Size matters:

How "small data" helps us understand the growing impact of electric vehicles on power networks.

A paper by TNEI

Released 27.02.2019

Big data and Electric Vehicles (EVs) are two hot topics in the energy sector in 2019. In this paper, we explore what one can tell us about the other. We suggest that, for the moment, EVs might actually be a "small data" problem and demonstrate how sophisticated statistical methods– using an approach known as Bayesian statistics – can help to handle such small data problems. Most importantly, we show how these methods can help network operators explicitly quantify the risk faced by their networks due to electric vehicles.

Data science and probabilistic modelling are growing specialisms for TNEI, complementing our existing expertise in network analysis and energy systems innovation. We think that now is the right time for the electricity industry to start thinking more deeply about risk and probability, to make sure the power system is ready for RIIO-2 and the electrification of heat and transport. Please get in touch for further information of if you have any questions or comments.

Not a big data problem - yet.

The increasing presence of electric vehicles (EVs) on our power networks – both in future projections and growth that has already occurred – is certainly a hot topic for the energy sector currently.

For example, at the end of 2018 the website *Current News*[1] invited key figures from the energy sector in the UK to make predictions about developments in 2019. Among a range of predictions, two topics stood out: EVs and big data. Headlines included: "EV uptake accelerates", "No brakes on electric car growth", "Unlocking the potential of data" and even "Data will become more valuable than energy".

"In the public domain there is currently surprisingly little data to help the energy industry and network companies understand electric vehicles, especially with respect to patterns of usage"

Such views are also held by IMB's global distributed energy resource leader, Clay Luthy, who is quoted by the technology website *ZDNet*[2] as stating that "Big data is going to be a big issue for electric cars" and that "One of the keys to electric vehicle success is ensuring the grid can support them".

The arrival of even a small number of EVs poses a challenge for networks, and low voltage (LV) networks in particular, as they might struggle to meet the power demands from more electric vehicle customers. It is therefore essential that distribution network operators (DNOs) are able to predict and detect where and when EVs are likely to appear, and in what quantities.

This knowledge must be coupled with an understanding of the demands arising from multiple EVs on a single LV network – particularly the magnitude and timing of their peak demands. This will be necessary for DNOs and many other industry players, in order to plan network upgrades and the procurement of flexibility services.

Doing this type of modelling and forecasting both accurately and precisely (see below) can require a substantial amount of data. Whilst it is likely that many avenues for generating such

1. https://www.current-news.co.uk/news/current-predicts-the-energy-transition-in-2019-part-one
2. www.zdnet.com/article/how-is-big-data-influencing-electric-vehicle-development/

data will exist in the future, particularly with the development of the internet of things and the roll-out of smart meters, some important system planning and regulatory decisions may need to be taken before a substantial amount of data exists.

In fact, in the public domain there is currently surprisingly little data to help the energy industry and network companies understand electric vehicles, especially with respect to patterns of usage. Various innovations projects, including Low Carbon London, Customer Led Network Revolution (*CLNR*)[3] and My Electric Avenue, have looked at patterns of residential off-street parking. However, in total this amounts to less than 500 records for slow chargers, covering around a year each. It's a similar story for public datasets that describe the behaviour of residential off-street fast chargers[4].

Industry is continuing to explore this, with two EV-focused Network Innovation Competition projects funded by Ofgem in 2018 to examine aspects like private hire vehicles and on-street charging. A Network Innovation Allowance project, *Electric Nation*, has trialled smart charging for almost 700 users and is currently assessing the results. The Government also publishes some information about vehicle registration and charge-point ownership at a reasonable level of geographic granularity. But it seems that, for the moment, there is still a fairly big gap in evidence about residential off-street consumption.

"Determining the impacts of EVs on distribution networks is, for the moment, best viewed as a "small data" problem"

In summary, at the moment there aren't that many datasets available to understand the network impacts of EVs, especially on LV networks. We therefore think that determining what these impacts are, is for the moment actually better treated as a "small data" problem. That is, there is relatively little data about EV consumption behaviour even at a national level, so when looking at very specific local geographic areas – including the level of individual specific LV networks (at most a few hundred customers) the data is particularly sparse. The challenge, then, is to draw as much insight from this data as possible using an appropriate model for EV demand.

3. www.networkrevolution.co.uk/project-library/insight-report-electric-vehicles/
4. If we've missed some, please let us know.

