Shortening Quasi-Static Time-Series Simulations for Cost-Benefit Analysis of Low Voltage Network Operation with Photovoltaic Feed-In

Claudio David López, Basem Idlbi, Thomas Stetz and Martin Braun
Fraunhofer Institute for Wind Energy and Energy System Technology
Kassel, Germany
Email: claudio.lopez@iwes.fraunhofer.de

Abstract—**Executing quasi-static time-series simulations is time consuming, especially when yearly simulations are required, for example, for cost-benefit analyses of grid operation strategies. Often only aggregated simulations outputs are relevant to grid planners for assessing grid operation costs. Among them are total network losses and power exchange through MV/LV substation transformers. In this context it can be beneficial to explore alternatives to running quasi-static time-series simulations with complete input data that can produce the results of interest with high accuracy but in less time. This paper explores two methods for shortening quasi-static time-series simulations through reducing the amount of input data and thus the required number of power flow calculations; one is based on downsampling and the other on vector quantization. The results show that execution time reductions and sufficiently accurate results can be obtained with both methods, but vector quantization requires considerably less data to produce the same level of accuracy as downsampling. In particular, when the simulations consider voltage control or when more than one simulation with the same input data is required, vector quantization delivers a far superior trade-off between data reduction, time savings, and accuracy. However, the method does not reproduce peak values in the results accurately. This makes it less precise, for example, for detecting voltage violations.**

Keywords—*Power flow calculation, quasi-static time-series, vector quantization, PV generation.*

I. INTRODUCTION

Distributed generation poses diverse challenges to distribution network operation and planning, such as increased voltage violations as well as branch and transformer overloading [1]. Overcoming these challenges requires assessing the effects distributed energy resources have on power grids and evaluating possible mitigation strategies, like voltage control or grid reinforcement, from a cost-benefit perspective.

In the presence of distributed generation, traditional grid analysis methods based on power flow simulations of extreme generation and load scenarios (e.g., maximum demand with minimum generation and vice versa) do not provide enough insight into grid behavior [2]. One of the recommended methods for quasi-static analysis of grids with distributed generation are quasi-static time-series simulations [3]. These simulations consist of consecutive power flow calculations using discrete load and generation profiles over a time horizon that in the case of cost-benefit analyses is typically in the order of a year.

Data at high time resolution is preferred due to accuracy concerns, which requires profiles with large numbers of time

steps, and consequently, time-series simulation that consist of a large number of power flow calculations. As reference, a yearly simulation at one minute resolution requires at least 525600 individual power flow calculations. Cost-benefit analyses are typically carried out through the comparison of a base case with variations of the base case, for example, operation with different voltage control strategies. As the number of scenarios to consider grows, the execution time can quickly become excessive. In this context, the computational challenge of quasi-static time-series simulations is twofold: their execution time is long and the required amount of input data is high (in the CSV file format it can reach the order of GB). Handling such amounts of data is, in itself, time consuming.

Much attention has historically been given to the computational performance of power flow calculations from the point of view of numerical accuracy and execution time (see for example [4], [5]), yet the computational performance of time-series power flows remains open for analysis. Perhaps the only explicit attempt to reduce their execution time so far consists in reducing a distribution feeder to a smaller equivalent circuit connected to the bus under analysis [6], which dramatically reduces the calculation time and keeps result accuracy high. The disadvantage of this method is that each simulation can only deliver results for one bus in the entire network.

Two aspects of quasi-static time-series simulations can be explored for improving their performance: In the case of cost-benefit analysis for LV grids, only a small set of aggregated results needs to be determined, such as total network losses, power exchange through the MV/LV transformer, and total power provision from distributed generators. These results can be translated into costs for the grid operator (see for example [7]). Additionally, the cyclical nature of load and generation profiles [8], [9] suggests that some time steps in the profiles represent load and generation scenarios that reappear over time and hence can be simulated only once.

The objective of this paper is to explore the feasibility of shortening quasi-static time-series simulations by means of reducing the amount of input data, and therefore, reducing the number of required power flow calculations. Two alternatives are compared in terms of data and execution time reduction, and result accuracy: one consists in reducing the time resolution of the input profiles through downsampling and the other in finding similar time steps in the input profiles through vector quantization and simulating them once.

This paper is structured as follows: Section II presents the characteristics of quasi-static time-series simulations for cost-benefit analysis. Section III introduces the shortening methods. Section IV defines the case studies used to test the shortening methods. Section V presents the results and their analysis. Finally, Section VI presents the conclusions and proposes future research.

