A novel non-intrusive method using design of experiments and smooth approximation to speed up multi-period load-flows in distribution network planning

Héloïse Dutrieux Baraffe, Marjorie Cosson, Julien Bect, Gauthier Delille, Bruno Francois

A novel non-intrusive method using design of experiments and smooth approximation to speed up multi-period load-flows in distribution network planning

Héloïse DUTRIEUX BARAFFE[a]*, Marjorie COSSON[ab], Julien BECT[b], Gauthier DELILLE[a], and Bruno FRANCOIS[c]

[a] EDF R&D, 7 boulevard Gaspard Monge, F-91120 Palaiseau, France

[b] Laboratoire Signaux et Systèmes (L2S), CentraleSupélec, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 3 rue Joliot-Curie, F-91192 Gif-sur-Yvette, France

[c] Univ. Lille, Centrale Lille, Arts et Métiers Paris Tech, HEI, EA 2697 – L2EP, Laboratory of Electrical Engineering and Power electronics, F-59000 Lille, France

* Corresponding author. Email: heloise.baraffe@edf.fr

Abstract

Alternative solutions to network reinforcement are now being investigated in distribution network planning studies to reduce the costs and periods for integrating renewable energy sources. However, a thorough techno-economic analysis of these solutions requires a large number of multi-period load-flow calculations, which makes it hard to implement in planning tools. A non-intrusive approximation method is therefore proposed to obtain fast and accurate multi-period load-flows. This method builds a surrogate model of the load-flow solver using polynomial regression and kriging, combined with Latin hypercube sampling. Case studies based on real distribution networks show that the proposed method is more efficient for distribution network planning in presence of renewable energy sources than time subsampling, line model simplification, and voltage linearization. In particular, accurate 10-minute profiles of voltages, currents, and network power losses are obtained in a satisfactory computation time.

Keywords

Approximation method; Distribution network planning; Kriging model; Multi-period load-flow; Polynomial regression model; Principal Component Analysis.

1. Introduction

To enable the expected intensive development of Renewable Energy Sources (RES) and new electrical usages (active demand, electric vehicles, etc.), distribution network planning needs to evolve quickly [1,2]. Today, most of the voltage/current constraints due to RES are removed by reinforcing the network, i.e., by replacing existing network infrastructures or adding new ones. As these network adaptations may be expensive and take time, several alternative solutions, such as Volt-VAR control and load/generation curtailment, are now being investigated to reduce the costs and periods for integrating RES while ensuring an acceptable level of risk and quality in the network. Unlike network

reinforcement, alternative solutions may have operating limits in energy or duration and important operational costs depending on the voltage/current constraints which have been avoided. Therefore, to assess the techno-economic impacts of these solutions and find the best ones, network planning methods should allow, not only to detect any risk of constraint as it is done today, but also to characterize the constraints in terms of depth, duration and frequency. This implies studying multi-year profiles of load and generation, and thus performing multi-period load-flow calculations [2].

The time uncertainty and power variability of intermittent RES must be taken into account in order to accurately assess the temporary constraints, and thus the techno-economic performance of the solutions. It is thus essential to study a large number of load/generation profiles, with a time step ΔT which is as small as possible (no more than 10 minutes, see Section 3.2). This nevertheless leads to an intensive use of a time-consuming "load-flow solver" method for solving the nonlinear load-flow equations. For instance, if $p = 100$ yearly load/generation profiles with a time step of $\Delta T = 10$ minutes are considered over a period of $A = 10$ years ($\approx 5.3 \times 10^6$ minutes), then $n = pA / \Delta T \approx 5.3 \times 10^7$ load-flow calculations are required to assess the performance of alternative solutions for a given network. This means around seven hours of computation[1] when using a Forward-Backward load-flow solver [20] under MATLAB for one of the radial 400-bus networks studied in Section 3.2. Such a computation time is not acceptable for Distribution System Operators (DSO), who are often responsible for the medium/long-term development of several hundred or even thousand primary substations.

The possible options to reduce the computation time needed for multi-period load-flow calculations can be classified into three categories: 1) time subsampling of the load/generation profiles, 2) intrusive approximation of the load-flow solver, and 3) non-intrusive approximation of the load-flow solver.

Time subsampling consists in using only a part of the available input data, by either increasing the time step ΔT of the yearly load/generation profiles or, in rarer cases, selecting only a few daily profiles for each year. Yearly profiles averaged over a time step of $\Delta T = 30$ minutes or $\Delta T = 1$ hour are commonly used to study alternatives to reinforcement [3-6]. Time subsampling is easy to implement, which accounts for its popularity, but the results are not very accurate compared to the time savings (see Section 3.2.3).

Intrusive approximation consists in simplifying the load-flow equations using physical assumptions and/or intrusive approximation techniques. This option is often used to study network stability [7] or the statistical impacts of load and generation power variations [8]. The most popular approaches are to linearize the load-flow equations at one or several operating points, to simplify the voltage equations by assuming voltage angles equal to zero, or to model the electrical lines by series resistances and reactances only. The effectiveness of this option strongly depends on the assumptions and intrusive approximation techniques used.

Non-intrusive approximation consists in building a surrogate model for the load-flow solver using

---

[1] All the computation times presented in this paper are obtained with a laptop with a 2.50-GHz processor (Intel Core i7 4710MQ) and a 16-GB RAM.

approximation techniques that treat the load-flow solver as a "black box". The effectiveness of this option depends on the sampling methods used to select the operating points where the exact model (the load-flow solver here) is evaluated, and on the approximation methods used to build a surrogate model based on the evaluation results. To our knowledge, non-intrusive approximation has rarely been used in network studies and, if so, only in a simple form like nearest-neighbor interpolation [9,10]. Smooth approximation techniques, such as polynomial regression or kriging, seem not to have been investigated in network planning studies so far.

