, A.R.J., 2010. Causal mapping and scenario building with multiple organisations. Futures 42, 219–229.

Huld, T., Müller, R., Gambardella, A., 2012. A new solar radiation database for estimating PV performance in Europe and Africa. Solar Energy 86, 1803–1815.

Iversen, G.R., 1984. Bayesian Statistical Inference. SAGE Publications.

Jensen, F.V., Olesen, K.G., Andersen, S.K., 1990. An Algebra of Bayesian Belief Universes for Knowledge-Based Systems. Networks 20, 637–659.

Koller, D., Pfeffer, A., 1997. Object-Oriented Bayesian Networks, in: Proceedings of the Thirteenth Annual Conference on Uncertainty in AI (UAI-97). pp. 302–313.

Martin, D., Nolan, A., Tranmer, M., 2001. The application of zone-design methodology in the 2001 UK Census. Environment and Planning A 33, 1949–1962.

Molina, J.L., Bromley, J., García-Aróstegui, J.L., Sullivan, C., Benavente, J., 2010. Integrated water resources management of overexploited hydrogeological systems using Object-Oriented Bayesian Networks. Environmental Modelling & Software 25, 383–397.

Neil, M., Fenton, N., Tailor, M., 2005. Using Bayesian networks to model expected and unexpected operational losses. Risk analysis 25, 963–72.

Neil, M., Tailor, M., Marquez, D., 2007. Inference in Hybrid Bayesian Networks using Dynamic Discretisation. Statistics and Computing 17, 219–233.

Pearl, J., 1990. Reasoning Under Uncertainty. Annual Review of Computer Science 4, 37–72.

Rowley, P., Gough, R., Doylend, N., Thirkill, A., Leicester, P.A., 2013. From Smart Homes to Smart Communities : Advanced Data Acquisition and Analysis for Improved Sustainability and Decision Making, Proceedings of the I-Society Conference, Toronto, June 2013.

Walker, G., 2008. Decentralised systems and fuel poverty: Are there any links or risks? Energy Policy 36, 4514–4517.