Digital Object Identifier

Optimal Power Flow with Regression-based Small-signal Stability Constraints

FRANCESCA ROSSI[1], EDUARDO PRIETO-ARAUJO[2], (Senior Member, IEEE,),MARC CHEAH-MANE[3], (Member, IEEE,), ORIOL GOMIS-BELLMUNT[4], (Fellow, IEEE,).

[1]Universitat Politecnica de Catalunya, Barcelona, 08028 Spain (e-mail: francesca.rossi@upc.edu)
[2]Universitat Politecnica de Catalunya, Barcelona, 08028 Spain (e-mail: eduardo.prieto-araujo@upc.edu)
[3]Universitat Politecnica de Catalunya, Barcelona, 08028 Spain (e-mail: marc.cheah@upc.edu)
[4]Universitat Politecnica de Catalunya, Barcelona, 08028 Spain (e-mail: oriol.gomis@upc.edu)

Corresponding author: Francesca Rossi (e-mail: francesca.rossi@upc.edu).

ABSTRACT This paper is focused on the design of a Small-signal Stability Constrained Optimal Power Flow (SSC-OPF), as a tool for power systems planning and operation, aimed to support the integration of Converter Interfaced Generators (CIGs) in modern power systems. The small-signal stability constraints are formulated through regression functions, which act as a data-driven surrogate model of the conventional state-space linear model eigenvalues. Such regressions are aimed to calculate the value of a numerical stability indicator, related to the oscillations damping. After a qualitative and quantitative comparative analysis among linear and non-linear regression techniques, Multivariate Adaptive Regression Splines proves to be suitable for this application.

The SSC-OPF implementation aims to verify the possibility of operating the system in optimal and stable conditions. To this end, two strategies are implemented and compared. One is by computing the best generator dispatch, the other includes the possibility of tuning controllers' parameters.

The two strategies are tested on a 9-bus power system, with two objective functions exemplifying the use of the SSC-OPF for power system operation and analysis. The possibility of tuning a Voltage Source Converter frequency droop controller effectively ensures an optimal and stable solution.

INDEX TERMS Power Systems Operation, Power Systems Analysis, Small-signal Stability, Multivariate Adaptive Regression Splines

I. INTRODUCTION

OPTIMAL power flow (OPF) studies are used to determine the best system operation that complies with the steady state magnitudes limits of grid components, such as buses, lines, or generators [1]. OPF computation is employed for power system operation, from day-ahead markets up to real-time operation, and for system planning, from mid-term to short-term optimal expansion planning [1]. The OPF formulation typically considers static stability constraints. To check that the dynamic stability limitations are not violated, System Operators (SOs), which are in charge of the secure and stable power system operation, have to conduct separate studies. In terms of small-signal stability analysis (the main focus of this study), SOs need to verify that the operating point reached by the system after a small disturbance (as a small load variation) does not show insufficient oscillation damping [2]. Moreover, in modern power systems, characterized by increased penetration of power electronics converters, oscillatory problems caused by converter controllers' interactions with other elements of the system could be observed [3], [4]. An example of these phenomena is the low-frequency and subsynchronous oscillations observed in the Texan and Chinese power systems, due to wind integration in a weak system and mainly affected by the Phase-Looked-Loop (PLL) control [5]. Therefore, considering small-signal dynamics in power system operation [6], [7] and planning studies [8], [9], and in converters' control tuning studies [10] is a relevant issue. Currently, in all the mentioned studies, small-signal stability assessment is typically carried out through separate analyses. In operation studies, the generation dispatch, resulting from market clearing, is subjected to a stability assessment, performed by SOs. Then, if it results

unstable, a re-dispatch is computed by SO [6]. The choice of which generators participate in re-dispatch is based on stability sensitivity analyses, considering, therefore, generators with a major impact on stability [11], [12]. The disadvantage of this procedure is that, as in the case of grid congestion problems [13], a re-dispatch implies an extra cost for the SO. In short-term planning studies, conventionally, a copper plate network model is used. If the small-signal assessment is considered, it is limited to the analysis of selected worst-case scenarios. They are chosen based on historical considerations about weather and loading. However, for integrating intermittent and uncertain renewable generators, future power systems planning has to include stability assessment [8], scanning a large number of operating scenarios with a sustainable computing burden [14]. Concerning converters' control tuning studies, control parameters can be modified in operation and planning studies to ensure stability. By way of example, consider the studies related to the integration of converters for High Voltage Direct Current (HVDC) links. If during operation, it is detected that the converter installation leads to instability in the network, the control algorithms need to be redesigned. This requires separate studies, carried out either by the converter manufacturer, due to intellectual property issues, or by the SO, if provided, at least, with a converter black-box model [10].

To integrate optimization and small-signal assessment studies, a small-signal stability-constrained OPF (SSC-OPF) is required, i.e. an OPF formulation with a small-signal stability limitation. Consider that the conventional mathematical formulation used for small-signal stability assessment is based on the eigenvalues analysis of the linearized system of the Differential and Algebraic equations (DAEs) that represent a power system [15]. Including such formulation into an OPF is complex [6]. Therefore, the use of approximation functions has been proposed. In [16], linear eigenvalues sensitivities are used to formulate the small-signal stability constraints. They are linear approximations of the eigenvalues sensitivities to small variations of generators' power injection. For large power systems, computing all the sensitivities-based functions is time-consuming. An emerging solution involves the use of data-driven Machine Learning (ML)-based algorithms [17]. Such ML models act as surrogate models of the exact mentioned small-signal stability mathematical formulation and allow the inclusion of the dynamic stability constraints into the optimization problem. The use of several ML algorithms has been proposed in the literature. In [18], a Neural Network (NN) is used for mapping the space of the feasible, small-signal stable and 'N-1' secure OPF solutions. Then, a mixed-integer reformulation of the NN is required to include the dynamic constraints in the OPF. In [6], the small-signal stability constraints are formulated as the conditional line flow limitations that lead to instability. They are estimated through the rules identified by a Decision Tree (DT) algorithm trained to predict the stability. In [19], the SSC-OPF uses Support Vector Machine (SVM) to obtain a linear kernel that describes the small-signal stability boundary of the system as a function of relevant generator bus voltages. In [20], the Extreme Gradient Boosting DT regression is used to predict the damping ratio of the system operating points. It is employed in the optimization problem, not as a constraint but in the objective function. Such optimization is aimed to compute the generator's re-dispatch that maximizes the damping ratio, in case of initial unstable dispatch.

In the mentioned works, Mixed Integer Linear Programming (MILP) is used for the OPF formulation, as the dynamic constraints lead to the presence of discrete variables. Handling discrete variables in an OPF is challenging and, for the solution of a large power system, not tractable [21]. On the contrary, the present study is intended to propose an SSC-OPF formulation that can employ a state-of-the-art optimization algorithm, such as the interior-point algorithm [22]. It is widely employed in large-scale nonlinear programming optimization. Interior-point algorithm requires that the nonlinear equality and inequality constraints included in the optimization formulation have to be expressed by smooth functions [22]. Therefore, to include the small-signal stability constraint in the OPF it is necessary to express such constraint by a continuous and derivable analytic expression. The linear models are among the regression techniques that fulfill such characteristics. However, such models perform poorly when the relationship between the independent and dependent variables is non-linear. In this case, the use of a non-linear regression is required. Among the non-linear regressions, Multivariate Adaptive Regression Splines (MARS) has the characteristics mentioned above. In fact, MARS provides a continuous and derivable analytic expression of the regression, suitable to be included in an optimization problem. It is also appropriate for fitting non-linear targets, in a high-multidimensional space [23]. MARS is often employed in computer vision and image processing. In the field of power systems, it has been used to predict the critical clearing time (CCT), as an indicator of transient stability [24]. To the best of the authors' knowledge, it has not been employed in small-signal stability analysis.

This paper proposes a methodology for developing an SSC-OPF. The contributions are:

- The formulation of the small-signal stability constraint through a regression function. It has to provide the fitting of a small-signal stability numerical indicator as a function of the OPF variables. Linear regressions and MARS compared. It is shown as MARS, which can capture the non-linear dependencies between the inputs and the output quantities, outperforms the linear models in terms of accuracy and robustness.
- An SSC-OPF algorithm that, for any objective function, can directly compute an optimal and small-signal stable solution. To achieve stability, the implementation of two strategies is presented. One possibility proposes to compute the generation dispatch (GD) whereas the other combines the GD and the tuning of controllers' parameters (such as controllers' gain, time constant, etc.).

The small-signal stability indicator is based on the damping ratio of the critical eigenvalues. In this work, it is shown how to properly handle this indicator, such that it results suitable to be fit by a regression. Concerning the two problem implementations, they are indicated in this paper as the GD-driven case and the GD and controllers tuning-driven (GD&CT-driven) case, respectively. In the GD-driven case, the OPF variables are only the operating point variables (buses voltage magnitudes and phases, active and reactive powers), computed by Power Flow (PF). In the GD&CT-driven case also the controllers' parameters are included in the OPF variables. Therefore, in the GD-driven case, the solution of the SSC-OPF provides all the PF quantities that ensure optimal and stable operation, for a certain power demand. In the GD&CT-driven case, the SSC-OPF solutions provide also the best tuning of the parameters of the converters' controllers involved in the analysis.