This paper investigates the latter option, i.e., non-intrusive approximation, to obtain fast and accurate multi-period load-flows. The proposed method builds a surrogate model of the load-flow solver using polynomial regression and kriging combined with Latin hypercube sampling. The paper is organized as follows. Section 2 describes a generic approximation procedure to estimate any function, then presents the proposed method for multi-period load-flows. Section 3 illustrates the performance of the proposed method, in terms of computation time and approximation errors, through several case studies based on real distribution networks and 10-minute profiles of load/generation. The effectiveness of the proposed method is compared with four other methods commonly used in network planning: the use of 30-minute averaged load/generation profiles, the computation of a single load-flow iteration, the modelling of electric lines by series resistances and reactances only, and the linearization of the voltage load-flow equations. Finally, Section 4 discusses the validity area of the proposed method.

2. Proposed method for the approximation of the load-flow solver

2.1. *A generic non-intrusive approximation procedure*

Let us consider a real *s*-dimensional variable $y = [y_1 \ ... \ y_s]$, which is the result of a function *f* when applied to the real *m*-dimensional variable $x = [x_1 \ ... \ x_m]$: $y = f(x)$. The purpose is to calculate the *n* values of *y*, gathered in the matrix *Y*:

$$Y = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{bmatrix} = \begin{bmatrix} y_1^{(1)} & \cdots & y_s^{(1)} \\ \vdots & \ddots & \vdots \\ y_1^{(n)} & \cdots & y_s^{(n)} \end{bmatrix},$$

that are associated with *n* given values of *x*, gathered in the matrix *X*:

$$X = \begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(n)} \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & \cdots & x_m^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(n)} & \cdots & x_m^{(n)} \end{bmatrix}.$$

Let us assume that: 1) the number of points *n* and the output dimension *s* are very large, and 2) the evaluation of *f* is time-consuming, which makes it difficult or even impossible to compute *Y* using *f* and *X* directly. An approximation procedure is therefore required to estimate *Y* precisely in an acceptable computation time.

Table 1 details a generic non-intrusive approximation procedure to estimate the output-value matrix *Y*. Different variants of this procedure have already been used in several application fields [11–13]. This procedure is based on a sampling method, a dimension reduction method, and an approximation method.

The sampling method selects a small set of input points, called design of experiments, so as to guarantee a high-quality approximation (step 2 in Table 1).

The dimension reduction method decreases the dimension of the output variable *y*, and thus reduces the number of times the approximation method is used in the procedure (step 4 in Table 1). We detail here the Principal Component Analysis (PCA), which is one of the most commonly used dimension reduction methods [14]. This unsupervised linear dimension reduction method eliminates the variables that are interdependent from *f* and keeps only q ($<< s$) linearly-uncorrelated non-physical variables, called principal components, for the approximation procedure.

The approximation method builds a fast surrogate model for each principal component (step 5 in Table 1). The value of *y* is finally computed based on the approximate values of its principal components (step 7 in Table 1).

The generic procedure detailed in Table 1 can be tailored to different application fields thanks to three degrees of freedom: the sampling method used in step 2, the approximation method used in step 5, and the number q of principal components from the PCA in step 4. The choice of the sampling and approximation methods largely depends on the properties of the function *f* to approximate (e.g., if its inputs/outputs are continuous or discrete, if its response is smooth or not, etc.). The number of principal components can be chosen in different ways, e.g., to represent a certain share of the total variance of *y*. Another degree of freedom, which is less obvious, is the set of variables to which the PCA is applied. More precise results can sometimes be obtained if the approximation procedure is applied to different sets of *f* outputs [15].

1) Select a sampling method and an approximation method.

2) Apply the sampling method to build a design of experiments: $X* = \begin{bmatrix} x^{(1*)} \\ \vdots \\ x^{(n*)} \end{bmatrix}$ where $n*$ is

substantially smaller than n.

3) Compute the output $Y*$ associated with $X*$ using the exact model f:

$$\forall i \in [1*; n*] \quad y^{(i)} = f(x^{(i)}).$$

4) Perform the Principal Component Analysis (PCA) to convert $Y*$ into a reduced set of the q first

principal components: $Z* = [Z_1 * \cdots Z_q *] = \begin{bmatrix} z_1^{(1*)} & \cdots & z_q^{(1*)} \\ \vdots & \ddots & \vdots \\ z_1^{(n*)} & \cdots & z_q^{(n*)} \end{bmatrix}$, as follows:

$$Z* = (Y* - \mathbf{1}.\bar{Y})W,$$

where $\bar{Y} = [\bar{y}_1 \cdots \bar{y}_s]$ is the empirical mean vector of $Y*$, $\mathbf{1}$ is the $n* \times 1$ vector of ones and W is the

$s \times q$ matrix composed of the weighting coefficients from the PCA (i.e., the columns of W are the

orthonormal eigenvectors of the empirical covariance matrix corresponding to the q largest

eigenvalues).

5) Create a surrogate model for each principal component: $f_1*, \ldots, f_q*$. For each principal component

k, estimate the parameters of the surrogate model f_k* based on the pair $(X*, Z_k*)$.

6) Compute the matrix $\tilde{Z}$ of approximate principal components associated with X using the surrogate

models $f_1*, \ldots, f_q*$:

$$\forall k \in [1; q] \ \forall i \in [1; n] \ \tilde{z}_k^{(i)} = f_k * (x^{(i)}).$$

7) Perform the inverse PCA, i.e., compute the matrix $\tilde{Y} = \mathbf{1}.\bar{Y} + \tilde{Z}.W^T$ of approximate output values,

where $\mathbf{1}$ is the $n \times 1$ vector of ones.

Table 1. Generic non-intrusive approximation procedure to estimate any s-dimensional variable $y = f(x)$.

2.2. *Interest of a non-intrusive approximation procedure for multi-period load-flows in distribution network planning*

As explained in Section 1, a thorough study of the solutions to integrate RES requires a large number of load-flow calculations, which is time-consuming. There is therefore a real need to use fast surrogate models instead of the load-flow solver.

The function to approximate f is the input-output mapping associated to the load-flow solver. This function f provides an abstract view of the resolution of the load-flow equations by the load-flow solver, for a given set of input parameters x. The output $y = f(x)$ depends not only on x but also, implicitly, on the definition of the network under study and the assumptions about power dispatching.

A large number s of load-flow outputs is involved in the techno-economic analysis of alternatives to reinforcement: the voltages at the Medium-Voltage (MV) buses U, the real and reactive powers in the MV lines P_l and Q_l, the real and reactive powers at the High-Voltage/Medium-Voltage (HV/MV) transformer P_0 and Q_0, the overall real power losses in the MV network P_{loss}. A dimension reduction method like the PCA is thus suitable.

