Citation


URL

https://oro.open.ac.uk/55835/

License

(CC-BY-NC-ND 4.0) Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

https://creativecommons.org/licenses/by-nc-nd/4.0/

Policy

This document has been downloaded from Open Research Online, The Open University’s repository of research publications. This version is being made available in accordance with Open Research Online policies available from Open Research Online (ORO) Policies

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding
Green neighbourhoods: the role of big data in low voltage networks’ planning

Danica Vukadinović Greetham and Laura Hattam

Abstract In this chapter, we aim to illustrate the benefits of data collection and analysis to the maintenance and planning of current and future low voltage networks. To start with, we present several recently developed methods based on graph theory and agent-based modelling for analysis and short- and long-term prediction of individual households electric energy demand. We show how maximum weighted perfect matching in bipartite graphs can be used for short-term forecasts, and then review recent research developments of this method that allow applications on very large datasets. Based on known individual profiles, we then review agent-based modelling techniques for uptake of low carbon technologies taking into account socio-demographic characteristics of local neighbourhoods. While these techniques are relatively easily scalable, measuring the uncertainty of their results is more challenging. We present confidence bounds that allow us to measure uncertainty of the uptake based on different scenarios. Finally, two case-studies are reported, describing applications of these techniques to energy modelling on a real low-voltage network in Bracknell, UK. These studies show how applying agent-based modelling to large collected datasets can create added value through more efficient energy usage. Big data analytics of supply and demand can contribute to a better use of renewable sources resulting in more reliable, cheaper energy and cut our carbon emissions at the same time.

Danica Vukadinović Greetham
Centre for the Mathematics of Human Behaviour, Department of Mathematics and Statistics, University of Reading, Reading RG6 6AX, UK e-mail: d.v.greetham@reading.ac.uk

Laura Hattam
Centre for the Mathematics of Human Behaviour, Department of Mathematics and Statistics, University of Reading, Reading RG6 6AX, UK e-mail: l.hattam@reading.ac.uk
1 Introduction

Our planet is heavily influenced by how electric energy is generated. Electricity comes mostly from fossil fuels, nuclear fuels and renewable resources. Both globally and in the UK, the energy mosaic is complex, there are many important factors that will reshape the production and use of electric energy in near future.

Among them are: the mass roll-out of energy smart meters expected in 2020 in the UK [19]; bigger percentages of renewable generation in national energy supply, especially solar and wind [37]; predicted electrification of transport and heating through electric vehicles and heat pumps [41]; more affordable energy storage [42], [13] and many others. The listed factors also mutually influence each other. For example, the combination of solar panels and energy storage [32] might make the uptake of both happen sooner than previously expected. All this results in a complex system difficult to model and predict.

In the last two decades, half-hourly or higher resolution measurements of individual households and commercial properties demand became available through different pilot projects [31], [7]. This resulted in more understanding in individual household energy demand and how different new technologies impact this demand. There is still a lot to be done in understanding energy use on individual level and creating accurate forecasts that can be used in so called smart grid.

This includes distributed generation and so called prosumage where the end-users, traditionally only consumers, are now also involved in production and storage of energy. Some other aspects of smart grid are adaptive control of devices, automatic demand response, optimal use of secondary generating sources and optimised charging and discharging of storage. For all of those aspects, efficient solutions cannot be obtained without big data collection and analysis.

Another issue that we encounter and where big data might help is that the energy objectives on the national level, where one is looking at the aggregated demand, and on the local level, where one is concerned with low voltage networks infrastructure, might be conflicting. For example, regionally or nationally, it might make sense to rectify the usual late afternoon demand peak by incentivising customers to charge their electric vehicles (EVs) during night. On the other hand, uncontrolled applications of time of use tariffs could be extremely detrimental to low voltage networks that were designed with behavioural diversity taken into account. In some cases, where there is a fast uptake of EVs concentrated in few neighbourhoods, local networks would not be able to cope with sudden night peaks without costly reinforcements. In these circumstances high resolution data can help by enabling modelling and constrained optimisation on a local level that can be scaled up to assess how the decisions made locally would impact the aggregated demand. Also, data analysis can help to react quickly in the cases where proposed policies would create inefficiencies or damage some of the stake-holders as in the example above.
2 Short-term forecasting of individual electric demand

Short-term forecasting of energy demand is usually taken to be forecasting one day up to one-two weeks in future[21]. While distribution network/system operators were doing aggregated demand forecasts for more than 60 years, the individual level forecasting is newer and came about when the individual data in higher resolution first became available.