The proposed SSC-OPF is intended to be applied both in operation studies and as an analysis tool to be included in planning studies. Therefore, in this work, both problem implementations are shown using two objective functions. For exemplifying an operation study the minimum transmission active power losses objective function is used. For power system analysis, the objective function used is minimizing the active power injection by non-renewable, synchronous generation. In all the cases a 9-bus power system case study is employed.

II. SMALL-SIGNAL STABILITY CONSTRAINED OPTIMAL POWER FLOW

A. GENERAL FORMULATION

Traditionally, OPF solutions consider static limitations of the network, whereas the stability assessment has to be studied in subsequent analyses. On the contrary, the proposed SSC-OPF is a single-step tool, for direct computation of the optimal, static, and small-signal stable OPF solution. Fig. 1 compares how the conventional and the proposed tools act when the OPF solution results unstable. It is worth noticing that a re-dispatch, aimed to reach stability, does not consider the OPF objective function minimization. Therefore, the employment of conventional tools might result in a sub-optimal solution. For completeness, consider that if the OPF solution results stable, then OPF and SSC-OPF solutions should coincide.

Focusing on the small-signal stability of power systems, it is usually assessed based on the eigenvalues analysis of the system state-space linear model [15]. Including the complete power system linear model into the OPF to consider stability restrictions would be intractable [18]. For large power systems, in particular, the state-space linear model involves hundreds to thousands of differential equations, states, and eigenvalues, which is impossible to digest as part of the optimization problem. Therefore, in this work, it is proposed to fit regression functions that act as surrogate models. These functions can predict the values of small-signal stability indicators (as the eigenvalues real parts, the eigenvalues damping ratios, etc.) as a function of the optimization decision

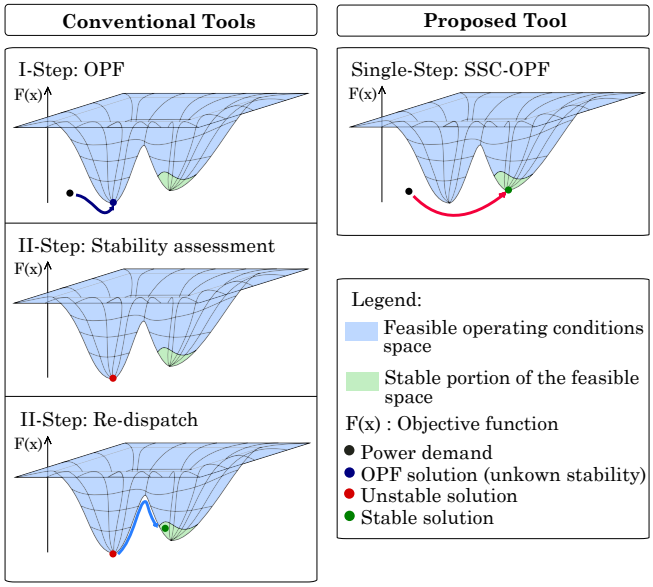


FIGURE 1: Comparison between the conventional multi-step procedure and the proposed single-step procedure for obtaining an optimal, small-signal stable OPF solution.

variables. The analytic expressions of such regressions are included in the OPF formulation as inequality constraints. That is, imposing the outcomes of the regressions to assume values below the threshold marking instability. Thus, the general SSC-OPF formulation is detailed as follows

$$\min_{x} \quad F(x) \quad \text{s.t.} \tag{1}$$
$$w_i(x) = 0 \quad i = 1, ..., N_w \tag{2}$$
$$z_j(x) \leq 0 \quad j = 1, ..., N_z \tag{3}$$
$$s_k(x) < \tau_k \quad k = 1, ..., N_s \tag{4}$$

where

- x collects the OPF variables, including the demand, state, and control variables [25].
- $F(x)$ in (1) is the optimization objective function.
- $w(x)$ in (2) is a set of N_h non-linear and linear equality constraints, formulating the PF equations and the demand variables constraints, respectively.
- $z(x)$ in (3) is a set of N_g non-linear and linear inequality constraints, imposing the static stability constraints, i.e. the apparent power flow limits, the buses voltage limits, and the generators active and reactive power limits, according to their capacity.
- $s(x)$ in (4) is a set of N_s small-signal stability constraints, formulated as nonlinear inequality constraints. Each k-th equation in (4) corresponds to the analytic expression of a small-signal stability indicator. Imposing the τ_k upper thresholds over indicators, small-signal stability is enforced.

The SCC-OPF problem expressed in equations (1)-(4) is a non-linear programming problem. To solve such opti-

mization, it is proposed to use the interior-point algorithm which combines Sequential Quadratic Programming (SPQ) and trust regions technique to handle the constraints' nonlinearities and the problem's non-convexity, respectively [22].

B. PROBLEM DEFINITION

In the scope of this study, the proposed SSC-OPF is intended to be applied to operate networks with a high presence of power electronics assets. This is reflected in the choice of the OPF variables and of the optimization objective function.

Consider a n-bus power system. Define $\mathcal{N}$ as the set of the system buses, $\mathcal{L}$ as the set of loads, $\mathcal{G}$ as the set of generators, which includes the sets of non-renewable and renewable synchronous generators (SGs), $\mathcal{S}_{NR}$ and $\mathcal{S}_R$, and the set of converter interfaced generators (CIGs), $\mathcal{C}$. Given that the small-signal stability is influenced by the operating point and by the controllers implemented in the system devices, the SSC-OPF is tested by implementing two strategies:

- GD-driven case: a stable OPF solution is achieved by changing the GD. In this case, the OPF variables x are:

$$x = [(V_i, \theta_i)_{i\in\mathcal{N}}, (P_i, Q_i)_{i\in\mathcal{L}}, (P_i, Q_i)_{i\in\mathcal{G}}] \quad (5)$$

  where V and θ are the bus voltage magnitude and phase, P and Q are the active and reactive power.
- GD&CT-driven case: a stable OPF solution is achieved by changing both the GD and the converters CT. In this case, the OPF variables x are:

$$x = [(V_i, \theta_i)_{i\in\mathcal{N}}, (P_i, Q_i)_{i\in\mathcal{L}}, (P_i, Q_i)_{i\in\mathcal{G}}, K] \quad (6)$$

  where, with respect to the OPF variables of the previous case, K is added. It collects the equipment's controllers' gains or parameters, of which the possibility of changing their tuning is considered.

Concerning the objective function, two options are considered: minimizing the transmission lines' active power losses and minimizing the power injected by non-renewable SGs.

- Power losses minimization: it is a common objective function used to optimize power system operation. For a power system with N_L transmission lines, the objective function is

$$F(x) = \sum_{m=1}^{N_L} G_{m,ij}[V_i^2 + V_j^2 - 2V_iV_j\cos(\theta_i - \theta_h)] \quad (7)$$

  where the subscripts i and j indicate the terminal buses of the m-th line, with $i, j \in \mathcal{N}$, $G_{m,ij}$ is the conductance of the m-th line, V_i, V_j and θ_i, θ_j are the voltage magnitude and voltage angle at bus i and j, respectively.
- Non-renewable SGs power injection minimization: this objective function is adopted for exemplifying the use of the SSC-OPF as a tool for power system analysis. It is expressed as:

$$F(x) = \sum_{i\in\mathcal{S}_{NR}} P_i \quad (8)$$

C. SMALL-SIGNAL STABILITY INDICATOR

Different quantities might be used as targets of the small-signal stability constraint regressions, as, among others, the real parts or the damping ratios of the power system linear model eigenvalues. Here it is proposed to use a simple indicator that can condense information about the oscillatory mode damping. Considering a system whose state-space linear model eigenvalues analysis returns N_e eigenvalues $(\lambda_i)_1^{N_e}$. For each of them, the damping ratio ξ is calculated as

$$\xi_i = \frac{-\Re(\lambda_i)}{\sqrt{\Re(\lambda_i)^2 + \Im(\lambda_i)^2}} \quad i = 1, ..., N_e \quad (9)$$

To condense such information in a unique indicator, the Damping Index (DI), calculated as in (10), can be used [26]. As follows from Lyapunov's first method [15], a $DI \geq 1$ indicates system instability.

$$DI = 1 - \min\{\xi_1, ..., \xi_{N_e}\} \quad (10)$$

For the formulation of the SSC-OPF, one possibility is to express the small-signal stability constraint by fitting a unique regression model, aimed to predict the unique, global system DI in (10). However, in the scope of a data-driven regression-based constraint formulation, the use of a unique DI implies some disadvantages. Even though the linearity of the small-signal stability problem, the DI might experience high nonlinear changes, hard to capture with a regression function. For this reason, it is proposed to compute the DI for groups of eigenvalues that are evaluated as interesting and critical for the small-signal analysis. The eigenvalues are grouped according to their location in the modal map plane and to their trajectories. The DIs calculated for such groups of eigenvalues have a smoother behavior, compared with the one of the global DI, leading beneficial impact on the accuracy of the trained regressions. To evaluate the stability it is sufficient to take into account only the critical eigenvalues, which are the ones that have high mobility, in relation to changes in PF variables or control parameters, and cross the imaginary axis of the modal map plane. This allows the exclusion of eigenvalues with negative real parts close to zero but with low mobility, as they do not provide valuable information about stability. One regression for each critical eigenvalues group is trained and the same number of small-signal inequality constraints are added in the SSC-OPF formulation (4). In this paper, it is assumed that DI has to be strictly lower than 1 ($\tau_k = 1$, $k = 1, ..., N_s$). However, this value might be adjusted, and, by imposing a lower threshold, improved dynamic performance can be obtained.

