Energy 238 (2022) 121993

Contents lists available at ScienceDirect

Energy

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

Machine learning-assisted outage planning for maintenance activities in power systems with renewables

Jean-François Toubeau [a,*], Lorie Pardoen [a], Louis Hubert [a], Nicolas Marenne [b], Jonathan Sprooten [b], Zacharie De Grève [a], François Vallée [a]

[a] Power Systems and Markets Research Group, University of Mons, Belgium
[b] Power System Planning, Grid Development, Elia Transmission Belgium, Belgium

ARTICLE INFO

Article history:
Received 18 February 2021
Received in revised form
1 September 2021
Accepted 2 September 2021
Available online 7 September 2021

Keywords:
Explainable AI
Machine learning
Maintenance
Reliability
Tree-based models

ABSTRACT

The optimal coordination of maintenances is becoming increasingly important to guarantee the security of supply in renewable-dominated power systems. However, current planning tools are plagued with tractability issues arising from the need to comply with operational security standards. The grid must indeed safely accommodate any unexpected contingency occurring during the scheduled maintenances, which requires simulating many different scenarios. To alleviate this computational burden, this paper proposes to leverage machine learning models to predict the outcome of contingency analyses in a fast and reliable manner. The methodology is tested on the full regional transmission grid of Belgium, covering the voltage levels from 150 kV down to 30 kV. Different models, including naive Bayes classifiers, support vector machines and tree-based models, are tested and compared. Outcomes reveal that random forests consistently outperform other benchmarks, by identifying with an accuracy higher than 90% the time periods during which maintenances can be safely performed. Also, we show that the expected rise in renewable generation will impact the maintainability of the future system, with an increase of up to 20% of unsuitable periods to perform maintenances in some grid areas.

1. Introduction

Following the liberalization of the electricity sector, generation and retail activities are now open to competition, and fully decoupled from the transmission and distribution of electricity. The Transmission System Operator (TSO) is thus the legal entity responsible for designing, building and maintaining the high-voltage electrical grid, and must therefore strive to guarantee the security and quality of electricity supply for all connected users (i.e. industrial and distribution grids), while facilitating the energy transition at the lowest costs possible. This task of efficiently managing the infrastructure is increasingly challenging due the emergence of decentralized and more complex to forecast generation, the ageing of assets as well as the increasing demand arising from the electrification of transportation and heating systems [1,2]. This leads to the need for more interventions on electricity systems, while fewer suitable periods are available to securely schedule these operations [3–5].

Planning maintenance outages in modern power systems is thus of critical importance, and requires to identify the best trade-off between reliability and costs [6,7].

1.1. Literature review

Historically, in a context of load-driven dispatch of centralized generation and limited market coupling between European countries, finding the optimal outage schedule on the different grid assets was based on the knowledge of grid experts. However, the increasing diversity of operating conditions (arising from, e.g., the complex interdependency between cross-border flows, weather-driven generation and demand-side management [8]), requires innovative outage scheduling tools to achieve reliable and flexible maintenance plans [9,10].

In particular, the deployment of metering devices in modern energy systems has led to the advent of data-driven predictive maintenance where the goal is to identify issues at early stages (by analyzing the operating conditions of grid components), thus taking timely maintenance decisions that prevent unexpected outages

* Corresponding author. National Fund of Scientific Research (FNRS) at the Electrical Power Engineering Unit, University of Mons, Belgium.
E-mail address: Jean-Francois.TOUBEAU@umons.ac.be (J.-F. Toubeau).

https://doi.org/10.1016/j.energy.2021.121993
0360-5442/

Nomenclature

Sets and Indices

$c \in \mathcal{C}$	: Set of grid assets (needing maintenance)
$t \in \mathcal{T}$	: Set of time steps in the training set.
$\tau \in \mathrm{T}$	: Set of time steps in the test set.
$d \in \mathcal{D}$	: Set of all predictor variables
$\mathcal{D}_c \subset \mathcal{D}$	: Set of important predictor variables for asset c

Parameters

$\mathbf{x}^{\text{orig}} \in \mathbb{R}^{\lvert\mathcal{D}\rvert \times \lvert\mathcal{T}\rvert}$	: Predictor variables
$\mathbf{x}_c \subset \mathbf{x}^{\text{orig}}$	: Predictor variables related to asset c
$y_{c,t} = \{0, 1\}$	: Maintenance feasibility of asset c at time t
$\hat{y}_{c,t}$	: Predicted maintenance feasibility of asset c at time t
θ_c	: Parameters of the model predicting the maintenance feasibility of asset c
θ_c^*	: Optimal values of parameters θ_c
f_{θ_c}	: Model predicting the maintenance feasibility of asset c
$\mathcal{L}$	: Loss function of the classification study

[11,12]. In this way, deep learning is used in Ref. [13] to provide information on the health status of components to human experts, thus helping them in deciding whether maintenance actions are needed. Then, a support vector machine (SVM) model is developed in Ref. [14] to predict (unplanned) outages of grid components in presence of extreme events (e.g., imminent hurricane). In Ref. [15], multiple classifiers are combined for improved maintenance recommendations, while the potential of tree-based models is demonstrated in Ref. [16] for predicting the likelihood of failure based on a continuous monitoring of machine conditions.

However, although these machine learning-based decision tools are useful to preventively identify defaults on specific assets, they are not able to anticipate the effect of the outages on the remaining system. This makes them poorly suited for outage scheduling in power grids [17] since the planned outages may result into stressed operating conditions (e.g., overloading of transmission lines), which can be very costly due to the need of redispatching decisions such as load shedding [18].

In this work, the objective is thus to complement current predictive maintenance strategies with a new machine learning tool able to identify the effect of a planned outage in the electricity transmission system. In other words, based on the outcome of predictive maintenance (deciding which components need servicing), the goal is to identify time periods during which maintenance operations can be performed without jeopardizing the security of the power system. A similar research question is tackled in Ref. [19], where a machine learning tool is built for predicting the power system operating conditions during the maintenance of grid components. In Ref. [20], a new sequential formulation for preventive maintenance in multi-energy microgrids is introduced, wherein maintenances and the optimal grid operation are jointly optimized.

Overall, these models fail to comply with European security standards that rely on the "N-1" criterion, whereby the loss of one major component (such as a transmission line or a generator) does not result into cascading events that would ultimately violate the grid operational security limits. In this paper, we develop a new methodology designed to predict the time windows during which planned and unplanned outages can be safely accommodated. This task is essential, but highly challenging [21]. Indeed, in the regional Belgian grid, covering the voltage levels from 150 kV down to 30 kV, about 7000 combinations of planned maintenance outages with unplanned contingencies need to be simulated to have an exact characterization of the system reliability at a given hour of the year. Since the scheduled outages need to be simulated for all hourly periods of the horizon of interest, the problem quickly becomes computationally intractable as the number of maintenance activities increases.

1.2. Contributions

In light of this context, this work, which results from a collaboration between university and the Belgian TSO (i.e., Elia), aims at constructing a generic machine learning proxy able to quickly and reliably identify the time periods during which the maintenance of a specific grid asset can be safely performed. Overall, the contributions of the work are fourfold.

First, we develop a data-driven methodology able to accurately predict (in a fast and reliable manner) the feasibility of maintenance activities given specific grid conditions. Specifically, the learning task is formulated as a binary classification problem, which identifies whether a maintenance outage can be safely scheduled based on (weather, load and generation) input information. Relevant classification algorithms (Bayesian models, support vector machines, and tree-based models) are used to solve the problem, and are compared not only in terms of accuracy but also regarding their practical benefits for industry.

