Received: 24 July 2022 | Revised: 27 September 2022 | Accepted: 23 October 2022

IET Generation, Transmission & Distribution

DOI: 10.1049/gtd2.12658

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Robust frequency constrained uc using data driven logistic regression for island power systems

Mohammad Rajabdorri | Enrique Lobato | Lukas Sigrist

IIT School of Engineering, ICAI Universidad Pontificia Comillas, Madrid, Spain

Correspondence
Mohammad Rajabdorri, IIT School of Engineering, ICAI Universidad Pontificia Comillas, Madrid, Spain.
Email: mrajabdorri@comillas.edu

Funding information
European Regional Development Fund, Grant/Award Number: RTI2018-100965-A-I00

Abstract

In the current practice of short-term power scheduling, online power reserves are used to address generation mismatches and contingencies. Neither online inertia nor the speed of the committed units is considered in the scheduling process. With the increasing injection of uncertain renewable energy sources, this practice is starting to fall short, especially in island power systems, where the primary frequency response is already scarce, and any contingency leads to potentially poor frequency response. This paper introduces a data-driven linear constraint to improve the post-fault frequency quality in island power systems. A coherent initial data-set is obtained by simulating the system frequency response of single outages. Then logistic regression is employed as a predictive analytic procedure to differentiate the acceptable and unacceptable incidents. To compare the conventional methods with the proposed approach and also to handle the uncertain nature of renewable energy generation, an adaptive robust unit commitment formulation is utilized. Results for the island power system of La Palma show that depending on the chosen cut-point on the logistic regression estimation the proposed method can improve the frequency response quality of the system while reducing the operation costs.

1 | INTRODUCTION

1.1 | Motivation

Variability and uncertainty are becoming a bigger concern in power systems due to the ever-increasing penetration of RES as a source of power generation. Among power systems, island power systems suffer more as they inherently possess less inertia and primary frequency control capacity. Inertia scarcity in island power systems makes them more susceptible to power outages and fluctuations in uncertain renewable energy sources (RES). Traditionally, online reserve power provided by conventional units has been the main tool to tackle unforeseen sudden changes in power balance and to maintain the frequency within a tolerable range. The current reserve assignment is such that the N-1 criterion is covered and expected load and RES variations can be absorbed, but it ignores available inertia and system response speed. This practice is falling short as (1) the conventional units are less utilized by increasing the share of RES, (2) the amount of available reserve might not be enough depending on the changes in RES infeed, which is exposed to forecast errors, (3) the system is left with insufficient amount of responsive resources facing outages and forecast errors. It's reasonable to propose a scheduling method that can tackle the uncertainties that the RES is bringing while ensuring the frequency response after the outages are tolerable. This is challenging, because both stochastic formulations and frequency constraints are quite complex, and impose a lot of computational burden on the scheduling procedure. This paper tries to address both of these issues while keeping the size of the UC problem intact.

1.2 | Framework

To address the volatile nature of RES and include the stochasticities in the scheduling process, usually stochastic and robust models are employed. Considering the pros and cons of different models, an adaptive robust UC is employed for the purpose of this paper. Some of the more recent usages and developments in the formulation can be found in [1–3], and [4]. A robust formulation is employed in this paper to include the uncertainties of RES. To ensure the provision of sufficient and fast reserves, different solutions are introduced in the literature ([5–8]). While new sources of the reserve are being introduced, it's also essential to make sure that the quality of frequency transitions is guaranteed in the scheduling process, in case of any abrupt contingency.

Following the higher injection of RES to the grid, larger frequency deviations are expected after any power mismatch. The amount of frequency control that is needed depends on system inertia, generation loss, and the speed of providing reserve. More attention is being paid to this issue. One obstacle is that frequency-related constraints, like frequency nadir, are highly non-linear, so it's hard to implement them in the scheduling process, which is usually solved by mixed-integer linear programming methods. In [9], a linear formulation is introduced that equips the unit commitment problem with information about inertial response and the frequency response of the system and makes sure that in case of the largest outage, there is enough ancillary service to prevent under frequency load shedding (UFLS). To linearize frequency nadir constraint, first-order partial derivatives of its equation with respect to higher-order non-linear variables are calculated. Then the frequency nadir is presented by a set of piecewise linearized constraints. In [10], different frequency services are optimized simultaneously with a stochastic unit commitment (SUC) approach, targeting low inertia systems that have high levels of RES penetration. The stochastic model uses scenario trees, generated by the quintile-based scenario generation method. To linearize frequency nadir, an inner approximation method is used for one side of the equation, and for the other side, a binary expansion is employed and linearized using the big-M technique. In [11], a stochastic unit commitment approach is introduced for low inertia systems, that includes frequency-related constraints. The problem considers both the probability of failure events and wind power uncertainty to compute scenario trees for the two-stage SUC problem. An alternative linearization approach is used to make sure the nadir threshold is not violated. Instead of piece-wise linearizing the whole equation, relevant variables including the nonlinear equation are confined within a plausible range that guarantees frequency drop after any contingency will be acceptable. Reference [12] has proposed a forecasting approach to model the uncertainties of RES to define upper and lower bounds and further implement them in a robust unit commitment (RUC). This study has assumed that frequency deviation is a linear function of the RoCoF, and has added it as a constraint to the RUC problem. In [13], a reformulation linearization technique is employed to linearize the frequency nadir limit equation. To address the uncertainties of wind generation, an improved interval unit commitment is used. Results show that controlling the dynamic frequency during the scheduling process decreases the operation costs of the system while ensuring its frequency security. In [14], first, a frequency response model is developed that provides enough primary frequency response and system inertia in case of an outage. All frequency dynamic metrics, including the RoCoF and frequency nadir, are obtained from this model, as analytic explicit functions of UC state variables and generation loss. These functions are then linearized based on a pseudo-Boolean theorem, so they can be implemented in linear frequency constrained UC problems. To find the optimal thermal unit commitment and virtual inertia placement, a two-stage chance-constrained stochastic optimization method is introduced in [15]. Frequency nadir is first defined with a bi-linear equation and then it's linearized with the help of the big-M approach. Although these methods are directly obtained from the dynamic equations, they are based on assumptions, and they increase the computational complexity of the UC problem.

In [16], instead of extracting analytical formulas from swing equation, a data-driven multivariate optimal classification trees (OCT) technique is used to extract linear frequency constraints. A robust formulation is proposed to address the uncertainties of load and RES. OCT is solved separately as an MILP problem. Because of that, the size of the training dataset should be moderately small, especially for deeper tree structures. A dynamic model is presented in [17] to generate the training data. The generated data is trained by the deep neural network. Trained neural networks are formulated so they can be used in an MIL problem and the frequency nadir predictor is developed, to be used in the UC problem. Then in [18] deep neural network (DNN) is trained by high-fidelity power simulation and reformulated as an MIL set of constraints to be used in UC. The number of constraints and variables that are required for MIL representation of DNN, can be overwhelming and increase the computational complexity of the problem. In [19], a revised support vector machine (SVM) based method is introduced to convexify the frequency nadir constraint. Then based on the method, an FCUC model is formulated as a tractable mixed integer quadratic problem. In [20] the gradient boosting decision tree algorithm is employed to build a frequency response model, which is then added to the UC problem to maintain the frequency within an acceptable range. In the same line of research, this paper is presenting a data-driven constraint to enhance the frequency response quality after outages. A summary of the reviewed FCUC-related papers is provided in Table 1.

Analytical formulations for frequency metrics are usually based on simplified models with respect to fully detailed power system models or linearizing the nonlinear behavior of power systems during large active power unbalance, and including them methodically in the UC problem To include the non-linear frequency metrics in linear UC, reviewed references are trying to employ a linearization technique. Eventually, the obtained linear lines are always a function of system dynamic constants, available inertia, and the amount of power imbalance. Although this serves the purpose of ensuring the quality of frequency response, it usually increases the size and complexity of the UC problem, in order to reach some level of accuracy. This paper employs Logistic regression (LR) as a dichotomous classification approach to classify the post-fault frequency drop as acceptable or unacceptable. LR is one of the most useful statistical procedures in healthcare analysis, medical statistics, credit

TABLE 1 A summary of frequency constrained UC references

#/year	Uncertainty model	Linearization technique	Case study
[9]/2018	Deterministic	First order partial derivatives	Great Britain 2030
[10]/2019	Stochastic	Inner approximation and binary expansion	Great Britain 2030
[11]/2020	Stochastic	Extracting bounds on relevant variables	IEEE RTS-96
[12]/2016	Robust	Assuming nadir is a linear function of RoCoF	Northern Chile
[13]/2020	Improved interval	Reformulation linearization technique	IEEE 6-bus
[14]/2020	Deterministic	Pseudo-Boolean theorem	IEEE RTS-96
[15]/2021	Chance-constrained	Binary expansion	China 196-bus
[16]/2021	Robust	Data-driven optimal classifier trees	Rhodes island and IEEE 118
[17]/2021	Deterministic	DNN trained by dynamic simulation	Modified 33-node system
[18]/2021	Deterministic	DNN trained by high-fidelity generated data	IEEE 39-bus system
[19]/2022	Deterministic	Convexifying by support vector machines	IEEE 24 and 118 bus systems
[20]/2022	Deterministic	Gradient boosting decision tree algorithm	Taipower system in Taiwan

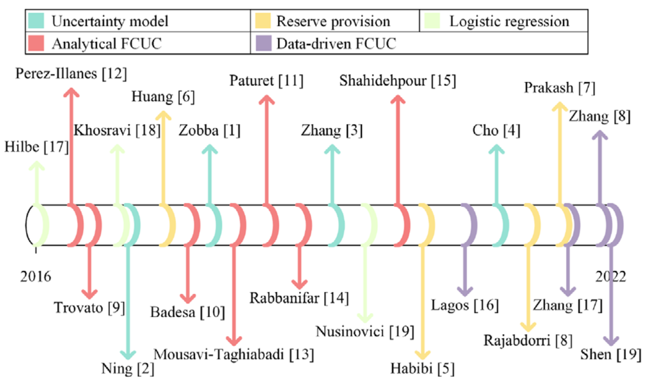

FIGURE 1 Summary of references

rating, ecology, social statistics, econometrics, etc. This procedure is important in predictive analytics, as it's able to categorize the outcome [21]. Considering the problem at hand and the purpose of this paper, this approach is promising. In [22], a framework is proposed that removes irrelevant features with no effect on classification and concludes that a training data-set with missing values can still generate sufficient explanations of LR classifications. These characteristics of LR make it an interesting option for FCUC application: (a) a linear constraint is derived from LR that can be directly used in the MIL formulation of UC. (b) in practice there are some generator outages that can be tolerated and some others that cannot. This type of dichotomous problem is what LR describes well. (c) in contrary to [14, 16], or [18] the obtained constraint from LR does not introduce any additional binary variables to the formulation, making it viable for more computationally demanding methods like robust and stochastic UC. (d) training data with LR is very fast, even for a big number of inputs to better represent the system behavior. (e) the LR gives a probabilistic interpretation of the classification. That helps the operator to choose the margin of security, depending on the requirements. A summary of all discussed papers is shown in Figure 1.

