On the number of representative days for sizing microgrids with an industrial load profile

Selmane Dakir, Sélim El Mekki and Bertrand Cornélusse
Montefiore Institute, Department of Electrical Engineering and Computer Science
University of Liège
Liège, Belgium
Email: {s.dakir, selim.elmekki, bertrand.cornelusse}@uliege.be

***Abstract*—The sizing process of microgrids requires to run multiple simulations that can be computationally intensive depending on the desired accuracy. An effective way to reduce the simulation time is to compress the available data by selecting representative days from the list of days to be evaluated, such as the 365 days of a year, and assigning them a weight. The aim of this paper is to determine a recommended number of representative days for the sizing of microgrids with an industrial load profile. To this end, real load profiles were collected and analyzed from 22 companies. A sensitivity analysis on the optimal sizing identified according to the number of representative days is carried out for two representative days selection methods. A reliability indicator is proposed and allows to show that, with an optimization-based selection method, 10 representative days are enough on average to characterize the system.**

***Index Terms*—Load profiles, microgrids, representative days, selection methods, sizing**

I. Introduction

A microgrid is a small but complete electrical network equipped with its own resources for electricity generation and consumption. The microgrids considered in this paper are Belgian companies that have a known consumption pattern and can decide to invest in photovoltaic panels (PV) and a battery energy storage system (BESS). To correctly size a microgrid, financial flows over its lifetime (e.g. 20 years) must be considered. These financial flows can be estimated accurately by simulating in detail the microgrid operations. The sizing methodology adopted in this paper is based on the use of a microgrid simulator and a sizing module in charge of navigating in the search space to find an optimum. This sizing module therefore progressively generates configurations to be evaluated which are transmitted to the simulator. The latter simulates the operations of these configurations over one year and forwards cost indicators to the sizing module as explained in more detail in Section III. However, applying this procedure can be very time consuming, especially when the microgrid operations are managed by an optimization-based controller that solves an optimization problem at every time step. There are essentially two ways to reduce computation time:

1) at the simulator level by reducing the simulation time of a configuration through the use of representative days;
2) at the sizing module level by reducing the number of configurations to be evaluated by navigating more efficiently within the search space through the use of more advanced optimization techniques.

This paper focuses on the use of representative days in order to speed up the process of sizing microgrids. A representative day selection step must therefore be added to the procedure. Each selected day is assigned a weight that represents the number of days in the year with similar characteristics. Thus, the sum of the weights of the selected days must therefore be equal to the number of days in the year. Consequently, the cost indicators of the evaluated configurations can be approximated and simulation times are drastically reduced. However, the lower the number of representative days (NRD), the less reliable the results are. A trade-off must therefore be made between simulation time and accuracy of the results.

Energy-related problems often use representative days to lighten computations. However, in many cases, the choice of the number of representative periods is not discussed or specified [1], [2]. This can be explained by the lack of criteria or because the choice was made to carry out a very specific study. In other cases, heuristic methods accompanied by random selection methods are applied [3]–[6]. More rarely, the choice of the representative days is made by clustering methods [7]–[10] or optimization-based models [11], [12].

The goal of this paper is to evaluate a recommended number of representative days, by establishing a reliability indicator, for the sizing of microgrids with an industrial load profile. For this purpose, we collected 22 annual consumption profiles from companies located in Belgium. As all these companies have different activities, the diversity of the solutions obtained is significant as illustrated in Section II. In Section III, the microgrid sizing problem and the methodology used to solve it are detailed. Section IV describes the two selection methods analyzed in this paper, i.e. a random method and an optimization-based method. Then Section V presents the methodology used to derive the optimal microgrid sizing as a function of the NRD for each company, using both selection methods. Finally, the established reliability indicator is presented in Section VI and the results are discussed in Section VII.

