# NOVEL ANALYSIS TECHNIQUES FOR LV NETWORK PLANNING USING SMART METER DATA

Diptargha Chakravorty
TNEI Services – UK
diptargha.chakravorty@tneigroup.com

Charlotte Higgins
TNEI Services – UK
charlotte.higgins@tneigroup.com

Gruffudd Edwards
TNEI Services – UK
gruffudd.edwards@tneigroup.com

Francis Shillitoe
WSP - UK
francis.shillitoe@wsp.com

Alan Creighton
Northern Powergrid – UK
alan.creighton@northernpowergrid.com

Gordon McFadzean
TNEI Services – UK
gordon.mcfadzean@tneigroup.com

## ABSTRACT

*The present design and modelling tools for LV networks are much more simplistic than those used for HV & EHV network planning. This paper argues that while this simplicity was appropriate while LV networks were load-centric and passive, the emerging paradigm – involving new loads such as electric vehicles and heat pumps, distributed generation and flexibility services – means that a more robust and flexible approach is required. A novel approach is proposed that addresses current challenges by combining:* (1) *the automated importation of network models into a load flow software package, (2) a Bayesian statistical model of demand, and* (3) *a method for independently modelling the response of the network to sets of demand values at crucial nodes.*

## INTRODUCTION

Increasing electrification of the heat and transport sectors, uptake of low carbon technologies and increasing customer flexibility is leading to more dynamic and complex demands being imposed by customers on distribution networks across all voltage levels. The impacts of changes will typically be most significant at low voltage (LV), because the demands on a network are less smooth and predictable as the number of customers supplied decreases, and the presence of technology more clustered. Greater visibility and understanding of power flows at LV will therefore be necessary for the efficient design and operation of LV networks.

Existing modelling approaches and tools such as After Diversity Maximum Demand (ADMD) [1], and DEBUT [2] are not designed to provide such comprehensive visibility – rather adopting 'typical' end-user aggregate demands for groups of customers of different types. Real instances of a small customer group are unlikely to behave in these average ways, and their impact on a specific feeder may be quite different from the modelled average.

The objective of such approaches is typically to establish a single "worst case" demand, rather than considering the range of network loading conditions that may occur due to the fundamentally random, yet periodic nature of demand. This may well be detrimental to the continuing suitability of these tools for the LV networks of the near future.

In addition, these approaches typically avoid rigorous calculation of power flows within networks by equating line utilisations with downstream aggregate demand, and implementing simple voltage calculations. This may not be appropriate in the future, given the potential for thermal and voltage violations driven by the connection of new types of demand whose impact on the network is poorly understood. However, AC load flows are computationally expensive, so assessing a wide range of network conditions for a large number of LV feeders could be prohibitive. The situation is even more challenging since network operators rarely have detailed power systems models of their LV networks.

Fortunately, the increased adoption of smart meters is providing a significant opportunity to collate and analyse data on customer load characteristics, enabling more differentiated characterisation of demand on specific networks.

We propose novel analysis techniques that can overcome the identified challenges by adopting a robust risk-based approach that utilises the new data sources, and includes efficient use of load-flow calculations.

This study has been carried out as part of the Network Innovation Allowance project "Smart Network Design Methodologies", sponsored by Northern Powergrid (NPg).

## OVERVIEW OF OUR APPROACH

Our approach comprises three novel analysis techniques:

1. Tools for automating the build of a power systems model from GIS records.
2. A Bayesian statistical approach to represent the demand patterns of specific small groups of customers, with explicit uncertainty modelling.
3. An approach that decouples the characterisation of the network response to demand from the representation of demand.

Table 1 demonstrates how these techniques address the challenges highlighted above, and Figure 1 shows how they integrate into a single end-to-end model.

Each of these techniques is discussed in more detail in the following sections of this paper.

*Table 1 – Challenges and novel solutions proposed*

| Modelling | Challenge | Novel analysis technique |
|---|---|---|
| **1 Network topology** | Network operators don't typically have detailed power systems models of their LV networks | **Automated LV model build** of a power systems model based on GIS data |
| **2 Customer demand** | Customer demand is subject to lots of uncertainty, even for a known mix of customer types. | A **Bayesian customer demand model**, to account for uncertainty in customer demand. |
| **3 Network condition** | AC load flow is computational expensive, making "Monte-Carlo" load flow unattractive. | **Decoupled network response** - analysis of power flow and voltage for varied demand conditions. |

