Optimized Time Reduction Models Applied to Power and Energy Systems Planning -Comparison with Existing Methods

Remy Rigo-Mariani

Optimized Time Reduction Models Applied to Power and Energy Systems Planning – Comparison with Existing Methods

Rémy Rigo-Mariani.[1*]

[1] *Univ. Grenoble Alpes, CNRS, Grenoble INP, G2Elab, 38000 Grenoble, France*

ARTICLE INFO

Keywords:
Power&Energy Systems,
Optimization,
Supervised Clustering,
K-medoid,
Hierarchical Clustering,
Dimensionality Reduction

Word Count : 12,034

ABSTRACT

The paper proposes a strategy for the time horizon reduction in power and energy studies. The method denoted Optimized Weighted Time Slices is compared with conventional approaches based on representative days that rely on unsupervised and supervised clustering as well as different strategies to reconstruct the problems. Those reference methods suffer from a lack of scalability when high numbers of dimensions are considered and their outputs strongly depend on the starting point in the partitioning process. The proposed strategy is based on a hierarchical clustering coupled with a least square minimization. The originality of the approach is that it works on individual time slices rather than on representative periods. Those representative time samples are furtherly optimized considering fitting criteria with the input time series thanks to a linearization of the duration curves. All the time modelling methods are tested on both a simple energy hub at a building scale (i.e. load, solar storage) and on a 33-buses distribution network with storage. The methods performances are assessed while comparing the results of the systems operation over the reduced time horizons with the outputs from full yearly simulations. In particular, complex objective functions are considered for the systems operation, as it is shown that they impact the accuracy of the time reduction as much as the systems complexity itself. The proposed strategy displays smaller errors (1 % - 5 % more accuracy) than the reference methods, is much more scalable (> 10 times faster), and systematically returns the same outputs.

HIGHLIGHTS

- A method for dimensionality reduction of time series based on the optimization of representative samples.
- Application to power and energy systems management or design problems.
- Comparison with supervised and unsupervised daily profiles clustering.
- Impact study of the management objective functions and problem reconstruction strategies after time reduction.
- Higher accuracy for nonlinear objective functions, faster computation and replicability.

1. Introduction

In order to face the limitation of fossil fuels and the need to reduce greenhouse gases emissions, a significantly increasing numbers of renewable-based distributed energy resources (DERs) have been connected to legacy power and energy systems in the past two decades. Those resources display many advantages: they actively participate in the decarbonization of the electricity sector, they allow to reduce system losses with generation closer to the points of consumption and they can defer grid investments in a context of growing demand [1]. However, renewable-based DER incur well identified challenges due to volatile production profiles in addition to a lack of controllability and the impossibility to predict them perfectly (around 80 % accuracy for intraday solar forecast in [2]). One envisioned way to address those emerging concerns is to introduce more flexibility in the considered systems in the form of storage units, controllable loads or partially controllable DERs. To that end, many works and projects investigate strategies for the operation and the design of hybrid power and energy systems including renewable and storage resources. Typical ‘energy systems’ or ‘energy hub’ refer to small scale microgrids at a building or district scale [3], that may or may not include various energy vectors (e.g. electricity, heat, gas) such as in [4] [5] [6] in order to take advantage of the interactions between different equipment and leverage additional flexibility. From the mathematical point of view, the optimal management of those systems is typically represented as power dispatch problems with energy exchanged between the various sources and sinks, subject to balance constraint at one ‘energy node’. ‘Power systems’ denote electrical networks for which the power balance shall be ensured at every node or bus in addition to the fulfilment of grid constraints (e.g. voltage and current limits) [7].

They correspond to more complex mathematical formulations due to the problem size that increases with the numbers of node and branches.

Conventional design problems for power or energy systems consist in finding the best DER configuration (e.g. type, size, location) in order to optimize objectives that could be economical (e.g. trade-off OPEX/CAPEX), technical (e.g. losses) or environmental (e.g. CO2 emissions) [8]. The main challenge encountered is to avoid prohibitive computational times while simulating the systems operations over long period of time and for different configurations before selecting the best investment plan. Obviously, the problem size and complexity may increase drastically due to the system dimensions (e.g. number of power system buses) and/or the need to simulate long time horizons. Especially, when storage resources are considered, the whole operating period shall be integrated in an optimal control problem in order to account for time coupled constraints (or 'cyclic constraints') that allow to compute the storage state of charge and fulfil its limits. In the vast majority of the encountered studies, one way to tackle this mathematical complexity is to make use of simplified models with convex optimization or mixed integer linear programming [5] [9] [10] [11].

Additional simplifications consist in reducing the simulation period. The main challenge is then to represent the whole input information in a shorter period with the profiles variability, the occurrence and concurrence of different events as well as potential correlations. The task relates to multivariate time series analysis and becomes more complex with increasing numbers of profiles that could also be heterogeneous in shape and scales (e.g. solar generation profiles, energy price profiles, load curves for different nodes in a power system). Time reduction strategies are not new and [12] [13] propose comprehensive comparisons of conventional methods with profiles downsampling (i.e. increased length of time steps), time slicing (i.e. limited representative time steps) and clustering strategies (i.e. selection of representative profiles). The latter approach is by far the most commonly applied in energy systems studies. It relies on three main characteristics, i) the distance computation between the profiles, ii) the selection algorithm and iii) the overall 'profile reconstruction' for the simulation [13]The main idea is then to select a set of representative periods, typically days, in the original data set with the use of hierarchical clustering [14], or partitional clustering, such as K-medoids [15] or K-means [16] [10]. Mixed heuristic approaches can be adopted in order to include days with extreme features (e.g. peak values in the input dataset) [12]. As pointed out in [17], the objective function targeted in the systems operation greatly impacts the performances of the time reduction techniques. Indeed, if classical strategies for the selection of daily profiles are accurate enough for energetic objectives (e.g. flat energy rates, constant efficiencies, etc), they may not be sufficient when the considered objectives and models embed strong nonlinearities (e.g. penalties for peak power consumption). Another way to enhance the performances of the time reduction techniques based on representative days is the implementation of supervised clustering methods. Similar to the proposition in [18], the main idea is to select the daily profiles while minimizing a set of metrics when comparing the reconstructed data, used for the simulation on the reduced space, with the input information (e.g. minimizing the error in the duration curves of the different time series and correlation coefficients). That 'profile reconstruction' step is indeed an important feature of time reduction techniques.. Typically, each representative day is weighted by the number of daily profiles in the corresponding cluster over the total number of profiles (e.g. 365 daily profiles for an initial yearly simulation). However, this approach is not deemed sufficient when seasonal storage units are considered with the need to represent long term energy variations, over the whole simulated period. Authors in [19] then propose a reconstruction strategy and an overall problem formulation that accounts for the state of charge computation in the inter periods between the selected representative days. Finally, time reduction techniques are conventionally tested and validated on simple case studies before being deployed on more complex systems or longer simulations up to decades [12][20]. Ultimately, a great diversity of scenarios and length of representative periods are encountered in the literature; heuristic selection of three representative days per month for the operation of a multi energy system in [21], from three up to seventy two daily profiles for similar application in [4], and up to three hundred days for the design of islanded energy system with solar and storage [8]. As already mentioned, the consideration of power systems incurs more complexity. Thus, smaller numbers of representative days are observed with around twenty profiles for capacity expansion problems in [12] or [18]. Also, it is worth mentioning that the time reduction performances are oftentimes assessed with regards to the results obtained in terms of system design such as the size of the assets or the final levelized cost of energy [12][18][19][22]. Such analysis could be misleading in cases where the installation costs have a much greater impact than the operating expenditures, and very distinct simulation profiles may lead to similar overall results.

The scope of this paper is the time reduction for both power and energy systems problems with storage units. Unlike the aforementioned studies, this paper proposes a time reduction techniques based on time slicing rather than the selection of representative daily profiles. Originally, time slicing in applied for long term prospective energy studies in macroscopic models such as TIMES [23] (e.g. only twelve time steps per year in [20]). The idea here is to compute the system state variables over a limited numbers of time steps while estimating the state of charge of the storage systems over the whole simulation horizon, similar to the inter period concept introduced by Kotzur et al. in [19]. The originality of the proposed approach is that the representative time steps are not taken among the existing values in the input time series but are optimized though a convex formulation that minimizes a set of criteria on mean square root errors and duration curves. The proposed approach is denoted Optimized Weighted Time Slices (OWTS). The main motivation is to address two concerns identified for conventional time reduction strategies based on representative days: i) the potential lack of scalability when the number of dimensions of the input profiles increases with higher combinatorial complexity (i.e. integer variables to select the days), especially for power system studies; and ii) typical sorting methods are not deterministic and do not systematically select the same days (i.e. outputs from conventional K-means or K-medoids depending on the starting points/centers). This lack of replicability is not necessary mentioned in the literature when K-means are applied, that is why the results presented in this paper will refer to mean performances over a given number of simulations for the reference clustering methods In order to decompose the design effect and better compare the proposed OWTS with the selected reference methods, the results are here analysed regarding the operating outputs only as in [13]. Finally, the main contributions of the paper are:

- A generic problem formulation for the operation of power and energy systems with storage, that is adapted for different time reduction and problem reconstruction strategies.
- A selection of reference methods for both the selection of representative days and the problem reconstruction: supervised and unsupervised clustering, weighted days with and without the representation of interperiods.
- The development of a scalable and deterministic time reduction method OWTS based on hierarchical clustering and convex optimization.
- The comparison of the proposed approach with the benchmarked methods on both power and energy system case studies and considering the operating aspects mainly.

The rest of the paper is organized as follows. Section 2 presents a generic problem formulation for the envisioned case studies with storage and details additional motivations for the proposed work after a set of preliminary tests is performed. The reference methods based on representative days for the time reduction and problem reconstruction are then presented in Section 3. The proposed OWTS is introduced in Section 4 before being compared with the references approaches in Section 5 along with a discussion on the reduction fitting performances. Conclusions are drawn in Section 6 and Appendix A gives the mathematical formulation of the two considered operational problems for power and energy systems.

2. Generic Problem Formulation and Paper Motivations

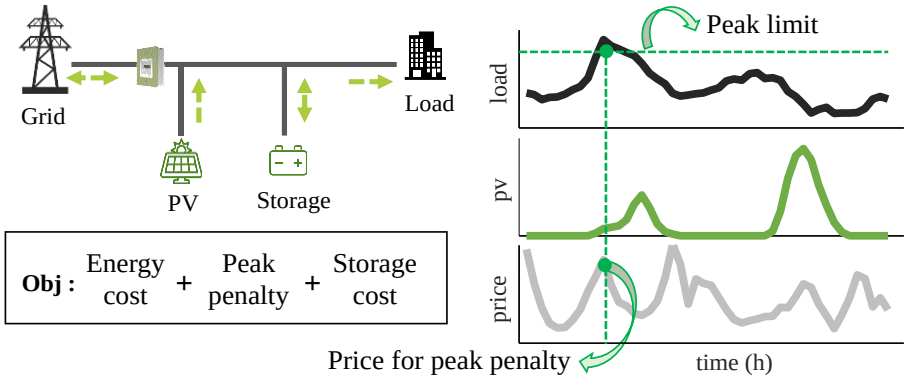

Figure 1: Generic energy system schematic

2.1. Optimization Problems

The objective of the work is to compare scenario reduction strategies without losing accuracy in the presence of storage units for both power and energy systems. The most generic energy system can be considered as a simple grid connected energy hub with a load, a photovoltaic (PV) generator and a storage unit (Figure 1). The obvious operational objective in such a case is the minimization of the energy bill with an arbitrage on the purchase price of the energy imported from the upstream grid, the local generation and potential penalties for peak load levels or storage degradation. Relying on typical linear formulations and models, such a problem can easily be solved over long period of times (i.e. years) with the set of equations detailed in the Appendix A.1. For the sake of genericity, and for a clearer explanation of the different profiles reconstruction strategies furtherly introduced, the problem formulation is simplified in this section. Input multi variate time series for decision making (e.g. load/generation profiles, prices) are denoted as $x_{t,n}$ and defined along a temporal set T with a given number of dimensions N. One assumption is that the considered objective f (e.g. energy bill) and constraints g (e.g. energy balance) can be computed at every time step depending on the input time series and system controls that are aggregated in a vector u_t. As already mentioned, a specificity of the problems considered in this paper is the integration of energy storage units. From the mathematical point of view, it introduces time coupled constraints with the computation of the state of charge s_t that depends on the controls and equipment efficiency (function h). Ultimately, a simplified and generic formulation of the encountered case studies can be summarized as in (1) for the optimization of the system controls within their bounds. Note that a constraint on the final value for the state of charge is often necessary in order to ensure energy conservation - the storage returns to its initial charge at the end of the simulated period. This periodicity constraint ensures that the system operation is consistent over the simulated period with no overall surplus or deficit of energy [4] [24].

