Engineering Applications of Artificial Intelligence 139 (2025) 109617

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Research paper

Data-driven estimation of the amount of under frequency load shedding in small power systems

Mohammad Rajabdorri [a,*], Matthias C.M. Troffaes [b], Behzad Kazemtabrizi [b], Miad Sarvarizadeh [a], Lukas Sigrist [a], Enrique Lobato [a]

[a] IIT, Comillas Pontifical University, Spain
[b] Durham University, UK

ARTICLE INFO

Keywords:
Novel regression tree
Tobit model
Data-driven model
Island power systems
Machine learning
Under frequency load shedding

ABSTRACT

This paper presents a data-driven methodology for estimating under frequency load shedding (UFLS) in small power systems. UFLS plays a vital role in maintaining system stability by shedding load when the frequency drops below a specified threshold following loss of generation. Using a dynamic system frequency response (SFR) model we generate different values of UFLS (i.e., labels) predicated on a set of carefully selected operating conditions (i.e., features). Machine learning (ML) algorithms are then applied to learn the relationship between chosen features and the UFLS labels. A novel regression tree and the Tobit model are suggested for this purpose and we show how the resulting non-linear model can be directly incorporated into a mixed integer linear programming (MILP) problem. The trained model can be used to estimate UFLS in security-constrained operational planning problems, improving frequency response, optimizing reserve allocation, and reducing costs. The methodology is applied to the La Palma island power system, demonstrating its accuracy and effectiveness. The results confirm that the amount of UFLS can be estimated with the mean absolute error (MAE) as small as 0.179 MW for the whole process, with a model that is representable as a MILP for use in scheduling problems such as unit commitment among others.

1. Introduction

Synchronous generators are being displaced with cleaner, albeit non-synchronously coupled alternatives (like wind and solar), which inadvertently has led to a reduction of inertia, hence incorporating frequency dynamics into the operational planning of power systems is more important than ever. Island power systems are already suffering from a lack of inertia because of their small size. There has been extensive research on how to include frequency dynamics in scheduling optimization problems. Both analytical methods (directly from the swing equation) and data-driven methods (based on dynamic simulations) have been proposed to obtain frequency constraints for inclusion in the operational planning process. Typical frequency response metrics after outages are the rate of change of frequency (RoCoF), quasi-steady-state frequency, and frequency nadir. However, calculating the frequency nadir is much more complicated than the other metrics. To derive the frequency nadir from the swing equation, some simplifying assumptions are needed, and still, the obtained equation is non-linear and non-convex, which makes it challenging to be used in MILP problem formulations. A common assumption in analytical frequency-constrained methods like Trovato et al. (2018), Badesa et al. (2019), Paturet et al. (2020), Shahidehpour et al. (2021), Ferrandon-Cervantes et al. (2022), and many other similar works, is that the provision of reserve increases linearly in time, and all units will deliver their available reserve within a given fixed time. Consequently, the ensuing complicated analytical methods are not necessarily accurate. On the other hand, more recently, ML-based methods have been proposed to incorporate frequency dynamics. For instance, optimal classifier tree is used in Lagos and Hatziargyriou (2021), deep neural network is used in Zhang et al. (2021), and logistic regression is used in Rajabdorri et al. (2022), among other approaches. An analytical frequency constrained unit commitment (UC) is compared with data-driven models with the help of ML in Rajabdorri et al. (2023) and their pros and cons are highlighted.

In smaller systems like islands the frequency can easily exceed the safe threshold after any contingency because usually online units are providing a considerable percentage of the whole demand. To maintain the frequency stability of an electrical power system, UFLS schemes are implemented to shed or disconnect a certain amount of load in predefined steps when the frequency drops below a specified

* Corresponding author.
E-mail address: mrajabdorri@comillas.edu (M. Rajabdorri).

https://doi.org/10.1016/j.engappai.2024.109617
Received 6 March 2024; Received in revised form 3 October 2024; Accepted 3 November 2024

Acronyms

ED	economic dispatch
KDE	kernel density estimate
MAE	mean absolute error
MILP	mixed integer linear programming
ML	machine learning
RES	renewable energy sources
RoCoF	rate of change of frequency
SFR	system frequency response
UC	unit commitment
UFLS	under frequency load shedding

threshold following disturbance events. This corrective protection measure helps to balance the power supply and demand and prevents a complete system blackout. Different methods have been introduced to tune and optimize the UFLS scheme for electrical power systems, which can be categorized into conventional and adaptive methods. Conventional UFLS schemes use fixed load shedding steps (Ketabi and Fini, 2014; Laghari et al., 2014; Wang et al., 2022; Kalajahi et al., 2021), while adaptive UFLS schemes dynamically adjust the load shedding amount based on real-time system conditions (Mehrabi et al., 2018; Li et al., 2019; Tofis et al., 2016; Silva and Assis, 2020). Although adaptive schemes provide a more optimized and flexible response, as they require more advanced monitoring systems and computational capabilities for real-time monitoring and optimization, they are still not used widely in practice. The performance of both conventional and adaptive methods can be improved by incorporating ML (Hooshmand and Moazzami, 2012; Golpîra et al., 2022).

Depending on the size of the system, it is possible to prevent UFLS activation. Many studies, like Zhang et al. (2021), Chang et al. (2012), Sedighizadeh et al. (2019), Pérez-Illanes et al. (2016), and others set the frequency dynamic thresholds high enough, so no outage leads to UFLS. This is not possible in a small system, where every online unit is providing a substantial percentage of the demand and any outage can be big enough to trigger the UFLS activation (Rajabdorri et al., 2022). In such systems, co-optimizing UFLS activation and scheduling of the units can have some benefits like:

- Having an estimate of UFLS in the scheduling optimization problem will prevent incidents with poor frequency responses.
- The estimated amount of UFLS can be deduced from the required reserve. There is no need to schedule reserve as much as the biggest outage if eventually the UFLS will be activated after the outage.
- UFLS can be monetized easily and added directly to the objective function of the optimization problem, to reduce the overall operation costs.

This paper tries to estimate the amount of UFLS, regardless of the UFLS scheme that is used for the system, through a learning process. The estimation of the amount of UFLS of conventional schemes is however very complicated because of its discrete nature and disturbance-dependent behavior. Given the widespread use of conventional schemes in real systems, it is the principal focus of this paper. The dataset used for the learning process is labeled with the UFLS of every sample generation combination in the dataset. The labels can be obtained by SFR models or any other power system simulator. The purpose is to use the estimation of UFLS in the operational planning of small power systems (such as generation UC, economic dispatch (ED), reserve allocation, ancillary service scheduling, renewable energy sources (RES) integration, and so on). The operational planning process is usually modeled and solved as an MILP problem. Therefore, it is convenient to limit the hypothesis space of the UFLS estimation models to models that are representable by MILP. Including UFLS estimation in the problem is in a sense equivalent to including frequency dynamics in the operational planning, because poor frequency response subsequently triggers the activation of UFLS.

To estimate the amount of UFLS, a dataset generation process is proposed to acquire a set of operating points (potential hourly generation schedules) that can properly describe the system under study. Every possible outage in each set of generation schedules is labeled with its corresponding UFLS. The obtained dataset is carefully analyzed to choose the representative features, and then a learning process is proposed. A regression tree model with a novel partitioning algorithm is suggested in this paper. As the UFLS is activated in steps and discretely sheds load, a regression tree seems most suitable. Our partitioning algorithm exploits the data structure to most effectively represent the regression tree as a MILP (Maragno et al., 2021). Also, the use of the Tobit model to estimate UFLS is proposed and studied. Although the Tobit model is typically used to describe censored data (Tobin, 1958), we found it to be effective to also describe the UFLS, which has a cluster around zero followed by a linearly increasing trend as a function of the features, very similar to zero censored datasets. Additionally, the Tobit model has a simple analytic structure that makes it easy to incorporate into a MILP. Both the suggested regression tree and the Tobit model, alongside their MILP representation, are demonstrated in the following.