II. Load profiles characterization

Given the variety of business activities considered, the load profiles analyzed in this paper are quite diverse as illustrated

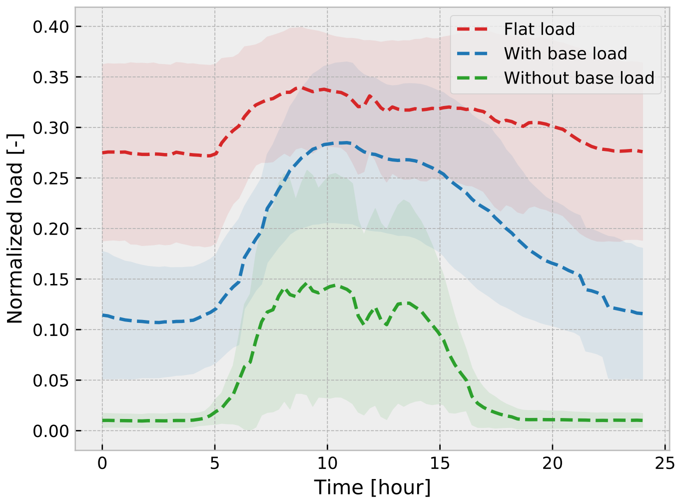

Fig. 1. Mean and standard deviation of the companies normalized load profiles grouped into 3 types.

in Fig. 1. Due to a lack of data, it was assumed that the withdrawal capacity contracted by the companies is equal to twice the highest peak consumption over the year. This value was therefore used to normalize the 22 load profiles, which have been classified into three types as presented in Fig. 1. The latter illustrates that some profiles are relatively flat and others have large changes in consumption between peak and off-peak times. In addition, there is a wide variety of peak and off-peak consumption and duration ratios and some profiles do not include a base load.

Similarly, the profiles show a great disparity in terms of power and annual energy consumed. The largest peak consumption varies from 3 kW to 2400 kW and the annual consumption from 10 MWh to 8000 MWh. Regarding the electricity purchase price, in order to fit as well as possible with the limited invoice data available, we assumed a price of 0.23€/kWh for a company consuming 10 MWh or less per year and 0.13€/kWh for a company consuming 2000 MWh or more per year. We also assumed that this price varies linearly between these two values as a function of the annual consumption. As regards the selling price, it is considered equal to 0.04€/kWh for all companies. The purchase price greatly influences the sizing. For example, two companies with exactly the same load profile but different purchase prices will not get the same optimal sizing. Indeed, the one with a high purchase price will have much more interest in investing in a storage system than the one with a low price.

III. The microgrids sizing problem

A microgrid sizing problem can be succinctly expressed as:

$$\max_{c,\lambda} \; NPV(c,\lambda) \tag{1a}$$

$$\text{s.t. } a_t = \lambda(v_t) \in \mathcal{A}(c_{[t-1]}, s_t),\ \forall t \tag{1b}$$

$$s_{t+1} = f(s_t, a_t, c_t),\ \forall t \tag{1c}$$

$$s_t \in \mathcal{S}_t(c_{[t]}),\ \forall t \tag{1d}$$

where c, a, and s are the vector of sizing decisions, control actions and state variables, respectively. The sub-vector $c_{[t]}$ of c denotes the sizing decisions from time 0 to time t. The actions available at time t are constrained to be in the set $\mathcal{A}$ that is a function of the investment made so far and the state of the system. The policy λ converts a state and observation of exogenous variables (e.g. the consumption and the renewable production) into an action that controls the system, of which the function f encodes the dynamics. NPV stands for "net present value" and can be expressed as:

$$NPV(c,\lambda) = \sum_{y=1}^{N} \frac{-I_y(c) + R_y(c,a) - O_y(c,a)}{(1+d)^y} \tag{2}$$

The numerator terms are dependent on the microgrid sizing decisions and represent annual costs. I stands for (re)investment costs (I_1 is the initial investment cost) R for revenues and O for overall operating costs. The present value of the cash flow is obtained by using a discount factor d that is applied over the microgrid lifetime in years N. The key challenge of microgrid sizing is the need to determine and simulate their operation. Furthermore, financial flows (R and O) are highly dependent on the control strategy applied. This problem is thus difficult because the NPV must be evaluated over a large period, the problem is non-convex, and we must optimize over the set of policies that satisfy the constraints of the system.

The sizing methodology used in this paper to solve this problem is based on the decomposition into three nested sub-problems:

1) From an initial sizing configuration (c_0, λ_0), a sizing module progressively generates new configurations (c_i, λ_i) to identify $\arg\max NPV(c,\lambda)$.
2) A simulator, taking as an input (c_i, λ_i), returns an evaluation of $NPV(c_i, \lambda_i)$ on a sample of representative days.
3) The control policy λ_i is evaluated for every simulated state in configuration c_i.