Electric vehicle demand is inherently uncertain

So, what should an EV demand model look like? Should it represent exactly what the (additional) demand at a certain time will be? TNEI's answer to that is 'no' – the demand is only partially predictable and should therefore be treated as fundamentally uncertain. We can, however, construct models which help to characterise the nature of this randomness for a given time of day using probability distributions[5]. These can then be used these to model things like the average values of demand, along with a range within which it will almost certainly fall.

In other words, we should aim for a model which is less precise, or reliable about its estimates but is much more accurate[6]. Figure 1 demonstrates why.

The nature of demand means we can't be in the top right quadrant, so we'd rather be in the bottom right than the top left quadrant. In our experience, this contrasts other demand modelling methods, which tend to make precise predictions that are not necessarily justified by their accuracy. In other words, in the context of Figure 1, we'd rather be in the bottom right quadrant than the top left, since the inherent difficulty in predicting demand means that we cannot be in the top right quadrant.

We believe that a good model should be transparent about the uncertainty in the demand level of a specific set of EVs at a given time.

In fact, a rather more radical proposal is that the EV demand model should also acknowledge that one set of EVs (e.g. a group of 10) often exhibit substantial differences in their demand patterns (i.e. probability distributions) compared to another set of 10 EVs.

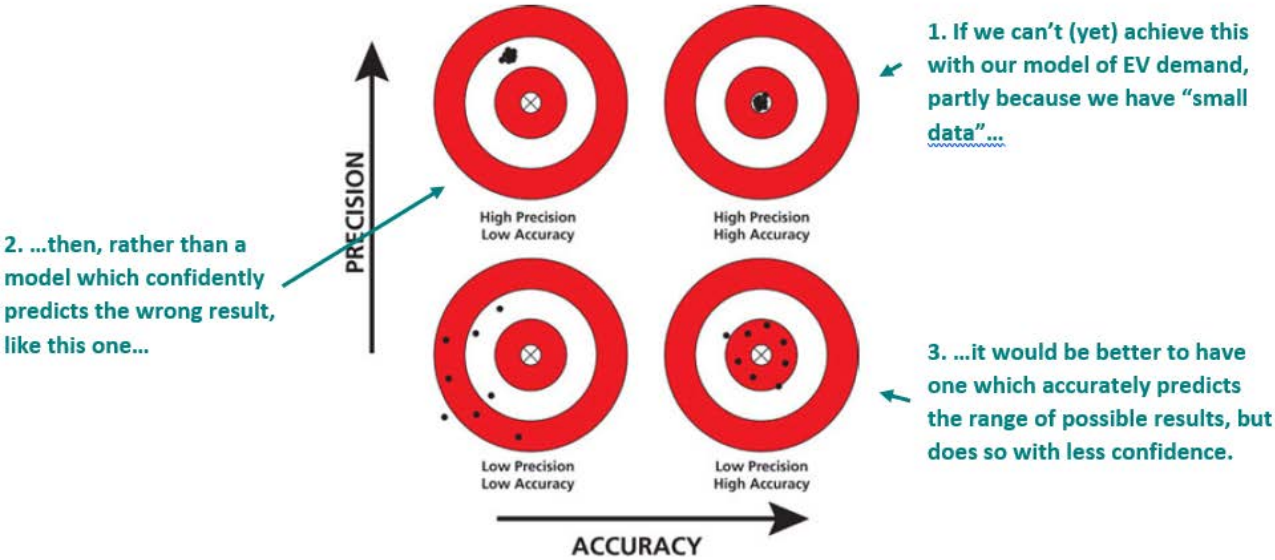

Figure 1: Accuracy vs precision, adapted from http://www.edvotek.com/Micropipet

5. Probability distributions have several equivalent definitions, but the most common is the following: for a random variable X, i.e. any quantity that cannot be precisely forecasted, the probability distribution – or more accurately the cumulative probability distribution, F_X (x), is the probability that the any single instance of X is less than or equal to x.
6. Although commonly treated as synonyms, in this context precision and accuracy are quite different. Precision (or reliability) refers to the level of certainty within the model. Accuracy measures whether the model's precision is justified by results.