To the best of the author's knowledge, there is little careful analysis on UFLS estimation or an MILP representation of UFLS in the existing literature. Teng and Strbac (2017) introduces a model for optimal stochastic scheduling, integrating energy production with operating reserve, frequency response, and UFLS. The optimal amount of UFLS is determined, based on the assumption of a linear increase in generation over time and a specified outage size. The authors in O'Malley et al. (2022) introduce a novel constraint based on the swing equation to manage frequency nadir in low-carbon power grids, incorporating fast frequency response, dynamically reduced largest loss, and UFLS. The research demonstrates that incorporating UFLS in frequency security planning can significantly reduce operational costs. A corrective frequency-constrained UC model for island power systems is introduced in Rajabdorri et al. (2024) that incorporates analytical constraints on UFLS, demonstrating its ability to lower generation costs while minimizing the expected UFLS.

In contrast to the methods previously presented in the literature, this paper proposes a model to estimate the real UFLS schemes used in most island power systems, which are step-wise and contain time delays, through data-driven methods. Table 1 gives a summary of the reviewed literature, highlighting the main differences compared to this paper.

The rest of the paper is organized as follows; in Section 2 the methodology of the paper is introduced, including the data generation (in Section 2.1), labeling the data (in Section 2.2), and the learning process (in Section 2.3). Then in Section 3 the results for the island under study (La Palma) are presented, including the data generation and analysis (in Section 3.1), the applied learning process, and its accuracy (in Section 3.2). Finally, conclusions are drawn in Section 4.

2. Methodology

2.1. Data generation

To estimate the UFLS, a proper set of data is necessary. The training dataset comprises features $x \in \mathcal{X}$ and labels $y \in \mathcal{Y}$. In the case of implementing the estimation of UFLS in the scheduling problems (like UC), features are any measurable quantities from the power system that might help predict the UFLS. These features are extracted from generation combinations, while the labels are obtained from dynamic simulations of UFLS after outages. These measurements can be obtained by solving high-order differential swing equations, or by using SFR

Table 1
Comparison of the reviewed literature.

Paper	UFLS modeled	MILP representation	Real scheme	Adaptive scheme
Trovato et al. (2018), Badesa et al. (2019), Paturet et al. (2020), Shahidehpour et al. (2021), Ferrandon-Cervantes et al. (2022), Lagos and Hatziargyriou (2021), Zhang et al. (2021), Rajabdorri et al. (2022, 2023)	✗	✗	✗	✗
Teng and Strbac (2017)	✓	✗	✗	✓
O'Malley et al. (2022)	✓	✗	✗	✓
Rajabdorri et al. (2024)	✓	✓	✗	✓
This paper	✓	✓	✓	✗

models. The features should be carefully chosen to represent a reasonable amount of information about their labels. Using an unnecessarily large number of features can be detrimental to both computational and statistical aspects. Selecting a large feature vector increases the dimensions of the problem, thereby requiring more resources for calculations. In addition, using a higher number of features makes the model more susceptible to overfitting. Therefore, it is beneficial to use only the features with the most relevant information to predict the label y (Jung, 2018). In this paper, y is the amount of UFLS for each outage. Several methods have been introduced in the literature to reduce the size of the feature vector. For this paper, the features must be accessible throughout the scheduling process. Therefore, variables that are most correlated with the label will be chosen as features. As shown later in Section 3, the selected features for predicting UFLS are available inertia ($\mathcal{H}_g$), weighted gain of turbine–governor model ($\mathcal{K}_g$), the amount of lost power (P_g), and the amount of available reserve ($\mathcal{R}_g$), after the outage of generator g.

To obtain a complete dataset, every combination of possible generation outputs of the units can be considered. However, many of these combinations are infeasible as they do not satisfy the constraints that are used in the scheduling process (power balance, reserve constraint, or maximum RoCoF), or are unappealing as the optimization problem will favor cheaper combinations. In this paper, a data generation method is used to only generate feasible control points that are cost-effective, and hence more likely to be scheduled in the real operation. The process is outlined in Algorithm 1.

Algorithm 1 Synthetic Data Generation

Inputs: $\overline{D}$, $\underline{D}$, $\overline{P}_i$, Δf, f_0, H_i, M_i
Output: $\mathcal{F}$: set of feasible power level vectors

$\mathcal{F} \leftarrow \emptyset$
for $\vec{p} \in \bigtimes_{i=1}^{I} \mathcal{P}_i$ **do** ▷ for all power level vectors
  for $i \in \{1, \dots, I\}$ **do** ▷ for every generator
    $u_i := 0$ if $p_i = 0$ else 1 ▷ status of unit
  end for
  $G := \sum_{i=1}^{I} p_i$ ▷ total generation
  $R_\ell := \left(\sum_{i=1}^{I} u_i(\overline{P}_i - p_i)\right) - u_\ell(\overline{P}_\ell - p_\ell)$ ▷ reserve after outage of ℓ
  $H_\ell^{\text{sys}} := (\sum_{i=1}^{I} H_i M_i u_i) - H_\ell M_\ell u_\ell$ ▷ inertia after outage of ℓ
  if $\underline{D} \le G \le \overline{D}$ and $R_\ell \ge p_\ell$ and $H_\ell^{\text{sys}} \ge \frac{p_\ell f_0}{2\Delta f}$ **then** ▷ feasible?
    $\mathcal{F} \leftarrow \mathcal{F} \cup \{\vec{p}\}$ ▷ add power level vector
  end if
end for
Sort $\mathcal{F}$ ascending by the quadratic generation cost function
Keep a reasonable number of cheaper combinations and remove the rest

$\overline{D}$ and $\underline{D}$ are upper and lower bounds on yearly thermal generation (MW), i is the index of unit $\in \{1, \dots, I\}$, $\overline{P}_i$ is the capacity of unit i (MW), Δf is critical RoCoF (Hz/s), f_0 is nominal frequency (Hz), $\mathcal{P}_i$ is the finite set of power levels of unit i including level 0 for not committed (MW), ℓ is the index of the lost unit (can be any i), H_i is the inertia of unit i, and M_i is the base power of unit i.

2.2. Labeling the data

In labeling the data, the SFR model is used to analyze the frequency stability of small isolated power systems, such as the La Palma Island system being studied. The SFR model can reflect the short-term frequency response of such systems, but other dynamic power system models could also be employed. The power-system model, which is typically used to design UFLS schemes for an island power system consisting of I generating units, is detailed in Fig. 1. A second-order model approximation is used to represent the turbine–governor system of each generating unit (i). The dynamic frequency responses are mainly influenced by the rotor and turbine–governor system dynamics, while excitation and generator transients are ignored due to their faster dynamics. The load-damping factor (D) is used to consider the overall response of the loads, provided that its value is known. The gain (k_i), which is the inverse of the droop, and the parameters ($a_{i,1}$, $a_{i,2}$, $b_{i,1}$, and $b_{i,2}$) of each generating unit (i) can be determined from more precise models or field tests. The gain (k_i) is an essential parameter to indicate the frequency response of unit i, and will influence the UFLS scheme activation. To have features that can reflect the amount of UFLS after the outage, a weighted gain is defined, which will be used as a feature for the training dataset. The equation,

$$\mathcal{K}_g = \sum_{i \in \mathcal{I}, i \neq g} k_i \mathcal{M}_i u_{t,i} \tag{1}$$

represents the weighted gain after the outage of unit g. Due to the limited primary spinning reserve, the units' power output is restricted by the power output limitations $\Delta p_{i,min}$ and $\Delta p_{i,max}$, and the ramp-up speed of the units should be constrained by the maximum ramping capacity of each respective unit. The complete model is explained in Sigrist et al. (2016).

