

Probabilistic analysis of solar photovoltaic self-consumption using Bayesian network models

ISSN 1752-1416

Received on 6th August 2015

Revised on 11th November 2015

Accepted on 28th November 2015

doi: 10.1049/iet-rpg.2015.0360

www.ietdl.orgPhilip A. Leicester^{1,2} ✉, Chris I. Goodier², Paul N. Rowley¹¹Centre for Renewable Energy Systems Technology (CREST), School of Electronic, Electrical and Systems Engineering, Loughborough University, Loughborough LE11 3TU, UK²School of Civil and Building Engineering, Loughborough University, Loughborough LE11 3TU, UK✉ E-mail: P.A.Leicester@lboro.ac.uk

Abstract: To assess the systemic value and impacts of multiple photovoltaic (PV) systems in urban areas, detailed analysis of on-site electricity consumption and of solar PV yield at relatively high temporal resolution is required, together with an understanding of the impacts of stochastic variations in consumption and PV generation. In this study, measured and simulated time-series data for consumption and PV generation at 5 and 1 min resolution for a large number of domestic PV systems are analysed, and a statistical evaluation of self-consumption (SC) carried out. The results show a significant variability of annual PV SC across the sample population, with typical median annual SC of 31% and inter-quartile range of 22–44%. About 10% of the dwellings exceed an SC of 60% with 10% achieving 14% or less. The results have been used to construct a Bayesian network model capable of probabilistically analysing SC given consumption and PV generation. This model provides a basis for rapid detailed analysis of the techno-economic characteristics and socio-economic impacts of PV in a range of built environment contexts, from single building to district scales.

1 Introduction

Over the course of the past decade, photovoltaic (PV) technology has become a mainstream form of electricity generation in many regions as PV energy costs have converged toward grid parity, whilst adoption has been further incentivised by subsidy schemes. This rapid increase in PV penetration has given rise to a need for a deeper understanding of PV's impacts and benefits, both on incumbent technical infrastructure and on non-technical aspects such as net household fuel costs or green house gas (GHG) emissions [1]. Analysis of such aspects requires a deeper insight than has been attained to date into the performance of PV in the context of building energy consumption, especially given the diversity of settings within which PV is found in the built environment. Therefore, in order to provide a valid basis for subsequent analysis of such factors as network impacts, potential energy storage requirements and return on investment, this paper aims to deliver a detailed statistical evaluation of PV self-consumption (SC), using the domestic sector in the UK as a case study. The quantification of SC (defined as the fraction of PV energy which is used to do meaningful on-site work, rather than exported [2]) is important for several key reasons:

- (i) It can influence the economic viability of consumer co-located PV.
- (ii) It has a direct impact on the low- and medium-voltage electricity grid.
- (iii) It supports decision makers in the appropriate design of PV-related regulatory and fiscal policy mechanisms.

In building contexts, techno-economic outcomes are largely influenced by the price differential between the value of exported electricity and the cost of imported electricity. If the latter is higher, SC is generally favourable to the bill payer, since the excess cost of imported electricity is avoided [3]. Accurate, case-specific techno-economic analysis of SC is thus particularly relevant where export payments are less than electricity tariffs.

This applies to a number of feed-in tariff regimes, which, despite their name, make payments for gross generation, rather than purely for exported generation [4]. This is the situation in this UK study where a generation tariff of £0.13 is augmented by an export tariff of £0.04 per kilowatt hour (kWh). Recently for some instruments such as the German Feed-in Tariff (EEG), SC is incentivised by premium payments when it exceeds 30% [5]; in contrast, maximising SC offers no economic advantage under full net-metering arrangements, as exist in some US states, since export and import have the same value [6]. Thus, accurate quantification of SC is required in order to assess the return for investors (since avoided electricity import costs can comprise a significant proportion of net revenue) and costs related to use-of-system charges for exported energy or network reinforcement overheads [7]. Furthermore, the detailed understanding of SC has a bearing on future energy system outcomes facilitated by smart meters and smart-grids such as the potential for 'intelligent' demand shifting, and time-of-use tariffs [8].

In general terms (and in the absence of co-located storage), the lower the aggregated SC, the higher is the aggregated electricity exported to the grid. This can have adverse impacts on grid infrastructure and on the quality of the supply such as over-voltage, phase unbalance and harmonic distortions [9]. However, studies frequently use normative models for both generation and SC based on a standard system configuration and consumption. A comprehensive and robust assessment of grid impacts and commensurate mitigation strategies requires the quantification of SC, and its variability, for assemblies of real dwellings. Storage helps increase SC, but the quantification of instantaneous SC is required for the effective modelling of the efficacy of storage, particularly in quantifying round trip losses [10].

In terms of socio-economic impacts, SC is pertinent across a number of domestic, commercial, or industrial contexts; for example, in tenanted properties, the financial benefit of SC often accrues solely due to the avoided electricity import costs. Thus, the nature of SC is pertinent within analyses of fuel affordability in the domestic context [11], and the overall economic evaluation

of PV [12]. For non-domestic contexts with significant load profile (and thus SC) diversity, such uncertainty can increase investment risk and affect the efficacy of policy or regulatory instruments.

Past studies have shown that SC varies with both generation and demand [13]. In a study of small PV systems, between 1 and 2 kWp, typical observed SC was 600 kWh/year (corresponding to $44 \pm 16\%$ of total yield), whilst for a larger, 4.5 kWp PV system, 770 kWh/year (19% of total yield) was observed in a simulation study to assess energy storage technologies [10]. In a study of 172 domestic PV installations in Germany, with system ratings from 1 to 5 kWp, and annual yields between 430 and 875 kWh/kWp, SC fractions ranging from 20 to 60% were observed [14]. A large number of studies have reported a similar wide variation in SC [15]. Each has been measured for specific domestic contexts with a fixed system rating and annual electricity consumption which do not necessarily capture the variability in the wider population. In the UK, recent measured specific annual yields have been recorded [16] from 600 to 1100 kWh/kWp and domestic PV system ratings over the past 5 years range from 1 to 4 kWp [17]. Meanwhile, measured domestic annual electricity consumption ranges widely up to 25 MWh/year [18]. It is thus apparent that the prediction of SC is not straightforward due to the wide variability of predictor parameters, which in turn are derived from the variety of contexts in which solar PV is installed. Given the stochastic nature of both demand and generation, a deterministic relationship of the form $SC = f(\text{consumption}, \text{generation})$ is inappropriate, not least because important parameters (such as user behaviour effects) are uncertain and not easily quantified [19]. Thus, an approach that endogenises uncertainties due to unknown variables, and in which the probabilistic relationships between SC, consumption and generation are quantified, offers distinct advantages. This can be represented by a conditional probability distribution (CPD) of the form as shown in (1)