Second, we present an input selection process to select (among hundreds of available features) the most relevant explanatory variables to feed the classification models. In addition to inform the TSO on the important factors affecting the maintainability of the grid [22], this reduced-size input vector enables to reduce the complexity of models, thus limiting overfitting risks.

Third, we adapt the classification threshold to reflect the risk appetite of the TSO. False positives (i.e., predicting that a maintenance is acceptable while it may lead to unsafe operating conditions) would minimize the complexity of outage planning but would require (costly) corrective actions during the network operation, while false negatives (which needlessly reduce the time periods available for maintenances) would increase the need to invest in the grid infrastructure to enable safe outage planning.

Fourth, we extend the procedure by using the trained models on predicted scenarios of future conditions, with the goal of assessing the long-term maintainability of the system. In particular, we develop two novel indicators to quantify the criticality and predictability of maintenances, which provides useful information in the context of grid development.

Overall, the outcomes from the procedure are inherently interpretable, which is an important aspect to foster the acceptability of the proposed framework by operational teams in the industry.

1.3. Organization of the paper

The rest of the paper is organized as follows. Section 2 presents the methodology used to predict the outcome of contingency analyses using machine learning. Then, section 3 conducts case studies on the full regional transmission grid of Belgium and show the superiority of tree-based models. In particular, we discuss how these models can be exploited by modern TSOs at a large scale to assess the maintainability of the grid. Finally, section 4 draws the main conclusions and potential outlooks.

2. Machine learning proxy for contingency analyses

When scheduling a maintenance, the TSO is preventively adapting the topology of the grid to secure it from unplanned contingencies. The goal is to ensure that the resulting electrical grid complies with the "N-1-1″ criterion (or quasi "N-2" criterion) imposing that its operating conditions (arising from both planned and unplanned outages) lie within security limits. This is evaluated by performing contingency analyses that compute the state of the grid for all possible outage events for each hourly time step of the maintenance job.

To avoid to perform these time consuming and cumbersome contingency studies, we develop a supervised learning proxy that uses available data to learn the relationship between $|\mathcal{D}|$ predictors $\mathbf{x}_t^{\text{orig}}$ (i.e., load, generation and weather conditions at time t) and the maintainability $y_{c,t}$ of an asset c. The trained surrogate model can then be used to reliably extrapolate (for any new system state) whether a maintenance can be safely performed.

The global procedure is summarized in Fig. 1, and is thoroughly described in the rest of the paper. In brief, we firstly construct the database (in section 2.1) to align available input features $\mathbf{x}_t^{\text{orig}} \in \mathbb{R}^{|\mathcal{D}|}$ with the maintenance feasibility $y_{c,t} = \{0, 1\}$ of all assets $c \in \mathcal{C}$ for all time steps $t \in \mathcal{T}$. The assets considered in this work are transmission lines and transformers. Since each grid asset c is affected by different parameters, a feature selection process is developed (in section 2.2) to retain the $|\mathcal{D}_c|$ most relevant predictors $\mathbf{x}_c \subset \mathbf{x}^{\text{orig}}$. Based on these features $\mathbf{x}_c$, a dedicated machine learning proxy f_{θ_c} is trained (in section 2.3) for each asset c.

To ensure the practical interest of the methodology for the TSO, the proxy needs to comply with three important constraints, (i) it must be easy to use (the model can be trained without complicated data pre-processing), (ii) the model quality must be robust to changes in the hyper-parameters (no strong expert knowledge and experience is necessary to properly fine-tune the model), and (iii) the outcomes must be readily interpretable by operational teams (the important features can be identified). In this work, different types of machine learning models (e.g., naive Bayes, support vector machine, tree-based models) will therefore be compared in regards to these three criteria.

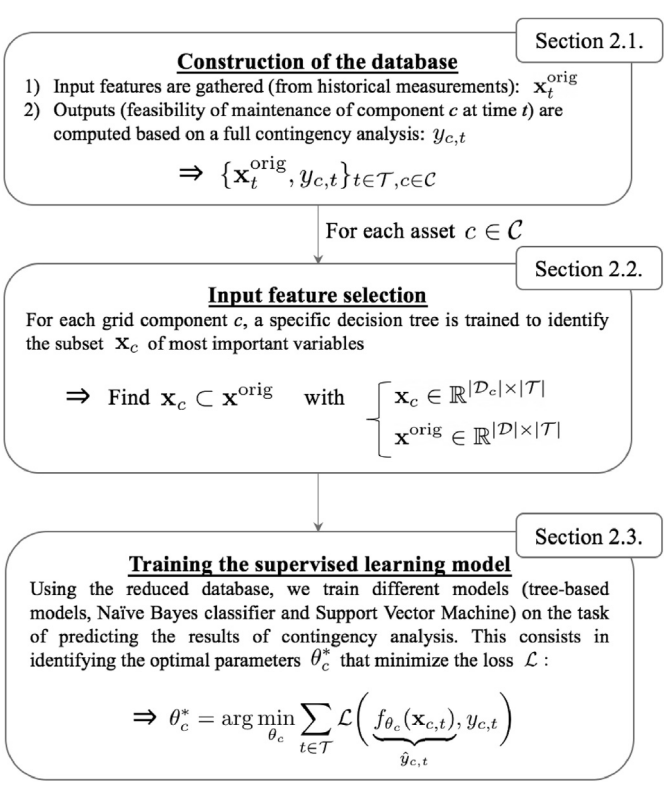

Fig. 1. Overview of the methodology for predicting outcomes of contingency analysis.

2.1. Construction of the database

The procedure to construct the database for training the machine learning models is depicted in Fig. 2.

We rely on a database $\mathbf{x}^{\text{orig}} \in \mathbb{R}^{|\mathcal{D}|\times|\mathcal{T}|}$ composed of variables $d \in \mathcal{D}$ measured at hourly time periods $t \in \mathcal{T}$. The $|\mathcal{D}|$ variables consist of both global and local data.

At the system scale, we have access to the total load (typically ranging from 6 GW during off-peak periods up to 13 GW in peak periods of winter). We also rely on the aggregated generation for different technologies (the total installed capacity is around 23 GW). The generation mix (over the year 2020) was the following: nuclear (39.1%), gas (34.4%), offshore (8.3%) and onshore (5%) wind, photovoltaic (5.3%), biogas (2.5%), coal (0%), imports (−0.3%), and others (5.7%). All these market data are continuous variables whose hourly values are publicly available on the website of Elia (the Belgian TSO) due to transparency obligations [23]. Weather information, including temperature, cloud cover and wind speed at one location in the centre of Belgium, is also used in this work but these data are not publicly available.

The local variables are the generation/consumption measurements at the different nodes of the transmission system (i.e., big industrial clients, power plants and interfaces with local distribution networks). Practically, there are more than 1000 nodes on the Belgian transmission grid, which significantly increases the dimensionality of the problem. These nodal energy exchanges are confidential data pertaining to Elia. It should also be noted that the topology of the Belgian grid, along with all its constitutive assets are also provided by Elia, thus enabling to perform meaningful studies on the Belgian transmission system in the case study (section 3).