2.3. Learning process

2.3.1. Proposed regression tree

A regression tree is suggested in this paper to estimate the amount of UFLS since conventional schemes shed load in a discrete manner. Typical regression trees split the feature space into rectangular cells and use a constant value within each cell for prediction. However, if we want to incorporate the model into a MILP, we prefer to have as few cells as possible. Therefore, a novel regression tree is proposed here, which is inspired by Verwer et al. (2017) but deviates from it in several ways:

- The data shows that using convex regions leads to a far more efficient representation. Therefore, we use linear functions to partition cells, rather than single features.
- Within each cell, a linear model is used instead of a constant, as this further reduces the number of cells.
- For prediction, it is important to estimate incidents with no UFLS as exactly zero and not a small number. The suggested tree structure can achieve that.

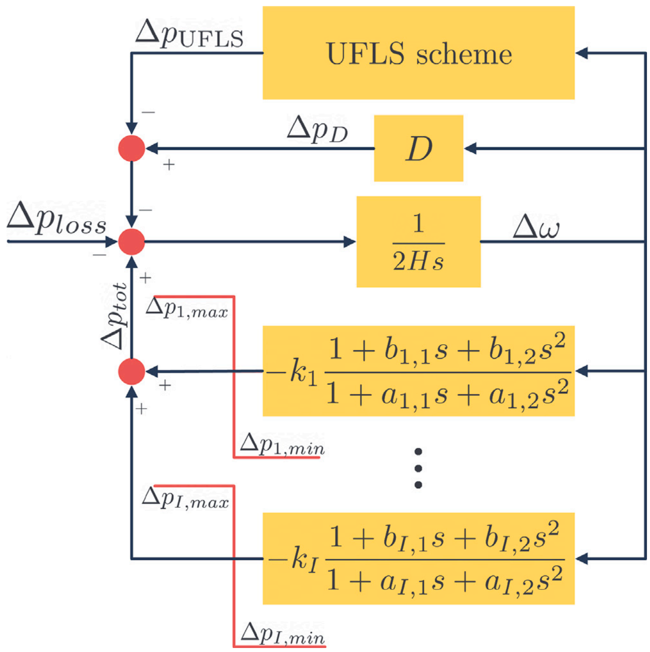

Fig. 1. SFR model.

Representing such regression tree as MILP is presented later. MILP-representability of many ML methods is exploited in Maragno et al. (2021).

The suggested regression tree is shown in Fig. 2. In this figure, N_0 is the root node. N_1 and N_2 are the nodes of the first layer. A linear function of the features (for example $f_0(x)$ for the root node) will split the nodes into two to classify the incidents with a threshold on the labels. Then on the last layer, there are the leaves L_1 to $L_{\mathcal{L}}$. Linear regression is applied to the samples within each of these leaves.

As the labels of the dataset are already accessible after the labeling process (Section 2.2) different methods can be applied to split each node. Each splitting function is found by solving a univariate optimization problem, as explained next. A grid search is performed to find the optimal cut-point for each split (c in Fig. 2). Splitting nodes is continued until the MAE of the new structure is higher than before splitting. From the results that are obtained from the grid search, the one with the best overall MAE will be picked.

Finding the optimal linear function $f(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$, where p is the number of features, to split each cell requires solving a $p+1$ dimensional optimization problem, which finds the coefficients β that minimizes the error of the local linear fits in each cell. This is a difficult optimization problem, and also a reason why typical regression trees split only on single features, i.e. only on functions of the form $f(x) = x_i - c$ for a single feature x_i. To reduce the problem of finding β to a univariate optimization problem, a model inspired by Verwer et al. (2017) is proposed, but for splitting the nodes logistic regression is used instead of assigning a threshold to a feature and instead of assigning a constant at the leaf nodes a linear regression is applied. Considering that the UFLS from the SFR model is either zero or a positive number, it is important to estimate the incident with no UFLS as zero and not a small number. This also can be achieved with this tree structure.

First, a univariate threshold variable c (over which optimization will be performed) is introduced, and a binary variable z, with $z = 0$ if $y < c$ (here, y is the amount of UFLS, i.e. the value being predicted) and $z = 1$ otherwise. Define the logistic function as follows:

$$p_\beta(x) := \frac{1}{1+\exp(-\beta_0 - \sum_{i=1}^{p} \beta_i x_i)} \tag{2}$$

The key idea now is that a good split for predicting y should be able to predict z from x. Therefore, $\beta_0, \ldots, \beta_p$ are found to maximize the log-likelihood,

$$\sum_{j\in N} z^{(j)} \log p_\beta(x^{(j)}) + (1-z^{(j)}) \log(1-p_\beta(x^{(j)})) \tag{3}$$

where N is the node (as a subset of sample indices) that is currently being split. For any given threshold c, denote the maximum likelihood estimate of β by $\hat{\beta}(c)$. Consequently, for each c, node N can be split into two sub-nodes $N'(c)$ and $N''(c)$. Now, by performing a one-dimensional grid search, c that minimizes the overall error of the local linear models, will be chosen. The splitting can be continued until the tree can predict the amount of UFLS with an acceptable accuracy. Note that a simpler tree structure is preferred for two reasons. Firstly, although the accuracy of the model for the dataset in hand might improve by adding more layers, the model will be more susceptible to overfitting. Secondly, the MILP representation becomes computationally burdensome for more complicated tree structures.

To predict values of the UFLS on each leaf, standard linear regression is used. The maximum likelihood estimate of the parameters of the linear model of leaf L_ℓ can be calculated by finding $\alpha_0, \ldots, \alpha_p$ that minimize

$$\sum_{j\in L_\ell} \left(\alpha_0 + \sum_{i=1}^{p} \alpha_i x_i^{(j)} - y^{(j)}\right)^2 \tag{4}$$

Note that, for this fit, only the samples that are assigned to the leaf L_ℓ of the UFLS data are included, resulting in the summation over indices $j \in L_\ell$.

Putting it all together, $\hat{y}$ from will be predicted from x, using the following piecewise linear function,

$$\hat{y}(x) = \begin{cases} \alpha_0^1 + \sum_{i=1}^{p} \alpha_i^1 x_i & \text{if } x \in L_1 \\ \vdots & \\ \alpha_0^\ell + \sum_{i=1}^{p} \alpha_i^\ell x_i & \text{if } x \in L_\ell \\ \vdots & \\ \alpha_0^{\mathcal{L}} + \sum_{i=1}^{p} \alpha_i^{\mathcal{L}} x_i & \text{if } x \in L_{\mathcal{L}} \end{cases} \tag{5}$$

Encoding the proposed regression tree as MILP: To encode the regression tree (i.e. Eq. (5)) as an MILP model, first a binary variable u_ℓ for each leaf L_ℓ needs to be defined, which is equal to 1 if x belongs to leaf L_ℓ. Since x can only belong to one leaf (see (5)), the sum of the binary variables u is equal to 1:

$$\sum_{\ell\in\mathcal{L}} u_\ell = 1 \tag{6}$$

Further, the binary variables u should be equal to 0 if any of the parent nodes fails. Finally, the decisions at the non-leaf nodes (N_0 to $N_{\mathcal{N}}$) directly influence the values of the binary variables of the downstream leaves (Verwer et al., 2017). For instance, if the decision at N_0 is $f_0(x) = \hat{\beta}_0^0 + \sum_{i=1}^{p} \hat{\beta}_i^0 x_i < 0$, then $u_\ell = 0$ for leaves in the upper subtree, and $u_\ell = 1$ for leaves in the lower subtree. The following two constraints force u_ℓ to take these values as a function of x:

$$\hat{\beta}_0^0 + \sum_{i=1}^{p} \hat{\beta}_i^0 x_i + \underline{\mathcal{M}}_0 \sum_{\ell\in\mathcal{L}'} u_\ell \geq \underline{\mathcal{M}}_0 \tag{7a}$$

$$\hat{\beta}_0^0 + \sum_{i=1}^{p} \hat{\beta}_i^0 x_i + \overline{\mathcal{M}}_0 \sum_{\ell\in\mathcal{L}''} u_\ell < \overline{\mathcal{M}}_0 \tag{7b}$$

where $\hat{\beta}^{(0)}$ are the obtained logistic regression coefficients for node N_0. $\underline{\mathcal{M}}_0$ and $\overline{\mathcal{M}}_0$ are lower and upper bounds for the values that $\hat{\beta}_0^0 + \sum_{i=1}^{p} \hat{\beta}_i^0 x_i$ can take for any x in N_0. $\mathcal{L}'$ and $\mathcal{L}''$ are the list of leaves in the upper and lower subtrees of the node N_0. A similar set of constraints must be defined for each node. Now that a binary variable pointing to the correct leaf is accessible, $\hat{y}$ in Eq. (5) can be calculated as:

$$\hat{y} = \sum_{\ell\in\mathcal{L}} u_\ell \times \left(\alpha_0^\ell + \sum_{i=1}^{p} \alpha_i^\ell x_i\right) \tag{8}$$