$$P(SC|\text{consumption}, \text{generation}) \quad (1)$$

This CPD can be used in a probabilistic graphical model based on a Bayesian network (BN) to evaluate the energy balance between electricity consumption, generation, import and export in a manner which captures the variety of building contexts via the use of appropriate building stock datasets. This model can then be utilised to assess a number of multi-domain system impacts, including network point-of-connection load-flow or socio-economic outcomes such as geographical variances in household net fuel spent [20].

2 PV self-consumption

PV SC (also known as the load match index or cover factor) [21] occurs when PV generation temporally matches or exceeds the building load. The temporal load profile is dependent on such factors as occupant practices and the associated use of appliances [22]. Similarly, the temporal profile of solar PV generation is subject to the predictable motion of the sun modified by unpredictable transient weather characteristics [23]. Figs. 1a and b show simulated demand and generation profile for an arbitrary day in the UK Midlands at two different temporal resolutions. Where demand exceeds generation, electricity is imported, and where generation exceeds demand excess electricity is exported. The resultant SCs are shown by the dashed line. The area under this, divided by the area under the generation curve, yields fractional SCs. These are seen to differ significantly, with an apparent 30% increase in SC when using 1 h data compared with 1 min temporal resolution.

Rapid fluctuations in demand occur due to the stochastic switching of electrical appliances, either automatically or due to occupant behaviour [24, 25]. Whilst under conditions of clear (or consistently overcast) skies, plane-of-array irradiance changes only slowly, under conditions such as partial cloud, rapid fluctuations in

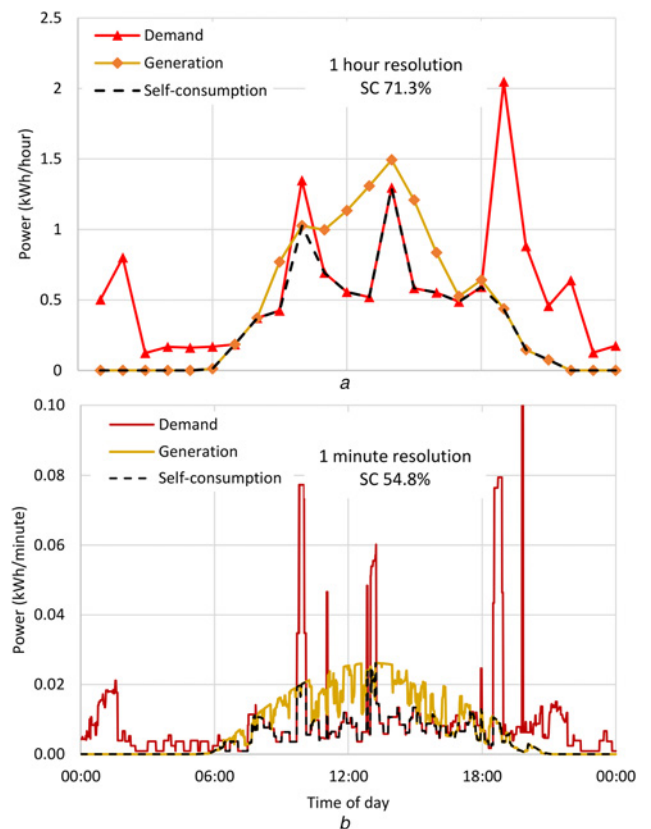


Fig. 1 Sample simulated domestic electricity demand and PV generation profiles at

a 1 min Resolution

b Same data aggregated to 1 h resolution

power output occur due to the attenuation of irradiance by transient clouds, particularly in high wind speed conditions [26].

A limited number of simultaneous direct measurements of PV generation, export and demand have been described [13, 14]. One approach to garnering more data with greater variety is to use demand profile data and estimate the SC fraction using separately obtained PV data. However, demand profile data is commonly measured at time resolutions of 5 min or greater, which has the effect of smoothing the spiky demand profiles. This results in apparent reduced peak loads as illustrated in Fig. 1b, where the sharp demand spikes, observed either side of midday, would have to be partially met by imported electricity. However, when aggregated to 1 h resolution (Fig. 1a), these spikes are no longer discernible and appear adequately covered by on-site PV generation. Thus, in this example, SC is over-estimated at 71.3% using 1 h data, compared with 54.8% when using 1 min data. This effect has been studied [2, 27] and errors as large as 80% have been reported. Thus data of high temporal resolution are required to accurately quantify SC.

There are several distinct features to consider when quantifying SC to adequately capture the probabilistic dependency on generation and consumption:

- The overlap between load and generation is determined by the magnitude of both the energy generation and the energy consumption. The greater the magnitude of either, the greater will be the probability of overlap and thus the magnitude of SC, though this will necessarily approach a limiting value.
- The stochastic nature of temporal demand and generation means that the load match is not readily modelled deterministically; rather it requires a probabilistic analysis and a large data sample to capture the requisite variety inherent in a population.
- The temporal frame for the stochastic events occurs over relatively short high-resolution time-frames, but the socio-economic impacts

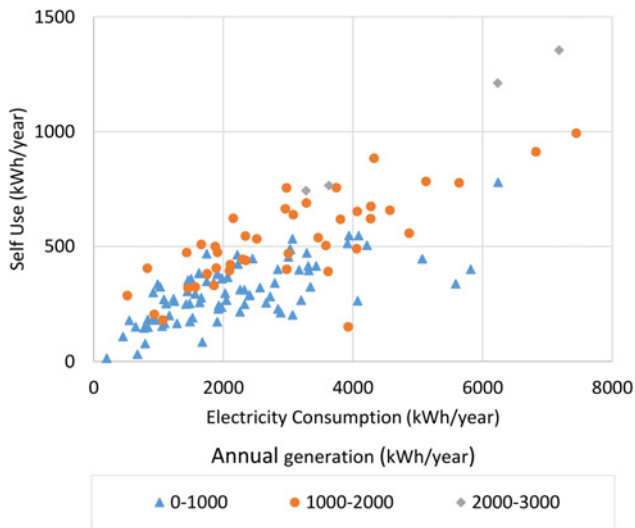


Fig. 2 Annual SC as a function of annual electricity consumption segmented by annual system yield (generation) from the DFT data

are generally modelled over much longer time-frames, typically 1 year.