In general, historical realizations $\{\mathbf{x}_t^{\text{orig}}\}_{t\in\mathcal{T}}$ (e.g., on the year 2020) are used to train machine learning models (of section 2.2 and section 2.3). However, these input scenarios can also be generated through market simulations, which can emulate a wide range of original grid conditions (not present in the historical database). By simulating expected market conditions (e.g., nuclear phase-out, large integration of renewables, and increase of interconnection capacities), one is able to construct a model that can properly generalize to unseen future conditions (e.g., for years 2025−2030). These aspects are further discussed in section 3.3.2, where we assess the future maintainability of the system, which yields relevant insights for identifying the necessary investments in the long-term planning of the grid.

Evidently, there are no missing data in the simulated scenarios. Also, there are no missing data in the historical information, since there are legal obligations that the market data are properly validated (due to economic implications). Hence, no strategy for consolidating the data had to be implemented in this work.

In complement to this database, the different grid assets that need maintenance are listed. For each asset $c \in \mathcal{C}$, the goal is the compute the feasibility of the maintenance for each system state $t \in \mathcal{T}$ of the database. This is achieved by performing an extensive "quasi N-2" contingency analysis, simulating all relevant unplanned outages concurring with the scheduled maintenance. Practically, the maintenance of asset c is allowed ($y_{c,t} = 1$) if the following 4 different conditions are satisfied:

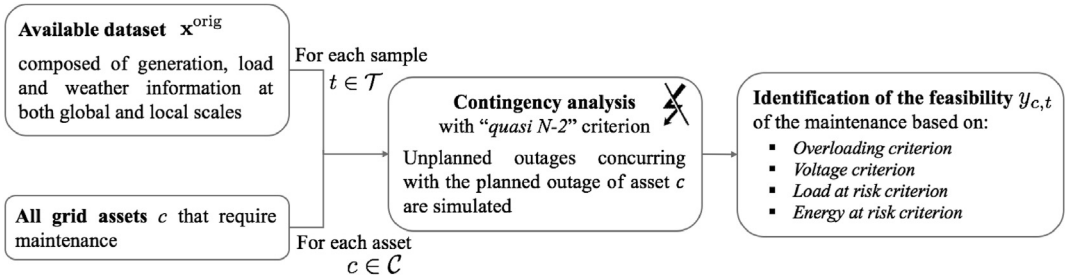

Fig. 2. Creation of the database used for training the machine learning models.

1. The overloading criterion: no asset within the remaining grid must be congested, since this could lead to a cascade effect and further negative impacts (e.g., loss of load) in the system.
2. The voltage criterion: each connection point of grid users should be kept within defined power quality standards (i.e., 90% and 105% of the nominal value).
3. The load at risk criterion: the aggregated power of grid users that could be lost following a contingency should be kept below a given MW threshold.
4. The energy at risk criterion: the total energy that could not be supplied to grid users cannot exceed a given MWh threshold.

If any of these criteria is violated, the maintenance cannot be scheduled ($y_{c,t} = 0$). This multicriteria approach allows to embed both reliability and economic perspectives to define the feasibility of a maintenance [24].

2.2. Selection of the relevant input variables

Since many different features $\mathbf{x}_t^{\text{orig}} \in \mathbb{R}^{|\mathcal{D}|}$ are available and since each asset $c \in \mathcal{C}$ is affected by a different set of variables $\mathcal{D}_c$, a feature selection is performed for each asset individually. The goal is to reduce the dimensionality of input data (that will be processed by the subsequent classifier in section 2.3) with a minimum loss of the initial information. In this way, we reduce model complexity (and thus training time), while avoiding overfitting risks.

Different approaches can be used for this purpose. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are well-known techniques that perform a linear mapping of the data into a new space wherein most of the relevant information can be contained in a reduced number of dimensions. A nonlinear extension is provided by kernel PCA. Another efficient nonlinear framework is given by auto-encoders, i.e., neural networks tailored to learn a reduced-space representation of data. However, the main issue with all those approaches is that the new (reduced) representation of the data has no physical meaning, i.e., it is a combination of features that cannot be interpreted and compared to the original data. In our work, this makes them irrelevant for grid experts that want to evaluate the practical pertinence of the selected features. In light of this context, decision trees have been naturally selected. Indeed, these algorithms gather valuable properties, i.e., they can identify the most important variables (from the original dataset), they do not require any feature scaling (owing to their invariance to monotonic transformations), and they do not rely on strong assumptions (i.e., they are inherently able to capture multivariate non-linear dependencies).

Overall, a decision tree algorithm is used to efficiently select (for asset c) this subset $\mathcal{D}_c$ of the most relevant variables, i.e., $\mathbf{x}_c \subset \mathbf{x}^{\text{orig}}$, with $\mathbf{x}_c \in \mathbb{R}^{|\mathcal{D}_c| \times |\mathcal{T}|}$. As discussed in section 2.3, another decision tree can also be used for the classification phase (to predict the feasibility of a maintenance).

A decision tree is a supervised learning method, which consists in partitioning the whole $|\mathcal{D}|$-dimensional feature space into smaller subspaces. As depicted in Fig. 3 for an illustrative example with 2 variables, a decision tree is trained through a hierarchical multi-stage process. At each stage, a new (chance) node m is created, and the algorithm selects the variable that provides the optimal separation among data (i.e., leading to the best differentiation among classes). The variable is split into two subspaces, respectively noted m_l and m_r, based on a threshold value M. The procedure continues until further partitioning the input space no longer brings value to the classification accuracy. The end nodes show the final classification outcome.

At each stage, the best model parameters (the best feature m to be split and the corresponding threshold value M) are based on an impurity function. In this work, the *Gini* impurity is used [25]:

$$Gini(m) = 2p_m(1 - p_m) \tag{1}$$

where p_m is the probability that a maintenance is classified as possible at node m. It can be seen that the *Gini* indice varies between 0 and 0.5. It is equal to 0 when the node m is pure, i.e., when all samples pertaining to that node belong to a single class. The indice is equal to 0.5 when the impurity of the node is maximum, i.e., when all samples of that node are equally distributed between both classes (feasible and infeasible maintenances).

Once the decision tree (corresponding to an asset c) is trained, it can be used to determine the D_c most important features, based on their position in the tree. In this way, the importance of a variable is calculated as the decrease in node impurity weighted by the probability of reaching that node. Since a single variable can be used at different nodes of the tree, its global importance is the sum of its individual contributions. However, this ranking of important variables does not provide information on the cut-off value under which variables have no real interest for classification. To identify this threshold value in a generic way (applicable for each asset), a random variable is added to the list of predictors $\mathbf{x}^{\text{orig}}$, and the tree is trained on this augmented dataset of size $\mathbb{R}^{(|\mathcal{D}|+1) \times |\mathcal{T}|}$. Since the random variable has no influence on the asset maintainability, all variables with a lower score are removed from the dataset. To ensure stability of the outcome, the score of each variable is

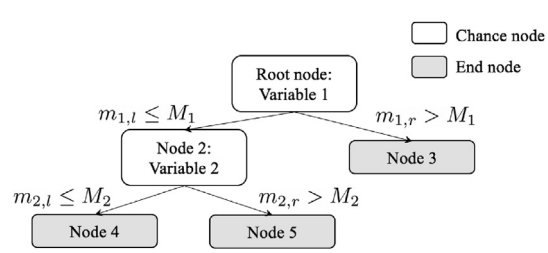

Fig. 3. Illustrative example of decision tree with 2 variables.

averaged over 10 repetitions of the procedure.