$$obj: \min_{\underline{u} \le u_t \le \overline{u}} \sum_{t \in T} f(x_{t,n}, u_t)$$
$$\text{s.t.} \begin{cases} g(x_{t,n}, u_t) \le 0 \\ s_t = s_0 + \sum_{i=1}^{i=t} h(u_t) \\ \underline{s} \le s_t \le \overline{s} \\ s_{\|T\|} = s_0 \end{cases} \quad \forall t \in T \tag{1}$$

2.2. Preliminary Tests and Motivations

Preliminary tests are run on the generic simple energy system depicted in Figure 1. The objective function consists in minimizing the energy bill. The first component of this function is the cost for the purchased energy from the grid computed with real time prices (at 1 h resolution). If the power imported from the upstream grid is higher than a limit (e.g. subscribed power) it incurs penalties that depend on the corresponding real time energy price value. Lastly, the storage degradation is considered based on a given number of cycles, the rated capacity and the charge/discharge profiles (taken from the system controls u_t). More details on the models and equations are given in Appendix A.1. With linear formulations, the optimal operation of the system can quickly be computed over a whole

year. The output results are here considered as a reference when estimating the performances of the time horizon modelling strategies introduced in the following sections. Note that many similar studies in the literature compare those performances with reference results in terms of installed capacities and overall costs for optimal planning/design problems [19] [18]. Those optimal designs s usually rely on a trade-off between the operation and capital expenditures, with the energy bill estimated for different system configurations before returning the best solution. As pointed out in [13], assessing the performances on the operational results only allows to avoid the impact of the sizing. Indeed, as shown in previous results [11] [25], the design of energy system with the inclusion of installation costs tends to "smoothen" the problem with flat optimal areas due to convex formulations and greater sensitivity of the design functions (e.g. total cost of ownership, net present value) with the investment costs.

At first, four simple heuristic strategies are considered to decompose the time horizon based on the selection of representative days. This selection is computed with regard to the Euclidian distance between the daily profiles and along all the dimensions (i.e. load, solar and energy price profiles here (Figure 1)). The four chosen approaches are *i*) one representative day per season , *ii*) three representative days per season, *iii*) one representative week per season and *iiii*) one representative day per month. The results displayed in Table 1 are compared in terms of total cost error ΔB (in %), computed as the difference between the reference yearly energy bill (B^T over the time set T) and the one returned when considering the reduced time horizon (BT' over the time set T'). Note that the error estimation requires to consider the different time lengths of the simulation with the cardinal function denoted ||.|| in (2). Also the error can be similarly computed over the different bill component (i.e energy, peak penalty and storage degradation). Results are displayed in Table 1. As could be expected, the approach with the smallest number of days (4 days) gives the worst results in terms of total cost error. However, increasing the number of profiles does not systematically improve the accuracy. For instance twelve days in *iiii*) give much better performances than the twelve selected days in *ii*) or than twenty eight days in *iii*). Those performances may be explained by the uniform distribution of the days along the year in *iii*) compared to the other methods. Also, one noticeable result is that the error significantly differs with the considered cost components. Indeed, for the energy component the accuracy is above 90 % in every case with the average consumed energy easily captured by the selection processes. Note that the simulation considers here dynamic price profiles and outputs results could even display smaller errors with simpler schemes such as time of use rates or flat prices. As for the error on the peak penalty, it is significant, up to 100 % with four representative days only that do not capture any excess of limit power. As explained in [17] such nonlinear component is obviously less easy to capture by conventional selection processes (K-medoid here). In practise, it can motivate heuristic approaches with the additional selection of days with extreme values as suggested in [12]. Strong deviations are also observable regarding the storage degradation (cost) that is also nonlinear with the introduction of the absolute value for the power exchanged by the storage (both charge and discharge flows - see Appendix A.1 for the detailed equations).

$$\Delta B = \frac{\left| B^T - B^{T'} \times \frac{\|T\|}{\|T'\|} \right|}{B^T} \times 100\% \tag{2}$$

Table 1: Operational cost errors with the reference results

Error	4 days (1/season)	12 days (3/season)	4 weeks (1/season)	12 days (1/month)
Energy cost	6.9 %	8.7 %	0.6 %	6.9 %
Penalty peak	100.0 %	50.6 %	72.7 %	30.5 %
Storage cost	67.0 %	40.5 %	19.1 %	3.8 %
Total cost (ΔB)	**30.5 %**	**19.4 %**	**17.8 %**	**2.2 %**

As a conclusion, it is obvious that the original time series profiles have a strong impact on the time reduction strategies to adopt (e.g. seasonality effect, geographic coordinates). However, the previous results show that the objective to optimize shall be considered as well, with high accuracy for 'energetic objective' and strong deviations when nonlinearities occur. The idea of this paper is not to rely on heuristics approaches and to propose a

scalable/replicable methodology, indiscriminately from the objective function of the considered case study (see Section 4).

3. Reference Methods Based on Representative Days

As detailed in [13], typical time reduction methods can be characterized by three main features i) the distance computation, ii) the selection process and iii) the reconstruction phase. As for the distance, the Euclidian distance is traditionally considered for time series reduction in power and energy problems [9] [10]. However, Dynamic Time Wrapping gained interest recently for time series analysis [26] [27] due to its ability to measure the alignment between profiles over successive time steps while matching the peaks and valleys. In this paper, the impact of distance computation is left aside for the sake of simplicity. The way the distance is computed can have an impact on the clustering results and performances indexes (up to 20 clustering). However, it is hard to estimate this impact a priori once the management problem is run on the reduced time horizon. Indeed, as it is furtherly demonstrated and discussed in the paper, a good clustering does not systematically corresponds to the most accurate results when the reconstructed problems are solved **(**Section 5) The Euclidian distance is then considered in this paper, and the work focuses on the two other features of the reduction process: the days selection and the problem reconstruction. Especially, unsupervised and supervised selection processes are investigated with different strategies to rebuild the whole time horizon and optimization problems. The methods introduced in this section will be considered as reference approaches to assess the performances of the proposed OWTS.

3.1. Day Sorting Strategies

3.1.1. Unsupervised Clustering : **CL₀**

Partitional clustering is traditionally considered when dealing with time reduction for power and energy problems. The main objective is to select a set of representative days with the use of K-means or K-medoid algorithms. In this paper K-medoid will be preferred as conventional K-means may lead to oversmoothed profiles that could incur a loss of accuracy and the impossibility to capture extreme values ([13]). Indeed, the objective functions considered in the encountered problems often display significant values in cases of extreme events – e.g. penalty for peak consumption in an energy bill minimization or significant marginal cost (or CO_2 emissions) in unit commitment problems. Typically, time series are arranged along a set of days D with $x^{d}_{t_d,n}$, and T_d denoting the daily time set (e.g. 1h,..,24h). K-medoid algorithm aims at sorting the days within K different clusters $D_k \subset D$ and while updating the clusters centers (medoids) until they remain constant over two successive iterations. Ultimately, a set C_k^* of optimum centers c_k^* is returned with the Euclidian distance computed along the daily time set and different dimensions ($l(.)$ in (3)).

$$c_1^*,...,c_K^* : \arg\min \sum_{c_k \in C_k} \sum_{d_k \in D_k} l\left(x^{d_k}_{t_d,n}, x^{c_k}_{t_d,n}\right)^2 \tag{3}$$

3.1.2. Supervised Clustering – Cluster Performances : **CL₁**

One identified drawback of conventional K-medoid is that the results in terms of returned clusters and medoids strongly depends on the starting points. In particular, the use of the method does not prevent to obtain many clusters with single profiles for some extreme events and few partitions that embed the majority of the other profiles. In order to mitigate those issues, a second reference method is considered here and relies on a supervised partitioning that maximizes indexes related to the clustering quality [8] [28]. The first metric is the cluster tightness index *CTI* in [0,1] that measures the extent to which each daily profile is proximate to the other profiles within the corresponding cluster $D_k(d)$. This index evaluates to one if all the profiles are in distinct clusters and to zero if only one cluster aggregates all the input data ((4)). The second metric is the cluster size index (*CSI* in [0,1]) that estimates the homogeneity between the different clusters sizes with the cardinal function denoted ||.|| in (5). Ultimately, the objective is to maximize those two metrics following (6) with priority given to the *CTI*. This maximisation is performed while integrating the convention K-medoid procedure into a genetic algorithm that generates the starting points and successively run different K-medoid along the iterations (generations) [29]. Especially, the number of generations and the size of the population shall increase with the predefined number of clusters *K*. Note that the outputs of this hybrid clustering will vary from one run to the other due to the stochastic nature of the genetic

algorithm. Thus, average results and confidence intervals over several computations will be displayed when the performances of the references methods are assessed in the further sections, which is not usually mentioned in similar studies.

$$CTI = 1 - \frac{\sum_{d \in D} \sum_{d_k \in D_k(d)} l(x_{t_d,n}^{d}, x_{t_d,n}^{d_k})}{\sum_{\{i,j\} \in D} l(x_{t_d,n}^{i}, x_{t_d,n}^{j})} \quad (4)$$

$$CSI = e^{-\left(\ln\frac{\|D\|}{K} - \ln\sum_{d \in D}\frac{\|D_k(d)\|}{\|D\|}\right)^2} \quad (5)$$

$$obj : \max_{c_k}\left(CTI^2 \times CSI\right) \quad (6)$$

3.1.3. Supervised Clustering – Profiles Performances : **CL2**

Following the hybrid strategy introduced in the previous subsection, a last reference method to select the representative days relies on a supervised clustering that compares the representative profiles with the original input data set. Overall time series profiles $\tilde{x}_{t',n}$ are generated as the aggregation of the representative days defined by the optimum clusters centers c_k^* ((7)). Those reconstructed profiles are then defined along a new temporal set $t' \in T'$, shorter than the original set T and with a length that depends on the numbers of representative days (i.e. the number of clusters K) and the time resolution of the daily profiles (i.e. $\|T'\| = K.\|T_d\|$). At first, the outputs from a regular K-medoid are analysed regarding the load duration curves denoted with the subscript $(.)^{DC}$ and defined along the N dimensions. This analysis involves the computation of the normalized root mean square error ($nRMSE$) – error with the duration curves of the original multi-variate time series ((8)). The duration curves are disaggregated in 100 bins and normalized with the extreme values of the input data set. Secondly, the correlation between the two time series dimensions $\{i,j\}$ in N is also computed using the Paerson coefficient $p(x_{t,i}, x_{t,j})$. Following [18], the average correlation error CE_{av} is defined as the difference between the Paerson coefficients computed for the original data set and for the reconstructed profiles ((9)). Similar to the previous subsection, the metrics to optimize are arranged as a single criteria ((10)) that defines the objective function of the genetic algorithm in the supervised clustering.

$$\tilde{x}_{t',n} = \bigcup_{c_k^* \in C_k^*} x_{t_d,n}^{c_k^*} \quad (7)$$

$$nRMSE = \frac{1}{\|N\|}.\sum_{n \in N}\sqrt{\frac{1}{100}\left(\frac{\left(x_{t,n}\right)^{DC} - \left(\tilde{x}_{t',n}\right)^{DC}}{\max\left(x_{t,n}\right) - \min\left(x_{t,n}\right)}\right)^2} \quad (8)$$

$$CE_{av} = \frac{2}{\|N\|.\left(\|N\|-1\right)}.\sum_{\{i,j\} \in N}\left|p\left(x_{t,i}, x_{t,j}\right) - p\left(\tilde{x}_{t',i}, \tilde{x}_{t',j}\right)\right| \quad (9)$$

$$obj : \min_{c_k}\left(nRMSE \times CE_{av}\right) \quad (10)$$

3.2. Time Series Reconstruction and Case Study Solving

3.2.1. *Successive Days :* **OPT0**

Once the representative days are selected, different strategies may be adopted to account for the reduced time horizon in the reformulation of the optimization problems. The simplest approach consists in solving a new problem over the shorter time set T'. The time series profiles $\tilde{x}_{t',n}$ in (7) aggregate the representative daily patterns arranged according to their order of occurrence in the original set of days D. The variables and constraints are then defined along this new temporal set and the objective function shall take into account the length ratio between the original problem and the reduced one ((11)) – problem re-formulation denoted **OPT0**.