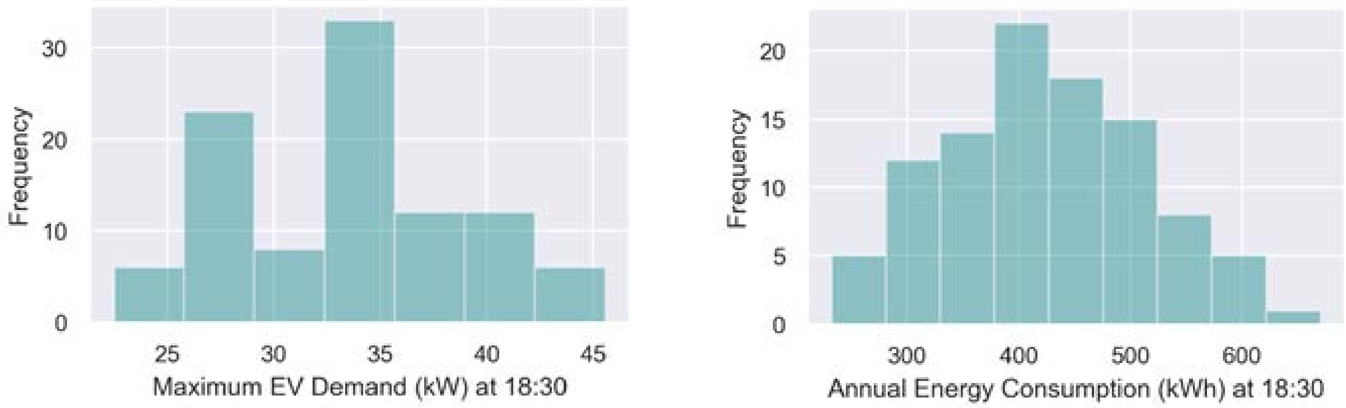

Figure 2: Variation in maximum EV demand and annual energy consumption

To demonstrate this principle, we examined data captured by the CLNR project, comprising: approximately 100 high-quality EV consumption records, for energy consumed in 10-minute intervals, for the period 25/06/2014 – 31/03/2015, where all chargers are 'slow', with a power rating of 3.8kW. We randomly selected 15 EVs from the full set and aggregated their consumption, then repeated with a further 99 random selections of 15 EVs. This gave us 100 series of data – one for every different random group of 15 EVs – each of which describes the total aggregate demand of that group of 15 electric vehicles.

> *"We believe that a good model should be transparent about the uncertainty in the demand level of a specific set of EVs at a given time."*

We then filtered out all data other than consumption between 6:30pm-6:40pm on weekdays, typically the current peak demand period for domestic customers. Examining the maximum aggregate demand observed across the trial period, we found that the average for 15 electric vehicles was 33.3 kW, but this varied between 22.5 kW and 45.6 kW for individual groups. Similarly, the 6:30pm consumption (across all weekdays) for an average group of 15 electric vehicles has a mean value (across all sampled groups) of 423.6 kWh, but this quantity varied between 232.3 kWh and 670.5 kWh for individual groups. Clearly, assuming that all groups of EVs of a given size will generate the same levels of additional demand on an LV network, and taking reinforcement decisions accordingly, is a sub-optimal strategy.

So, what methods can we use to address a problem like this? We think that any method should have the following characteristics:

- Demand is treated as uncertain, described by probability distributions.
- Customers are not assumed to all behave in the same way, and as a result each group of **n** EVs display distinct, but related patterns of demand (probability distributions).
- The method needs to work well with small data sets, and possibly integrate insights from a variety of different small data sets, as they become available.
- Due to the limited data, it should allow engineers and other specialists to incorporate their existing expert knowledge and intuition.
- It should be easy to use and understand, providing meaningful information to a network operator, rather than just a black box.

On the surface, this sounds like a very challenging set of criteria to meet. But thankfully, there is an approach to statistical modelling that ticks all of these boxes– the approach known as Bayesian statistics. We have in fact been exploring the application of Bayesian statistics to modelling demands in LV networks as part of an ongoing network innovation project with Northern Powergrid[7].

7. http://www.smarternetworks.org/project/npg_nia_020

A tried and tested approach: Bayesian statistics

The key things[8] to know about Bayesian statistics are:

1. Although it has applications as a practical tool, it requires the adoption of a slightly different philosophy about probabilities, where these are viewed as representing subjective beliefs.
2. Initial beliefs are formalised mathematically as 'prior' probability distributions, which represent the subjective belief before having access to data from a new trial or experiment. These priors are formed based on the existing knowledge of the modeller, or through consulting other engineering experts.
3. When new data becomes available, it should be used to update the "prior", according to an equation known as Bayes' Law, to form a 'posterior" probability distribution.
4. This process can be repeated indefinitely, with the 'old' posterior becoming the new prior -i.e. beliefs being updated and refined every time new data becomes available. Eventually, the initial prior belief has very little influence on the distribution.

