



HAL
open science

Autocalibration of a bottom-up methodology for long term electricity consumption forecasting

Kevin Bellinguer, B. Grept, A. Chevalier, Robin Girard, Alexis Bocquet

► **To cite this version:**

Kevin Bellinguer, B. Grept, A. Chevalier, Robin Girard, Alexis Bocquet. Autocalibration of a bottom-up methodology for long term electricity consumption forecasting. CIREC 2023 - 27th International Conference on Electricity Distribution, CIREC, Jun 2023, Rome, Italy. pp.2129-2133, 10.1049/icp.2023.1190 . hal-04533548

HAL Id: hal-04533548

<https://minesparis-psl.hal.science/hal-04533548v1>

Submitted on 17 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

AUTOCALIBRATION OF A BOTTOM-UP METHODOLOGY FOR LONG TERM ELECTRICITY CONSUMPTION FORECASTING

Kevin BELLINGUER
Mines Paris – France

kevin.bellinguer@minesparis.psl.eu

Benjamin GREPT
Mines Paris – France

benjamin.grept@etu.minesparis.psl.eu

Antoine CHEVALIER
ENEDIS – France

antoine.chevalier@enedis.fr

Robin GIRARD
Mines Paris – France

robin.girard@minesparis.psl.eu

Alexis BOCQUET
Mines Paris – France

alexis.bocquet@minesparis.psl.eu

ABSTRACT

In the coming years, the electrical grid will have to deal with deep changes in the way we consume and produce electricity. These changes result from the democratization of the electric vehicle, the rise of household production of photovoltaic energy, and the implementation of building retrofitting policies to name but a few. To support the grid to face these evolutions, the development of prediction tools for planning studies becomes vital for stakeholders and distribution system operators. It is in that context that the load curve simulator MOSAIC, developed by ENEDIS and Mines Paris, comes into play. This bottom-up approach simulates the current and future electric consumption of a wide range of uses (e.g. electric heating, cooking) for residential and tertiary customers. MOSAIC has been validated in the scope of residential customers during peak and average demand. Today the objective is to improve the accuracy of the simulator by providing a better calibration of the different uses. To this end, we propose in this paper a calibration process based on simplified versions of the models implemented in MOSAIC. Then an optimization approach is conducted to find the best parameters to fit observations. Scores improvements are observed not only with traditional error criteria, but also with new metrics proposed in this study.

NOMENCLATURE

T_{in}	Temperature inside the dwelling [K]
T_{out}	External temperature [K]
T_c	Target temperature [K]
T_0	Heating temperature threshold [K]
C	Thermal capacity of the building [J/K]
U_n	Coef. of heat transfer of the building [W/Km^2]
u_{i,j_n}	Thermal transmittance of the building part i , and age category j_n [W/Km^2]
COP	Coef. of performance of the device [\emptyset]
P	Power of the EH at the building level [W]
S	Surface of windows facing south [m^2]
E	Solar irradiance [W/m^2]
p_f^d	Daily power at the feeder level [W]
T_{out}^d	Daily outdoor temperature [K]
N_f	Number of customers at feeder f
F	Number of feeder
I	Set of the building parts (e.g. wall, floor, window)
E_f^{Max}	Maximum load observed at feeder f

INTRODUCTION

Transitioning to a low-carbon energy economy requires the development and the management of a resilient electrification system associated with appropriate long-term local investment strategies. These strategies range from the renovations of old dwellings, the development of electric car charging infrastructures, to the growth of rooftop photovoltaic energy production systems. In such a changing environment the projection of long-term electricity consumption becomes of great importance for supporting urban strategic decision-making and assessing their impact on the grid. For those reasons, Enedis, France's main electricity distribution system operator (DSO), and MinesParis are developing a new method based on a bottom-up electric load curve simulator. This simulator, named MOSAIC [1], [2], will enable Enedis to assess the impacts of different city-development-scenarios on the electrical grid for the household, industrial, and services sectors, at various time, and spatial scales.