III. METHODOLOGY

In this section, the methodology for developing the proposed SSC-OPF is presented. The main steps are summarized in Fig. 2. Steps A-D are the phases executed *off-line* and are required to train the DIs regressions models. Step E is the application of SSC-OPF, as it would be used in operation studies or analysis tools.

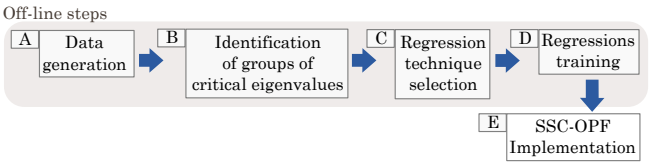


FIGURE 2: Scheme of the proposed methodology

A. DATA GENERATION

Data required for training the regressions and, therefore, formulating the SSCs have to gather information about the small-signal stability behavior of the system for different operating conditions. Such data have to be generated by computation because the amount and quality of data that might be collected during a system's actual operation are typically not sufficient to implement ML-based tools [27]. Even though the proliferation of smart grids is increasing the number and the sampling rate of the measurements, such data do not capture the information related to unstable operating conditions, which are not likely to occur in actual operation [27]. Therefore this information is obtained by performing the small-signal stability assessment, through the use of the mathematical tools conventionally employed for this analysis. Such tools are the PF calculation and the state-space linear model representation [15]. For each equilibrium point, computed by PF and provided to the state-space linear model, the eigenvalues analysis of the system characteristics matrix reveals if the system is stable, as mentioned in Section II-C.

To train regressions with good predictive performances, the selection of the instances to be included in the database (DB) assumes a relevant role. Training data have to extensively describe the small-signal power system behavior in the whole operable region but with a finer granularity into the stability margin, [28]. The choice of exploring the whole power system operable region, and not only the area that would be involved by the solutions of the OPF, is also motivated by the objective of generating a unique DB, that can be employed to train the DI regressions for any SSC-OPF problem, regardless of the problem objective function. However, power system operable space exploration is an NP-hard problem [27]. Therefore, a strategy aimed at avoiding curse-of-dimensionality is needed. This work uses the data generation strategy proposed in [29]. Such a strategy combines the use of Latin Hypercube Sampling (LHS) [30] and of the Entropy function [31]. The first is used for obtaining a limited number of samples, representative of the space under investigation. The Entropy function is used to detect the portions of the operable region affected by the stability margin and to bias the sampling process.

Hence, the sampling strategy is applied in a search space whose dimensions are defined according to the input variables required by the computing tools. Such variables are, in the GD-driven case, the known quantities of the PF, according to its formulation (i.e. how the buses are defined). For the GD&CT-driven case, additional dimensions have to be considered, related to the controllers' parameters for which it is intended to apply a different tuning.

Finally, the DB collects the variables describing each sampled instance and the corresponding stability assessment result. The former are the PF quantities and the control parameters. These variables correspond to the ones collected in x (5)-(6). The latter are the damping ratios (9).

B. IDENTIFICATION OF CRITICAL EIGENVALUES GROUPS

Data collected in the DB are post-processed to obtain the DIs of the groups of critical eigenvalues. Therefore, first, such critical eigenvalues have to be identified. According to the number of eigenvalues N_e, which directly increase with the size of the power system, it might be possible to show the eigenvalues in the modal plane. This way, by a visual inspection, the groups of critical eigenvalues can be identified straightforwardly. An alternative method can be grouping the eigenvalues based on their position in the modal plane. To this aim, clustering algorithms, such as K-means clustering [32] or Gaussian Mixture Models clustering [33], can be used. Then, the groups of eigenvalues are identified as critical according to their mobility, which can be measured by the coefficient of variation of the real part of the eigenvalues, and whether eigenvalues belonging to the group have a positive real part. Finally, for such groups of eigenvalues, the DIs are computed as in (10).

C. REGRESSION TECHNIQUE SELECTION

The regression aim is to compute the DIs as a function of the OPF variables x. Many regression techniques can fulfill this task but they do not perform in the same manner in terms of accuracy, scalability, robustness, and interpretability. The choice of which regression technique to adopt should consider the nature of the data to be fitted and the application in which the regression will be used. In this study, the training DB has a multi-dimensional scattered, not-ordered structure. The optimization problem in which the regression will be employed suggests the use of a technique that returns a continuous and derivable analytic expression. The above-mentioned issues exclude the use of interpolation techniques, whether they are based on gridded-structured data or scattered-structured data [34]. The first does not agree with the DB structure and both turn out to be computationally unfeasible for high-dimensional problems. Among the regression techniques, liner models might be feasible for this application as they fit the characteristics mentioned above. However, they cannot capture non-linearity in the target behavior, which might lead to poor accuracy. Some popular linear regression models are linear regressions (LR), Ridge, and Lasso regressions. Concerning non-linear regressions, most of them have been discarded as they do not provide an analytic expression of the fitted model. Therefore, they would require the inclusion of discrete variables and a mixed-integer formulation of the optimization, as proposed in [6], [18], where NNs and DT are used. Such kind of formula-

tion complicates the resolution of the optimization problem, especially for large power systems [21]. Among non-linear regressions, MARS is identified as a feasible possibility. Such a model is suitable for non-linear high-dimensional problems and provides an analytic expression of class $\mathcal{C}^1$ [23].

Given the mentioned considerations, a qualitative comparison of the interpolation and regression techniques is provided. The results are summarized in Table 1. Then, among the feasible techniques, i.e. the linear regression models and MARS, a quantitative comparison has to be performed, measuring the accuracy of fitting the target, to select the regression to be employed.

D. REGRESSION TRAINING

Generally speaking, both MARS and the linear models present a training procedure based on minimizing the Residual Sum-of-Squares (RSS) (12) between the exact and predicted (i.e. outcomes of the regression) values of the target. Therefore, the training of a generic k-th regression $s_k(x)$, with $k = 1, ..., N_s$, for the formulation of one of the N_s small-signal stability constraints, is expressed as

$$s_k : x \longrightarrow DI_k \quad \text{s.t.} \tag{11}$$

$$\min \sum_{i \in DB} (DI_{k,i} - s_k(x_i))^2 \tag{12}$$

where x are the OPF variables, set according to the OPF problem formulation, DI_k is the DI of the k-th critical group of eigenvalues, and DB collects the DB training instances.

Next, more details about the training process and the form of the resulting analytic expression of the regression models considered in this study are provided. In particular, MARS, LR, Ridge, and Lasso regressions are presented.

1) MARS

The MARS expression of the k-th small-signal stability constraint $s_k(x)$ is a non-linear function, obtained as a combination of piece-wise linear functions:

$$s_k(x) = \sum_{b=0}^{B} \beta_b h_b(x) \tag{13}$$

where B is the total number of terms added to the model, β are constant coefficients evaluated through the application of the least-square criterion, and $h(x)$ are single basis functions or products of more basis functions. The basis functions involved are the *hinge functions*, a pair of reflected piecewise linear splines. For the j-th variable in x, the form of the *hinge functions* is $(x_j - t)_+$ and $(t - x_j)_+$, which can be equivalently expressed as $\max(0, x_j - t)$ and $\max(0, t - x_j)$, respectively. In such expressions, t is one of the values assumed by x_j in the training data set, called a knot.

The model's building process involves two phases. The first is a forward phase, in which the basis functions that lead to an error decrease are added to the model. At the end of this phase, a large and complex model is obtained, which tends to overfit the training data. Therefore, the second phase is a backward pruning that is aimed at improving the generalization ability of the model by removing those terms that decrease its accuracy. Next, the forward and backward phases are described as they have been implemented in [23].

a: Forward phase

Each iteration of the forward phase selects the term to be added to the model to decrease the RSS as in (12). As mentioned above, the terms to be added to the model are single reflected pairs of splines basis functions or products of them, multiplied by constant coefficients estimated by the least-square criterion. Before describing the forward procedure, let's define the set of candidate pairs of basis functions $\mathcal{H}$. It is made up of $2|x||DB|$ functions, where $|x|$ is the number of input variables and $|DB|$ is the number of samples contained in the training DB, i.e. the observed points. Therefore, for each variable, a pair of linear splines is computed using as knots all the observed points of the DB.

At iteration zero, a constant hinge function $h_0 = 1$ is added to the model. The first iteration takes one-by-one the pairs of basis functions from $\mathcal{H}$, multiply them for h_0 obtaining a pool of $|x| \cdot |DB|$ candidate terms to be added to the model, with the form

$$\begin{aligned} h_1(x) = \beta_{1,(j,i)} h_0(x)(x_j - t_{j,i})_+ + \\ + \beta_{2,(j,i)} h_0(x)(t_{j,i} - x_j)_+ \\ j = 1, ..., |x| \quad i = 1, ..., |DB| \end{aligned} \tag{14}$$

that with $h_0(x) = 1$ is equal to

$$\begin{aligned} h_1(x) = \beta_{1,(j,i)}(x_j - t_{j,i})_+ + \beta_{2,(j,i)}(t_{j,i} - x_j)_+ \\ j = 1, ..., |x| \quad i = 1, ..., |DB| \end{aligned} \tag{15}$$

Among all the candidates in $\mathcal{H}$, the $h_1(x)$ added to the model is the one with the minimum RSS. Then, the pairs of basis functions in $h_1(x)$ are removed from $\mathcal{H}$.