2.1 Methodology

As we mentioned before, load forecasting was traditionally done at the medium voltage (MV) to high voltage (HV) networks level. At that level of aggregation, the demand is more smooth and regular [45].

There is a plethora of methods for short-term forecast of aggregated demands, mostly coming from artificial intelligence [48](such as artificial neural networks, fuzzy logic, genetic algorithms, expert systems and similar) and statistics [46] (linear regression, exponential smoothing, stochastic time series, autoregressive integrated moving average models (ARIMA) , etc.). More recently, probabilistic methods allowing for ensemble forecasts are being developed [21].

On the other hand, methods for short term individual demand forecast are somewhat scarcer. Historically, only quarterly meter readings were available, and distribution network operators were more concerned about commercial (large) customers. This started to change when the half-hourly or higher resolution data became available and when new technologies such as EVs and photovoltaics became a cause of concern for future in distribution networks, because of increased demand but also because of issues with voltage, thermal constraints, network balance etc. [8, 39, 9, 11]. At the LV network level (from an individual household to a feeder or substation), the demand is less regular and normally has frequent but irregular peaks [3]. Therefore new methods need to be developed that work well for this kind of inputs.

2.2 Recognised Issues

Several issues are recognised in the literature when predicting the individual level demand. Firstly, errors are normally much larger than with aggregated demands. In [9] it is claimed that aggregated forecast errors are in brackets of 1-2% when Mean Absolute Percentage Error (MAPE) is considered, while at the individual level errors raise to 20% and higher. Secondly, in [39] it has been noted that while most forecasts concentrate on the mean load, they actually have significant errors predicting peaks. Predicting peaks accurately, on the other hand, is of utmost importance for distribution network (system) operators for all kind of applications: demand side
response, efficient use of battery storage etc. Thirdly, it was noticed in [12] that on individual level, demand is influenced by many different types of behaviour resulting in frequent but irregular peaks, and therefore volatile and noisy data.

2.3 Forecasts quality - measuring errors

Above mentioned issues combined bring us to the problem of measuring quality of different individual short-term forecasts. If we denote with $y_t$ an observed demand at time $t$ and $f_t$ is a forecast of $y_t$ then the forecast error is commonly defined as $e_t = y_t - f_t$ [23]. Mean square error $MSE = \text{mean}(e_t^2)$ and its variants, root mean square error, mean absolute error etc, are all examples of scaled errors. That means that their scale depends on the scale of data [23]. The percentage error is given by $p_t = \frac{100e_t}{y_t}$ and it is used to define scale independent errors, such as mean absolute percentage error $\text{MAPE} = \text{mean}(|p_t|)$, and other variants. It was noted in [39] that predicting peaks is very challenging. For that reason in operational settings, variants of $MSE$ are sometimes replaced with $p-$norms, thus taking higher powers of $p$ (e.g. $p = 4$ instead of $p = 2$) in order to highlight peaks amplitudes. So, errors are calculated as $\sqrt[p]{\sum t e_t^p}$, for some $p \in \mathbb{N}$.

2.4 Adjusted errors

For noisy and volatile data, a forecast that predicts peaks accurately in terms of size and amplitude, but displaced in time, will be doubly penalised once for predicted and once for missed peak. Thus peaky forecasts can be easily outperformed by smooth forecasts, if using point-wise error measures.