2.3. Classification tools

In this part, the goal is to develop a quick and reliable tool to predict the maintainability $y_{c,t}$ of an asset c at time t given the grid conditions specified by $\mathbf{x}_{c,t}$. The problem is formulated as a classification task with a binary outcome, i.e., $y_{c,t} = 1$ when the asset can be safely put out of operation, and $y_{c,t} = 0$ when the maintenance results into stressed conditions for the remaining system.

The objective is to optimize the parameters θ_c of these (asset-specific) surrogates f_{θ_c} such that the output $y_{c,t}$ can be accurately predicted based on the given inputs $\mathbf{x}_{c,t}$ [26]:

$$\theta_c^* = \arg\min_{\theta_c} \sum_{t\in\mathcal{T}} \mathcal{L}\left(\underbrace{f_{\theta_c}(\mathbf{x}_{c,t})}_{\hat{y}_{c,t}}, y_{c,t}\right), \tag{2}$$

where $\mathcal{L}$ is the loss function used to reduce the classification error between the predicted $\hat{y}_{c,t}$ and actual $y_{c,t}$ outputs during the learning phase. At the end of the training, the optimal parameters θ_c^* of the model are obtained, and it can be used (in the online stage) to predict the outcome $\hat{y}_{c,\tau}$ of a contingency analysis under new system conditions $\mathbf{x}_{c,\tau}$ (when grid asset c is in maintenance). This procedure is described in Fig. 4, and is tested for different machine learning models.

2.3.1. Decision tree

As previously explained, the training phase of a decision tree (DT) consists in a greedy search, in which the optimal model parameter $\theta_{c,m}^*$ at a node m is the optimal split of the feature m. In this part, the model is trained with a particular attention to avoid overfitting issues (e.g., by controlling the depth of the tree) to ensure optimal generalization capabilities on unseen data. The decision tree is then compared with useful benchmarks in statistical classification, i.e., a naïve Bayes classifier and a support vector machine (SVM) framework.

2.3.2. Naive Bayes classifier

The Naive Bayes classifier relies on the assumption that the $|\mathcal{D}_c|$ inputs $\mathbf{x}_{c,t} = (x_{1,t}, \ldots, x_{|\mathcal{D}_c|,t})$ are fully independent, and have an equal contribution to predict the outcome $y_{c,t} = \{0, 1\}$. Hence, the Bayes' theorem can be expressed as [27]:

$$P\left(y_{c,t}\middle|x_{1,t}, \ldots, x_{|\mathcal{D}_c|,t}\right) = \frac{P(y_{c,t})\prod_{i=1}^{|\mathcal{D}_c|} P(x_{i,t}|y_{c,t})}{P(x_{1,t}, \ldots, x_{|\mathcal{D}_c|,t})} \tag{3}$$

Since $P(x_{1,t}, \ldots, x_{|\mathcal{D}_c|,t})$ is constant for a given database, the following classification rule can be used:

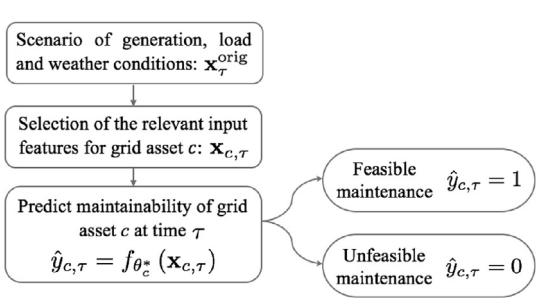

Fig. 4. Online utilization of the trained classification model.

$$\hat{y}_{c,t} = \arg\max_{y_{c,t}} P(y_{c,t}) \prod_{i=1}^{|\mathcal{D}_c|} P(x_{i,t}|y_{c,t}) \tag{4}$$

where $P(y_{c,t})$ and $P(x_{i,t}|y_{c,t})$ are easily computed using the Maximum A Posteriori estimation. In this way, $P(y_{c,t} = 0)$ is the relative frequency of $y_{c,t} = 0\ \forall t$ in the training set, such that $P(y_{c,t} = 0) + P(y_{c,t} = 1) = 1$.

2.3.3. Support Vector Machine

Support Vector Machine (SVM) aims to find a $|\mathcal{D}_c| - 1$-dimensional hyperplane in the D_c-dimensional space (defined by the $|\mathcal{D}_c|$ input features) that distinctly classifies the data points $y_{c,t} = 0$ and $y_{c,t} = 1$. Among all possible hyperplanes, the goal is to find the one that leads to the maximum distance between points of both classes ($y_{c,t} = 0$ and $y_{c,t} = 1$) such that future outcomes can be classified with more confidence [28]. The problem is formulated as the following optimization model:

$$\max_\alpha = \sum_{t=1}^{|\mathcal{T}|} \alpha_t - \frac{1}{2}\sum_{t=1}^{|\mathcal{T}|}\sum_{u=1}^{|\mathcal{T}|} \alpha_t \alpha_u y_{c,t} y_{c,u} K(\mathbf{x}_{c,t}, \mathbf{x}_{c,u}) \tag{5}$$

s.t.

$$0 \le \alpha_t \le C \tag{6}$$

$$\sum_{t=1}^{|\mathcal{T}|} \alpha_t y_{c,t} = 0 \tag{7}$$

where C is a constant employed to penalize training errors, and K is the kernel function used to map the input features into a high-dimensional space, which allows to efficiently handle complex classification problems. In this work, the linear kernel function $K(\mathbf{x}_{c,t}, \mathbf{x}_{c,u}) = \mathbf{x}_{c,t}^\top \mathbf{x}_{c,u}$ was found to yield good performance. By solving this problem (5)-(7), we obtain the regularization parameters α that provides the wider margins between samples of both classes.

2.3.4. Preliminary model selection

The models are tested for predicting the maintainability of a 70 kV line in the Belgium transmission system in different conditions. In particular, we use a full year of data, which is subdivided into a training set (composed of 70% of the database) and a test set (consisting in the remaining 30%, i.e., 2621 hourly periods). The accuracy of the methods are compared using the success rate r, which is defined as the number of accurately classified samples (i.e., true positives and true negatives) on the 2621 points of the test set. Interestingly, the decision tree and the Naïve Bayes classifier do not require data pre-processing, while the SVM relies on the Min-Max technique (8) to normalize the data into the [0, 1] interval before training the model.

$$x_{d,t}^{\text{orig}} = \frac{x_{d,t}^{\text{orig}} - \min(x_d^{\text{orig}})}{\max(x_d^{\text{orig}}) - \min(x_d^{\text{orig}})} \quad \forall d \in \mathcal{D} \tag{8}$$

where x_d^{orig} is the vector of values (for all time steps t of the database) associated with variable $d \in \mathcal{D}$. Then, to make a fair comparison between classifiers, an extensive search is performed to find the best combination of hyper-parameters of each model, in order to maximize generalization capabilities and classification accuracy [29]. For each technique, different models were trained (with different values of hyper-parameters), and the best one is selected

Table 1
Confusion matrix of the different studied classifiers.

	Bayes		SVM		DT	
	$\hat{y}_0$	$\hat{y}_1$	$\hat{y}_0$	$\hat{y}_1$	$\hat{y}_0$	$\hat{y}_1$
y_1	112	2114	18	2508	16	2510
y_0	92	3	35	60	70	25
r	95.6%		96.4%		98.4%	

based on its accuracy achieved through a cross-validation procedure [30]. Outcomes are summarized in Table 1.