Therefore, the challenges of garnering sufficient data to evaluate SC and its variability are considerable; the range of annual electricity consumption needs to match that of typical empirical domestic electricity consumption, and for each annual demand value a commensurate range of solar PV generation values needs to be sampled in order to deliver a suitably granular CPD of the form given in (1). Two approaches, one using empirical field data, and another using simulated data have been used in this paper and are presented in the following sections.

3 Measured (empirical) datasets

The UK PV domestic field trial (DFT) dataset [13], comprising 23 months of 5 min time resolution data from 135 domestic PV installations, was initially used to calculate annual SC. System ratings (typically 1 kWp) are considerably lower than those now deployed in the UK, whilst specific yields are lower than those reported for today's systems [16]. In addition, the households in the sample exhibited somewhat lower electricity consumption compared with the general population [18].

However, whilst these data are within a limited parameter space of low generation and low consumption compared with contemporary PV deployment contexts, SC analysis of the dataset demonstrates relevant trends. Figs. 2 and 3 illustrate an increase in SC with both annual household electricity demand and PV generation. The scatter of SC data indicates a high degree of stochasticity, resulting

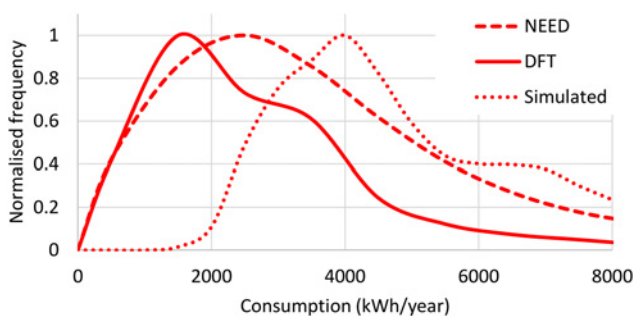


Fig. 3 Comparison of the distribution of electricity consumption in the DFT, simulated data and the NEED framework

Table 1 Starting parameters and ranges used in simulations

Parameter	Lower	Upper
number of residents	1	5
PV system rating, kWp	1	5
azimuth, deg	-90	90
slope, deg	20	45

largely from a wide variety of building occupant behaviours, together with varying PV generation.

This dataset did not comprise a wide, random sample of PV adopters, but a purposeful selection of new-build, tenanted social housing properties. Thus, though expected dependencies and variability are evident, the results may not be representative of those for the wider population.

4 Stochastic simulation of SC

To garner more data at higher generation and consumption, a hybrid stochastic/probabilistic model [22] was used to simulate daily electricity load and generation profiles. This generates 1 min time-step demand data from a set of household appliances randomly assigned to the dwelling, based on published statistics of appliance ownership and ratings. Appliances are categorised into groups: those that run all the time, consuming a base load (e.g. a freezer), and those operated by an active occupant performing a particular activity (e.g. leisure). Active occupancy and activity profiles within the dwelling are simulated stochastically using temporal probabilities derived from a time-use survey for between 1 and 5 residents. Separate probability data were used for weekdays and weekends. The model also predicts the load due to lighting using a seasonally linked lighting simulation module [28].

The software simultaneously creates a PV generation profile [29] from a calculated minute-by-minute clear-sky irradiance attenuated using a clearness index. The latter is stochastically generated for each minute using a transition probability matrix constructed using 1 min time-series of empirical horizontal irradiance data recorded in Loughborough, England, over one calendar year [30]. Array tilt and azimuth are used to calculate irradiance in the plane-of-array (PoA) [31] and a simple system efficiency method is applied to convert PoA irradiance into an estimation of the minute-by-minute AC electrical output of the system.

The simulation cycles through every day of a whole year with a fixed set of start parameters randomly allocated from a uniform distribution between the upper and lower limits (Table 1). In addition a random allocation of appliances and lighting loads is generated.

About 25,000 such simulations are required to generate 100 samples for each 1 MWh/year interval over the consumption range 0–10 MWh/year, generation range 0–5 MWh/year and SC range 0–5 MWh/year. Demand, generation and export were aggregated for each minute of each day and added to a running total for the year. After the last day of the year, the annual totals were saved along with the start parameters and the simulation repeats for another whole year with a new set of random start parameters.

5 Joint probability distribution (JPD) for SC

The data for simulated and empirical annual consumption and PV generation were organised into 500 kWh/year intervals (bins), and SC data into 200 kWh/year intervals. A 3-way contingency table was created with the cell frequency corresponding to the number of samples satisfying the corresponding intervals. To be representative, the marginal distribution for electricity consumption should correspond to the empirical distribution observed in the general population. A comparison of the distribution of electricity consumption in the DFT, simulated data and the National Energy Efficiency Data (NEED) framework [18] is as shown in Fig. 4. It

is evident that the simulation model under-represents low electricity consumers, possibly due to an over-estimation of either active occupancy, the probability of specific energy consuming behaviours, or appliance ownership. All these factors can mitigate against the observation of low electricity consumption. However, the measured DFT data covers this gap, and thus the combined dataset exhibits a realistic range of domestic electricity consumption as observed in the general population.

A graphical representation of the resultant dataset is shown Fig. 4, in which SC is plotted against annual consumption, and with annual PV generation represented by colour coded intervals of 1000 kWh. This quantifies the dependence and the uncertainty of SC on both PV generation and consumption, with a hitherto unreported level of detail. It confirms that SC increases as both consumption and PV generation increase, whilst the high level of variability is indicative of uncertainty inherent in specific parameters ('missing' variables) [32]. To investigate the influence of household occupancy, simulations with a typical 3 kWp system and using six idealised occupancy archetypes [19] were conducted. This allowed the dependency of simulated SC to occupancy parameters to be quantified (Table 2). Little difference between morning and afternoon occupancies (SCs of 44 and 47%, respectively) is observed. Home-comers at 16:00 have 25% SC compared with 19% for those that return at 18:00. In contrast, all-day occupancy achieves 69% SC, compared with an average SC of 15% for all-day unoccupied. The use of idealised occupancy archetypes strongly suggests that the variability in SC is due to different occupancy behaviours, and it is significant that even all-day occupied dwellings rarely achieve 100% SC with a 3 kWp system.