In [12] the authors suggested an adjusted error measure that allows slight displacements in time. This displacement in time can be reduced to an optimisation problem in bipartite graphs, namely the maximum weight perfect matching in bipartite graphs [36] that can be solved in polynomial time. A graph $G = (V,E)$ is bipartite if its vertices can be split into two classes, so that all edges are between different classes. In this case observations $y_t$ and forecasts $f_t$ make two natural bipartite classes. Errors between observations and forecasts are used as weights on the edges between the two classes. Instead of just looking at errors $e_t = y_t - f_t$ (i.e. solely considering the edges between $y_t$ and $f_t$), also taken into account are differences between

$$y_t - f_{t-1}; y_t - f_{t+1}; y_t - f_{t-2}; y_t - f_{t+2}; \ldots; y_t - f_{t-w}; y_t - f_{t+w},$$

for some time-window $w$. These differences are added as weights and some very large number is assigned as the weight of all the other possible edges between two classes, in order to stop permutations of points far away in time. Now, we want
to find the perfect matching that minimises the sum of all weights, therefore allowing possibility of slightly early or late forecasted peaks to be matched to the observations without the double penalty. Using errors as weights we can solve this as the minimum weighted perfect matching in polynomial time [29], but also allow localised permutations in forecasts without large computational cost as in [5]. The Hungarian algorithm with a time complexity of $O(n(m + n\log n))$ for graphs with $n$ nodes and $m$ edges ([29, 47]) when $n$ is quite large, for example when forecasting in very high resolution, can be replaced by much faster suboptimal alternatives that use parallelisation. The auction algorithm [35] is able to find matchings comparable with ones produced by Hungarian algorithm, but is simpler to implement and much faster [20] for large $n$.

2.5 Short-term forecasts based on adjusted errors

The idea of an adjusted error resulted in two new forecasts that allow for the ‘peakiness’. This can be useful for electricity network distribution operators for different demand side response and storage dynamic control applications.

A new forecast, ‘adjusted average’, attempting to preserve ‘peakiness’ by allowing slight shifts in time in the historical data is presented in [12]. Given $N$ historical profiles (for example $N$ half-hourly demands during previous $k$ Thursdays) the aim is to predict demand for coming Thursday. The initial baseline $F_{1}$ can be defined as a median of $N$ profiles (or similar). Then, the $k^{th}$ historical demand profile $G(k)$ is matched through constrained permutations with the current baseline so that nearby peaks (e.g. an hour and a half in past and future) match up and in this way a new matched profile $\hat{G}_{k}$ is obtained. The new baseline is then calculated as

$$F_{k+1} = \frac{1}{k+1}(\hat{G}_{k} + kF_{k}).$$

(1)

This process is repeated $k$ times and the final forecast is

$$F_{N} = \frac{1}{N+1}(\sum_{k=1}^{N} \hat{G}_{k} + F_{1}).$$

(2)

Another forecast, ‘permutation merge’ is presented in [5]. Starting from the same premise as ‘adjusted average’, instead of averaging baseline and a matched profile in each step to obtain the new forecast, here a profile is chosen instead that minimises the sum of the adjusted errors in each step. This results in a faster forecast when allowed permutations are close to the forecasted point, i.e. a permutation window is small$^{1}$.

$^{1}$ For example, for half-hourly data if predicted peaks are early/late up to one hour and a half that means that a permutation window is 3.
On Fig 1 adjusted average and permutation merge forecast were compared (the benchmark was mean forecast that, based on \( k \) historical profiles for a day in question, e.g. Tuesday, forecasts each half hour of that day as a mean of \( k \) previous Tuesdays).

![Fig. 1 A comparison of adjusted average, permutation merge and mean forecast. From [5].](image)

The data was Irish Smart meter data [7]. The adjusted error measure (with \( w = 3 \) and \( p = 4 \), see 2.2) was used to calculate the errors. The PM forecast improved as more historical profiles were added, and was also much faster than adjusted average on this dataset for all \( w < 6 \).

### 2.6 Applications

Accurate short-term forecasts of individual demand are important for distribution network and systems operators (DNO, DSO) for efficient planning and managing of low voltage networks.