We acknowledge that some readers may initially object to the notion that statistics should be viewed as representing subjective beliefs, and that initial beliefs should be combined with objective data. But Bayesian statistics is a tried and tested approach which is increasingly being used in other sectors, such as quantitative finance and pharmaceuticals.

Here's how the Bayesian approach could help us understand the impact of EVs on a LV network. Imagine a network planner working for a DNO, who's assessing the possible need to reinforce an LV network due to the recent addition of 15 EVs to the feeder's demand, where again the chargers are all 3.8 kW. Before these electric vehicles turned up, the headroom on this feeder was 14 kW, and the peak demand was estimated to occur at 6:30pm on a weekday evening. The question to be addressed by the planner is: what are the chances of these 15 electric vehicles exceeding the remaining capacity headroom and overloading this feeder?

"Bayesian statistics is a tried and tested approach which is increasingly being used in other sectors, such as quantitative finance and pharmaceuticals."

In our demonstration of the approach, we continue to use the CLNR dataset described above. We have found that we can describe these demand values very well using a family of probability distributions called Beta distributions. For Beta distributions, the modelled quantity (demand) can only take on values within some range, which in the case of our 15 chargers is 0 to 57 kW[9].

Figure 3 below shows a histogram of the total demand from a random group of 15 electric vehicles, overlaid with a fitted Beta distribution as a probability density function[10]. Next to this, we have shown the same Beta distribution in the form of a "survival function", where the vertical axis gives the probability that a single instance of the aggregated demand takes a value that is greater than or equal to the value on the horizontal axis. Also included in this plot is the 'descending cumulative histogram' of the data, i.e. the blocks of the histogram piled on top of each other cumulatively moving from right to left.

For the example group shown on the left, there is an 80% chance that the combined demand of these EVs will be equal to or higher

8. Rather than get into a comprehensive explanation of the Bayesian statistical approach within this article, we'd suggest the Wikipedia page as a good starting point for any readers who want a more thorough explanation en.wikipedia.org/wiki/Bayesian_inference
9. This aspect of the distribution is in contrast to e.g. Normal distributions, where the random variable (in this case, demand) can take arbitrarily large positive or negative values. A model that consistently predicted EV demands that exceeded the charge-point rating would not be as useful.
10. A probability density function, f_X (x) is defined as follows: the probability that a single instance of the random variable X takes a value in the range x to x+"d" x, where "d" x is extremely small, is given by f_X (x)·"d" x.

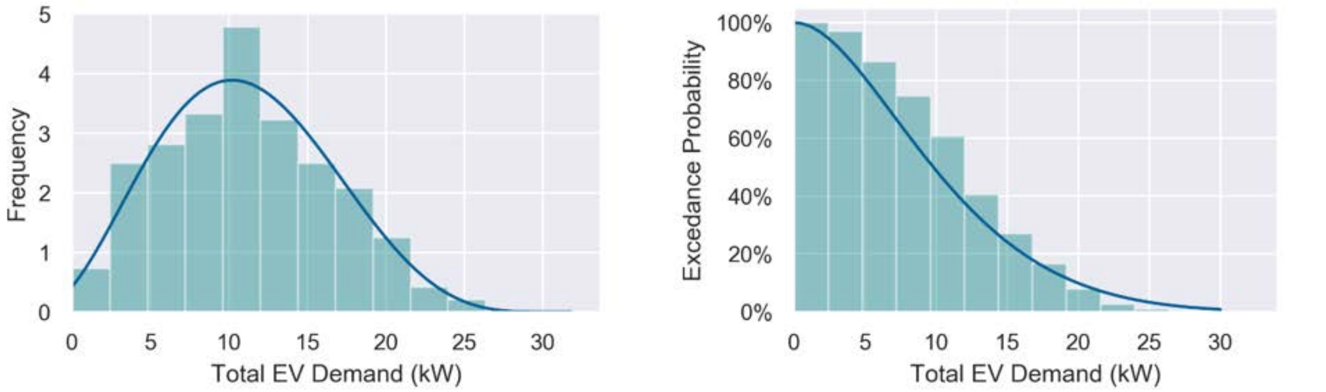

Figure 3: The curved line in the two plots show the fitted Beta distribution, first as a probability density function, then as a survival function (probabilities of demands being exceeded). The histograms show the analogous quantities for the raw CLNR data.

than 5 kW. In other words, in the long term, the demand from these 15 EVs will be greater than or equal to 5 kW on four weekdays out of five, between 6:30pm and 6:40pm.

