Logic-Based Explanations of Imbalance Price Forecasts using Boosted Trees

J. Bottieau[1], G. Audemard[2], S. Bellart[2], J-M. Lagniez[2], P. Marquis[2,3], N. Szczepanski[4], and J.-F. Toubeau[1]
[1]Power Systems & Markets Research Group, Electrical Power Engineering Unit, University of Mons, Belgium.
[2]CRIL, University of Artois & CNRS, France.
[3]Institut Universitaire de France, France.
[4]IRT SystemX, Saclay, France.

***Abstract*—Explainability is one of the keys to foster the acceptance of Machine Learning (ML) models in safety-critical fields such as power systems. Given an input instance $\mathbf{x}$ and a complex ML model f, the driving features of the corresponding output are commonly derived using model-agnostic approaches such as SHAP. Although being generic, such approaches offer limited guarantees about the quality of the explanations they provide. In this paper, we opt for a logic-based approach to derive post-hoc explanations. Our approach provides formal guarantees about the explanations t that are generated for input instances $\mathbf{x}$ given an interval I containing $f(\mathbf{x})$ and representing the admissible imprecision about $f(\mathbf{x})$. Thus, our approach ensures that the prediction $f(\mathbf{x}')$ on every instance $\mathbf{x}'$ covered by t belongs to I as well. In our work, f is a boosted tree, which is accurate and associated with an equivalent logical representation. The forecasted variable is the imbalance price, which is an important market signal for trading strategies of energy traders. The outcomes –using data from the Belgian power system– shed light on the input patterns that drive a high or low imbalance price prediction, while investigating whether such input patterns are intelligible for a human explainee.**

***Index Terms*—Automated reasoning, Electricity price forecasting, Explainable AI, Real-time price.**

I. Introduction

Due to the necessity to continuously keep a balance between generation and consumption, the real-time trade and pricing of electricity is central in liberalized power systems [1]. In this way, market participants, such as producers and suppliers, trade quantities of energy in forward markets based on their expectation of (future) real-time electricity prices. For each delivery period, if they produce or consume quantities that differ from what they have contracted, imbalance volumes are settled at a real-time price, called the imbalance price in Europe. Depending on the system conditions, imbalance payments may be extremely high, which can lead to bankruptcy for energy traders if they are not appropriately hedged. On the other hand, such high prices may also offer valuable temporal arbitraging opportunities for fast-ramping units.

The electricity price forecasting literature has shown evidence that imbalance prices might be very hard to predict at the day-ahead stage [2]. Indeed, before the closure of the day-ahead market, imbalance volumes of market actors, i.e., metered deviations from forward market transactions, are inherently random at this stage. Yet, as we get closer to the delivery period, imbalance volumes and prices can be more accurately predicted [3]. In particular, authors in [4] found that balancing prices in Great Britain are predictable one hour-ahead and that a Markov non-linear approach outperforms a linear model. The use of a regime-switching model was motivated by the imbalance price formation, whose prices can be low or high according to whether the net imbalance volumes are positive or negative. Still on a British case study, non-linear models based on Machine Learning (ML) techniques have been compared in [5], where XGBoost exhibits a higher performance than other methods. This observation was confirmed in [6], which also showed the superior forecasting accuracy of deep learning compared to econometrics models.

However, for facilitating the adoption of ML techniques in power systems, ML models should not only be very accurate but also trustworthy, i.e., robust and explainable [7], [8]. In this direction, probabilistic forecasters have been developed for imbalance price prediction in [9], [10], where ML models, from tree-based ensemble methods to advanced neural network architectures, demonstrate impressive, but not guaranteed, accuracy of quantile forecasts. While very accurate predictions can be achieved in practice, it must be kept in mind that 100% correct predictions cannot be ensured in general. To improve confidence in the predictions, an efficient solution is to derive explanations about outcomes, i.e., providing pieces of evidence that allow the practitioner to possibly reject a prediction when the corresponding explanations do not comply with expectations.

In this line of research, both [9], [10] try to enrich their models with notions of explainability, based on the scores obtained from attention layers of neural forecasters. While providing insights into the model's inner logic, such attention scores still fail to generate explanations from which one could reason, which limits their usefulness for users [11]. Moreover, such attention scores are model-specific. Alternatively, additive feature attribution methods score each feature relative to its importance for a particular outcome. Such approaches are often model-agnostic and thus widely applicable. In particular, the SHAP method was proposed as a unified measure of feature importance based on Shapley values [12]. Such values are derived from cooperative game theory, deciphering the marginal contribution of each feature to a specific ML outcome. Nevertheless, ignoring the underlying ML model,

Submitted to the 23rd Power Systems Computation Conference (PSCC 2024).

The work is supported via the energy transition funds project 'ADABEL' organized by the FPS economy, S.M.E.s, Self-employed and Energy.

and proposing an explanation given the training data, and an estimated distribution of the input still has some severe limitations. In this way, Shapley values do not offer any formal guarantee, i.e., no conclusion about predictions from the ML model can be deductively drawn. Recent works [13] have even demonstrated that Shapley values for explainability may provide misleading information about feature importance. Finally, common explainable approaches have been developed for computer vision applications. Though naively applying such methods to regression can occasionally yield useful results, several works show the importance of tailoring explanation techniques to the specificities of the regression problem [14].

Formal explainable methods (also called logic-based methods) can tackle these two limitations for regression problems. They provide the underlying conditions leading to a particular outcome, which is achieved by leveraging the logical structure of the model. Among useful explanations are the abductive ones. Such explanations are subsets t of the characteristics of the input instance $\mathbf{x}$ intended to reflect why $\mathbf{x}$ has been mapped to its corresponding values $f(\mathbf{x})$ by the regression function f.

