LEVERAGING DAILY SMART METER DATA TO ESTIMATE THE PEAK LOAD OF DISTRIBUTION TRANSFORMERS

Florent Cadoux[1*], Damien Frésier[2], Joris Garcia[2], Sébastien Julien[2], Sébastien Vallet[1]

[1]Roseau Technologies, Meylan, France
[2]GreenAlp, Grenoble, France
**florent.cadoux@roseautechnologies.com*

Keywords: PEAK LOAD ESTIMATION, SMART METER DATA, DISTRIBUTION TRANSFORMERS

Abstract

Accurate distribution transformer sizing requires reliable peak load estimation. Traditional methods, such as applying standard coincidence factors to subscribed powers, are often imprecise as they fail to reflect a DSO's specific customer behaviour. To improve accuracy, the French DSO GreenAlp leveraged its smart metering infrastructure. However, while smart meters can technically provide detailed 10- or 30-minute load curves, French regulations generally restrict DSOs to using only daily consumption data, except for specific projects. To address this constraint, GreenAlp analysed detailed interval data, i.e. 10- or 30-minute load curves, from 10% of its customers over a limited duration only; and relied on power subscriptions and daily metering data for the remaining customers and time periods. To this aim, a supervised learning model was developed in order to predict peak loads based solely on the legally accessible data. Cross-validation ensured the model's reliability, and a methodology was designed for handling incomplete customer data. The new approach proved significantly more accurate than the method previously used by GreenAlp, while remaining straightforward to implement once the learning model is trained. GreenAlp has integrated this model into its grid modelling software, replacing the previous load estimation method. It is now used for daily operations, providing a practical and scalable solution for optimizing transformer load assessments.

1 Introduction

1.1 Estimating Transformer Load at GreenAlp: Current Practices and Challenges

GreenAlp, a French Distribution System Operator (DSO) serving approximately 130,000 customers in and around the city of Grenoble, seeks to estimate the peak load of its distribution transformers. Load estimation is essential for operational management, to identify and address overloaded, but also potentially underloaded, assets. It is important also for planning purposes, such as determining whether a transformer upgrade is required before connecting new customers.

Measuring the peak load directly for each transformer is currently not feasible. No permanent sensors are installed on the transformers, and while mobile measurement equipment can be used, it is constrained by practical limitations: installing and removing data loggers requires significant effort, and only a few units are available.

To address this, GreenAlp has historically estimated peak loads using a methodology based on the notion of *coincidence factor* [1] or its inverse, the *diversity factor*. This approach involves multiplying the total subscribed power of considered customers by a specific factor that is determined by the number of customers. This factor is usually read from standardized tables, such as the one shown in Table 1. For instance, consider a group of five customers: three with subscribed powers of 6 kVA each and two with subscribed powers of 9 kVA each. According to this table, the peak power drawn from the grid by this group would be calculated as ~28 kW; this is because the total subscribed power is 36 kVA (3*6+2*9=36), the coincidence factor of a group of five customers is supposed equal to 0.78, and 36*0.78≈28.

Table 1 Example of coincidence factors (excerpt from the French C14-100 standard* [2])

Number of customers	Coincidence factor
1 to 4	1
5 to 9	0.78
10 to 14	0.63
15 to 19	0.53
…	…

** The specific coincidence factors used by GreenAlp to date slightly differ from the standard.*

GreenAlp, however, has found this methodology insufficiently accurate. Indeed, measurements conducted using mobile sensors and data loggers have repeatedly shown that the method of coincidence factors, with its current factor values, tends to overestimate peak loads: even when measurements were carried out over extended intervals of time, and during peak consumption periods, the recorded peak values were consistently and significantly lower than those estimated using the method of coincidence factors. This discrepancy led GreenAlp to suspect that the methodology was insufficiently reliable and that it resulted in potentially overly conservative decisions regarding transformer upgrades.

To address this issue, GreenAlp has initiated a project aimed at improving its method for estimating transformer peak loads.

1.2 Directions for Improvement

It was first observed that the accuracy of the method of coincidence factors could likely be improved by relying not on standardized values for the coefficients such as the ones presented in Table 1, but on values tailored to reflect the behaviour of GreenAlp's actual customers. In other words, a first and relatively straightforward improvement would thus be to retain the current methodology but to find a way to update and improve the value of its coefficients.

Additionally, it was noted that alternative methods, distinct from the method of coincidence factors, could also be explored. Specifically, the method of coincidence factors only relies on customers' power subscriptions to predict peak load, leaving aside other potentially useful descriptors such as customers' consumption history. It is thus desirable not only to update the coefficients of the method of coincidence factors, as suggested in the previous paragraph, but also to generalize this method to incorporate other available and potentially useful data that currently remain unused.

This raised several key questions:

1. Which new methodology, more general than the method of coincidence factors, could be used to estimate the peak load of distribution transformers? This includes in particular the question of which input data the methodology should rely on.

2. How to determine the parameters of the chosen methodology? For instance, in the context of the method of coincidence factors, how to tune the values of the coefficients to better align with GreenAlp's specific customer base, using data that is currently available to GreenAlp?