II. QUASI-STATIC TIME-SERIES SIMULATIONS

Quasi-static time-series simulations are consecutive power flows calculated on each time step of a set of load, generation and slack voltage profiles. The results of such simulations are profiles that represent the behavior of the grid over the simulated period. Traditionally, the slack voltage is considered constant (1 p.u.), but in time-series simulations it can be a profile as well. This allows to simulate the influence of the MV level on the LV level.

A block diagram of a quasi-static time-series simulation is shown in Fig. 1. Its main aspects are described in this section.
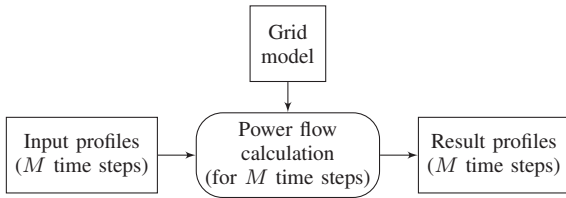

Figure. 1.   Block diagram of a quasi-static time-series simulation.

A. Simulation Inputs: Load, Generation and Voltage Profiles

In an LV network, loads and PV generators are associated with either residential or commercial buildings. The simulation inputs for a network with N load/PV buses are the following discrete time-dependent profiles:

- P_{L_n}: active load at the nth load/PV bus
- Q_{L_n}: reactive load at the nth load/PV bus
- P_{PV_n}: active PV generation at the nth load/PV bus
- Q_{PV_n}: reactive PV generation at the nth load/PV bus
- U_{slack}: slack voltage (MV side of the MV/LV transformer)

with $n = 1, 2, \ldots, N$. If each profile has M time steps, the set of input profiles can be represented as a matrix $\mathbf{X}$ with at most $4N + 1$ columns and M rows. Each column in $\mathbf{X}$ represents a profile and each row represents a time step. Table I shows an example of input profiles at one minute resolution for a network with two loads and two PV generators. In the example, the PV generators are operating at unity power factor.

TABLE I.   EXAMPLE OF SIMULATION INPUTS (MATRIX $\mathbf{X}$)

t [min]	P_{L_1} [kW]	P_{L_2} [kW]	Q_{L_1} [kvar]	Q_{L_2} [kvar]	P_{PV_1} [kW]	P_{PV_2} [kW]	Q_{PV_1} [kvar]	Q_{PV_2} [kvar]	U_{slack} [p.u.]
1	0.40	0.58	0.10	0.18	-0.31	-0.21	0.0	0.0	1.03
2	0.41	0.60	0.11	0.17	-0.32	-0.22	0.0	0.0	1.03
3	0.20	0.41	0.05	0.10	-0.40	-0.32	0.0	0.0	0.99
4	0.39	0.60	0.09	0.20	-0.29	-0.23	0.0	0.0	1.03
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
M	0.21	0.43	0.05	0.09	-0.41	-0.30	0.0	0.0	0.98

B. Simulation Outputs: Result Profiles

The simulation outputs to consider when the target application is a cost-benefit analysis are the following discrete time-dependent profiles:

- P_X: active power exchange through the MV/LV transformer
- Q_X: reactive power exchange through the MV/LV transformer
- P_{losses}: total network power losses
- P_{PV}: total active power provision from PV generators
- U_n: voltage at the nth load/PV bus

with $n = 1, 2, \ldots, N$. The total energy losses E_{losses} are of interest as well. The outputs have M time steps and can be described as a matrix $\mathbf{Y}$ where each column represents a result profile and each row represents a time step. Table II shows an example of result profiles at one minute resolution.

TABLE II.   EXAMPLE OF SIMULATION OUTPUTS (MATRIX $\mathbf{Y}$)

t [min]	P_X [kW]	Q_X [kvar]	P_{losses} [kW]	P_{PV} [kW]	U_1 [p.u.]	...	U_n [p.u.]	...
1	0.47	0.28	0.01	0.52	0.99	...	1.00	...
2	0.48	0.28	0.01	0.54	0.98	...	0.99	...
3	-0.08	0.15	0.03	0.72	0.95	...	0.94	...
4	0.49	0.29	0.02	0.52	0.98	...	0.99	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
M	-0.04	0.14	0.03	0.71	0.94	...	0.95	...