In our work, f is a boosted tree derived using the XGBoost algorithm. Practically, given an input instance $\mathbf{x}$ and a user-defined interval range I such that $f(\mathbf{x}) \in I$, an abductive explanation aims at answering, in the most concise fashion, why the forecaster makes a prediction within the predefined interval I. Interval I enables the explainee to specify how he/she cares about the exact value of $f(\mathbf{x})$, by representing the imprecision on $f(\mathbf{x})$ the explainee is ready to tolerate. Pointing out t may allow the human explainee to decide whether or not the prediction $f(\mathbf{x})$ can be trusted. Note that works in [15], [16] show the possibility of extending abductive explanations to neural networks, but empirical findings suggest that scalability remains a significant challenge for neural networks with more a few dozen neurons.

In the following, we show the usefulness of generating abductive explanations for predicting imbalance prices, taking advantage of recent developments in formal explainable AI using boosted regression trees [17]. Formally explaining these forecasts can provide valuable insights into the factors that drive the imbalance price regime, while increasing the user's trust on the prediction. This complements the current explainable methods, wherein the importance of the driving factors are revealed from either linear regression parameters or SHAP values, see e.g., [18] and [19]. Overall, the contributions of this paper can be summarized as follows:

- We present a logic-based method for explaining imbalance price forecasts. The method ensures a level of rigor as explanations are directly extracted from the logical representation of the prediction model, and tailors the explanations according to the use-case of the explainee.
- Through an academic case study, we identify how logic-based explanations can help understanding and debugging a faulty model, compared to SHAP values. Although both types of explanations are local, derived from specific instances under examination, logic-based explanations offer a distinct advantage: they can be generalized across different instances exhibiting the same pattern. This generalizability enables the making of robust deductions about the functioning of the model.
- Finally, the applicability of logic-based explanations is analysed on a real-life case study. Besides the size of explanations and computational aspects, we investigate the driving factors that conduct the prediction model to infer a high or low imbalance price regime.

TABLE I: Imbalance price mechanism.

P = Production C = Consumption SI = System Imbalance		Transmission System Operator (TSO) control area	
		Excess (P>C) Positive SI	Shortage (P<C) Negative SI
BRP area	Excess	MDP (low price) TSO pays BRP	MIP (high price) TSO pays BRP
	Shortage	MDP (low price) BRP pays TSO	MIP (high price) BRP pays TSO

The rest of the paper is organized as follows. Section II presents key concepts involved in the determination of imbalance electricity price in Europe. Section III focuses on the methodology used to compute explanations for the predictions. Section IV provides a simple example. Section V presents a case study. Section VI concludes the paper, and gives perspectives for further research.

II. Imbalance Price Mechanism

In Europe, imbalance prices are intended to reflect the real-time value of energy. Following neoclassical economics, the real-time value of energy is currently defined as the cost of providing the marginal amount of energy needed to balance the system. This is exemplified in Table I, which shows the single price imbalance mechanism. We can observe that the imbalance price is formed according to two price regimes: i) the marginal incremental price (MIP), which amounts, for a given period, to the price of the most expensive upward balancing service activated, and ii), the marginal decremental price (MDP), which amounts, for a given period, to the price of the least remunerative downward balancing service. When the control area is lacking generation, i.e., the system imbalance (SI) is SI < 0, the Transmission System Operator (TSO) mainly activates upward balancing services, which drives the imbalance price towards a high price regime (typically above the day-ahead market price). On the other hand, when the system has a production surplus (SI > 0), downward balancing services are activated, driving the imbalance price towards a low price regime (typically below the day-ahead market price). Hence, when a market agent (named Balance Responsible Party, abbreviated BRP thereafter) is in imbalance, it is charged at an imbalance price. Depending on whether its imbalance position helps or worsens the global system imbalance, the BRP receives economic gain opportunities (green boxes of Table I) or financial penalties (red boxes of Table I).

We focus on the Belgian case study, whose imbalance volumes are settled every 15 minutes using a single imbalance

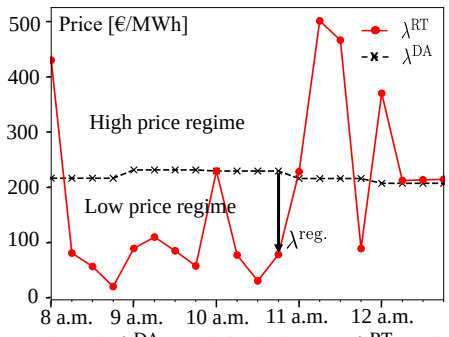


Fig. 1: Day-ahead (λ^{DA}) and imbalance (λ^{RT}) prices on January 8, 2022 between 8 a.m. and 1 p.m.. The black arrow is the imbalance premium ($\lambda^{reg.}$) between both market floors.

price mechanism based on marginal pricing (as in Table 1). The only difference is that the Belgian TSO may use adders on the imbalance price for increasing the incentive to BRPs in helping the system in case of structural imbalances. In our models, the imbalance premium is hence considered instead of the imbalance price. The imbalance premium is defined as the difference between the imbalance and the day-ahead market prices. Fig. 1 showcases these three prices for the day of January 8 2022 between 8 am and 1 pm. The imbalance price (λ^{RT}) and the corresponding day-ahead market price (λ^{DA}) are respectively plotted in red and black, while the imbalance premium (denoted $\lambda^{reg.}$) is showed by a black arrow. We can observe the regime switching behavior of the imbalance price, where a high imbalance price regime corresponds to a positive imbalance premium, while a low imbalance price regime is characterized by a negative imbalance premium.

III. A Logic-Based Approach to Explanations

This section presents how logic-based explanations can be derived from an already trained boosted tree f. The flowchart of the logic-based explainer is described in Fig. 2, where explanations are extracted ex-post based on i) the input instance $\mathbf{x_e}$, ii) the logical representation of f, and iii) the interval I about $f(\mathbf{x_e})$ admissible by the explainee. In a nutshell, given an instance $\mathbf{x_e}$ and an interval I, such that $f(\mathbf{x_e}) \in I$, the logic-based explainer aims at extracting a minimal subset of conditions that are sufficient to explain why $f(\mathbf{x_e}) \in I$. Section III-A provides a formal definition of the abductive explanations being sought, while Section III-B presents briefly the generation process of such explanations.