Today, MOSAIC is calibrated through a set of empirical rules and physics-based knowledge for various uses ranging from electric heating (EH) to cooking needs. The ability of the model to simulate peak and average demand is validated on electricity consumption datasets from the household sector. The next step consists in improving the robustness, and the accuracy of the simulator by proposing an automated calibration process of the different modelling parameters. For this purpose, a large validation dataset composed of residential and tertiary consumers is considered. A straightforward optimization of the various parameters is not conceivable due to the tremendous computation time required to fit MOSAIC on this dataset.

This paper proposes to consider metamodels, which are simplified versions of the physics-based models implemented in MOSAIC, to reduce the computational complexity of the calibration process. The purpose of the metamodel is to mimic locally the behavior of the complex model as a function of the parameters that we want to calibrate. This methodology has to be practical with large numbers of parameters and observations. In this case, the estimation procedure implies a high dimensional optimization problem. The aim of this metamodel is both

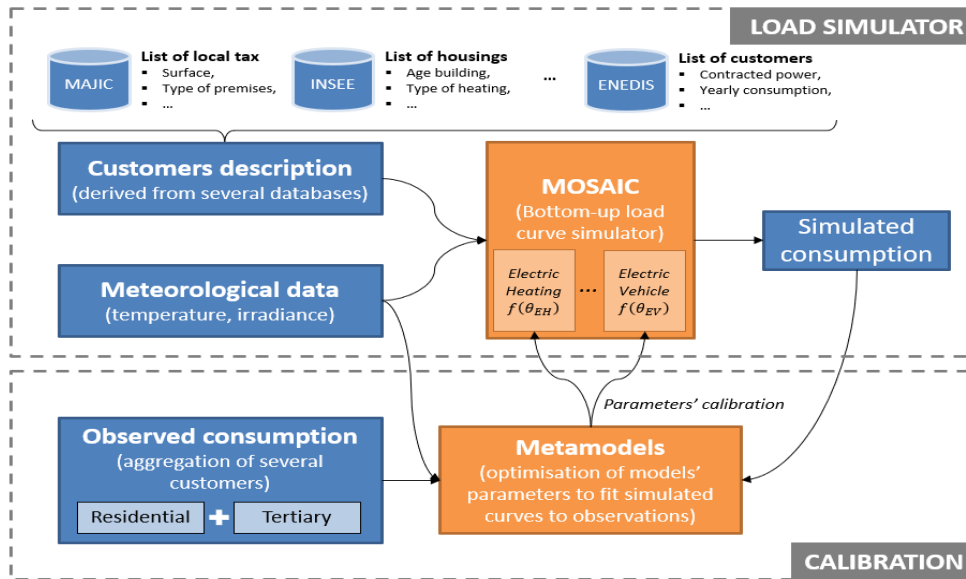


Figure 1 Overview of the simulation chain

to increase the speed of computation and to be compatible with linear high dimensional optimization algorithms. In a nutshell, these metamodels mimic the behavior of MOSAIC but at higher temporal and spatial scales (i.e. not all customers and temporal instances are simulated). Then during the calibration step, an optimization algorithm is used to adjust the metamodels' parameters to fit observations. The proposed calibration method permits to achieve higher accuracy compared to simulations generated from default values. Scores improvements are observed not only with traditional error criteria, but also with new metrics proposed in this study.

The paper is organised as follows. First, the load curve simulator MOSAIC is introduced. Then, the calibration methodology is detailed, while the next section discusses the outcomes. Finally, conclusions are drawn.

MOSAIC

This section introduces the load curve simulator MOSAIC. To assist the reader, Figure 1 provides an overview of the simulator, its inputs, and the way the calibration process is performed.