The successive iteration takes one-by-one the hinge functions in $\mathcal{H}$ and multiplies them for the functions already added to the model, i.e. $h_0(x)$ and $h_1(x)$. The terms candidate to be $h_2(x)$ are

$$h_{2,1}(x) = \beta_{3,(j,i)}(x_j - t_{j,i})_+ + \beta_{4,(j,i)}(t_{j,i} - x_j)_+ \tag{16}$$

and

$$h_{2,2}(x) = \beta_{3,(j,i)} h_1(x)(x_j - t_{j,i})_+ + \beta_{4,(j,i)} h_1(x)(t_{j,i} - x_j)_+ \tag{17}$$

with $(x_j - t_{j,i})_+$ and $(t_{j,i} - x_j)_+ \in \mathcal{H}$ and considering $h_0(x) = 1$. The term selected to be added to the model is the $h_2(x)$ that leads to the highest error decrease.

Such a procedure is repeated until a pre-set maximum number of terms is reached or until the addiction of terms has a negligible effect on the error decrease.

TABLE 1: Qualitative comparison of some regression and interpolation techniques which might be used for the small-signal stability constraints formulation.

		Expression of class C^1	Fitting scattered data	Fitting in multi-dimensional space	Fitting non-linear target
Interpolation	Scattered	✓	✓	✗	✓
	Gridded	✓	✗	✗	✓
Non-linear Models	NNs	✗	✓	✓	✓
	DTR	✗	✓	✓	✓
	MARS	✓	✓	✓	✓
Linear Models	LR	✓	✓	✓	✗
	Ridge	✓	✓	✓	✗
	Lasso	✓	✓	✓	✗

b: Backward phase

The model at the end of the forward phase is a linear combination of a large number of terms, here indicated as B_{max}. To select which terms have to be pruned, the following procedure is applied: the number of terms in the model (π) is reduced from B_{max} to 1, eliminating one term per iteration. At each n-th iteration n, the term to be removed is selected after having tested all the possible combinations of $\pi = B_{max} - n$ terms. The test consists of measuring the error of the model in predicting the exact value of the training DB. The strategy used to estimate such an error is the generalized cross-validation (GCV) (18) that considers the RSS of the $s_k^{\pi}(X)$ model, penalized by a factor that accounts for the complexity of the model. This penalty factor, i.e. the denominator of the expression in (18) is a function of the effective number of parameters involved in the model $M(\pi) = r + c(B_{max} - 1)$, where r is the number of linear independents basis functions of the model and $c = 3$ is the number of parameters involved in the optimal knots selection during the forward phase [23].

$$GCV(\pi) = \frac{\frac{1}{|DB|}\sum_{i=1}^{|DB|}(DI_{k,i} - s_k^{\pi}(x_i))^2}{(1 - M(\pi)/|DB|)^2} \tag{18}$$

After having computed all the iterations, the best model, i.e. the one that has the highest $GCV(\pi)$, is identified.

Such a backward pruning phase, based on the GCV criterion, makes the MARS model able to automatically perform *Features Selection* and provide information about *Features Importance*. The GCV estimates the error reduction of whether a feature is included in the model. The highest the GCV associated with the inclusion of a feature, the highest its importance within the model. On the other side, pruned features have a null GCV. Therefore, the importance of the unpruned features can be estimated according to the value of GCV associated with them and valuable information about which are the physical quantities having a major impact on the regression target can be drawn. It is worth noticing that to properly apply MARS with *Feature Importance* purpose, it is necessary to first remove correlation among the training variables.

2) Linear Models

a: Linear Regression

The LR obtained as k-th small-signal stability constraint has the form

$$s_k(x) = \beta_0 + \sum_{i=1}^{N_x} \beta_j x_j \tag{19}$$

where N_x is the number of input variables x. The fitting of the βs coefficients of the k-th model is performed by minimizing the RSS:

$$\beta^{LR} = \arg\min_{\beta}\{\sum_{i\in DB}(DI_i - \beta_0 - \sum_{j=1}^{N_x}\beta_j x_j)^2\} \tag{20}$$

Such models can suffer the effect of variables collinearity which leads to high model variance [35]. To alleviate this problem, shrinkage methods can be used, which consist of calculating the coefficients of the model by minimizing a penalized RSS. Linear models that exploit this possibility are the Ridge and Lasso regressions.

b: Ridge Regression

In Ridge regression, an L_2 penalty is involved. Therefore, the βs coefficients are calculated by minimizing the following expression:

$$\beta^{Ridge} = \arg\min_{\beta}\{\sum_{i\in DB}(DI_i - \beta_0 - \sum_{j=1}^{N_x}\beta_j x_j)^2 + \lambda\sum_{j=1}^{N_x}\beta_j^2\} \tag{21}$$

The parameter λ controls the amount of shrinkage. The larger the value of λ, the lower the values of β^{Ridge}s, which increase their robustness [35].

c: Lasso Regression

In Lasso regression, the βs coefficients are calculated as

$$\beta^{Lasso} = \arg\min_{\beta}\{\frac{1}{2}\sum_{i\in DB}(DI_i - \beta_0 - \sum_{j=1}^{N_x}\beta_j x_j)^2 + \lambda\sum_{j=1}^{N_x}|\beta_j|\} \tag{22}$$

where, with respect to the Ridge regression, the L_2 penalty is replaced by the L_1 penalty. Such penalty term has the effect of forcing some of the βs coefficients to be equal to zero,

performing an automatic feature selection. In this case, the parameter λ controls the sparsity of the coefficients [35].

E. SSC-OPF IMPLEMENTATION

The model implementation is the application of the SSC-OPF for computing the optimal and stable solution of a new power system operation scenario, included in the operable region investigated during the data generation phase but not coincident with any of those instances. Therefore, for a generic power demand $(P_i, Q_i)_{i\in\mathcal{L}}$, the SSC-OPF returns the solution x^{OPT} and the prediction of the DIs associated with that solution, $s(x^{OPT})$. If the GD-driven case is implemented

$$x^{OPT} = [(V_i, \theta_i)^{OPT}_{i\in\mathcal{N}}, (P_i, Q_i)_{i\in\mathcal{L}}, (P_i, Q_i)^{OPT}_{i\in\mathcal{G}}] \quad (23)$$

whereas, for the GD&CT-driven case

$$x^{OPT} = [(V_i, \theta_i)^{OPT}_{i\in\mathcal{N}}, (P_i, Q_i)_{i\in\mathcal{L}}, (P_i, Q_i)^{OPT}_{i\in\mathcal{G}}, K^{OPT}] \quad (24)$$

The variables $(P_i, Q_i)_{i\in\mathcal{L}}$, as demand variables, maintain a constant value [25]. The solution of the remaining variables is computed by solving the PF equations in (2), according to the objective function in (1) and complying with the static and dynamic stability constraints in (3)-(4). Therefore, $(P_i, Q_i)^{OPT}_{i\in\mathcal{G}}$ corresponds to the optimal GD and K^{OPT} to the optimal CT.

IV. CASE STUDY

In this Section, the proposed methodology is applied to an example power system. In Section IV-A the power system is described. Then, two cases are studied: the GD-driven case, in which the OPF variables are only the operating point variables, and the GD&CT-driven case, which includes among the OPF variables also a control parameter. By way of example, the converter frequency droop controller characteristic is considered. For each case, two SSC-OPFs are formulated, changing the objective function of the problem. As introduced in Section II-B, the objective functions are minimizing the power losses and the non-renewable SGs power injection.

A. POWER SYSTEM TEST CASE

The proposed methodology is applied to a modification of the IEEE 9-bus power system example [26], shown in Fig. 3. It is composed of two SGs, one Voltage Source Converter (VSC), and eight loads. It is assumed that the SGs represent in an aggregated manner non-renewable thermal-based generators. Whereas, the VSC represents, in an aggregated manner, renewable CIGs. Therefore, renewable SGs are not considered. According to the results in [26], which discuss the impact of the SGs rated power on the system stability, the rated power of the SGs installed at bus 6 (SG_1) and bus 5 (SG_2) has been set equal to 50 MVA and 30 MVA, respectively. They have been modeled with their corresponding exciters, governors, and frequency droop controllers. The CIG has a rated power equal to 500 MVA and, as a conventional VSC, it includes a current control, a voltage control, an active power control, a PLL, and a frequency droop control. The impact of the frequency droop controller on the small-signal stability of the system is studied. Therefore, given the following frequency droop controller formulation

$$P_{CIG} = P^*_{CIG} + k_{fd}(f^* - f) \quad (25)$$

where P_{CIG} is the CIG active power output, f is the network frequency, P^*_{CIG} and f^* are the corresponding reference values, and k_{fd} is the controller gain. Such gain is changed by varying the value of the frequency droop characteristic R, with

$$k_{fd} = \frac{S_{CIG,rated}}{f^* \cdot R}, \quad R = \frac{\Delta f(\%)}{\Delta P(\%)} \quad (26)$$

where $S_{CIG,rated}$ is the CIG rated power. Typically the value of R is SO defined, but, in this work, a valid range is provided to study the impact of potential variation of the controller on the stability of the SSC-OPF solution. Therefore, R is set equal to 0.05 in the GD-driven case, whereas in the GD&CT-driven it varies in the range between 0.05 and 0.10.