The decision tree (DT) outperforms both other classifiers in term of success rate r with a high score of 98.4%. The decision tree predicts that maintenance cannot be performed (i.e. $\hat{y}_{c,\tau} = 0$) for 16 h during which it could have been safely done (i.e., $y_{c,\tau} = 1$), and recommends a secure maintenance operation for 25 h during which reliability issues can be actually created in the transmission system. The SVM model is biased towards predicting safe maintenances, while the Naive Bayes classifier is globally less efficient but provides safer recommendations (with only 3 false positives). It is also worth mentioning that, in contrast with other models, the naive Bayes classifier is able to determine the maintainability $y_{c,t}$ even in the absence of some predictor values. However, due to the perfect state of the available database, this property is not useful in the context of this work.

Overall, the decision tree is selected as a basis for the rest of the study, which arises from their high accuracy (that lead to few false positives while suggesting sufficient time slots to perform maintenance actions), combined with their useful properties for field experts. Indeed, such techniques can be easily and intuitively understood, and can be fed by any type of input feature (continuous, integer) without the requirement to normalize these data. This bypasses the need to pre-process the database, which may lead to poor results if it is not properly realized.

2.3.5. Improvements of the classifier

In order to further improve the performance of the basic decision tree algorithm (without altering its ease of use and interpretability), different variants have been investigated. The idea is to combine multiple decision trees in order to decrease the variability of the resulting model. Such tree-based models have shown high predictive accuracy, even in high-dimensional problems with highly correlated features [31]. Two different learning strategies can be used.

On the one hand, bagging methods, such as random forest (RF), create (in a single step) N^{RF} independent decision trees, but each tree is built on a random subset of features, and each split of each tree is constructed based on a random sub-sample of the remaining data set. The final prediction is determined by averaging the results of the N^{RF} individual trees.

On the other hand, boosting methods, such as Gradient Boosting Decision Trees (GBDT) and eXtreme Gradient Boosting Trees (XGBoost), sequentially creates new models (in an additive fashion) to forecast the residuals of the global model obtained at the previous stage.

These models are tested in section 3, to illustrate their practical value for the TSO.

3. Case study

The proposed methodology is used on real data of Belgium. All classification tools have been implemented using the (open-source) R programming language. The network simulations have be carried out using the 'Power Factory' software, which performs contingency analysis, i.e., load-flow computations estimating the state of the power system in different contingency (outage of grid assets) scenarios.

In this part, we firstly present the metrics used to evaluate the performance of the different models in section 3.1. Then, we discuss the outcomes from the feature selection and the subsequent training of machine learning proxies in section 3.2 for two representative assets. Finally, the results are generalized, and the global maintainability of different electrical zones is studied in section 3.3.

3.1. Evaluation metric

For TSOs worldwide, false positives that erroneously predict safe outage windows should remain limited. It is thus necessary to control the trade-off between true and false positives, depending on the cost of misclassification for the TSO. This can achieved by adapting the threshold ρ from which a maintenance is classified as feasible. Typically, this threshold is fixed at $\rho = 0.5$ for binary classification tasks (i.e., a maintenance is deemed feasible if the classifier yields a probability higher than 50% that no contingency scenarios leads to unreliable or uneconomic grid conditions). However, more conservative values $\rho > 0.5$ can be privileged to decrease the prevalence of false positives. To guide the optimal selection of the decision threshold ρ, a receiver-operating characteristic (ROC) analysis is performed.

ROC curves display the rate of true positives (i.e., sensitivity) against the rate of false positives (1-specificity) for different cut-off values ρ. A perfect classifier has a ROC curve including the point (0, 1), which corresponds to no false positive and 100% of true positives. In this work, the optimal decision threshold ρ^* is thus defined as the one leading to the smallest Euclidean distance between the ROC curve and the optimal point (0, 1). In addition, it should be noted that the area under the ROC curve (which is referred to as AUC) also provides a valuable performance indicator. Indeed, a random classifier has a linear ROC curve passing though the point (0.5, 0.5) where the model cannot discriminate classes. This corresponds to an AUC of 0.5. A perfect classifier passing through the optimal point (0, 1) has an AUC of 1. This AUC metric, pertaining to the [0.5, 1] interval, is thus used to compare the accuracy of different classification models.

3.2. Performance of tree-based models

In this part, we evaluate the performance of the methodology in predicting outcomes from contingency analyses (to assess whether the maintenance of a specific asset can be performed without jeopardizing the operational security of the grid). To that end, we use one year of data, i.e., $|\mathcal{T}| = 8736$ hourly states, which are divided into a training set (including 70% of the samples) and a test set (with the remaining data). The analysis is performed for two different assets.

3.2.1. Asset 1

First, we investigate the maintainability of a major high-voltage transmission line of the Belgian system. To that end, we firstly select the most relevant explanatory variables (to predict the feasibility of maintenance actions in different conditions) by training a decision tree on the whole database. It should be noted that the trained model cannot be graphically represented due to the high-dimensionality of input data (i.e., more than 1000 variables). However, thanks to the methodology described in section 2.2, it is possible to quantify the importance of variables from the trained model. As a reminder, the most significant variables are those with an importance level higher than the one achieved by a fully random variable. Outcomes are depicted in Fig. 5.

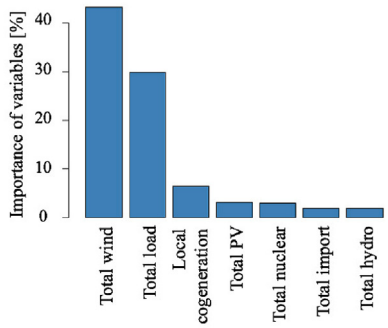

Fig. 5. Ranking of important variables affecting the feasibility of the maintenance of asset 1.

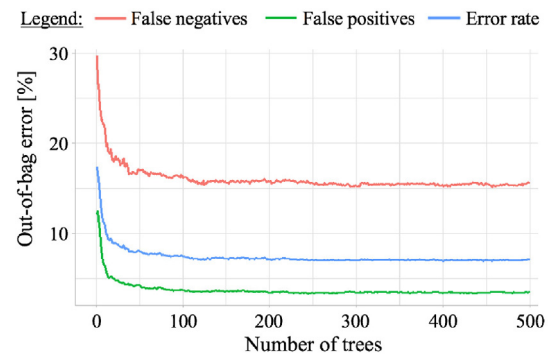

Fig. 7. Out-of-bag error as a function of the number of trees in the random forest.

In particular, we observe that the maintainability of the studied asset is strongly influenced by global grid conditions (i.e., the total Belgian load as well as the total wind and PV generation). Those results are coherent since the studied asset is a backbone overhead line of the transmission system, and it is thus logical that it is mainly affected by aggregated power exchanges. From these outcomes, it can be concluded that defining the maintenance plan of the studied line is not a simple task since the main explanatory factors cannot be reliably predicted over long horizons.

Then, we analyze the accuracy of the decision tree, by quantifying its performance on both training and test data. To that end, we use the ROC performance curve (presented in section 3.1), and the results are represented in Fig. 6.

In accordance with the current literature, we observe that the decision tree has a high variance, which leads to overfitting with a significant decrease of accuracy between training and test data, i.e., the AUC drops from 0.88 (in training) down to 0.8 when unseen data are fed into the trained model. Hence, in this work, decision trees are only used for their ability to identify the most important variables during the training (and to quantify the importance of those features), and the actual predictions are performed with advanced tree-based models.

