Stochastic Effects of Customer Behaviour on Bottom Up Load Estimations

Jacco Heres[1], *Werner van Westering*[1,2], *Gerben van der Lubbe*[3], *Daan Janssen*[3]

1. Alliander N.V., the Netherlands, e-mail: jacco.heres@alliander.com
2. Delft University of Technology, the Netherlands, e-mail: w.h.p.vanwestering@tudelft.nl
3. Radboud University Nijmegen, the Netherlands, e-mail: gerbenvanderlubbe@gmail.com, daanjanssen@science.ru.nl

Abstract

Since the volatility of the power load is expected to keep increasing due to new energy technologies, modeling the stochastic properties of the power loads becomes increasingly important for DSOs. Due to limited measurements in these grids, often bottom up methods are used to create load estimations with which the peak load of the power customers is calculated. However, in average electricity consumption profiles, as used in most bottom up methods, the stochastic behavior of the customers energy consumption is neglected.

In this paper, the effect of neglecting the stochastic behaviour is investigated and is shown to be particularly strong in situations with a low number of consumers. To cope with this problem, several efficient methods to quantify the uncertainty and to determine peak loads have been evaluated. These methods were applied and validated on a data set with nearly thousand consumer measurement series, measured over 3.5 years on a 15 minute resolution. In low voltage networks with less than 10 power consumers, the conventional methods are shown to be at least a factor two too low. The suggested 'individual rescaling method' is accurate within 10 percent.

1 Introduction

For planning and operating purposes, DSOs aim to estimate the current and future load on their medium and low voltage assets and create accurate predictions of the load in the distant future. New energy technologies, such as electric vehicles and photovoltaic power generation are causing the loads in the electricity network to become more and more volatile, making these prediction more difficult, but also more important.

Traditionally, Strand-Axelson and Velander load calculations are used for modeling power loads. Recently, bottom up approaches are starting to be used to determine peak loads in the electricity grid as they can provide more thorough results[1][2][3]. In a bottom up modeling approach, the general goal is to predict the power consumption as accurately as possible to be able to calculate the impact of this consumption on the load of the network. This approach is chosen because of the expectation that the statistics of the residential and commercial loads will shift dramatically, which makes the traditional methods unusable [4]. The possibility to apply this methodology to large networks has arisen because of the increasing availability of customer data and computing power. Another advantage of the bottom up models is that they are able to produce full time series instead of a single value of the maximum load. These time series can be used in risk-models, condition-models or business cases for batteries and demand-shifting policies.

For large scale bottom up model such as [3], detailed time series of each user's consumption are used. Since Alliander DSO has over more three million customers, obtaining measured time series for all customers is currently infeasible due to economical, technical and privacy reasons. However, Dutch DSOs do have access to the annual electricity consumption (AEC) data of each customer in their grid. Furthermore normalised user consumption profiles are available for groups of customers from NEDU, based on measurement data.

By multiplying such a consumption profile with the annual electricity consumption, an estimate time series for the load is created. At Alliander [3] consumption profiles of NEDU [6] are used for residential customers and profiles for commercial customers are constructed from own data. As these consumption profiles are always constructed via averaging, they are well-suited to describe the behaviour of a large group of comparable customers, but fail to describe the relative large fluctuations of a single or few customers. When looking for extremes in the load, such as the maximum and minimum, this problem becomes enormous.

2 Scope and data

This paper describes the first major steps to quantify and deal with this problem of the stochastic effects of customer behaviour on bottom up load estimations. The two main question it tries to answer are:

1. How can the uncertainty of a load estimation via a bottom up approach at each point in time be quantified and calculated?
2. How can the expected maximum load over a year be calculated?

Profile	$>$	$\leq$	other	#	# used
E1A	-	3x25A	single tariff	205	101
E1B	-	3x25A	night tariff	222	40
E1C	-	3x25A	evening tariff	81	34
E2A	3x25A	3x80A	single tariff	71	32
E2B	3x25A	3x80A	double tariff	369	99
E3A	3x80A	100 kW	$UT \leq 2000$h	75	32
E3B	3x80A	100 kW	$UT \leq 3000$h	36	18
E3C	3x80A	100 kW	$UT > 3000$h	17	5

Table 1: The eight categories of NEDU profiles, including the the lower capacity limit ($>$) and upper capacity limit ($\leq$), other profile specifications and the number of measured time series that were available and used.

The focus in this paper is on consumption profiles in the residential area, but most of it is applicable to commercial profiles as well. To apply and validate the different proposed methods, measurements of different Dutch DSOs are used. These measurements were used to construct the consumption profiles of NEDU [6]. The only available information was in which of the eight categories the customer was placed. The definition and number of available measurements is shown in Table 1. The data was gathered over a period of 3.5 years. Due to data quality issues, only a smaller number of measurements could be used for all the methods.

