Improved Peak Load Estimation from Single and Multiple Consumer Categories

(article starts on next page)

CIRED 2020 Berlin Workshop (CIRED 2020)

22-23 September 2020

Theme 1: Integrating New Flexibility Tools and Principles for Planning

Improved peak load estimation from single and multiple consumer categories

ISSN 2515-0855
doi: 10.1049/oap-cired.2021.0300
www.ietdl.org

Kristoffer Fürst[1] ✉, Peiyuan Chen[1], Irene Yu-Hua Gu[1], Lang Tong[2]

[1]*Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden*
[2]*School of Electrical & Computer Engineering, Cornell University, Ithaca, USA*
✉ *E-mail: kristoffer.furst@chalmers.se*

Abstract: Velander's formula and coincidence factors have traditionally been used to estimate peak load for new connections in the distribution grid. By re-evaluating their underlying assumptions, this study proposes two improved models for aggregated peak load estimation (PLE). For single-category load aggregation, the proposed coincidence factor model, by incorporating an average correlation coefficient, improves the model fitting by 76–96% as compared to the standard Rusck model. For multiple-category load aggregation, the proposed joint Gaussian regression model reduces the PLE bias from 3–34% to 0.2–3% compared to the traditional approach.

1 Introduction

When planning for new connections of block loads such as residential or commercial developments, there is no available measurement data. Instead, the distribution system operators (DSOs) need to estimate the peak load by relying on accessible information such as the type and area of buildings, number of apartments, fuse sizes etc. Conventional methods used by DSOs for estimating peak load and sizing requirements are based on the typical load curves [1], coincidence factor [2], and the Velander–Strand–Axelsson's method [3]. With the increasing electrification process especially in the transport sector and other industries, it is crucial for DSOs to re-evaluate the validity of their methods for peak load estimation (PLE).

In Sweden, Velander's formula-based method is used for PLE. The main disadvantage of this method is that it assumes independently identically distributed (i.i.d.) loads with Gaussian distribution [3]. Such an assumption does not reflect the diverse mixture of loads under a low-voltage substation. Recent research effort on PLE for new connection abandons Velander's method completely and resorts to for example detailed load profiles as with a bottom-up approach [4, 5]. Such a model can serve multiple purposes, e.g. PLE and estimation of demand-side management potential. It has the potential of capturing individual load behaviour more accurately. However, they require detailed consumer data for model building, e.g. the number of inhabitants, end-user behaviour, weather-dependent energy consumption etc. On the other hand, the main benefits of Velander's formula are its simplicity with only two model parameters for a specific consumer category and the use of annual energy as its input.

This paper will present improved models for estimating aggregated loads from single and multiple consumer categories. The assumptions of Velander's formula and Rusck's coincidence factor are first re-evaluated. In the proposed model, we remove the independent Gaussian load assumption, assuming instead that the loads are jointly Gaussian with correlation within and between load categories. With the same methodology, it can also be applied with a typical load curve approach.

2 Review: Velander's formula for individual loads

Assume that we have N loads with a *Gaussian distribution* where P_i is a random variable representing the hourly power of the ith load with mean value $\bar{P}_i$ and standard deviation σ_i, $P_i \sim \mathcal{N}(\bar{P}_i, \sigma_i^2)$. The peak power $P_i^{(X)}$ of the Xth percentile is then

$$P_i^{(X)} = \bar{P}_i + K^{(X)}\sigma_i. \tag{1}$$

The coefficient $K^{(X)}$ gives the degree of probability of the peak power $P^{(X)}$, where $K^{(X)} = \sqrt{2}\mathrm{erf}^{-1}(2X-1)$ and erf^{-1} is the inverse of the Gaussian error function. Define two new parameters $E_i = T\bar{P}_i$ and $\vartheta_i^2 = \sigma_i^2/\bar{P}_i$, where T is the number of hours, and substitute them into (1). This gives

$$P_i^{(X)} = E_i T^{-1} + K^{(X)}\vartheta_i\sqrt{E_i T^{-1}}. \tag{2}$$

The parameter ϑ_i^2 is the so-called variance-to-mean ratio (VMR). With $T = 8760$ h, the parameter E_i can be estimated by the annual energy from the load data. In practice, Velander's formula for PLE is expressed as

$$P_i^{(X)} = \alpha E_i + \beta\sqrt{E_i}. \tag{3}$$

Comparing (2) and (3), $\alpha = T^{-1}$ and $\beta = K^{(X)}\vartheta T^{-1/2}$ if the load is Gaussian distributed. According to (3), β is a constant parameter, which implies a *constant VMR* for a given consumer category, i.e. $\vartheta^2 = \vartheta_1^2 = \cdots = \vartheta_N^2$. Note that other definitions of α and β exist where it is based on a stricter assumption of equally sized loads rather than constant VMR [6]. However, in practice, the load is not strictly Gaussian, and the parameters α and β can be taken directly from literature or estimated by regression from measurement data. With parameters from literature, there is a risk that the parameters are not valid for the considered geographical area, or the parameters can be outdated or not transparent [6, 7]. On the other hand, by fitting the parameters in (3) to the peak rather than to the mean and variance of the data according to (1), the PLE can be substantially improved for non-Gaussian loads as the Gaussian parameters are shifted to fit to the peak.