A. Formal definition of an abductive explanation

Practically, a boosted tree f can be seen as a a linear combination of n binary trees (T_i, for $i \in [n]$), whose internal nodes are labeled by Boolean conditions related to m attributes and their leaves are labeled by real numbers. For a given instance $\mathbf{x_e}$, the weight $T_i(\mathbf{x_e}) \in \mathbb{R}$ of a tree T_i is given by the label of the leaf reached from the root as follows: at each node go to the left or right child depending on whether or not the condition labeling the node is satisfied by $\mathbf{x_e}$. The prediction f of an instance $\mathbf{x_e}$ is given by the sum of the weights of all trees for this instance.

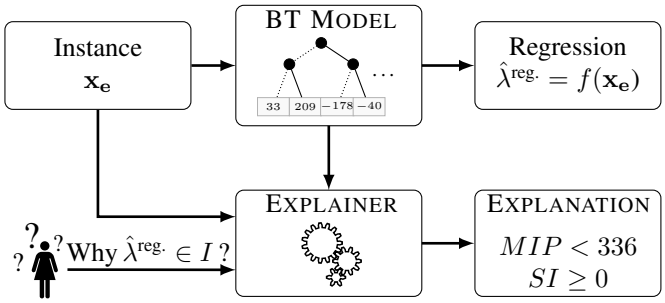


Fig. 2: Flowchart of the logic-based explainable method

.

Given this logical representation, the boosted tree $f : \mathbb{R}^m \rightarrow \mathbb{R}$ can be alternatively viewed as a mapping from $\{0,1\}^p$ to $\mathbb{R}$, where p denotes the cardinality of the set of Boolean conditions (over attributes) $\mathcal{B}$ used in the model. We denote $t_{\mathbf{x_e}}$ a conjunctively-interpreted set of p Boolean literals (one per condition) such that the literal associated with a condition is positive (resp. negative) whenever $\mathbf{x_e}$ satisfies (resp. does not satisfy) the condition.

For the regression task, an abductive explanation for an instance $\mathbf{x_e}$ is intended to explain why the regression value $f(\mathbf{x_e})$ belongs to I [17].

More formally, a subset-minimal abductive explanation for an instance $\mathbf{x_e}$ given f and an interval I is a subset of $t_{\mathbf{x_e}}$ (the Boolean conditions $\mathcal{B}$) such that any instance $\mathbf{x'_e}$ sharing the same conditions satisfies $f(\mathbf{x'_e}) \in I$, and no proper subset of this explanation satisfies this last property. The interval I allows a human explainee to tailor his/her question for the regression task at hand. Hence, an interval equals to the model prediction outcome, i.e., $I = f(\mathbf{x_e})$, explains exactly why we obtained this prediction value, but at the cost of generating a lengthy explanation (the term $t_{\mathbf{x_e}}$ in the worst-case scenario). On the other hand, a very large interval containing the lower and upper bounds of the prediction model necessitates none explanation. Note that all these formal notions are illustrated in Section IV-A for a boosted model with 3 trees.

B. Generation process of an abductive explanation

The general approach for generating abductive explanations consists of a greedy algorithm which tries to eliminate iteratively the Boolean characteristics from $t_{\mathbf{x_e}}$, while satisfying a coverage test (here induced by I). In logic, this covering test is coNP-complete implying that it is computational hard task in general. For tackling this task timely, the greedy algorithm is first run with approximate, yet efficient, coverage tests, before using exact, but expensive, ones on the remaining Boolean characteristics.

More particularly, approximate coverage tests are verified by summing the minimal (resp. maximal) leaf values of each tree T_i that could be reached based on the tested term t. This approximation acts as worst-case scenarios for the coverage test, and can be run in polynomial-time. The output is an abductive explanation, but without any guarantees about subset-minimality, i.e., some Boolean characteristics can be kept even if their removal does not necessarily violate the

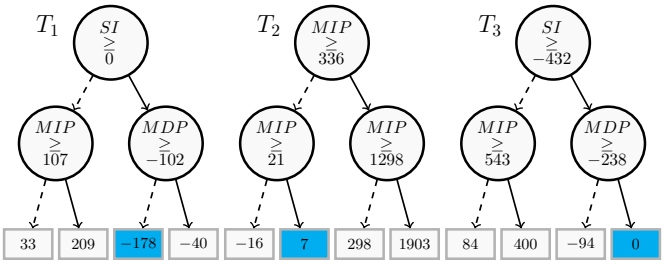


Fig. 3: The boosted model with three trees T_1, T_2, and T_3.

exact coverage test. That is why, based on this first explanation, the greedy algorithm is run a second time with exact coverage tests, ensuring the subset-minimality property. The exact coverage test is achieved by solving a set of MILP constraints encoding the boosted model, the tested term t, and the coverage condition I. This two-step strategy leads to substantial computational improvements, which renders the generation of abductive explanations for boosted trees scalable in practice [17], [20]. Note that a python library dedicated to formal explanations can be found in [21].

IV. An academic example

This section illustrates the value of computing abductive explanations via an academic example. The academic example is a simplified version of the imbalance premium prediction task $\hat{\lambda}^{\text{reg.}}_{t_{0+1}} = \mathbb{E}(\lambda^{regime}_{t_{0+1}}|\mathbf{x})$, where perfect knowledge of the real-time market operations is assumed, i.e., the imbalance of the system ($\text{SI}_{t_{0+1}}$), the marginal incremental price ($\text{MIP}_{t_{0+1}}$), and the marginal decremental price ($\text{MDP}_{t_{0+1}}$) are known[1]. Recalling Table I, we observe that all the required information are available for modelling the imbalance price mechanism, at the exception of the imbalance price adders.