Based on the selected features, a Random Forest (RF) can be trained. In RF, an important hyper-parameter is the number of decision trees that are averaged to reduce the variance of the model. When the number of trees increases, the resulting model takes more time to train, and it is thus valuable to find the optimal trade-off between accuracy and simulation time. In Fig. 7, we therefore depict the out-of-bag error of the RF, which is computed by feeding each tree (of the RF) with the samples that were not used in the learning procedure, and by averaging the classification errors.

One can see that the optimal solution is to use around 100 decision trees in the forest. Indeed, when this number increases, the performance gains become marginal, while the simulation time inherently increases. The number of false positives and negatives are those associated with a default value of the decision threshold, i.e., $\rho = 0.5$.

It is also important to optimize the other hyper-parameters of the RF model (to maximize performance while limiting overfitting). Here, individual trees (forming the RF) are trained with 75% of the feature space, and a minimum of 5 samples are required to split an internal node. However, we fix no limit on the depth of each tree (i.e., we do not limit the maximum number of allowed splits of the input space). For GBDT, the number of boosting stages is fixed at 100 (to reach a good trade-off between accuracy and computation time). An additional hyper-parameter to be optimized is the learning rate (which weights the contribution of each tree). Here, a value of 0.1 is selected. Once the hyper-parameters are fixed, both RF and GBDT models are trained, and are then evaluated on the test set. The resulting outcomes are shown in Fig. 8.

We see that both tree-based models exhibit a high accuracy, with the RF (AUC = 0.93) slightly outperforming the GBDT

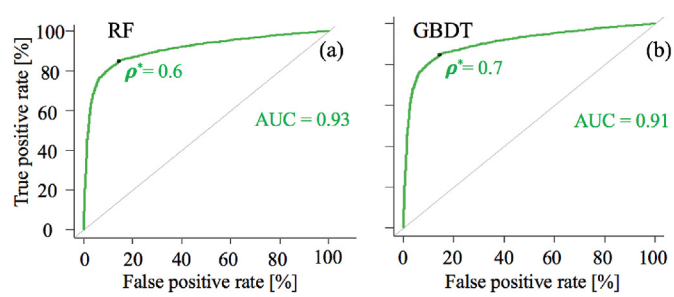

Fig. 8. ROC curves of the random forest (a), and gradient boosting decision tree (b), using the features selected by the decision tree.

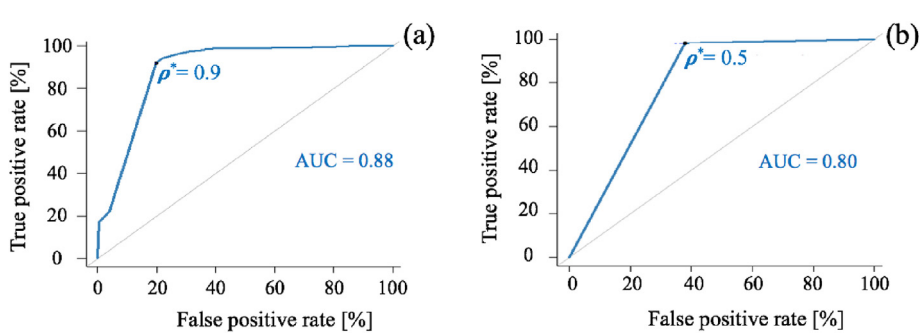

Fig. 6. Performance of the decision tree, measured on both training data (a), and on the test set (b).

(AUC = 0.91). In that regard, the ROC curves of both models are very close to the theoretical optimum point (0, 1) that represent a perfect classifier. Interestingly, the models are characterized by an optimal decision threshold ρ^* (from which maintenances can be classified as feasible) more conservative than the traditional cut-off point of $\rho = 0.5$. Such results are consistently observed across simulations (for different samples $t \in \mathcal{T}$, for different assets $c \in \mathcal{C}$).

Since RF is more robust to changes in the hyper-parameters (such that it requires less hand tuning in the training phase), and has calculation times roughly twice as small as the GBDT, only RF is kept by the TSO for the following of the study.

Finally, to quantify the added value of our feature selection procedure (presented in section 2.2), the obtained solution (Fig. 5) is compared with two other benchmarks. In the first one, we only select the two most important variables, while the second one considers all input features. The features selected by the different methods are then separately fed to the RF classification tool (to identify the feasibility of maintenances). Interestingly, we observe an AUC of 0.87 for the model with 2 inputs (for an AUC of 0.93 with the reference RF using the optimal number of variables). This accuracy loss (due to the loss of explanatory information) is not accompanied with significant time gains. Then, the RF with all variables achieves an AUC of 0.94, but treating the high-dimensional input space (>1000) significantly complicates the training process. Moreover, this strategy prevents identifying the most important features, which is highly valuable information for expert teams.

3.2.2. Asset 2

The methodology is now illustrated for a power transformer localized in a different area of the grid. A dedicated decision tree is used to determine the most significant input features, whose ranking is presented in Fig. 9.

The outageability of this asset is mainly affected by the load consumption (i.e. total load). The influence of other variables (PV, hydro and biomass generation) is very limited. The selected explanatory variables are leveraged to train a random forest, and the resulting ROC curve is shown in Fig. 10.

The accuracy of the RF model is very high, with an AUC of 0.97. The optimal decision threshold is obtained for a conservative value of $\rho^* = 0.9$. However, this threshold still leads to around 10% of false positives, and the TSO can thus use the ROC curve to select an even more conservative cut-off ρ, while taking care that sufficient outage windows are still available for the maintenance activities.

This framework (building fast and reliable surrogates for contingency analyses) has been tested on a large scale (for many assets in different areas), and similar results were consistently achieved. These models can thus be reliably integrated into planning tools. In parallel, as further discussed in the following section 3.3, these models can also be used (following an aggregation of results) to quickly predict the global maintainability of the grid.

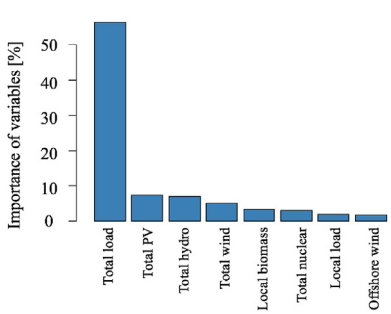

Fig. 9. Ranking of the most important variables influencing the operation of asset 2.

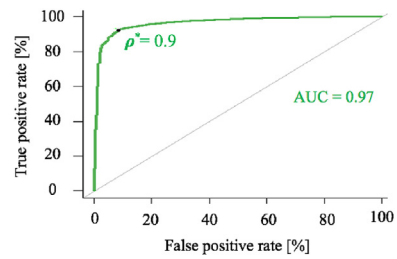

Fig. 10. ROC curve of the random forest (on the test set) for the second studied asset.

3.3. Model exploitation

The machine learning proxy developed in this work may serve two main purposes. First, from an operational perspective, the model can be used to directly inform maintenance planning tools with time periods during which grid assets can be safely put out of operation. In this way, there is no need to simulate all contingencies within the decision procedure. The resulting time gains can bring a lot of value for operational teams that have to frequently make use of such planning tools for dynamically accommodating the outage schedule as new information becomes available (such as delays in previous maintenances) [32,33]. Second, given the high number of grid assets (i.e., transmission lines and transformers), it is of high interest for the TSO to have an overview of the maintainability of all assets at an aggregated level. This second aspect is further investigated in the rest of the paper.