> *"Different groups of EV owners, of the same size, behave quite distinctly."*

If we pick a different group of 15 electric vehicles (i.e. owned and used by a different group of customers), we get a different distribution, as shown in Figure 4. The data is a little messier for these EV owners – there are spikes at around 3.8 kW and 7.6 kW corresponding to 1 electric vehicle at full output, 2 electric vehicles at full output and so-on. However, as the exceedance probability curve shows, the Beta distribution still provides a reasonably good model.

We can also see that two different random groups of electric vehicles behave quite differently. For example, the first group will have a demand higher than 15 kW about 25% of the time, while the second group will only have demand higher than this about 10% of the time. This is a very significant difference, but our network planner has no way of knowing which group the 15 connected to their specific network are most similar to. The impact of the "true" group of EV owners on the network is very uncertain.

We believe that the optimal tool that we can provide to the network planner is a "hierarchical Bayesian model" – which we demonstrate here (without getting into too many technical details). Given a data set such as our CLNR trial, a model-fitting algorithm would randomly select a large number of groups of 15 electric vehicles (e.g. 100 or even up to 500), and fit a model which accounts for the difference between the groups, using the principles of Bayesian statistics.

The method requires as an input a formalised "prior belief" about electric vehicle demands – i.e. what

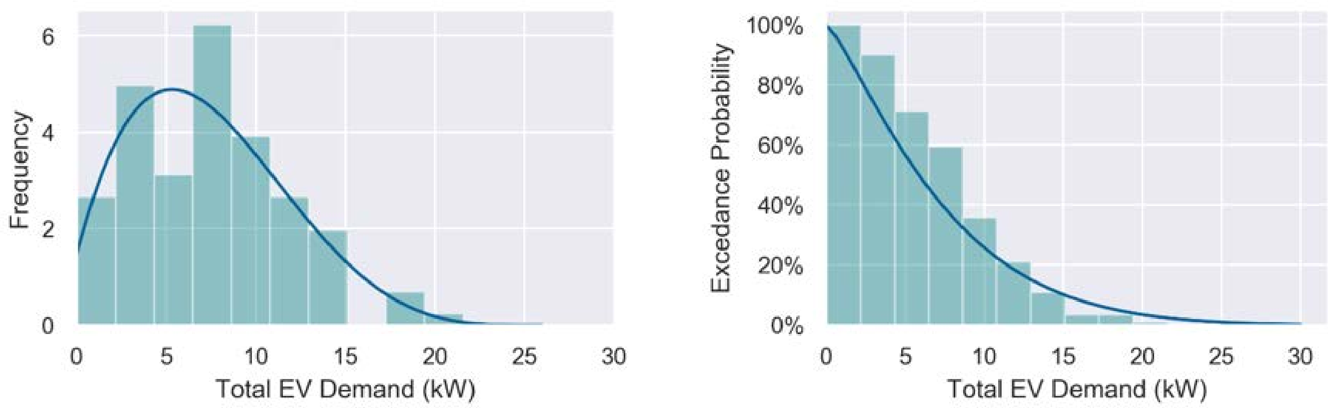

Figure 4: The curved line in the two plots show the fitted Beta distribution, first as a probability density function, then as a survival function, for a new group of 15 EVs. The histograms show the analogous quantities for the raw CLNR data

we expect the result to look like before we see any of the CLNR data, based on previous knowledge and experience. In our example, we adopt a prior belief that any instance of the aggregate EV demand could take any value between 0 kW and 57 kW with equal probability, i.e. we believe that the demand is very hard to predict (and as such this type of prior is often described as weak). Such a prior might be described as weak, as it does not represent a strong conviction that a specific level of demand will be observed. The model-fitting algorithm would pass the data for the 100 simulated instances of groups of 15 EVs through our production-line of Bayesian updating[11] to obtain a 'posterior' estimate, which accounts for how the data should rationally update our beliefs. The posterior distribution represents the Bayesian model's prediction about the probability distribution of a new instance of 15 EVs (i.e. the network being considered for reinforcement), including robustly calculated measures of uncertainty.

The posterior distribution for our CLNR data set is shown in Figure 5 below, in the form of an exceedance function. The solid line is the average exceedance probability for every level of electric vehicle demand. This is the weighted average over every possible type of customer group. The shaded area around it is the standard deviation of these probabilities, as a measure of the model's uncertainty. This uncertainty exists primarily because different EV owners use their chargers in different ways, and the planner does not know whether the specific set of 15 on the feeder will exhibit e.g. higher or lower demands than the average group of this size[12]. Understanding this range gives the planner a more thorough understanding of the risk to the network associated with the presence of these 15 EVs.