IV. Representative days selection algorithms

As mentioned in the introduction, there are different methods for selecting representative days. The heuristic algorithms generally consist of picking the days where the load is the highest and lowest over the year. When the required number of representative days is high enough, the algorithm will also look for the period with the highest and the lowest renewable energy production for each season. Clustering algorithms [7], [13] gather similar observations over the year in a same cluster by defining a distance threshold between observations. In this way, each representative day can be extracted from a cluster. The weight associated to each representative day is then proportional to the number of observations contained in the cluster from which it is extracted.

Random selection algorithms [11] randomly select days and evaluate the associated error according to the load duration curve. It iterates until the error is low enough (threshold constraint). Alternatively, it can also consider a set of randomly selected subsets of representative days, and pick the subset leading to the lowest error (set size constraint).

Mixed-integer linear optimization models [11] select the representative days and their weights in order to minimize the difference between real duration curves and duration curves reconstructed from the representative days. More precisely, the duration curves are discretized into a finite set of levels called bins. In the present case study, only the load and PV generation features f are considered. $L_{f,b}$ represents the number of periods of a time series where the value of feature f is higher than the value held by the bin b. In that sense, $L_{f,b}$ values represents the discretized form of the real duration curve of the feature f. The latter are compared to $\hat{L}_{f,b}$ values which are the values reconstructed from representative days. $\hat{L}_{f,b}$ is computed as follows:

$$\hat{L}_{f,b} = \sum_{d \in \mathcal{D}} \frac{w_d}{N_{tot}} A_{f,b,d}, \ \forall f \in \mathcal{F}, b \in \mathcal{B} \tag{3}$$

where $A_{f,b,d}$ is the number of periods during the day d where the value of feature f is higher than the value held by the bin b. In other words, it is a discretized duration curve but only based on the day d. w_d is the weight attributed to the day d and N_{tot} is the total number of days considered in the time series. The objective function of the optimization problem is to minimize the difference between the values of $L_{f,b}$ and $\hat{L}_{f,b}$ by acting on two decision variables types: the weights w_d and u_d, the latter being binary variables that indicate if the day d is selected or not. The optimization problem can therefore be cast as a mixed-integer linear problem as expressed below:

$$\min_{u_d, w_d} \sum_{f \in \mathcal{F}} \sum_{b \in \mathcal{B}} \mid L_{f,b} - \hat{L}_{f,b} \mid \tag{4a}$$

$$\text{s.t. } \hat{L}_{f,b} = \sum_{d \in \mathcal{D}} \frac{w_d}{N_{tot}} A_{f,b,d}, \qquad \forall f \in \mathcal{F}, b \in \mathcal{B} \tag{4b}$$

$$\sum_{d \in D} w_d = N_{tot} \tag{4c}$$

$$\sum_{d \in D} u_d = n \tag{4d}$$

$$w_d \leq u_d N_{tot}, \qquad \forall d \in \mathcal{D} \tag{4e}$$

$$u_d \in \{0, 1\}, \qquad \forall d \in \mathcal{D} \tag{4f}$$

$$w_d \in \mathbb{R}^+ \tag{4g}$$

Constraint 4c ensures that the sum of the weights corresponds to the total number of days in the original time series. The number of representative days n is imposed to the optimization problem through constraint 4d. Constraint 4e attributes a zero weight to a day d that is not selected.

There exists many error metrics to quantify the quality of the selected representative days [11]:

- the error due to the difference in mean between the real observations and the estimated ones.
- the error related to the duration curves.
- the correlation difference error between loads and productions observations. If these observations are not well correlated in the real data, then the same should apply to estimated loads and productions.
- the error associated to the ramp duration curves which is the duration curve of the differentiated observations. This metric is used to capture short-term fluctuations.