C. Power Flow Calculation

The solution to the power flow equations of a network is denoted by a (multivariate) function $\mathbf{f}_{pf}$ of a given time step in the simulation inputs. If $\mathbf{x}_m$ is the mth time step (row) in the input matrix $\mathbf{X}$, then the mth time step (row) in the output matrix $\mathbf{Y}$ is given by

$$\mathbf{y}_m = \mathbf{f}_{pf}(\mathbf{x}_m). \tag{1}$$

Thus, a quasi-static time-series simulation consists in calculating $\mathbf{y}_m$ from (1) for $m = 1, 2, \ldots, M$.

III. SHORTENING METHODS

A block diagram of a shortened quasi-static time-series simulation is shown in Fig. 2. The main aspects of this diagram are detailed in this section.
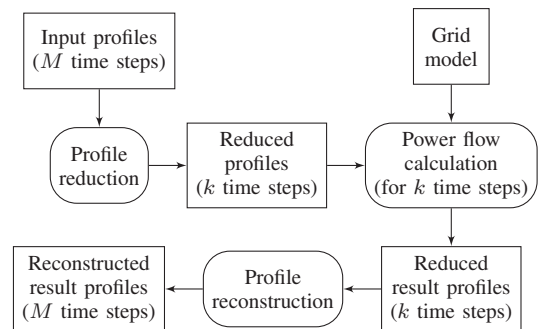

Figure. 2.   Block diagram of a shortened quasi-static time-series simulation.

A. Profile Reduction

1) Downsampling: Downsampling a profile consists in reducing its time resolution. If the original resolution is not a whole divisor of the new resolution, this procedure requires interpolation. To keep the execution time short, linear interpolation has been chosen. Downsampling, for example, the profiles from Table I to two minute resolution is equivalent to selecting the time steps at $t = 1, 3, 5, \ldots$, minutes. Interpolation is not required in this case.

2) Vector Quantization: Applying vector quantization to the input profiles means partitioning the input matrix $\mathbf{X}$ into disjoint sets of similar time steps and mapping each set to a representative time time step. The classic algorithm for vector quantization is k-means [10]. A modified version of k-means called k-means++ has been chosen as it produces better partitions [11]. With k-means++ $\mathbf{X}$ can be partitioned into k sets of similar time steps, also called clusters. A value for k must be given. The similarity between time steps is established through Euclidean distance. The algorithm maps the ith cluster to its centroid, denoted by $\mathbf{c}_i$, with $i = 1, 2, \ldots, k$. Each cluster centroid is calculated as the average of all the time steps in the cluster. For convenience, a vector l of labels that indicate to which cluster each time step in $\mathbf{X}$ belongs is defined.

Table III shows an example of the mapping of clusters to centroids and the structure of vector l. In this example, similar time steps can be recognized by simple inspection. In a good partition each centroid represents each time step in its corresponding cluster well, so power flows calculations for M time steps are no longer required, but only for k centroids.

B. Profile Pre-processing for Vector Quantization

1) Standardization: There is empirical evidence of the benefits of standardizing data before applying clustering algorithms based on Euclidean distance when the variables lay on different scales (e.g., P_{L_n} between $100\,\text{W}$ and $1\,\text{kW}$, U_{slack} in the vicinity of $1\,\text{p.u.}$). In this paper the standardization is achieved dividing each profile by its standard deviation [12].

2) Dimensionality Reduction: The execution time of k-means++ is influenced by the dimensionality of the data [13], defined in this case by the number of profiles in $\mathbf{X}$. If a set of profiles is highly correlated with one another, it is sufficient to use only one of them when searching for similar

time steps. The profiles used in this paper consider constant power factor loads and generators, which produce completely correlated active and reactive power profiles, and PV profiles from reduced geographical areas, which are highly correlated (see Section IV-A5). This property can thus be exploited for reducing the dimensionality of $\mathbf{X}$, so k-means++ can be applied to a profile subset composed of the active power load profiles, one active power PV profile and the slack voltage profile, reducing the dimensionality from $4N + 1$ to $N + 2$.

C. Result Reconstruction

1) From Downsampled Profiles: If, for example, the input profiles from Table I downsampled to two minute resolution are used to calculate the time-series power flow, the results will only consist of $\{\mathbf{f}_{pf}(\mathbf{x}_1), \mathbf{f}_{pf}(\mathbf{x}_3), \mathbf{f}_{pf}(\mathbf{x}_5), \ldots\} = \{\mathbf{y}_1, \mathbf{y}_3, \mathbf{y}_5, \ldots\}$. To obtain the remaining time steps, that is $\{\mathbf{y}_2, \mathbf{y}_4, \mathbf{y}_6, \ldots\}$, interpolation is required. To keep the execution time short, linear interpolations has been chosen.