Several important applications arise through demand side response (DSR), a set of measures that helps DNOs to manage load and voltage profiles of the network. Under this regime, consumers are incentivised to lower or shift their electricity use at peak times. For example, Time-of-Use tariffs can be implemented in order to reduce aggregated demand peaks (such as well-known early evening peak for instance). Individual forecasts then allow network modellers to predict consumers reactions on tariffs[1] and to avoid undesired effects such as move of the excessive demand overnight or a loss of natural diversity of demand that could potentially exacerbate peaks instead of mitigating them. Another obvious application is in control of small energy storage (at household or residential level).
Set point control is the simplest and most common way of controlling battery storage, but its simplicity is paid by inefficiency as it rarely succeeds in reducing the peak demand. Efficiency can be improved significantly through accurate household-level forecasts and dynamic control. This would help to optimally control charging and discharging of batteries[34, 27].

3 Long-term forecasting: modelling uptake of low carbon technologies

As more data is getting available on individual households’ level energy consumption and low carbon technologies, such as electric vehicles, photovoltaics and heat pumps, we are facing new challenges and opportunities.

The individual and combined impact of different low carbon technologies on electric networks presents us with numerous challenges for accurate mathematical modelling and forecasting. Again, several inter-dependent factors come into play: tariffs, socio-demographics, occupancy, life styles, etc. (see [30], [26], [49], [22]). The time of use will differ for different technologies - while photovoltaics are physically limited to a certain period of day, the use of an electric car will be determined by social and environmental patterns. Considering the availability of electric storage and vehicle to grid technologies, we might become more flexible. However, to capture this behaviour, quite complex approaches to modelling, forecasting and optimisation problems are needed.

Uptake of low carbon technologies can be modelled as innovation diffusion. One of most influential early studies on innovation diffusion within a social system was provided by [33]. The author suggested that the spread of innovations over time resulted from individuals belonging to a social network communicating with one another.

This theory has been further developed from a top-down (macroscopic) and a bottom-up (microscopic) perspective, which can both be applied to predict the adoption of low-carbon technologies (LCTs) within a sample population. A macroscopic approach uses differential equations to determine the diffusion process, whereas, a microscopic viewpoint applies agent-based modelling techniques.

3.1 Macroscopic approaches - differential equation models

Initially, it was aggregate models that followed the workings of [33]. In the seminal work [2], Bass proposed the following nonlinear ordinary differential equation to govern the adoption process

\[
\frac{dF}{dt} = (p + qF)(1 - F),
\]
where $F$ is the adopter ratio, $T$ is time, $p$ is the innovation coefficient and $q$ is the imitation coefficient. The equation parameters $p$ and $q$ correspond to external and internal effects respectively such that innovators adopt independent of their social interactions, whilst imitators are influenced by these exchanges. Example external pressures include government incentives and advertising.

From (3), the solution is then derived as

$$F(T) = \frac{1 - \exp\left(-(p+q)T\right)}{(q/p)\exp\left(-(p+q)T\right) + 1},$$

which is plotted in Figure 2 for $p = 0.01, q = 1$ (black) and $p = 0.1, q = 1$ (red-dashed). These are S-shaped adopter curves, which describe the temporal dependence of the adoption behaviour and how this varies with $p$ and $q$.

Later, this model was extended in [18] by also including spatial variables with a diffusion coefficient. As a result, innovations propagate over time and space, like a wave. This comparison to wave motion was made by [28], where real-life data was used.

The equation that was developed by [18] can be used to forecast the uptake of LCTs across specific geographies. For instance, [38] applied this nonlinear system to estimate the adoption behaviour for hybrid cars within Japan. The innovation, imitation and diffusion coefficients were chosen by fitting the overall adopter curve to real information. [24] predicted the uptake of photovoltaics within southern Germany, although they focussed on the diffusive effects, excluding the innovation and imitation influences, which meant that instead, the heat equation was used. A semi-
A hypothetical case was presented, where the diffusive coefficient was modified over the domain of interest so to reflect a neighbourhood’s likelihood of adoption.