Modelling strategies

Typically, energy consumption modelling techniques are split into two fundamental categories: top-down and bottom-up approaches [3]. This wording can be seen as a shortcut for aggregated and disaggregated models. On the one hand, the top-down strategy is based on high-level information (i.e. macroscopic data) collected regarding an activity and its energy consumption, and often rely on statistical tools [4]. On the other hand, the bottom-up approach needs an exhaustive inventory (i.e. microscopic data) of all the equipment and its consumption pattern [4], then an extrapolation is performed to determine the consumption of the whole system (here a dwelling for instance). Therefore, this last philosophy is more detailed

but also more data intensive. In that respect, this high computation cost can be an issue when outputs are expected rapidly. The electric load curve simulator MOSAIC belongs to this last class of models.

Structure of the simulator

MOSAIC is based on a set of grey box models that combine complex physical approaches and fast statistical tools. As each model is devoted to a specific energy use, a compromise has been sought between models' complexity and computational time. Thus, for some uses such as EH and air conditioning (AC), MOSAIC computes a "RC" model (Equation 1) at each time step, while for other devices like the washing machine, we use blocks of power profiles (i.e. the first block represents the heating of water, the second, the washing cycle, and the third, the spinning step) associated with probabilistic laws to determine the number and duration of each instance. In the end, the simulator can generate in a few tens of seconds the yearly consumption of hundreds of customers at an hourly resolution. The architecture of the simulator is flexible enough to allow the integration of new elementary models that stand for additional consumption devices. One strength of MOSAIC is its capability to model the electricity consumption of a wide range of customers and uses. Thus, the load curve simulator is able to model the consumption of various devices found in the residential sector (i.e. houses and apartments): EH, AC, domestic water heater (DWH), freezer, refrigerator, oven, cooktop, washing machine, dryer, dishwasher, swimming pool, computer, TV, light, and miscellaneous. In addition, MOSAIC can also model the load consumption of tertiary sectors such as offices, teaching/research, shops, shops with food, restaurants, hotels, and charging stations for electric vehicles. The number of parameters associated with all the elementary models is tremendous. Up to now, the calibration of these parameters has been performed with data from the literature.

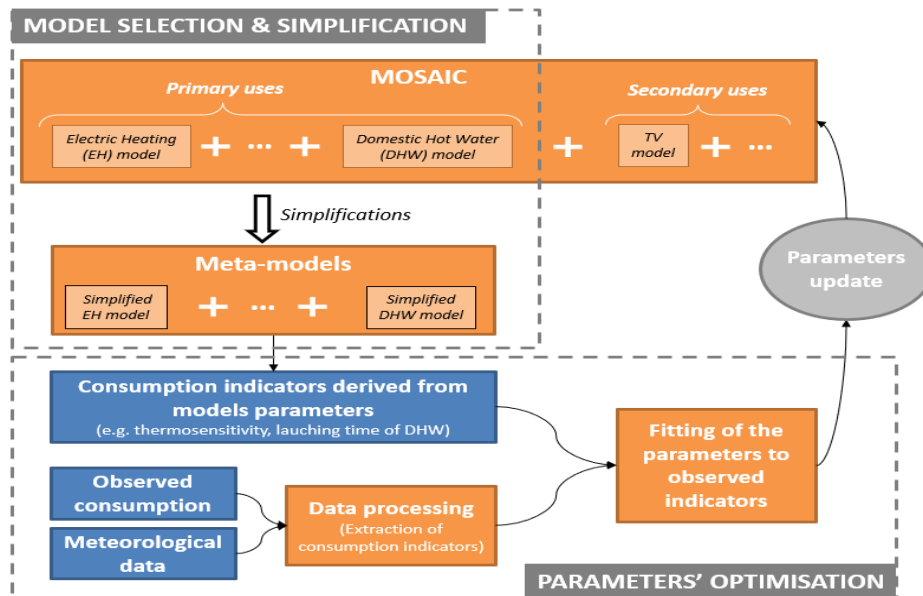


Figure 2 Overview of the optimization strategy.