$$
\begin{aligned}
&obj: \min_{\underline{u} \le u_{t'} \le \overline{u}} \frac{\|T\|}{\|T'\|} . \sum_{t' \in T'} f(\tilde{x}_{t',n}, u_{t'}) \\
&\text{s.t.} \begin{cases} g(\tilde{x}_{t',n}, u_{t'}) \le 0 \\ s_{t'} = s_0 + \sum_{i=1}^{i=t'} h(u_{t'}) \\ \underline{s} \le s_{t'} \le \overline{s} \\ s_{\|T'\|} = s_0 \end{cases} \quad \forall t' \in T'
\end{aligned}
\tag{11}
$$

3.2.2. *Weighted Days :* **OPT₁**

For better accuracy, typical approaches in the literature account for the weight w_k of each representative day c_k^*, computed as the ratio between the corresponding cluster size $\|D_k\|$ and the original numbers of days $\|D\|$ [6] This leads to a modification of the problem, discretized along the different representative days. Especially, the control variables are arranged in daily profiles $u_{t_d}^{c_k^*}$ and the constraints are defined along the set of representative days and daily time set T_d. However, the computation of the state of charge shall remain the same as in the previous formulation (i.e. along T') in order to capture the yearly evolution with the days arranged in their order of occurrence in the original data set ((12)) – problem re-formulation denoted **OPT₁.**

$$
\begin{aligned}
&obj: \min_{\underline{u} \le u_{t_d}^{c_k^*} \le \overline{u}} \sum_{c_k^* \in C_k^*} w_k . \sum_{t_d \in T_d} f(x_{t_d,n}^{c_k^*}, u_{t_d}^{c_k^*}) \\
&\text{s.t.} \begin{cases} g(x_{t_d,n}^{c_k^*}, u_{t_d}^{c_k^*}) \le 0 & \forall \{t_d, c_k^*\} \in T_d, C_k^* \\ s_{t'} = s_0 + \sum_{i=1}^{i=t'} h(u_{t'}) & \forall t' \in T' \\ \underline{s} \le s_{t'} \le \overline{s} & \forall \{t_d, c_k^*\} \in T_d, C_k^* \\ s_{\|T'\|} = s_0 \end{cases}
\end{aligned}
\tag{12}
$$

3.2.3. *Inter-Periods :* **OPT₂**

The main drawback of the previous formulations is that the potential of the energy storage units cannot be fully harnessed in case of significant capacities. For instance, if seasonal storage is considered in the targeted applications, the proposed formulations **OPT₀** and **OPT₁** do not allow to represent the evolution of the state of charge s_t between the selected periods. In other words, the initial state of charge for each representative day is equal to the final value computed for the previous day in the sequence. It is then impossible to account for the integration of any surplus/deficit of energy over periods between two successive representative days c_k^*. Similar to [19], the last reference strategy to reconstruct the profiles **OPT₂** consists in a representation of the inter periods dependencies for the state of charge estimation. The objective function and constraints remain the same as in **OPT₁**, and the degrees of freedom for the storage (i.e. charge/discharge) are computed along the sets T_d and C_k^*. However, the state of charge is estimated all along the original time set T. The storage controls for every day in D are deduced from the controls $u_{t_d}^{c_k^*}$ for the corresponding cluster representative profile noted $c_k^*(d)$.

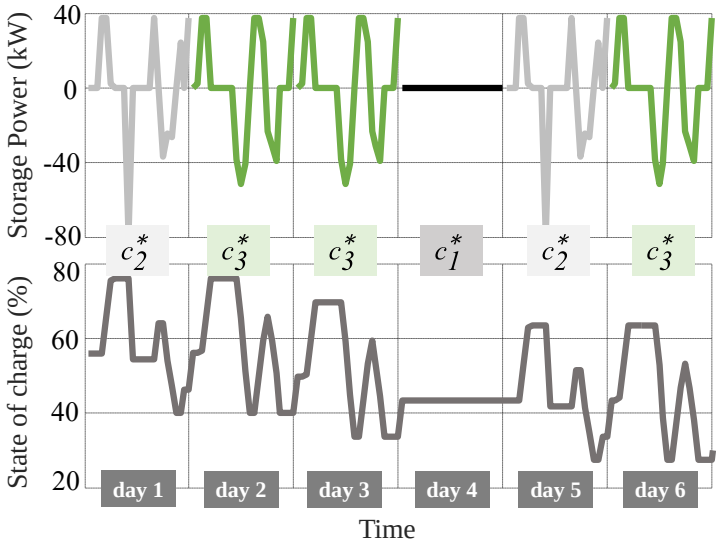

Figure 2: State of charge modelling in OPT$_2$

Figure 2 gives a graphic representation of the computation for the state of charge based on the controls for the representative profiles corresponding to each day in the year (i.e. day 1, day 2, etc). Note that the figure corresponds to an arbitrary storage system for illustrative purposes only. The mathematical formulation of the reconstruct problem is defined in (13) with the necessity to map each day d to its representative profile $c_k^*(d)$. In practice, this is simply done with an appropriate matrix issued from the clustering processes. The state of charge variable can then be defined along the set of total days D with the successive values arranged in the variable $s_{t_d}^d$. In addition, there is a need to ensure the state of charge continuity along the days with the initial value ($t_d = 0$) at a given day d equals to the final values ($t_d = ||T_d||$) at the previous day d–1. Note that this formulation incurs an increased numbers of variables in the optimization compared to the previous formulations, due to the state of charge described along the whole original time horizon T. However, the control variables (i.e. the degrees of freedom) remains defined along the reduced time horizon and the complexity of the problem is still reduced compared to the reference yearly simulation, especially for case studies with few storage units connected to large power systems.

$$
\begin{aligned}
&obj: \min_{\underline{u} \le u_{t_d}^{c_k^*} \le \overline{u}} \sum_{c_k^* \in C_k^*} w_k \cdot \sum_{t_d \in T_d} f(x_{t_d,n}^{c_k^*}, u_{t_d}^{c_k^*}) \\
&\text{s.t.} \begin{cases}
g(x_{t_d,n}^{c_k^*}, u_{t_d}^{c_k^*}) \le 0 & \forall \{t_d, c_k^*\} \in T_d, C_k^* \\
s_{t_d}^d = s_0^d + \sum_{i=1}^{i=t_d} h(u_i^{c_k^*(d)}) & \forall \{t_d, d\} \in T_d, D \\
\underline{s} \le s_{t_d}^d \le \overline{s} & \forall \{t_d, d\} \in T_d, D \\
s_0^d = s_{\|T_d\|}^{d-1} & \forall d \in D - \{1\}
\end{cases}
\end{aligned}
\tag{13}
$$

4. Optimized Weighted Time Slices – OWTS

4.1. Time Steps Hierarchical Clustering

Similar to [18], the approach proposed in this paper consists in a two stage procedure with a first clustering step that is further refined thanks to an optimization process. As previously introduced, most of the reviewed literature performs clustering while selecting sets of representative periods in the original time set - typically representative days as in the reference methods in the previous section. Inspired by the long term macroscopic energy models (e.g. TIMES, EUTCAD), the approach proposed in this paper considers the selection of individual time slices in order to find a set of K representative time steps $x_n^{k^*}$ similar to the approaches in [30] [31]. The idea is to provide more flexibility in the time representation with no predefined sequences and while decoupling the occurrences and concurrences of the different events - independently over the different dimensions of the original data set. In order to ensure the replicability of the results, a conventional hierarchical approach is considered in the clustering phase - as

opposition to K-medoid whose results depend on the initial conditions. Hierarchical clustering processes in successive pairwise groupings of the studied samples based on their distance (Euclidian distance still considered here). The samples are first normalized with the maximum values along the different dimensions *n*. They are then iteratively aggregated in clusters, with the inter cluster distance computed based on their centroid (i.e. representative sample). The clustering results in a typical dendrogram as displayed in Figure 3, with the vertical axis denoting the distance between the groups. Clusters can then be easily defined based on the successive inter distances computed (inter 'leaf nodes' distance) – in Figure 3, four clusters are defined with the partitioning line crossing four vertical edges.

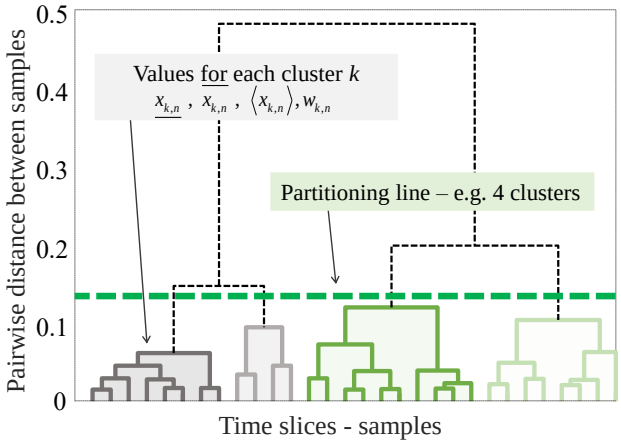

Figure 3: Hierarchical Clustering Outputs

Note that the time index of each sample $t \in T$ (i.e. $\{1, 2, \ldots, \|T\|\}$) is also considered in the hierarchical classification as an additional dimension in the original time series data set. Similar to the other data dimensions, this feature is entered as a normalized time series, divided by the number of elements in the temporal set *T*. The typical clustering aims at representing the occurrences of the different events (time samples) in the original data set. Considering the temporal dimension in the distance computation between the samples then allows to account for the concurrence between the events. In other words, this allows to decouple two similar samples (in terms of data values) that would occur in two "significantly" different periods of time. Indeed, in case of seasonal storage, similar values of load and generation may not lead to the same optimal storage usage (i.e. charge/discharge) in winter or summer where the stage charge levels may differ greatly.

4.2. Optimized Weighted Time Slices as a Least Square Problem

From the previous partitioning step, it would be possible to define a representative sample for all the time steps within a cluster along the different dimensions *N* (e.g. the mean value $\langle x_{k,n} \rangle$). However, similar to the supervised clustering introduced in the previous section, and in order to further improve the performances, the representative time slices shall be adjusted. The objective is to reach the best values $x^*_{k,n}$ that represent each original time step *t* and that allows to minimize fitting criteria compared to the original profiles $x_{t,n}$ once the approximated profiles $\tilde{x}_{t,n}$ are reconstructed. The selection of those representative time slices does not consist here in sorting the best values from the original set. Indeed, such an approach in the form of combinatorial optimization problem would require the use of metaheuristics (e.g. genetic algorithm as in **OPT₁** and **OPT₂**) or mixed integer problem formulations ([18]). Those strategies would suffer from a lack of scalability, especially when facing high-dimensional time series as it could be the case in power system modelling (i.e. as many profiles as there are buses in the electrical network). Thus, the best values $x^*_{k,n}$ that have to be determined are here optimization variables defined in the search space delimited by the lower/upper bounds in each cluster $\underline{x_{k,n}}$, $\overline{x_{k,n}}$, derived from the previous hierarchical partitioning. The time slice adjustment is finally formulated as a linearly constrained least square problem ((14)). The two criteria considered correspond to the error between the original and reconstructed profiles in terms of time series and duration curve. Especially a coefficient α [0,1] allows an arbitrage between those two normalized indexes [33].

$$obj: \min_{\underline{x_n^k} \le x_n^k \le \overline{x_n^k}} \alpha.\sigma_{RMS} + (\alpha - 1).\sigma_{DC} \tag{14}$$

The profiles $\tilde{x}_{t,n}$ are reconstructed based on the mapping of each time step t to a cluster k and thanks to a linear constraint with a matrix computed from the hierarchical partitioning ((15)). The first metric to minimize noted σ_{RMS} is homogeneous to a normalized root mean square error between the original and the reconstructed time series. A penalty coefficient is introduced here in order to enhance the fitting of the extreme time series values around the average ($\langle . \rangle$). The underlying idea is to better represent extreme events that can have great impact on the final performances for the considered case studies (e.g. penalty for peak consumption, marginal cost/emissions) This coefficient is then an optimization parameter computed from the original data set for every time step along all the dimensions ((16)).

$$\tilde{x}_{t,n} : \left\{ x_{k,n} \mid x_{t,n} \in \text{cluster } k, \forall t \in T \right\} \tag{15}$$

$$\sigma_{RMS} = \frac{1}{\|T\|.N} . \sum_{t \in T} \sum_{n=1}^{n=N} \left(\frac{x_{t,n} - \langle x_{t,n} \rangle}{\langle x_{t,n} \rangle} \right)^2 . \left(x_{t,n} - \tilde{x}_{t,n} \right)^2 \tag{16}$$