At this point, it should be noted that in this paper we have not sought to introduce an occupancy archetype parameter, but rather, to endogenise the occupancy-related uncertainty for SC predicted solely by consumption and generation. Thus, in the JPD defined by (1), the occupant behaviour remains a hidden variable. In the next section, the use of this JPD in probabilistic modelling is explored.

6 Integrated BN model for domestic solar PV

BNs are used to model causal relationships between variables using conditional probabilities [33]. A BN is represented by a directed acyclic graph (DAG), in which 'nodes' represent variables, and directed 'arcs' from parent nodes to child nodes represent conditional probability relationships between them (Fig. 5). A BN is an expedient representation of a JPD over the whole parameter space, which, using the chain rule (2), can be factorised into the marginal distributions of the leaf nodes (those without any parents such as node A in Fig. 5), and CPDs of each child node V given their parents, π_V [34]. These distributions can be learnt from data or provided directly using an empirical JPD

$$(V_1, V_2, \dots, V_n) = \prod_{i=1}^n P(V_i | \pi_{V_i}) \quad (2)$$

Observations made at one or more nodes can be used to update the probability distributions on target nodes of interest using belief propagation algorithms [34]. Observations applied to one or more nodes are known as 'evidence'; of these there are two key forms. First, a variable V is instantiated to state v_x given evidence e such that $P(V=v_x|e)=1$, referred to as 'hard evidence', or a 'hard finding'. Second, probabilistic evidence may be applied to an observed variable which establishes a new local probability distribution on the variable [35]. Once evidence has been applied to a variable, probabilistic reasoning algorithms update the probability distributions of all dependent nodes to yield new posterior distributions. This facilitates both prognostic and diagnostic inference, depending on whether evidence is applied to a key input or output parameter. A BN can be used as a tool for multi-criteria decision support [36] and reasoning under uncertainty [37] and represent a powerful transdisciplinary knowledge representation and inference tool [38].

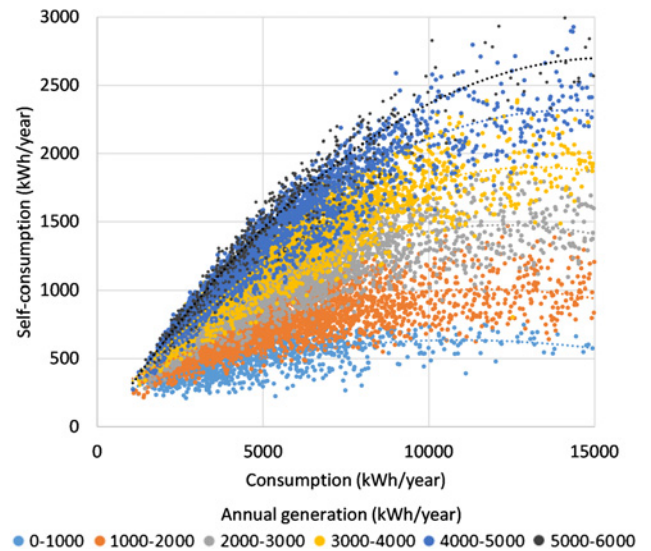


Fig. 4 Annual PV SC as a function of annual electricity consumption segmented by annual system yield

The current work involves the creation of a BN model which utilises the CPD for SC discussed above (Fig. 6a). Here, the consumption and generation input parameters have been set to a uniform distribution since the marginal distributions obtained in the simulations do not conform to any specific empirical evidence; rather they serve to quantify the conditional probabilities. The BN also includes a node to display SC as a percentage. The model can be used to enter hard evidence for consumption and generation in order to generate a new posterior distribution for SC. In Fig. 6b, hard evidence for consumption (4–4.5 MWh/year), and generation (2.5–3 MWh/year) results in a 58% probability of yielding an SC of 800–1000 kWh/year. The posterior distribution, presented as percentage SC is shown in Table 3. In this manner, the BN can be adapted and used to estimate SC for any interval of consumption and generation likely to be observed in a specific building context.

Table 2 Typical appliance load profiles for average domestic household related to occupancy archetypes with the average simulated SC observed

Occupancy archetype	Average SC, %
unoccupied 9:00–13:00	44
unoccupied 9:00–16:00	25
unoccupied 9:00–18:00	19
unoccupied 13:00–18:00	47
all-day occupied	69
all-day unoccupied	15