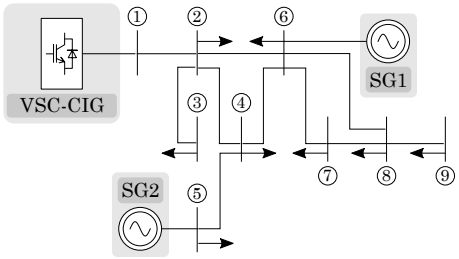


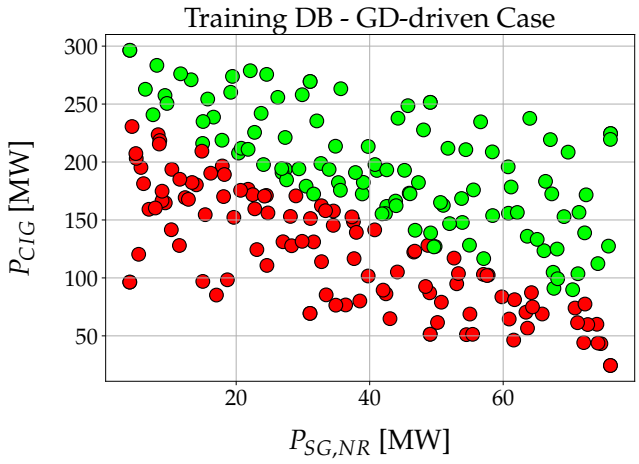
FIGURE 3: Scheme of the power system [26].

Concerning the demand, some assumptions have been made to limit and define the variable space to be explored during the data generation phase. Supposing that the load profile could be drawn from real data, the total power demand P_D is assumed to vary between 100 MW and 300 MW. Also, it is assumed that the loads absorb almost always the same share of the total demand. That is, $[P_{L_2}, ..., P_{L_9}]$=[10, 15, 20, 10, 5, 10, 5, 25]% of P_D. Then, it is considered that each load can vary by a 30% of its value ($\Delta P_{L_i} = P_{L_i} \pm 30\% P_{L_i}, \forall i \in \mathcal{L}$).

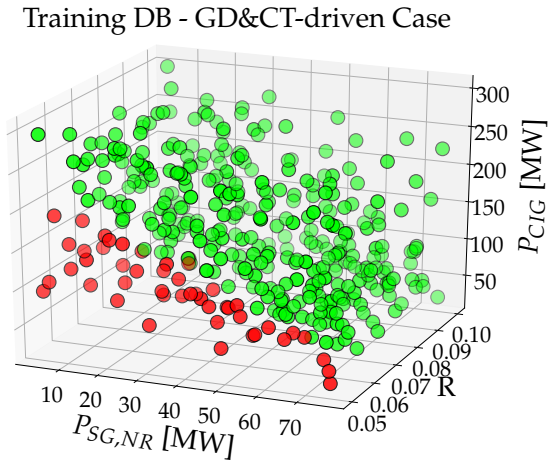
In the PF and OPF computations, the buses to which SGs are connected, are modeled as PQ buses: the SGs reactive power is set equal to zero while the active power is arbitrarily set between 5% and 95% of the SG rated power, changing the active power reference. The bus with the CIG is modeled as the slack bus: the active power and the reactive power are computed by PF equations, considering that the active power reference is set equal to the remaining part of the power demand not covered by SGs.

B. DATA GENERATION - GD-DRIVEN CASE

The following details how data have been generated for the study of the GD-driven case. First, the generation of the training DB is described, which follows the procedure presented in [29]. Then, it is explained how test instances



(a)



(b)

FIGURE 4: Samples of the GD-driven case (Fig. 4a) and GD&CT-driven case (Fig. 4b) training DBs. Markers color indicates the result of the small-signal stability assessment: green for stable operating points and red for unstable operating points.

have been generated, which, in the scope of this work, are used for validating the performance of the proposed SSC-OPF.

1) Training DB

The proposed data generation strategy is applied to a bi-dimensional space, whose dimensions are $P_{SG,NR}$ and P_{CIG}. Therefore, the space boundaries are identified by the generators' capacity limits and by the values of the minimum and maximum total power demand. In this way, the 216-sample DB shown in Fig. 4a is obtained.

2) Test DB

Two test DBs have been built, one for each optimization problem, i.e. the minimization of the power losses and the minimization of the power injected by non-renewable SGs. They are used both for the validation of regressions accuracy and for evaluating the effectiveness of the SSC-OPF. Each of them contains samples corresponding to simulation instances generated as follows:

1) 100 P_D values are selected randomly in the interval between 100 MW and 300 MW and each of them is associated with a different load distribution, applying LHS.
2) For each of these power demand instances, the OPF solution is computed to calculate the generators' power dispatch and the values of V and Θ at each bus.
3) The obtained equilibrium points are used as inputs for the state-space linear model and the damping ratios are computed, from which the DIs of the groups of critical eigenvalues are calculated.

C. DATA GENERATION - GD&CT-DRIVEN CASE

As for the GD-driven case, data for training and testing are generated as explained next.

1) Training DB

When the CIG frequency droop controller characteristic R is considered as an OPF variable, the space to be explored is a 3-dimensional space. The dimension R is limited between the selected values, 0.05 and 0.1. Applying the Entropy-based strategy, the 376-sample DB of Fig. 4b is obtained.

2) Test DB

In this case, for each optimization problem, two test DBs are generated, one considering a constant value of R and the other considering a variable value of R. Therefore, these DBs collect the damping ratios of system equilibrium points computed by OPF (with the proper objective function and for random values of power demand, selected as explained in Section IV-B2), and by evaluating the state-space linear model setting R either equal to a constant value (e.g. 0.05) or to a random value (between 0.05 and 0.1). If R is set equal to a constant value, the same instances of the test DBs generated for the validation of the GD-driven case can be used. Such test DBs are used to prove the performance of the SSC-OPF implementation. On the other side, the test DBs generated with a variable value of R are used for validating the accuracy of the DIs regressions.

D. IDENTIFICATION OF CRITICAL EIGENVALUES GROUPS

Analyzing the results of the small-signal stability assessment of the operating points of the training DBs, it is possible to identify the groups of critical eigenvalues. Such results can be shown and inspected through the representation of the eigenvalues in the modal plane. In this paper, for reasons of space, only the modal plane showing the results of the small-signal stability assessment of the instances of the GD&CT-driven case training DB is provided (Fig.5). The modal plane

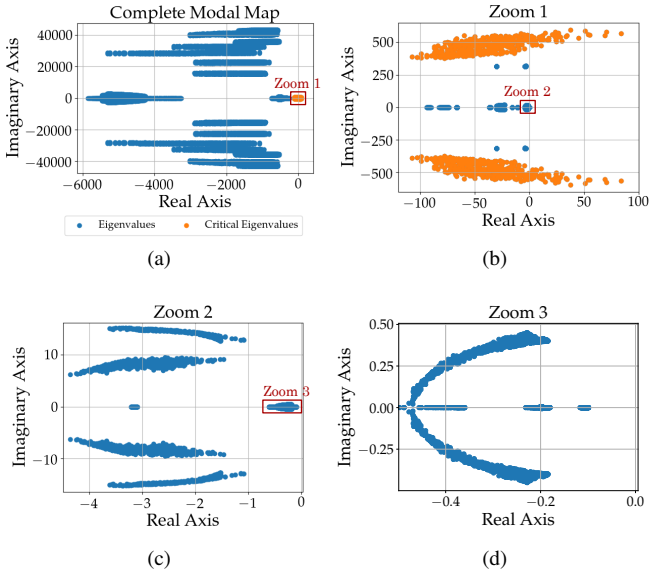


FIGURE 5: Modal maps showing the eigenvalues resulting from the small-signal stability assessment of instances of the GD&CT-driven case training DB. 5a Complete modal map, showing all the N_e eigenvalues for all the DB instances. 5b Zoom on the region of the plane interested by the critical eigenvalues. 5c-5d Zoom on the low-damped, non-critical eigenvalues.

related to the GD-driven case is omitted because Fig. 5 can be considered representative of both cases. The eigenvalues of the GD-driven case are obtained from the small-signal stability assessment of a set of operating points included in the range investigated for the generation of the training DB of the GD&CT-driven case. Therefore, observing Fig. 5, it is possible to assume that both for the GD-driven case and the GD&CT-driven case, there is one group of critical eigenvalues. Such a group is characterized by a couple of complex conjugate eigenvalues, crossing the imaginary axis and with high mobility. These eigenvalues are the ones with the orange markers in Fig. 5a-5b. In Fig. 5c-5d, it is possible to observe that the other eigenvalues with low damping, close to the imaginary axis, actually have low mobility and they never cross the imaginary axis causing instability. Therefore they can be considered not critical. Table 2 summarizes the values of the minimum, maximum, and standard deviation (std) of the DIs of the two cases implemented. The comparison between the global DI and the DI of the group of critical eigenvalues shows that the second has a larger std, which will positively affect the accuracy of the regression.

E. REGRESSION TECHNIQUE SELECTION

Given the results of the qualitative, comparative analysis provided in SectionIII-C, the final quantitative comparison is performed among the linear regressions and MARS. Therefore, the final technique selection is aimed at identifying

TABLE 2: Global DI and critical eigenvalues DI features for the GD-driven and GD&CT-driven cases.