2) From Vector-Quantized Profiles: Once the power flow of each centroid has been calculated, result profiles with k time steps are obtained. To transform these profiles into profiles with M time steps, the k time steps need to be reordered according to the labels in vector l. An example of the reconstruction process is shown in Table IV.

IV. CASE STUDIES

In order to evaluate and compare the considered shortening alternatives, a set of case studies using a benchmark network is developed.

A. Simulation Assumptions

1) Shortening Methods: Shortening through downsampling and vector quantization is tested.

2) Distribution Grid: The "Kerber *Dorfnetz*", designed to represent a German rural grid [14], is considered . Its one-line diagram is shown in Fig. 3.

3) PV Penetration: The low PV penetration scenario considers $7 \times 5\,\text{kW}_\text{p}$, $15 \times 10\,\text{kW}_\text{p}$ and $2 \times 30\,\text{kW}_\text{p}$ PV generators. The high PV penetration scenario considers $13 \times 5\,\text{kW}_\text{p}$, $26 \times 10\,\text{kW}_\text{p}$ and $3 \times 30\,\text{kW}_\text{p}$ PV generators. All generators are specified by their nominal power. The generators are randomly

TABLE III.   EXAMPLE OF VECTOR QUANTIZATION

l	$\mathbf{x}_m$	P_{L_1} [kW]	P_{L_2} [kW]	Q_{L_1} [kvar]	Q_{L_2} [kvar]	P_{PV_1} [kW]	P_{PV_2} [kW]	U_{slack} [p.u.]
1 ←	$\mathbf{x}_1$	0.40	0.58	0.10	0.18	-0.31	-0.21	1.03
1 ←	$\mathbf{x}_2$	0.41	0.60	0.11	0.17	-0.32	-0.22	1.03
2 ←	$\mathbf{x}_3$	0.20	0.41	0.05	0.10	-0.40	-0.32	0.99
1 ←	$\mathbf{x}_4$	0.39	0.60	0.09	0.20	-0.29	-0.23	1.03
:	:	:	:	:	:	:	:	:
2 ←	$\mathbf{x}_M$	0.21	0.43	0.05	0.09	-0.41	-0.30	0.98

$\mathbf{c}_i$	P_{L_1} [kW]	P_{L_2} [kW]	Q_{L_1} [kvar]	Q_{L_2} [kvar]	P_{PV_1} [kW]	P_{PV_2} [kW]	U_{slack} [p.u.]
$\mathbf{c}_1$	0.40	0.59	0.10	0.18	-0.31	-0.22	1.03
$\mathbf{c}_2$	0.21	0.42	0.05	0.10	-0.41	-0.31	0.99
:	:	:	:	:	:	:	:
$\mathbf{c}_k$	0.32	0.33	0.02	0.05	-0.21	-0.43	0.98

TABLE IV.   EXAMPLE OF RESULT RECONSTRUCTION

$\mathbf{c}_i$	P_{L_1} [kW]	P_{L_2} [kW]	Q_{L_1} [kvar]	Q_{L_2} [kvar]	P_{PV_1} [kW]	P_{PV_2} [kW]	U_{slack} [p.u.]
$\mathbf{c}_1$	0.40	0.59	0.10	0.18	-0.31	-0.22	1.03
$\mathbf{c}_2$	0.21	0.42	0.05	0.10	-0.41	-0.31	0.99
:	:	:	:	:	:	:	:
$\mathbf{c}_k$	0.32	0.33	0.02	0.05	-0.21	-0.43	0.98

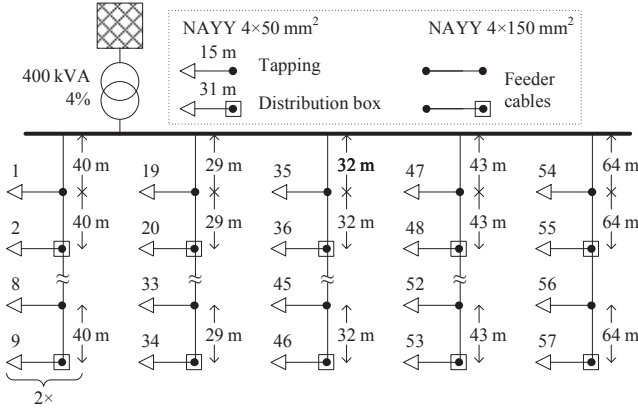| l	$\mathbf{y}_m$	P_X [kW]	Q_X [kvar]	P_{losses} [kW]	P_{PV} [kW]	U_1 [p.u.]	...
1 →	$\mathbf{f}_{pf}(\mathbf{c}_1) \approx \mathbf{y}_1$	0.48	0.28	0.01	0.53	0.98	...
1 →	$\mathbf{f}_{pf}(\mathbf{c}_1) \approx \mathbf{y}_2$	0.48	0.28	0.01	0.53	0.98	...
2 →	$\mathbf{f}_{pf}(\mathbf{c}_2) \approx \mathbf{y}_3$	-0.06	0.15	0.03	0.72	0.95	...
1 →	$\mathbf{f}_{pf}(\mathbf{c}_1) \approx \mathbf{y}_4$	0.48	0.28	0.01	0.53	0.98	...
:	:	:	:	:	:	:	
2 →	$\mathbf{f}_{pf}(\mathbf{c}_2) \approx \mathbf{y}_M$	-0.06	0.15	0.03	0.72	0.95	...