3 Proposed approaches

3.1 Single category load aggregation by the coincidence factor model

For the aggregation of N loads of the same category, different approaches can be utilised. In [1], Velander's formula for aggregated loads is presented. However, as this assumes i.i.d., it is not suitable for aggregation of correlated loads. Considering instead a joint Gaussian distribution, $P_\Sigma \sim \mathcal{N}(P, \boldsymbol{\Sigma})$, where $P = [P_1, P_2, \ldots, P_N]$ and $\boldsymbol{\Sigma}$ is the covariance matrix, the PLE is given as

$$P_\Sigma^{(X)} = \sum_{i=1}^{N} \bar{P}_i + K^{(X)} \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{i,j}}, \quad (4)$$

where $\sigma_{i,j}$ is the covariance between the ith and the jth load. Using (4) directly for the aggregated PLE of new loads, $N^2 + N$ parameter needs to be estimated. Furthermore, the data may not be of the same size or quantity as for the new loads, hence assumptions are required. To deal with these two issues, a common practice in the industry for single category load aggregation is to apply a coincidence factor [1]. The coincidence factor is defined as the ratio between the peak of the aggregated load and the sum of the individual peak loads

$$c(N) = \frac{P_\Sigma^{(X)}}{\sum_{i=1}^{N} P_i^{(X)}}. \quad (5)$$

This formulation does not require any assumptions. However, for new loads (5) can be inflexible, e.g. when N is larger than the data set used for parameter estimation. As described in the next section, a one-parameter estimation of the coincidence factor is also commonly used in practice, which can be used for $N \to \infty$.

3.1.1 Rusck's model: Under the assumption of i.i.d., the peak load in (4) is simplified as

$$P_\Sigma^{(X)} = \sum_{i=1}^{N} \bar{P}_i + K^{(X)} \sqrt{\sum_{i=1}^{N} \sigma_i^2}. \quad (6)$$

Consequently, with (6) and (1), the coincidence factor in (5) can be given as

$$c(N) = \frac{\sum_{i=1}^{N} \bar{P}_i + \sqrt{\sum_{i=1}^{N} \left(P_i^{(X)} - \bar{P}_i\right)^2}}{\sum_{i=1}^{N} P_i^{(X)}}. \quad (7)$$

With the assumptions of all loads of the same category having the same mean value $\bar{P}_0$ and peak power $P_0^{(X)}$, (7) can be simplified as

$$c(N) = c_\infty + \frac{1}{\sqrt{N}}\left(1 - c_\infty\right), \quad (8)$$

where $c_\infty = \bar{P}_0 / P_0^{(X)} = \lim_{N\to\infty} c(N)$. This estimate of the coincidence factor is commonly seen in literature and was derived by Rusck in 1956 [2]. Given (8), the peak power can then be estimated according to (5).

3.1.2 Proposed method – accounting for correlation: The loads have a relatively high correlation. To address this, we propose to improve (8) by considering the correlation between loads. Relaxing the assumptions of independent loads, the aggregated peak load in (6) is expressed as

$$P_\Sigma^{(X)} = \sum_{i=1}^{N} \bar{P}_i + K^{(X)} \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{i,j}\sigma_i\sigma_j}, \quad (9)$$

where $\rho_{i,j}\sigma_i\sigma_j = \sigma_{i,j}$ and $\rho_{i,i} = 1$. By substituting (9) and (1) into (5) and by assuming the loads have the same average power $\bar{P}_0$ and same peak power $P_0^{(X)}$ as in [2], with the additional assumption of equal correlation between loads of the same consumer category, $\rho = \rho_{1,2} = \cdots = \rho_{N,N-1}$, the coincidence factor can be expressed as

$$c(N) = \frac{\bar{P}_0}{P_0^{(X)}} + \left(1 - \frac{\bar{P}_0}{P_0^{(X)}}\right)\sqrt{\frac{1 + \rho(N-1)}{N}}. \quad (10)$$

The aggregated peak can then be estimated with (5). An alternative approach is to calculate the joint Gaussian directly, where the loads do not have to be of equal size. PLE using the joint Gaussian approach for single and multiple categories will be presented in (14).