Smooth approximation methods are appropriate here since the load-flow outputs are relatively smooth. Moreover, the number m of load-flow inputs is generally limited in planning network studies, thanks to assumptions about power dispatching (see Sections 3.1 and 4.2 for examples). Consequently, advanced smooth approximation methods, such as kriging, can be considered here.

2.3.   The proposed non-intrusive approximation procedure

The proposed Non-Intrusive Approximation procedure (NIA) results from a thorough comparative study of different variants of the generic procedure applied to different distribution networks [15]. NIA includes three steps to reach a satisfactory trade-off between computation time and accuracy (Fig. 1).

Step 1: surrogate models of the load-flow solver are built using a n^*-point maximin Latin Hypercube Sample (LHS) [16,17] as a design of experiments and Polynomial Regression of order 3 (PR3) [17] as an approximation method. To do this, the procedure in Table 1 is applied four times, to estimate separately P_0, Q_0, P_{loss} and $y_{upq} = [U, P_l, Q_l]$. Applying the procedure to these subsets of outputs leads to more accurate results than to $y = [U, P_l, Q_l, P_0, Q_0, P_{loss}]$, in return for a short increase in time [15]. The PCA is performed on the non-scalar output only, i.e., on y_{upq}. Before performing the PCA on y_{upq}, each set of outputs of the same nature (U, P_l, and Q_l) is centered around zero by subtracting its empirical mean, then is divided by the square root of its total empirical variance. The number q of principal components is set so as to represent more than α percent of the total variance of the variable to estimate. The accuracy of the surrogate models is assessed on a k-point test sample. The surrogate models are considered precise if the errors committed on the test sample are smaller than thresholds defined for each output: ε_U, ε_{Pl}, ε_{P0}, ε_{Q0}, and ε_{Ploss}. If the surrogate models are precise, they can be used instead of the load-flow solver to estimate the output profiles.

Step 2: if the surrogate models are not precise enough, new surrogate models are built as in step 1, but using Ordinary Kriging with a 5/2 Matérn covariance (OK52) [17-19] instead of PR3. OK52 is used in step 2 because it is generally slower but more accurate than PR3 [15]. The parameters of the Matérn covariance function are estimated by the approximation method itself so as to maximize the likelihood of the output values from the design of experiments in the covariance model. The surrogate models are assessed on the same test sample.

Step 3: in the rare case where the surrogate models built in step 2 are not precise enough, the output profiles are computed using the exact load-flow solver.
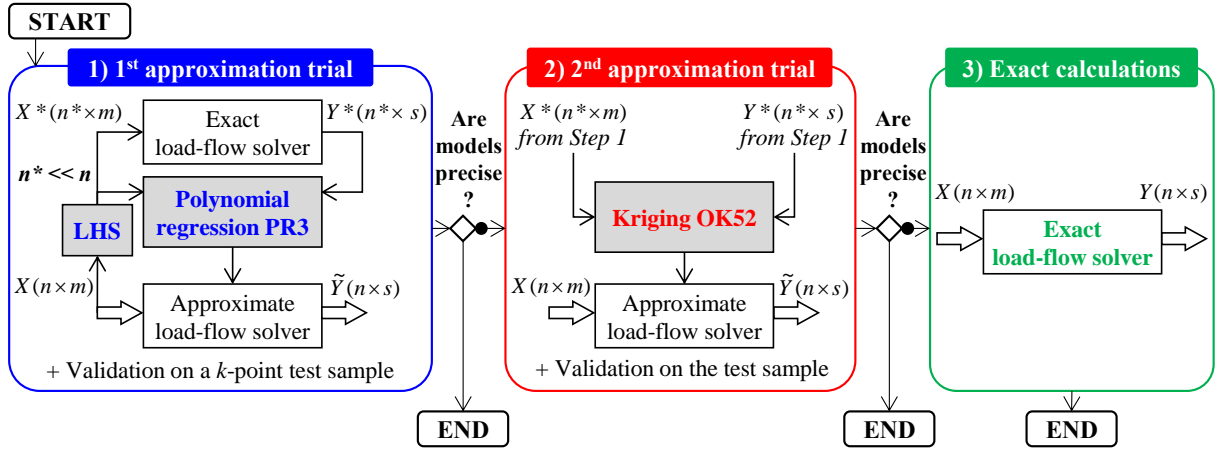
Fig. 1. The proposed non-intrusive approximation procedure to approximate multi-period load-flows.

If NIA stops at step 1 or 2, only $(n* + k)$ exact load-flow calculations are required.

The accuracy-time ratio for building and validating the surrogate models strongly depends on $n*$, α, and k. Increasing $n*$ or α enhances the accuracy of the surrogate models but also raises the computation time required for building them. Increasing k improves the assessment of the models' accuracy but, at the same time, leads to more exact load-flow calculations to validate the models.

NIA can also be used to estimate other load-flow outputs, e.g., the currents in all the MV lines I and the apparent power at the HV/MV transformer S_0. Because of power flow reversing, I and S_0 are not smooth, which makes them difficult to approximate accurately using polynomial regression or kriging. In order to build precise surrogate models for these outputs, it is actually better to approximate P_l, Q_l, P_0, and Q_0 rather than I and S_0, and then to compute approximations of I and S_0 as follows:

$$\forall \text{ line or transformer } i, \ S_i = \sqrt{P_i^2 + Q_i^2}, \tag{1}$$

$$\forall \text{ line } i, \ I_i = S_i / U_j \text{ , where } U_j \text{ is the end bus of the line } i. \tag{2}$$

3. Case studies: comparing the proposed method with four commonly used methods to speed up load-flow calculations

3.1. General information about the case studies presented in Sections 3.2 and 3.3

The purpose is to compute the 10-minute profiles of load-flow outputs quickly and accurately for different radial distribution networks. This means performing $n = 52\,560$ exact load-flow calculations per year. A Forward-Backward algorithm is used as a benchmark for solving load-flow equations [20]. This load-flow solver is especially efficient for radial networks. It takes 25 seconds on average to compute $n = 52\,560$ load-flow calculations under MATLAB for the radial 400-bus networks studied in Section 3.2.