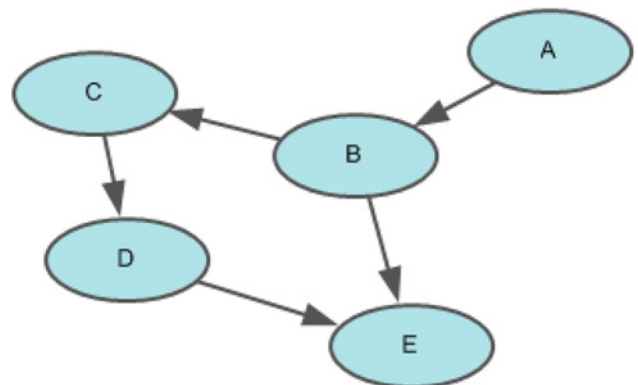


Fig. 5 Example of a DAG

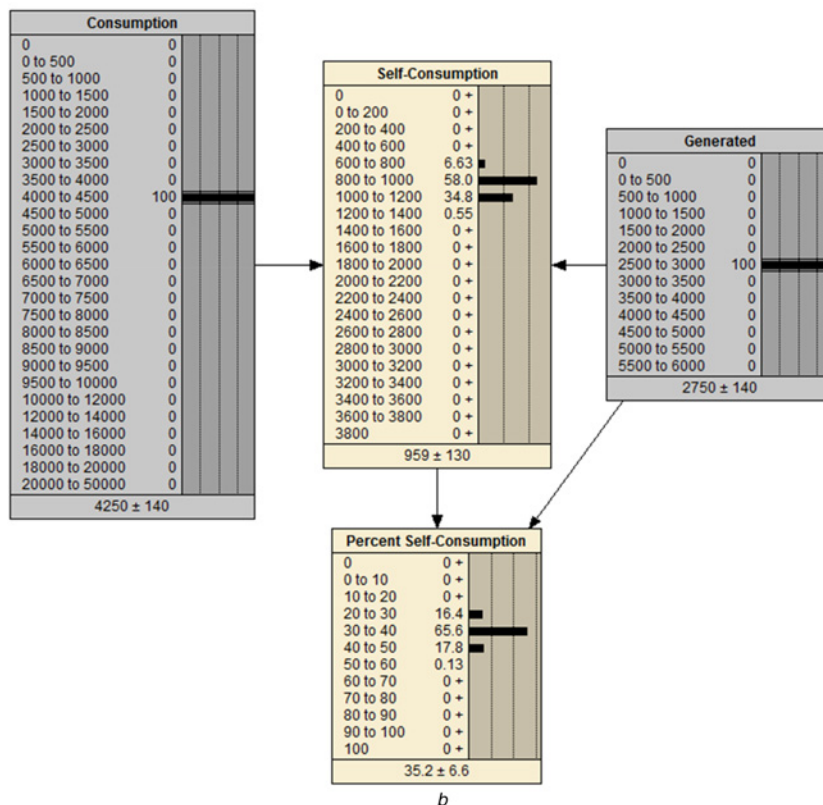
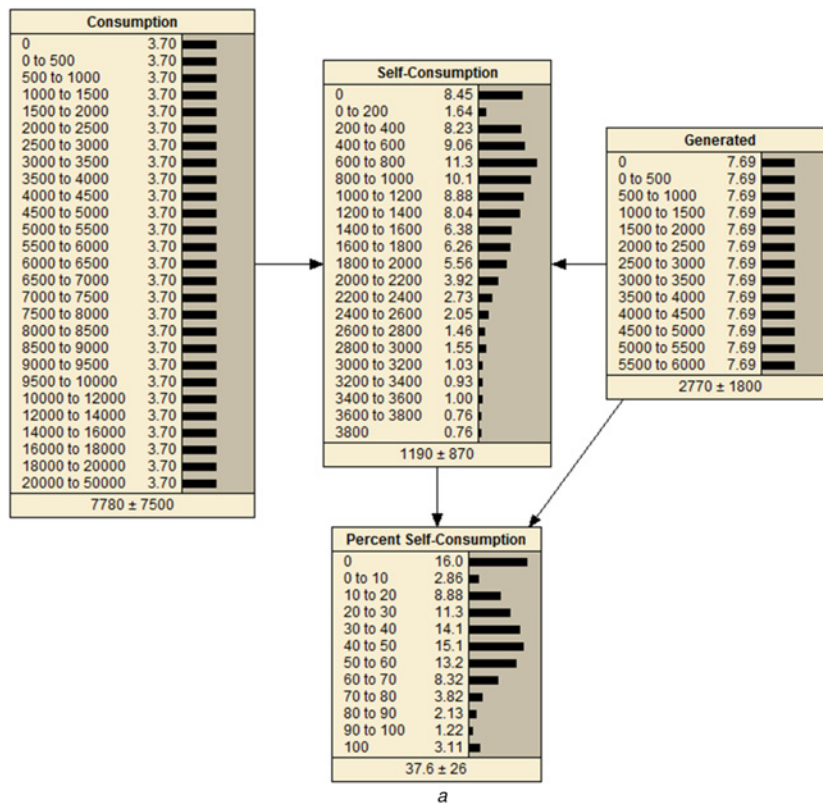


Fig. 6 BN model for SC

a With uniform distribution for consumption and generation
 b Hard evidence for consumption (4–4.5 MWh/year) and generation (2.5–3 MWh/year)

The model proves most versatile when instead of hard evidence for PV yield and electricity consumption, these are given as probabilistic evidence. To this end, two BN models were developed which probabilistically predict building energy consumption and solar PV yield, for four urban areas in the UK

[39]. As inputs, these models take building stock parameters from a third BN model. This model includes data for floor area, building age, building form and geographical region. These in turn are used to predict electricity and gas consumption, whilst roof pitch, orientation and area are used to predict PV yield. The data

Table 3 Example of the posterior probability distribution for SC given hard evidence for generation and PV yield as in Fig. 6b

PV SC, %	Probability, %
20–30	16.4
30–40	65.6
40–50	17.8

have been calibrated using empirical data from the NEED framework [18] and the UK Microgeneration Database [16].

The SC BN model was integrated with these models such that electricity consumption (derived from the building energy consumption model) and PV yield (derived from the PV yield model) are used as its inputs. This is represented schematically in Fig. 7, with each BN shown as an entity (or object), and the interfaces shown as relationships, in an ‘entity relationship diagram’ [40]. This concept of connecting autonomous BNs through common interfaces is referred to as an object-oriented BN [41].

In the present analysis, four urban census areas in England, each comprising ~600 dwellings, have been used to populate the building stock model. Building parameters were obtained by integrating data from the photointerpretation of high-resolution digital aerial photography [42], ordnance survey mapping data [43] and commercial lidar data which provided roof geometries and orientations [44]. The benefit of utilising census areas is that they are used as a basis for various socio-economic studies, and a range of statistical data can be sourced and integrated into the BN model. In the present paper, household income has been integrated into the building stock parameters using an iterative proportional fitting technique [45], which supports the evaluation of socio-economic impacts of solar PV.

7 Quantification of SC

Having furnished the building stock BN model object with probabilistic evidence parameters, posterior distributions are delivered for electricity consumption (Fig. 8a) and PV generation (Fig. 8b). When these distributions are further propagated through

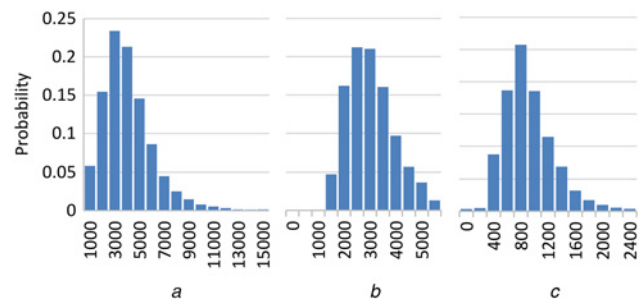


Fig. 8 Posterior distributions for electricity consumption, PV yield and SC (kWh/year) showing EV obtained using the building stock parameters for census areas in Camborne, Cornwall and England

a Consumption (kWh/year)
b PV yield (kWh/year)
c SC (kWh/year)

the BN model, the posterior distribution for SC is delivered (Fig. 8c). These specific distributions are for a census area in Camborne, in the southwest of England.