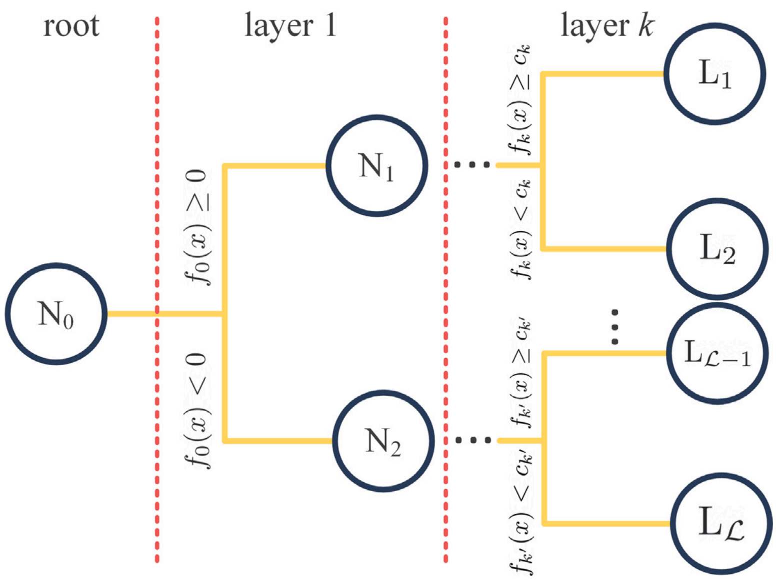

Fig. 2. Proposed regression tree.

Eq. (8) is non-linear due to the product of binary and continuous variables. To linearize (8), the following sets of constraints need to be defined for each leaf ℓ,

$$\alpha_0^\ell + \sum_{i=1}^{p} \alpha_i^\ell x_i - \overline{\mathcal{M}}_\ell(1-u_\ell) \leq r_\ell \tag{9a}$$

$$\alpha_0^\ell + \sum_{i=1}^{p} \alpha_i^\ell x_i - \underline{\mathcal{M}}_\ell(1-u_\ell) \geq r_\ell \tag{9b}$$

$$\overline{\mathcal{M}}_\ell u_\ell \geq r_\ell \tag{9c}$$

$$\underline{\mathcal{M}}_\ell u_\ell \leq r_\ell \tag{9d}$$

where $\overline{\mathcal{M}}_\ell$ and $\underline{\mathcal{M}}_\ell$ are upper and lower bounds of the term $\alpha_0^\ell + \sum_{i=1}^{p} \alpha_i^\ell x_i$ for all $x \in L_\ell$, and r_ℓ is an auxiliary variable. Now the linear equation for $\hat{y}$ can be simply written as,

$$\hat{y} = \sum_{\ell \in \mathcal{L}} r_\ell \tag{10}$$

Considering the proposed MILP encoding method, after training the regression tree with $\mathcal{N}$ nodes and $\mathcal{L}$ leaves with the proposed method, to estimate a new observation in a MILP model, $2 \times \mathcal{N}$ constraints are needed to present the tree structure, one constraint to one-hot encode the binary variables u, $4 \times \mathcal{L} + 1$ constraints to calculate the linearized $\hat{y}$. Also for each observation, $\mathcal{L}$ new continuous variables and $\mathcal{L}$ binary variables are defined.

2.3.2. Tobit model

Instead of making use of a binary tree and linear regression, the standard Tobit model can be considered (Tobin, 1958). The model considers $y^* = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_i + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$, but instead of y^* the following is observed,

$$\hat{y} = \begin{cases} y^* & \text{if } y^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The MILP representation is straightforward. Same as before, one binary variable is defined for each side.

$$\alpha_0 + \sum_{i=1}^{p} \alpha_i x_i + \underline{\mathcal{M}} u_R \geq \underline{\mathcal{M}} \tag{12a}$$

$$\alpha_0 + \sum_{i=1}^{p} \alpha_i x_i + \overline{\mathcal{M}} u_L < \overline{\mathcal{M}} \tag{12b}$$

$$u_R + u_L = 1 \tag{12c}$$

$$\hat{y} = (\alpha_0 + \sum_{i=1}^{p} \alpha_i x_i) \times u_R + 0 \times u_L \tag{12d}$$

As seen in Eq. (12d) a variable in binary multiplication appears, which can be linearized as stated before. After training the dataset with the Tobit model, to calculate $\hat{y}$ for each observation in the MILP model, 12 constraints (Eq. (12a), Eq. (12b), Eq. (12c), and 9 constraints to linearize Eq. (12d) as explained before), 2 binary variables (u_R and u_L), and 2 continuous variables (auxiliary variables for right and left leaf) are added.

3. Results

The process of data generation, labeling, data analysis, learning model, and the obtained accuracy for the process are presented in this section. The case study is La Palma Island. The data regarding the island is presented in Rajabdorri et al. (2022).

3.1. Data generation and analysis

The algorithm presented in algorithm 1 is utilized to construct a training dataset for La Palma Island. The power levels are defined using increments of 0.5 MW to form a vector, and all possible combinations of the generators are listed. However, any combinations exceeding the annual thermal generation peak or falling below the annual thermal generation minimum are excluded. The historical thermal generation data for La Palma island indicates that the thermal generation ranges between 36 MW and 16 MW throughout the year. Therefore, the training dataset should only include generation combinations between the maximum and minimum thermal generation limits. Generation combinations that violate the technical requirements are not feasible and are excluded.

The remaining operation points are then sorted by the total value of their quadratic generation cost functions, and the cheaper ones are retained for every thermal generation level. The reason is that these operation points are more likely to appear in the optimized solution. As previously explained, all data points must be labeled with the SFR model. The SFR model is used to calculate the amount of UFLS for each respective outage, with over 110,000 possible outages for the training dataset. The aim is to label all outages with their expected amount of UFLS. Please note that the synthetic data generation method introduced here utilizes exhaustive enumeration, which can become computationally intensive if the number of generators is large. However, it is

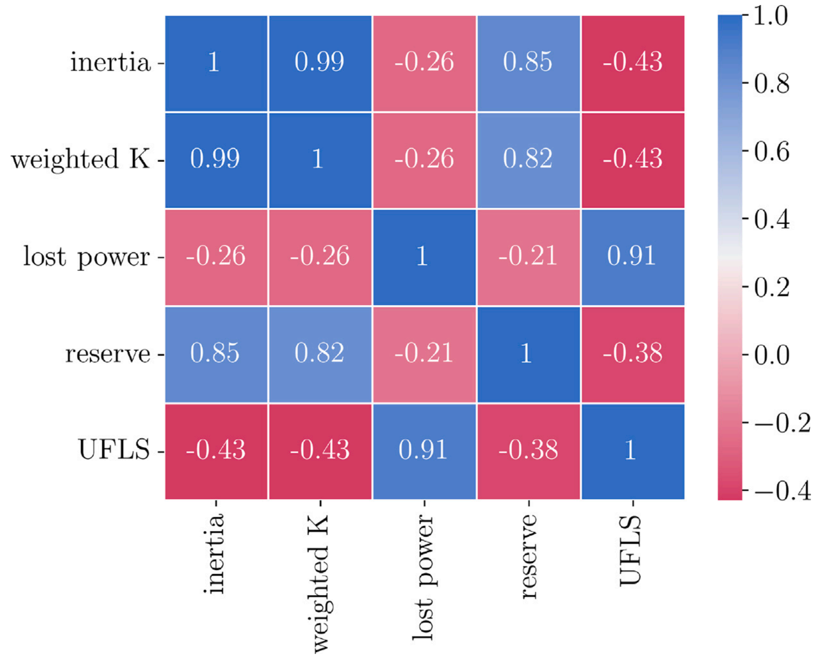

Fig. 3. Pearson correlation between inertia, weighted K, lost power, power reserve, and the amount of UFLS.