3 The dependency between different customers

For both questions, it is important to know whether the energy consumption of each customer can me modeled as an independent stochastic variable. If this is the case, a large number of statistical methods could be used to compute quantities such as the variance of the load of a group customers, by using the variance of each of the individual customers. Also, a single distribution could be created for a heterogeneous group of customers to model the uncertainty in the load estimation.

To study the correlation, a virtual group has been created from all the measured customers of the K customers that are behind a single secondary substation H. We normalise each of the measurement series by dividing it by the total energy consumption per year. This is done because we know each customer's AEC and want to separate it from the unknown partition of the energy consumption over time. For each customer i we thus get an sample of the load,

$$\tilde{v}_i(t) = \mathrm{AEC}_i \cdot \tilde{w}_i(t). \tag{1}$$

Here w is a normalised measurement series in h^{-1}, AEC is the annual electricity consumption in kWh, and v is the load sample. To distinguish the quantities that are drawn from a sample and the known quantities, the first ones are marked by a tilde. The sample of the load of H becomes thus

$$\tilde{H}(t) = \sum_{i=1}^{K} \mathrm{AEC}_i \cdot \tilde{w}_i(t). \tag{2}$$

The estimation of the load by the bottom up method is

$$H_0(t) = \sum_{i=1}^{K} \mathrm{AEC}_i \cdot u_{g(i)}(t) = \sum_{i=1}^{K} \mathrm{AEC}_i \cdot u_{\mathrm{E1A}}(t), \tag{3}$$

Where u_{E1A} is the user consumption profiles of category E1A (see Table 1. Here we assume that the group $g(i)$ for all customers is E1A. This simplifies the notation, all the results can be generalised straightforward.

The variance of this estimation can be calculated by taking many samples $\tilde{H}$ and compute the expectation value $E((\tilde{H}(t) - H(t))^2)$. The expression in terms of the original measurement series becomes quite cumbersome, but in can be summarised by the expression

$$\begin{aligned} \mathrm{Var}(H(t)) = &\sum_{i=1}^{K} \mathrm{AEC}_i^2 \,\mathrm{Var}(p)(t) + \\ &\sum_{i=1}^{K}\sum_{j=1}^{K} \mathrm{AEC}_i \,\mathrm{AEC}_j \,\mathrm{Cov}(p,p')(t), \end{aligned} \tag{4}$$

where $\mathrm{Var}(p)(t)$ is defined as the variance of a single customer, and $\mathrm{Cov}(p,p')(t)$ as the covariance between two customers. It can be stated that for large K the covariance term will always dominate the variance term, as it grows by K^2 instead of K.

One is interested in the ratio of the two terms described in equation 4, which naturally depends greatly on number and combination of customers behind asset H. As an example the ratio is calculated for 100 customers of classes E1A, E1B and E1C. This resulted that on average over time the total variance consisted for 62% of the individual variances and 38% of the covariances. The conclusion is that the energy consumption of each customer can not be treated as an independent stochastic variable, with the possible exception of a very low number and a high diversity of customers. In the rest of this paper it is considered impossible to treat the energy consumptions as independent.

4 Quantifying the uncertainty

For network operating purposes, the most important reason for estimations of the load in the medium and low voltage network is to identify possible overloaded assets. To this end, the variance is not informative. It is more useful to know the probability that the load tress-passes a certain boundary. Formalised in a mathematical way, this becomes

$$O = P\left(\frac{W(t)}{V(t)} \leq c\right), \tag{5}$$

here $W(t)$ is the true load at time t, $V(t)$ is the load estimated by the bottom up method, and c is a parameter without a unit. It is assumed that O is not (strongly) dependent on the time or the AEC of the customer.

The probability density of $W(t)/V(t)$ for customers of the E1A category computed from 1000 simulations is shown in

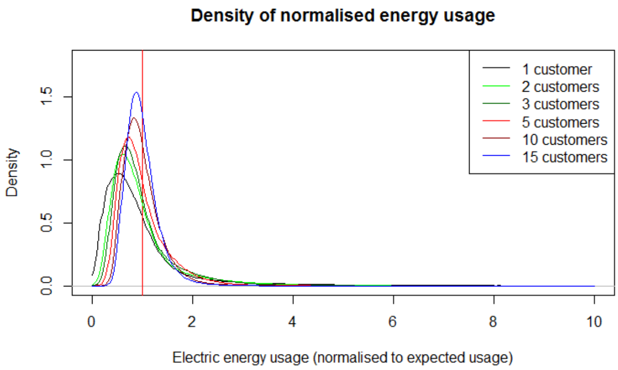

Fig. 1: Density of the quantity $W(t)/V(t)$ for a given number of customers from the category E1A. The red vertical line indicates $c = 1$, that is the density where the true load is equal to the expected load, $V(t) = W(t)$