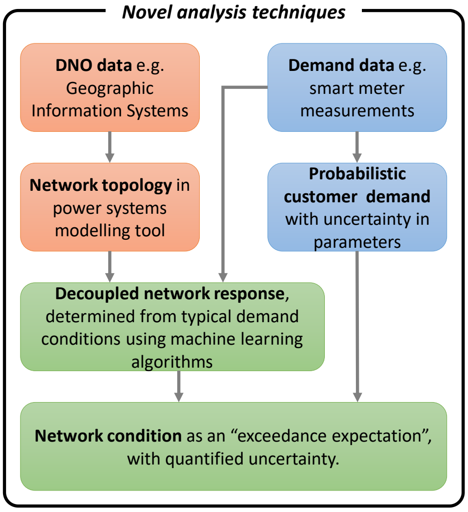


*Figure 1: Overview of the proposed Novel Analysis Techniques*

## AUTOMATED LV MODEL BUILD

Models of unbalanced LV networks have been built in the IPSA power systems analysis software using Python scripts, which involved the conversion of network data from NPg's GIS system. The objective is to accurately represent the LV network in the power-flow analysis software, using an automated process. These tools allow NPg to rapidly create reliable LV network models in the future. The LV model build is a two-step process. The first step involves the export of the network connectivity and asset data from the GIS database (Oracle) using Structured Query Language (SQL) into flat CSV (comma separated values) files. The second step, the IPSA model build, is performed by a suite of Python scripts, resulting in a topological and electrical representation of the network. Balanced or unbalanced representations of LV network can be created with these scripts, with the latter possibly involving arbitrary phase allocation for some single-phase customers, where the true phase is not known.

Once the network is built, scripts assign the phase of the customers, based on DNO data, and note the presence of any embedded generators. In the future, other information that the DNO has in its records could be utilised, such as the presence of electric vehicle charging points.

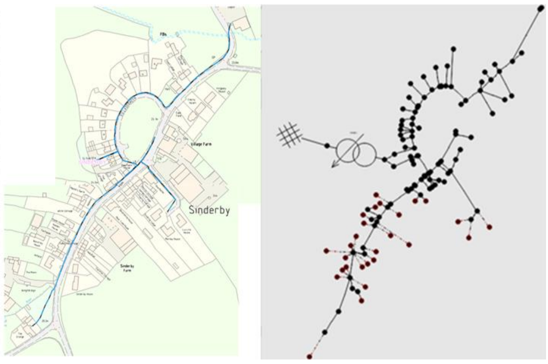


*Figure 2: Example of an automated IPSA power-flow model generated from NPg's GIS database*

Three NPg LV networks have been developed for testing using these tools: a rural network (Sinderby) with 66 customers, and two urban networks (Cranwood and Crandyke) each with around 630 customers.

## BAYESIAN CUSTOMER DEMAND MODEL

### <u>Current Approach</u>

There are two approaches which are commonly used to model the aggregate demand for power from a group of customers on a distribution network.

The first is to adopt a relatively simple probabilistic model, which will, by necessity, require some strong assumptions to be made. For example, as network operators might have access to annual, but not half-hourly resolution energy (kWh) metering data from customers, they may use a model which makes strong assumptions about the links between the customers' peak demand and their annual energy consumption. For simplicity, such models often "pre-select" an acceptable level of risk, such as designing for the level of demand expected to be observed once every ten years. This approach is used frequently in GB, with a model based on the ACE49 standard [2] forming the basis of the DEBUT LV modelling tool.

The second common approach is to model how the ADMD per customer decreases with increasing customer numbers, typically using a simple linear or power law relationship. ADMDs are typically determined from direct (but infrequent) analysis of demand data and therefore do not rely on strong assumptions. However, since they do not involve a statistical model, it is impossible to translate an ADMD into a risk-metric. Even if the underlying conditions driving customer demands were to remain unchanged over many years, the maximum observed over any e.g. two-year period will always be unique, reflecting the fundamentally random nature of such metrics.

A typical feature of both approaches is that they assume that all groups of a given size and mix of customer types have identical demand characteristics, even if they account for differences between types. This is contrary to evidence [4] particularly where the demand of a small number of customers is to be assessed. Therefore, even if a model is constructed that perfectly captures all of the salient probabilistic characteristics of an *average* specified group of customers of a certain type, it might lead to inappropriate decisions being taken for a *specific* network with both over- and under-engineering possible.

Another limitation of these common approaches is that they do not easily allow for information available to DNOs about customers on a specific feeder to be incorporated into demand models. This limits planners' benefits from the potentially substantial insights contained within smart meter data and new network monitoring data.