Table 2
The summary of the dataset.

	Inertia (MW s)	Weighted K (MW)	Lost power (MW)	Reserve (MW)	UFLS (MW)
count	133,717	133,717	133,717	133,717	133,717
mean	93.02	936.68	4.76	10.02	1.98
Std Dev	17.16	158.53	1.97	2.89	2.35
min	39.26	450.00	2.50	0.50	0.00
25%	81.21	830.00	3.00	8.50	0.00
50%	97.13	960.00	4.00	10.50	1.26
75%	102.78	1031.00	7.00	12.00	4.67
max	133.18	1327.00	10.00	19.00	7.04

important to emphasize that this method is specifically tailored for small power systems with a limited number of units. In larger systems, UFLS following single unit outages is generally not a significant issue and can typically be avoided. In Table 2 a summary of the generated dataset is presented. The table includes the count of the samples, mean value, standard deviation, min value, 25th percentile (the value below which 25% of the data falls), 50th percentile (the value below which 50% of the data falls i.e. median), 75th percentile (the value below which 75% of the data falls), and the maximum value of each feature and label in the dataset. The interested reader is referred to Rajabdorri (2024) to access the full dataset.

First, let's look at the correlations between the features and the labels. In Fig. 3 the Pearson correlation between available inertia, weighted K, lost power, power reserve, and the amount of UFLS is shown on a heatmap.

In Fig. 4 the histogram plot of inertia, weighted K, lost power, power reserve, and the amount of UFLS is depicted in diagonally. The KDE for each two combinations is depicted in respective squares in the bottom left (the carves). In the upper right part of the figure, the scatter plot of the same quantities is depicted (the dots). The outages that do not lead to any UFLS are shown in black, and outages with positive UFLS are shown in red. Fig. 4 gives a good insight into the distribution of the data and how smooth the data is. This figure clearly shows the complexity of the problem at hand. As the final purpose is to use the estimation of UFLS in the operational planning process, it is important to estimate the black incidents in Fig. 4 as exactly zero, and not a small number. Although the general relation between the features shown in Fig. 4 and the amount of UFLS is complex and non-linear, some trades can be spotted. It seems that the incidents with no UFLS (in black), and the incidents with some UFLS (in red) cannot be easily distinguished with only one feature. The combination of all features will distinguish black and red dots with better accuracy. That is another reason to use methods like logistic regression for splitting the nodes, rather than decision trees that rely on one feature to apply the splits. In Fig. 5 a histogram of UFLS is presented. Both of the methods that were introduced in the methodology (Tobit model and proposed regression tree) are applied to the dataset in order to estimate UFLS.

3.2. Learning process

To train and evaluate the models, the dataset from Section 3.1 is divided randomly into a training dataset (80% of the data) and a test dataset (20% of the data). The learning process is done using the training dataset, and the evaluation is done using the test dataset.

3.2.1. With regression tree

A grid search is performed to find the optimum tree structure. Different UFLS thresholds for splitting are tried in a loop, starting from zero and with 0.1 MW steps, and the one that leads to the overall minimum MAE is chosen. Looking at the distribution of the amount of UFLS in Fig. 4 three groups of data can be distinguished: incidents with zero UFLS, incidents with small UFLS (between 0 to 4 MW), and incidents with big UFLS (between 4 to 8 MW). Considering this observation and after performing a grid search to find c, the tree structure shown in Fig. 6 achieves small MAE while being simple.

On the node N_0 the data is classified into positive UFLS and zero UFLS. The split containing zeros does not need any further classification, as all of them are equal to zero. On the node N_1 the remaining points are classified into UFLS bigger than 3.1 MW and smaller than 3.1 MW. The estimated amount of UFLS is presented by its corresponding linear regression on each leaf. The obtained data of this tree structure is presented in Table 3. The scores that are shown in this table are the result of applying the model trained by the training dataset, on the test dataset.

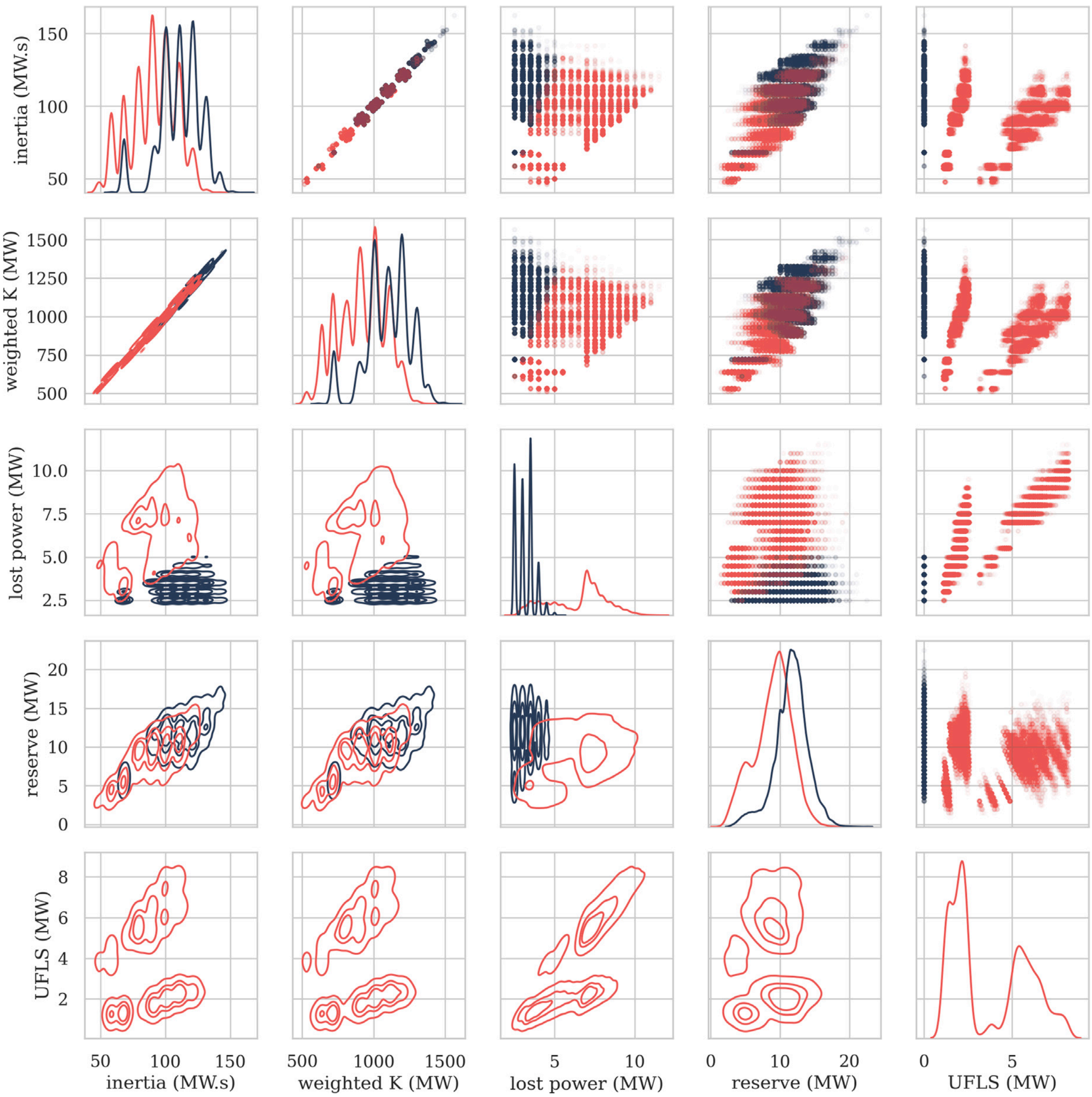

Fig. 4. Histogram plot (diagonally), KDE (curves in the bottom left) and scatter plot (dots in the top right) of the features and the labels.