For techno-economic issues, four sets of load-flow outputs have to be estimated: U, I, P_{loss}, and S_0. For the radial 400-bus networks studied in Section 3.2, this means estimating $s = 1201$ output variables

(since P_l, Q_l, P_0, and Q_0 have to be approximated to compute I and S_0, see Eq. 1-2). The load-flow solver takes $m = 4$ inputs here: the overall wind generation P_w, the overall photovoltaic generation P_{pv}, the overall consumption P_c, and the busbar voltage U_{ref}. Powers P_w, P_{pv}, and P_c are respectively dispatched between all the generators and loads proportionally to their rated real power (for producers) or peak apparent power (for consumers). Reactive powers are computed using power factors.

NIA is set to reach a satisfactory trade-off between speed and accuracy: $n^* = 200$, $\alpha = 99.999$ and $k = 200$ [15]. This means that only 400 exact load-flow calculations are required if NIA stops at step 1 or 2. The maximal errors tolerated on the test sample are: $\varepsilon_U = 0.0025$ pu, $\varepsilon_I = 5$ A (i.e., < 5 % of rated currents of the MV lines), $\varepsilon_{S0} = 100$ kVA (i.e., < 1 % of rated apparent powers of HV/MV transformers), and $\varepsilon_{Ploss} = 1$ % of yearly energy losses.

The performance of NIA is compared with four other methods commonly used in network planning: 1) 30-minute Subsampling (S30), which consists in performing exact load-flows after averaging load/generation profiles over a 30-minute range, 2) Single load-flow Iteration (SI), which consists in setting the initial values of voltage at the final values from the last load-flow calculation and performing only one iteration of the exact load-flow solver, 3) Line Model simplification (LM), which consists in modeling the electric lines by series resistances and reactances only (without considering shunt capacitances), and 4) Voltage Linearization (VL), which consists in linearizing the voltage equations at the mid-range operating point and applying the resulting approximate load-flow solver on the 10-minute profiles of load/generation. The case studies below have been implemented in MATLAB using the STK toolbox for kriging [21].

3.2. General performance in terms of speed and accuracy

3.2.1. Purpose of the study

This study aims at analyzing the general performance of S30, SI, LM, VL, and NIA over 100 different scenarios. Each scenario is composed of a 20-kV 400-bus radial network and a set of 10-minute profiles of P_w, P_{pv}, and P_c, and U_{ref} over one year. All the studied networks are possible evolutions of a same real network after accommodating from 0 to 20 MW of RES (see Ref. [22] for an example).

3.2.2. Performance indicators

Speed is assessed using the computation time T, i.e., the total time of the considered method for building surrogate models and computing the 10-minute profiles of U, I, S_0, and P_{loss} over one year. Another speed indicator used here is the time efficiency G_T of the method, i.e., the ratio between computation time from the exact load-flow solver (around 25 seconds for $n = 52\,560$ load-flows) and computation time from the method.

Accuracy is assessed using the output errors, i.e., the absolute deviations between the 10-minute output profiles from the considered method and the ones from the exact load-flow solver. For each set of output variables (U, I, S_0, or P_{loss}), three error indicators are introduced to characterize the distribution of absolute errors for each scenario: the mean of absolute errors err_{mean}, the 90-percent quantile of

absolute errors err_{Q90}, and the maximum of absolute errors err_{max} over all the time steps and, for voltage/current errors, over all the buses/lines.

3.2.3. Results of the study

Fig. 2 focuses on six randomly chosen hours of the voltage at a given bus i and the associated absolute errors obtained by S30, VL, and NIA with respect to the exact load-flow solver. This example shows clearly that the voltage profiles obtained by VL and NIA are closer to the exact one than that of S30.
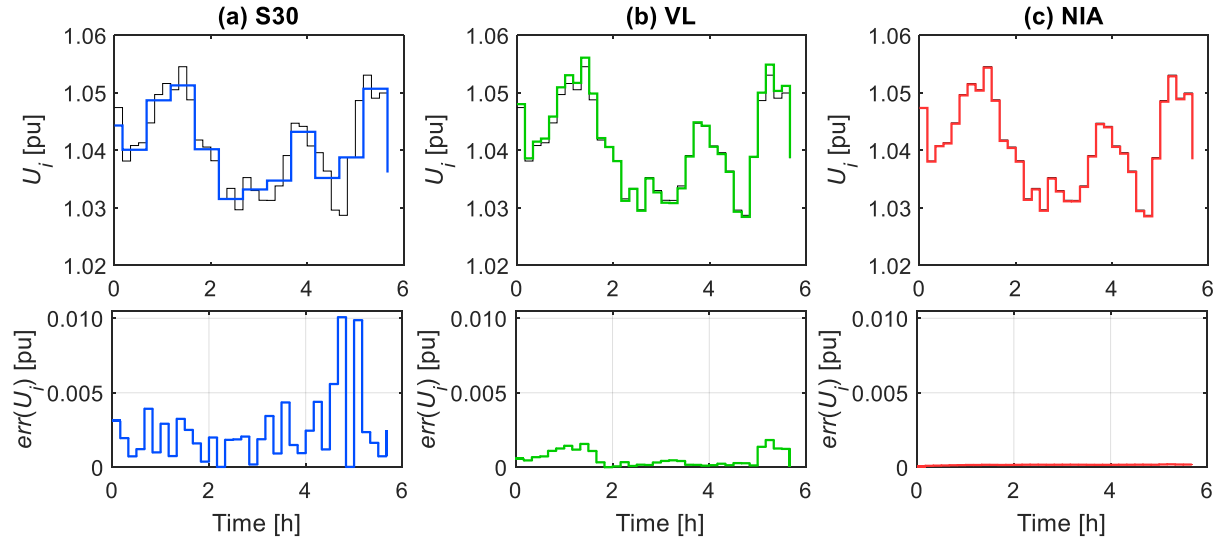

Fig. 2. Top: examples of voltage computed by the exact load-flow solver (black line on a, b and c), S30 (blue line on a), VL (green line on b), and NIA (red line on c). Bottom: voltage error from S30 (blue line on a), VL (green line on b), and NIA (red line on c) with respect to the exact load-flow solver. Note that the nominal voltage is 20 kV here.