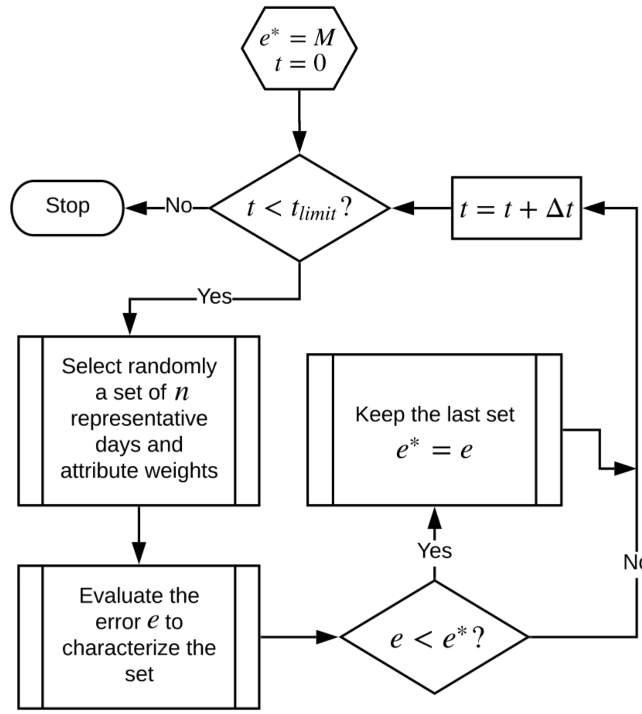

Fig. 2. DaysXtractor random selection algorithm.

To conduct this study, we compared a random selection method and an optimization-based method[1]. The implemented random selection method uses a set size constraint where the termination condition is imposed by a computation time limit t_{limit} as shown in Fig. 2. Each set of representative days is linked to an error e that is computed similarly to the objective function 4a and M is a large enough value. This error can only be determined by first assigning a weight to each randomly selected day. The latter reflects the number of days in the time series with similar characteristics. Hence, the sum of weights of the selected days is equal to 365, if one year of data is considered. Therefore, the algorithm passes over each day of the real data, identifies the selected day that most closely matches it and increments the weight assigned to that day by one. Finally, the selected set of days is that with the smallest error e^*.

Obviously, the longer the search time allocated to the random selection method, the more likely it is to find a good set of representative days. Nevertheless, we notice that after a given time limit, the precision of the selected days no longer increases significantly and tends to stabilize. This is illustrated in Fig. 3 where the evolution of the mean of the

[1]Implemented in the DaysXtractor tool available at https://github.com/sebMathieu/daysxtractor

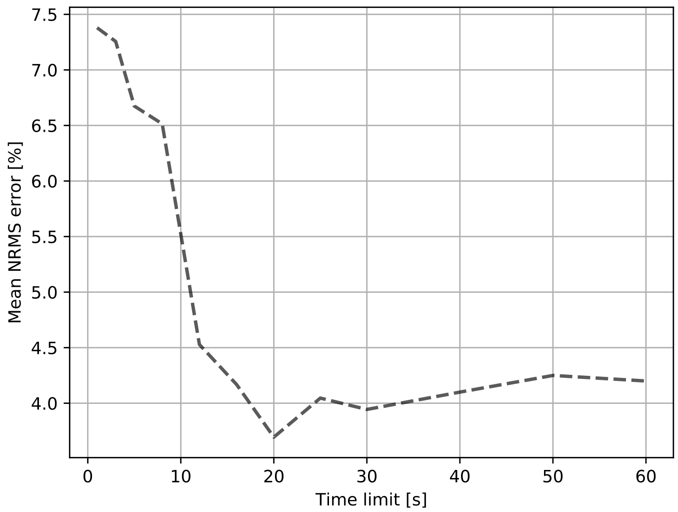

Fig. 3. Evolution of the mean NRMS error as a function of the time limit using the random selection method with NRD = 10.

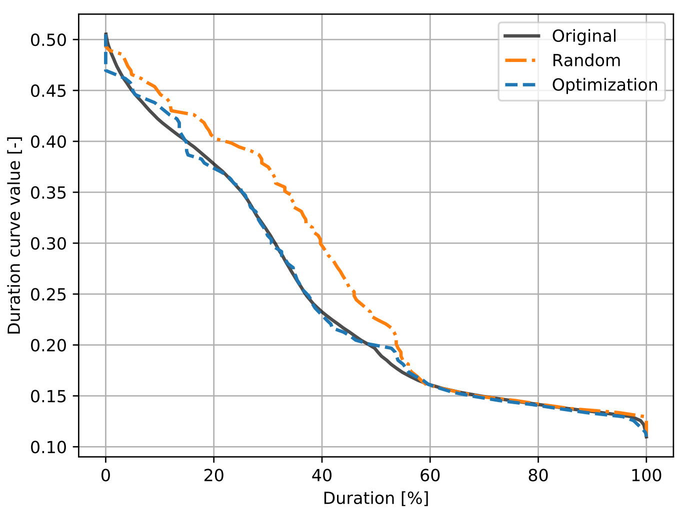

Fig. 4. Original and reconstructed load duration curves using the random method and the optimization-based method with NRD = 4 and a time limit of 15 s.