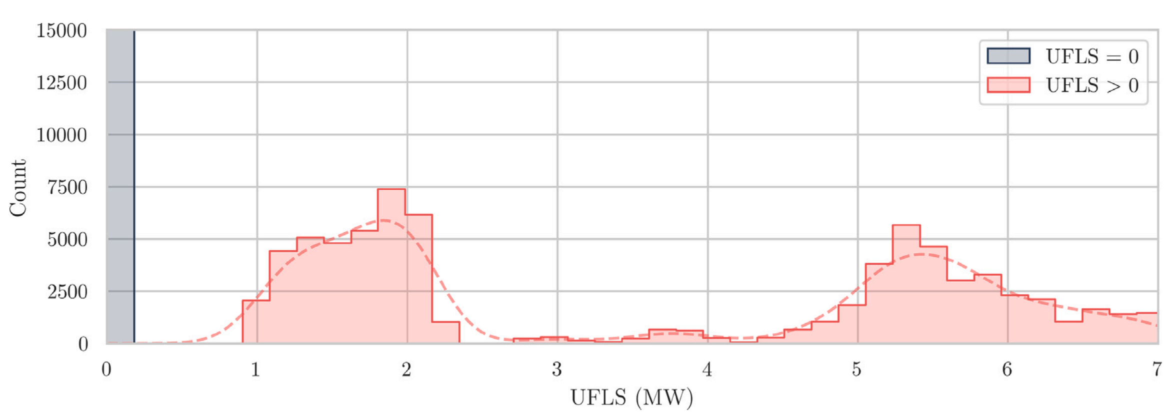

Fig. 5. Histogram of UFLS amount.

The final MAE of this process will be partly due to the classification errors on the nodes and the regression error on the leaves. In Fig. 7 the classification error on different leaves is presented.

The diagonal squares show the percentage of true positive classifications for each class. For example, 96.43% of the whole data has been classified correctly (sum of the diagonal squares). 0.54% of the

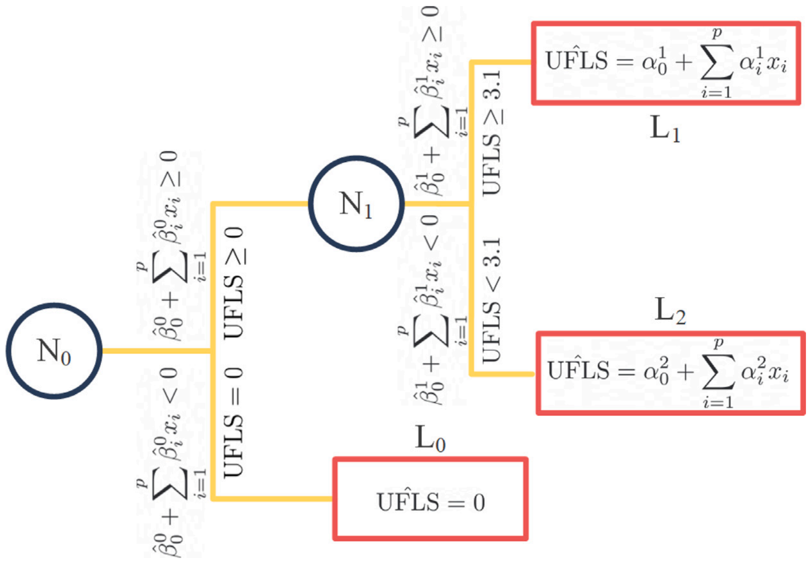

Fig. 6. Proposed regression tree with minimum overall MAE.

Table 3
The data of the trained tree.

	Intercept	$\times\mathcal{H}_g$	$\times\mathcal{K}_g$	$\times P_g$	$\times\mathcal{R}_g$	Score
N_0	$\beta_0^0 = 1$	$\beta_1^0 = 0.337$	$\beta_2^0 = -0.099$	$\beta_3^0 = 18.669$	$\beta_4^0 = -0.889$	acc = 98.9%
N_1	$\beta_0^1 = 1$	$\beta_1^1 = -0.029$	$\beta_2^1 = 0.025$	$\beta_3^1 = -3.498$	$\beta_4^1 = 0.324$	acc = 95.5%
L_0	$\alpha_0^0 = 0$	$\alpha_1^0 = 0$	$\alpha_2^0 = 0$	$\alpha_3^0 = 0$	$\alpha_4^0 = 0$	MAE = 0
L_1	$\alpha_0^1 = 0.269$	$\alpha_1^1 = 0.022$	$\alpha_2^1 = -0.0007$	$\alpha_3^1 = 0.132$	$\alpha_4^1 = -0.055$	MAE = 0.048 MW
L_2	$\alpha_0^2 = 0.194$	$\alpha_1^2 = 0.013$	$\alpha_2^2 = -0.001$	$\alpha_3^2 = 0.878$	$\alpha_4^2 = -0.168$	MAE = 0.256 MW

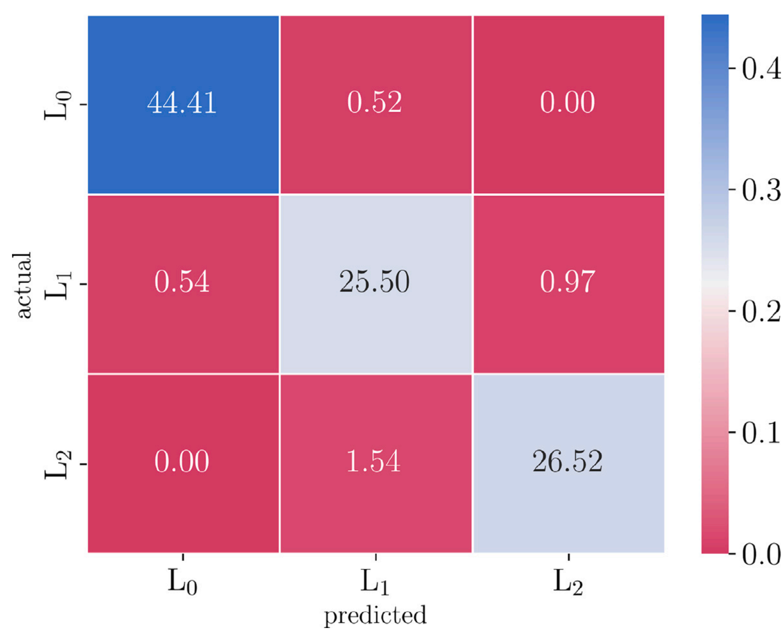

Fig. 7. The classification confusion matrix for each class.

samples on L_1 are incorrectly classified as L_0. 0.52% of L_0 samples are incorrectly classified as L_1, and so on.