> *"Bayesian models allow very clear analysis of uncertain risks."*

Since 14kW (the amount of remaining headroom) is the only demand level of interest in our example, we can take a vertical line 'cut' through the shaded area in Figure 5 for a demand value of 14kW, and examine the probabilities in detail. This is presented in Figure 6, where the horizontal axis gives probabilities that a demand level of 14kW is exceeded on any given day (6:30pm-6:40pm), while the vertical axis presents the corresponding probability that a random instance of a group of 15 EVs will exceed 14kW with the given probability. The black lines show the average probabilities, while the red lines demonstrate the 5th and 95th percentiles, together forming a 90% prediction interval. This concept is, admittedly, quite abstract, as we're talking about probabilities of probabilities. But it is still

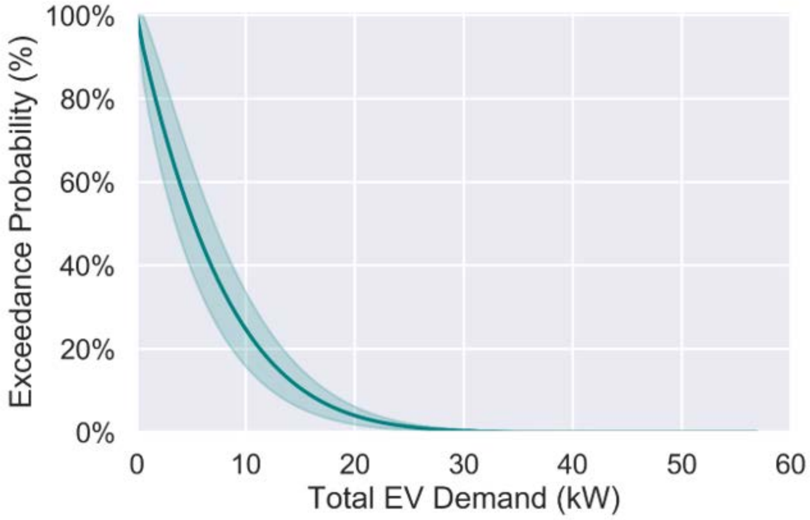

Figure 5: Range of exceedance probabilities for demands from 15 electric vehicles

11. We used a Python module called PYMC3 to set up a Hierarchical Bayesian model to infer the parameters for groups of different customers.
12. Although, if the planner had access to relevant data for that specific instance of 15 EV's, the algorithm could use it to produce a new posterior through Bayesian updating.

worthwhile, as it is a true reflection of what the data can actually tell us about EV demand from a statistical perspective.

The figure shows that our best estimate for a typical group of 15 EVs is that there is an 13% chance of their demand being equal to or higher than 14 kW on any given weekday. In other words, the circuit headroom will be exceeded once every ~8 years in the long term for an average group of EVs[13].

However, the planner might be more risk averse, and might want to see what will happen if the EV owners on their network are among those that have a pattern of greatest demand for EV charging at 6:30pm (among all patterns of demand across groups). More specifically, they might then be interested in the probability of exceeding a demand level of 14kW on any given day, for a group of customers whose particularly high-demand is in the top 5% of possible values of demand (known as the 95th percentile). (Or, perhaps this might be required as part of a network planning policy).

Figure 6 shows that the 95th percentile result is a 22% chance of exceeding 14 kW on any given weekday (or once every five years for breaching the headroom). The figure also shows the exceedance probability for the EVs whose demand patterns are in the lowest 5% (among all possible groups), which is around 5% (or once every 20 years for breaching the headroom). As seen previously, the expected value over all possible customer groups is 13%.

This highlights a brilliant feature of Bayesian models – they enable very straightforward analysis of uncertain risks. However, this might then require DNOs to more actively consider what level of risk is acceptable on their networks – currently, this is left to deterministic planning standards like P2 and ACE49.

After the introduction of 15 EVs, the average over all groups is that the circuit is overloaded every 1-in-8 years. But there is a 5% chance that an unknown group will make such onerous demands that the headroom is exceeded at least as often as once every 5 years. But, there is equally a 5% chance that they make such light use of chargers that the frequency of exceedance is less than once every 20 years. (More realistically, there will be uncertainty due to differences in the behaviour of groups of domestic customers of the same size, that is exacerbated by the uncertainty in EV charging behaviour). A Bayesian model can reflect this very considerable uncertainty, and we strongly believe that this knowledge is of considerable value to the DNO.