### 3.2 Agent-based modelling

Aggregate models are informative tools for studying the overall dynamics, however sometimes they fail to capture the more intricate interactions of individuals within a social system. Consequently, agent-based models (ABMs) have been also developed to determine innovation diffusion. These can characterise an agent’s decision process whilst obtaining a macroscopic understanding by aggregating the agents’ behaviours. An early ABM was proposed by [6], where individuals were assigned a probability of adoption that corresponded to their awareness of the innovation price and performance. Starting at a micro level, they then examined the subsequent dynamics from an aggregate perspective, and therefore, conducting a bottom-up approach.

In [25] the authors concluded that by assuming agent $i$ has the probability of innovation uptake

$$P = \left( p + \frac{\sum_{i=1}^{M} x_i}{M} q \right) (1 - x_i),$$

where $x_i = 1$ if agent $i$ has adopted and $x_i = 0$ if not, and $M$ is the total number of agents, then the behaviour observed overall is consistent with the Bass model. More specifically, the aggregate adopter trajectories that result are of the same form as the S-curves derived by Bass and depicted in Figure 2. Moreover, they are also dependent upon the external and internal parameters, $p$ and $q$ respectively, which were first introduced by [2].

More recently, [10] outlined an agent-based technique for calculating an individual’s likelihood of technology adoption, where the overall behaviour was determined with multiple numerical simulations. This model incorporated the effects of advertising, social interactions and a resistance to change at a microscopic level. In addition, agents could reverse their decision of uptake and revert to using the original technology. This unique model feature led to interesting adoption dynamics, including cyclical uptake results.

### 3.3 Hybrid models

Recent developments in the field of innovation diffusion have combined the bottom-up and top-down approaches. These models are referred to as hybrid simulations. For instance, [44] forecasted the spread of technologies across the world by applying a hybrid model. More specifically, the theory of [2] determined the adoption timings for various groups of countries. Then, agent-based techniques were employed to
model the behaviour of individual countries, which were influenced by neighbouring countries and their own preferences.

Another example of a hybrid technique was proposed by [16]. This study focused upon a local electricity network located in the UK that consisted of multiple feeders, where households were linked to the network by their feeder. The adoption process at a macro-level was governed by the equation derived by [18]. Initially, the network was depicted with a finite element grid and then the finite element method was applied to model innovation diffusion with Haynes’ equation. This resulted in feeder uptake curves as a function of time being identified. By following these estimated feeder trends, the adoption behaviour at a household level was next simulated.

Hence, the macro and micro level methods are viable forecasting tools for the uptake of technologies across a sample population. More importantly, the merging of the two approaches provides many varied and complex solutions to the problem of innovation diffusion modelling. An overview and guide for this simulation technique is detailed by [43], where three different types of hybrid models are highlighted and various example studies are described.

4 Case studies

Now we present two case studies of above mentioned techniques used with big data obtained from smart meters.

4.1 Modelling uncertainty - uptake of low carbon technologies in Bracknell

As a part of the Thames Valley Vision project (www.thamesvalleyvision.co.uk), a realistic low voltage (LV) network was modelled using agent-based modelling. The LV network is situated in Bracknell, UK, comprising of around 600 substations. In this case study, we considered 1841 properties/44 feeders, where 7 were households with PVs installed and 71 were non-domestic properties. The household count connected to each feeder varied, where the minimum number of households along one feeder was 4 and the maximum was 114. Note that here each feeder corresponded to one neighbourhood and all households along a particular feeder were considered neighbours i.e. they were not necessarily adjacent properties. Data sets for 17 selected days (representatives of different seasons, weekdays, weekends and bank holidays) were created using metered data from this LV network. The data sets consisted of a combination of metered and predicted daily demand energy profiles (kWh) for every household, where a genetic algorithm was used to allocate monitored endpoints to unmonitored customers. These profiles had readings every half hour and therefore for each household a load profile with 48 data points was provided. These data sets were referred to as baseloads.
The combined scenario of EV and PV uptake was considered. In this model, an agent corresponded to a household. In particular, agents were presumed to be more likely to adopt if their neighbour had a LCT. To analyse the result from a macroscopic viewpoint, multiple runs were undertaken and then the distributions of LCTs were compared at a neighbourhood level. The following outlines the clustering algorithm applied to forecast LCT uptake: firstly, the percentage of households in the sample population that will adopt LCTs and the number of years it will take is established (here, the time horizon was 8 years to simulate the current distribution network operator’s regulatory planning period). An initial distribution of LCT seeds was performed. Council tax band (CTB) information and existing PV properties were used to inform the initial seed distribution. The assumption was that larger homes corresponded to higher CTBs and therefore, CTBs were used to identify neighbourhoods with a higher proportion of larger properties.