The UFLS for all the samples that are assigned to L_0 is estimated as zero and for L_1 and L_2 linear regression is applied. The residuals (predicted value minus observed labels) of the estimation on the test dataset are shown in Fig. 8.

Considering the complexity of the problem at hand, and being limited to using linear models, the accuracy is acceptable. The prediction error for the samples in L_1 is rarely more than 0.2 MW, and for L_2 is rarely more than 1 MW. Note that the samples on L_2 are bigger than the samples on L_1.

Now, considering all the classifications and regression applied in the suggested tree structure, it is possible to look at the residuals for the whole process, shown in Fig. 9.

Other than regression errors, errors due to misclassification are evident. The bigger residuals are because of the misclassification. That is why more complicated tree structures would not improve the overall accuracy in this case. Note that although the estimation error might be

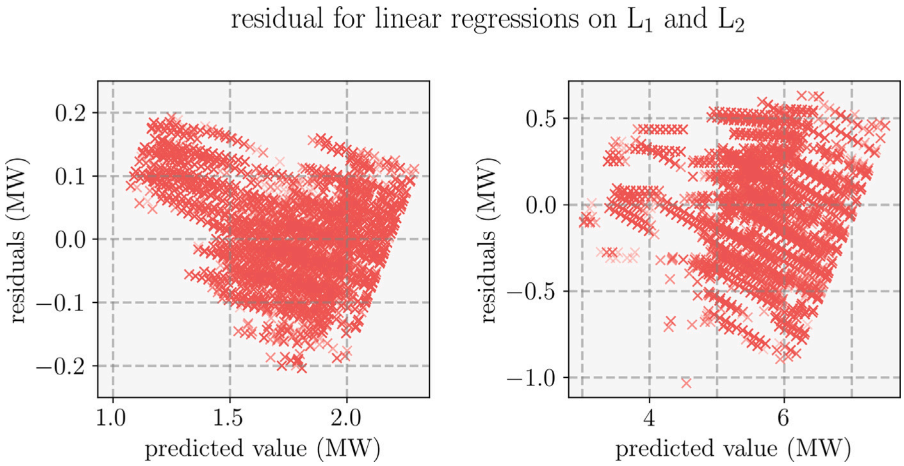

Fig. 8. The residual for the regression applied on L_1 and L_2.

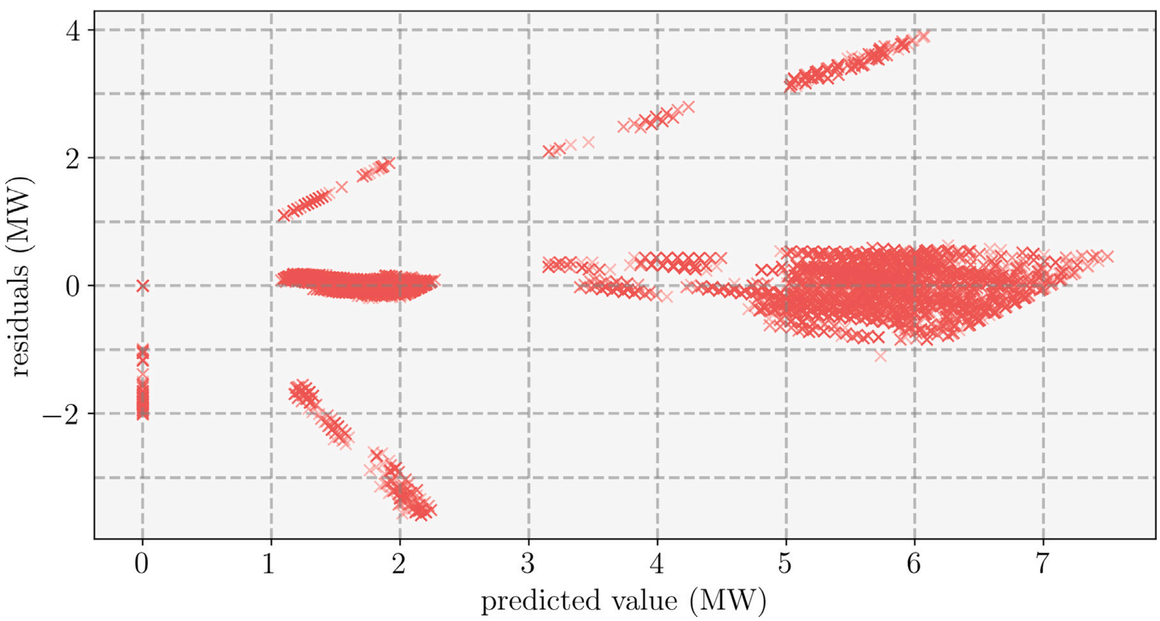

Fig. 9. The residual for the suggested regression tree.

high for some incidents, it does not endanger the stability of the system as the UFLS scheme will ensure the stability of the system. It is expected that the benefits from correct estimations will outweigh the downsides of errors. The MAE of the whole process on the test dataset is 0.179 MW. The trained tree can be represented as MILP with the addition of 18 new constraints, 3 continuous variables, and 3 binary variables for every observation.

3.2.2. With Tobit model

The training dataset is trained with the Tobit model (Section 2.3.2). Then it is applied to the test dataset. The residuals are shown in Fig. 10. The model has successfully distinguished most of the zero UFLS incidents and pushed them to the negative side so they will be equal to zero in the model. As the model tries to fit all the positive UFLS incidents with one line, the error is high for some incidents. The overall MAE of the model on the test dataset is 0.405 MW. The advantage of this model is being easy to implement as MILP. Here is the trained Tobit model,

$$\hat{\text{UFLS}} = \begin{cases} 1.003 - 0.077\mathcal{H}_g - 0.011\mathcal{K}_g \\ +1.164P_g - 0.172\mathcal{R}_g & \text{if it's} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

According to Eq. (12) the term in Eq. (13) can be represented as MILP with the introduction of 2 new binaries, 2 new continuous variables, and 12 constraints for each observation.

4. Conclusion

In this paper, a ML-based approach for estimating UFLS in power systems is presented. By leveraging a carefully generated dataset and applying two suggested ML algorithms, the proposed regression tree, and the Tobit model, the relationship between relevant features and UFLS labels is learned. The trained model demonstrated accurate and effective UFLS estimation, providing valuable insights for operational planning, that will lead to frequency response improvement, reserve allocation optimization, and cost reduction. Applying the methodology to the La Palma island power system showcased its practicality and reliability, highlighting the potential for integrating UFLS estimation into the scheduling optimization problem. While the MILP representation of the Tobit model is computationally simpler, the accuracy of the suggested binary tree structure is superior. Future research avenues may focus on integrating this methodology into the actual operational planning problems like UC and ED.

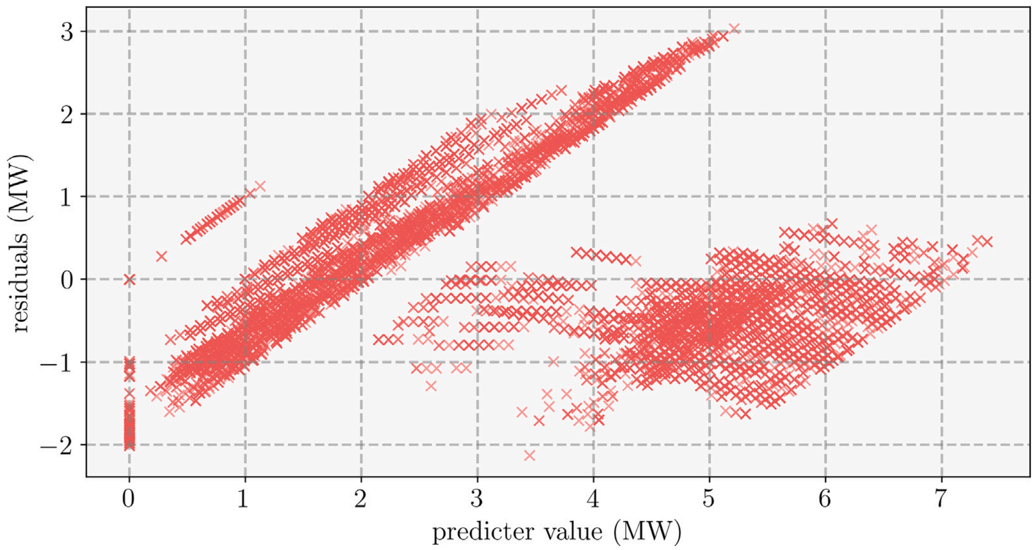

Fig. 10. Residuals of the test dataset. The training dataset is trained with Tobit model.