In order to overcome these issues, our proposed model uses a Bayesian statistical approach to represent customer demands.

## Bayesian Approach

The Bayesian approach to statistics describes nothing other than people's state of knowledge about the phenomenon or quantity of interest. The main consequence of this is that everything which is uncertain is treated as a random variable. This means that random quantities such as customer demand are often modelled as having probability distributions belonging to established parametric 'families', and characterised by a set of parameters. However, the distribution parameters are themselves modelled as random variables, characterised by their own set of *hyper-parameters*.

As a result, it is natural to explicitly quantify the uncertainty within a model – such as the uncertainty associated with modelling the demand of a *specific* group of customers rather than a *typical* group, or the uncertainties associated with the demand from electric vehicles or heat pumps.

Bayesian statistics is unique in treating probability distributions as subjective, so that the statistical modeller's prior beliefs constitute a valid and necessary part of their model, even when no specific data is available. As a result, the approach is well suited to problems where there is limited data, or data from multiple sources.

The statistical modeller's prior beliefs are translated into mathematical form through the construction of prior probability distributions. For example, if the modeller believes that the quantity of interest is likely to lie within a certain range, but knows nothing else, the equivalent prior distribution has a constant probability density within that range, and zero density outside of the range.

There is then a well-defined updating process (based on Bayes' rule) which dictates how any new data that becomes available should be used to update the prior distribution, to obtain what is known as a posterior distribution. If additional information becomes available at a later time, the posterior distribution becomes the new prior, which is updated by the new sample. This process is clearly ideal for DNOs in the emerging new paradigm described above, particularly for incorporating new data from smart meters into the models of demand.

A convenient way of representing a Bayesian parametric distribution with uncertain parameters is as a set of distributions (each with fixed parameters), where the parameter values are samples from their respective distributions. Figure 3 shows a stylised example of this, where a sample of 100 parameter sets have been drawn, and the corresponding distributions are shown in the form of exceedance curves. For each level of demand on the horizontal axis (or whatever the quantity may be), the exceedance curves give the probability that the observed demand value for a specific 30-minute period turns out to be greater than that value.

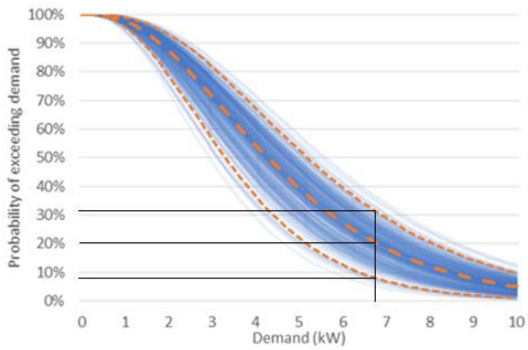


*Figure 3: Example of uncertain Bayesian distributions*

In the current context of LV network demands, the figure can represent the typical planning assumptions where no specific data about demand patterns on particular network is available. Each curve represents a plausible probability distribution for the demand of the group of customers.

The central orange line shows the average probability associated with each level of demand, while the outer orange lines show the range of probabilities which include 95% of the plausible customer behaviours. In the illustrative example of the figure, the black lines help to demonstrate that the expected demand that is exceeded 20% of the time is 6.8kW, and 95% of samples have the probability of occurrence for this demand within the range 8% to 31%. One can also take a fixed level of probability and obtain a range of demand values.

In the long-run, as new data is incorporated within the model, the uncertainty around the parameters in Figure 3 will decrease, with the range of possible values "tightening" around the average. However, the process will not always be one-way, i.e. sudden changes in demand patterns following the uptake of new technologies, may temporarily increase uncertainty.

## Case Study

The Bayesian statistical approach was applied to the northern feeder of the Sinderby LV network, which contains 17 customers. This network was also used to validate the automatic network-building process, and is shown in Figure 2. For brevity, the presentation does not cover all aspects of the developed model.

The prior distribution we constructed was from a very large dataset of 30-minute resolution domestic customer demand, collected over a period of three summers and two winters as part of the NPg's Customer Led Network Revolution project (CLNR) [4]. The adopted model actually takes the form of a set of Bayesian probability distributions, accounting for the variation in demand patterns across both time-of-day and season of the year. With 48 half-hours per day and four seasons per year, the model consists of 192 probability distributions.

The Gamma and three-parameter-Weibull families of probability distributions were both found to be appropriate to model demands for each combination of time-of-day and season, but Gamma is better than Weibull for some times-of-day, and vice versa. It was found that these parametric families have the flexibility to account for many features of observed demand series. One such property is the "*skewness*" of the distribution, i.e. asymmetric shapes around the central cluster of frequently observed values.