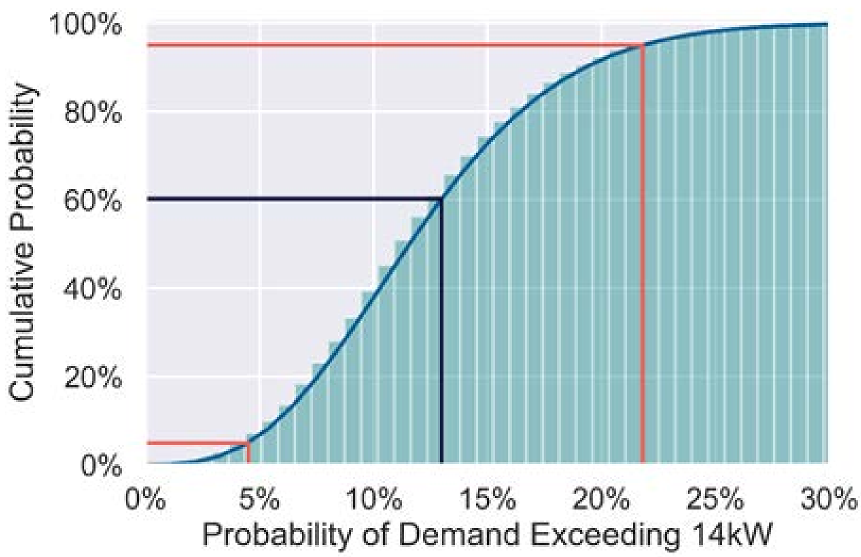

Figure 6: Probability of demand from 15 electric vehicles exceeding 14 kW

13. This relies on a lot of assumptions, which we've made for the purposes for this demonstration only, and which we'll discuss in more detail at the end of paper. It assumes that (i) the headroom is known to be exactly 14 kW, (ii) that this headroom is never breached by the existing customer demand, (iii) that this condition definitely occurs at 18:30 on a weekday and (iv) that we can ignore other times of the year when the EV demand might be greater than 14 kW.

Existing beliefs matter for small datasets

What if we had different initial beliefs about EV demand patterns? For example, another network planner might use a version of the model with a different ‘prior’ belief, e.g. a strongly pessimistic view that 15 electric vehicles are certain to be charging at, or very close to, 57 kW at 6:30pm every day. The results presented in Figure 5 were re-calculated with the equivalent prior, with the results presented in Figure 7. The green line is the prior distribution (that the EVs are almost certain to be charging at 57 kW) and the blue line is the posterior distribution.

Clearly the results look quite different now. The estimated distribution sits around lower values of demand than the prior belief, which is very close to 57 kW. However, the data is insufficient to convince our pessimist (based on the robust mathematics of Bayesian inference) that demand can be as low as 10 kW to 20 kW, as the data on its own might suggest. Instead, they should now believe that demand is likely to be around 40 kW to 50 kW – lower, but not as low as the data would tell them if looked at on its own. In addition, because the data is so contradictory to their initial prior belief, the uncertainty has actually increased.

> *“The initial belief adopded by the network planner has a huge impact on the result - this perfectly demonstrates why modelling EVs on LV networks is currently a “small data” problem.”*

This perfectly demonstrates why modelling EVs on LV networks is currently a “small data” problem – the existing belief has a huge impact on the result. The data we have used for this demonstration (~200 points for each of our 100 groups of EVs) is insufficient to gain a complete understanding of the EV behaviour and to ‘overrule’ the prior beliefs. In a similar way, our network planners could incorporate data and beliefs from any number of sources, and still get a logically consistent understanding of the electric vehicles’ behaviour. If, at a later date, some new EV consumption data becomes available, we could run the model again – we’d probably end up with an answer much closer to Figure 5.

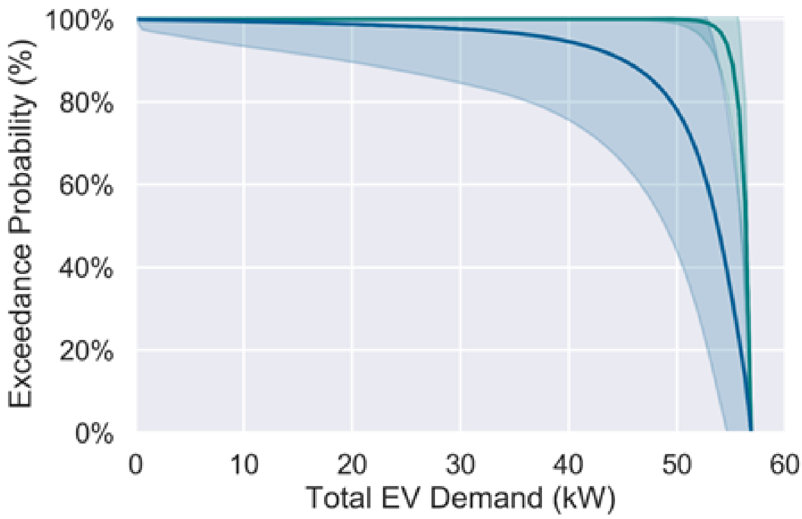

Figure 7: Range of exceedance probabilities for demands for 15 electric vehicles, with a different prior belief