The data spans from January 5, 2020 to January 10, 2022, with a 15-minute time granularity. Each day is thus composed of 96 instances, resulting in 70,656 instances for the whole dataset. The first two years are used to train and validate the prediction models (69,792 instances), while the last 10 days are used for out-of-sample analysis (864 instances). The boosted model is selected because such models are quite accurate and formal explanations can be derived efficiently from them, as explained in the previous section. For the sake of clarity, the boosted model owns 3 regression trees, which are limited to a depth of 2, with a learning rate of 1. Section IV-A first illustrates the formal notions of an abductive explanation on the boosted model presented in Section III-A, while Section IV-B compares our abductive explanations with mathematical guarantees to those obtained with SHAP.

A. Formal description of the boosted model

Fig. 3 exhibits the already trained boosted model with 3 trees. The instance $\mathbf{x_e}$ is thus defined by 3 attributes $(\text{SI}, \text{MIP}, \text{MDP})$. The set of Boolean conditions associated with the boosted model contains $p = 9$ conditions, i.e., $\{(\text{SI} \geq -432), (\text{SI} \geq 0), (\text{MIP} \geq 21), (\text{MIP} \geq 107), (\text{MIP} \geq 336), (\text{MIP} \geq 543), (\text{MIP} \geq 1298), (\text{MDP} \geq -238), (\text{MDP} \geq -102)\}$. Hence, for an instance $\mathbf{x_e} = (324.5, 103.5, -140)$, $t_{\mathbf{x_e}} = (1, 1, 1, 1, 0, 0, 0, 1, 0)$. Indeed, $(\text{SI} \geq -432)$, $(\text{SI} \geq 0)$, $(\text{MIP} \geq 21)$ and so on. The prediction for this instance is $f(\mathbf{x_e}) = -178 + 7 + 0 = -171$ (see the blue leaves in Fig. 3).

We can observe that tree T_1 is aligned with Table I showcasing the imbalance price mechanism. The root condition of T_1 relates to the sign of the system imbalance ($\text{SI} \geq 0$). The price regime is low if this condition is verified, i.e., reaching the leaves $\{-178, -40\}$ depending on a condition on the MDP attribute, otherwise the price regime is high, i.e., reaching the leaves $\{33, 209\}$ following a condition on the MIP attribute. However, regarding the trees T_2 and T_3, the relation between Table I and their respective conditions is less obvious. For instance, all conditions of tree T_2 relate exclusively to the MIP attribute, with a leaf value that can go up to 1903 €/MWh. This extreme leaf value only covers 4 instances in the training dataset, indicating a possibility of overfitting leading to a significantly wrong model of the imbalance price mechanism.

B. Abductive vs. SHAP explanations

Flaws in the understanding of the prediction model are further analysed in Table II. Table II presents formal and SHAP explanations for two instances of the test set. The first instance, i.e., 1st quarter hour of January 1, 2022, corresponds to a correct prediction of the price regime (Table IIa), while an incorrect prediction of the price regime occurs at the 819^{th} instance, i.e., January 9 at 12:30 pm (Table IIb). For each instance, the left column (denoted by AX) gives us a minimum-size abductive explanation for the instance, while the right column provides the SHAP values of the features for this instance. More specifically, the abductive explanation in Table IIa provides a set of conditions explaining why the predictor infers a low price regime, i.e., why the imbalance premium prediction is negative ($I =]-\infty, 0]$), while the abductive explanation in Table IIb explains why the predictor indicates a high price regime, i.e., why the imbalance premium prediction is positive ($I = [0, +\infty[$). Note that open intervals are denoted by non-closed brackets, whereas closed intervals are indicated by closed brackets.

For the first instance, the inputs are SI_{t_0+1} of 324.5 MW, a MIP_{t_0+1} of 103.5 €/MWh, and a MDP_{t_0+1} of -140 €/MWh, and the target value $\lambda^{\text{reg.}}_{t_0+1}$ is -143 €/MWh. The system imbalance being positive, the imbalance price is driven by the MDP with a negative adder of €3/MWh applied by the Belgian TSO. The boosted model with 3 trees infers a prediction $\hat{\lambda}^{\text{reg.}}$ of -171 €/MWh, obtained by summing the leaves labeled by -178, 7 and 0. The abductive explanation (AX) pinpoints two Boolean conditions for explaining why $\hat{\lambda}^{\text{reg.}} \in]-\infty, 0]$: $\text{SI}_{t_0+1} \geq 0$ and $\text{MIP}_{t_0+1} < 336$. Based on this explanation, a user with some expertise may already question the inner working of the model on this instance, as the prediction of a low price regime is conditioned by the value of the MIP attribute. Indeed, Table I shows us that a low imbalance price regime only depends of the sign of the system imbalance. Such a reasoning is not possible with SHAP values,

[1]Note that the MIP and MDP have also been differentiated w.r.t. the day-ahead market price (λ^{DA}).

TABLE II: Formal explanations and SHAP values for the prediction made on January 1, 2022 at 12:00 am (Table IIa) and January 9 at 12:30 pm (Table IIb).

(a) **1st instance:** $\lambda^{\text{reg.}}_{t_0+1}$ = -143 and $\hat{\lambda}^{\text{reg.}}_{t_0+1}$ = -171 [€/MWh].

Inputs	AX	SHAP [€/MWh]
SI_{t_0+1} = 324.5 MW	SI ≥ 0	-86
MIP_{t_0+1} = 103.5 €/MWh	MIP < 336	3
MDP_{t_0+1} = -140 €/MWh	-	-87

(b) **819th instance:** $\lambda^{\text{reg.}}_{t_0+1}$ = -204 and $\hat{\lambda}^{\text{reg.}}_{t_0+1}$ = 120 [€/MWh].