Table 2 depicts the averages of time efficiency and of mean/quantile/max absolute errors over the 100 scenarios for S30, SI, LM, VL, and NIA. It turns out that:

- S30 and LM are the worst methods. They are time-consuming (Fig. 3.a) in addition to generally committing the highest errors on the load-flow outputs (Fig. 3.b and Table 2).

- SI obtains accurate outputs (Fig. 3.b and Table 2) but it is rather slow (Fig 3.a).

- VL and NIA get high accuracy on the outputs in a satisfactory computation time.

- NIA is the most efficient method. It gets the highest accuracy on the outputs (Table 2). Moreover, it is 18 times faster than 10-minute exact load-flows whereas the other methods are only 2 to 4 times faster. Note that building the surrogate models by PR3 (step 1) and by PR3 then OK52 (steps 1-2) takes respectively around 81 % and 98 % of the total computation time here.

This case study over 100 scenarios (with $n = 52\ 560$) highlights the high performance of the proposed method to approximate multi-period load-flows. With $n^* = 200$, $\alpha = 99.999$ and $k = 200$, NIA performs fast and accurate load-flow calculations for several variations of a radial network with different overall generation and power dispatching. Although commonly used in network planning studies, time

subsampling is not efficient to estimate 10-minute profiles of load-flow outputs.

		S30	SI	LM	VL	NIA
CPU time [s]	T	8.3	11.7	19.3	7.5	1.3
Time efficiency [-]	G_T	3.0	2.1	1.3	3.3	18.4
Voltage errors [pu]	$err_{mean}(U)$	1.3E-3	<1E-4	7.9E-4	<1E-4	<1E-4
	$err_{Q90}(U)$	3.2E-3	<1E-4	1.5E-3	2.2E-4	<1E-4
	$err_{max}(U)$	1.4E-2	8.3E-4	2.2E-3	1.5E-3	1.3E-4
Current errors [A]	$err_{mean}(I)$	0.4	< 0.1	0.7	< 0.1	< 0.1
	$err_{Q90}(I)$	0.7	< 0.1	2.3	< 0.1	< 0.1
	$err_{max}(I)$	110.8	0.5	15.4	0.2	0.3
App. power errors [kVA]	$err_{mean}(S_0)$	224.3	3.6	729.6	0.4	< 0.1
	$err_{Q90}(S_0)$	530.5	7.3	1248.4	0.9	< 0.1
	$err_{max}(S_0)$	4 860.4	47.5	1466.7	4.7	0.1
Power loss errors [kW]	$err_{mean}(P_{loss})$	4.3	3.0	3.6	0.4	< 0.1
	$err_{Q90}(P_{loss})$	11.6	5.7	6.7	0.9	< 0.1
	$err_{max}(P_{loss})$	131.2	13.8	10.6	4.0	0.1

Table 2. Average performance of S30, SI, LM, VL, and NIA over 100 scenarios with *n* = 52 560. Note that the Forward-Backward load-flow solver used as a benchmark takes 25 s.
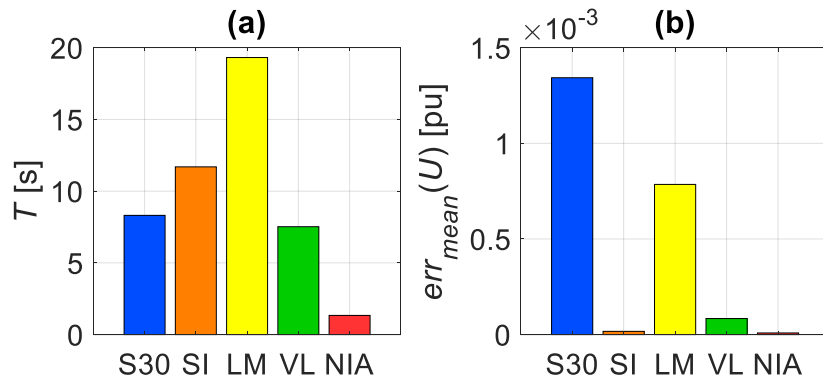

Fig. 3. a) Average of time efficiency and b) mean voltage error over 100 scenarios with *n* = 52 560 for S30, SI, LM, VL, and NIA with respect to the exact load-flow solver.

3.3.   Illustration of approximation-error impacts on a generation curtailment case

According to EN 50160 [23], 10-minute averages of any delivery point voltage in distribution networks must remain within ± 10 % of the nominal value for 95 % of the week. A good accuracy of load-flow outputs at a small time step ($\Delta T \le 10$ minutes) seems thus necessary to estimate the voltage constraints and the techno-economic performance of the alternatives to reinforcement, regardless of the network study duration. This case study aims to support this statement and to show the interest of using a precise approximation method for load-flow calculations through a simple example.

In this case study, we suppose that a 5-MW wind producer requests to be connected to a 20-kV 126-bus feeder (part of one of the networks studied in Section 3.2). This feeder has already hosted 3.39 MVA of loads, a 1-MW MV wind producer and 0.15 MW of Low-Voltage (LV) photovoltaic producers. The study of the worst-case scenario "low consumption and high production" shows that the new producer may cause overvoltage on the existing feeder. Consequently, the DSO proposes two connection

solutions to remove this voltage problem: 1) a firm-connection solution, which consists in reinforcing the feeder to totally remove the overvoltage risk (Fig. 4.a), and 2) a non-firm connection solution, which consists in curtailing the producer's real power to avoid overvoltage in real-time if need be (Fig. 4.b). For the latter solution, the DSO is assumed to contract with the producer a maximal annual curtailed energy (or curtailment duration). It is essential to assess accurately the contractual value of curtailment since, on the one hand, the producer uses it to carry out his cost-benefit analysis and to find the cheapest connection solution, and, on the other hand, the DSO has to compensate the producer for any extra curtailment. That is why the annual curtailed energy (or curtailment duration) should be sized on several profiles of load and generation. We assume here that the DSO studies 100 sets of 10-minute profiles of load/generation over one year ($n = 52\,560$) and chooses, as a contractual value, the annual curtailed energy (or curtailment duration) that removes overvoltage for 90 % of the profiles.