Figure. 3.   One-line diagram of the Kerber rural network.

assigned to the 57 possible PV connection points that are shown in Fig. 3.

4) Voltage Control: The $Q(U)/P70\%$ voltage control strategy is a distributed strategy that uses the reactive power provision capabilities of some PV inverters. Reactive power is provided once the voltage at the PV bus reaches a lower threshold, and increases until said voltage reaches an upper threshold. If the voltage continues to increase after this point, the reactive power provision remains constant. This strategy limits the active power provision to 70% of the peak power of the PV module [15].

5) Input Profiles: The load profiles are generated using the methodology introduced in [16] and have a constant power factor of 0.98. An irradiation profile measured in Munich, Germany, is scaled to match the order of magnitude of the rated power of each PV module [17]. The slack voltage profile is obtained from the yearly simulations of MV grids carried out in [7]. The time resolution of all profiles is one minute.

6) Number of Time Steps After Profile Reduction: The number of time steps in the reduced profiles (k) must be specified. This number affects simulation accuracy and duration. Given the nonlinearity of the power flow equations, relating k and result accuracy is challenging, especially in the case of vector quantization [18]. This relationship is thus investigated experimentally running every simulation for different values of k.

7) Computer and Software Specifications: The simulations are run on a computer cluster. Each node has two Intel Xeon CPUs (8 cores at $2.60\,\text{GHz}$), $126\,\text{GB}$ of working memory shared between all cores, and runs Scientific Linux 2.6.32 (64 bits). All simulations are run on the same node, sixteen in parallel. The Pypower power flow solver, a Python port of Matpower [19], is used.

B. Cases

The cases defined in Table V are investigated under the conditions described in Section IV-A. Each case study consists of a quasi-static time-series simulation (reference simulation) and a set of shortened simulations using different values of k. Cases 1 and 2 consider yearly simulations, but their shortened versions are broken down into twelve monthly simulations. Since the execution time of vector quantization also depends on the number of time steps in each profile [13], this prevents

TABLE V.   SUMMARY OF CASE STUDIES

Case	Method	PV penetration	Voltage control	Month
1	Downsampling	Low	No control	January to December
2	Vector quantization	Low	No control	January to December
3	Downsampling	High	No control	June
4	Vector quantization	High	No control	June
5	Downsampling	Low	$Q(U)/P70\%$	June
6	Vector quantization	Low	$Q(U)/P70\%$	June

an execution time escalation and allows a seasonal analysis. The remaining simulations consider only the month of June, as it has the highest PV feed-in.

V.   RESULTS

This section presents only a selection of results, however, their analysis is valid for all the results specified in Section II-B. Result accuracy is measured by the coefficient of determination (R^2) and the RMS error (RMSE). Accurate results have an R^2 close to one and a low RMSE. The RMSE is more sensitive to large deviations and can increase steeply even if the duration of the deviations is short and the profiles under comparison have similar overall shapes [20], therefore the R^2 is also calculated.

A. Influence of Seasonal Profile Variations

Fig. 4 and 5 indicate that result accuracy depends on seasonal profile characteristics. Regardless of the shortening method, the most accurate months are December and January, both winter months, while the least accurate months are June and July, both summer months. Seasonal variations are present in the load, PV and slack voltage profiles, thus, the accuracy variations cannot be attributed in this case to a single factor.

B. Influence of PV Penetration Level

Fig. 6 shows that result accuracy changes between PV penetration scenarios. This result, together with the yearly simulation from Fig. 5, show that the accuracy is affected by the level of PV feed-in, however, comparing only two scenarios it is not possible to establish with generality in what way the PV penetration level affects the accuracy of the results.