Figure 1. One can see that the density is asymmetric around its expectation value ($c = 1$), and that the variance decreases when the number of customers increases. The tail of the density is quite long, as there are a few customers that use sometimes much more electricity than on average. It turned out to be not feasible to fit this density to a parametrised distribution with satisfying results. All of the distribution that were tried out failed to cover the behaviour of the density around $c = 1$ or the long tail. So even for a low number of customers, it was not feasible to model the stochastic effects of customer behaviour by distributions.

The conclusion is that the only way to deal with the problem is to to compute O for each combination of customers in the grid that exists in a bottom up model. From the densities in figure 1 one can deduce for instance that $O = 65\%$ for five E1A customers with arbitrary E1A. This means that the probability that the load at a certain time is higher than calculated in a bottom up model is 35%. This is of average and if load profile for a customer category shows a high variability this value can be misleading. Therefore the error of O is calculated by evaluating again 1000 simulations and compute for each simulation n the expression for the error

$$\tilde{e}_n = \max\left\{\left|P\left(\frac{V(t)}{W(t)} \le c\right) - P(\left(\frac{\widetilde{V}_n(t)}{\widetilde{W}_n(t)} \le c\right)\right| |c \ge 1\right\}. \tag{6}$$

The requirement $c \ge 1$ is set because the focus is on the chances of overloading an asset. An example of this is shown in Figure 2.

For all simulations the 50th, 75th, 90th and 95th percentiles of e_n are calculated. The results for E1A are shown in Table 2. This table can be used as follows: when investigating the load on a low voltage cable with 8 customers E1A, one can estimate a value for the load and a value for $O = P\left(\frac{V(t)}{W(t)} \le 1\right)$. Now in 95% of the cases the real value of O will be within $P\left(\frac{V(t)}{W(t)} \le 1\right) \pm 0.08$

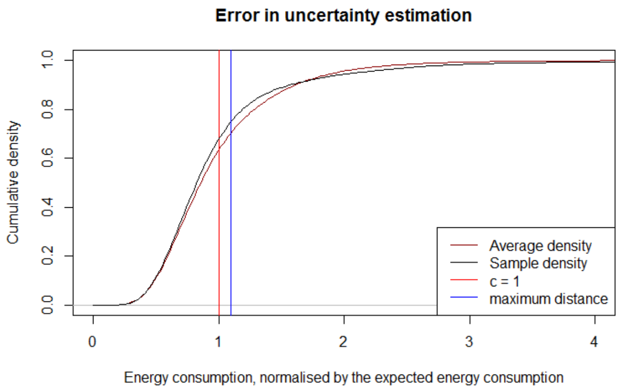

Fig. 2: Example of error calculation of $O = P\left(\frac{W(t)}{V(t)} \le c\right)$. The estimation of O for five customers of E1A is shown in dark red as the average cumulative density of the normalised load. A sample of a density for a group of five customers of E1A is shown in black. The maximum of the distance between the two graphs is at $c \approx 1.1$ and shown in blue, this distance, which is defined as the error of this sample is 0.045 .

#	50%	75%	90%	95%
1	0.054	0.078	0.121	0.155
2	0.046	0.063	0.086	0.112
4	0.039	0.058	0.075	0.089
8	0.035	0.050	0.069	0.080
16	0.035	0.049	0.066	0.074
32	0.036	0.048	0.062	0.074
64	0.033	0.045	0.057	0.067
128	0.032	0.043	0.055	0.061
256	0.030	0.041	0.050	0.057

Table 2: Results of the error calculation of O defined in eq. 5

5 Improving computation times

The drawback of the method described in section 4 is the computation time: for each asset or collection of assets within the bottom up model that have the same load, 2000 simulations have to be evaluated instead of 1 in the bottom up model. Even when using sophisticated parallelization and efficient programming language and code, this becomes problematic when one wants to study the load of for instance the more than 40.000 secondary substations in the Alliander Grid at each of the 35.040 quarters of an hour within a year. However, the advantage of the very quantitative method is that the effect of evaluating less simulations and time-steps can be studied in detail. For reducing the number of time-steps only prime numbers larger then 10 were used. For instance for the prime number 11 only the time-steps 10, 21, 32, 43, 54, ... were used in the estimation of O and the error in O. It is important to use only prime number to avoid that certain times within a day are always selected or passed over.