3.2 Multiple category load aggregation by joint Gaussian regression

The following expressions are derived for the aggregation of N loads from $k \in \{A, B, \ldots, \Omega\}$ different categories.

3.2.1 Traditional approach: For the aggregation of loads of different categories using Velander's formula, it is suggested in [1] to either assume no or full correlation between categories. However, it still assumes i.i.d. for loads of the same category, thereby it is not suitable for aggregation of correlated loads. An alternative approach is, to sum up, the peaks from each category, assuming a full correlation between the aggregated peak of each category.

$$P_\Sigma^{(X)} = \sum_{k=1}^{\Omega} P_k^{(X)}, \quad (11)$$

where $P_k^{(X)}$ is the aggregated peak of the kth category estimated by an arbitrary method, e.g. Velander's formula with coincidence factor.

3.2.2 Proposed method – accounting for correlation: To take the correlation between loads from single and multiple consumer categories into account, we propose to improve Velander's formula for aggregated loads. For the aggregation of N loads from k different categories, the covariance matrix of the joint Gaussian in (4) can be expressed as blocks of covariance of loads from the same category and between loads of two different categories. The PLE of aggregation of loads from different categories can then be expressed with

$$P_\Sigma^{(X)} = \sum_{k=A}^{\Omega}\left(\sum_{i=1}^{N_k} \bar{P}_{ki}\right) + K^{(X)}\sqrt{Y} \quad (12)$$

where

$$Y = \sum_{k=A}^{\Omega}\left(\sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \rho_{ki,kj}\sigma_{ki}\sigma_{kj}\right) + \sum_{k=A}^{\Omega} \sum_{\substack{m=A \\ m\neq k}}^{\Omega} \left(\sum_{i=1}^{N_k}\sum_{j=1}^{N_m} \rho_{ki,mj}\sigma_{ki}\sigma_{mj}\right) \quad (13)$$

and $\rho_{ki,ki} = 1$. Assume that the correlation between loads of the *same* category is equal, $\rho_{ki,kj} = \rho_k$ for all $i \neq j$, and that the correlation between loads from two *different* categories is equal $\rho_{ki,mj} = \rho_{k,m}$ for all i and j. Furthermore, assume constant VMR and given the Velander coefficients and definition of annual energy in (3) and (12) can be estimated as

$$P_\Sigma^{(X)} = \sum_{k=A}^{\Omega}\left(\sum_{i=1}^{N_k} \alpha_k w_{ki} E\right) + \sqrt{E\gamma}, \quad (14)$$

where the annual energy, $E = \sum_{k=A}^{\Omega} \sum_{i=1}^{N_k} E_{ki}$, the weight factor $w_{ki} = E_{ki}/E$,

$$\gamma = \sum_{k=A}^{\Omega} \beta_k^2 \left(\sum_{i=1}^{N_k} w_{ki} + \sum_{i=1}^{N_k} \sum_{\substack{j=1 \\ j \neq i}}^{N_k} \rho_k \sqrt{w_{ki} w_{kj}} \right) + \sum_{k=A}^{\Omega} \sum_{\substack{m=A \\ m \neq k}}^{\Omega} \left(\beta_k \beta_m \rho_{k,m} \sum_{i=1}^{N_k} \sum_{j=1}^{N_m} \sqrt{w_{ki} w_{mj}} \right). \quad (15)$$

4 Measured data and parameter estimation

Three different heating systems/categories of 1–2 family houses were considered for this analysis, (i) direct electricity, (ii) ground source heat pump and (iii) district heating. Smart meter data were collected from Göteborg Energi, the DSO in Gothenburg, Sweden. The heating system and building category were collected from buildings energy declaration through Boverket [8]. In total, 100 customers of each category were chosen for 3 years of hourly data. After removing faulty smart meters, 89, 87 and 41 m remained for each category, respectively.

For the PLE of individual loads according to (3), α and β need to be estimated for each consumer category. For the aggregation of consumers using coincidence factor, either $c(N)$ as in (5), c_∞ as in (8) or $\bar{P}_0/P_0^{(X)}$ and ρ as in (10) needs to be estimated. For the aggregation using joint Gaussian method directly using (14), ρ_k and $\rho_{k,m}$ need to be estimated in addition to α_k and β_k. The following will present two parameter estimation approaches:

- *Non-parametric* – the peak load $P_\Sigma^{(X)}$ is estimated by applying the percentile function to the data, and $c(N)$ according to its definition in (5) directly.
- *Regression* – the model parameters in (3), (8), (10) and (14) are estimated by least square estimate with estimated mean and/or peak power for an individual or aggregated loads as an input.