C. Influence of Voltage Control

The accuracy comparison of simulations with and without voltage control shown in Fig. 6 reveals that the results with voltage control are always more accurate by a small margin.
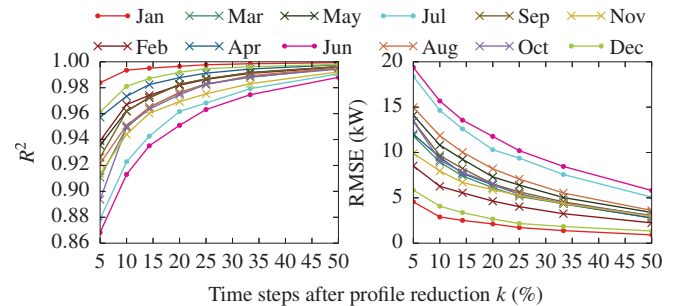

Figure. 4.   Accuracy comparison of P_X profiles for every month in a year. Data reduced through downsampling (case 1). Value of k in percentage of M.
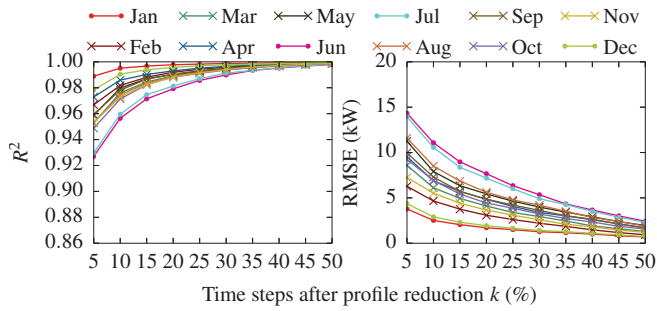
Figure. 5. Accuracy comparison of P_X profiles for every month in a year. Data reduced through vector quantization (case 2). Value of k in percentage of M.

An explanation can be found in Fig. 7, which shows a section of reference and reconstructed profiles, without voltage control (left) and with voltage control (right) from cases 2 and 6. Following the dashed line at $U = 1.04$ p.u. it is apparent that the most difficult section to reconstruct are the fast-occurring peaks between approximately 13:30 and 15:00 hours. Shortening simulations through vector quantization implies calculating power flows of averaged time steps, which eliminates fast profile variations and clips peaks. A similar clipping effect is found due to the interpolation required for result reconstruction from downsampled data. Fig. 7 shows how voltage control reduces the highest peaks, allowing the reconstructed profile to get closer to the reference profile.

Fig. 8 shows that, with and without voltage control, the execution time of the shortened simulations is similar even though the reference simulations differ in about 8 minutes. As an example, using vector quantization the time savings with voltage control for $k = 20\%$ of M are of about 52% while 40% savings are registered without voltage control.

D. Vector Quantization versus Downsampling

According to Fig. 4, 5 and 6, vector quantization substantially outperforms downsampling in terms of accuracy. However, considering the execution time shown in Fig. 8 and the accuracy shown in Fig. 6, without voltage control downsampling yields an accuracy of $R^2 = 0.97$ in 17 minutes ($k = 50\%$ of M) while vector quantization requires 20 minutes to reach the same accuracy ($k = 25\%$ of M). Despite the slight superiority of downsampling in terms of execution time in this specific case, vector quantization requires much less data to reach the same accuracy. This is beneficial in terms
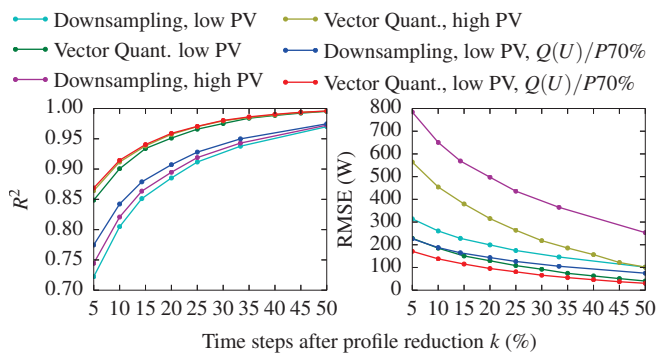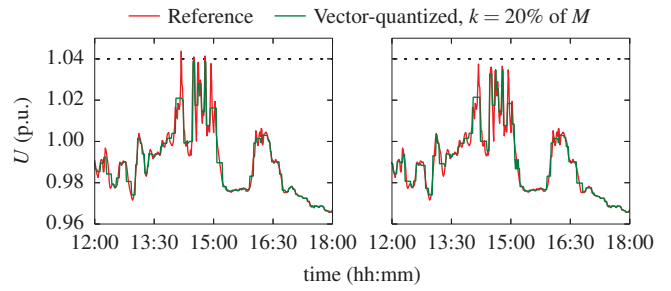
of data handling and when more than one simulation needs to be run using the same input profiles, for example, for comparing different voltage control strategies. Fig. 9 shows the time savings achieved when more than one simulation is run with the same input profiles with a P_{losses} target accuracy of $R^2 = 0.97$, and the reduction process is run only once. Vector quantization yields higher time savings than downsampling if at least two simulations without voltage control are run. If five simulations are run, the time savings rapidly increase to 65% with vector quantization while with downsampling they remain at 44%. The benefits are larger with voltage control. According to Fig. 9, five simulations yield savings of 70% for vector quantization and of 43% for downsampling. It should be noted that with voltage control, vector quantization yields the highest time savings regardless of the number of simulations.