	GD-driven case		GD&CT-driven case	
	Global DI	Critical Eigenvalues DI	Global DI	Critical Eigenvalues DI
min	0.9861	0.9020	0.9861	0.7336
max	1.1463	1.1463	1.1464	1.1464
std	0.0354	0.048	0.0209	0.072

TABLE 3: Comparison of the accuracy, measured by the R^2, between linear models and MARS.

		LR	Lasso	Ridge	MARS
GD-driven Case	min P_{losses}	0.7177	-0.8748	0.9292	0.9899
	min $P_{SG,NR}$	0.9922	-0.4976	0.9132	0.9939
GD&-driven Case	min P_{losses}	0.9056	-0.3615	0.4499	0.9965
	min $P_{SG,NR}$	-7.7527	-0.3189	0.5280	0.9961

the most accurate and robust technique. To do this, each technique is trained using the training DBs and then tested on the testing DBs obtained by solving the OPF problems. Therefore, each technique is trained twice, once with the GD-driven case training DB and once with the GD&CT-driven case training DB. Then, for each technique, both trained models are tested with the two corresponding testing DBs, which differ in the objective function of the OPF used to generate the data.

Before training, data are pre-processed: correlation among variables is removed by discarding the correlated variables (with a Pearson coefficient $> 90\%$) poorly correlated with the target variable. The result of such a process leads to select as input variables only the loads power demand ($P_{L,2}$-$P_{L,9}$), the power injected by the SGs (P_{SG_1},P_{SG_2}), and, in the GD-driven case the voltage angle at the bus number 9 (θ_9), while, in the GD&CT-driven case the voltage angle at the bus number 3 (θ_3) and the frequency droop characteristics (R).

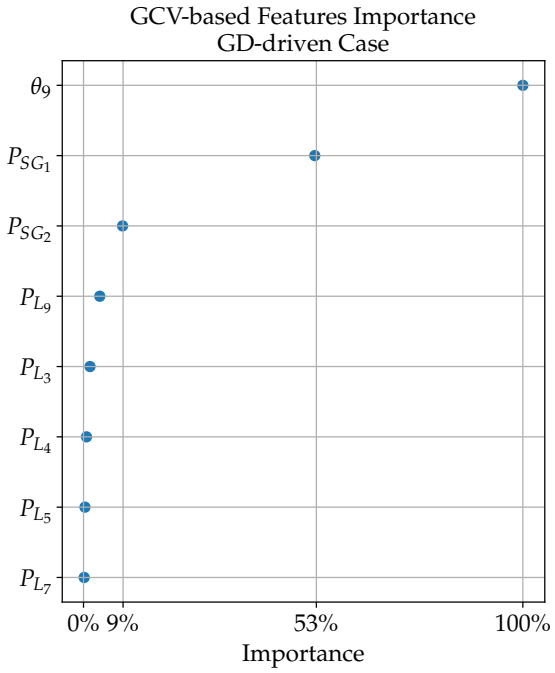
The metric used to assess the accuracy of the models is the R^2. Table 3 summarizes the obtained performances showing that MARS in all cases equals or outperforms the linear models, ensuring also high robustness.

F. REGRESSION TRAINING

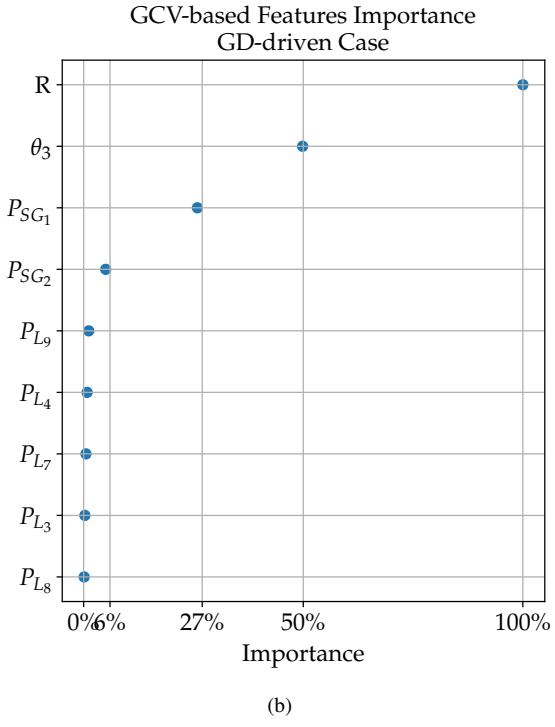
Two MARS models are trained, one for the GD-driven case and one for the GD&CT-driven case. Figures 6a and 6b show the results of the GCV-based *Features Importance* analysis obtained from the MARS models training.

G. SSC-OPF IMPLEMENTATION

The SSC-OPF is implemented using the *fmincon* function of the Matlab Optimization Toolbox [36]. This Section shows the results obtained implementing the SSC-OPF using the two proposed objective functions, both for the GD-driven case and for the GD&CT-driven case. To estimate the SSC-OPF performance, its solutions are compared with the ones obtained by the corresponding OPF, formulated with the same objective function but without the small-signal stability



(a)



(b)

FIGURE 6: *Features Importance* results obtained by training the MARS model and shown as the cumulative GCV score.

constraint. Therefore, the SSC-OPF and the OPF are executed for the same instances, which belong to the testing DBs with constant R tuning, both for the GD-driven case and the GD&CT-driven case. The solutions are compared in terms of DI, the value of the objective function (P_{losses} and $P_{SG,NR}$), generators power dispatch ($P_{SG,NR}$) and tuning of R, in the GD&CT-driven case. In the plots shown in Figs. 7-8, such quantities are represented in relation to the value of the total power demand P_D of the operating point. This is motivated by the fact that the test DBs have been generated imposing a different total power demand for each instance. In contrast, the value of the power injected by generators is calculated according to the objective function and the small-signal stability constraint. Concerning the graphs in which the DI is plotted, three typologies of results are shown and compared. They are indicated with the following labels and refer to:

- OPF: DI of the OPF solutions. It is computed through the analysis of the eigenvalues, using the OPF solution as input of the state-space linear model.
- SSC-OPF: exact DI of the SSC-OPF solutions. It is computed through the analysis of the eigenvalues, using the SSC-OPF solution as input of the state-space linear model.
- Predicted SSC-OPF DI: predicted DI of the SSC-OPF solutions. It is computed as the outcome of the regression, using the SSC-OPF solution quantities as inputs.

1) GD-driven Case

Fig. 7 shows the results obtained applying the GD-driven strategy both to the active power losses minimization problem (Fig. 7a) and the non-renewable SG power injection minimization problem (Fig. 7b). In both optimization problems, all the OPF solutions, computed for values of P_D between 100 MW and 300 MW, are feasible. In fact, there is no violation of the static limits.

Consider the results of the active power losses minimization problem shown in Fig. 7a. The plot related to $P_{SG,NR}$ shows that the solutions of the OPF tend to values of the SGs injected power close to their maximum limit. Therefore, the CIG is used to cover the remaining part of the power demand. The OPF solutions of the operating points with values of $P_D \geq$150 MW fulfill both the static and the small-signal stability limits. In fact, for such operating points, the DI values of the OPF solutions, shown in the top graph of Fig. 7a, are all lower than the stability threshold $\tau = 1$. Therefore, for these cases, the OPF solutions provide optimal and stable operating points. Hence, the corresponding SSC-OPF solutions should be equal to the ones of the OPF, in terms of GD, the value of the objective function, and operating point DI. It is observed that this is true for almost all the test operating points. The plot related to the $P_{SG,NR}$ quantity shows that the solutions of the SSC-OPF of 13 (over 100) test cases provide for a GD dispatch different from the one obtained by computing the OPF. Such different GD is caused by the presence of the small-signal stability constraint, which drives the solution

toward a different operating point. In particular, lower values of SGs power injection are observed. However, such SSC-OPF solutions do not show worse DI and higher power losses. Concerning the operating points with a P_D $\leq$150 MW, in Fig. 7a only the solutions of the OPF are shown. The SSC-OPF solutions are omitted as they do not converge. The unfeasibility of the SSC-OPF is due only to the non-fulfillment of the small-signal stability constraint. In fact, for such operating points, the OPF solutions converge, respecting all the static limitations, but are unstable (DI $\geq$1 for P_D $\leq$150 in Fig. 7a). The graph related to $P_{SG,NR}$ (bottom of Fig. 7a) shows that the OPF solutions of such unstable points provide for the injection of maximum SG power. The maximum SG power injection corresponds both to the minimum power loss and to the minimum DI conditions. Hence, the GD-driven strategy is not sufficient for computing stable solutions, as even when injecting the maximum SG power, stability is not achieved. This is what happens when the SSC-OPF is employed. Convergent solutions can be obtained by relaxing the small-signal stability constraint: in this way, the algorithm would converge to the same solutions provided by the OPF. In terms of accuracy, comparing the exact and predicted DI of the convergent SSC-OPF solutions, R^2 equal to 0.9823 is achieved.