1.3 | Gaps and contributions

To the best of the authors' knowledge, logistic regression has not been used as an analytic tool in the UC problem and has never been employed to estimate the quality of frequency response in island power systems. Considering the presented background, this paper proposes a predictive analytic approach to enhance post-fault frequency quality in a robust UC model. The idea is to avoid dispatches that lead to poor frequency responses by scheduling only those generators whose outage would not violate acceptable frequency deviations, thus reducing the potential UFLS.

This paper proposes a novel data-driven constraint, by analyzing a coherent data-set, using a logistic regression procedure. To build an initial set of data to train the LR model, an adaptive robust UC formulation with reserve constraint is employed and solved for different levels of the reserve requirement. The idea of using different levels of the reserve is to simultaneously determine the level actually needed. The commitment variables of the robust UC solution for different levels of reserves are used to solve the economic dispatch (ED) problem for day-ahead stochastic scenarios. Every single outage of the obtained results is simulated by an SFR model, which makes the training dataset a proper representative of all acceptable and unacceptable frequency responses. From the training dataset, a new constraint is derived using the logistic regression procedure and then included in robust UC instead of conventional reserve constraint to ensure both frequency quality after outages, and feasibility of the result in case of any realization of the stochastic variable. Although the linearization happens in the training process, the new constraint does not add to the number of constraints in the UC problem, hence keeping the problem size intact. The system operators can use this method to include frequency dynamics in the short-term scheduling process (weekly UC, day-ahead UC, hours-ahead UC etc.) while keeping the size of the UC problem unaffected. To compare the proposed approach with recent data-driven methods that are introduced in the literature, OCT is also used to train a linear constraint.

Both methods are compared in the results and their computational run-time and improvements in the frequency quality are highlighted. Key contributions and merits compared to the current state of the art can be summarized as

- This paper introduces logistic regression as a tool to train output data of the SFR model, and develops a new constraint to be used in UC problems instead of the reserve constraint.
- Proposed formulation does not add any new binary, integer, or free variables to the UC problem and does not enlarge the number of UC constraints, conserving the size and complexity of the problem.
- The procedure of training the new constraint is very fast and can be done, using any computer algebra system.

The rest of the paper is organized as follows. In Section 2, the required methodology of the proposed approach is presented, starting with the robust UC with reserve constraint in Section 2.1. Then the SFR model is presented in Section 2.2, which takes the UC solutions as input. The outputs of the SFR model are used as the training data set for the LR model. How the LR works, and how the LR constraint is obtained is presented in Section 2.3. The adaptive robust UC formulation with LR can be found Section 2.4. The results are demonstrated in Section 3, and conclusions are drawn in Section 4.

2 | METHODOLOGY

This section presents the methodology. The main argument for using LR is that instead of trying to methodically linearize highly non-linear terms, it is possible to use historic or synthetic data to represent frequency metrics with a line that is a function of system dynamic constants, available inertia, available reserve, and the amount of lost power. Such a procedure does not jeopardize accuracy through linearization and does not introduce unnecessary complexity and computational burden. The methodology is valid for active power unbalances in general, including generation outages. The proposed method tries to distinguish between outages that potentially violate tolerable frequency levels and the ones that do not. This type of problem can be dealt with with dichotomous classification approaches like LR. The first step is to build a comprehensive set of data to train an accurate constraint. An adaptive robust UC with reserve constraint is used in this paper to obtain this data-set, which is explained in Section 2.1. The UC problem is solved for different levels of the reserve requirement, and ED is solved for all of the stochastic scenarios. The obtained results predominantly picture the possible feasible solutions that might be encountered in real-time. Using these data dynamic simulations are carried out to see the quality of frequency response in case of all potential outages. To perform the dynamic simulations an SFR model including the UFLS scheme is used (Section 2.2). As the inputs of the SFR model have different levels of reserve and the amount of inertia is ignored, the simulation results will be a broad-ranging mix of tolerable frequency responses, poor responses, and even unstable cases. Analyzing the correlation between inputs and outputs of the SFR model facilitates the training of the LR model (Section 2.3), so it can distinguish the tolerable cases and the ones which will lead to poor frequency responses in case of outages. Note that any other power system simulation tool can be used instead of the SFR model. The obtained estimation of the LR model is further used in an adaptive robust UC formulation as an alternative constraint instead of the current reserve constraint (Section 2.4). Such formulation is inherently equipped with a constraint that is able to control the quality of frequency response of potential outages.

2.1 | Adaptive robust UC with reserve constraint

The Unit Commitment (UC) problem is a mixed-integer problem and is usually solved with MIL Programming solvers after the linearization of nonlinear terms. To solve the UC problem with uncertainty, an adaptive robust formulation is employed in [23] and [24]. The formulation is robust, because it considers all of the possible realizations of the uncertain input, and makes sure that the chosen commitment status of the units, which is decided at the master level, will be feasible for any realization of the uncertain variable. The formulation is adaptive because the subproblem level is a function of the uncertain variables and can adapt the master level decision variable, depending on the different realizations of the uncertain variable. A general representation of the UC problem with reserve constraint and uncertain wind power injection is provided here,

$$\min_{x,p(w)} suc(x_{t,i}) + gc(p_{i,t}), \tag{1}$$

$$x_{t,i} - x_{t-1,i} = y_{t,i} - z_{t,i} \quad t \in \mathcal{T}, i \in \mathcal{I}, \tag{1a}$$

$$y_{i,t} + z_{i,t} \leq 1 \quad t \in \mathcal{T}, i \in \mathcal{I}, \tag{1b}$$

$$\sum_{tt=t-UT_i+1}^{t} y_{tt,i} \leq x_{t,i} \quad t \in \{UT_i, \dots, \mathcal{T}\}, \tag{1c}$$

$$\sum_{tt=t-DT_i+1}^{t} z_{tt,i} \leq 1 - x_{t,i} \quad t \in \{UT_i, \dots, \mathcal{T}\}, \tag{1d}$$

$$p(w)_{t,i} \geq \underline{\mathcal{P}}_i . x_{t,i} \quad t \in \mathcal{T}, i \in \mathcal{I}, w \in \mathcal{W}, \alpha, \tag{1e}$$

$$p(w)_{t,i} + r(w)_{t,i} \leq \overline{\mathcal{P}}_i . x_{t,i} \quad t \in \mathcal{T}, i \in \mathcal{I}, w \in \mathcal{W}, \beta, \tag{1f}$$

$$p(w)_{t-1,i} - p(w)_{t,i} \leq \underline{\mathcal{R}}_i \quad t \in \mathcal{T}, i \in \mathcal{I}, w \in \mathcal{W}, \gamma, \tag{1g}$$

$$p(w)_{t,i} - p(w)_{t-1,i} \leq \overline{\mathcal{R}}_i \quad t \in \mathcal{T}, i \in \mathcal{I}, w \in \mathcal{W}, \delta, \tag{1h}$$

$$\sum^{i \in \mathcal{I}} \left(p(w)_{t,i} \right) + wg(w)_t = d_t \quad t \in \mathcal{T}, w \in \mathcal{W}, \zeta, \tag{1i}$$

$$wg(w)_t \leq w_t \quad t \in \mathcal{T}, w \in \mathcal{W}, \eta, \tag{1j}$$

$$\sum_{ii\neq i}^{ii\in\mathcal{I}} \left(\bar{\mathcal{P}}_i - p(w)_{t,ii}\right) \geq p(w)_{t,i} \quad t \in \mathcal{T}, i \in \mathcal{I}, w \in \mathcal{W}, \mu. \tag{1k}$$

The aim is to solve Equation (1) subject to Equations (1a) to (1d), which only depends on binary variables, and Equations (1e)–(1k), which depend on both binary and real variables. $gc(.)$ is usually a quadratic cost function, which will be piecewise linearized to be utilized in a MIL problem. Equations (1a) and (1b) represent the binary logic of the UC problem. Equations (1c) and (1d) are the minimum up-time and minimum downtime constraints of the units. Equation 1e is the minimum power generation constraint, with dual multiplier α. Equation 1f is the maximum power generation constraint with dual multiplier β, and states that the summation of power generation and power reserve of every online unit, should be less than the maximum output of the unit. Equations (1g) and (1h) are ramp-down and ramp-up constraints, with dual multipliers γ and δ, respectively. Equation 1i is the power balance equation with dual multiplier ζ. Equation 1j with the dual multiplier, η makes sure that the scheduled wind power is always equal to or less than the uncertain forecasted wind. Equation 1k is the current reserve constraint with dual multiplier μ, and makes sure that in case of any contingency, there is enough headroom to compensate for the lost generation. Note that all the decision variables from Equations (1e) to (1k) are a function of uncertain wind power realization. In practice, an iterative delayed constraint generating Benders' decomposition algorithm is used to solve this problem [25]. The problem is broken to a master problem minimization subjected to Equations (1a) to (1d), and a subproblem with max-min form subjected to Equations (1e) to (1k).