E. Accuracy of Energy Losses

The energy losses (E_{losses}), calculated as the area under the P_{losses} profile, are estimated with very high accuracy with both methods and under every tested condition. Even with $k = 5\%$ of M, E_{losses} can be estimated with an error lower than 0.7%. The error quickly drops below 0.2% for $k = 25\%$ of M.

F. Choosing the Number of Time Steps After Profile Reduction

The accuracy curves shown in Fig. 5 and 6 have an elbow located before $k = 20\%$ of M. This elbow represents the best possible trade-off between accuracy and execution time reduction, nevertheless, its location cannot be generalized from these results, and a computationally inexpensive method for finding it needs to be developed.

Figure. 7. Comparison of U_{33} profiles from reference and vector quantized simulations on June 22nd without voltage control (left) and with voltage control (right) in the low PV penetration scenario (cases 2 and 6).

Figure. 6. Accuracy comparison of P_{losses} profiles from simulations of the month of June (cases 1 to 6). Value of k in percentage of M.
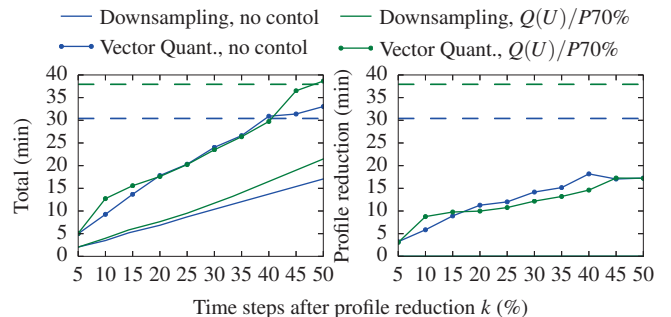

Figure. 8. Comparison of the execution time of simulations with low PV penetration, month of June (cases 1, 2, 5 and 6). The data reduction time of downsampling is negligible. Dashed lines represent the execution time of the reference simulations at 1 min. resolution. Value of k in percentage of M.
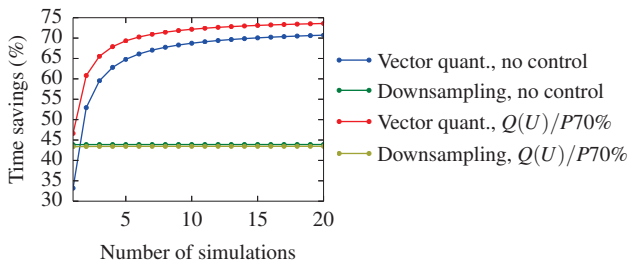
Figure. 9. Execution time savings running multiple simulations with the same input profiles in the low PV penetration scenario for a P_{losses} target accuracy of $R^2 = 0.97$ (cases 1, 2, 5 and 6). For downsampling $k = 50\%$ of M and for vector quantization $k = 25\%$ of M.