Inputs	AX	SHAP [€/MWh]
SI_{t_0+1} = 58 MW	-	-133
MIP_{t_0+1} = 484 €/MWh	MIP ≥ 336	340
MDP_{t_0+1} = -204 €/MWh	-	-87

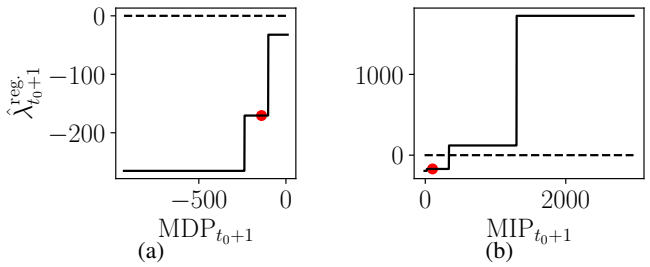


Fig. 4: Model's response to variations in MDP (see Fig. 4a) and MIP (see Fig. 4b), highlighting the current value of the prediction with a red point

TABLE III: Error metrics of a linear and boosted models for the academic case study.

	RMSE [€/MWh]	MAE [€/MWh]
Linear model	71	54
Regression tree	86	73
Boosted model with 3 trees	50	40
Boosted model with 100 trees	16	11

which mainly highlight the SI and MDP attributes as important drivers. This explanation blinds the user about the importance of the MIP value even when inferring a low imbalance price regime. Indeed, if we fix only the values of SI and MDP, the possible prediction outcomes can still range from -194 to 1725 €/MWh. By not providing mathematical guarantees about a prediction interval, SHAP values cannot be used here to detect a faulty model. This is visually demonstrated in Fig. 4, which depicts the model's response to variations in MDP or MIP.

The 819th instance confirms that the prediction model has a flaw in its modeling. The observed price regime is -204 €/MWh, and the prediction is 120 €/MWh (obtained with the leave values -178, 298 and 0). This prediction is driven by the inputs $\text{SI}_{t_0+1} = 58$ MW, $\text{MIP}_{t_0+1} = 484$ €/MWh, and $\text{MDP}_{t_0+1} = -204$ €/MWh. The abductive explanation that has been computed about this erroneous price regime is given by the Boolean condition MIP ≥ 336, which is complementary to the abductive explanations derived for instance 1. From this observation, we can deduce that, each time the MIP input is greater than 336 €/MWh, the predictor infers a high price regime, regardless of the system imbalance value. For this instance, SHAP correctly pinpoints the MIP input as the main driver of the prediction, but it is unable to derive the associated threshold value that triggers the imbalance price regime.

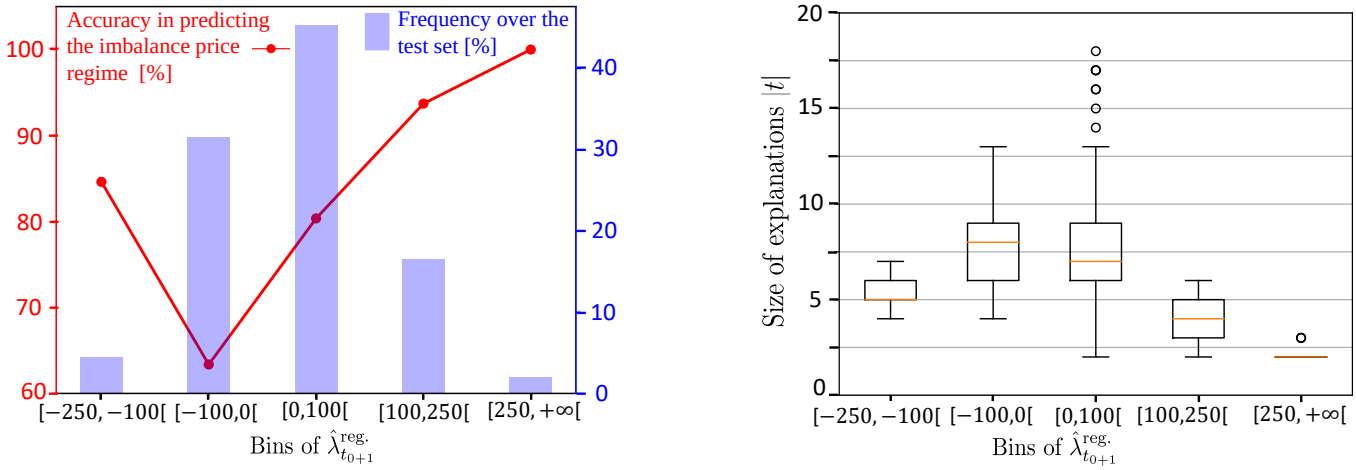
This wrong understanding of the boosted model with three trees is also corroborated by Table III, showing the root mean square error (RMSE) and the mean absolute error (MAE) over the test set, i.e., from January 1 to 10, 2022. Such results are compared with a linear model, and a boosted model with 100 trees. First, we may observe the benefits of adding a large number of trees for refining the prediction, where the boosted model with 100 trees has a reduction of 58% and 72.5% of RMSE and MAE compared to the boosted model with 3 trees, respectively. Furthermore, we may also observe that the boosted model with 3 trees still exhibits a decrease in the RMSE and MAE metrics of 30% and 26% compared to the linear model, respectively. This shows the difficulty of linear models in capturing the regime switching behavior of imbalance prices, even with perfect knowledge of the real-time market conditions. Note that an error signal still exists even with a boosted model with 100 trees due to the adders applied on imbalance prices in Belgium, whose information is not provided in the input data.

V. A CASE STUDY

This section shows results for the prediction task on real-life operating conditions. A practical case study is presented, and the benefits offered by abductive explanations in this practical case are highlighted. As in Section IV, the conditional expectation of the imbalance premium is computed, but with other inputs. The set of input features is composed of i) intra quarter hour dynamics of the system imbalance, characterized by averaged values over the last previous minutes ($\text{SI}_{1'}$, $\text{SI}_{5'}$, $\text{SI}_{15'}$, where minutes are denoted in subscript), ii) past dynamics of the imbalance premium ($\lambda^{\text{reg.}}_{t_0}$, $\lambda^{\text{reg.}}_{t_0-1}$, $\lambda^{\text{reg.}}_{t_0-2}$, $\lambda^{\text{reg.}}_{t_0-3}$), and iii) calendar indicators (the hour and quarter hour, denoted H and Qh respectively). Fulfilling its role of market facilitator, the Belgian TSO provides each minute updates about the real-time market operations via its website platform, hence no initial lag needs to be introduced. The training, validation, test set and intervals (i.e., $I =]-\infty, 0]$ or $I = [0, +\infty[$) used are the same as in Section IV. Section V-A shows that the boosted tree model used achieves top accuracy compared to other well-established forecasting techniques. Section V-B presents a global overview of the abductive explanations, while Section V-C focuses specifically on some instances.