5 Experimental results

This paper presents results on the model comparison and parameter estimation using all data for parameter estimation. For single-category load aggregation, the V – Velander's formula for individual loads (3) is used in combination with the coincidence factor, which can be estimated using three different approaches: c_0 – non-parametric coincidence factor (5), c_1 – coincidence factor assuming i.i.d. (8), c_2 – proposed coincidence factors considering correlation (10). These will be referred to as the V and c_0, V and c_1, and V and c_2 models, respectively, later. The fourth model is the joint Gaussian approach, referred to as the JG model, according to (14) with only one load category. For the aggregation of loads from multiple categories, the two evaluated models include T – the V and c_0 model together with the sum of the peak load from each category as in (11), JG – joint Gaussian model considering correlation as in (14). Table 1 summarises the number of parameters to estimate for each model.

The annual energy was estimated as the sample mean times T = 8760 h. The reference peak is calculated directly from the data as the 99.87th percentile of individual/aggregated loads, which corresponds to $K^{(X)} = 3$ for a Gaussian distribution. For the parameter estimation of V, all loads are used whereas for c_0, c_1, c_2, JG, and the model comparison, 1000 combinations of $N \in \{1, 2, \ldots, N_k\}$ randomly selected loads of each category are used. Note that for N loads from a subset of N_{tot} loads, there are $N_{\text{tot}}!/(N!(N_{\text{tot}} - N)!)$ unique combinations.

Table 1 Parameters to estimate where M denoted the number of categories and N the range of aggregated loads {2,3,...}

Model	c_0	c_1	c_2	V	JG
no. of parameters	NM	$1M$	$2M$	$2M$	$3M + \left(\frac{M^2 - M}{2}\right)^*$

*Number of pairwise categories.

5.1 Single-category peak load estimation

To compare different models to estimate coincidence factors, c_0 can be seen as the reference for coincidence factor estimation, as shown in Fig. 1. When assuming i.i.d. (c_1), the coincidence factor is not able to replicate the trend of c_0. By adding the correlation, the coincidence factor can give a much better fit to c_0. The average mean percentage error (MAPE) of the coincidence factor estimation can on average be reduced by 76–96% for the different categories.

Fig. 2 presents the mean percentage error of the four different PLE models. As for the coincidence factor estimation in Fig. 2, the V and c_1 model is not able to give an unbiased estimation of the peak load. The other three methods show an unbiased result when a higher number of loads are aggregated. However, there is a bias when only a few loads are aggregated, especially when the load number is less than 5. For fewer loads, the V and c_0 and the V and c_2 models outperform the JG model. This is because the probability distribution of the aggregated loads is changing when only a few loads are aggregated, and the JG model assumes a specific Gaussian distribution and thus cannot capture it. It is observed that the distribution function of the aggregated loads becomes more uniform with an increasing number of aggregated loads. Therefore, when optimised for one category, there can be a bias when only a few loads are aggregated. The V and c_2 model has one additional parameter as compared to the JG, which can reduce this bias further. The V and c_0 model shows the best fit but the number of parameters estimated increases proportional to the number of loads aggregated.

5.2 Multiple-category PLE

For multiple-category PLE, aggregating loads from the three categories with a 1:1:1 ratio is presented. The T model does not require further parameter estimation. For the JG model with

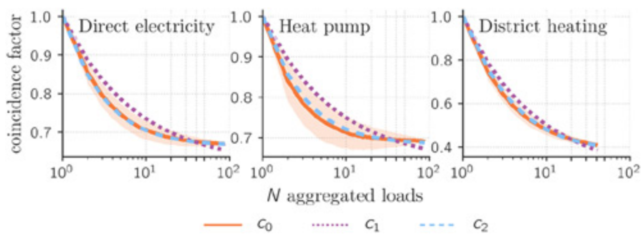

Fig. 1 *Coincidence factor with a standard deviation of c_0*

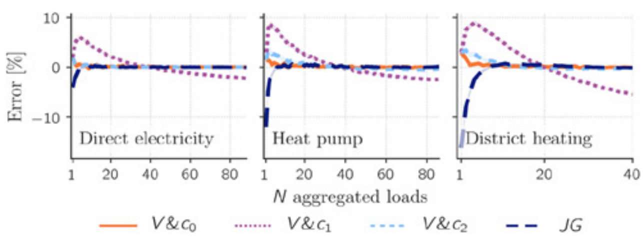

Fig. 2 *Mean percentage error of peak load estimation with 99.87th percentile of the load combination as reference*