$$\min_{x} \left(suc(x_{t,i}) + \max_{w\in\mathcal{W}} \min_{p} gc(p_{i,t}) \right). \tag{2}$$

The minimization on the master level is subjected to Equations (1a) to (1d), and the subproblem level minimization is subjected to Equations (1e) to (1k). The subproblem minimization problem determines the ED cost for a fixed commitment $\hat{x}$, and then it's maximized over the uncertainty set $\mathcal{W}$. Here the concept of duality in linear problems can be used. As the strong duality suggests, *the dual has an optimal solution if and only if the primal does, and the solutions are equal.* Taking the dual of subproblem converts the max-min form into a maximization problem. Considering the decomposed form of the problem, the feasible region of subproblem maximization is independent of x. So the subproblem maximization can be described as a set of extreme points and extreme rays of the solution region. Let $\mathcal{O}$ be the complete set of possible extreme points, and $\mathcal{F}$ be the complete set of possible extreme rays. These properties will later be used to define the decomposed master problem. In the iterative solution process, the binary variable, $\hat{x}_{t,i}$, is obtained from the masters' problem, hence it is fixed. With that in mind, and defining the auxiliary variable ϕ, as an understimator of optimal subproblem objective value, the dual form of the subproblem is defined as follows,

$$\phi \geq \max_{p} \begin{pmatrix} \sum^{t\in\mathcal{T}} \sum^{i\in\mathcal{I}} \alpha_{t,i}(\underline{\mathcal{P}}_i.\hat{x}_{t,i}) \\ -\sum^{t\in\mathcal{T}} \sum^{i\in\mathcal{I}} \beta_{t,i}(\overline{\mathcal{P}}_i.\hat{x}_{t,i}) \\ -\sum^{t\in\mathcal{T}} \sum^{i\in\mathcal{I}} (\gamma_{t,i}.\underline{\mathcal{R}}_i + \delta_{t,i}.\overline{\mathcal{R}}_i) \\ -\sum^{t\in\mathcal{T}} (\zeta_t.d_t + \eta_t.w_t) \\ -\sum^{t\in\mathcal{T}} \sum^{i\in\mathcal{I}} \mu_{t,i} \left(\sum_{ii\neq i}^{ii\in\mathcal{I}} \left(\overline{\mathcal{P}}_i\right)\right) \end{pmatrix}, \tag{3}$$

$$C_i - \alpha_{t,i} + \beta_{t,i} + \gamma_{t,i} + \delta_{t,i} + \zeta_t +$$

$$\eta_t + \sum_{ii}^{ii\in\mathcal{I}} \mu_{t,ii} \geq 0 \quad t \in \mathcal{T}, i \in \mathcal{I}, \tag{3a}$$

$$\zeta_t + \eta_t \geq 0 \quad t \in \mathcal{T}, \tag{3b}$$

$$\alpha, \beta, \gamma, \delta, \eta, \mu \geq 0 \text{ and } \zeta \text{ is free.} \tag{3c}$$

The dual form is Equation (3) subject to Equations (3a) to (3c). ζ is a free variable, because Equation (1i) is an equality. To find out more about writing a standard form of a problem, and taking the dual, have a look at [26]. The term $\eta_t w_t$ in the dual objective function is nonlinear, so an outer approximation approach [27] is employed to cope with it. The objective function of subproblem dual is a function of all dual variables and fixed $\hat{x}_{t,i}$ from the master problem in the previous iteration. Let us define the set of dual variables as u and the dual objective solution as $f(\hat{x}_{t,i}, \hat{u})$. Then the master problem is defined as follows,

$$\begin{aligned} \min_{x} \quad & suc(x_{t,i}) + \phi, \\ \textbf{s.t.} \quad & (1a) \text{ to } (1d), \\ & \phi \geq f(\hat{x}_{t,i}, \hat{u}) \quad \forall u \in \mathcal{O}^{'} \\ & 0 \geq f(\hat{x}_{t,i}, \hat{u}) \quad \forall u \in \mathcal{F} \end{aligned} \tag{4}$$

The iterative solution process starts with empty sets of $\mathcal{O}$ and $\mathcal{F}$. Then if the subproblem solution corresponding to $\hat{x}_{t,i}$ $(f(\hat{x}_{t,i}, \hat{u}))$ is feasible, an optimality cut is generated and added to $\mathcal{O}'$. And if the subproblem solution corresponding to the $\hat{x}_{t,i}$ is infeasible, $f(\hat{x}_{t,i}, \hat{u})$ is unbounded and a feasibility cut is generated and added to $\mathcal{F}'$. The iterations continue until ϕ is converged enough. The iterative algorithm is presented in Algorithm 1.

The UC problem is solved for different levels of the reserve requirement. The optimal commitment variables are then used to solve the ED problem for various stochastic wind

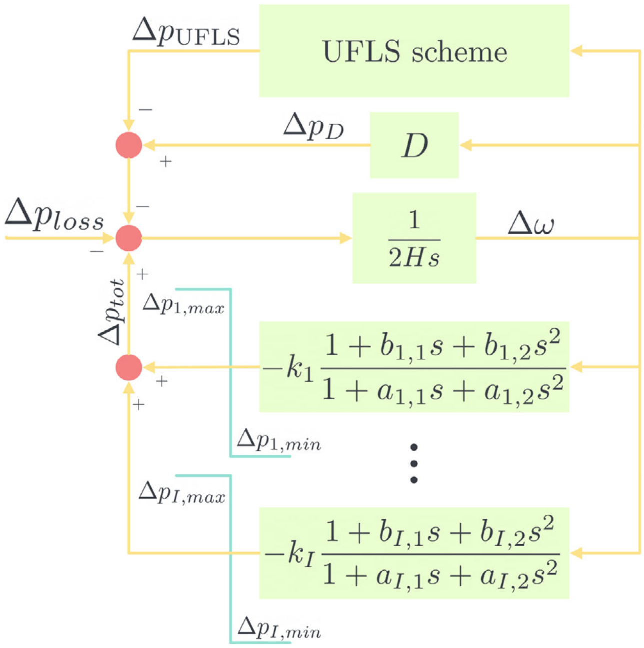

FIGURE 2 SFR model

scenarios to build an initial dataset, which will be implemented in the SFR model.

2.2 | System frequency response (SFR) model

This section briefly presents SFR models used to analyze the frequency stability of small isolated power systems. The model is able to reflect the underlying short-term frequency response of small isolated power systems. Figure 2 details the power-system model typically used to design UFLS schemes for an island power system, consisting of $\mathcal{I}$ generating units. Each generating unit i is represented by a second-order model approximation of its turbine-governor system. In fact, dynamic frequency responses are dominated by rotor and turbine-governor system dynamics. Excitation and generator transients can be neglected for being much faster than the turbine-governor dynamics. Since frequency can be considered uniform, equivalent normalized system inertia $\widetilde{\mathcal{H}}$ can be defined as follows,

$$\widetilde{\mathcal{H}}_t = \sum^{i\in\mathcal{I}} \frac{\mathcal{H}_i \mathcal{M}_i^{base} x_{t,i}}{\mathcal{S}^{base}}. \tag{5}$$

The overall response of loads can be considered by means of a load-damping factor D if its value is known. The gain k_i and parameters $a_{i,1}$, $a_{i,2}$, $b_{i,1}$ and $b_{i,2}$, of each generating unit i can be deduced from more accurate models or field tests. Since primary spinning reserve is finite, power output limitations $\Delta p_{i,min}$ and $\Delta p_{i,max}$ are forced. So the units can only participate as much as their available reserve. The complete model is explained in [28].

In practice, the UFLS scheme is designed to stabilize the system after large outages. For the purpose of building a data set to train the LR model, the UFLS scheme should be deactivated so the results capture the free frequency responses, including the ones that lead to instability quantified by unacceptable low-frequency nadir and steady-state frequency. Note however that the UFLS scheme will be considered to quantify the expected amount of UFLS when comparing the new reserve constraints with the current one in Section 3.

2.3 | Logistic regression (LR)

Regression methods are used for data analysis, concerned with describing the relationship between a response variable and one or more explanatory variables. Sometimes the output variable needs to be discrete, taking one or more possible values. In these instances, logistic regression is usually used. Consider a collection of m independent variables denoted by the vector $\xi' = (\xi_1, \xi_2, \ldots, \xi_m)$ related to a dichotomous dependent variable υ, where υ is typically coded as 1 or 0 for its two possible categories. Considering that for a (0,1) random variable, the expected value of υ is equal to the probability of $\upsilon = 1$ (i.e. $\pi(\upsilon = 1)$), and is defined here,

$$\pi(\upsilon = 1) = \frac{1}{1 + e^{-(c_0 + c_1\xi_1 + c_2\xi_2 + \cdots + c_m\xi_m)}}. \tag{6}$$

The regression coefficients c_0 to c_m in the logistic model Equation (6) provide important information about the relationships of the independent variables in the model to the dichotomous dependent variable. For the logistic model, these coefficients are used to estimate the odds ratio. Odds are defined as the ratio of the probability that some event will occur divided by the probability that the same event will not occur. Thus the odds for the event $\upsilon = 1$ is,

$$odds(\upsilon = 1) = \frac{\pi(\upsilon = 1)}{1 - \pi(\upsilon = 1)}. \tag{7}$$

Generally the conditional probability that the outcome presents is denoted by $\pi(\upsilon)$. The logit transformation of the probability $\pi(\upsilon = 1)$ is defined as natural logarithm of the odds of event $\upsilon = 1$, and considering Equation (6) is defined as,

$$\begin{aligned} logit\left(\pi(\upsilon = 1)\right) &= ln\left(\frac{\pi(\upsilon = 1)}{1 - \pi(\upsilon = 1)}\right) \\ &= c_0 + c_1\xi_1 + c_2\xi_2 + \cdots + c_m\xi_m. \end{aligned} \tag{8}$$

This is the *logit form* of the model and is given by a linear function [29]. The logit transformation is primarily applied to convert a variable that is bounded by 0 and 1 (i.e. probabilities) to a variable with no bounds [30]. When $logit(\pi(\upsilon = 1))$ goes toward $+\infty$, the probability of event $\upsilon = 1$ gets closer to 1, and when $logit(\pi(\upsilon = 1))$ goes toward $-\infty$, the probability of event $\upsilon = 1$ gets closer to 0. Usually $logit(\pi(\upsilon = 1)) = 0$ is considered as a cut-point, that separates those events with the probability of more than 0.5 on the positive side, and those events with the probability of less than 0.5 on the negative side. Depending

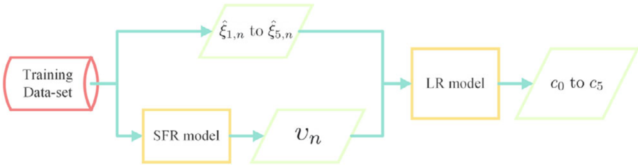

FIGURE 3 Flowchart of calculating LR coefficients

on the required accuracy for the model, different cut-points can be chosen.