A. Performance of the boosted tree model

We first compare the performance of our boosted model with well-established forecasting techniques. Our empirical protocol was as follows. Five different forecasters were initially compared: i) a linear model, ii) a regression tree, iii) a random forest, iv) a feed-forward neural network and v) a



(a) (b)

Fig. 5: Accuracy in predicting the imbalance price regime and histogram of the predicted imbalance premium (Fig. 5a) and size of the explanations per different bins of predicted imbalance premium (Fig. 5b).

TABLE IV: Error metrics of different predictors for the practical case.

	RMSE [€/MWh]	MAE [€/MWh]
Linear model	98	70
Regression tree	109	80
Random forest	93	69
Neural network	93	69
Boosted tree	93	68

boosted tree. We conducted a hyperparameter optimization to identify the most suited model complexity of each forecaster on the 10% validation set. This was achieved through a grid search, where the same number of iterations is used across all benchmarks. The regression tree has a depth of 12. The random forest has a depth of 12 with 100 trees. XGBoost has 100 estimators, a learning rate of 0.1, and a max depth of 4. The neural network has two hidden layers with 20 and 12 neurons. The neural network has been optimized via the Adam algorithm.

In Table IV, we report the error metrics RMSE and MAE for the different prediction models. We can observe that XGBoost, random forest and neural network have obtained the same error metrics. We can note that the ML models have a RMSE reduced by 5% compared to a linear model. We can observe also the benefits of using ensemble methods based on trees as demonstrated by the reduction of 15% in terms of RMSE compared to the single regression tree.

B. Global overview of the explanations generated

Concerning scalability concerns, the computational time for extracting abductive explanations using the algorithm described in III-B is on average 13s, with only 22 instances above 60s. Note that, for those 22 instances, the algorithm can be stopped, and still outputs abductive explanations (but without subset-minimality guarantee). This is aligned with the experiments in [17], which shows that the computational time across different datasets is manageable (not exceeding 90 seconds on average) up to a substantial number of Boolean conditions (up to 800). Figure 5 shows two pictures. The leftmost picture (Fig. 5a) depicts the predicted imbalance premium categorized in different bins. The left y-axis (in red) shows the accuracy of predicting a low ($\hat{\lambda}^{\text{reg.}}_{t_0+1} < 0$) or high ($\hat{\lambda}^{\text{reg.}}_{t_0+1} > 0$) imbalance price regime for the different categories. The right y-axis (in blue) shows the proportion of each bin over the whole test set. We can observe that the accuracy of predicting the correct imbalance price regime is about 63% and 80% for the intervals $[-100, 0[$ and $[0, 100[$, respectively. Both classes represents 77% of the test set. Interestingly, the model is more efficient for predicting a positive imbalance price regime rather than a negative imbalance price regime. In the same vein, we can see also that the model accuracy increases when the model predicts extreme values of the imbalance price regime. The corresponding rightmost picture (Fig. 5b) shows the size of the abductive explanations per different categories. The size of explanations is computed based on the number of Boolean conditions they contain. We can observe that for predictions ranging into $[-100, 0[$ (resp. $[0, 100[$), explanations are quite large: on average, they consist of 8 (resp. 7) conditions with a maximum that could go towards 18 conditions (using thus all 9 features) for some instances. Interestingly, we can observe that the explanations that have been derived contain more Boolean conditions regarding low imbalance price regime than high imbalance price regime. Furthermore, we can observe that for "more extreme" predictions, i.e., between $[-250, -100[$, $[100, 250[$ and $[250, +\infty[$, the size of explanations are reduced. This complies with our expectations as the corresponding predictions are further from the baseline 0 than the ones considered with the other intervals. This is even more pronounced for the interval $[250, +\infty[$, where only two conditions are sufficient for predicting a high price regime.

TABLE V: Frequency of appearance of attributes $[\%]$ in the explanations for the different categories of predictions made.

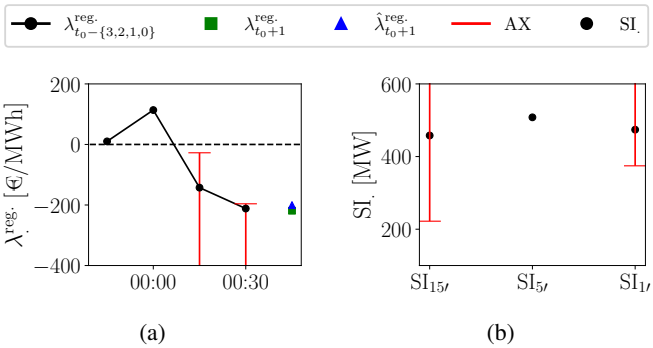
	[-250,-100[	[-100,0[	[0,100[	[100,250[	[250,+∞[
$SI_{1'}$	100	100	100	100	100
$SI_{5'}$	28	82	84	37	0
$SI_{15'}$	85	90	42	13	0
$\lambda^{\text{reg.}}_{t_0}$	100	100	100	99	100
$\lambda^{\text{reg.}}_{t_0-1}$	100	100	59	6	5
$\lambda^{\text{reg.}}_{t_0-2}$	15	76	83	55	5
$\lambda^{\text{reg.}}_{t_0-3}$	5	53	85	29	5
Qh	64	62	54	23	0
H	0	28	54	5	0