Then, during the remaining years, LCTs were assigned to households according to the fitness score - this assured that agents with more neighbours who adopted LCT in previous rounds will adopt LCT with higher probability than the agents with none or few neighbours with LCT. The number of LCT households (households that adopted a LCT) increased linearly every year until the specified amount was attained. This allowed to compute impact for different percentages of LCT uptake.

The details of the algorithm can be found in [15]. Once the household obtained LCT, an LCT profile was added (or subtracted in the case of PVs) to its baseload. LCT profile was chosen uniformly at random from the set of available profiles, respecting the seasonal rule: if the baseload applied was representative of spring, summer or winter then the LCT profile chosen also corresponded to spring, summer or winter respectively.

Multiple runs of the model are performed, in order to capture the model uncertainty. The aggregate feeder load resulting from each run is recorded so that eventually a distribution forms at each feeder site. Then, the 10% and 90% quantile are calculated. This gives confidence bounds [15], a lower and upper bound for the model response that indicate a minimal and maximal possible daily load. The quantile with the baseload subtracted represents the variation in LCT load at the feeder. Then, dividing the quantiles by the number of households along the feeder, we can compare all the feeders and their LCT loads.

On Fig 3 one can see results obtained for two feeders (feeder 39 with 82 properties, and among those 54% with CBT D or higher on the left and on the left and feeder 40 with 86 properties, and among those 0% with CBT D or higher. Although the number of properties on two feeders is similar and they might be close geographically, there is a notable difference in confidence bounds obtained, resulting in much higher troughs and peaks for feeder 39.

The feeder upper (EVs) and lower bounds (PVs) for the model response can be used as input to a network modelling environment. Consequently, potential voltage and thermal issues that result from this LCT uptake can be assessed. Furthermore, vulnerable network sites can be highlighted. This allows for priority rankings for future investments across LV networks, and represents significant contribution to LV networks planning. This framework also allows for quantifying demand uncertainty
of different uptake scenarios. Scenarios can be updated subsequently, when new data becomes available or a trajectory of a new technology uptake takes an unpredicted turn. In this case, data collection and analysis allow us to mitigate some of risks and uncertainties related to new technologies.

### 4.2 Global versus local: electric vehicles charging incentives

With the accelerated uptake of electric vehicles (EV), significant increase in load is expected in local networks that obtain many vehicles in a short time span.

At national level, electric energy providers would prefer those loads to miss existing afternoon/early evening peak. Therefore, some schemes were piloted were the owners of EVs are incentivised to charge over night. On the other hand, simultaneous charging over night can be detrimental to local networks, as they are designed expecting quieter periods over night.

We used the data from two trials [17], E-mini where the users were incentivised to charge over night, and My Electric Avenue, without such incentive, where the charging was naturally diverse. While this setting has some similarity with electric heating, (for example, in the UK, the existing Economy 7 is a restrictive tariff that encourages household electric storage heating to occur during the night) one must note that the current local networks were designed with Economy 7 in mind, while EV demand is being retrofitted.