3. How to validate the methodology, including deciding between various methodological options and evaluating the confidence intervals associated with the results provided by the chosen method?

1.3 Smart Meter Data and Their Legal Framework

Regarding the question of which data to use — both as input for the load estimation method and as a basis for determining its parameters (e.g., tailoring the values of coincidence factors) during the training phase — smart meter data immediately emerged as an appealing source. Indeed, GreenAlp now has smart meters installed for most customers, and these meters provide a vast amount of new data that, to date, was not being leveraged for peak load estimation.

More precisely, the smart meters that are currently in use are technically capable of collecting 10-minute load curves for commercial and industrial customers with a subscription of more than 36 kVA; and 30-minute load curves for customers under 36 kVA. They also collect daily consumption data, which is technically redundant with 10- or 30-minute load curves. This daily data, however, is important for regulatory reasons: indeed, while daily consumption indices may freely be collected by the DSO in France, the collection of load curves is only permitted:

- for commercial and industrial customers with a subscription of more than 36 kVA,
- and for customers with a subscription of 36 kVA or less *provided that the customers give the DSO their explicit consent.*

Yet, at this point, few customers with a subscription of 36 kVA or less did provide their explicit consent, and there is little perspective that this situation will change substantially in the near future. Therefore, the 30-minute load curves of customers with a subscription of 36 kVA or less must be considered mostly as forbidden data, although it is technically possible to access it.

There is however an exception to this statement; the French regulator, CRE (Commission de Régulation de l'Energie), allows DSOs to access additional data, such as the 30-minute load curves, provided that this is motivated by a specific project of the DSO and approved by the regulator. In such a framework, additional smart meter data may be collected by the DSO, for a limited period and a limited number of customers as dictated by the needs of the project.

In the context of the project described herein, GreenAlp was thus able to collect 30-minute load curves from 10% of its customers over a duration of about one year.

The problem can now be stated as follows: based (1) on this partial data, and (2) on the remaining data that can be accessed freely, as summarized in Table 2, which methodology could be derived to assess the peak load of distribution transformers with improved accuracy?

Table 2 Summary of available data

Data field	% of customers with data available
Power subscription	100%
Upstream MV/LV transformer ID	~100%
30-minute load curve (≤ 36kVA)	~10% (over ~1 year)
24-hour load curve (≤ 36kVA)	53% (*)
10-minute load curve (> 36 kVA)	~100%

(*) The metering data was collected while the AMI deployment was ongoing. This ratio would now be close to 100% (not 53%).

2 Methodology

2.1 Modelling of Customers ≤ 36 kVA

Modelling customers with power subscriptions ≤ 36 kVA is the core challenge of our approach. This issue arises because detailed data (30-minute load curves) is available for only a limited period (~1 year) and for a small subset (~10%) of customers. Consequently, we must extrapolate from this limited dataset, which is accomplished using supervised learning, specifically regression analysis.

Using only the detailed dataset (10% of customers over 1 year), we began with data preparation by creating 12,500 random aggregates of customers. For each aggregate, we calculated the peak demand by summing individual 30-minute load curves (which, by construction, are available for these customers).

Next, we split the data into training and test sets, a standard technique to ensure model robustness and avoid overfitting.

After experimenting with various descriptors (or "features"), we selected the following two:

- first, the power subscription P_i (for instance, 6 kVA) of customer "i";
- and second, the parameter E_i that we defined as the total energy consumed by this customer during the 30 days of last year where the customer's daily consumption was the highest (for instance, ~500 kWh for a typical customer with a 6 kVA subscription).

Based on these descriptors, we opted for a predictor of the following form:

$$P_{estimated} = a\,\frac{\sqrt{\sum_{i=1}^{n}(P_i)^2}}{\sum_{i=1}^{n}P_i} + b\,\frac{\sum_{i=1}^{n}E_i}{\sum_{i=1}^{n}P_i} + c.$$

We then proceeded with model training, adjusting parameters a, b and c to minimize the Root Mean Squared Error (RMSE). This was done exclusively using the training set. We relied on the Scikit-Learn [3] library.

Finally, we validated the model's performance on unseen data from the validation set to assess its performance.

The fact that some customers have no consumption history was taken into account using the following method. To assess the peak consumption of an aggregate of customers, each with a power subscription under 36 kVA,

- for most customers, a consumption history over at least one year, consisting of daily load curves, was available; based on this history, we identified the 30 days where the customer's daily consumption was the highest, and calculated E_i;
- for some customers (e.g. new customers), no consumption history was available. In this case, we used a standard value for E_i; this value was calculated as the average value for the customers having the same power subscription.
- based on this input data, the regression model was used to estimate the peak load of the aggregate of customers.