In Table V, the frequency of appearance of each attribute within the abductive explanations is shown over the whole test set. First, the system imbalance over the last previous minute, and the imbalance premium over the last quarter hour are almost always present in the abductive explanations. In relation with Fig. 5b, we can also see that many features are present in explanations for the prediction intervals $[-100, 0[$ and $[0, 100[$. More specifically, we can observe that the imbalance prices at t_{0-2} and t_{0-3} are frequently used for predicting the imbalance price regime. In addition, the dynamics of the system imbalance over 5 and 15 minutes is used as well. For low imbalance price regime, the imbalance premium at $t_0 - 1$ is used as well as the system imbalance over the last 15 min. When the prediction lies in $[250, +\infty[$, $SI_{1'}$ and $\lambda^{\text{reg.}}_{t_0}$ drive the prediction of a high price regime. We can also observe that for the interval $[-250, -100[$ the four main features are $SI_{1'}$, $SI_{15'}$, $\lambda^{\text{reg.}}_{t_0}$ and $\lambda^{\text{reg.}}_{t_0-1}$. Both calendar information have an impact on the prediction of the price regime, when the prediction is not far from 0.

C. A focus on three instances

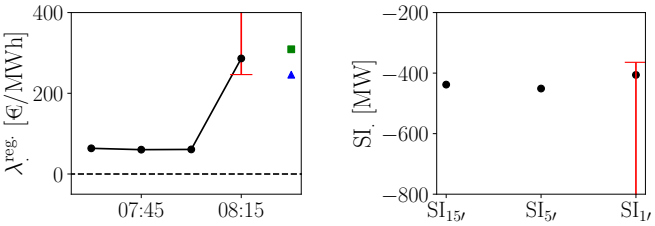
Fig. 6 illustrates the sufficient conditions leading to a negative imbalance premium for instance 3, predicting an imbalance premium of -200 €/MWh for January 1 at 12:30 am, compared to the observed value of -219 €/MWh. More particularly, as depicted in Fig. 6a, the conditions on the past observed imbalance premium are $\lambda^{\text{reg.}}_{t_0} < -196$ €/MWh and $\lambda^{\text{reg.}}_{t_0-1} < -28$ €/MWh, while, in Fig. 6b, past system imbalance dynamics are required to be $SI_{15'} \geq 222$ MW and $SI_{1'} \geq 374$ MW. In contrast, instance 34, on January 1 at 8:15 am, corresponds to a prediction of 246 €/MWh for an observed imbalance premium value of 309 €/MWh. For this instance, the two sufficient conditions for the predictor to be in the positive interval are that $\lambda^{\text{reg.}}_{t_0} \geq 246$ €/MWh and $SI_{1'} < -364$ MW (as shown respectively in Fig. 7a and Fig. 7b). These examples demonstrate how extreme values in specific input attributes drive the imbalance price regime.

One of the difficulties for the predictor is to make a change in price regime compared to the past quarter of an hour. This is what is happening at instance 225 corresponding to 01/03/22 at 8:00 am, which is showcased in Fig. 8. The imbalance premium $\lambda^{\text{reg.}}_{t_0+1}$ is -121 €/MWh, while the imbalance premium observed over the past quarter hour ($\lambda^{\text{reg.}}_{t_0}$)



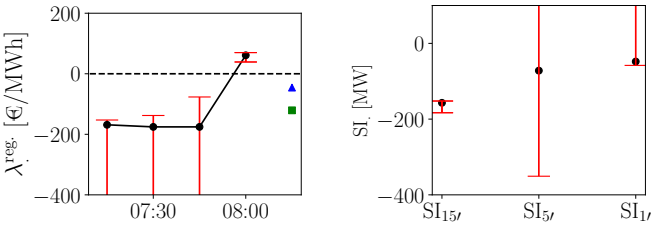
(a) (b)

Fig. 6: Visualization of explanations for instance 3, on January 1 at 12:30 pm



(a) (b)

Fig. 7: Visualization of explanations for instance 34, on January 1 at 8:15 am



(a) (b)

Fig. 8: Visualization of explanations for instance 225, on January 3 at 08:00 am

is 61 €/MWh. The value of the prediction is -45 €/MWh, showing a regime switching behavior of the boosted model. To predict this regime switching, an abductive explanation must use 10 Boolean conditions covering 8 inputs. Concerning the dynamics of the regime prices (shown in Fig. 8a), we can observe that the imbalance premium of the past quarter of an hour must not be too high, i.e., $\lambda^{\text{reg.}}_{t_0} \in [39, 70]$. Concerning the imbalance of the system (detailed in Fig. 8b), the average value over the 15 min should not be too low, i.e., between -152 and -350 MW. In addition, two conditions appear about the average imbalance over the last 5 minutes and the last minute. They indicate that the values of those features should be respectively greater than -351 MW and -58 MW. In addition, we see that it is important for the predictor that the price dynamics from the past quarters 1 to 3 is lower than -76, -138, and -153 €/MWh. Finally, the role of the calendar entry, indicating whether it is the first quarter hour, is also important. This may be due to the ramp phenomena which occur due to the hourly granularity of the day-ahead market and can therefore change the price

prediction.

VI. Conclusions and Perspectives

Imbalance price prediction is a risky task. Though ML models are the best approaches that have been provided so far for this task, significantly wrong predictions may happen. Accordingly, it is important to develop explanation methods about the predictions made for enabling users to gain trust in the predictions made when the predictions cohere with what the user expects and to reject them otherwise. In this perspective, this work focuses on formal methods for deriving abductive explanations and understanding why the model used predicts a high or low imbalance price regime. The prediction is made every quarter hour and the method of generation is computed below 13s on average.