As the frequency response of the system after contingencies is highly nonlinear, different approaches are employed in the literature to somehow linearize and include them in the UC problem. Some of these approaches are mathematically complicated and often tremendously burdensome for the solver. The argument here is that instead of linearizing the complex nonlinear equations, the output of developed SFR models can be analyzed to drive a linear constraint. To do so, the frequency response after each contingency can be marked as acceptable or unacceptable, depending on whether it violates the predefined limits or not. Then logistic regression is employed here to analyze the data and separate acceptable and unacceptable results with a trained line. This line is added later to the UC problem as a constraint by replacing the current reserve constraint Equation (1k). Such constraint can improve the frequency response quality and reduce the amount of load shedding due to unexpected outages, as it takes into account the expected dynamic behavior of the system.

As it's going to be further discussed in Section 3, a training data-set consisting of either synthetic or historic UC solutions is created to train the LR model. The independent variables of the LR model are assigned for every possible generator outage n in the training dataset. The independent variables that are defined should have a good correlation with the frequency response metrics after outage n. In this paper, the defined independent variables are the weighted summation of online inertia ($\xi_{1,n}$), the summation of inverse droop of the online units ($\xi_{2,n}$), lost power ($\xi_{3,n}$), lost power divided by the corresponding demand of that hour ($\xi_{4,n}$), and remaining of the reserve power after generator outages ($\xi_{5,n}$). Then every possible generator outage n in the training dataset is simulated with the SFR model to obtain a frequency response for each outage. Depending on the results of the SFR model, outage n will be tagged as 1 (acceptable) if the frequency response is within tolerable boundaries or 0 (unacceptable) if it's not. These binary tags will be employed as dependent variables of each outage for the LR model (v_n). Now that both independent and dependent variables of the LR model are formed, the training process can be performed to obtain coefficients c_0 to c_5. The process is shown in Figure 3 and explained again later Section 3 for the case study. The general form of the trained constraint estimated by the LR model is presented as follows,

$$c_0 + c_1\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{H}_{ii}\mathcal{M}_{ii}^{base} x_{t,ii}\right) + c_2\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{K}_{ii} x_{t,ii}\right) + c_3 p_{t,i} + \frac{c_4}{d_t} p_{t,i} + c_5\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} (\overline{\mathcal{P}}_{ii} x_{t,ii} - p_{t,ii})\right) \geq \psi \quad t\in\mathcal{T}, i\in\mathcal{I}. \tag{9}$$

This constraint enables the UC problem to also take into account the inertia and time constants of the system. The purpose is to improve the quality of frequency response with these measures. To obtain c_0 to c_5, the size and coherency of the training set should be good enough to describe the system. Although a small training data set can be classified with high precision, it might be insufficient to reflect the behavior of the system. Figure 4 shows a hypothetical example of the distribution of two variables. According to the distribution in Figure 4a all the lines, L1 to L3, can perfectly separate the two variables. But with a bigger size distribution sample, Figure 4b shows that L2 is a better candidate.

2.4 | Adaptive robust UC with LR constraint

The general formulation is similar to Equation (1), but reserve constraint in Equation (1k) is replaced by the LR constraint in Equation (9). The subproblem dual with the new constraint will become as follows,

$$\phi \geq \max_{p} \left(\begin{array}{l} \sum^{t\in\mathcal{T}}\sum^{i\in\mathcal{I}} \alpha_{t,i}(\underline{\mathcal{P}}_i.\hat{x}_{t,i}) \\ -\sum^{t\in\mathcal{T}}\sum^{i\in\mathcal{I}} \beta_{t,i}(\overline{\mathcal{P}}_i.\hat{x}_{t,i}) \\ -\sum^{t\in\mathcal{T}}\sum^{i\in\mathcal{I}} (\gamma_{t,i}.\underline{\mathcal{R}}_i + \delta_{t,i}.\overline{\mathcal{R}}_i) \\ -\sum^{t\in\mathcal{T}} (\zeta_t.d_t + \eta_t.w_t) \\ -\sum^{t\in\mathcal{T}}\sum^{i\in\mathcal{I}} \rho_{t,i} \\ \left(c_0 + c_1\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{H}_{ii}\mathcal{M}_{ii}^{base} x_{t,ii}\right)\right. \\ + c_2\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{K}_{ii} x_{t,ii}\right) \\ \left. + c_5\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} (\overline{\mathcal{P}}_{ii} x_{t,ii})\right)\right) \end{array} \right), \tag{10}$$

$$C_i - \alpha_{t,i} + \beta_{t,i} + \gamma_{t,i} + \delta_{t,i} + \zeta_t + \eta_t + (c_3 + \frac{c_4}{d_t})\rho_{t,i} + c_5 \sum_{ii\neq i}^{ii\in\mathcal{I}} \rho_{t,ii} \geq 0 t\in\mathcal{T}, i\in\mathcal{I}, \tag{10a}$$

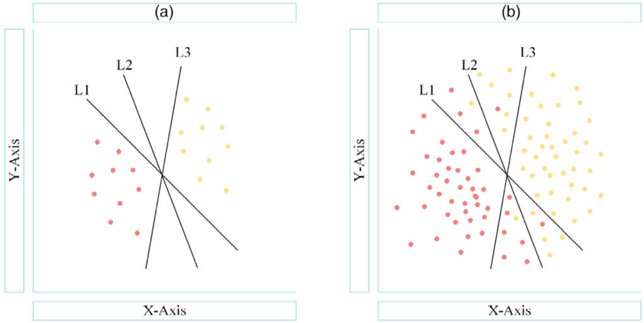

FIGURE 4 A hypothetical example to show the importance of the size of the training data set. a) small size. b) big size.

ALGORITHM 1 Iterative UC with reserve

Input: System specifications, wind uncertainty set, power demand

Output: ϵ-optimal solution

$j = 0$

while $|\phi^j(\hat{x}^j_{t,i}, \hat{w}^j) - \phi^j(\hat{x}^j_{t,i}, \hat{w}^{j-1})| < \epsilon$ **dodo**

  Solve master, minimizing $suc(x^j_{t,i}) + \phi^j(x^j_{t,i}, \hat{w}^{j-1})$ to get $\hat{x}^j_{t,i}$

  Solve subproblem using outer approximation, maximizing $f(\hat{x}^j_{t,i}, w^j)$ to get $\hat{w}^j$

  If $f(\hat{x}^j_{t,i}, \hat{w}^j)$ is bounded $\rightarrow \mathcal{O}' \cup \{\hat{u}_j\}$

  If $f(\hat{x}^j_{t,i}, \hat{w}^j)$ is unbounded $\rightarrow \mathcal{F}' \cup \{\hat{u}_j\}$

  $j = j + 1$

end while

$$\zeta_t + \eta_t \geq 0 \quad t \in \mathcal{T}, \tag{10b}$$

$$\alpha, \beta, \gamma, \delta, \eta, \rho \geq 0 \text{ and } \zeta \text{ is free.} \tag{10c}$$

As the objective function in the primal form and all the constraints that only involve binary variables are the same, the master problem remains the same as Section 2.1. The iterative solution procedure here is the same as Algorithm 1. A flowchart of the different steps of the proposed method is presented in Figure 5.

3 | RESULTS

3.1 | Case study and inputs

Simulations for the proposed methodology are carried on the real power system of La Palma island, one of Spain's Canary Islands. The yearly demand in 2018 is reported as about 277.8 GWh (average hourly demand of 31.7 MWh), supplied by eleven Diesel generators pre-dominantly. According to [31], the installed capacity of the La Palma island power system mounts to 117.7 MW, where about 6% of the installed capacity belongs to wind power generation. RES covers about 10% of the yearly demand. The capacity of the generators is shown in Table 2.