The second performance index is the error between the duration curves (reconstructed and original) over the different dimensions n. The duration curves of the original time series are computed from the input data set and entered as parameters in the least square problem. However, the duration curves for the reconstructed data set shall be computed in the course of the optimization, each time different values for the representative time slices $x_{k,n}$ are investigated – i.e. with the computation of the reconstructed profiles $\tilde{x}_{t,n}$ based on the optimization variables $x_{k,n}$. The calculation of the load duration curves normally incurs non-linearities with the need to sort the values of the reconstructed profiles $\tilde{x}_{t,n}$ in descending order. Typically, a hundred bins b^{DC} are defined along the data range and the load duration curve computation consists in counting the number of occurrences above each bin values. Non-linearities then occur due to both the ranking and enumeration processes and shall be leveraged similarly to the approaches in [18] and [32] . The optimization proposed here consists in a linearized formulation of the load curve for the reconstructed profiles based on the outputs from the hierarchical clustering (Figure 4). Especially, this load curve computation does not require to reconstruct the entire profiles and only considers the representative variables $x_{k,n}$ ranked in descending order and attached to a weight that represent their occurrence – i.e. the number of time samples in the corresponding cluster over the total number of samples. To do so, in a first step, the weighs $w_{k,n}$ off all the clusters returned by the partitioning are arranged in the order corresponding to the average clusters values $\langle x_{k,n} \rangle$ ranked from the biggest to the smallest – which could differ from one dimension to another. Also, a cumulative sum from the half value of the ranked weights allows to define the bin values $w_{k,n}^{DC}$ in the duration curves displayed in the graphical example of Figure 4 ((17)). Those bin values derived from the weights and are then used to identify $x_{k,n}^{DC}$, the corresponding value in the original duration curve for each dimension of the input multi variate time series ((18)). In the convex optimization problem, a first linear constraint ranks the variables $x_{k,n}$ according to the descending order of the average clusters values $\langle x_{k,n} \rangle$ (matrix-vector computation). Remind that in constraint (19), $x_{k,n}$ are optimization variables while the average values are parameters obtained from the hierarchical partitioning. A last constraint is implemented to ensure the order of those ranked variables denoted $x_{k,n}^{DC}$ ((20)). Indeed, this order may not be guaranteed in case of a lower bound $\underline{x_{k,n}}$ of a cluster k is higher than the upper bound $,\overline{x_{k',n}}$ of the previous cluster k' in the descending order. Ultimately, the error in the duration curve between the original and reconstructed profile is estimated over K samples and computed following (21). Figure 4 gives a graphical representation of the introduced problem that could be broadly considered as a "fitting" of the horizontal steps in the reconstructed duration curves relatively to the original ones.

$$w_{k,n}^{DC} : \left\{ \sum_{i=1}^{i=k} \frac{w_i}{2} \mid \langle x_{k,n} \rangle \geq \langle x_{i,n} \rangle, \forall \{i,k\} \in \{1,..,K\}^2 \right\} \tag{17}$$

$$x_{k,n}^{DC} : \left\{ b^{DC} = w_{k,n}^{DC}, \forall k \in \{1,..,K\} \right\} \tag{18}$$

$$\tilde{x}_{k,n}^{DC} : \left\{ x_{k,n} \mid \langle x_{k,n} \rangle \geq \langle x_{i,n} \rangle, \forall \{i,k\} \in \{1,..,K\}^2 \right\} \tag{19}$$

$$\tilde{x}_{k,n}^{DC} \geq \tilde{x}_{k-1,n}^{DC} \quad \forall k \in \{2,..,K\} \tag{20}$$

$$\sigma_{DC} = \frac{1}{K.N} \sum_{k=1}^{k=K} \sum_{n=1}^{n=N} \left(\frac{x_{k,n}^{DC} - \tilde{x}_{k,n}^{DC}}{\langle x_{t,n} \rangle} \right)^2 \quad (21)$$

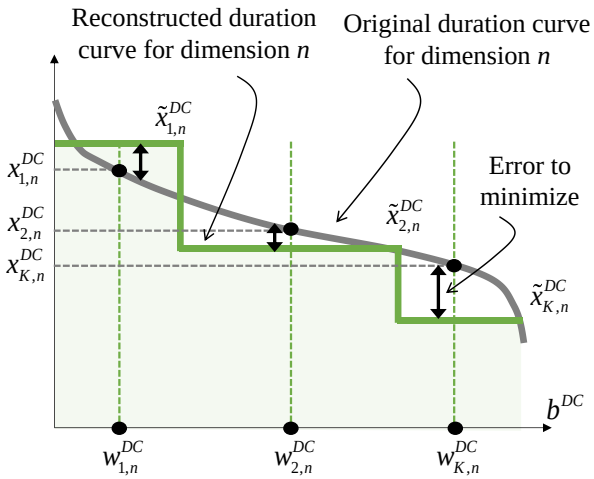

Figure 4: Duration curve error minimization along every dimension *n*

A set of preliminary validation tests is performed while fitting a weekly profile (at 1 h time step) for a univariate time series (office load profile here). The results displayed in Figure 5 show the reconstructed time series and duration curves with $K = 6$ clusters and α set to 0.5. The OWTS outputs are compared to the profiles reconstructed after the first step with the hierarchical clustering only. Especially, the extreme values are better approximated with the OWTS procedure while capturing higher peak powers (Figure 5*a*). The cumulated period with higher load values is obviously more accurate when computing the duration curve (Figure 5*b*). However, with six representative samples only, it is not possible to fully capture the small variations during the week end hours which results in a single value to represent around 50 % of the time steps in the duration curve – i.e. in Figure 5b, the duration curve of the reconstructed profiles displays a single value 50 % of the time corresponding to a single sample ,with 200 kW to represent the night hours and week-end.Increasing the number of representative time steps will obviously improve the fitting performances and the scalability of the procedure is ensured by the convexity of the problem as opposition to metaheuristics [18].

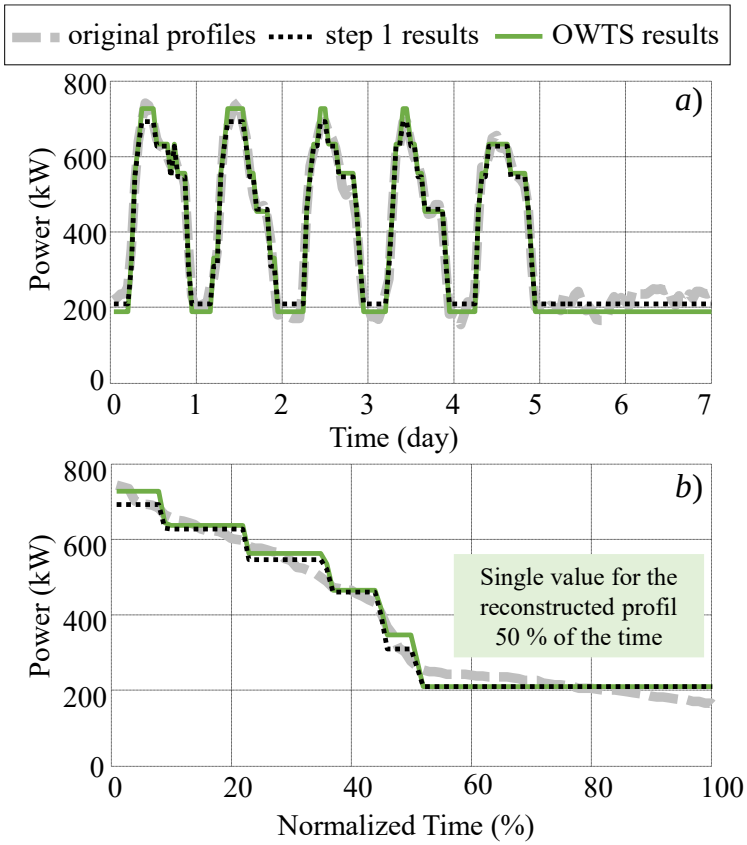

Figure 5: OWTS results compared to original profiles – *a)* original and reconstructed time series- *b)* original and reconstructed duration curves

4.3. Time Series Reconstruction and Problem Solving

Similar to the **OPT₂** approach with inter-periods representation, the system controls are optimized over the K representative sets only $x^*_{k,n}$ with the optimization variables denoted u_k. The objective function of the reconstructed problem accounts for the cluster weights extracted from the hierarchical clustering. As in **OPT₂**, the state of charge for the storage unit(s) is computed over the entire original time set while considering the mapping of each time step to a representative sample ((15)) – the corresponding controls are denoted $u_k(t)$ – and the storage returns to its initial charge at the end of the simulated period for energy conservation purposes. The problem is finally formulated following ((22)).

$$obj: \min_{\underline{u} \le u_k \le \overline{u}} \sum_{k=1}^{k=K} w_k . f(x^*_{k,n}, u_k)$$

$$\text{s.t.} \begin{cases} g(x^*_{k,n}, u_k) \le 0 & \forall k \in \{1,..,K\} \\ s_t = s_0 + \sum_{i=1}^{i=t} h(u_k(t)) & \forall t \in T \\ \underline{s} \le s_0 \le \overline{s} & \forall t \in T \\ s_0 = s_{\|T\|} \end{cases} \tag{22}$$

5. Obtained Results and Discussions

5.1. Generic Energy System

5.1.1. Considered Case Studye and Objectives

At first, a simple grid connected energy system (load, solar, storage) is considered with an energy bill minimization problem. As already discussed in Section 2.1, this objective consists in the cost for the purchase energy, the penalty for peak power excess and the cost of using the storage. The main constraints refer to the operational

limits for the storage (i.e. charge/discharge and state of charge) as well as the power balance that shall be ensured at every time step. Detailed equations are given in Appendix A.1. The case study consists in an office building (peak load at 1.2 MW) located in Grenoble, France, and associated with a 300 kWp solar generator. Actual load and generation data are given at one hour resolution for the year 2018 along with the corresponding real time prices of energy in France. For the reference scenario, the storage size is initially set to 300 kW / kWh. Details of the simulated case study are given in Table 2 with the subscribed power above which penalties for excess have to be paid.

Table 2: Case study details

System Load	Solar	Storage	Energy Price	Subscribed power
Peak load 1.2 MW	3 kWp	300 kW/kWh	Real-time	590 kW

5.1.2. Obtained Performances

A first set of simulations is performed with the references methods and the performance are assessed in terms of error on the total operational cost compared to the yearly reference. Especially, the results are analysed with the error mean values and confidence interval (at 95 %) after ten runs. For every run, the cost error is computed with regard to the energy bill, similar to the preliminary tests of Section 2.2 with (2). All the combinations of approaches to select the representative days (**CL_0**, **CL_1**, **CL_2**) and to reconstruct the operational problem (**OPT_0**,**OPT_1**, **OPT_2**) are tested for different cluster sizes (i.e. number of days) K:{5,10,15,20,25,30,40,50,60,90} – for a total of nine hundred simulations. The results in Figure 6 show the obvious tendency of improved performances when the number of representative days increases, no matter the selection process or the reconstruction strategy. Globally, for all the investigated methods, the results tend to be the best for 40 representative days or more according to the obtained results (mean errors around 5 %). The first reconstruction strategy (**OPT_0**) leads to larger confidence intervals for the different selection procedures considered. This confirms the need to integrate weighting factors in order to better represent the occurrences of the different values in the original time series. As already mentioned, the outputs of the unsupervised clustering **CL_0** strongly depends on the starting point. This also ultimately incurs slightly larger confidence intervals compared to the supervised approaches. Regarding the cost error, the best results for small numbers of representative days are obtained with the supervised clustering based on the load duration curves (**CL_2**). Associated to the reconstruction with the representation of the inter periods (**OPT_2**), this supervised clustering leads to error values close to 5 % for ten representative days or more.