Our approach to abductive explanations presents two key advantages than are not shared by model-agnostic approaches (as SHAP): the user can choose the precision of the prediction he/she is ready to tolerate, and the explanations are provably correct, allowing the user to reason from them for making the final decision (believing or not in the prediction made). These two advantages were exemplified through an academic case study. In particular, abductive explanations clearly identifies the input conditions for which the boosted model infers a wrong imbalance price regime. Such tailored reasoning is not possible from SHAP explanations, which falsely reinforce the explainee's beliefs about the correct behavior of the model for instance 1. Then, the applicability and usefulness of abductive explanations is assessed on a real-life case study. For this case study, ML models (i.e., the boosted tree, random forest and neural network models) achieve similar performance and show a decrease of 5% in RMSE compared to a linear model. As expected, the size of abductive explanations decreases on average as the prediction value is further from the baseline 0. Features that appear the most frequently in the explanations are the system imbalance over the last minute, and the past quarter hour imbalance premium. Both features are sufficient for explaining an extremely high price regime, but need to be complemented with other system imbalance dynamics and imbalance premiums for prediction near 0. Finally, a deeper investigation about one instance exemplifies the usefulness of such explanations in understanding the input conditions that trigger the model to predict a price regime switching from an observed positive imbalance premium. For this instance, the derived abductive explanation (using 8 features out 9 available) highlights that the previously observed imbalance premium should be not too high while others past imbalance premiums must be negative, an increasing trend in the intra quarter hour dynamics of the system imbalance, and a shift to the first quarter of the next hour.

This work calls for many perspectives. The most important one surely is to integrate in a deeper way humans in the prediction/explanation loop so as to create interactive systems that are usable, useful and enjoyable to use. Explanations as viewed by AI researchers and by Human-Computer Interaction researchers are quite different. Both communities had a long way to go for bridging the gap. Developing a protocol of interaction with actual BRPs would be a step toward this goal, making the explanation process more dynamic and user-centric.

References

[1] J. Cartuyvels and A. Papavasiliou, "Calibration of operating reserve demand curves using a system operation simulator," *IEEE Trans. Power Syst.*, vol. 38, no. 4, pp. 3043–3055, 2023.

[2] G. Klaeboe, A. L. Eriksrud, and S.-E. Fleten, "Benchmarking time series based forecasting models for electricity balancing market prices," *Energy Syst.*, vol. 6, no. 1, pp. 43–61, 2015.

[3] J. Browell and C. Gilbert, "Predicting Electricity Imbalance Prices and Volumes: Capabilities and Opportunities," *Energies*, vol. 15, no. 10, pp. 36–45, 2020.

[4] D. W. Bunn, J. N. Inekwe, and D. MacGeehan, "Analysis of the fundamental predictability of prices in the british balancing market," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1309–1316, 2021.

[5] A. Lucas, K. Pegios, E. Kotsakis, and D. Clarke, "Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression," *Energies*, vol. 13, no. 20, pp. 1–16, 2020.

[6] S. Deng, J. Nkwoma, V. Smirnov, A. Wait, and C. Wang, "Machine Learning and Deep Learning Forecasts of Electricity Imbalance Prices," *SSRN Electronic Journal*, 2023.

[7] J. Stiasny, S. Chevalier, R. Nellikkath, B. Sævarsson, and S. Chatzivasileiadis, "Closing the loop: A framework for trustworthy machine learning in power systems," in *Proceedings of IREP 2022*, 2022.

[8] R. Machlev, L. Heistrene, M. Perl, K. Levy, J. Belikov, S. Mannor, and Y. Levron, "Explainable artificial intelligence (xai) techniques for energy and power systems: Review, challenges and opportunities," *Energy AI*, vol. 9, p. 100169, 2022.

[9] J. Bottieau, Y. Wang, Z. De Grève, F. Vallée, and J.-F. Toubeau, "Interpretable transformer model for capturing regime switching effects of real-time electricity prices," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2162–2176, 2023.

[10] V. N. Ganesh and D. Bunn, "Forecasting imbalance price densities with statistical methods and neural networks," *IEEE Trans. Energy Markets, Policy and Regulation*, pp. 1–10, 2023.

[11] S. Jain and B. C. Wallace, "Attention is not explanation," in *Proceedings of the NAACL-HLT 2019*, 2019, p. 3543–3556.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of NeurIPs 2017*, 2017, p. 4768–4777.

[13] J. Marques-Silva and X. Huang, "Explainability is not a game," 2023, arXiv preprint, 2307.07514.

[14] S. Letzgus, P. Wagner, J. Lederer, W. Samek, K.-R. Müller, and G. Montavon, "Toward explainable artificial intelligence for regression models: A methodological perspective," *IEEE Signal Process. Mag.*, vol. 39, no. 4, pp. 40–58, 2022.

[15] N. Narodytska, S. Kasiviswanathan, L. Ryzhyk, M. Sagiv, and T. Walsh, "Verifying properties of binarized deep neural networks," in *Proceedings of AAAI 2018*, 2018, pp. 6615–6624.

[16] A. Ignatiev, N. Narodytska, and J. Marques-Silva, "Abduction-based explanations for machine learning models," in *Proceedings of AAAI 2019*, 2019, pp. 1511—-1519.

[17] G. Audemard, S. Bellart, J. Lagniez, and P. Marquis, "Computing abductive explanations for boosted regression trees," in *Proceedings of IJCAI 2023*, 2023, pp. 3432–3441.

[18] L. M. Lima, P. Damien, and D. W. Bunn, "Bayesian predictive distributions for imbalance prices with time-varying factor impacts," *IEEE Trans. Power Syst.*, vol. 38, no. 1, pp. 349–357, 2023.

[19] J. Trebbien, L. Rydin Gorjão, A. Praktiknjo, B. Schäfer, and D. Witthaut, "Understanding electricity prices beyond the merit order principle using explainable ai," *Energy AI*, vol. 13, p. 100250, 2023.

[20] G. Audemard, S. Bellart, L. Bounia, F. Koriche, J. Lagniez, and P. Marquis, "Trading complexity for sparsity in random forest explanations," in *Proceedings of AAAI*, 2022, pp. 5461–5469.

[21] CRIL, University of Artois & CNRS,, "Pyxai project," https://www.cril.univ-artois.fr/pyxai/, 2023.