normalized root mean square (NRMS) error between the 22 companies original load duration curves and their respective one reconstructed with their set of 10 representative days as a function of the time limit is presented. Figure 3 shows that the NRMS error drops rapidly during the first 12 seconds of search and then stabilizes below 4.5%. The slight error increments after 20 seconds are due to the fact that the days selection process was carried out considering consumption and photovoltaic production data. Therefore, with more time, it is possible to find a more representative set of days that maximizes the overall accuracy of both duration curves but may slightly penalizes one or the other. For the rest of this study, we consequently used a time limit of 15 seconds.

Examples of load duration curves with 4 representative days using both selection methods are presented in Fig. 4. As expected, the load duration curve reconstructed with the

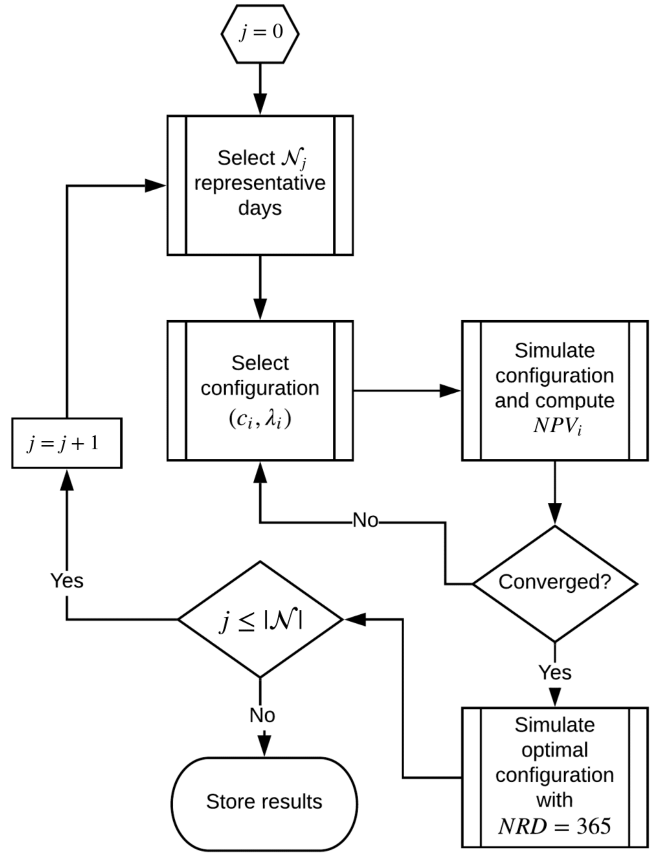

Fig. 5. Methodology used for the sensitivity analysis.

optimization-based method fits better the original curve.

V. Sizing as a function of the number of representative days

A series of simulations and sizing processes were carried out in order to quantify the impact of the number of representative days and their selection method on the companies microgrid sizing. Each company is considered to have sufficient space to install as much PV and BESS as necessary. The BESS is modeled as a tank of a given capacity, with charge and discharge efficiencies, and maximum charge and discharge power rates. Regarding PV generation, a typical generation profile for Belgium has been obtained through the PVGIS platform [14] and used for all companies. The controller employed applies a rule-based strategy designed to maximize self-consumption. The simulation time step is one hour and the investment horizon is 20 years.

The methodology used is described in Fig. 5. The set $\mathcal{N}$ gathers the evaluated NRD and is fixed to: $\mathcal{N} = \{1, 3, 5, 10, 20, 30, 40, 50, 75, 100, 150\}$. For each evaluated NRD, the optimal sizing is identified using a grid-search algorithm. Then, a complete simulation of the optimal configuration over the 365 days of the year is performed. The financial indicators derived from this simulation are used to compute the reliability indicator stated in the following section. The