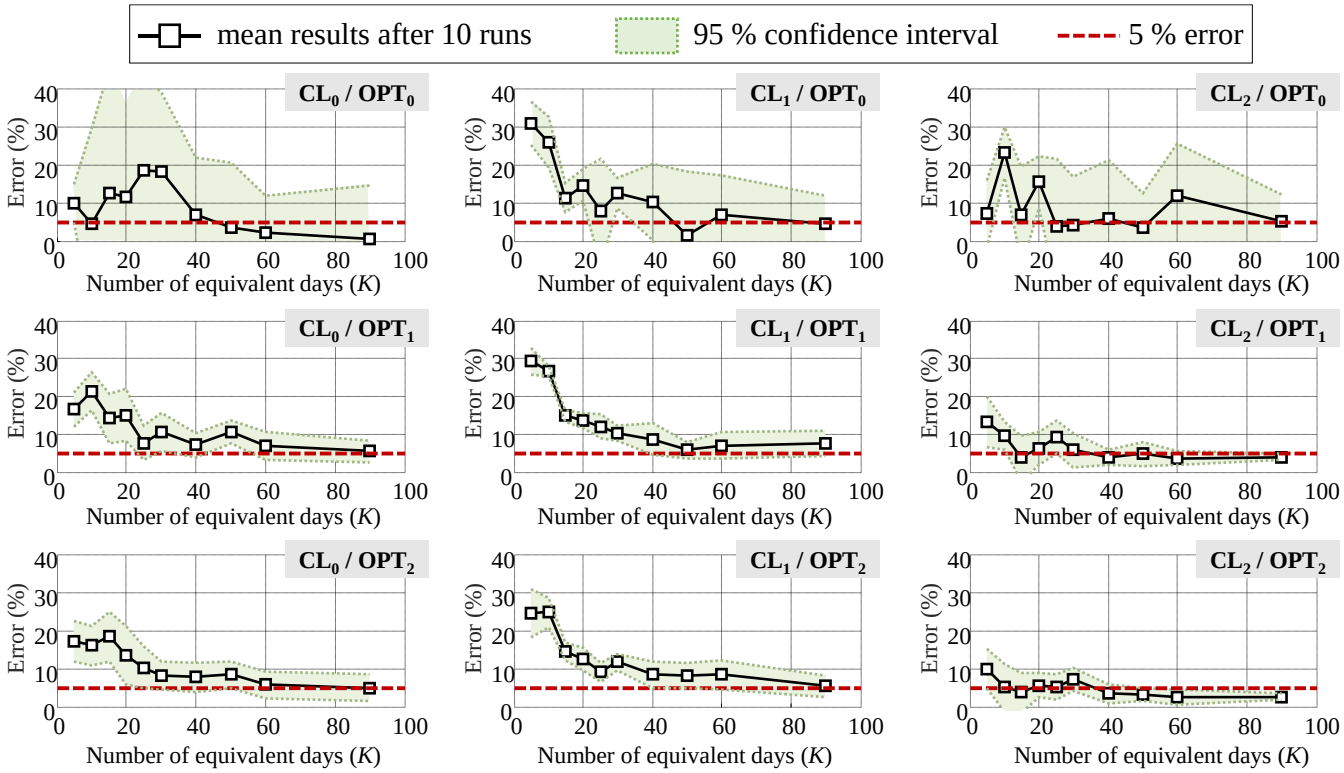

Figure 6: Results obtained with the reference methods for the profile selection (CL_i) and problem reconstruction (OPT_i), total cost error with the reference value for different numbers of representative days (K) and 5 % error level

The coupled methods **CL_2 / OPT_2** are further considered as a baseline to assess the performances of the proposed OWTS. The results of Figure 7 display the errors observed for all the cost components (i.e. total cost, energy, peak penalty and storage). Especially, the OWTS is run for different values of the coefficient α that allows an arbitrage between the error with the original time series profiles or with their duration curves. The results are plotted for different numbers of equivalent representative days (K). In practice, the OWTS is performed using $24\times K$ representative samples at 1 h resolution for the input time series in order to have a consistent comparison in terms of 'numbers of equivalent days' with the reference methods. The OWTS returns better results in terms of overall cost error for 20 equivalent representative days or more. The setting $\alpha = 1$ is systematically the best method (Figure 7*a*). As already observed in the preliminary results (Table 1), the deviations with the reference are small when dealing with energy aspect only, with errors below 3 % in most of the cases (Figure 7*b*). The errors for the peak penalty are in an order of magnitude ten times greater than the deviations on the energy expenditures. However, the proposed OWTS allows reduced errors, especially for 20-50 equivalent representative days compared to the mean results obtained after several runs of the supervised clustering **CL_2 / OPT_2** (Figure 7*c*). The storage cost estimated with the OWTS time reduction differs from the reference more than the supervised clustering with 35 representative days or less (Figure 7*d*). In addition to those improved overall performances in the total cost, the main advantage of the OWTS approach is that it systematically returns the same results with a heuristic hierarchical clustering associated to the solving of a convex problem for which a global solution can be found. Also this convexity allows a scalable method when high dimensions multi variate time series are investigated which may not be guaranteed due to the use of a metaheuristic (genetic algorithm) in the reference supervised clustering **CL_2**.

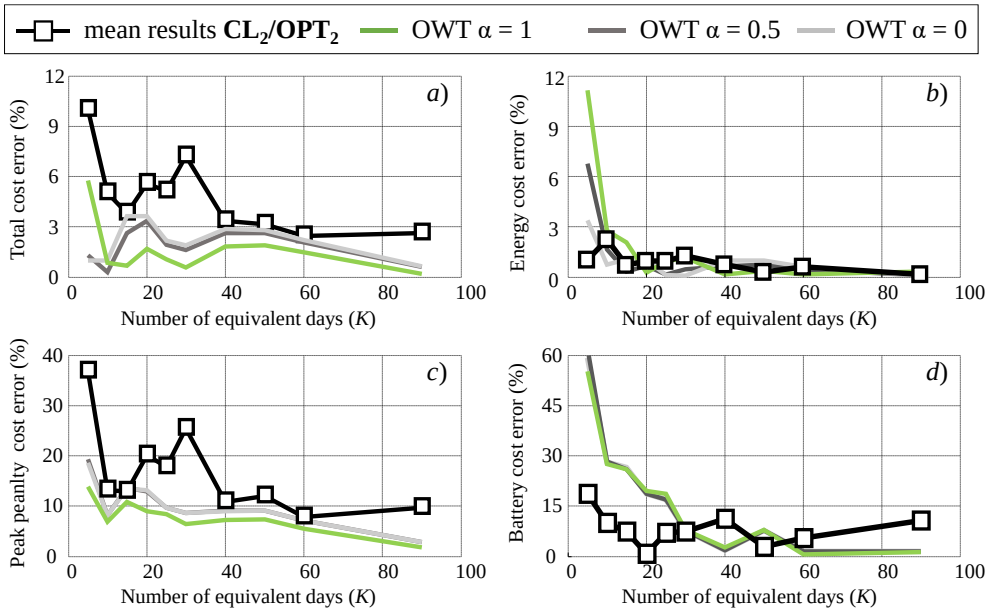

Figure 7: Comparison of the cost error obtained with the reference CL_2 / OPT_2 and the proposed OWTS – *a*) total cost error – *b*) energy cost error – *c*) peak penalty error – *d*) battery cost error

5.1.3. Impact of Case Study

In this subsection, the energy case study configuration is changed and the results of the introduced OWTS are still compared with the outputs form the reference **CL_2 / OPT_2**.(mean results after ten runs) At first, the capacity of the considered storage is changed with E^s (in kW / kWh) ={0,300,600,900,1200,1500}. The results of both time reduction methods are displayed in terms of total cost error with the reference computed over the whole year. Obviously, this reference cost tends to decrease with the size of the installed storage capacity due to the reduction of the peak penalty but it is not reflected in results displayed here. The objective is merely to compare the performances of the proposed OWTS with the benchmark **CL_2 / OPT_2** for different number of representative days (i.e. equivalent representative days for the OWTS). For both methods, the results are better with longer representative periods up to 90 days here, even if this trend is less obvious for the **CL_2 / OPT_2** (Figure 8*a*). The proposed OWTS is significantly better with errors systematically lower than 4 % and even for 20 equivalent days only (i.e. 24×20 time slices are deemed sufficient to account for seasonal storage contributions). Additional simulations are performed while changing the threshold "subscribed power" above which a penalty shall be paid in case of excess of power imported from the upstream grid. In the previous tests, this value was set based on the duration curve of the yearly load profile with power excesses occurring 5 % of the whole time horizon. This percentage of occurrence is here changed and the total bill is again computed with time reduction approaches and compared to the reference. The OWTS slightly outperforms the **CL_2 / OPT_2** and displays much better results with increased number of days (K = 90) (Figure 8*b*). However, the error tends to increase when more power excesses occur (up to 10 % in the reference load profile) but remain small for the investigated scenarios and below 2 % only for ninety equivalent days to represent the original time series. This would suggest that the fitting of the extreme values in the OWTS process (and reference methods as well) shall be further penalized in the objected functions of the considered optimization problems of Section 3 and 4. This would correspond to penalize more the deviations in the duration curves occurring at the lowest levels of normalized time (e.g. left part of the horizontal axis in Figure 5).

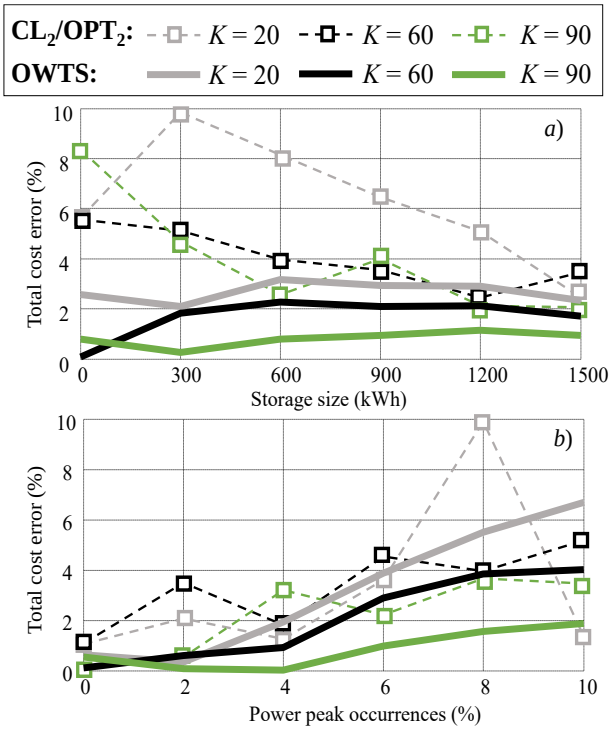

Figure 8: Total cost error for OWTS and CL_2/OPT_2 and different numbers of equivalent representative days – *a*) for different storage sizes – *b*) for different occurrences of peak power exceeds

5.1.4. Optimal Design

As previously mentioned, this paper proposes to estimate the performance of the investigated time reduction techniques with regard to the errors on the system operating expenditure comparatively to reference yearly simulations. The idea is to decompose the effect of the equipment sizing that can ultimately have a great impact in case of significant installation costs – the operation would not matter too much in the optimized criteria for the design (e.g. levelized cost of energy [12]). In this subsection, and similarly to most of the encountered studies ([18] [19]), the reduction techniques are tested when integrating them in an optimal design tool for the considered energy system. The objective is to optimize the Total Cost of Ownership (TCO) over the system lifetime with a typical arbitrage between the operating cost and the capital expenditures. This capital cost is computed with regards to the installation price for the equipment – values of 1500 €/kWp of installed solar and 1000 €/kWh for storage system are considered here. The management strategies, both for the yearly reference simulation and reduced profiles, are then integrated in a design methodology. At first, an exhaustive search is performed while varying the installed capacities in predefined ranges: 0 – 2000 kWp for the solar generator and 0 – 1000 kWh for the storage system. For each tested configuration, the TCO (in k€) is computed with the yearly simulation and the reduced OWTS profiles with 20 equivalent days. The operating cost over the system lifetime is then computed by multiplying the output of the management strategies by the number of estimated years (20 years here).

Obtained results are displayed on Figure 9*a* with similar overall shape of the TCO over the investigated search space with slightly smoother curves when the operation is simulated over a reduced horizon (Figure 9*b*). Typically, the TCO decreases when solar and storage capacities are installed, thanks to reduced operating costs. However, oversized solar and/or storage systems incur greater installation expenditures that are not compensated by the billing reduction and ultimately lead to increased TCO values. Thus, the design problem is characterized by a unique optimum. The best results returned in terms of TCO are close - less than 4 % error with 8278 k€ for reference yearly simulation and 8604 k€. However, and despite the curve shape similarities, the corresponding solutions differ when comparing the solar and storage capacities. In the scope of this paper, this difference can be explained by the approximation induced by the time reduction. Nonetheless, it is worth noticing that the optimized objective function displays a wide flat area around the optimal solution, and many different configurations return similar TCO values in the reference case (Figure 9*a*). Thus, integrating the optimal capacities obtained with reduced horizon (OWTS in

Figure 9*b*) in the reference yearly simulation leads to a TCO of 8327 k€ that remains very close to the value obtained for the reference sizing over the yearly sets (less than 1 % error).