The full implementation of a parametric Bayesian model using these families has not been completed at the time of writing. Instead, a demonstration of a set of exceedance functions such as those in Figure 3 was conducted, where each curve was obtained directly from the CLNR dataset by random sampling (i.e. with no parametric modelling).

Rather than reproduce a very similar plot to Figure 3, we present the approximated results for a fixed exceedance expectation: 0.5 events per year. This is equivalent to an expected frequency of exceedance of once every two years. Figure 4 shows the results of 100 different samples from the CLNR data as a smoothed histogram.

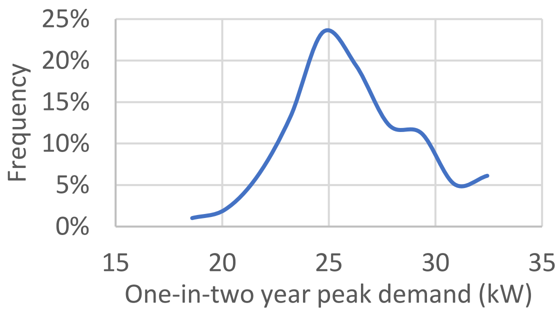


*Figure 4: Distribution of peak demands for Sinderby case study*

The horizontal axis values are demand levels that are expected only once every two years, while the vertical axis show the number of sampled customer groups (from a total of 100 groups) where the peak demand was close to that level. The values represented in the figure are approximate, as they are derived from finite samples composed of only two winters and three summers.

## DECOUPLED NETWORK RESPONSE

A stylised example of how decoupled network response modelling is combined with the Bayesian customer demand model is presented below, with the text supported by Figure 5.

**Step 1:** A Bayesian model of the probability distribution of the aggregate demand for a group of customers downstream from a network node is constructed, assuming a parametric family of distributions for demand.

**Step 2**: Probability distributions with known parameters are sampled from this Bayesian model, as in Figure 3. These probabilities can be converted to the expected number of exceedance events in e.g. a year, by multiplying with the number of time-steps in a year. This makes use of the linear property of expected values, i.e. the fact that the expected value of $n$ instances of *any* random variable is simply $n$-times the expected value of that variable. The relationship between demand and expected exceedances per year is referred to as the exceedance expectation curve. In our case study, we assumed a total of 148 distinct distributions (48 half-hours, 4 seasons), so each distinct probability was multiplied by either 90, 91 or 92 instances of that distribution to obtain the expected annual total.

An illustrative example of such an exceedance expectation curve is shown in Figure 5, labelled (2). For clarity, only one curve is shown, corresponding to one sample per parameter from the Bayesian model, although we really produced an 'envelope' of curves, such as in Figure 3.

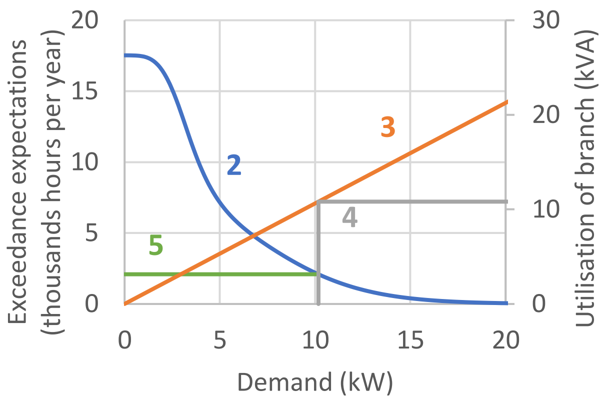


*Figure 5: Determining the exceedance expectation*

**Step 3:** Structured "trials" to explore the relationship between the LV network current and voltage values and aggregated demand values are conducted by passing thousands of systematically chosen values of demands to the power-flow model. From this, an equation can be found which describes the utilisation of a feeder section (or voltage at a particular location) for any value of downstream aggregated demand. It has been established, through experimentation, that the relationships of interest can often be represented reasonably well by a linear function of only one variable (e.g. aggregate downstream demand) as shown in Figure 6. For more complex networks, other machine learning algorithms could be used

to characterise nonlinear responses. The set of demands used in the trial does not need to reflect the probabilities of demands occurring, simply to represent plausible values.
For other networks, the state of some network components cannot be determined with sufficient accuracy by a single aggregated demand, e.g. where some customers lie at the end of a long section of cable. In such cases, the customer demands must be aggregated into two or more groups, and the network utilisation would be a function of these multiple variables. While our methodology has been extended to deal with such situations, a demonstration is beyond the scope of this paper.
Figure 6 shows the results of the power-flow calculations for the load on the Sinderby North feeder case study, with demands sampled from the CLNR dataset. A simple linear relationship dependent on the sum of the downstream load was tested and found to explain essentially all of the variation in the thermal utilisation of this section ($R^2 = 1$). That is, there is a very simple relationship between the feeder load and total downstream demand. This relationship is presented in Figure 6.