VI. CONCLUSION

In this paper, the feasibility of shortening the execution time of quasi-static time-series simulations through downsampling and vector quantization of input profiles is studied in terms of result accuracy and execution time savings. The results show that the accuracy of shortened simulations is affected by the level of PV feed-in. The use of $Q(U)/P70\%$ voltage control has a positive albeit marginal effect on result accuracy. The highest time savings and accurate results can be achieved thorough downsampling only when one simulation without voltage control is run and the input profiles are not drastically reduced. Accurate results and time savings, as well as a considerable profile reduction, can be obtained through vector quantization under every tested condition. For a total power losses profile with an accuracy of $R^2 = 0.97$, the smallest time saving registered with vector quantization is of 34%. At constant accuracy, the time savings with vector quantization increase if either voltage control is used or if more than one simulation is run using the same input profiles, as in both cases the time proportion spent on profile reduction decreases. With voltage control, time savings reach 70% if five simulations are run for a total power losses profile with an accuracy of $R^2 = 0.97$. Total energy losses can be estimated with errors in the order of 0.2% or lower. The results suggests that vector quantization can be particularly useful for cost-benefit analyses that compare scenarios with variations in the network model but use the same input profiles. Such is the case of cost-benefit analyses of different voltage control strategies. The main drawback of this method is its inaccuracy for reproducing peak values in the results profiles. This makes the method unsuitable, for example, for detection of voltage violations.

Future work should establish the generality of the results for other networks and other voltage control strategies. A clear relationship between result accuracy and PV feed-in must be determined. A computationally inexpensive procedure for choosing the level of profile reduction that guarantees the desired trade-off between time savings and result accuracy needs to be developed. Investigating improvements that allow accurate reconstruction of profile peaks would make the method suitable for more diverse applications. A possible research path is developing a criterion for selecting a time step from each cluster instead of calculating power flows of cluster centroids.

ACKNOWLEDGMENTS

REFERENCES

[1] J. Peças Lopes, "Integration of dispersed generation on distribution networks-impact studies," in *IEEE Power Engineering Society Winter Meeting, 2002*, vol. 1, 2002, pp. 323–328 vol.1.

[2] L. Ochoa, A. Padilha-Feltrin, and G. Harrison, "Time-series-based maximization of distributed wind power generation integration," *IEEE Transactions on Energy Conversion*, vol. 23, no. 3, pp. 968–974, Sep. 2008.

[3] "IEEE guide for conducting distribution impact studies for distributed resource interconnection," *IEEE Std 1547.7-2013*, pp. 1–137, Feb. 2014.

[4] B. Stott and O. Alsac, "Fast decoupled load flow," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 3, May 1974.

[5] C.-P. Ng, K. Jabbour, and W. Meyer, "Loadflow analysis on parallel computers," in , *Proceedings of the 32nd Midwest Symposium on Circuits and Systems, 1989*, Aug. 1989, pp. 10–15 vol.1.

[6] M. Reno, K. Coogan, R. Broderick, and S. Grijalva, "Reduction of distribution feeders for simplified PV impact studies," in *PV Specialists Conference (PVSC), 2013 IEEE 39th*, Jun. 2013, pp. 2337–2342.

[7] T. Stetz, M. Kraiczy, M. Braun, and S. Schmidt, "Technical and economical assessment of voltage control strategies in distribution grids," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 6, pp. 1292–1307, Sep. 2013.

[8] J. Bank and B. Mather, "Analysis of the impacts of distribution connected PV using high-speed datasets," in *2013 IEEE Green Technologies Conference*, Apr. 2013, pp. 153–159.

[9] E. Wieben, "Multivariates Zeitreihenmodell des aggregierten elektrischen Leistungsbedarfes von Standardverbrauchern für die probabilistische Lastflussberechnung," Ph.D., Technische Universität Clausthal, 2008.

[10] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[11] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '07. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[12] G. W. Milligan and M. C. Cooper, "A study of standardization of variables in cluster analysis," *Journal of Classification*, vol. 5, no. 2, pp. 181–204, Sep. 1988.

[13] S. Har-Peled and B. Sadri, "How fast is the k-means method?" in *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '05. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005, pp. 877–885.

[14] Georg Kerber, "Aufnahmefähigkeit von Niederspannungsverteilnetzen für die Einspeisung aus Photovoltaikkleinanlagen," PhD, Technische Universität München, München, Mar. 2011.

[15] J. von Appen, T. Stetz, M. Braun, and A. Schmiegel, "Local voltage control strategies for PV storage systems in distribution grids," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1002–1009, Mar. 2014.

[16] Von Appen, Haack, and Braun, "Erzeugung zeitlich hochaufgelöster Stromlastprofile für verschiedene Haushaltstypen," in *Proc. 2014 IEEE Power and Energy Student Summit*, Stuttgart, Jan. 2014.

[17] J. Appen, M. Braun, B. Zinßer, D. Stellbogen, "Leistungsbegrenzung bei PV-Anlagen," in *Proc. 2012 27*, pp. 47–52.

[18] B. Mirkin, "Choosing the number of clusters," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 252–260, May 2011.

[19] R. Zimmerman, C. Murillo-Sanchez, and R. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

[20] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.