2.2 Modelling of Customers > 36 kVA

For customers with a power subscription > 36 kVA, the situation is much simpler:

- for most of these customers, a consumption history over at least one year, consisting of 10-minute load curves, was readily available to GreenAlp. Downstream of each distribution transformer, we simply added those 10-minute load curves to estimate the peak power of these customers.
- For a fraction of customers (the ones that were recently connected to the grid, plus a few specific cases), there was no or insufficient consumption history to apply the technique of simply summing their 10-minute load curves. By default, we considered that those customers would potentially reach their power subscription.

To summarize, we considered each distribution transformer; identified the list of customers >36 kVA downstream of that transformer; and added their 10-minutes load curves (when available) or power subscription (otherwise). This yielded a total consumption profile, whose maximum was retained as the peak load of the group of customers understudy.

2.3 Modelling of a Combination of Both Types of Customers

The previous sections described how we estimated the peak load of each category of customers (≤ 36 kVA on one hand and > 36 kVA on the other hand) separately. We now turn to the question of estimating the peak load of an aggregate of both types of customers.

To address this issue, we randomly constituted 10,000 aggregates of both types of customers, using only customers for which the 10- or 30-minute load curve was available;

this allowed us to calculate the total load curve unambiguously, hence to calculate the coincidence factor between both groups of customers. The results are presented in Figure 1.

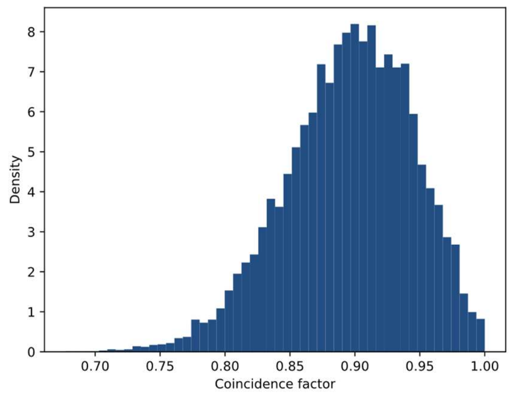

Fig. 1 Statistical distribution of the coincidence factor between both types of customers

This analysis showed that the coincidence factor was typically around 0.9 and could occasionally reach 1. For this reason, and for the sake of simplicity, we decided to apply the following conservative rule: the peak load of an aggregate of both types of customers was calculated by simply summing the peak loads of each of the two groups. In other words, we assumed that the coincidence factor between a group of customers <36 kVA and a group of customers >36 kVA was systematically equal to 1.

3 Results

We applied the methodology described in the previous section to evaluate the load of each of the distribution transformers that are currently operated by GreenAlp, and compared it with their nominal power. Figure 2 shows the results, in the form of a histogram of the statistical distribution of the load factor (i.e. the ratio of the peak load divided by the nominal power of the transformer).

The figure shows that for most transformers, the value of the peak load currently lies between 20% and 60% of their nominal power, which seems satisfactory. A handful of transformers, on the right of the *x* axis (load factor >1), were identified as slightly overloaded; those will be equipped with mobile measurement devices and data loggers for confirmation. A few transformers were also identified as surprisingly underloaded and were investigated as well.

By contrast, the method that was previously used by GreenAlp was leading to the conclusion (disproved, as explained above, by actual measurements based on mobile data loggers) that about 30% of distribution transformers were overloaded, sometimes by a vast amount. The new method, on the contrary, shows that the loading state of the distribution transformers operated by GreenAlp is overall very satisfactory.

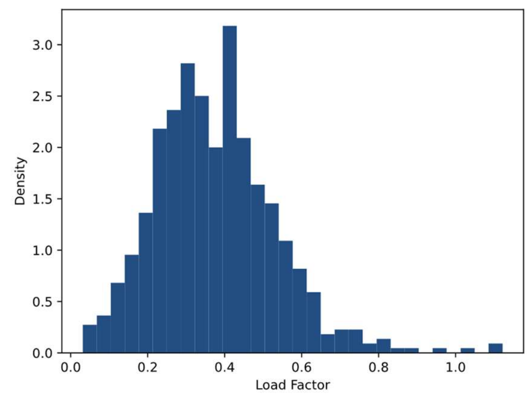

Fig. 2 Statistical distribution of the loading ratio of GreenAlp's transformers

4 Conclusion and perspectives

The approach presented in this paper allows GreenAlp to benefit from improved estimates of the peak load of distribution transformers, while remaining easy to implement. It was immediately applied in operation, where the new method helped target transformers that were potentially overloaded or underloaded and needed specific investigation. It was also immediately applied for planning purposes, to assess whether connecting new customers would potentially overload existing transformers.

Future work will include periodically re-running the supervised learning algorithm to update the coefficients based on new smart metering data. This is particularly important in the context of quickly evolving loading patterns; for instance, the development of electric vehicles in the city of Grenoble could substantially modify the relationship between input data (ie, power subscription and daily consumption data for customers <36 kVA) and output data (ie, peak power) and would probably require a readjustment of the coefficients.

5 References

[1] Grainger, J. J., & Stevenson, W. D. (1994). Power System Analysis. McGraw-Hill Education.
[2] NF C14-100: Installations de branchement à basse tension, 2008
[3] Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825–2830, 2011.