Our simulations used the data from a real LV network of 98 feeders in Bracknell. Clustered uptake of EVs was simulated so that households with higher council tax bands had a proportionally higher initial probability to obtain an EV. Social influence was then modelled by a household having a higher probability of uptake in each round of simulation if more neighbours already adopted one. This created clus-
tered uptake through LV. EV load was then added to the households that adopted an EV, using random samples from two trials. Running many simulations allowed to quantify confidence bounds.

In Figure 4, we calculated a feeder index, by taking the maximum value of the 90% feeder quantile (over 500 simulations). The base-load is then subtracted and the result normalised dividing by the number of households along the feeder. In this way we get an index of EV load for each feeder. The blue and red trends relate to the Mini-E trial and the My Electric Avenue trial respectively. One can see that there is a significant increase for incentivised night charging.

Fig. 4 A comparison of the feeder index (for 50% EV uptake). Blue: Mini-E trial, red: My Electric Avenue trial. From [17].

The results[17] have shown that simple incentives that favour charging overnight in order to smooth early evening peak might potentially cause extensive problems for local networks, and that diversity needs to be reintroduced in order to prevent damage of existing local networks in so called green neighbourhoods where the uptake of LCTs is expected to be higher sooner. This is an example where a simple policy solution (incentivising overnight EV charging) that optimises one cost function (reducing evening peak) has detrimental effect on a connected optimisation problem - to maintain LV network at the minimal cost. This just reinforces the fact that a complex system is more than a sum of its components, and needs to be perceived and simulated in its entirety.
5 Policy implications

Energy availability and price plays extremely important role in daily life and subsequently in politics, as we are often reminded. There is an adage that no government can survive continuous power cut.

Big data can help energy policy makers to make more informed decisions, but it should be used in parallel with smaller in-depth studies. This is especially important because of ever-changing technology and human behaviour that interplay and create a complex system.

In [14], based on an in-depth small scale study, the use of smart meters as tools per se for demand reduction is dismissed. It was found that the variability of users demand behaviour is large, confirming the natural diversity observed in larger studies. They also noticed that seeing the patterns of own consumption actually might make them more entrenched (as opposed to reducing demand). While we fully agree that smart meters use cannot help with the reduction demand without additional interventions, we argue that it will be much easier to measure future interventions with smart meters and to spot if they are going in unwanted directions much quicker. Also, methodologies that work from bottom up such as agent-based modelling and uncertainty measures developed for the aggregated profiles will allow for better planning and design of new interventions for the reduction and/or shifting of demand.

The ways to increase efficiency of renewables are discussed in [4]. Three types of policies are discussed, noting that implementations differ across different states:

- net metering policies, where the production and consumption of energy in a household is based on a single meter, and utilities are required to buy back generated, not consumed energy; this encourages owners to generate renewable energy because it decreases financial barriers for distributed generation; again, in our opinion no proper assessments of costs and benefits can be done without smart meter data collection and analysis;
- interconnection standards for distributed generation - this is a more infrastructural point that is not directly connected to big data;
- dynamic pricing policies - here the disadvantage is that households are exposed to price variability, so it is more difficult to e.g. plan an annual budget, but the advantages are that through critical peak pricing and time-of-use tariffs, this can help with reducing total energy price by smoothing peaks and troughs in demand, which again relies critically on big data collection and analysis.

In addition the ownership of smart-metered data and of consumer owned renewables generation is just starting to be regulated, so new rules and regulations could help reduce uncertainties associated with smart-grid investment.

In [40] implications of smart meters installation were discussed and 67 potential benefits and obstacles across technical and socio-economic (vulnerability, poverty, social background, consumer resistance and ambivalence) dimensions are given. They conclude that while the centralised national roll-out through big players assumes a rational consumer with a complete info on single technology at its end, the
reality is a very complex system of stakeholders with different needs, and emotional
agent that progressively influences the future states of the system.

Both of these issues, the efficiency of renewables and the data ownership are
quite complex, and we believe that transparent methods of analysis and public do-
main data available creating repeatable case-studies can help shedding light to those
problems and will result in more efficient and fairer systems for all involved actors.