Inputs

The inputs used to populate MOSAIC are large databases that have been matched together (the matching process is tackled in [2]). Some of these main databases are DSO customer information, national census, and meteorological information (i.e. temperature and irradiance). Thus, ENEDIS provides customer information regarding subscribed power, spatial location, and type of customer (residential or tertiary client). The BDTOPO database [5] from the National Institute of Geographic and Forest Information gathers buildings characteristics (e.g. text addresses and geometry of buildings), while the National Institute of Statistics and Economic Studies collects households information (e.g. type of heating, age of the building, number of inhabitants). Concretely, the simulator needs the following inputs to generate the load consumption of a residential customer: (1) the household characteristics (e.g. ownership rate of the different consumption devices, number of inhabitant), and (2) the building characteristics (e.g. surface, geometry, age, thermal properties). Then, the power of the devices, the frequency and the duration of use are determined by dedicated models embedded in MOSAIC from these inputs (e.g. the number of people living in the dwelling leads to the power of the DWH).

Outputs

As a bottom-up modelling-based strategy, MOSAIC provides as outputs the energy consumption time series of all uses selected by the user. This fine granulometry enables us to calibrate each use individually. It is worth mentioning that MOSAIC delivers load time series with a fine temporal scale up to the minute resolution. Such a resolution is vital to simulate realistic peak consumption and engagement of devices associated with off-peak hours (e.g. DWH).

Scope of application

MOSAIC is designed to simulate the current and future electric consumption of a wide range of uses for thousands of residential and tertiary customers. This tool can be used to support planning studies by assessing, for instance, the impact of building's thermal properties retrofitting strategies over electric consumption. In that sense, MOSAIC offers the possibility to policy makers to estimate energy savings and financial return associated with building retrofit policies. Its flexibility enables us to integrate new uses (e.g. electric vehicles) or upgrades of existing consumption devices, which provides valuable knowledge regarding the evolution of the electricity demand. Such information is of prime importance for the DSOs regarding the sizing of the power network, the demand side management, and the smart control of storage devices. MOSAIC, by simulating the energy need of a whole city, can also assist councillors in determining which suburbs have to be renovated and how.

DATA

Thanks to the large-scale deployment of smart meters, DSOs record the electric consumption at strategic points of the grid. As a result, for the purpose of this study we have at our disposal a three-year dataset of electricity consumption from around 940 distribution substations at a 10-minutes temporal granularity. Each substation gathers around 1000-10000 low-voltage electricity customers. In addition, ENEDIS provides information regarding the different clients (e.g. subscribed power) associated with each feeder. The local temperature and the irradiance data is also considered to account for thermal-related uses.

Before considering the calibration process, it is necessary to process these consumption time series to remove outliers. First, we identify and reject frozen observations (i.e. values that remain constant for several time steps) and

data outside the physical constraints imposed by the number of customers. In a second step, we look for potential changes in the consumption regimes. Such variations can result from an alteration in the number of customers connected to the feeder. To achieve this goal, we resort to a clustering algorithm fed with features such as the daily consumption, the daily temperature and the day of the year.

METHODOLOGY

In this section, we detail the approach developed to calibrate parameters that intervene in MOSAIC's modelling. In this regard, Figure 2 illustrates the optimisation strategy implemented. First, the calibration process focuses only on uses that have a significant impact on the load curve (e.g. EH, AC, and DWH). These primary uses consume more energy and usually have noticeable effects on the load observations. For instance, EH and AC devices impact the thermosensitivity of the consumption (i.e. the variation of electricity consumption due to the variation of the outdoor temperature). For the sake of conciseness, in this paper we only consider the EH. Then, we need to assess the use consumption sensitivity to the model's parameters, and to optimize only the most influential parameters. We highlight that the thermal resistance of the building has a genuine impact on the load curve.

$$C \frac{dT_{in}(t)}{dt} + U_n [T_{in}(t) - T_{out}(t)] = COP \cdot P(t) + S \cdot E(t) \quad (1)$$