To have a global quantification of the maintainability of the system, two complementary indicators are developed (section 3.3.1). The first one yields the prevalence of safe time slots during which planned outage is expected to minimize reliability and economic factors, while the second one characterizes the complexity of accurately forecasting suitable periods for outage scheduling. These indicators are computed for three different zones of the Belgian transmission system (section 3.3.2), which allows to evaluate the health status of these zones.

3.3.1. Criticality and unpredictability indicators

First, we define a criticality score characterizing the difficulty of finding adequate periods for scheduling safe outages. This indicator is defined as the number of working days during which the maintenance of an asset is defined as unsafe (for at least 1 h) on the total number of working days during the year. This indicator therefore varies between 0 (when no opportune moment is available to plan an outage for maintenance) and 1 (when the asset can be put out of operation during all working days without compromising the grid's security).

This value is very useful for planning teams that need to define optimal outage schedules, while assessing the robustness of the transmission system.

In complement to the criticality score, we define a score quantifying the unpredictability of maintenances, based on the importance of the (selected) predictor variables. The goal is to identify assets for which safe outage windows are significantly affected by variables that are complex to forecast (e.g., wind, solar irradiance, etc.). Practically, this indicator is computed as the aggregated contribution of the unpredictable variables divided by the contributions of all input variables. A score of 0 means that the maintainability of an asset only depends on features that are easy to predict (such as the nuclear generation), and a score of 1 is imputed to assets whose operation conditions are influenced by more

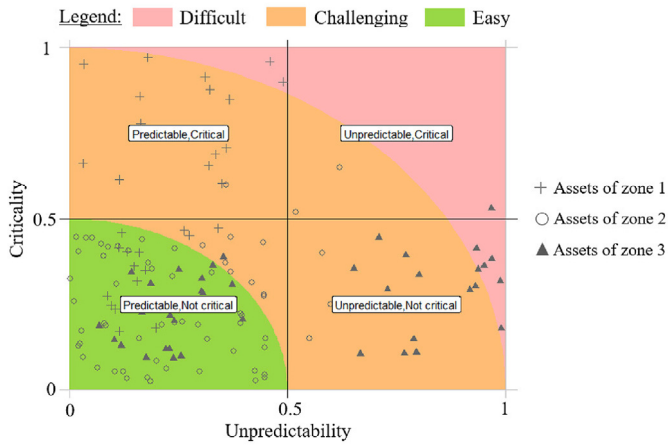

Fig. 11. Criticality and unpredictability scores for all assets within three different zones of the Belgian system.

complex to forecast variables.

This information is highly valuable for the TSO that can differentiate assets which can be incorporated into long-term maintenance planning from assets to be scheduled in the short-run (to rely on a consistent prediction of the stochastic variables influencing their loading conditions). Moreover, the results are inherently interpretable, which can strengthen the acceptability of the proposed framework by operational teams.

3.3.2. Global maintainability of specific zones

The criticality and unpredictability indicators can be used to assess the outageability for maintenance of the whole system, and they are here computed for three different zones of the Belgian transmission grid. To that end, all assets (i.e., transmission lines and transformers within each of the studied zones) are displayed in Fig. 11 in a two-dimensional space defined by the indicators. A different symbol is used for assets located in different zones. For ease of interpretation, the graph is divided into three different parts (separated by iso-risk concentric circles defined by the Belgian TSO) that yield an expert-based categorization between assets that are "easy" and "difficult" to maintain, with an intermediate "challenging" category.

We observe that most assets are in the lower-left part of the graph, which is an indicator of the good maintainability of the three studied zones. Interestingly for the TSO, many assets are mostly influenced by predictable variables (mainly the total Belgian load). Moreover, those maintenance actions are likely to be performed without compromising the security of the power grid (since many safe time slots are available throughout the year). This clearly highlights the robustness of the transmission system in these studied areas. However, some assets are located far away from the optimal point (0, 0), i.e., with high values of unpredictability and criticality indicators, and the TSO should thus keep a particular focus on the zones containing such grid assets.

A solution is to investigate the evolution of the maintainability of the zones over the next few years. This is achieved by relying on representative scenarios of the future climate years (for estimating more closely the real situations faced in operation, such as years with long heat waves, or years with a lot of wind), along with projection of the future load behavior and generation mix (including the renewable energy uptake). These scenarios are then used as inputs for the trained tree-based models in order to predict the outageability of each asset for each hourly period of the studied years. This procedure is performed for all grid assets (of the three studied zones) for three different years, i.e., 2023, 2025 and 2030. The outcomes are depicted in Fig. 12. It should be noted that the light computation burden of using (trained) random forests, those results are obtained in less than 15 min.

It can be seen in Fig. 12 that the maintainability of the three studied zones are expected to follow different trends. In Fig. 11, it was observed that many assets of zones 3 were affected by renewable-based generation. Hence, with the projected increase of these resources [34], predicting the feasibility of outages for maintenance activities will become more challenging. However, the future maintainability of assets pertaining to zones 1 and 2 is predicted to remain stable (with only a small decrease over the next decade). In that regard, the maintenance of more than 75% of assets of zone 2 will still be easily scheduled in 2030. It is therefore not necessary for the TSO to consider reinforcing the network in this zone in the near future. By applying this procedure at the grid scale, operational teams can thus improve their knowledge of the grid, and easily identify the problematic zones as well as the specific challenges therein. This information may then serve as a reliable basis to take adequate investment solutions.

4. Conclusions

In this paper, a new framework to improve security-constrained outage planning in modern power systems is presented. Different machine learning techniques are tested to provide fast and accurate surrogates of contingency analyses, with the goal of predicting the feasibility of a maintenance in specific grid conditions. To reduce overfitting risks, while identifying the most important explanatory variables, the tools are enriched with a tailored input selection process.

The methodology is applied to different assets pertaining to the Belgian transmission system. First, it is observed that the feature screening procedure provides relevant outcomes (typically

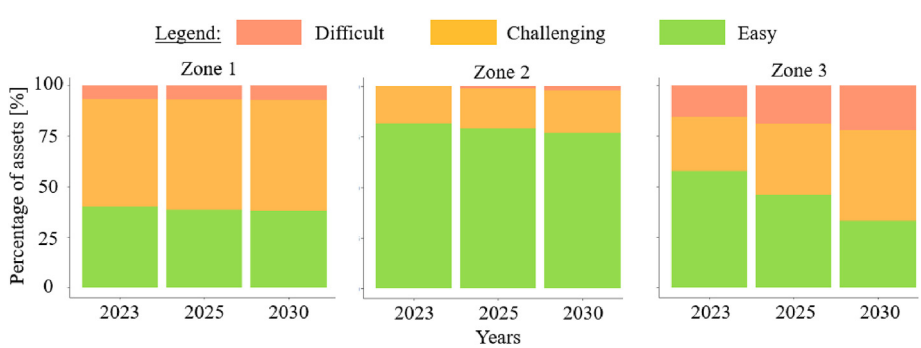

Fig. 12. Evolution of the maintainability of zone 1 up until 2030 based on the expected input scenarios of the Belgian TSO.

reducing the initial set of more than 1000 features down to 5 to 15 variables), which is illustrated by the performance of subsequent classification tools (dedicated to predicting the feasibility of maintenance actions). In particular, random forests achieve the highest prediction accuracy, with a success rate steadily higher than 90%. We also show that further reducing the input space (down to 2 features) may strongly affect the prediction performance, with a loss in accuracy higher than 5%.