6 Conclusions

In this chapter, we hope that we made our case to show the benefits of data collection
and analysis to the maintenance and planning of current low voltage networks and to
describe potential future use that would benefit not only distribution system opera-
tors, but the whole ecosystem of stakeholders around energy distributed generation,
storage and consumption. Beside individual benefits for each end-user, through en-
ergy efficiency the greater good is expected also by cutting carbon emissions.

After considering several standard and recently developed methods for analysis
and short- and long-term prediction of individual households electric energy de-
mand we reviewed some applications of graph theory that can be used for mea-
suring quality of and producing better short-term forecasts. In particular maximum
weighted perfect matching in bipartite graphs is highlighted and recent implement-
tations of this method that allow applications on very large datasets are shown.

We then discussed different modelling techniques for uptake of low carbon tech-
nologies starting with equation-based, moving to agent-based modelling and finally
reviewing hybrid approaches. While equation-based models use big data for param-
eter calibration, agent based-models need to take into account socio-demographic
characteristics of local neighbourhoods and for this reason big data is even more
valuable for those. This includes, besides smart meter data, publicly available data
on postcode, incomes, maps, etc. that can significantly improve the accuracy of
models’ predictions. While some of these techniques are relatively easily scalable,
measuring the uncertainty of their results is more challenging. We have presented
confidence bounds that allow us to measure uncertainty of the uptake based on dif-
ferent scenarios.

Furthermore, two case-studies are reported, describing applications of these tech-
niques to energy modelling on a real low-voltage network in Bracknell, UK. These
studies show how applying agent-based modelling to large collected datasets can
create added value through more efficient energy usage. Big data analytics of sup-
ply and demand can contribute to a better use of renewable sources resulting in more
reliable, cheaper energy and cut our carbon emissions at the same time.

There is a lot of space for improvement on current methodologies. Current lim-
itations include data availability, missing data and data ownership issues which in-
fluence models’ calibration and validation. This is especially the case in markets
like UK, where several different (competing) entities might collect and manage data
(even just one type, for example smart meters) for households on the same street.
Also a complex system of stakeholders, who range from distributed systems operators through centralised and distributed energy suppliers to the end users makes it difficult to measure fairness and costs/benefits analysis for all of the actors.

In addition, some of the main challenges currently experienced (and expected at least in near future) will be data security, privacy and ownership. In [40] some of the main barriers for data collection are technical glitches of the measuring components (such as smart meters), incompatibility between suppliers (in UK market switching suppliers might mean changing a smart meter too) then security and privacy issues such as hacking and cyber-terrorism, the perception of utilities ‘spying’ on households etc. and vulnerability issues, such as consumer misunderstanding of smart meters’ use, financial burden assuming that the cost is passed to the consumers, health anxieties based on beliefs that WiFi communication of smart meters data back to centre produces non-ionizing electromagnetic fields bad for health and peripheralisation of groups not covered by smart meters roll-out (rural areas, basements and similar) and so on.

We believe that the system efficiency and ways to improvements of current methods will be obtained through a careful balance of all of these factors when collecting the data. More show-cases of benefits brought by analysing the data and transparency with types of data and methods used in analysis will evidently help with this balancing act.

Finally, we lay out some of the future trends. As the data from various sources becomes available, we expect more accurate hybrid and agent-based models to develop, allowing for better spatio-temporal predictions and therefore a better modelling and forecasting input into low voltage network planning and design.

For instance, combining data obtained from Google maps that shows the available surface and orientation for photovoltaics, availability of off-road parking for EV charging and similar with smart meter, high resolution meteorological, satellite and transport data could mitigate current uncertainties around LCT uptake and improve projections of off-grid PV supply locally and nationally.

To conclude, accurate and efficient individual or feeder level short and long term forecasts of electrical energy demand are needed in order to plan for greener future and big data collection and analysis is facilitating further developments in this area.

**Acknowledgements** This work was supported by Scottish and Southern Electricity Networks through the New Thames Valley Vision Project (SSET203 New Thames Valley Vision), and funded by the Low Carbon Network Fund established by Ofgem.

**References**