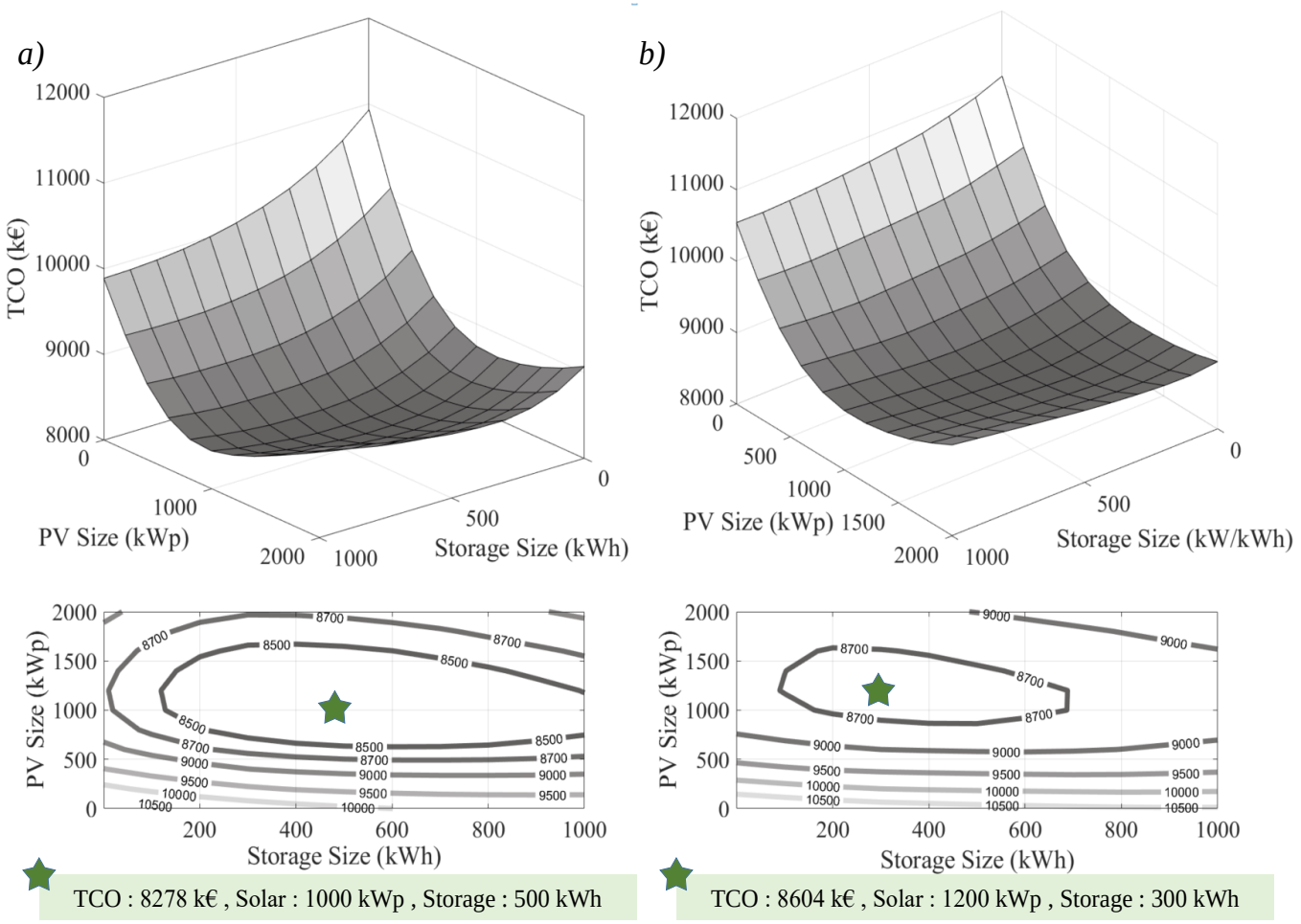

Figure 9: TCO versus installed capacities and best results - *a)* operation simulated over a full year – b) operation simulated over a reduced period generated with OWTS and 20 equivalent days.

In a second step, the management strategies (along the whole year and reduced periods) are integrated in a Particle Swarm Optimization algorithm [34] that generates different sizes to test in order to find the best system configuration – i.e. for the minimum TCO. The objective is to compare the optimum values obtained with the reference yearly operation (denoted as Ref. in Table 3) and the ones returned with reduced simulated periods for both OWTS and CL_2/OPT_2 methods. The obtained results are displayed in Table 3. The TCO values for the reduced simulations are very close to the reference one which implies good performances of both reduction techniques with the different numbers of equivalent days considered. The values in terms of installed capacities are closer to the reference when using the purposed OWTS for 60 and 90 equivalent days. However, and as observed previously, the capacities values returned with reduced time horizons globally differ from the optimal ones obtained with the reference yearly simulation. The last row of Table 3 estimates the TCO with yearly simulations for the capacities returned by the optimization on shortened profiles. The errors are below 1 % compared to the reference TCO value, and the OWTS returns identical results for 60 and 90 equivalent days. As mentioned above, this is mainly explained by the overall shape of the optimized problem with wide area displaying similar TCO values [11]. This is due to the nature of the design problem which ultimately complexifies the performances assessment of the different reduction methods. That effect, motivated the approach in this paper with a comparison of the methods regarding the system operation performances only, in order to decompose the impact of the sizing.

Table 3: Optimal system designs with yearly and reduced profiles for the system operation

	Ref.	CL_2/OPT_2 K =20	CL_2/OPT_2 K =60	CL_2/OPT_2 K =90	OWTS K =20	OWTS K =60	OWTS K =90
TCO (k€)	8276	7923	8151	8030	8599	8459	8362
Solar (kWp)	1019	1092	1124	976	1163	1058	1047
Storage (kWh)	494	603	403	310	340	535	505
TCO (k€) yearly profile	N.A.	8286	8288	8321	8308	**8276**	**8276**

5.2. Scalability on a Generic Power System

5.2.1. Considered Case Study and Objectives

One originality of the paper is that the reduction strategies are tested on both energy and power system. Thus, in this subsection, the management of a distribution network is considered as a case study - a generic IEEE 33-buses network with a storage unit connected to bus 10. Similar to the previous case study, the storage is operated to minimize the electrical bill computed at the point of common coupling with the upstream grid and with the consideration of the battery degradation costs. In order to further increase the complexity of the cost function, penalties are introduced if the buses voltage is under 0.95 p.u. or above 1.05 p.u. The detailed equations can be found in Appendix A.2 B, the main idea being the linearization of the load flow equations ([35]) in order to allow yearly simulations that return the reference values furtherly used for the performances assessment of the time reduction techniques. The objective of this case study is to test the scalability of the reductions strategies. The input data set that has to be reduced include as many yearly load profiles as there are network nodes – i.e. N = 67 with both active and reactive power profiles for each bus in addition to the real time prices of electricity.

5.2.2. Obtained Performances

Figure 10 displays the results obtained with the different reference methods (mean results after 10 independent runs) as well for the proposed OWTS. The $\mathbf{CL_0}$ / $\mathbf{OPT_1}$ approach corresponds to the most widely encountered strategy with weighted K representative days selected thanks to an unsupervised K-medoid. The observed errors in term of overall cost (Figure 10*a*) remain very high above 25 % even with 90 representative days. As previously observed, supervised clustering strategies (i.e. $\mathbf{CL_1}$ and $\mathbf{CL_2}$) allow better performance with deviations below 10 % for 50 representative days or more. The proposed OWTS displays the best results with deviations below 5 % for 20 equivalent representative days or more. As for the previous energy case study, these performances are mainly explained with the cost component related to the peak penalty that is much better approximated with the OWTS (Figure 10*b*) while the penalty for the voltage deviations are well estimated even with the supervised clustering strategies (Figure 10*c*). In addition to the higher accuracy observed, one main advantage of the proposed OWTS is its replicability (same results over successive runs) and its computational time (< 30 s) shorter than the calculation of supervised clustering methods $\mathbf{CL_1}$ and $\mathbf{CL_2}$ (> 5min for a single run). The performances of the supervised clustering approaches may be improved (in terms of clustering/fitting criteria) while increasing the numbers of generations or the population size in the considered genetic algorithm. It is not investigated here as it would ultimately incur longer computational times, and a single run of the supervised clustering methods based on genetic algorithm already require more computational time than the proposed OWTS. See also in Section 5.3.2, the best fitting criteria in the reductions processes do not necessarily guarantee the smallest errors with the reference yearly simulation.

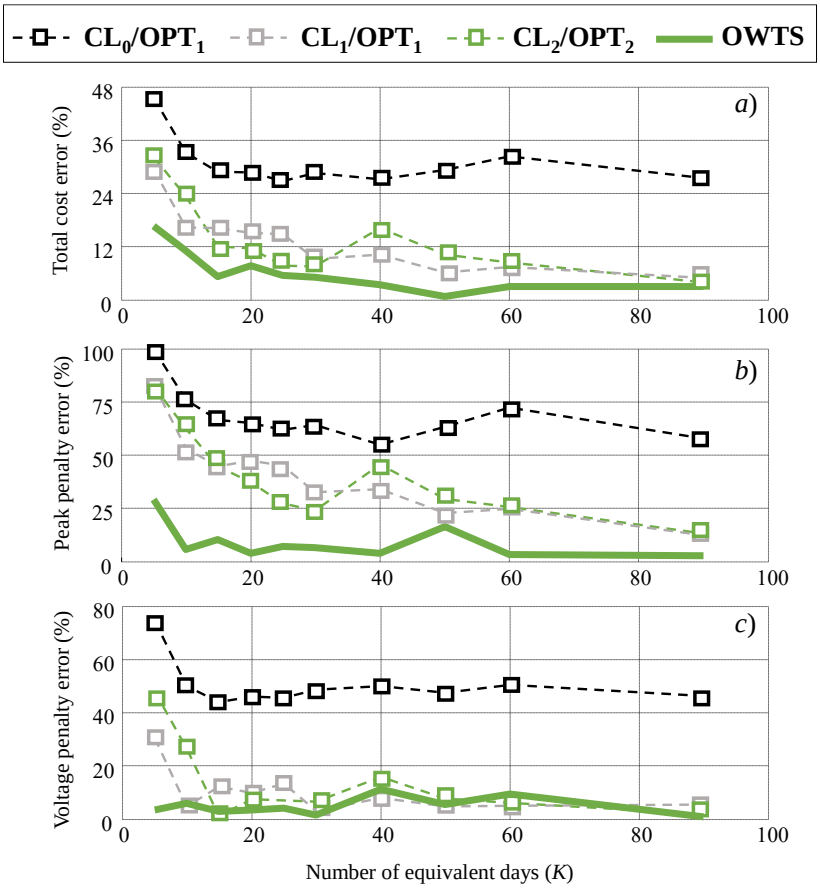

Figure 10: Cost errors for reference methods and proposed OWTS and different numbers of equivalent representative days – *a*) total cost error – *b*) peak penalty error – *c*) voltage deviations error

5.3. Discussion

5.3.1. Complexity Reduction

As already mentioned, the main obvious objective for the use of time horizon reduction techniques is to decrease the optimization problem complexity with a smaller search space. In order to better understand that reduction magnitude, Table 4 displays the amount of variables and constraints for different numbers of representative days K and depending on the reconstruction methods (equivalent representative days for OWTS). The two first reconstruction methods **OPT$_0$** and **OPT$_1$** display the same problem size, the only difference being that the contribution of each representative day is weighted in the second approach. However, the method **OPT$_2$** with inter periods representation and the proposed OWTS requires more variables (and thus constraints) as they need to represent the state of charge over the whole simulation horizon. However, if the number of storage units remains small compared to the system size (only one unit considered here), the difference in terms of complexity with the methods **OPT$_0$** and **OPT$_1$** is small with much better performances obtained as shown in Section 5.1.2. Obviously, the power system case study corresponds to much higher numbers of variables/constraints due to the introduction of variables for every node and branch as detailed in Appendix A.2.

Table 4: Number of variables (var.) and constraints (cst.) of the reconstructed problems for the different investigated methods and case studies compared with the yearly simulations

	Energy System		Power System 33-buses	
	$K = 20$	$K = 90$	$K = 20$	$K = 90$
OPT_0	3k var. 7k cst.	15k var. 34k cst.	144k var. 288k cst.	645k var. 1,298k cst.
OPT_1	3k var. 7k cst.	15k var. 34k cst.	144k var. 288k cst.	645k var. 1,298k cst.
OPT_2	12k var. 33k cst.	22k var. 53k cst.	152k var. 314k cst.	652k var. 1,319k cst.
OWTS	12k var. 33k cst.	22k var. 54k cst.	152k var. 297k cst.	652k var. 1,246k cst.
Year	61k var. 140k cst.		2,619k var. 4,949k cst	

5.3.2. Time Series Fitting Performances

As discussed along the paper, a recurrent objective targeted in the investigated approaches for time reduction is the minimization of the error in terms of duration curves compared with the input multi variate time series. It is especially the case for the supervised clustering **CL_2** and the proposed OWTS while the partitioning **CL_1** aims at maximizing clustering performances indexes (Section 3.1.2). Figure 11 displays the total cost error (with the yearly simulation) versus the error in the duration curves fitting obtained with different methods and numbers of equivalent representative days for the energy case study. Those results are issued from the performances assessment of Section 5.1.2 obtained after 10 runs of the supervised clustering strategies (**CL_1** and **CL_2**). There is an overall tendency of better accuracy of the time reduction when the duration curves fitting error decreases but it is not systematic as could have been expected. As already observed, the errors tend to reduce for greater numbers of representative days (> 20 in the displayed results). The supervised clustering **CL_2** that targets minimum duration curves error returns slightly smaller errors than the clustering **CL_1** with partitioning indexes maximisation. The most noticeable results is that the OWTS displays the smallest deviations by far, even for 20 representative days only, and the obtained points are close to the bottom left corner of the chart – i.e. "small fitting error and small cost error" (zoom in Figure 11).