To illustrate the interest of a precise approximation method, the contractual curtailment energy/duration is computed using the exact load-flow solver and three of the five approximation methods presented above: S30, VL, and NIA.
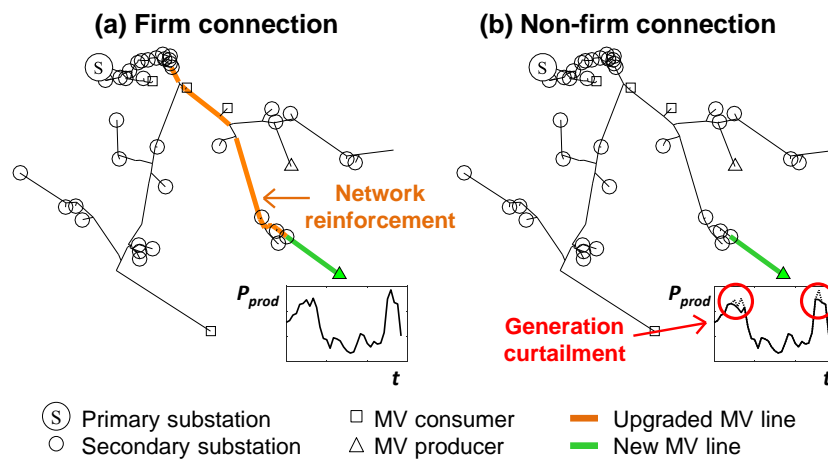

Fig. 4. Studied MV feeder with the new producer connected through: a) a firm connection (with network reinforcement), b) a non-firm connection (with generation curtailment).

The results of the study clearly show that the more accurate the approximation method is, the more precise the generation curtailment is on a 10-minute scale. Indeed, as depicted in Fig. 5, a voltage error may lead to under/overestimating the violations of high voltage limit (fixed to 1.05 pu here) and thus the curtailment duration/energy required to remove these violations.

As shown in Table 3, S30 obtains a good estimation of the mean annual constraint rate but also gets high positive and negative constraint detection errors. VL tends to overestimate the risk of overvoltage; it detects overvoltage periods very well but also finds overvoltage when there is no one. The proposed method, NIA, is effective to detect whether overvoltage occurs or not.
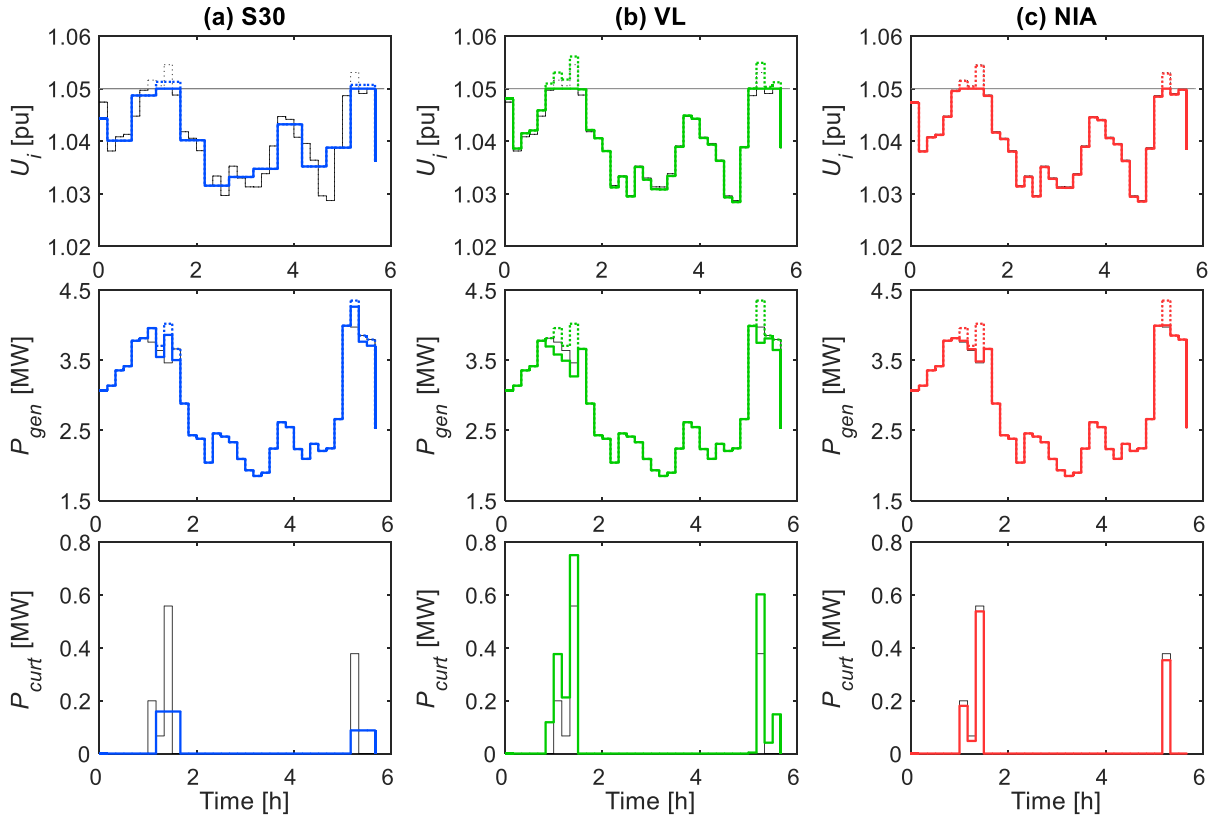
Fig. 5. Examples of voltage (top), generation power (middle) and curtailed power (bottom) computed by the exact load-flow solver (black lines on a, b and c), S30 (blue lines on a), VL (green lines on b), and NIA (red lines on c) before/after generation curtailment (dashed/solid lines).

	Ref.	S30	VL	NIA
Mean constraint rate over the year [%]	8.49	8.22	10.17	8.46
Mean share of true constraints [%]	-	86.5	100	99.6
Mean share of false constraints [%]	-	10.4	19.9	0.1

Table 3. Average overvoltage detection performance of S30, VL, and NIA over 100 load/generation profiles with n = 52 560.