Fig. 7b refers to the optimization problem whose objective function is to minimize $P_{SG,NR}$. In this case, the DI and the objective function value are plotted. The solutions of the OPF, which do not exhibit any convergence problem, show that it is possible to minimize the SG power injection without violating the static stability constraints. For values of P_D $\geq$246 MW, the OPF solutions are also dynamically stable. Therefore, for those cases, the solutions obtained by SSC-OPF coincide with the ones of the OPF. This is shown by the fact that for P_D $\geq$246 MW, both OPF and SSC-OPF solutions have a DI $<$ 1 and that the SGs power dispatch coincide. For values of P_D $<$246 MW the solutions of the OPF, with a minimum SGs power injection, are not stable. The SSC-OPF can find operating points with lower DI, increasing the power provided by SGs. In Fig. 7b, it is observed that the solutions of the SSC-OPF for the operating points with P_D between 150 MW and 246 MW have a DI slightly lower than 1 and a $P_{SG,NR}$ that ranges between its minimum value (at P_D= 246 MW) and maximum value (at P_D= 150 MW). Then, for P_D lower than 150 MW, the SSC-OPF solutions are not convergent, and for these reasons omitted in the graphs of Fig. 7b. By relaxing the small-signal stability constraint, it would be possible to observe that for such operating points, the SSC-OPF solutions would provide for the maximum possible SGs power injection. However, this GD is not sufficient for ensuring stability: DI values lower than the ones of the OPF solutions can be achieved, but still larger than 1. In this case, the R^2 between the exact DI and the predicted DI of the SSC-OPF solutions is equal to 0.8496.

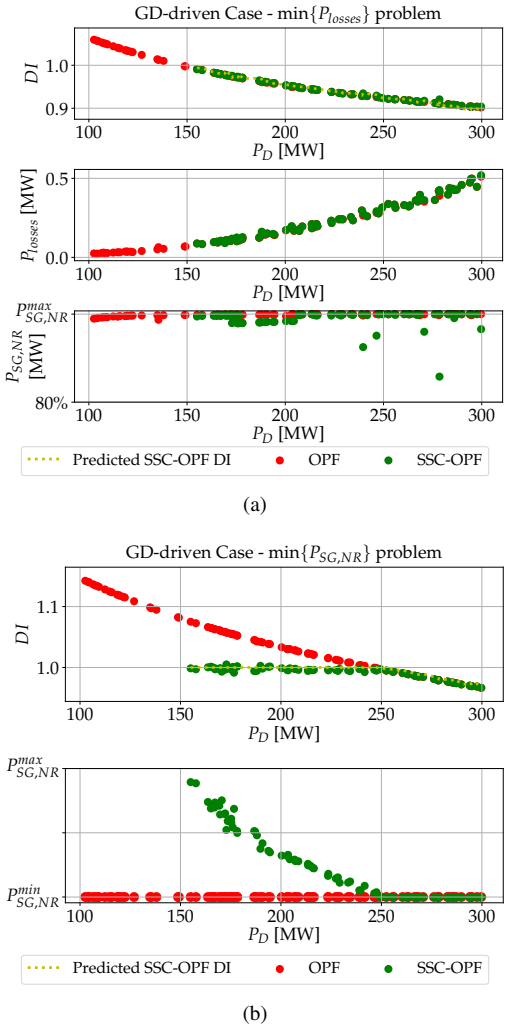


FIGURE 7: GD-driven case: comparison between the SSC-OPF and the OPF solutions of the minimum power losses (7a) and minimum SGs power injection (7b) problems.

2) GD&CT-driven Case

In this case, the SSC-OPF can compute a different operating point and a different tuning of the selected converter frequency droop controller, to reach a DI $<$1. These solutions are compared with the OPF solutions, computed with a fixed value of R=0.05.

The results in Fig. 8a refer to the optimization problem in which the power losses have to be minimized. The OPF solutions are equal to the ones shown in Fig. 7a and described previously. The SSC-OPF solutions show a small-signal stability behavior better than the one of the OPF solutions for all the test instances. In fact, in the DI graph of Fig. 8a it can be observed that the DI of the SSC-OPF solutions is always lower than the one of the OPF solutions. This is particularly relevant for those operating points that have unstable OPF solutions. Such points are the ones with P_D $\leq$150 MW, for which the SSC-OPF can find a solution with a DI $<$1. In general, such better small-signal stability behavior of the

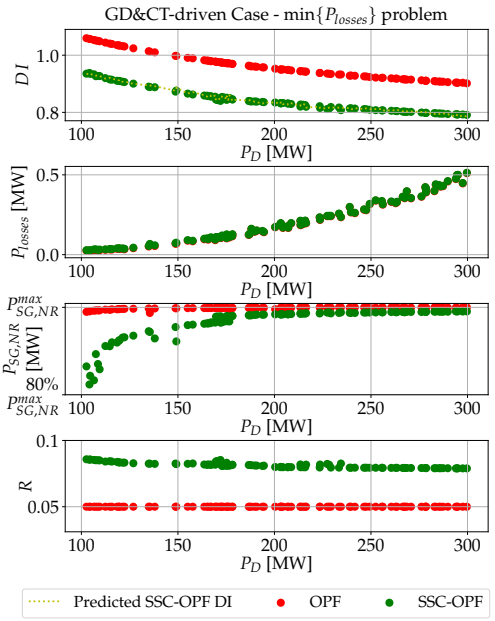
solutions of the SSC-OPF is achieved by providing a tuning of R equal to values larger than 0.05. Also in this case, the presence of the small-signal stability constraint leads to solutions of the SSC-OPF different in terms of GD from what might be expected. In the graph related to the $P_{SG,NR}$ in Fig. 8a, it is shown that, for some test cases, the SSC-OPF solutions provide for a GD with a $P_{SG,NR}$ lower than its maximum value (a condition that corresponds to the minimization of the system power losses). However, this is not caused by the inaccuracy of the regression-based small-signal stability constraint, which corresponds to a R^2 =0.9921, and does not lead to an increase in the system power losses.

Finally, in Fig. 8b the results obtained when the optimization problem is aimed to minimize the power injected by SGs are shown. In the graph related to the DI, it is shown that also in this case, tuning the R parameter, the SSC-OPF can provide optimal solutions with DI <1 for all the operating points, including the ones that have unstable OPF solutions. The graph related to the objective function value shows that the solutions are optimal as they are characterized by $P_{SG,NR}$ equal to its minimum value. The regression R^2 is 0.9909, calculated between the exact and predicted DI of the SSC-OPF solutions.

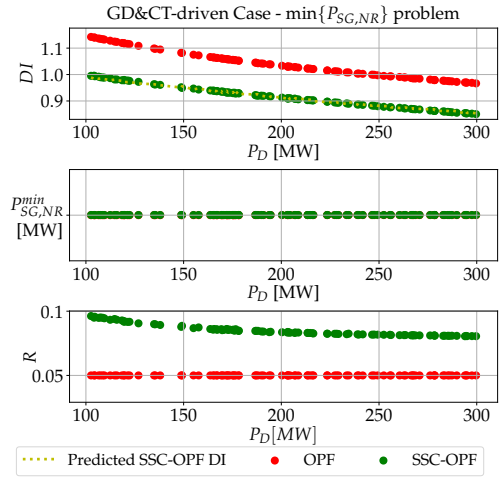
V. CONCLUSIONS

In this paper, a data-driven SSC-OPF is proposed. After a comparison among linear regressions and MARS, the latter has been selected to formulate the small-signal constraint for its better accuracy and robustness. Such regression-based constraint computes the value of a small-signal indicator, i.e. the DI of the critical eigenvalues. Two possible problem implementations are investigated: computing GD (GD-driven case) or combining GD and converters CT (GD&CT-driven case). In particular, the possibility of tuning the frequency droop controller of a converter is considered. Therefore, in the GD&CT-driven case, the solution of the SSC-OPF provides the CT that ensures the small-signal stability of the optimal operating point. From the analysis of the case study, the GD-driven case results in a less effective strategy: to achieve stable solutions the operating points are moved toward sub-optimal points. Moreover, this is not always sufficient for reaching stability. In contrast, including the possibility of tuning converters' controllers shows that it is technically feasible to achieve optimal, stable solutions. Both GD-driven and GD&CT-driven SSC-OPF problems have been tested with two objective functions: minimizing transmission active power losses and minimizing non-renewable sources power injection. Such objective functions have been selected to exemplify the use of the proposed SSC-OPF as a tool to be employed both for power system operation and analysis. Possible applications are identified:

- Market software adopted by an SO, to avoid re-dispatch due to small-signal instability.
- Analysis tool for grid planning, to be used for verifying the optimal, stable operation of new system topologies.



(a)



(b)

FIGURE 8: GD&CT-driven case: comparison between the SSC-OPF and the OPF solutions of the minimum power losses (7a) and minimum SGs power injection (7b) problems.

ACKNOWLEDGMENT

This work was supported by FEDER / Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación, under the project REFORMING (PID2021-127788OA-I00). The work of O. Gomis was supported by the ICREA Academia program. The work of E. Prieto and M. Cheah was supported by Serra Húnter Program.