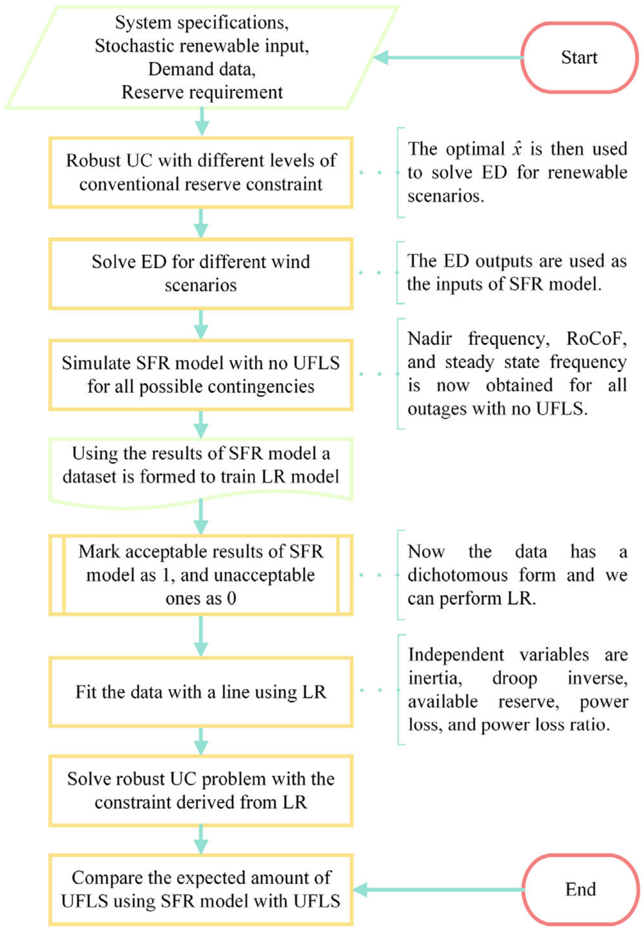

FIGURE 5 Flowchart of the proposed method

TABLE 2 Generator capacities

#	$\overline{P}_i [MW]$	$\underline{P}_i [MW]$
1	2.35	3.82
2	2.35	3.82
3	2.35	3.82
4	2.82	4.30
5	3.30	6.70
6	3.30	6.70
7	6.63	11.50
8	6.63	11.20
9	6.63	11.50
10	6.63	11.50
11	4.85	21

The input data for solving the UC problem is obtained from real data. Different scenarios of forecasted wind generation data of a sample day are chosen, which also provide the upper bound and the lower bound of the wind availability for the robust formulation. Wind data with 10 scenarios is shown in Figure 6.

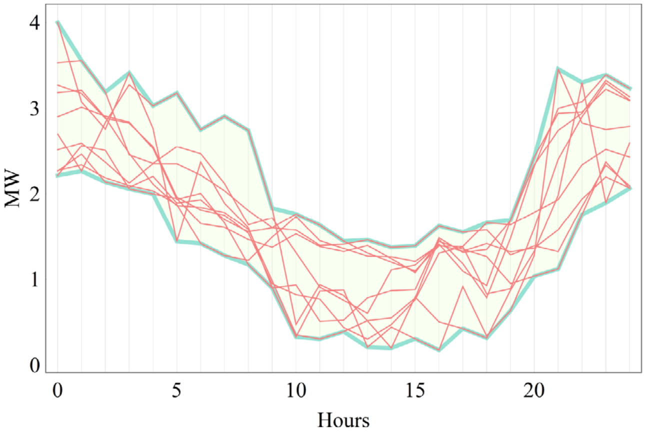

FIGURE 6 Wind data uncertainty set

The scenarios are from [32] and are scaled for La Palma island installed wind capacity.

An initial data set is required to train the LR model. The training data set should be able to represent the system, even in extreme circumstances, so it should include incidents with bad and unstable frequency responses too. Here it's decided to consider different reserve levels and different probable wind scenarios, to provide a wide range of plausible data to train the LR model, so it can reliably distinguish acceptable and unacceptable results. To achieve such a training data set, the conventional day-ahead robust UC is solved for ascending reserve requirements levels, starting from zero requirements until the problem becomes infeasible. By doing so a wide range of plausible UC solutions will be obtained, including generator outages that lead to tolerable frequency responses, poor frequency responses, and even unstable responses. In the conventional UC, the reserve requirement is typically the largest generation source under moderate RES penetration. A multiplier is defined here for the reserve requirement starting from 0, with 0.1 ascending steps, until 1.5, which is the point that problem becomes infeasible in this case. The RUC commitment schedule and the corresponding frequency response of every single outage for a sample hour are shown in Figure 7, for minimum reserve level (reserve level multiplier = 0) and maximum feasible reserve level (reserve level multiplier = 1.4). Then the ED is solved for 10 probable wind scenarios, using the robust UC solution for each reserve requirement level. At this stage, independent variables for the LR model can be picked up from ED results. Every possible generator outage n in the obtained ED solution is simulated with the SFR model, resulting in the corresponding frequency responses, to form the LR training data set. Considering the solved hourly RUCs and then for each of them hourly EDs of 10 wind scenarios, there will be around 20,000 possible single outages in total (every possible outage in every hour), building up the diverse training data set. The first RUC with a reserve level of zero is very fast. The corresponding ED problem receives fixed binary values from RUC, hence it's a linear problem with no integer values involved, so it is also very fast even when the quadratic cost function is used. As all the wind scenarios (Figure 6) are within the upper bound and lower bound of the RUC, it's certain that the ED for any scenario is feasible. To reduce solution time for higher levels of the reserve, the lower bound of the objective function can be set as the objective function of the previous level and a feasible binary solution can be attained from the previous level to give the current level a warm start. Obtained results confirm that other system characteristics, like online inertia, lost power, lost power percentage, and the droop of the turbine-governor system are more correlated with the quality of the frequency response, compared to the amount of reserve. The reason is that in systems with low inertia (like islands) frequency drops so fast after outages that the UFLS scheme is activated, although there is enough headroom in the remaining units (enough reserve). Table 3 shows the Pearson's correlation between mentioned characteristics and frequency response metrics, around 20,000 single outages that are simulated by the SFR model. For our case study, the whole process of building up the training data set and carrying out the dynamic simulations takes around 1 h. The SFR simulations are done by

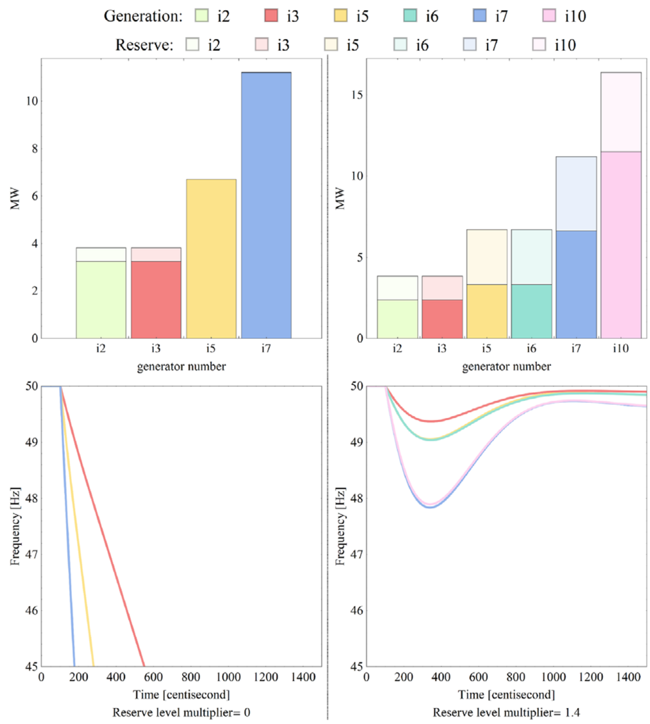

FIGURE 7 Schedule generation and reserve for different reserve levels and their corresponding frequency response after single outages for a sample hour

TABLE 3 Pearson's correlation between parameters

	f^{nadir}	f^{qss}	*RoCoF*
$\sum \mathcal{H}$	0.568	0.558	0.668
$\sum \mathcal{K}$	0.286	0.283	0.319
p^{loss}	−0.561	−0.532	−0.876
$\frac{p^{loss}}{d}$	−0.617	−0.588	−0.965
$\sum r$	0.506	0.516	0.269

TABLE 4 Logistic regression coefficients

Independent variable		LR coefficient	
–	–	c_0	26.577
ξ_1	$\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{H}_{ii}\mathcal{M}_{ii}^{base}x_{t,ii}$	c_1	−0.366
ξ_2	$\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{K}_{ii}x_{t,ii}$	c_2	0.102
ξ_3	$p_{t,i}$	c_3	1.484
ξ_4	$\frac{p_{t,i}}{d_t}$	c_4	−173.995
ξ_5	$\sum_{ii\neq i}^{ii\in\mathcal{I}} (\overline{\mathcal{P}_{ii}}x_{t,ii} - p_{t,ii})$	c_5	2.356

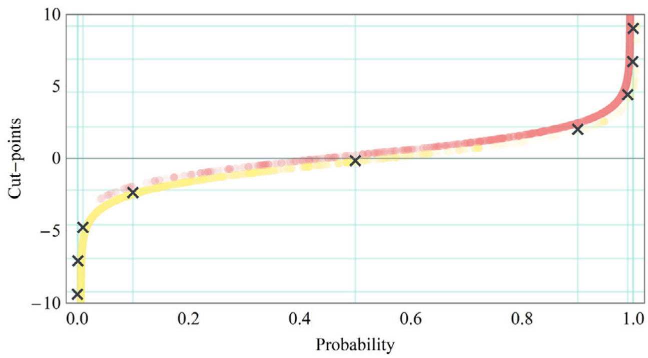

FIGURE 8 Logistic regression approximation

MATLAB Simulink.As expected, the ratio of lost generation to hourly demand has the best correlation with frequency metrics, as the big outages relative to the whole generation tend to disturb frequency considerably. Interestingly enough, the sum of the available reserves has a weaker correlation with frequency metrics, compared to the others. Meaning that fulfilling reserve criteria do not guarantee the quality of frequency response in small power systems with low inertia, as the remaining units are not fast enough to compensate for the power mismatch, while the frequency is dropping fast due to lack of inertia. So other parameters like total available inertia and power loss ratio are better representatives of the system dynamics.

From the results of the SFR model for every single outage, it is possible now to determine acceptable and unacceptable outages. At this stage generator outages n, which are followed by bad frequency responses are tagged with 1, and the outages that are followed by tolerable frequency outages are tagged with 0. This forms the dependent variable υ_n in the LR model. For the purpose of this paper, any generator outage incident which leads to frequency nadir less than 47.5 Hz, or a RoCoF higher than 1.5 Hz/s, or steady-state frequency less than 49.6 Hz, is considered an unacceptable incident and is tagged with 0. Other incidents are considered acceptable and are tagged with 1.