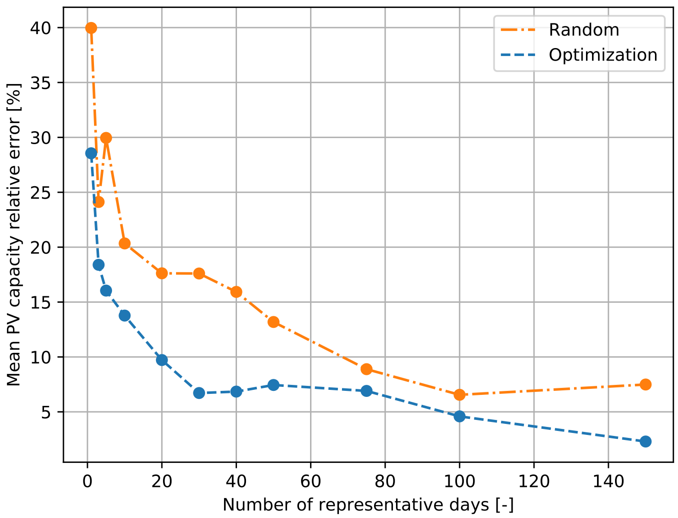

Fig. 6. Mean relative error of the PV capacity as a function of the NRD with both selection methods.

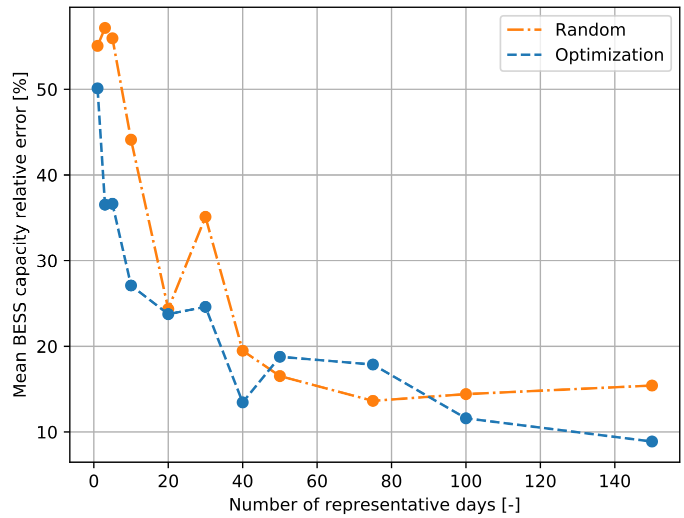

Fig. 7. Mean relative error of the BESS capacity as a function of the NRD with both selection methods.

convergence criterion corresponds to the maximum number of iterations, which was chosen independently for each company according to the search intervals selected. The larger the search intervals, the higher the number of iterations must be in order to guarantee the accuracy of the results. These intervals were identified beforehand by analyzing the companies consumption profiles and carrying out a preliminary sizing. This procedure was performed for each company and using the two previously discussed selection methods.

VI. Reliability Indicator

The proposed reliability indicator r is expressed as:

$$r_{n,s} = 1 - \frac{|NPV_{n,s} - NPV_{365,s}|}{|NPV_{365,s}|} - \frac{|NPV_{365,s} - NPV_{365,s^\star}|}{|NPV_{365,s^\star}|} \tag{5}$$

The first term quantifies the relative error obtained on the evaluation of the optimal NPV using n representative days. And the second term penalizes the reliability factor when the optimal sizing s obtained using n representative days differs from the actual optimal sizing $s^\star$, i.e. that achieved by simulating the 365 days of the year.

VII. Results and Discussion

The mean relative errors obtained for the PV and BESS capacities as a function of the NRD using the procedure presented in Fig. 5 are illustrated in Figs. 6 and 7. The reference values are obtained by sizing the system with 365 representative days. As expected, it can be noted that in general the selection method based on optimization gives better results than the random method. Furthermore, we can also notice a decrease in the relative error as the NRD increases. Nevertheless, for the random method some oscillations are observed.

The evolution of the mean $(\bar{r})$ and standard deviation of the reliability indicator as a function of the number of representative days is shown in Fig. 8. It can be seen that the selection method used greatly influences the outcomes. While convergence is achieved with 10 representative days using the optimization-based method, the random method converges only when this number reaches 50. We consider that there is convergence when:

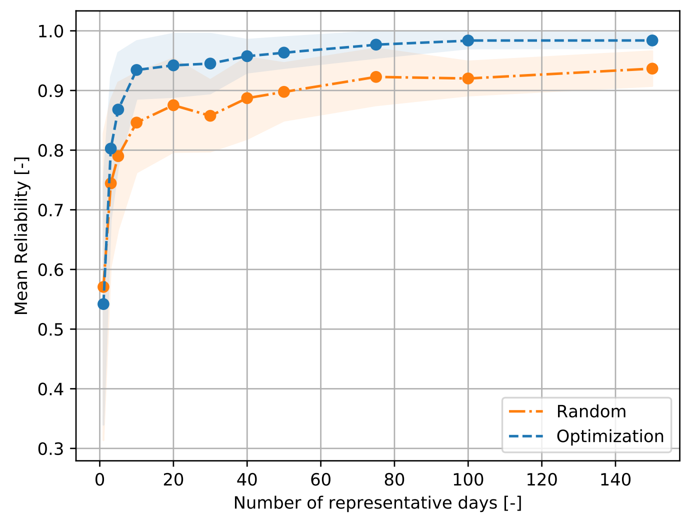

Fig. 8. Reliability indicator as a function of the NRD with both selection methods.

$$\bar{r}_n \geq 90\% \ \forall \ \text{NRD} \geq n \tag{6}$$

Moreover, the standard deviation obtained with the random selection method is larger. Therefore, unlike the optimization-based method where the reliability of the results is sufficient from 10 NRD (93% on average), a similar conclusion cannot be drawn with the random method. The latter must be used with 150 representative days to obtain the same level of reliability.

VIII. Conclusion

This paper presents a methodology to determine a recommended number of representative days for the sizing of microgrids with an industrial load profile. A sensitivity analysis was carried out on the sizing of microgrids as a function of the number of representative days using 22 actual Belgian consumption profiles. Two methods for selecting representative days, the first based on random selection and the second based on optimization, were analyzed and used. A reliability factor was established and made it possible to illustrate that beyond 10 days, the optimization-based selection method allows to obtain results with a degree of reliability higher than 93% on average.

References

[1] M. Fürsch, S. Hagspiel, C. Jägemann, S. Nagl, D. Lindenberger, and E. Tröster, "The role of grid extensions in a cost-efficient transformation of the european electricity system until 2050," *Applied Energy*, vol. 104, pp. 642–652, 2013.

[2] D. J. Swider and C. Weber, "The costs of wind's intermittency in germany: application of a stochastic electricity market model," *European Transactions on Electrical Power*, vol. 17, no. 2, pp. 151–172, 2007.

[3] M. Fripp, "Switch: a planning tool for power systems with large shares of intermittent renewable energy," *Environmental science & technology*, vol. 46, no. 11, pp. 6371–6378, 2012.

[4] C. Yuan, C. Gu, F. Li, B. Kuri, and R. W. Dunn, "New problem formulation of emission constrained generation mix," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4064–4071, 2013.

[5] E. K. Hart and M. Z. Jacobson, "A monte carlo approach to generator portfolio planning and carbon emissions assessments of systems with large penetrations of variable renewables," *Renewable Energy*, vol. 36, no. 8, pp. 2278–2286, 2011.

[6] M. Haller, S. Ludig, and N. Bauer, "Decarbonization scenarios for the eu and mena power system: Considering spatial distribution and short term dynamics of renewable generation," *Energy policy*, vol. 47, pp. 282–290, 2012.

[7] M. ElNozahy, M. Salama, and R. Seethapathy, "A probabilistic load modelling approach using clustering algorithms," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[8] W. A. Omran, M. Kazerani, and M. M. Salama, "A clustering-based method for quantifying the effects of large on-grid pv systems," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2617–2625, 2010.

[9] M. Nick, R. Cherkaoui, and M. Paolone, "Optimal allocation of dispersed energy storage systems in active distribution networks for energy balance and grid support," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2300–2310, 2014.

[10] J. Sachs and O. Sawodny, "Multi-objective three stage design optimization for island microgrids," *Applied Energy*, vol. 165, pp. 789–800, 2016.

[11] K. Poncelet, H. Höschle, E. Delarue, A. Virag, and W. D'haeseleer, "Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1936–1948, 2016.

[12] F. J. De Sisternes and M. D. Webster, "Optimal selection of sample weeks for approximating the net load in generation planning problems," *Massachusetts Institute of Technology Engineering Systems Division Working Paper Series: Citeseer*, pp. 1–12, 2013.

[13] P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf, "Carpe diem: a novel approach to select representative days for long-term power system models with high shares of renewable energy sources," 2014.

[14] M. Suri, T. Huld, E. Dunlop, and T. Cebecauer, "Geographic aspects of photovoltaics in europe: contribution of the pvgis website," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 1, pp. 34–41, 2008.