It turns out that when performing only 250 simulations and selecting from all time-steps T_i the one that satisfy

$$T^{'} = \{T_i | i \equiv 22 \bmod 23\}, \tag{7}$$

this barely worsens the stability and error of the method. This improves the performance of the method by 2 orders of magnitude. Even higher prime number can be chosen, but this is only at the cost of small loss of stability or increase of the errors.

6 Estimating the maximum load

Currently, the load profiles created by bottom up models are also used to estimate the maximum load at an asset. However, because the user consumption profiles are always constructed via averaging over a large group of users, spikes in individual customers energy usage are smeared out over time. This might be reasonable when the load at a certain time is constructed, for the estimation of the maximum load these spikes are very important. Without proper adjustment, the maximum load estimates can be far below the real maximum load when the number of customers is low. Four methods are performed and tested to find out which estimate is closest to the real maximum load.

6.1 Individual rescaling method

The measurement series are rescaled and used as a sample for the customers that are feeded via an asset, like in equation 1. From all the samples for the load for a single customer the total load of H is created just as in equation 2, the maximum of this load sample is calculated. This is done N times, now the estimation for the load is simply the average of the maximum of all the samples,

$$m_1 = \frac{1}{N} \sum_{n=1}^{N} \mathrm{max}_t \tilde{H}_{1,n}(t). \tag{8}$$

6.2 Overall rescaling method factor

The rescaling can also be done on the total load, the sampling is done on the original measurement series, without normalisation. The sample for the load of H is now

$$\tilde{H}_{2,n}(t) = \frac{\sum_{i=1}^{K} \mathrm{AEC}_i}{\sum_{j=1}^{K} \sum_t \tilde{s}_j(t)} \cdot \sum_{j=1}^{K} \tilde{s}_j(t), \tag{9}$$

where s_j is a measurement series of the same category as it's corresponding consumer i. Now the estimate for the maximum is again the average of the maximum from each sample,

$$m_2 = \frac{1}{N} \sum_{n=1}^{N} \mathrm{max}_t \tilde{H}_{2,n}(t). \tag{10}$$

6.3 Relative method

This method is inspired by the method for the quantification of the uncertainty in section 4. The following steps are followed to create this estimation:

1. Create the load profile in the usual way with the bottom up method $H_0(t)$ as in equation 3.
2. Select the 5% points in time with the highest loads.
3. Calculate via simulations the maximum of $c_n = \left(\frac{\widetilde{V}_n(t)}{\widetilde{W}_n(t)}\right)$ for each simulation.
4. The maximum estimation is created by the maximum of the bottom up model scaled by the mean of these c_n,

$$m_3 = \frac{1}{N} \sum_{n=1}^{N} c_n \mathrm{max}_t H_0(t) \tag{11}$$

6.4 Maximum of expectation value

Finally, for comparison, the maximum of the bottom up model is calculated directly. Because the bottom up model gives the expected load at each point in time, this method is method is called *Maximum of expectation value*.

$$m_4 = \mathrm{max}_t H_0(t) \tag{12}$$

6.5 Results

The methods described above were all tested at a sample secondary substation with 132 E1A customers and 12 E2B, so quite some households and a lower number of medium sized commercial customers. 500 simulations were performed for methods 1 to 3. The choice of the customers of the 'real' secondary substation was also done 100 to stabilise the results. The errors are calculated with the Root Mean Squared Error (RMSE) and the median of the absolute error. The results are shown in Table 3. The RMSE are all dominated by a few samples for which all estimates are far off. Therefore the choice for the best method is done on basis of the median of the absolute error. The rescaling method clearly performs the best.

Measure	m_1	m_2	m_3	m_4
RMSE	450	434	447	462
Median of absolute error	7.2	60.4	24.2	11.2

Table 3: Results of the evaluation of the methods for estimating the load in bottom up models. The Root Mean Squared Error is dominated by a few large errors, so the choice for the best method is based on the lowest Median of the absolute error. Clearly method 1, the individual rescaling method, is the best in this measure.

The original method (m_4) is also compared to the best method $m1$ for a varying number of customers. Reference situations were simulated with measurement series from year 1, the maximum was then estimated with series from year 2. The results for E1A customers is shown in Figure 3. It is clear that the proposed method achieves much better results, and that th original method is highly inaccurate for a low number of customers. The results for the median of the ratio and the results for E2B customers are very similar.