CIRED, Open Access Proc. J., 2020, Vol. 2020, Iss. 1, pp. 178–181

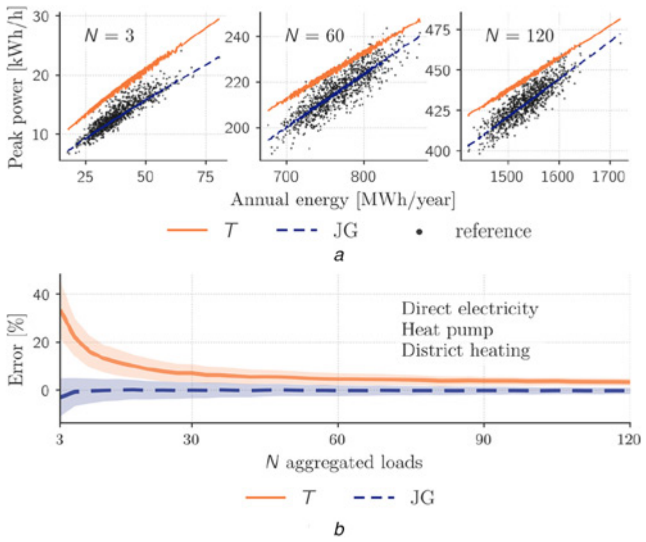

Fig. 3 *Aggregation of three categories with a 1:1:1 ratio with 99.87th percentile of the load combination as reference*

a PLE for $N = \{3, 60, 120\}$,
b Mean percentage error of PLE

parameters obtained from Section 5.1, $\rho_{k,m}$ is estimated with (14) by aggregating all loads in categories k and m.

Fig. 3 shows the results of aggregating loads from the three categories, where Fig. 3*a* shows the PLE for $N = \{3, 60, 120\}$ and Fig. 3*b* shows the percentage error. As the T model sums the peak of each category, it overestimates the aggregated peak from the three load categories by 3–34% due to the positive correlation between categories. In contrast, the bias of the PLE using the *JG* model is substantially reduced to 0.2–3%. However, note that the variance of the error in Fig. 3*b* is more or less unaffected by the different PLE methods. It is observed from the analysis that the assumption of constant VMR has the biggest impact on the variance of the error.

6 Conclusions and further work

By incorporating the correlation coefficient into Rusck's coincidence factor, the coincidence factor estimation shows a great potential of reducing the estimation error. However, the coincidence factors are only applicable to loads of the same category. The greatest benefit is seen when aggregating customers of different categories with the use of the joint Gaussian model considering the correlation between all loads. With only the annual energy and consumer category as input, the joint Gaussian model can give a less biased estimation when multiple loads are aggregated, especially for $N > 5$. However, as only 1–2 family households with different heating systems were considered, further analyses of a larger set of loads and with more diverse categories are needed to verify this approach. As most load patterns are not Gaussian distributed in real scenarios, interesting future work would be to extend to other methods, such as load curve with a joint Gaussian approach, Gaussian mixture models and machine learning.

7 Acknowledgments

The authors thank Ferruccio Vuinovich (Göteborg Energi AB, Sweden) for useful data and discussions.

8 References

1 Hemmingsson, M., Lexholm, M.: 'Dimensioning of smart power grids for the future: within ELFROSK program smart grids' (Elforsk, Stockholm, Sweden, 2013)
2 Rusck, S.: 'The simultaneous demand in distribution network supplying domestic consumers', *ASEA J.*, 1956, **29**, pp. 59–61
3 Velander, S.: 'Fördelningen av ett elverks fasta konstnader på olika abonnenter eller abonnentgrupper', *Tek. Tidskr.*, 1935, **65**, pp. 103–107
4 Bottaccioli, L., Di Cataldo, S., Acquaviva, A., *et al.*: 'Realistic multi-scale modeling of household electricity behaviors', *IEEE Access*, 2019, **7**, pp. 2457–2489
5 Sandels, C., Widén, J.: 'End-user scenarios and their impact on distribution system operators' (Energiforsk, Stockholm, Sweden, 2018)
6 Dickert, J., Schegner, P.: 'Residential load models for network planning purposes'. 2010 Modern Electric Power Systems, Wroclaw, Poland, September 2010, pp. 1–6
7 Sartori, I., Ortiz, J., Salom, J., *et al.*: 'Estimation of load and generation peaks in residential neighbourhoods with bipv: bottom-up simulations vs. Velander'. World Sustainable Building Conf., Barcelona, Spain, October 2014, pp. 17–24
8 SFS2006:985: 'Lag om energideklaration för byggnader', 2006