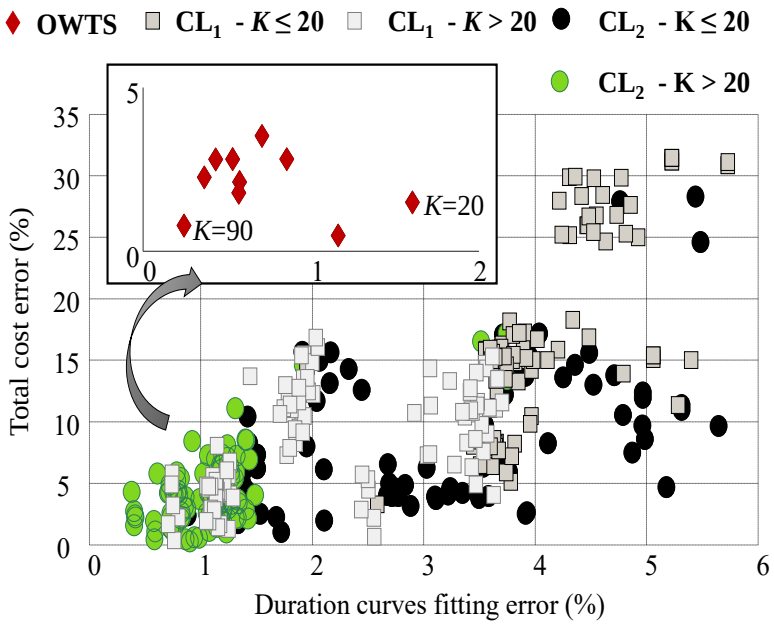

Figure 11: Total cost errors Vs the error in the fitting of the duration curves for the energy system case study

6. Conclusions

This paper proposes a method denoted OWTS for time horizon reduction applicable to power and energy system studies with storage units. The main objective is to avoid prohibitive computational times when dealing with the design of the systems over long periods of time. The proposed approach is compared to a wide range of conventional strategies based on clustering methods for the selection of representative days. The performances are assessed when comparing the results with reference yearly simulation. Before implementing the different methods, the paper provides a generic formulation for operation problems applied to power and energy systems with storage. Especially, a first set of simulations points out the impact of the operational objective function considered. Indeed, non linearities (e.g. power excess penalty) are more complex to capture compared to conventional energy costs (i.e. linearity with generated power) in time reduction techniques. Also, one originality of the paper is to assess the performances of the reduction methods regarding the operational results only and while decomposing the effects of the system design (discussions in Section 5.1.4).

The proposed OWTS method proceeds on the optimization of individual time slices, rather than on the selection of extended representative periods (typically days). In addition to faster computational times (less than 30 seconds to reduce 67 dimensions input data sets) and better scalability, the approach systematically returns the same results thanks to the convex nature of the modelled problem. For the energy hub case study, the error on the overall bill compared to the reference yearly simulation is below 3 % even for small numbers of representative samples. For both power and energy scenarios, the OWTS better captures the contribution of the non-linear terms in the objective function – less than 10 % error for the peak penalty in the energy cast study.

Despite those promising results, the scalability shall be further investigated with greater dimensions in the input time series or increased complexity in the encountered case studies. Further works may consider emerging advanced techniques derived from artificial intelligence for time series classification [36] while comparing them with the reference methods and proposed OWTS with an emphasis on scalability issues. Lastly, investigations shall be led on other distance computations to discriminate the samples, which could impact the output of the hierarchical clustering phase and the overall performances.

7. References

[1] R. Jordehi, “Allocation of distributed generation units in electric power systems: a review”, Renew. Sustain. Energy Rev., vol. 56, pp. 893-905, Apr. 2016.

[2] G. Zhang, D. Yang, G. Galanis, E. Androulakis, “Solar forecasting with hourly updated numerical weather prediction“, Renew. Sustain. Energy Rev., (in press), 2021,

[3] C. Ghenai, T. Salameh, A. Meraet, “Technico-economic analysis of off grid solar PV/ Fuel cell energy system for residential communityin desert region”, Int. J. Hydrog. Energy, vol. 45, pp. 11460-11470, 2020.

[4] P. Gabrielli, M. Gazzani, E. Martelli, M. Mazzottia, “Optimal design of multi-energy systems with seasonal storage”, App. Energy, vol. 219, pp. 408-424, 2018.

[5] M. Qadrdan, M. Cheng, J. Wu, N. Jenkins, “Benefits of demand-side response in combined gas and electricity networks“, App. Energy, vol. 192, pp. 360-369, 2017.

[6] T. Schütz, X. Hu, M. Fuchs, D. Müller, “Optimal design of decentralized energy conversion systems for smart microgrids using decomposition methods”, Energy, vol. 156, pp. 250-263, 2018.

[7] P. Vilaça Gomes, J. Tomé Saraiva, “A two-stage strategy for security-constrained AC dynamic transmission expansion planning”, Electr. Power Syst. Res., vol. 180, 2020.

[8] J. Sachs, O. Sawodny, “Multi-objective three stage design optimization for island microgrids”, App. Energy, vol. 165, pp. 789-800, 2016.

[9] C.F. Heuberger, I. Staffell, N. Shah, N. Mac Dowell, “A systems approach to quantifying the value of power generation and energy storage technologies in future electricity networks”, Comput Chem Eng, vol. 107, pp 247-256, 2017.

[10] I. Blanco, J. M. Morales, “An Efficient Robust Solution to the Two-StageStochastic Unit Commitment Problem”, IEEE Trans. Power Syst., vol. 32, no. 6, pp. 4477-4488, 2017.

[11] R. Rigo-Mariani, S.O. Chea Wae, S. Mazzoni, A. Romagnoli, “Comparison of optimization frameworks for the design of a multi-energy microgrid”, App. Energy, vol. 257, 2020.

[12] S. Pfenninger, “Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability”, App. Energy, vol. 197, pp. 1-13, 2017.

[13] H. Teichgraeber, A. R. Brandt, “Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison”, App. Energy, vol. 239, pp. 1283-1293, 2019.

[14] P. Nahmmacher, E. Schmid, L. Hirth, B. Knopf, “Carpe diem: A novel approach to select representative days for long-term power system modelling”, Energy, vol. 112, pp. 430-442, 2016.

[15] F. Domínguez-Muñoz, J. M. Cejudo-López, A. Carrillo-Andrés, M. Gallardo-Salazar, “Selection of typical demand days for CHP optimization”, Energy Build. ,vol. 43, no. 11, pp. 3036-3043, 2011.

[16] C. Elsido, A. Bischi, P. Silv, E. Martelli, “Two-stage MINLP algorithm for the optimal synthesis and design of networks of CHP units”, Energy, vol. 121, pp. 403-426, 2017.

[17] K. Fahy, M. Stadler, Z. K. Pecenak, J. Kleissl, “Input data reduction for microgrid sizing and energy cost modeling: Representative days and demand charges”, J. Renew. Sustain. Energy, vol. 11, 2019.

[18] K. Poncelet, H. Hoschle, E. Delarue, A. Virag, W. D’haeseleer, “Selecting Representative Days for Capturing the Implications of Integrating Intermittent Renewables in Generation Expansion Planning Problems”, IEEE Trans. Power Syst., vol. 32, no. 3, pp. 1936-1948, 2017.

[19] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, “Time series aggregation for energy system design: Modeling seasonal storage”, App. Energy, vol. 213, pp. 123-135, 2018.

[20] M. Welsch, P. Deane, M. Howells, B. Ó Gallachóir, F. Rogan, M. Bazilian, H.H. Rogner, “Incorporating flexibility requirements into long-term energy system models – A case study on high levels of renewable electricity penetration in Ireland”, App. Energy, vol. 135, pp.600-615, 2014.

[21] L. Guo , W. Liu, J. Cai, B. Hong, C. Wanga , “A two-stage optimal planning and design method for combined cooling, heat and power microgrid system”, Energy Convers. Manag., vol. 74, pp. 433-445, 2013.

[22] T. Schütz, M. H. Schraven, M. Fuchs, P. Remmen, D. Müller, “Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis“, Renew. Energy, vol. 128, 2018.

[23] R. Loulou, M. Labriet, “ ETSAP-TIAM: the TIMES integrated assessment model Part I: Model structure” Comput Manage Sci., vol. 5, no. 1, pp. 6-40, 2008.

[24] Y. Ding, Q. Xu, Y. Xia, J. Zhao, X. Yuan, J. Yin, “Optimal dispatching strategy for user-side integrated energy system considering multiservice of energy storage”, Int. J. Electr. Power Energy Syst., vol. 129, 2021.

[25] R. Rigo-Mariani, S.O.C. Wae, S. Mazzoni, “Impact of the Economic Environment Modelling for the Optimal Design of a Multi-Energy Microgrid”, 46th Annual Conference of the IEEE Industrial Electronics Society (IES), IECON, Singapore (Online), 2020.

[26] T. Teeraratkul_, D. O’Neill, S. Lall, “Shape-Based Approach to Household Load Curve Clustering and Prediction”, IEEE Trans. Smart Grid, vol. 9, no. 5, 2018.

[27] Y. Jin, Z. Bi, “Power Load Curve Clustering Algorithm Using Fast Dynamic Time Warping and Affinity Propagation”, 5th International Conference on Systems and Informatics (ICSAI), 2018.

[28] C. Sanchez E, Hines P, Barrows, Blumsack S, Patel M. “Multi-attribute partitioning of power networks based on electrical distance”, IEEE Trans Power Syst., vol. 28, no. 4, pp. 4979–4987, 2014.

[29] R. Rigo-Mariani, K. V. Ling, J. Maciejowski, “A Unit Commitment with Loss Linearization and Zonal Decomposition for Distributed Generation Planning”, Int. J. Electr. Power Energy Syst., Vol. 113, pp. 9-22, 2019.

[30] B. Bahl, A. Kuempel, M. Lampe, A. Bardow, “Time-series aggregation for synthesis of distributed energy supply systems by bounding error in operational expenditure. In: Kravanja Z, editor. 26th European symposium on computer aided process engineering. Elsevier B.V; 2016.

[31] J.H Merrick, “On representation of temporal variability in electricity capacity planning models”, Energy Econ vol. 59, pp. 261-274, 2016.

[32] W. Ko, J.K. Park, M.K. Kim, J.H. Heo, “A Multi-Energy System Expansion Planning Method Using a Linearized Load-Energy Curve: A Case Study in South Korea”. Energies, vol. 10, 2017,

[33] H. Mausser, “Normalization and Other Topics in MultiObjective Optimization”, Proceedings of the Fields–MITACS Industrial Problems Workshop, Toronto, Canada, August, 2006.

[34] J. Kennedy, R. Eberhart. “Particle swarm optimization”, IEEE Proceedings of the international conference on neural networks. 1995.

[35] Z. Wang, B. Chen, J. Wang, J. Kim, “Decentralized Energy Management System for Networked Microgrids in Grid-Connected and Islanded Modes”, IEEE Trans. Smart Grid, vol. 7, no. 2, pp. 1097-1105, Mar. 2016.

[36] Y. Lei, Z. Wu, “Time series classification based on statistical feature”, EURASIP Journal on Wireless Communications and Networking, vol. 46, 2020.