Exploring additional features, investigating alternative ML algorithms, and considering the impact of varying system configurations can further enhance the accuracy and applicability of UFLS estimation. To follow up on the findings of this paper, the proposed models should be implemented in operational planning problems, like UC, to further prove its benefits. Also, the proposed regression tree can be used for various applications, as an alternative to regular regression trees.

CRediT authorship contribution statement

Mohammad Rajabdorri: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Matthias C.M. Troffaes:** Writing – review & editing, Validation, Methodology. **Behzad Kazemtabrizi:** Writing – review & editing. **Miad Sarvarizadeh:** Validation, Investigation. **Lukas Sigrist:** Writing – review & editing, Supervision, Funding acquisition. **Enrique Lobato:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been funded by grant PID2022-141765OB-I00 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”.

Data availability

Data is linked in the manuscript.

References

Badesa, L., Teng, F., Strbac, G., 2019. Simultaneous scheduling of multiple frequency services in stochastic unit commitment. IEEE Trans. Power Syst. 34 (5), 3858–3868.

Chang, G.W., Chuang, C.-S., Lu, T.-K., Wu, C.-C., 2012. Frequency-regulating reserve constrained unit commitment for an isolated power system. IEEE Trans. Power Syst. 28 (2), 578–586.

Ferrandon-Cervantes, C., Kazemtabrizi, B., Troffaes, M.C., 2022. Inclusion of frequency stability constraints in unit commitment using separable programming. Electr. Power Syst. Res. 203, 107669.

Golpîra, H., Bevrani, H., Messina, A.R., Francois, B., 2022. A data-driven under frequency load shedding scheme in power systems. IEEE Trans. Power Syst. 38 (2), 1138–1150.

Hooshmand, R., Moazzami, M., 2012. Optimal design of adaptive under frequency load shedding using artificial neural networks in isolated power system. Int. J. Electr. Power Energy Syst. 42 (1), 220–228.

Jung, A., 2018. Machine Learning: The Basics. [Online]. Available arXiv:1805.05052.

Kalajahi, S.M.S., Seyedi, H., Zare, K., 2021. Under-frequency load shedding in isolated multi-microgrids. Sustain. Energy Grids Netw. 27, 100494.

Ketabi, A., Fini, M.H., 2014. An underfrequency load shedding scheme for islanded microgrids. Int. J. Electr. Power Energy Syst. 62, 599–607.

Laghari, J., Mokhlis, H., Karimi, M., Bakar, A.H.A., Mohamad, H., 2014. A new under-frequency load shedding technique based on combination of fixed and random priority of loads for smart grid applications. IEEE Trans. Power Syst. 30 (5), 2507–2515.

Lagos, D.T., Hatziargyriou, N.D., 2021. Data-driven frequency dynamic unit commitment for island systems with high RES penetration. IEEE Trans. Power Syst. 36 (5), 4699–4711.

Li, C., Wu, Y., Sun, Y., Zhang, H., Liu, Y., Liu, Y., Terzija, V., 2019. Continuous under-frequency load shedding scheme for power system adaptive frequency control. IEEE Trans. Power Syst. 35 (2), 950–961.

Maragno, D., Wiberg, H., Bertsimas, D., Birbil, S.I., Hertog, D.d., Fajemisin, A., 2021. Mixed-integer optimization with constraint learning. arXiv preprint arXiv:2111.04469.

Mehrabi, K., Afsharnia, S., Golshannavaz, S., 2018. Toward a wide-area load shedding scheme: Adaptive determination of frequency threshold and shed load values. Int. Trans. Electr. Energy Syst. 28 (1), e2470.

O’Malley, C., Badesa, L., Teng, F., Strbac, G., 2022. Probabilistic scheduling of UFLS to secure credible contingencies in low inertia systems. IEEE Trans. Power Syst. 37 (4), 2693–2703.

Paturet, M., Markovic, U., Delikaraoglou, S., Vrettos, E., Aristidou, P., Hug, G., 2020. Stochastic unit commitment in low-inertia grids. IEEE Trans. Power Syst. 35 (5), 3448–3458.

Pérez-Illanes, F., Álvarez-Miranda, E., Rahmann, C., Campos-Valdés, C., 2016. Robust unit commitment including frequency stability constraints. Energies 9 (11), 957.

Rajabdorri, M., 2024. UFLS Estimation Dataset. Zenodo.

Rajabdorri, M., Kazemtabrizi, B., Troffaes, M., Sigrist, L., Lobato, E., 2023. Inclusion of frequency nadir constraint in the unit commitment problem of small power systems using machine learning. Sustain. Energy Grids Netw. 101161.

Rajabdorri, M., Lobato, E., Sigrist, L., 2022. Robust frequency constrained uc using data driven logistic regression for island power systems. IET Gener. Transm. Distr. 16 (24), 5069–5083.

Rajabdorri, M., Rouco, A., Sigrist, L., Lobato, E., 2024. Unit commitment with analytical under-frequency load-shedding constraints for island power systems. IEEE Access 12, 72337–72344.

Sedighizadeh, M., Esmaili, M., Mousavi-Taghiabadi, S.M., 2019. Optimal joint energy and reserve scheduling considering frequency dynamics, compressed air energy storage, and wind turbines in an electrical power system. J. Energy Storage 23, 220–233.

Shahidehpour, M., Ding, T., Ming, Q., Catalao, J.P., Zeng, Z., 2021. Two-stage chance-constrained stochastic unit commitment for optimal provision of virtual inertia in wind-storage systems. IEEE Trans. Power Syst..

Sigrist, L., Lobato, E., Echavarren, F.M., Egido, I., Rouco, L., 2016. Island Power Systems. pp. 1–272.

Silva, Jr., S.S., Assis, T.M., 2020. Adaptive underfrequency load shedding in systems with renewable energy sources and storage capability. Electr. Power Syst. Res. 189, 106747.

Teng, F., Strbac, G., 2017. Full stochastic scheduling for low-carbon electricity systems. IEEE Trans. Autom. Sci. Eng. 14 (2), 461–470.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 24–36.

Tofis, Y., Timotheou, S., Kyriakides, E., 2016. Minimal load shedding using the swing equation. IEEE Trans. Power Syst. 32 (3), 2466–2467.

Trovato, V., Bialecki, A., Dallagi, A., 2018. Unit commitment with inertia-dependent and multispeed allocation of frequency response services. IEEE Trans. Power Syst. 34 (2), 1537–1548.

Verwer, S., Zhang, Y., Ye, Q.C., 2017. Auction optimization using regression trees and linear models as integer programs. Artificial Intelligence 244, 368–395.

Wang, C., Chu, S., Ying, Y., Wang, A., Chen, R., Xu, H., Zhu, B., 2022. Underfrequency load shedding scheme for islanded microgrids considering objective and subjective weight of loads. IEEE Trans. Smart Grid.

Zhang, Y., Cui, H., Liu, J., Qiu, F., Hong, T., Yao, R., Li, F., 2021. Encoding frequency constraints in preventive unit commitment using deep learning with region-of-interest active sampling. IEEE Trans. Power Syst. 37 (3), 1942–1955.