The second step consists in deriving consumption indicators that constitute a bridge between the theoretical modelling and the observations. For each customer, MOSAIC computes a "RC" model (Equation 1) at each time step to determine the indoor temperature and the energy needs to reach the target temperature. First, we impose some simplifying assumptions to reduce the computational complexity of the model: we consider (1) a steady state, (2) $T_{in}(t) = T_c \forall t$, (3) $E(t) = 0 \forall t$, and (4) $COP = 1$. For all the dwellings of the considered feeder, Equation (1) becomes:

$$P_f^d = -T_{out}^d \sum_{n=1}^{N_f} U_n + \sum_{n=1}^{N_f} (U_n \cdot T_c(t)) \quad T_{out}^d \leq T_0 \quad (2)$$

Then, the term $\sum_{n=1}^{N_f} U_n$ represents the thermosensitivity of the consumption at the feeder level. From a theoretical perspective U_n is the thermal transmittance of the building and is computed as $U_n = \sum_{i=1}^I u_{i,j_n}$. The term u_{i,j_n} represents the thermal transmittance of the different walls of the building envelop (e.g. floor, windows, ceiling) for several classes of thermal performances based on the building's age. These values are derived from the 3CL-DPE diagnostic tool [6], which is an energy performance rating method. A straightforward linear regression between the daily consumption at the feeder level and the outdoor temperature leads to the observed thermosensitivity, α_f^d (Figure 3). We introduce β , a

corrective parameter of the thermal transmittance parameters. As a result, the optimisation problem is:

$$\min_{\beta} \sum_{f=1}^F \left| \alpha_f^d - \left(\sum_{i=1}^I \sum_{n=1}^{N_f} \beta_{i,j_n} \cdot u_{n,f,i,j_n} \right) \right| \quad (3)$$

As it is formulated, this problem can be solved with a least-square regression approach. Nevertheless, we turn to linear programming [7] to provide a generic framework that can be applied with other optimisation parameters, and that allows the consideration of constraints (e.g. thermal performances of new building are better than older ones).

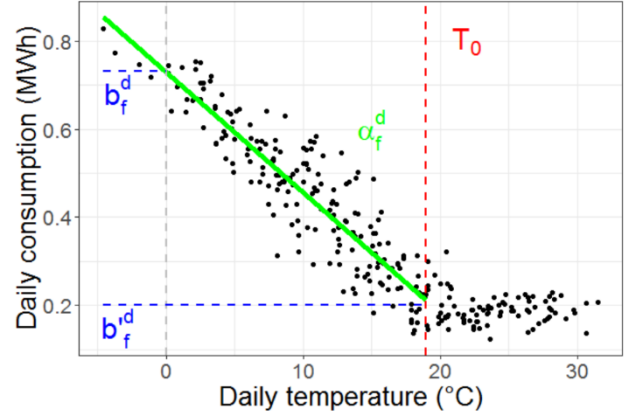


Figure 3 Observed consumption w.r.t. the outdoor temperature measured at the feeder f level. T_0 represents the threshold temperature below which customers used their EH, while α_f^d stands for the daily thermosensitivity obtained via a linear regression. b_f^d is the daily consumption measured when $T_{out}^d = 0^\circ\text{C}$, while b_f^d stands for the daily consumption without thermal-related uses.

VALIDATION

Two distinct simulations are run. First, we consider the default parameters of MOSAIC (this constitutes our reference model), then the parameters dedicated to the thermal properties of the building envelop are updated following: $U_n^* = \beta \cdot U_n$. For both approaches, all other uses (e.g. AC, DWH, cooking) are simulated with the default parameters of MOSAIC, and then aggregated with the EH consumption. In other words, only the load associated with the EH differs between the two simulations.

Data pre-processing

We have at our disposal a large range of observations that may represent distinct customer behaviours. For instance, the energy consumption of some feeder is driven by large scale companies that are not thermo-sensible, or by rural area highly depending on wood fires. In that context, such observations are not relevant for the EH calibration, and are then excluded.

To allow inter-comparison between the measured and simulated consumption time series, the latter are normalised with the maximal observed load at the substation level.