Table 4 shows the comparative data for the four census areas in this paper. The median value of percentage SC ranges from 31 to 34% with an inter-quartile range of circa $\pm 12\%$. The distribution is positively skewed, with upper and lower decile values of circa 60 and 13% SC, respectively.

The influence on the variability of SC by other nodes can be tested using a variance reduction sensitivity analysis (Table 5). This tests how much the variability of SC is reduced by fixing the values of the other variables in the BN [46]. The results of the sensitivity analysis show that the most influential parameter in this context is electricity consumption, with a 50% SC variance reduction. PV generation in contrast exhibits only a 6% influence on the variability of SC.

Fig. 4 shows that absolute SC increases as PV yield increases, but as a percentage of the yield, it decreases. The median value of SC is lower than that reported in the DFT study [13] at $44 \pm 16\%$. This is attributed to the lower system ratings in the DFT study than those typically deployed today. In comparison, for a system rating node

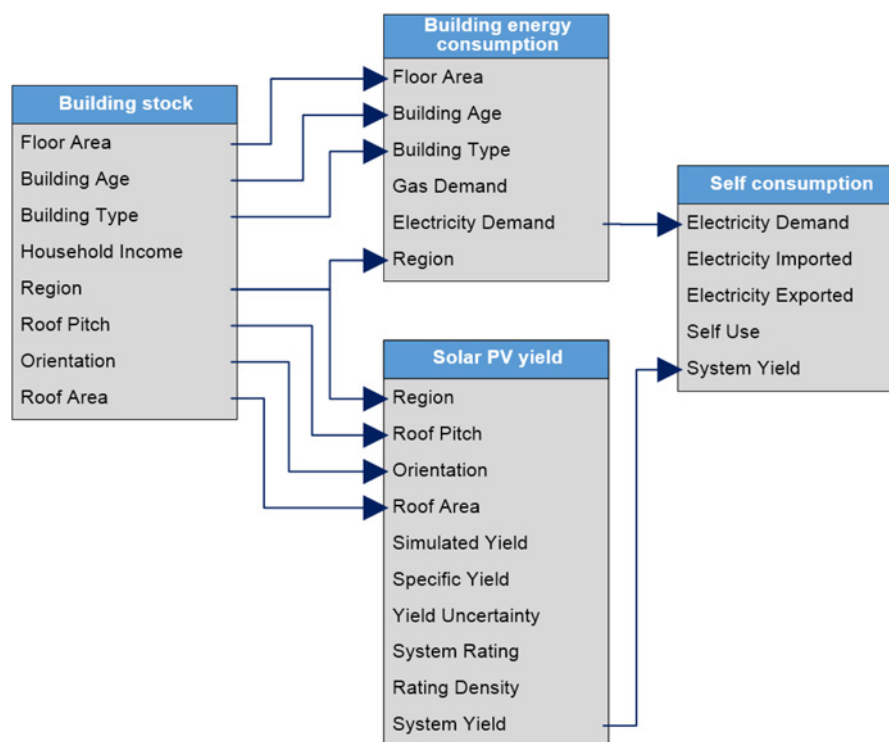


Fig. 7 Entity relationship diagram showing the SC BN model integrated into an object-oriented BN to model the energy system for defined building stock models

Table 4 Results for four census areas in England, UK, showing expected value (EV), standard deviation (SD), coefficient of variation (CV) and percentile points for consumption, generation, SC and per cent SC

		EV	SD	CV, %	Percentile				
					10	25	50	75	90
Camborne	consumption	3623	2371	65	1268	2160	3250	4622	6201
	generated	2814	1192	42	1658	2092	2683	3364	4086
	SC	872	516	59	405	566	769	1044	1375
Loughborough	percentage SC	33	18	54	13	21	30	41	54
	consumption	3689	3198	87	1102	1918	3053	4634	6655
	generated	2609	952	36	1549	1941	2494	3160	3881
Huddersfield	SC	869	573	66	358	523	741	1036	1434
	percentage SC	34	20	57	13	21	31	44	58
	consumption	3741	2693	72	1206	2155	3297	4741	6556
Newcastle	generated	2319	835	36	1325	1710	2225	2825	3452
	SC	830	524	63	371	528	727	976	1319
	percentage SC	37	19	53	15	24	34	47	60
Newcastle	consumption	3264	2254	69	1078	1905	2944	4189	5624
	generated	2390	846	35	1401	1775	2297	2887	3493
	SC	808	553	68	345	501	698	935	1270
	percentage SC	35	20	57	14	22	31	43	58

Table 5 Sensitivity of 'SC' to a finding at another node using a variance reduction sensitivity analysis

Node	Variance reduction	Percent
consumption	1.47×10^5	49.7
generation	1.83×10^4	6.18
roof area	1.28×10^4	4.33
floor area	5524	1.87
specific yield	2244	0.76
census area	706.7	0.239
orientation	571.3	0.193
pitch	183.9	0.0622

value of 1.5–1.75 kWp, the new posterior distribution for SC has a median of $44 \pm 20\%$. Similarly, when the electricity consumption node is set to 2500–3000 kWh/year, and the system rating set to 4.25–4.5 kWp, the mean percentage SC is $20.6 \pm 6.1\%$. This compares with the value obtained by Parra *et al.* [10] of 19%. In contrast, a Swedish study of 200 single family dwellings [3] with a mean electricity consumption and PV generation of 13.8 and 2.8 MWh, respectively, gave an average SC of 63%. When these values are approximated in the relevant BN nodes (consumption of 13.9 and generation of 2.9 MWh) an SC value of $57 \pm 9\%$ is obtained. Thus, the BN constructed in the present paper deliver comparable results to all these various studies.