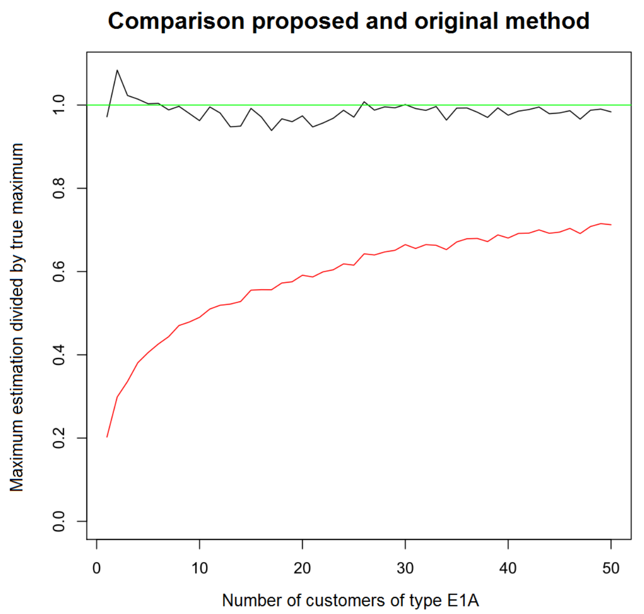

Fig. 3: Comparison of the proposed method for the maximum load estimation (6.1, black) with the original method (6.4 red). The ratio of the estimation and the true maximum is calculated, the average over 250 simulations is given.

It can be shown that the average deviation between the original maximum load estimation and the true maximum can be very well modeled by the Velander formula [7]. The original maximum can be modelled by $c_1 \cdot AEC_{\text{mean}}K$ and the true maximum by $c_1 \cdot AEC_{\text{mean}}K + k_2\sqrt{AEC_{\text{mean}}K}$ Therefore the proposed method combines the advantages of both Velander and bottom up models.

7 Possible applications

The quantification of the uncertainty and the improved maximum load estimation can be used directly to gain more insight and a estimation of the maximum load per year. There are however more ways to use these quantities, such as:

- For most DSOs, there are current and voltage measurements at the start of each feeder in the medium voltage grid. At this moment it is not possible to incorporate these measurements in the bottom up load for the MV grid, because it would imply scaling of the load of the secondary substations. With the quantification of the uncertainty at hand, this scaling becomes possible. The addition of real measurements could strongly improve the quality of the load estimation of the MV grid.
- If the scaling described above is performed over the whole grid, the scaling factors of all feeders can be studied. By looking at feeders with large scaling factors, defects in the grid topology registration could be detected and repaired.
- Sensors can be placed optimally trough the grid to reduce the uncertainty by the highest possible amount with a limited number of measurement devices.

8 Conclusions

This paper describes the first major steps to quantify and deal with the stochastic effects of customer behaviour on bottom up load estimations. The methods were developed and tested with real measurement data. It has been shown that the load of a single customer cannot be modeled as an independent stochastic variable because of the correlations with other customers. A quantification of the uncertainty has been defined as the probability that the load exceeds the expected load at least by a constant factor. The error of this uncertainty can also be calculated. An improved method for estimating the maximum load is proposed for bottom up models. It is shown that this method gives much better results than the original one. For less then 10 residential customers, the original method is at least a factor two too low, the proposed method is accurate within 10 percent.

References

[1] J. Dickert, P. Schegner, 2010. “Residential Load Models for Network Planning Purposes” *Modern Electric Power Systems 2010, Wroclaw, Poland*. Wroclaw, Poland

[2] I. Sartori, J. Ortiz, J. Salom, U. Dar, 2014. “Estimation of load and generation peaks in residential neighbourhoods with BIPV: Bottom-Up simulations vs. Velander” *World SB14*. Barcelona, Spain.

[3] W. van Westering et al., 2016 “Assessing and Mitigating the Impact of the Energy Demand in 2030 on the Dutch Regional Power Distribution Grid” *IC on Networking, Sensing, and Control*. Mexico City, Mexico.

[4] Klaassen et al., 2015, “Analyzing the impact of smart grids on load profiles using multiple regression”, *IEEE Power Engineering Conference (UPEC), 2015 50th International Universities*, United Kingdom

[5] E. Klaassen, J. Frunt, H. Slootweg, 2015. “Assessing the impact of distributed energy resources on lv grids using practical measurements” *23rd International Conference on Electricity Distribution (CIRED)* Lyon, France

[6] NEDU Consumer power load profiles [data file], http://nedu.nl/portfolio/verbruiksprofielen/, accessed: 05-01-2017

[7] S. Velander, 1952. “Methods of operational analysis applied to distribution of electric power”, *Teknisk Tidskrif* vol. 82, pp. 293299.