Independent variables that are chosen should have a good correlation with the frequency response metrics and have the ability to be used in the linear constraint. For this study, the presented parameters in Table 3 are defined as independent variables. Now that both independent variables and their associated dependent variables are acquired, the LR model can be trained to calculate the coefficients of Equation (9). The LR coefficients for the case study of La Palma island are presented in Table 4. These coefficients can be implemented to Equation (9), with an adjustable cut-point ψ to set up a new constraint. As discussed in Section 2.3, the logit form is a transformation of probabilities. In this case, incidents that are more probable to be acceptable should have a positive logit and a probability close to 1. On the other hand, incidents that are more probable to be unacceptable should have a negative logit and a probability close to 0. There will also be some errors, mainly around 0.5 probability, meaning that some acceptable incidents might end up possessing a negative logit value and vice versa. Depending on the preferred outcome, a proper cut-point can be chosen to create a more conservative or less conservative constraint. Figure 8 shows how accurately the applied logistic regression can separate acceptable and unacceptable results. Acceptable results are in red and unacceptable ones are in yellow.

Depending on the required conservativeness a cut-point is chosen. For example, $\psi = 0$ corresponds to $\pi(\upsilon = 1) = 0.5$. Putting $\psi = 0$, means all the incidents that their probability of being unacceptable is more than 0.5, will be eliminated, hence it's very conservative. A less conservative approach is to only eliminate the instances with the probability of being unacceptable more than 0.9 ($\pi(\upsilon = 1) \leq 0.1$). Then ψ should be set equal to −2.12 (considering Equation 8). Some probabilities and their corresponding cut-points are shown with the cross sign in Figure 8.

In Figure 9, it is shown how different independent variables, ξ_1 to ξ_5 (as defined in Table 4), are described by the logistic regression approximation. Those incidents that are marked as acceptable before are the red dots, and unacceptable incidents are the yellow dots. There are some errors, especially close to the *logit* $(\pi = 1)$ line, but the overall accuracy is justifiable. The summation of online inertia, ξ_1, is depicted in the Figure 9I. Acceptable results are more concentrated on the top side which are the incidents with higher online inertia, and as the online inertia drops, the dots move towards unacceptable results. A similar conclusion can be drawn for the summation of the droops of online turbine-governor systems, ξ_2, shown in Figure 9II. The amount of lost generation, ξ_3, is depicted in Figure 9III. As expected, larger outages tend to result in unacceptable incidents and as the figure goes toward smaller outages, the concentration of acceptable incidents grows. The same conclusion is derived from Figure 9IV, which shows the ratio of lost generation to hourly demand, ξ_4. The available reserve is depicted in Figure 9V. Generally incidents with a higher amount of online reserve tend to lead to better results, but still there are a considerable number of incidents that lead to unacceptable results, although they have a relatively high available reserve. This confirms that the available reserve is not the best indicator to ensure the quality of dynamic response after outages. The goal is to improve the quality of frequency response by

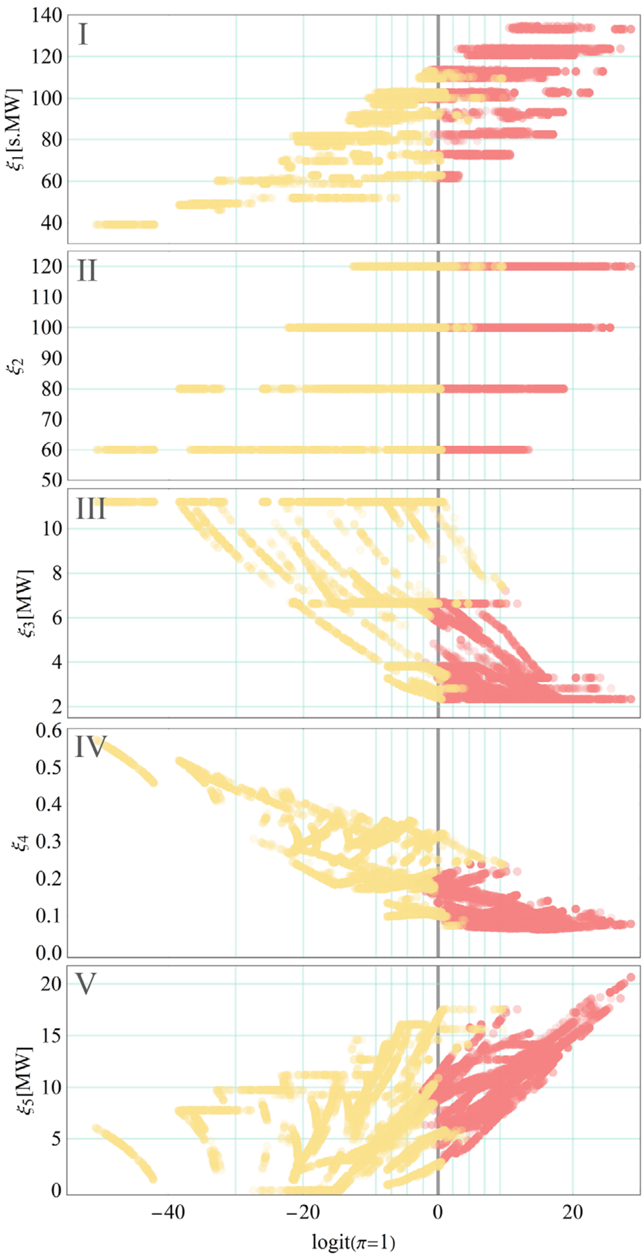

FIGURE 9 Logit transformation - variables. I) ξ_1, II) ξ_2, III) ξ_3, IV) ξ_4, V) ξ_5.

including all of these independent variables, each of them weighted carefully with logistic regression coefficients.

3.2 | Comparison of different methods

Simulations are carried out for three different methods:

3.2.1 | Conventional approach

The conventional formulation of robust UC, that the frequency response after an outage is only guaranteed by reserve criteria. Reserve requirement is the biggest online generation infeed.

3.2.2 | LR

The proposed logistic regression method. Reserve criteria are substituted with a constraint that is trained by the LR model. Different cut-points (ψ) are considered to assess the effectiveness of the proposed method, when the LR constraint is looser (smaller ψ) or tighter (bigger ψ).

3.2.3 | OCT

To also compare the proposed method with other recent data-driven methods in the literature, optimal classification trees are implemented to train a constraint, as introduced in [16]. The outputs of the SFR model are classified into acceptable and unacceptable incidents, using the MIL solution method of [33]. As solving the optimization problem for classification becomes very hard with a big set of inputs and a high depth of trees, only the biggest hourly outage of a limited number of scenarios is fed to the OCT problem as input, with the maximal depth of one and two.

3.2.4 | RoCoF

The RoCoF estimation from the swing equation is linear and can be directly used in MIL formulation. Different critical RoCoFs are considered to make comparison easier. More detail can be found in [10] and [12]. The added constraint to the conventional RUC is as follows,

$$2\left(\sum_{ii\neq i}^{ii\in\mathcal{I}} \mathcal{H}_{ii}\mathcal{M}_{ii}^{base}x_{t,ii}\right)\times\Delta f_{crit}\geq p_{t,i}t\in\mathcal{T},i\in\mathcal{I}. \quad (11)$$

A comparison of frequency response indicators for the conventional approach, LR, and OCT is presented in Table 5. Frequency-response quality indicators are the average amount of UFLS which is obtained from SFR with UFLS active, average frequency nadir, average RoCoF, and average quasi-steady-state frequency, which is obtained from SFR with UFLS deactivated. The changes in average UFLS and operation costs relative to the conventional approach are presented in percentage too. To better compare the overall frequency response quality of all presented methods, HQFR and LQFR are defined in percentage. LQFR is the percentage of incidents with low-quality frequency response, which are incidents with RoCoF higher than 1.5 Hz/s or frequency nadir lower than 47.5 Hz or quasi-steady-state frequency lower than 49.6 Hz. The rest of the incidents are counted as HQFR, or High-quality frequency response. Cut-points beyond $\psi = 2.12$ make the problem infeasible, so $\psi = 2.12$ is presented in the table as the most conservative cut-point that is feasible. The results assert that more conservative approaches lead to higher operation costs. But depending on the chosen cut-point, the proposed approach can sometimes lead to better frequency response quality, while keeping the operation costs relatively low. As it can be seen in Table 5,