8 Discussion and conclusions

A probability distribution for SC, conditional on annual electricity consumption and PV yield, has been created with a high granularity, such that it permits the probabilistic prediction of PV SC for the wide variety of consumption/generation scenarios found in various building and occupancy contexts. The model predicts the ranges and typical values of SC found in a number of empirical studies, and accurately quantifies probability distributions for SC given input distributions of PV generation and consumption produced by calibrated models for a number of defined census areas for which the building stock has been parameterised. Calibration of the model for other geographic regions and building stock may be readily conducted using similar empirical PV yield and generation data.

The distributions reported for each census area are similar, despite variations in median PV generation and electricity consumption for each census area. In Camborne, the median generation is 200 kWh/year higher than the next most southerly area, Loughborough. Camborne however, also exhibits a higher median electricity consumption. Similar SCs for all four sites, despite their differing locations, arise as a result of a complex interaction of predictor parameters, including differing PV

installed area, socio-economic aspects (impacting on total power consumption) and irradiation. Together, these act to balance positive and negative influences on SC.

The model has been validated, delivering results comparable with published studies, for cases of both low and high annual domestic electricity consumption. The BN-derived mean SC (~30%) differs significantly from the 50% SC value assumed within the UK feed-in-tariff programme in the absence of an export meter. Only 22% of the dwellings in the housing stock models used in this paper had an SC between 40 and 60%. The implications for the magnitude and distribution of SC are significant in relation to the techno-economic analyses of deployed PV, especially where the value of avoided electricity imports represents a significant cash flow. Simplistic SC estimates should be avoided in future studies of techno-economic analyses involving energy storage and demand shifting. Thus the magnitude and variability of exported electricity available for exploiting by demand shifting or energy storage has been quantified; here, a mean PV export value of 70% has been derived. Significantly, only 10% of dwellings had an SC parameter above 60%, with the occupancy archetype analysis implying that this is a limiting factor for dwellings that are heavily occupied, in this case with 40% of PV generation exploitable using demand side management or energy storage.

The application of a Bayesian modelling approach in the context of building integrated PV, demonstrated for the first time in the current study, illustrates the power of BNs as a means to endogenise uncertainties inherent in such complex systems that feature many variable parameters such as building stock characteristics, PV attributes and occupant behaviours. By propagating parameter uncertainties, the model rapidly calculates the resultant distribution of SC given a set of input conditions. This posterior distribution can then be used as an input to further BN models which evaluate specific outcomes such as techno-economic or socio-economic impacts. For example, the determination of probability distributions for investment returns and household energy spending facilitates effective risk analysis and options appraisal for investors and policy makers alike.

Finally, as well as SC, this paper enables the evaluation of the variability of import and export for grid-connected PV across varying scales, from district to national levels. Thus the model allows a more thorough techno-economic analysis of such technologies as energy storage, and of PV integration with the low-voltage network. With such improved predictions of the energy flux at the point of connection, the impacts on the grid and the requirement of mitigation strategies such as reinforcement, demand shifting and curtailment required under high PV penetration can be assessed. In terms of wider implications, such a detailed knowledge of the variability of import and export on smart-grids is useful for the development of new market paradigms such as local (peer-to-peer) energy trading and time-of-use tariffs.

9 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). We are grateful to Dr. Alastair Buckley for the sharing of a sample of anonymised monthly yield and system meta-data from the Microgen Database, University of Sheffield.