Tables 4 and 5 show the curtailment values obtained from the three methods. "Error cost" refers to the share of curtailment cost due to the error on the contractual curtailment energy/duration. The closer to zero the error costs are, the more effective the approximation method is. Indeed, on the one hand, positive error costs mean that the DSO is more likely to compensate the producer at the end of each year due to unplanned extra curtailed energy. On the other hand, negative error costs mean that the producer is more likely to refuse a non-firm connection, which is actually cheaper than a firm one.

The error costs depicted in Tables 4 and 5 result from just one connection study and should be appreciated with respect to the increasing total number of RES connections required year by year. Given the expected intensive development of RES, it is therefore crucial to use an approached load-flow solver that is as fast and as accurate as possible. In this case study, NIA has proven to be more efficient than S30 and VL.

	Ref.	S30	VL	NIA
Contractual curtailed energy [MWh]	510.3	461.5	732.5	510.1
Curtailed energy error [%]	-	−9.6	+43.5	−0.1
Error cost [€ with c_{curt} = 40 €/MWh]	-	1953	−8888	8
Error cost [€ with c_{curt} = 80 €/MWh]	-	3907	−17777	15

Table 4. Curtailed energy errors and associated error costs obtained by S30, VL, and NIA to remove all the constraints for 90 % of cases, with respect to the exact load-flow solver.

	Ref.	S30	VL	NIA
Contractual curtailment duration [h]	809	779	973	806
Curtailment duration error [%]	-	−3.7	+20.2	−0.4
Error cost [€ with c_{curt} = 40 €/MWh]	-	2910	−12732	334
Error cost [€ with c_{curt} = 80 €/MWh]	-	5821	−25464	668

Table 5. Curtailment duration errors and associated error costs obtained by S30, VL, and NIA to remove all the constraints for 90 % of cases, with respect to the exact load-flow solver.

4. Discussion

4.1. Influence of the number of load-flows to approximate

As shown in Fig. 6, the interest of using a specific method to approximate the load-flow solver depends on the number of load-flows n to compute.

Firstly, it is better to perform exact load-flow calculations when n is small. Methods for approximating the load-flow solver are appropriate when building the surrogate models and computing them n times are faster than performing n exact load-flow calculations (i.e., when $G_T > 10^0$). For instance, for the 400-bus radial networks studied in Section 3.2, NIA with $n^* = 200$ becomes time-effective from $n = 500$ to 10 000 load-flows, depending on the need to use OK52 for accuracy (Fig. 6).

Secondly, the time-efficiency G_T of the methods depends on n. This stems from the allocation of computation time between the building and computing phases of the surrogate models. NIA certainly needs a great investment in the model building phase, but its surrogate models are faster than the exact load-flow solver. As a result, NIA is a suitable option when a large number n of load-flows has to be computed (Fig. 6).
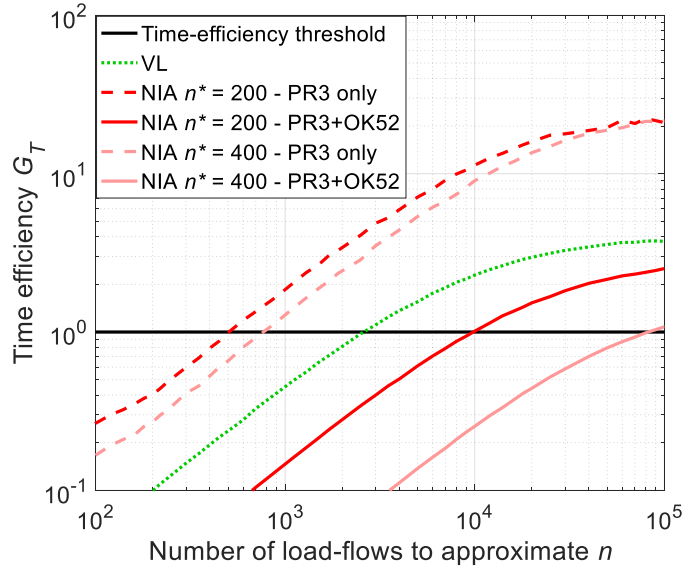
Fig. 6. Time efficiency of VL (green line), NIA with *n** = 200 and 400 using PR3 models only (dashed red lines), and NIA with *n** = 200 and 400 using both PR3 and OK52 models (solid red lines) depending on the number of load-flows to approximate, for the networks studied in Section 3.2.

4.2. Influence of the number of load-flow inputs

As power variations are generally measured at the primary substations, assumptions about power dispatching are made for distribution network planning studies. Thus the load-flow solver takes only a small number of inputs ($m = 4$ in Section 3), which makes the building of the surrogate models easier.

Considering a large number of load-flow inputs ($m \geq 20$) may strongly impact:

- the number $n*$ of exact load-flow calculations required for building the surrogate models (see Fig. 6 to compare time efficiencies obtained by NIA with $n* = 200$ and $n* = 400$);

- the number q of principal components from the PCA, which is equal to the number of surrogate models to be built;

- computation time needed to fit the parameters of the surrogate models.

If so, NIA may take much more time and/or commit higher approximation errors. Therefore, NIA is not appropriate if a specific power profile is defined at each delivery bus. However NIA is usable if loads and generators in the network under study can be modeled by about twenty different real/reactive power profiles, which is generally satisfactory for MV distribution network planning studies.

4.3. Network reconfigurations

The surrogate models obtained from NIA are valid for a given network configuration. It is therefore necessary to build as many approximate load-flow solvers as network configurations to study in planning studies. NIA is suitable when many load-flow calculations are required for a given network reconfiguration. For instance, it can speed up calculations to provide a probabilistic diagnosis of network constraints in a normal situation (as in Section 3.3) as well as the empirical probability distribution of non-distributed energy when a fault occurs in the primary substation (by taking into account different

levels of load and generation). On the other hand, NIA is not appropriate when many network configurations with few load-flow calculations have to be studied.