TABLE 5 Results of the simulations on La Palma island

	HQFR (%)	LQFR (%)	Average f^{qss}	Average f^{nadir}	Average RoCoF	Average UFLS	Operation cost	Run-time (s)
Conventional approach	67.6%	32.4%	49.61 Hz	48.29 Hz	−1.31 Hz/s	2.30 MW	140.61 k€	$1,422''$
LR with $\psi = 2.12$	81.8%	18.2%	49.81 Hz	48.81 Hz	−1.00 Hz/s	1.14 MW (−50.5%)	145.26 k€ (+3.3%)	$1,819''$
LR with $\psi = 0$	81.3%	18.7%	49.82 Hz	48.77 Hz	−1.03 Hz/s	1.23 MW (−46.5%)	143.68 k€ (+2.1%)	$2,092''$
LR with $\psi = -2.12$	78.9%	21.1%	49.81 Hz	48.68 Hz	−1.09 Hz/s	1.47 MW (−36.1%)	142.90 k€ (+2.3%)	$2,345''$
LR with $\psi = -4.95$	78.2%	21.8%	49.74 Hz	48.57 Hz	−1.12 Hz/s	1.75 MW (−23.9%)	141.32 k€ (+0.5%)	$1,965''$
LR with $\psi = -5$	73.4%	26.6%	49.73 Hz	48.48 Hz	−1.18 Hz/s	2.03 MW (−11.7%)	140.78 k€ (+0.1%)	$1,411''$
LR with $\psi = -6.91$	76.8%	23.2%	49.66 Hz	48.52 Hz	−1.13 Hz/s	2.06 MW (−10.4%)	139.83 k€ (−0.6%)	$1,015''$
LR with $\psi = -9.21$	69.8%	30.2%	49.26 Hz	48.09 Hz	−1.24 Hz/s	2.20 MW (−4.3%)	138.53 k€ (−1.5%)	$794''$
LR with $\psi = -10$	64.2%	35.8%	47.97 Hz	46.96 Hz	−1.28 Hz/s	2.44 MW (+6.1%)	136.86 k€ (−2.7%)	$847''$
LR with $\psi = -11.51$	61.9%	38.1%	47.91 Hz	46.85 Hz	−1.36 Hz/s	2.61 MW(+13.5%)	136.67 k€ (−2.8%)	$761''$
OCT, $d = 1$, $N = 1001$	79.2%	20.8%	49.80 Hz	48.75 Hz	−1.05 Hz/s	1.31 MW (−43.0%)	144.33 k€ (+2.6%)	$1,750''$
OCT, $d = 2$, $N = 1001$	79.1%	19.9%	49.81 Hz	48.77 Hz	−1.04 Hz/s	1.29 MW (−43.9%)	145.09 k€ (+3.2%)	$3,144''$
OCT, $d = 1$, $N = 2800$	79.2%	20.8%	49.80 Hz	48.76 Hz	−1.05 Hz/s	1.30 MW (−43.5%)	144.12 k€ (+2.5%)	$1,807''$
RoCoF ($\Delta f_{crit} = 1.5$)	88.4%	11.6%	49.80 Hz	48.81 Hz	−0.91 Hz/s	1.17 MW (−49.1%)	151.41 k€ (+7.1%)	$1,104''$
RoCoF ($\Delta f_{crit} = 2$)	75.3%	24.7%	49.79 Hz	48.60 Hz	−1.16 Hz/s	1.41 MW (−36.8%)	144.40 k€ (+2.7%)	$1,984''$
RoCoF ($\Delta f_{crit} = 2.5$)	69.3%	30.7%	49.78 Hz	48.42 Hz	−1.26 Hz/s	2.01 MW (−12.6%)	142.10 k€ (+1.1%)	$1,620''$

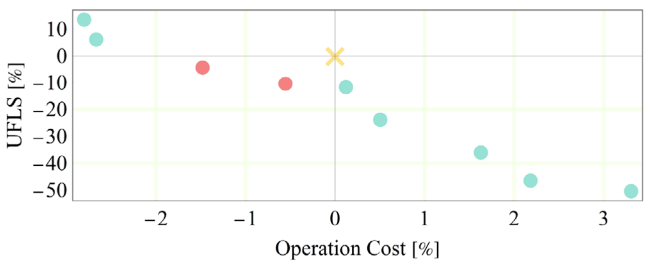

FIGURE 10 Average UFLS and operation cost in percentage

TABLE 6 Comparison of the training process

Method	N	Inaccuracy	Run-time
LR	19860	3.71%	00'03"
OCT, $d = 1$	1001	1.15%	00'32"
OCT, $d = 2$	1001	0.1%	28'07"
OCT, $d = 1$	2800	2.07%	42'06"

more conservative cut-points lead to less percentage of LQFR. Each column in the table is compared with the conventional approach. The ones that perform better than the conventional approach are underlined with red lines, and the ones that perform worse are underlined with yellow lines. The results also show that the proposed approach can guarantee a better frequency response quality if a proper cut-point is chosen. Depending on the required level of cautiousness, the operator can choose a cut-point. For the La Palma island, a probability assurance of $\psi = -6.91$ seems appealing, because both frequency response quality and operation cost are improved.

To better compare and choose the best ψ, all the simulated cases of La Palma island are compared with the conventional approach (highlighted with a yellow cross) in Figure 10. Although the operation costs go higher by choosing ψ closer to zero, the average UFLS is decreased considerably. Also, there are cases that lead to improvement in both operation cost and average UFLS, which are highlighted in red.

The results for OCT in Table 5 show improvements in the quality of frequency response compared to the conventional approach and LR with some cut-points. d is the depth of the tree structure. OCT with $d = 1$ leads to one set of constraints (so the size of the UC problem will remain the same), and OCT with $d = 2$ leads to two sets of constraints. Although OCT is very accurate in classifying the inputs, the run-time of the optimization problem relies heavily on the number of inputs and the depth of the tree structure. For that reason making the training set smaller was necessary. Solving OCT with a full set of training set (around 20,000 points) can take many days. So only the biggest hourly outages of some scenarios are considered (like in [16]), creating two training data-sets, one smaller with 1001 points, and one bigger with 2800 points. A comparison between the accuracy of representing the data set and solution run-time is presented in Table 6. The downside of a small training set for this practice is that more unacceptable incidents might be flagged as acceptable and vice versa. As it can be seen in Table 6, the advantage of OCT compared to LR is the superior accuracy in classifying the training set and the OCT disadvantage compared to LR is the computational burden of the training process, which effectively limits the size of the training set. Also, tuning the initial values in the OCT optimization problem is hard, and time-consuming. More discussion about this can be found in [33].

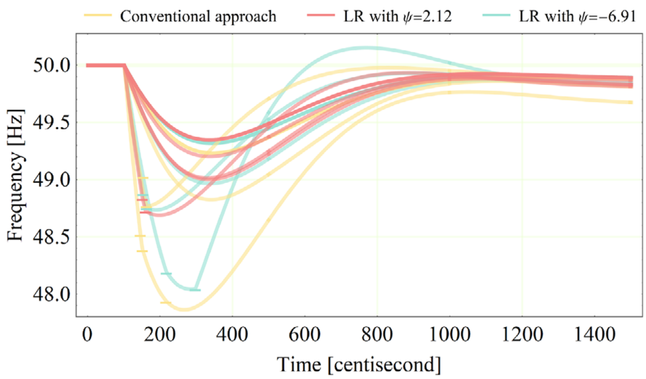

FIGURE 11 Frequency response after outages for a random hour with UFLS

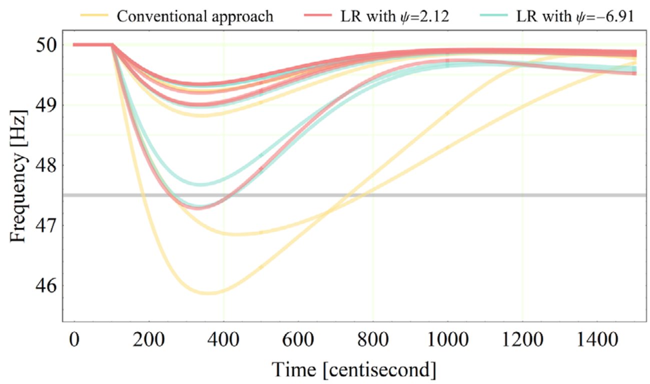

FIGURE 12 Frequency response after outages for a random hour without UFLS

Simulations are also carried out for RUC with reserve constraint, plus the linear RoCoF constraint, considering different critical values for RoCoF. The problem is infeasible for Δf_{crit} below 1.5 Hz/s. This criterion keeps the RoCof of any outage, under the defined Δf_{crit}. Yet it cannot prevent some of the unacceptable f^{qss} and f_{nadir} from happening. Also, the operation cost is considerably higher, because so many units should be online at the same time to keep RoCoF within boundaries after outages. Higher Δf_{crit} results are less conservative, but as expected they will cost more than the conventional approach. As it can be concluded from the results, even for the most restricted criteria, the percentage of LQFR is impossible to be zero for La Palma island. There is no feasible solution that can guarantee only HQFR incidents. Simply starting up more units would not solve the issue, as the units should maintain the minimum power output constraint, which effectively limits the maximum number of online units. For all the simulations in this paper, a computer with Intel core i7-8700 CPU and 32 GB installed RAM is used. All of the MILP problems (RUC and OCT), and ED which are quadratic, are solved by CPLEX solver in GAMS. The run-time of each respective method is presented in Table 5 as well. Different things might affect the solution-time of an MILP problem, including: tightness of the solution domain, compactness of the problem (problem size), lower bound of the objective function that the solver finds, number of feasible cuts, optimality gap, and etc.

It's also interesting to see and compare the dynamic frequency responses obtained from the SFR model. In Figures 11 and 12, the frequency response for a period of 15 s after outages are presented, for every single outage of online units in a random hour. In Figure 11 the UFLS scheme is activated, and Figure 12 shows frequency responses with no UFLS. The simulations for the conventional approach are in yellow, the most conservative case with $\psi = 2.12$ in red, and one of the preferred cases with $\psi = -6.91$ in green. The moments that the UFLS scheme has operated are also highlighted with dashes. The better performance of the conservative case is noticeable. Also, the case with $\psi = -6.91$ outperforms the conventional approach. The minimum allowed frequency nadir is shown with the gray line in Figure 12.

4 | CONCLUSION

This paper proposes a novel procedure to schedule short-term unit commitment in island power systems. Island power systems usually suffer from a lack of inertia and frequency response capacity, complicating containing frequency within an acceptable range during large disturbances. The proposed method uses an initial data set to train a linear constraint that takes into account the dynamic response of the system. For the purpose of training this constraint, logistic regression is employed to avoid incidents with undesirable frequency responses as much as possible. Then the logistic regression constraint is included in an adaptive robust formulation. Results show that by choosing a proper cut-point, the proposed method improves the frequency response, as well as the operation costs. As training data with the LR model is very fast, the size of the training set is not an issue. A complete training data set can better represent the system, leading to a more reliable frequency constraint. The proposed approach is compared with OTC as a method to classify the training set and also with linear RoCoF constraint. In both cases, the proposed method is more cost-efficient. The operator can use the proposed constraint in the UC problem to improve the frequency response of the system after outages, instead of lengthy analytical approaches.