[37] J. Lofberg, "YALMIP: a toolbox for modelling and optimization in MATLAB", IEEE International Symposium on Computer Aided Control Systems Design, New-Orleans, USA,, pp. 282-289, Sep. 2004

A. Appendix

A.1 Energy System Optimization

This subsection describes the main model equations for the optimal operation of the simple energy system considered in this paper. The input multi variate time series consist in the profiles for the load P_t^l, the solar generation P_t^{pv} and energy prices π_t^e for the import from the upstream grid - disaggregated in two power flows P_t^{gd+} and P_t^{gd-} in order to only account for the positive component. The degrees of freedom are the charge P_t^{st+} and discharge P_t^{st-} power flows for the storage at every time step. In addition to the operational bounds $\overline{P^{st}}$, this storage is represented by a capacity E^{st} (in kWh) and an efficiency η^{st} in the estimation of the state of charge (s^t in %). The cost of using the storage is computed with the installation cost C^s (in €/kWh) and the maximum number of cycles N^c. This estimation of the storage degradation requires to compute the total power exchanged by the storage with the system while integrating both charge and discharge power along the simulated time horizon. The penalty for peak power is paid every time P_t^{gd+} exceeds a given value P^m. To mitigate this nonlinearity, and allow fast computation for yearly simulations, the grid import power is here disaggregated in two flows $P_t^{gd<P^m}$ and $P_t^{gd>P^m}$. In the objective function, any excess (i.e. $P_t^{gd>P^m} > 0$)will incur a cost computed with the instantaneous energy price (multiply by 20 here). Ultimately, the whole operational problem is formulated following (Aeq.1) for a given time step Δt (1 h in the simulations presented in the paper).

$$obj: \min \sum_{t\in T}\left(\pi_t^e.P_t^{gd<P^m} + 20.\pi_t^e.P_t^{gd>P^m} + \frac{C^s.E^s}{2.N^c}.\left(P_t^{st+}+P_t^{st-}\right)\right).\Delta t$$

$$\text{s.t.}\begin{cases} 0 \le P_t^{st+}, P_t^{st-} \le \overline{P^{st}} \\ 0 \le P_t^{gd<P^m}, P_t^{gd>P^m} \\ 0 \le P_t^{gd<P^m} \le P^m \\ P_t^{gd+} = P_t^{gd<P^m} + P_t^{gd>P^m} \\ P_t^{gd+} + P_t^{pv} + P_t^{st-} = P_t^l + P_t^{gd-} + P_t^{st+} \quad \forall t \in T \\ s_t = s_0 + \sum_{i=1}^{i=t}\left(\frac{\eta^s.P_t^{st+} - P_t^{st-}/\eta^s}{E^s}\right).100.\Delta t \\ \underline{s} \le s_t \le \overline{s} \\ s_{\|T\|} = s_0 \end{cases} \quad \text{(Aeq.1)}$$

A.2 Power System Optimization

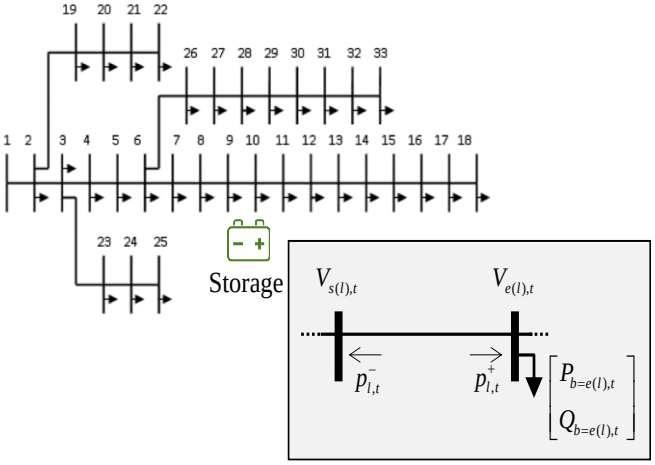

Figure 12: IEEE 33-buses system and DistFlow representation

The power system case study is the IEEE 33-buses network with a storage unit is connected at node 10. Similar to the previous energy system, the objective is to minimize the bill which is computed with the active power at the point of common coupling P_t^{pcc} and while considering the set of line L and buses B that represent the network. Part of the objective function still consists in an energy part plus the penalties for power excess and the cost of cycling the storage. The problem formulation lies on the DistFlow equations for radial networks Especially, their linearized version [35] is considered to allow a computation over an entire simulated year. The DistFlow equations consist in an iterative computation of the line active/reactive flows $p_{l,t}$ and $q_{l,t}$ from the slack bus down to the termination buses of the considered radial networks. Especially, the line flows are disaggregated along their positive contributions ($p^+_{l,t}$, $p^-_{l,t}$ for the active power), as the billing considers the positive component only at the point of common coupling. In practice, disaggregating the positive and negative components of the line flow can also allow to compute the losses, indiscriminately from the flow direction in case of linear formulation, similar to [29] - e.g. piece wise linearization to approximate the square of the active and reactive power (both negative and positive contribution). It is not considered here. The main idea of the DistFlow is that every branch should supply or absorb the active $P_{b,t}$ and reactive $Q_{b,t}$ power balance at all its downstream buses (set $D_{b(l)}$). This balance is equal to the bus load profiles $P^l_{b,t}$ and $Q^l_{b,t}$that represent the input time serie to reduce. For the bus {10}, to which the storage unit is connected, the active power balance is modified in order to account for the charge P_t^{st+} and discharge P_t^{st-} profiles that are optimized in the considered bill reduction problem (i.e. system degrees of freedom). All the lines are assigned to a start bus ($s(l)$) and end bus ($e(l)$) that allow to compute the voltage drop along the branch (Figure 12). The voltages $v_{b,t}$ are represented by a variable $V_{b,t}$ for their square values. Also, a cost component is introduced in the objective function in order to penalize voltage deviations over the limits ($V_{b,t}$ below 0.95^2 p.u. or above 1.05^2 p.u.). This requires the introduction of two variables to represent those positive ($V^+_{b,t}$) and negative ($V^-_{b,t}$) deviations. The voltage at the slack bus (b=1) is set at 1 p.u. and the power at the point of common coupling is taken equal to the positive flow (i.e. import) at the immediate downstream line (l = 1). The overall problem formulation is summarized in (Aeq.2) and can be typically considered in a resource allocation problem targeting system efficiency or maximum renewable integration. Figure 13 displays the yearly results with the power imported at the point of common coupling (Figure 13*a*), the voltage profiles at bus 18 (Figure 13*b*) with and without storage unit. The storage operations whose state of charge is plotted on Figure 13*c* allows to reduce the under voltage while bringing the values above 0.95 p.u. when possible. Also, the peak power for the imported power is significantly reduces with a threshold power P^m set at 6 MW.

$$obj : \min \sum_{t\in T}\left(\pi_t^e.P_t^{pcc<P^m} + 20.\pi_t^e.P_t^{pcc>P^m} + \frac{C^s.E^s}{2.N^c}.\left(P_t^{st+} + P_t^{st-}\right) + \sum_{l\in L}\frac{\left(V_{e(l),t}^{+} + V_{e(l),t}^{-}\right)}{r_l}.\left\langle \pi_t^e \right\rangle \right).\Delta t$$

$$\text{s.t. } V_{b,t}^0 \begin{cases} 0 \le P_t^{st+}, P_t^{st-} \le \overline{P^{st}} \\ 0 \le P_t^{pcc<P^m}, P_t^{pcc>P^m} \\ 0 \le P_t^{pcc<P^m} \le P^m \\ P_t^{pcc} = P_t^{pcc<P^m} + P_t^{pcc>P^m} \\ 0 \le p_{l,t}^{+}, p_{l,t}^{-}, q_{l,t}^{-}, q_{l,t}^{-} \le \overline{S} \quad \forall l \in L \\ p_{l,t} = \sum_{b\in D_b(l)\subset B} P_{b,t} \quad , \quad q_{l,t} = \sum_{b\in D_b(l)\subset B} Q_{b,t} \quad \forall l \in L \\ P_{b,t} = P_{b,t}^{l} \quad , \; Q_{b,t} = Q_{b,t}^{l} \quad \forall b \in B - \{10\} \\ P_{b,t} = P_{b,t}^{l} + P_{b,t}^{st+} - P_{b,t}^{st-}, \; Q_{b,t} = Q_{b,t}^{l} \quad \forall b \in \{10\} \\ V_{e(l),t} = V_{s(l),t} - 2.\left(r_l.p_{l,t} + x_l.q_{l,t}\right) \quad \forall l \in L \\ V_{b,t} = V_{b,t}^0 + V_{b,t}^{+} - V_{b,t}^{-} \quad \forall b \in B \\ 0.95^2 \le V_{b,t}^0 \le 0.95^2 \; , \; 0 \le \; V_{b,t}^{+}, V_{b,t}^{-} \quad \forall b \in B \\ V_{b=1,t} = 1 \text{ p.u.} \\ P_t^{pcc} = p_{l=1,t}^{+} \\ s_t = s_0 + \sum_{i=1}^{i=t}\left(\frac{\eta^s.P_t^{st+} - P_t^{st-}/\eta^s}{E^s}\right).100.\Delta t \\ \underline{s} \le s_t \le \overline{s} \\ s_{\|T\|} = s_0 \end{cases} \quad \forall t \in T \qquad \text{(Aeq.2)}$$

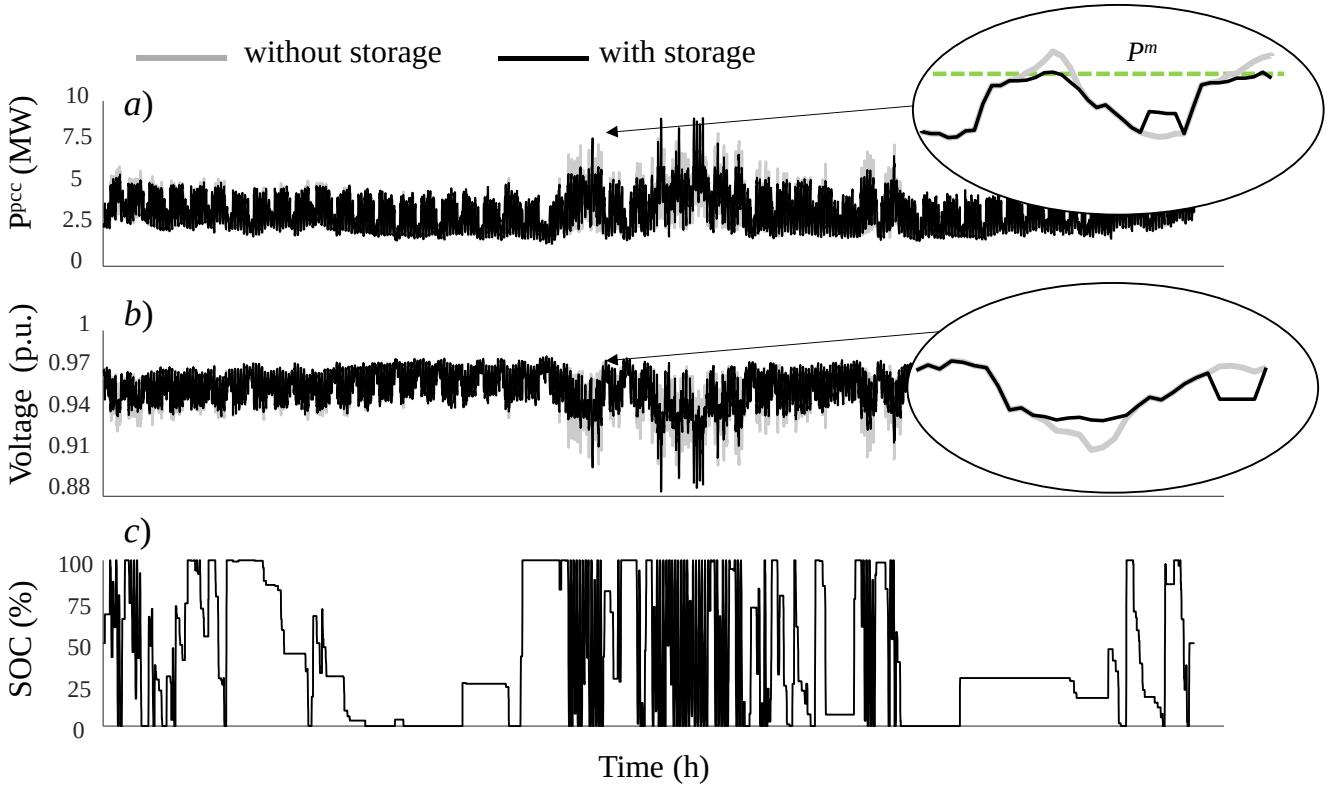

Figure 13: Reference results for a yearly simulation – *a*) power imported from the upstream grid , *b*) voltage profile at bus 18, *c*) storage state of charge

A.3 Simulation/Hardware Specifications

All the implemented problems are written using YALMIP toolkit [37] in MATLAB 2018b and solved with CPLEX 12.10.0. on a 4 cores i5-825OU processor with 8 Go of RAM and up to 8 threads in parallel. In order to reduce the computational time required to write and solve the problems, all the considered equations (linear constraints here) make use of matrix formulation.