4.4. Non-smooth input-output mappings

NIA is accurate when load-flow variables are smooth. However, some network equipment, such as fusegears and automatic reclosers, and alternatives to reinforcement, such as Volt-VAR control and load/generation curtailment, make some load-flow outputs non-smooth. Indeed, the (de)activation of these levers to remove constraints may cause a cusp of the load-flow outputs. Smooth approximation methods have difficulty in estimating such cusps. Further work is therefore necessary to consider the non-smooth effects of these levers from the design of the surrogate models.

4.5. Parallel computing

Parallel computing is a common practice to reduce computation time. It consists in carrying out independent tasks simultaneously after dispatching them over several cores, processors or computers. The two most expensive tasks of NIA, which are the n^* exact load-flow calculations on the design of experiments and the building of the q surrogate models of the principal components, can easily be parallelized. As a result, if only a moderate number of parallel computers are available (typically smaller than n^* and q), NIA remains faster than the parallelization of the n exact load-flow calculations.

5. Conclusion

Alternative solutions to network reinforcement are now being investigated in distribution network planning studies to reduce the costs and periods for the integration of RES. However, a thorough techno-economic analysis of these solutions requires a large number of multi-period load-flow calculations, which makes it hard to implement in DSO's planning tools.

A non-intrusive approximation method has therefore been proposed to obtain fast and accurate multi-period load-flows. This method builds a surrogate model of the load-flow solver using polynomial regression and kriging, combined with Latin hypercube sampling. The case studies presented in Section 3 highlight the fact that the proposed method is more efficient for distribution network planning in presence of RES than time subsampling, line model simplification, and voltage linearization. In particular, accurate 10-minute profiles of load-flow outputs are obtained in a satisfactory computation time.

In addition to the results presented in this paper, the proposed method has also proven to be effective for the techno-economic analysis of some alternatives to reinforcement over several hundred 10-year scenarios of RES integration [15,22].

With the proposed parameter values, the method performs fast and accurate load-flow calculations for several variations of a radial network with different overall generation and power dispatching. The proposed parameter values are now to be tested on more distribution networks of different sizes,

structures, and characteristics.

Further work will focus on the extension of the proposed method to consider the non-smooth effects of alternatives to reinforcement from the design of the surrogate models.

Acknowledgments

References

[1] EEGI, Research and innovation roadmap 2013-2022, 2013.

[2] evolvDSO, D1.3: Preliminary assessment of the future roles of DSOs, future market architectures and regulatory frameworks for network integration of DRES, 2014.

[3] E. Haesen, A.D. Alarcon-Rodriguez, J. Driesen, R. Belmans, G. Ault, Opportunities for active DER management in deferral of distribution system reinforcements, Proc. IEEE PES Power Syst. Conf. Expos., Paper 000232, 2009.

[4] Z. Hu, F. Li, Cost-benefit analyses of active distribution network management part 1: annual benefit analysis, IEEE Trans. Smart Grid, 3 (3) (2012) 1067–1074.

[5] S.N. Liew, G. Strbac, Maximising penetration of wind generation in existing distribution networks, Proc. IEE Gen. Transm. Distrib. 149 (3) (2002) 256–262.

[6] V. Silva, V. Stanojevic, D. Pudjianto, G. Strbac, Benefits of storage and DSM in transmission congestion management in systems with high wind penetration, Proc. 7th Wind Integration Workshop, Paper 81, 2008.

[7] H. Mori, S. Yamada, Continuation power flow with the nonlinear predictor of the Lagrange's polynomial interpolation formula, Proc. IEEE/PES Transm. Distrib. Conf.: Asia Pacific, 2 (2002) 1133–1138.

[8] A.M. Leite da Silva, V.L. Arienti, Probabilistic load flow by a multilinear simulation algorithm, IEE Proc. C Gener. Transm. Distrib. 137 (1990) 276–282.

[9] L.F. Ochoa, C.J. Dent, G.P. Harrison, Distribution network capacity assessment: variable DG and active networks, IEEE Trans. Power Syst. 25 (2010) 87–95.

[10] A. Pagnetti, G. Delille, A simple and efficient method for fast analysis of renewable generation connection in active distribution networks, Electr. Power Syst. 125 (2015) 133–140.

[11] T. Chen, K. Hadinoto, W. Yan, Y. Ma, Efficient meta-modelling of complex process simulations with time-space-dependent outputs, Comput. Chem. Engrg. 35 (2011) 502–509.

[12] G. Jia, A.A. Taflanidis, Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment, Comput. Methods Appl. Mech. Engrg. 261–262 (2013) 24–38.

[13] F. Douarche, S. Da Veiga, M. Feraille, G. Enchéry, S. Touzani, R. Barsalou, Sensitivity analysis and optimization of sulfatant-polymer flooding under uncertainties, Oil Gas Sci. Technol. 69 (4) (2014), 603–617.

[14] T. Jolliffe, Principal Component Analysis, second ed., Springer, New York, 2002.

[15] H. Dutrieux, Méthodes pour la planification pluriannuelle des réseaux de distribution ; application à l'analyse technico-économique des solutions d'intégration des énergies renouvelables intermittentes, Ecole Centrale de Lille, Villeneuve d'Ascq, France (2015) (Ph. D. Dissertation).

[16] M.E. Johnson, Minimax and maximin distance designs, J. Stat. Plan. Infer. 26 (1990) 131–148.

[17] A. Forrester, A. Sobester, A. Keane, Engineering Design via Surrogate Modelling: A Practical Guide, Wiley, Chichester, 2008.

[18] M.L. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York, 1999.

[19] E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, Cambridge, 2006.

[20] D. Shirmohammadi, H.W. Hong, A. Semlyen, G.X. Luo, A compensation-based power flow method for weakly meshed distribution and transmission networks, IEEE Trans. Power Syst. 3 (2) (1998) 753–762.

[21] J. Bect, E. Vazquez, et al. (2014). STK: a Small (MATLAB / GNU Octave) Toolbox for Kriging. Release 2.1: http://kriging.sourceforge.net

[22] H. Dutrieux, G. Delille, B. Francois, G. Malarange, Assessing the Impacts of Distribution Grid Planning Rules on the Integration of Renewable Energy Sources, Proc. IEEE PowerTech, Paper 464270, 2015.

[23] CEN, CSN EN 50160 ed.3 - Voltage characteristics of electricity supplied by public electricity networks, 2003.