NOMENCLATURE

Acronyms

DNN	Deep neural network
ED	Economic Dispatch
GAMS	General algebraic modeling system
HQFR	High Quality Frequency Response
LQFR	Low Quality Frequency Response
LR	Linear Regression
OCT	Optimal Classification Trees
RES	Renewable Energy Sources
RoCoF	Rate of Change of Frequency
RUC	Robust Unit Commitment
SFR	System Frequency response Model
SVM	Support vector machine

UC Unit Commitment
$UFLS$ Under Frequency Load Shedding

SFR model

Δd Load deviation [p.u.]
Δp Mechanical power deviation [p.u.]
$\Delta \dot{w}_i$ frequency deviation [p.u.]
$\mathcal{H}$ Inertia [s]
$\mathcal{K}$ Normalized gain of turbine-governor model
$\mathcal{M}$ Base rated power of units [MW]
$\mathcal{S}$ Base power of the system [MW]
$\widetilde{\Delta d}$ Total load deviation [p.u.]
$\widetilde{\Delta p}$ Total mechanical power deviation [p.u.]
$a_.$ Poles of the second order system
$b_.$ Zeros of the second order system
k Inverse of the droop [p.u.]
N Total number of contingencies
n Index of contingency

LR model

$\pi(.)$ Probability of .
ψ Regression cut-point
υ Dependant variable
ξ Independent variable
c Regression coefficient
f^{nadir} The minimum value of frequency reached during the transient period
f^{qss} Quasi steady-state frequency

Robust UC

α Dual variable of minimum power constraint
β Dual variable of maximum power constraint
Δf_{crit} Critical RoCoF [Hz/s]
δ Dual variable of up ramp constraint
η Dual variable of power balance constraint
γ Dual variable of down ramp constraint
$\mathcal{I}$ Set of all generators
$\mathcal{T}$ Set of all time intervals
$\mathcal{W}$ Set of Wind generation uncertainty
μ Dual variable of minimum reserve constraint
$\overline{\mathcal{P}_i}$ Maximum power output of generator i [MW]
$\overline{\mathcal{R}_i}$ Maximum ramp-up of generator i [MW]
ρ Dual variable of LR constraint
$\underline{\mathcal{P}_i}$ Minimum power output of generator i [MW]
$\underline{\mathcal{R}_i}$ Maximum ramp-down of generator i [MW]
ζ Dual variable of maximum wind constraint
DT Minimum down-time of generators [hours]
gc Generation costs [€]
i Index of generators
ii Alias index for generators
p Power generation variable [MW]
r Online reserve power variable [MW]
$suc(.)$ Start-up costs [€]
t Index of time intervals
tt Alias index for time intervals
UT Minimum up-time of generators [hours]
w Available forecasted wind power [MW]
wg Wind generation variable [MW]
x Commitment variable [$\in\{0,1\}$]
y Start-up variable [$\in\{0,1\}$]
z Shut-down variable [$\in\{0,1\}$]

AUTHOR CONTRIBUTIONS

Mohammad Rajabdorri: Investigation, methodology, validation, visualization, writing - original draft. Enrique Lobato Miguéelez: Supervision, validation, writing - review and editing. Lukas Sigrist: Supervision, validation, writing - review and editing.

ACKNOWLEDGEMENTS

This study is funded by European Regional Development Fund (ERDF), Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación, Project RTI2018-100965-A-I00.

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest to mention.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mohammad Rajabdorri https://orcid.org/0000-0002-4042-7442
Lukas Sigrist https://orcid.org/0000-0003-2177-2029

REFERENCES

1. Zobaa, A.F., Aleem, S.A.: Uncertainties in Modern Power Systems. Academic Press, San Diego (2020)
2. Ning, C., You, F.: Data-driven adaptive robust unit commitment under wind power uncertainty: A bayesian nonparametric approach. IEEE Trans. Power Syst. 34(3), 2409–2418 (2019)
3. Zhang, M., Fang, J., Ai, X., Zhou, B., Yao, W., Wu, Q., Wen, J.: Partition-combine uncertainty set for robust unit commitment. IEEE Trans. Power Syst. 35(4), 3266–3269 (2020)
4. Cho, Y., Ishizaki, T., Imura, J.I.: Three-stage robust unit commitment considering decreasing uncertainty in wind power forecasting. IEEE Trans. Ind. Inf. 18(2), 796–806 (2021)
5. Habibi, M., Vahidinasab, V., Aghaei, J., Mohammadi.Ivatloo, B.: Assessment of energy storage systems as a reserve provider in stochastic network constrained unit commitment. IET Smart Grid 4(2), 139–150 (2021)
6. Huang, H., Zhou, M., Li, G.: An endogenous approach to quantifying the wind power reserve. IEEE Trans. Power Syst. 35(3), 2431–2442 (2019)
7. Prakash, V., Kushwaha, P., Sharma, K.C., Bhakar, R.: Frequency response support assessment from uncertain wind generation. Int. J. Electr. Power Energy Syst. 134, 107465 (2022)
8. Rajabdorri, M., Sigrist, L., Lobato.Miguélez, E., Prats.Soriano, M.d.C., Echavarren.Cerezo, F.: Viability of providing spinning reserves by res in spanish island power systems. IET Renewable Power Gener. 15(13), 2878–2890 (2021)
9. Trovato, V., Bialecki, A., Dallagi, A.: Unit commitment with inertia-dependent and multispeed allocation of frequency response services. IEEE Trans. Power Syst. 34(2), 1537–1548 (2018)

10. Badesa, L., Teng, F., Strbac, G.: Simultaneous scheduling of multiple frequency services in stochastic unit commitment. IEEE Trans. Power Syst. 34(5), 3858–3868 (2019)
11. Paturet, M., Markovic, U., Delikaraoglou, S., Vrettos, E., Aristidou, P., Hug, G.: Stochastic unit commitment in low-inertia grids. IEEE Trans. Power Syst. 35(5), 3448–3458 (2020)
12. Pérez.Illanes, F., Álvarez.Miranda, E., Rahmann, C., Campos.Valdés, C.: Robust unit commitment including frequency stability constraints. Energies 9(11), 957 (2016)
13. Mousavi.Taghiabadi, S.M., Sedighizadeh, M., Zangiabadi, M., Fini, A.S.: Integration of wind generation uncertainties into frequency dynamic constrained unit commitment considering reserve and plug in electric vehicles. J. Cleaner Prod. 276, 124272 (2020)
14. Rabbanifar, P., Amjady, N.: Frequency-constrained unit-commitment using analytical solutions for system frequency responses considering generator contingencies. IET Gener. Transm. Distrib. 14(17), 3548–3560 (2020)
15. Shahidehpour, M., Ding, T., Ming, Q., Catalao, J.P., Zeng, Z.: Two-stage chance-constrained stochastic unit commitment for optimal provision of virtual inertia in wind-storage systems. IEEE Trans. Power Syst. 36(4), 3520–3530 (2021)
16. Lagos, D., Hatziargyriou, N.D.: Data-driven frequency dynamic unit commitment for island systems with high res penetration. IEEE Trans. Power Syst. 36(5), 4699–4711 (2021)
17. Zhang, Y., Chen, C., Liu, G., Hong, T., Qiu, F.: Approximating trajectory constraints with machine learning-microgrid islanding with frequency constraints. IEEE Trans. Power Syst. 36(2), 1239–1249 (2021)
18. Zhang, Y., Cui, H., Liu, J., Qiu, F., Hong, T., Yao, R., Li, F.F.: Encoding frequency constraints in preventive unit commitment using deep learning with region-of-interest active sampling. IEEE Trans. Power Syst. 37(3), 1942–1955 (2021)
19. Shen, Y., Wu, W., Wang, B., Yang, Y., Lin, Y.: Data-driven convexification for frequency nadir constraint of unit commitment. J. Mod. Power Syst. Clean Energy (2022)
20. Chiu, H.W., Chang Chien, L.R., Wu, C.C.: Construction of a frequency compliant unit commitment framework using an ensemble learning technique. Energies 14(2), 310 (2021)
21. Hilbe, J.M.: Practical Guide to Logistic Regression. CRC Press, Boca Raton (2016)
22. Khosravi, P., Liang, Y., Choi, Y., Broeck, G.V.d.: What to expect of classifiers? reasoning about logistic regression with missing features. arXiv preprint arXiv:190301620 (2019)
23. Bertsimas, D., Litvinov, E., Sun, X.A., Zhao, J., Zheng, T.: Adaptive robust optimization for the security constrained unit commitment problem. IEEE Trans. Power Syst. 28(1), 52–63 (2012)
24. Morales.España, G., Lorca, Á., de Weerdt, M.M.: Robust unit commitment with dispatchable wind power. Electr. Power Syst. Res. 155, 58–66 (2018)
25. Taşkin, Z.C.: Benders decomposition. Wiley Encyclopedia of Operations Research and Management Science (January 2011)
26. Lahaie, S.: How to Take the Dual of a Linear Program. Columbia University, New York (2008)
27. Fletcher, R., Leyffer, S.: Solving mixed integer nonlinear programs by outer approximation. Math. Program. 66(1), 327–349 (1994)
28. Sigrist, L., Lobato, E., Echavarren, F.M., Egido, I., Rouco, L.: Island Power Systems. CRC Press, Boca Raton (2016)
29. Kleinbaum, D.G., Kupper, L.L., Nizam, A., Rosenberg, E.S.: Applied Regression Analysis and Other Multivariable Methods. Cengage Learning, Boston (2013)
30. Huang, F.L., Moon, T.R.: What are the odds of that? A primer on understanding logistic regression. Gifted Child Q. 57(3), 197–204 (2013)
31. de Canarias, A.E.: Consejería de transición ecológica, lucha contra el cambio climático y planificación territorial. Gobierno de Canarias (2019)
32. Pandzic, H., Dvorkin, Y., Qiu, T., Wang, Y., Kirschen, D.S.: Unit commitment under uncertainty -gams models (2015)
33. Bertsimas, D., Dunn, J.: Optimal classification trees. Mach. Learn. 106(7), 1039–1082 (2017)

How to cite this article: Rajabdorri, M., Lobato, E., Sigrist, L.: Robust frequency constrained uc using data driven logistic regression for island power systems. IET Gener. Transm. Distrib. 16, 5069–5083 (2022). https://doi.org/10.1049/gtd2.12658