Evaluation Metrics

To assess the forecasting performances of the different approaches under consideration, it is necessary to define some metrics to quantify differences between observed and simulated values. We turn to a widely used metric, namely the normalised Mean Absolute Error (nMAE), which is described in Equation (4). In addition, we also compare the thermosensitivity exhibited by each feeder via Equation (5). This score makes it possible to assess only the EH influence. Scores are computed individually for the different feeders, before being averaged for ease of understanding.

$$nMAE_f = \frac{1}{N_f} \sum_{n=1}^{N_f} \left| \frac{E_n^{Obs} - E_n^{Sim}}{E_f^{Max}} \right| \quad (4)$$

$$nTS_f = \left| \alpha_f^{Obs} - \alpha_f^{Sim} \right| \quad (5)$$

Performances Assessment

From an energy perspective, Table 1 shows that the proposed optimization approach has a slight positive impact on the nMAE score: an averaged improvement of around 0.5% is observed for all considered substations. The proposed methodology leads to a 5.6% enhancement of the thermosensitivity-based score w.r.t. the reference approach. This significant outcome validates the proposed optimisation strategy regarding the parameters of the thermal envelop of the buildings.

Table 1 Average score

Simulation	nMAE (% of E_f^{Max})	nTS (% of $E_f^{Max}/^{\circ}C$)
Reference	0.263	0.0101
Optimised	0.262	0.00953

CONCLUSION AND PERSPECTIVES

In this paper, a new optimization framework is developed to calibrate the parameters of MOSAIC, a bottom-up energy demand model. This need for calibration originates from the fact that default parameters used in the model result from technical literature, survey, census, and other modeling strategies that may partially represent the effective customers' behavior. On the contrary, in this work, optimized parameters are derived from a large consumption databased composed of hundreds of measurements at the substation level during 3 years. As a first step, the proposed methodology is applied to calibrate the thermal properties of the building envelop. To do so, we rely on the thermosensitivity parameter that is easily computed from observations and derived from the heat transfer coefficients of the building used in MOSAIC. The optimization strategy improves the consistency of the thermal properties between the observed dataset and simulated load curves, while having a very slight influence on the total amount of consumed energy.

The following step consists in extending the proposed approach by including the optimization of target temperature parameter. Then, this approach will be applied

to optimize the parameters of the AC, before tackling the calibration of the DWH.

REFERENCES

- [1] N. Kong *et al.*, "Long-term forecast of local electrical demand and evaluation of future impacts on the electricity distribution network," *CIREN - Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 2401–2405, Oct. 2017, doi: 10.1049/oap-cired.2017.0743.
- [2] T. Barbier, R. Girard, F.-P. Neirac, N. Kong, and G. Kariniotakis, "A novel approach for electric load curve holistic modelling and simulation," in *MedPower 2014*, Nov. 2014, pp. 1–8. doi: 10.1049/cp.2014.1704.
- [3] T. Fleiter, E. Worrell, and W. Eichhammer, "Barriers to energy efficiency in industrial bottom-up energy demand models—A review," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 6, pp. 3099–3111, Aug. 2011, doi: 10.1016/j.rser.2011.03.025.
- [4] Y. Vogt, "Top-down Energy Modeling," *Strategic Planning for Energy and the Environment*, vol. 22, no. 4, pp. 64–79, Jan. 2003, doi: 10.1080/10485230309509626.
- [5] National Institute of Geographic and Forest Information, "BD TOPO® | Géoservices." <https://geoservices.ign.fr/bdtopo> (accessed Jan. 12, 2023).
- [6] Ministère de l'égalité des territoires et du logement, "Légifrance - Publications officielles - Journal officiel - JORF n° 0262 du 10/11/2012," Oct. 17, 2012. <https://www.legifrance.gouv.fr/download/pdf?id=qwfNFH6cKxfjW8YjmdKLcpMD0LeooqkyAbZzfe5bjzs=> (accessed Jan. 20, 2023).
- [7] L. Gurobi Optimization, "Gurobi Optimizer Reference Manual." 2023. [Online]. Available: <https://www.gurobi.com>