Then, simulations on scenarios of the future Belgian grid (for years 2025–2030) reveal that the maintainability of the system may be significantly impacted by the penetration of renewable generation, with a reduction of down to 20% of the safe maintenance periods in some grid areas.

Overall, the random forests developed in this work have attractive characteristics, by combining high performance, easy of use and interpretability properties, thus bridging the gap between machine learning and field expertise.

As a interesting perspective, the models could be trained to deal with uncertainties, translating the fact that the values of input features throughout the planning horizon are not perfectly known. Moreover, the integration of these models into outage planning tools is an important future step to improve the cost-efficiency of grid maintenances.

Credit author statement

Conceptualization, L. Hubert, N. Marenne and J. Sprooten; Methodology, J.-F. Toubeau, L. Pardoen, L. Hubert; Validation, J.-F. Toubeau, Z. De Grève and N. Marenne; Writing–original draft preparation, J.-F. Toubeau and L. Pardoen; Writing–review and editing, J.-F. Toubeau, Z. De Grève and F. Vallée; Supervision, F. Vallée and J. Sprooten; Project administration, F. Vallée and J. Sprooten. All authors have read and agreed to the published version of the manuscript.

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.

References

[1] Toubeau J-F, Vallée F, De Grève Z, Lobry J. A new approach based on the experimental design method for the improvement of the operational efficiency in Medium Voltage distribution networks. Int J Electr Power Energy Syst 2015;66:116–24.

[2] Fu J, Nunez A, De Schutter B. A short-term preventive maintenance scheduling method for distribution networks with distributed generators and batteries. *IEEE Trans. Power Syst.*; 2020.

[3] Schneider Joachim, Gaul AJ, Neumann C, Hogräfer J, Wellßow W, Schwan M, Schnettler A. Asset management techniques. Elect. Power Energy Syst. 2006;28:643–54.

[4] Khuntia RS, Rueda JL, Bouwman S, Meijden AM. A literature survey on asset management in electrical power [transmission and distribution] system. Int Trans Electr Eng Syst 2016;26:2123–33.

[5] Bertling L, Allan R, Eriksson R. A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. IEEE Trans Power Syst Feb 2005;20(1):75–82.

[6] Beehler ME. Reliability centered maintenance for transmission systems. IEEE Trans Power Deliv April 1997;12(2):1023–8.

[7] Klonari V, Toubeau J-F, Lobry J, Vallée F. Photovoltaic integration in smart city power distribution: a probabilistic photovoltaic hosting capacity assessment based on smart metering data. In: 2016 5th international conference on smart cities and green ICT systems. Rome: SMARTGREENS; 2016. p. 1–13.

[8] Toubeau J-F, Morstyn T, Bottieau J, Zheng K, Apostolopoulou D, De Grève Z, Wang Y, Vallée F. Capturing spatio-temporal dependencies in the probabilistic forecasting of distribution locational marginal prices. IEEE Trans Smart Grid May 2021;12(3):2663–74.

[9] Tang Y, Liu Q, Jing J, Yang Y, Zou Z. A framework for identification of maintenance significant items in reliability centered maintenance. Energy 2016;118:1295–303.

[10] Rusin A, Bieniek M. Maintenance planning of power plant elements based on avoided risk value. Energy 2017;134:672–80.

[11] Wang J, Liang Y, Zheng Y, Gao RX, Zhang F. An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. Renew Energy 2020;145:642–50.

[12] Cauchi N, Macek K, Abate A. Model-based predictive maintenance in building automation systems with user discomfort. Energy 2017;138:306–15.

[13] Shin W, Han J, Rhee W. AI-assistance for predictive maintenance of renewable energy systems. Energy 2021;221:1197.

[14] Eskandarpour R, Khodaei A. Leveraging accuracy-uncertainty tradeoff in SVM to achieve highly accurate outage predictions. IEEE Trans Power Syst Jan. 2018;33(1):1139–41.

[15] Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A. Machine learning for predictive maintenance: a multiple classifier approach. IEEE Trans Ind Inform June 2015;11(3):812–20.

[16] Kaparthi S, Bumblauskas D. Designing predictive maintenance systems using decision tree-based machine learning techniques. Int J Qual Reliab Manag 2020;37:659–86.

[17] Li W, Korczynski J. A reliability-based approach to transmission maintenance planning and its application in BC hydro system. IEEE Trans Power Deliv Jan. 2004;19(1):303–8.

[18] Liu J, Kazemi M, Motamedi A, Zareipour H, Rippon J. Security-constrained optimal scheduling of transmission outages with load curtailment. IEEE Trans Power Syst Jan. 2018;33(1):921–31.

[19] Dalal G, Gilboa E, Mannor S, Wehenkel L. Chance-constrained outage scheduling using a machine learning proxy. IEEE Trans Power Syst 2019;34(4):2528–40.

[20] Gargari MZ, Hagh MT, Zadeh SG. Preventive maintenance scheduling of multi energy microgrid to enhance the resiliency of system. Energy 2021;221:119782.

[21] Shayesteh E, Yu J, Hilber P. Maintenance optimization of power systems with renewable energy sources integrated. Energy 2018;149:577–86.

[22] Wang Y, Li Z, Shahidehpour M, Wu L, Guo CX, Zhu B. Stochastic Co-optimization of midterm and short-term maintenance outage scheduling considering covariates in power systems. IEEE Trans Power Syst Nov. 2016;31(6):4795–805.

[23] Elia NV. Grid data. available at: http://www.elia.be/en/grid-data; 2021.

[24] Carnero MC, Gómez A. Maintenance strategy selection in electric power distribution systems. Energy 2017;129:255–72.

[25] Raileanu LE, Stoffel K. Theoretical comparison between the Gini index and information gain criteria. Ann Math Artif Intell 2004:77–93.

[26] Toubeau J-F, Bottieau J, Vallée F, De Grève Z. Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. IEEE Trans Power Syst March 2019;34(2):1203–15.

[27] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn November 1997;29(2–3):131–63.

[28] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl July-Aug. 1998;13(4):18–28.

[29] Bottieau J, Hubert L, De Grève Z, Vallée F, Toubeau J-F. Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement. IEEE Trans Power Syst March 2020;35(2):1218–30.

[30] Schaffer C. Selecting a classification method by cross-validation. Mach Learn 1993;13:135–43.

[31] Liefferinckx C, Bottieau J, Toubeau J-F, Thomas D, Rahier J-F, Louis E, et al. Collecting new peak and intermediate infliximab levels to predict remission in inflammatory bowel diseases. Inflamm Bowel Dis Mar 30 2021. https://doi.org/10.1093/ibd/izab042.

[32] Burke EK, Smith AJ. Hybrid evolutionary techniques for the maintenance scheduling problem. IEEE Trans Power Syst Feb. 2000;15(1):122–8.

[33] Heo J, et al. A reliability-centered approach to an optimal maintenance strategy in transmission systems using a genetic algorithm. IEEE Trans Power Deliv Oct. 2011;26(4):2171–9.

[34] Toubeau J-F, Bottieau J, De Grève Z, Vallée F, Bruninx K. Data-driven scheduling of energy storage in day-ahead energy and reserve markets with probabilistic guarantees on real-time delivery. IEEE Trans Power Syst July 2021;36(4):2815–28.