10 References

- 1 Rowley, P., Gottschalg, R., Betts, T., *et al.*: 'Multi-domain analysis of photovoltaic impacts via integrated spatial and probabilistic modelling', *IET Renew. Power Gener.*, 2015, **9**, (5), pp. 424–431
- 2 Cao, S., Sirén, K.: 'Impact of simulation time-resolution on the matching of PV production and household electric demand', *Appl. Energy*, 2014, **128**, pp. 192–208
- 3 Widén, J.: 'Improved photovoltaic self-consumption with appliance scheduling in 200 single-family buildings', *Appl. Energy*, 2014, **126**, pp. 199–212
- 4 Campoccia, A., Dusonchet, L., Telaretti, E., *et al.*: 'An analysis of feed-in tariffs for solar PV in six representative countries of the European Union', *Sol. Energy*, 2014, **107**, pp. 530–542
- 5 Steuerrechtliche Auswirkungen des Direktverbrauchs von Strom aus Photovoltaikanlagen (§ 33 Abs. 2 EEG 2009) (Ministry of Environment (Bundesumweltministerium), Germany, 2008)
- 6 Black, A.: 'Financial payback on California residential solar electric systems', *Sol. Energy*, 2004, **77**, (4), pp. 381–388
- 7 van der Welle, A., de Joode, J.: 'Regulatory road maps for the integration of intermittent electricity generation: methodology development and the case of The Netherlands', *Energy Policy*, 2011, **39**, (10), pp. 5829–5839
- 8 McKenna, E.: 'Demand response of domestic consumers to dynamic electricity pricing in low-carbon power systems'. PhD thesis, Loughborough University, 2013
- 9 Patsalides, M., Stavrou, A., Efthymiou, V., *et al.*: 'Towards the establishment of maximum PV generation limits due to power quality constraints', *Int. J. Electr. Power Energy Syst.*, 2012, **42**, (1), pp. 285–298
- 10 Parra, D., Walker, G., Gillott, M.: 'Modeling of PV generation, battery and hydrogen storage to investigate the benefits of energy storage for single dwelling', *Sust. Cities Soc.*, 2014, **10**, pp. 1–10
- 11 Walker, G.: 'Decentralised systems and fuel poverty: are there any links or risks?', *Energy Policy*, 2008, **36**, (12), pp. 4514–4517
- 12 Pillai, G., Putrus, G., Georgitsioti, T., *et al.*: 'Near-term economic benefits from grid-connected residential PV (photovoltaic) systems', *Energy*, 2014, **68**, pp. 832–843
- 13 Munzinger, M., Crick, F., Dayan, E.J., *et al.*: 'Domestic photovoltaic field trials'. Final Technical Report, URN: 06/2218, BRE and DTI, London, 2006
- 14 Decker, B., Jahn, U.: 'Performance of 170 grid connected PV plants in northern Germany – analysis of yields and optimization potentials', *Sol. Energy*, 1997, **59**, (4–6), pp. 127–133
- 15 Luthander, R., Widén, J., Nilsson, D., *et al.*: 'Photovoltaic self-consumption in buildings: a review', *Appl. Energy*, 2015, **142**, pp. 80–94
- 16 Colantuono, G., Everard, A., Hall, L., *et al.*: 'Monitoring nationwide ensembles of PV generators: limitations and uncertainties. The case of the UK', *Sol. Energy*, 2014, **108**, pp. 252–263
- 17 DECC: 'Solar photovoltaics deployment': Feed-in Tariff Statistics, June, 2015. Available at <https://www.gov.uk/government/statistics/solar-photovoltaics-deployment>
- 18 Department of Energy and Climate Change: 'National Energy Efficiency Data-Framework, 2014 [computer file]'. SN: 7518: 10.5255/UKDA-SN-7518-1, UK Data Archive [distributor], Colchester, Essex, July 2014
- 19 Yao, R., Steemers, K.: 'A method of formulating energy load profile for domestic buildings in the UK', *Energy Build.*, 2005, **37**, (6), pp. 663–671
- 20 Leicester, P.A., Goodier, C., Rowley, P.: 'Using a Bayesian network to evaluate the social, economic and environmental impacts of community deployed renewable energy'. Proc. of CISBAT, Clean Technology for Smart Cities and Buildings, Lausanne, 4–6 September 2013, p. 10. Available at <https://www.dspace.lboro.ac.uk/2134/14472>
- 21 Voss, K., Sartori, I., Napolitano, A., *et al.*: 'Load matching and grid interaction of net zero energy buildings' (EuroSun, Graz Austria, 29 September–October 1 2010)
- 22 Richardson, I., Thomson, M., Infield, D., *et al.*: 'Domestic electricity use: a high-resolution energy demand model', *Energy Build.*, 2010, **42**, (10), pp. 1878–1887
- 23 Wieder, S.: 'An introduction to solar energy for scientists and engineers' (Wiley, New York, 1982, 1st edn.)
- 24 Firth, S., Lomas, K., Wright, A., *et al.*: 'Identifying trends in the use of domestic appliances from household electricity consumption measurements', *Energy Build.*, 2008, **40**, (5), pp. 926–936
- 25 Urquhart, A., Thomson, M.: 'Impacts of demand data time resolution on estimates of distribution system energy losses', *IEEE Trans. Power Syst.*, 2015, **30**, (3), pp. 1483–1491
- 26 Bright, J., Smith, C., Taylor, P., *et al.*: 'Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data', *Sol. Energy*, 2015, **115**, pp. 229–242
- 27 Wright, A., Firth, S.: 'The nature of domestic electricity-loads and effects of time averaging on statistics and on-site generation calculations', *Appl. Energy*, 2007, **84**, (4), pp. 389–403
- 28 Richardson, I., Thomson, M., Infield, D., *et al.*: 'Domestic lighting: a high-resolution energy demand model', *Energy Build.*, 2009, **41**, (7), pp. 781–789
- 29 Richardson, I., Thomson, M.: 'Integrated simulation of photovoltaic micro-generation and domestic electricity demand: a one-minute resolution open-source model', *Proc. Inst. Mech. Eng. A, J. Power Energy*, 2012, **227**, (1), pp. 73–81
- 30 Betts, T.R., Gottschalg, R.: 'Irradiance data: solar irradiation data for Loughborough', 2007 (Centre for Renewable Energy Systems Technology (CREST), Loughborough University, UK)
- 31 Dusabe, D., Munda, J., Jimoh, A.: 'Modelling of cloudless solar radiation for PV module performance analysis', *J. Electr. Eng.*, 2009, **60**, (4), pp. 192–197
- 32 Pearl, J.: 'Reasoning under uncertainty', *Annu. Rev. Comput. Sci.*, 1990, **4**, (1), pp. 37–72
- 33 Pearl, J.: 'Bayesian networks. [Los Angeles, Calif.]' (UCLA, Computer Science Department, 1985)
- 34 Jensen, F., Nielsen, T.: 'Bayesian networks and decision graphs' (Springer, New York, 2007)
- 35 Mrad, A., Delcroix, V., Piechowiak, S., *et al.*: 'An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence', *Appl. Intell.*, 2015, **43**, (4), pp. 802–824
- 36 Delcroix, V., Sedki, K., Lepoutre, F.: 'A Bayesian network for recurrent multi-criteria and multi-attribute decision problems: choosing a manual wheelchair', *Expert Syst. Appl.*, 2013, **40**, (7), pp. 2541–2551
- 37 Fenton, N., Neil, M.: 'Risk assessment and decision analysis with Bayesian networks' (Boca Raton, Taylor & Francis, 2012)
- 38 Weidl, G., Madsen, A., Israelson, S.: 'Applications of object-oriented Bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes', *Comput. Chem. Eng.*, 2005, **29**, (9), pp. 1996–2009
- 39 Leicester, P.: 'The development of Object Oriented Bayesian Networks to evaluate the social, economic and environmental impacts of solar PV'. PhD thesis, Loughborough University, 2015
- 40 Kesh, S.: 'Evaluating the quality of entity relationship models', *Inf. Softw. Technol.*, 1995, **37**, (12), pp. 681–689
- 41 Koller, D., Pfeffer, A.: 'Object-oriented Bayesian networks'. 13th Annual Conf. on Uncertainty in Artificial Intelligence, Providence, RI, 1997, pp. 302–313
- 42 UK Buildings.: The Geoinformation Group, UK website. Available at <http://www.geoinformationgroup.co.uk/products/ukbuildings>, 2013
- 43 OS MasterMap: Topography Layer [GML geospatial data], Coverage: Various, Updated, Ordnance Survey, GB. Using: EDINA Digimap Ordnance Survey Service. Available at <http://www.edina.ac.uk/digimap>, downloaded: June 2014
- 44 Bluesky LTD: 2014. Available at web address <http://www.bluesky-world.com/>
- 45 Anderson, B.: 'Estimating small area income deprivation: an iterative proportional fitting approach' (Springer, 2013, vol. 6), pp. 49–68
- 46 Marcot, B.: 'Metrics for evaluating performance and uncertainty of Bayesian network models', *Ecol. Model.*, 2012, **230**, pp. 50–62