Not so simple

In order to make this example work for a short article, we've had to cut lots of corners.

Firstly, we've talked abstractly about the "headroom" of the circuit, and calculated the probability of EV demand exceeding this headroom at 6:30pm on a weekday. In practice, the headroom on the circuit changes dynamically as the demand varies throughout a day, and over the course of the year. Our network planner is more likely to be interested in the risk of the circuit rating being exceeded at least once across the whole year, or once every ten years. To estimate that, we need to understand the combined risk associated with electric vehicles and the existing demand on the circuit – which means adopting a probabilistic model of both. This is exactly what we've been exploring with Northern Powergrid and we've found that, like EVs, no group of customers has the same domestic demand. Bayesian approaches allow us to explicitly quantify the uncertainty in our estimates of feeder overloads and voltage excursions, potentially helping planners make much more robust decisions when designing the networks.

Secondly, we've assumed the network planner knows with certainty that there are 15 electric vehicles on the circuit. In practice, they might not know anything so clearly. However, we could extend our Bayesian model to look at the possible range of numbers of electric vehicles on the circuit, based on, for example, data linking EV ownership to demographics. This would give us an estimate of the range of possible numbers of electric vehicles on each circuit, with an "average" expected value and some distribution around this.

Thirdly, we've used a single reasonably small dataset from the CLNR trials. This only accounts for a relatively small pool of customers (likely to be early adopters), a relatively small number of electric vehicle models, and a small set of 'external' conditions. Other customers using different types of EV could have very different patterns of demand. This could also change depending on the availability of on-street charging options, which is changing quite rapidly.

> *"In principle, all of these factors could be accounted for and built into a model which robustly deals with uncertainty, handles small datasets, and enables DNOs to understand and quantify the risks for their networks associated with transport electrification. "*

But, in principle, all of these factors could be accounted for and built into a model which robustly deals with uncertainty, handles small datasets, and enables DNOs to understand and quantify the risks for their networks associated with transport electrification. Eventually, when we have large volumes of smart meter (and other) data, we might be able to leave all of this to other "big data" and artificial intelligence methods, like neural networks. But until then, our recommendation is that Bayesian statistics could help network planners manage some of the uncertainty around EVs as the industry heads into the next distribution price control, and beyond.

Get in touch

Gordon McFadzean

Gordon has a broad background covering energy, engineering and economics, with extensive experience of techno-economic modelling and cost benefit analysis of power systems and innovative technologies. He has worked with network operators, regulators and governments to consider issues related to system planning, price controls, and network charges. Recently, he has focused on topics like low voltage networks, demand side response, electrification of heat and transport and smart meters.

Email: gordon.mcfadzean@tneigroup.com

Gruffudd Edwards

Gruffudd is a data scientist with considerable experience in mining, analysing, interpreting, predicting and visualising data. His academic background spans statistics, electrical engineering and physics, and his research has often bridged these disciplines, covering topics like generation capacity adequacy, the operation of energy storage, and the role of smart meters in low voltage planner. He has strong expertise in stochastic power system modelling and the application of operational research methods to power system problems.

Email: gruffudd.edwards@tneigroup.com

TNEI is an independent specialist energy consultancy providing technical, strategic, environmental and consenting advice to organisations operating within the conventional and renewable energy sectors. Our consultants have industry leading expertise in grid code compliance studies, noise assessment and analysis of innovative, smart grid technologies.

Our clients range from large utilities and network operators to regulators and community groups We are a specialist, independent company, and that's why we can offer a flexible, personal service and help our clients quickly and efficiently. And most importantly of all, we love to solve problems.

To find out more about how we can assist in this sector or to share feedback on this article, please contact us:

info@tneigroup.com
+44 (0)141 428 3180

Queens House
19 St. Vincent Place
Glasgow
G1 2DT

tnei

www.tneigroup.com
info@tneigroup.com