REFERENCES

[1] M. B. Cain, R. P. O'neill, A. Castillo, et al., History of optimal power flow and formulations, Federal Energy Regulatory Commission 1 (2012) 1–36.
[2] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. Van Cutsem, V. Vittal, Definition and classification of power system stability ieee/cigre joint

task force on stability terms and definitions, IEEE Trans. Power Syst. 19 (3) (2004) 1387–1401. doi:10.1109/TPWRS.2004.825981.
[3] N. Hatziargyriou, J. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal, et al., Definition and classification of power system stability–revisited & extended, IEEE Trans. Power Syst. 36 (4) (2020) 3271–3281.
[4] M. Cheah-Mane, A. Egea-Alvarez, E. Prieto-Araujo, H. Mehrjerdi, O. Gomis-Bellmunt, L. Xu, Modeling and analysis approaches for small-signal stability assessment of power-electronic-dominated systems, Wiley Interdisciplinary Reviews: Energy and Environment (2022) e453.
[5] L. Fan, Z. Miao, Wind in weak grids: 4 hz or 30 hz oscillations?, IEEE Trans. Power Syst. 33 (5) (2018) 5803–5804. doi:10.1109/TPWRS.2018.2852947.
[6] F. Thams, L. Halilbasic, P. Pinson, S. Chatzivasileiadis, R. Eriksson, Data-driven security-constrained opf, in: Proc. 10th Bulk Power Syst. Dyn. Control Symp., 2017, pp. 1–10.
[7] M. E. Bento, A method for monitoring the load margin of power systems under load growth variations, Sustainable Energy, Grids and Networks 30 (2022) 100677.
[8] M. A. Nezhadpashaki, F. Karbalaei, S. Abbasi, Optimal placement and sizing of distributed generation with small signal stability constraint, Sustainable Energy, Grids and Networks 23 (2020) 100380.
[9] J. L. Rueda, W. H. Guaman, J. C. Cepeda, I. Erlich, A. Vargas, Hybrid approach for power system operational planning with smart grid and small-signal stability enhancement considerations, IEEE Trans. on Smart Grid 4 (1) (2013) 530–539. doi:10.1109/TSG.2012.2222678.
[10] I. Jahn, L. Bessegato, J. Bjork, F. Hohn, S. Norrga, N. Svensson, K. Sharifabadi, O. Despouys, A proposal for open-source hvdc control, in: 2019 IEEE PES ISGT-Europe, 2019, pp. 1–5. doi:10.1109/ISGTEurope.2019.8905635.
[11] S. Mendoza-Armenta, I. Dobson, Applying a formula for generator redispatch to damp interarea oscillations using synchrophasors, IEEE Trans. Power Syst. 31 (4) (2015) 3119–3128.
[12] A. A. Lerm, Control of hopf bifurcation in power systems via a generation redispatch, in: 2001 IEEE Porto Power Tech Proceedings (Cat. No. 01EX502), Vol. 2, IEEE, 2001, pp. 6–pp.
[13] K. Van den Bergh, D. Couckuyt, E. Delarue, W. D'haeseleer, Redispatching in an interconnected electricity system with high renewables penetration, Electric Power Systems Research 127 (2015) 64–72.
[14] R. Liu, G. Verbič, J. Ma, A machine learning approach for fast future grid small-signal stability scanning, in: 2016 IEEE POWERCON, IEEE, 2016, pp. 1–6.
[15] P. S. Kundur, O. P. Malik, Power system stability and control, McGraw-Hill Education, 2022.
[16] R. Zárate-Miñano, F. Milano, A. J. Conejo, An opf methodology to ensure small-signal stability, IEEE Trans. Power Syst. 26 (3) (2010) 1050–1061.
[17] F. Hasan, A. Kargarian, A. Mohammadi, A survey on applications of machine learning for optimal power flow, in: 2020 IEEE TPEC, IEEE, 2020, pp. 1–6.
[18] A. Venzke, D. T. Viola, J. Mermet-Guyennet, G. S. Misyris, S. Chatzivasileiadis, Neural networks for encoding dynamic security-constrained optimal power flow to mixed-integer linear programs, arXiv preprint arXiv:2003.07939 (2020).
[19] J. Liu, Z. Yang, J. Zhao, J. Yu, B. Tan, L. Wenyuan, Explicit data-driven small-signal stability constrained optimal power flow, IEEE Trans. Power Syst. (2021).
[20] S. Asvapoositkul, R. Preece, Decision tree-based prediction model for small signal stability and generation-rescheduling preventive control, Electric Power Systems Research 196 (2021) 107200.
[21] F. Capitanescu, J. M. Ramos, P. Panciatici, D. Kirschen, A. M. Marcolini, L. Platbrood, L. Wehenkel, State-of-the-art, challenges, and future trends in security constrained optimal power flow, Electric power systems research 81 (8) (2011) 1731–1741.
[22] R. H. Byrd, M. E. Hribar, J. Nocedal, An interior point algorithm for large-scale nonlinear programming, SIAM Journal on Optimization 9 (4) (1999) 877–900.
[23] J. H. Friedman, Multivariate adaptive regression splines, The annals of statistics (1991) 1–67.
[24] M. Rahmatian, Y. C. Chen, A. Palizban, A. Moshref, W. G. Dunford, Transient stability assessment via decision trees and multivariate adaptive regression splines, Electric power systems research 142 (2017) 320–328.
[25] H. Glavitsch, R. Bacher, Optimal power flow algorithms, Analysis and control system techniques for electric power systems 41 (1991) 135–206.
[26] C. Collados-Rodriguez, M. Cheah-Mane, E. Prieto-Araujo, O. Gomis-Bellmunt, Stability and operation limits of power systems with high penetration of power electronics, International Journal of Electrical Power & Energy Systems 138 (2022) 107728.
[27] F. Thams, A. Venzke, R. Eriksson, S. Chatzivasileiadis, Efficient database generation for data-driven security assessment of power systems, IEEE Trans. Power Syst. 35 (1) (2019) 30–41.
[28] V. Krishnan, J. D. McCalley, S. Henry, S. Issad, High information content database generation for data mining based power system operational planning studies, in: IEEE PES General Meeting, 2010, pp. 1–8. doi:10.1109/PES.2010.5589760.
[29] F. Rossi, E. P. Araujo, M. C. Mañe, O. G. Bellmunt, Data generation methodology for machine learning-based power system stability studies, in: 2022 IEEE PES ISGT-Europe, IEEE, 2022, pp. 1–5.
[30] M. D. Shields, J. Zhang, The generalization of latin hypercube sampling, Reliability Engineering & System Safety 148 (2016) 96–108.
[31] E. A. Unger, L. Harn, V. Kumar, Entropy as a measure of database information, in: [1990] Proceedings of the Sixth Annual Computer Security Applications Conference, IEEE, 1990, pp. 80–87.
[32] L. Kaufman, P. J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, John Wiley & Sons, 2009.
[33] C. M. Bishop, N. M. Nasrabadi, Pattern recognition and machine learning, Vol. 4, Springer, 2006.
[34] V. Skala, Fast interpolation and approximation of scattered multidimensional and dynamic data using radial basis functions (2013).
[35] J. Franklin, The elements of statistical learning: data mining, inference and prediction, The Mathematical Intelligencer 27 (2) (2005) 83–85.
[36] T. M. Inc., Optimization toolbox version: 9.4 (r2022b) (2022).
URL `https://www.mathworks.com`

FRANCESCA ROSSI received the degree in Energy and Nuclear Engineering from the Polytechnic of Turin, Italy, in 2019. She joined the CITCEA-UPC research group in 2019 and she is currently pursuing a Ph.D. degree in electrical engineering. Her research interest includes data science and machine learning, power systems stability, and power power electronics-dominated power systems.

EDUARDO PRIETO-ARAUJO (S'12-M'16-SM'21) received the degree in Industrial Engineering from the School of Industrial Engineering of Barcelona (ETSEIB), Technical University of Catalonia (UPC), Barcelona, Spain, in 2011 and the Ph.D. degree in electrical engineering from the UPC in 2016. He joined CITCEA-UPC research group in 2010 and currently he is a Serra Húnter Associate Professor at the Electrical Engineering Department, at UPC. During 2021, he was a visiting professor at the Automatic Control Laboratory, ETH Zurich. In 2022, he co-founded eRoots, which is a spin-off company of CITCEA-UPC, focused on the analysis of modern power systems. His main interests are renewable generation systems, control of power converters for HVDC applications, interaction analysis between converters, and power electronics-dominated power systems.

MARC CHEAH-MANE (S'14-M'18) received the degree in industrial engineering from the School of Industrial Engineering of Barcelona (ETSEIB), Universitat Politecnica de Catalunya (UPC), Barcelona, Spain, in 2013, and the PhD degree in Electrical Engineering from Cardiff University, Cardiff, the U.K. in 2017. From 2017 to 2020 he was a research associate at CITCEA-UPC, Barcelona, Spain. Since 2020 he has been a Lecturer under the Serra Húnter program at the Electrical Engineering Department of UPC and since 2022 he has been co-founder of eRoots, which is a spin-off company of CITCEA-UPC. His research interests include power systems with power electronics, high-voltage direct current systems, wind, and photovoltaic generation.

ORIOL GOMIS-BELLMUNT (S'05-M'07-SM'12-F'21) received the degree in Industrial Engineering from the School of Industrial Engineering of Barcelona (ETSEIB), Technical University of Catalonia (UPC), Barcelona, Spain, in 2001 and the Ph.D. degree in electrical engineering from the UPC in 2007. In 1999, he joined Engitrol S.L. where he worked as a Project Engineer in the automation and control industry. Since 2004, he has been with the Electrical Engineering Department, at UPC, where he is a Professor and participates in the CITCEA-UPC Research Group. Since 2020, he has been an ICREA Academia researcher. In 2022, he co-founded eRoots, which is a spin-off company of CITCEA-UPC, focused on the analysis of modern power systems. His research interests include the fields linked with electrical machines, power electronics, and renewable energy integration in power systems.