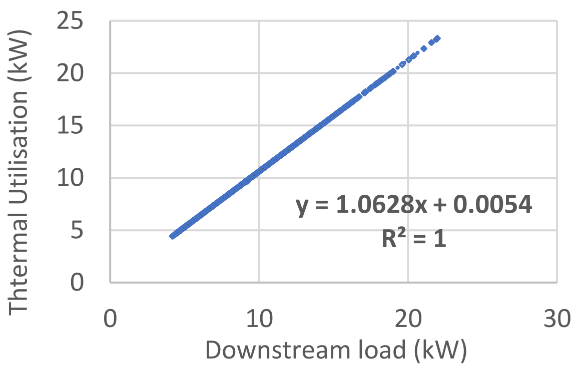


*Figure 6: Network response trials for Sinderby case study*

**Step 4:** The network planner can determine what level of aggregate customer downstream demand leads to a utilisation of 100% (or any other utilisation) from Step 3. Note again that only *plausible*, rather than actual recorded, demand levels are used here. For the Sinderby feeder the two values are equal, but this is not always the case. The process of converting a line utilisation to a level of aggregate demand is depicted in Figure 5 (labelled 4), where it can be seen that a demand of approximately 10 kW leads to a utilisation of 11 kVA.
**Step 5:** The exceedance expectation curve can be used to convert the demand level into an exceedance expectation value (and ultimately an exceedance expectation range).
In the stylised example of Figure 5, the demand level of 10 kW has an exceedance expectation of 2,000 periods per year. This is likely to be far in excess of a planning standard, and thus the line clearly requires upgrading.
As previously stated, Figure 5 shows a single probability distribution with known parameter values, i.e. curve 2 is a single line. However, the "Bayesian" customer demand model would involve re-running steps (2), (4) and (5), so that the initial uncertainty in demand is passed on correctly to the output metric of exceedance expectations.
For the Sinderby North feeder case study, using the results shown in figures Figure 4 andFigure 6 indicates that for an expected risk-level of one-in-two-winters, (i.e. a demand level that is only exceeded, on average, once in every two winters), the feeder load has an expected value of 27 kW, and it is 90% certain that the load with this risk-level will be between 22 kW and 34 kW. This is divided into a 5% chance of being below 22kW and a 5% chance of being above 34kW.
If this expected risk-level of one-in-two winters were adopted as policy for feeder ratings, our model states that there is a 95% probability that a rating of 34kW would be compliant with policy. The uncertainty in the demand reflects the fact that one group of 17 customers will display somewhat different patterns of consumption and hence peak demands from another group of the same size. A method for assessing the load on the feeder that does not consider this variability would probably conclude that a 34kW rating is sufficient, masking a small but significant chance that it is not.
Nonetheless, the probabilistic nature of the final output can be collapsed into a single deterministic relationship for the sake of clarity or simplicity. While some useful information would be lost, such an approach would be much more rigorous than collapsing the uncertainty at the demand modelling stage, as is current practice.

## CONCLUSION

The novel analysis techniques we have developed are formulated based on advanced statistical techniques that apply to both customer demand and corresponding network load. It incorporates the following features:

- Sophisticated data-driven statistical modelling, with a risk-based output.
- Accurately reflects the real and unavoidable uncertainties associated with specific small groups of customers served by LV networks.
- Dynamic and flexible, can be updated to incorporate new data, such as from smart meters, and learning about new technologies and their uptake.

This approach facilitates more efficient network planning in a changing energy system, maximising use of increasingly granular customer and network monitoring data, and adjusting to the increasing presence of low carbon technologies.

## REFERENCES

[1] Online:http://www.networkrevolution.co.uk/project-library/diversity-maximum-demand-admd-report/
[2] Electricity Networks Association, 1981, ACE Report No. 49, Report on Statistical Method for Calculating Demands and Voltage Regulation on LV Radial Distribution Systems
[3] Online:http://www.networkrevolution.co.uk/project-library/insight-report-domestic-baseline-profile/
[4] Online: http://www.networkrevolution.co.uk