Electricity Price Forecasting based on Order Books: a differentiable optimization approach

Léonard Tschora
Univ Lyon, INSA Lyon, LIRIS, UMR5205
BCM Energy, F69003 Lyon, France

Tias Guns
KU Leuven
Leuven, Belgium

Erwan Pierre
BCM Energy
F69003 Lyon, France

Marc Plantevit
EPITA Research Laboratory (LRE)
FR-94276 Le Kremlin-Bicêtre, France

Céline Robardet
Univ Lyon, INSA Lyon, LIRIS, UMR5205
F-69621 Villeurbanne, France

***Abstract*—We consider day-ahead electricity price forecasting on the European market. In this market, participants can offer electricity for sale or purchase for a specific price by submitting overnight orders. Market operators determine the market clearing price – the price at which the amount of electricity supplied equals the amount of electricity demanded – using the Euphemia balancing algorithm. EUPHEMIA is a quadratic optimization problem that maximizes the social welfare defined as the sum of the supplier surplus and consumer surplus while ensuring a null energy balance. This mechanism deeply influences the price calculation, but has so far been little considered in electricity price forecasting algorithms. Existing models are generally based on identifying relationships between exogenous characteristics (consumption and production forecasts) and the market clearing price to be predicted. A few studies have examined the EUPHEMIA mechanism during prediction, by doing costly manual transformations on order books. In this article, we overcome this limitation by considering the pricing mechanism during model training. For this, we use a predict-and-optimize strategy with differentiable optimization. We design a fully differentiable and scalable solving method for the EUPHEMIA optimization problem and apply it on real-life data from the European Power Exchange (EPEX). We design different model architectures using our differentiable solver and empirically study the impact of taking into account the optimal calculation of prices within the training of the neural network.**

***Index Terms*—Electricity Price Forecasting, decision-focused learning, differentiable optimization, predict-and-optimize.**

I. INTRODUCTION

Electricity Price Forecasting (EPF) is a problem that has interested specialists in the field for many years [1], [2] in order to better adjust electricity production to the needs of consumption for this resource that is difficult to store. This also allows stock market activities, such as making it possible for owners of renewable energy production facilities to make their investments profitable by anticipating price movements and promoting intelligent applications such as self-consumption [3] or battery optimization [4]. These issues have become increasingly crucial in recent months with high price volatility.

EPF is complex as many factors influence it, both at the level of production and consumption. Different methods have been used so far, such as auto-regressive methods [5]–[10], but also augmented machine learning models for the EPF problem [6], [11]–[21]. The two approaches focus on finding relationships between exogenous features (electricity consumption and production forecasts) and price histories with day-ahead prices, and machine learning models have recently proven to be superior to auto-regressive models [22].

However, European market prices are set by the EUPHEMIA algorithm on the basis of order books that contain purchase and sale offers, and not directly on the basis of consumption and production forecasts. EUPHEMIA [23] maximizes social welfare defined as the sum of the difference between the price paid to the supplier and the minimum price stated in the sales order, as well as the difference between the maximum price defined in purchase order and the price actually paid. By solving a mixed-integer quadratic programming optimization problem, EUPHEMIA ensures the highest price for producers, the lowest price for purchasers and a constant energy balance. This pricing mechanism is very specific and thus, it seems relevant to take it into account in the prediction model because it directly influences prices. For this, it is necessary to integrate the optimization problem as formalized in EUPHEMIA into the price prediction problem.

Various works have studied the coupling of a prediction model with an optimization task. In the predict-then-optimize approach [24] a predictive model is first built and then used to optimize decision-making. However, the learning of the model is not guided by the prediction errors on the final task related to the optimization problem. Conversely, the predict-and-optimize framework proposes to learn a predictive model by directly minimizing the error related to the downstream decision-making task [25].

In this paper, we explore the impact of coupling the EUPHEMIA optimization problem, with machine learning on price prediction accuracy. Three models are considered. First, a standard machine learning approach predicts day-ahead prices based on exogenous variables. The loss used minimizes the difference between predicted and real prices. The second model predicts order books by minimizing the difference between predictions and actual order books. The last model predicts order books with a neural network and then solve the EUPHEMIA optimization problem. The resulting prices are then compared to real prices. During training, the derivation of the loss *after optimisation* is used to adjust the parameters of the neural network. Finally, by making these three models share a common neural network, it is possible to combine them using a loss resulting from the linear combination of the three previous losses. We can then evaluate empirically and on different data sets the impact of each model on the accuracy of the predictions. We aim to show that end-to-end predict-and-optimize methods are beneficial for this problem.

Our contributions are summarized as follows:

1) We formalize EUPHEMIA optimization problem, a Mixed-Integer Quadratic Programming problem with linear orders, as the maximization of social welfare measured by the difference between market price and order book prices, for all accepted orders (supply orders with prices inferior to the market price, and demand orders prices superior to the market price).
2) We present the dual optimization problem, that expresses EUPHEMIA as the optimization of market prices. This problem can be solved by setting the derivative to zero using a dichotomy search.
3) We explain how to integrate the optimization problem into the neural model, by deriving the calculations of the backward pass. Several neural networks and their associated loss allow the model to be adjusted according to real prices or order books.
4) Finally, we present experiments performing a deep analysis of seven configurations on four datasets. We provide an extensive discussion of our results in context.

The rest of the paper is structured as follows. In Section II the EUPHEMIA optimization problem is framed based on its definition. Assuming that model predictions can be improved by tightly coupling the problem with order book prediction from exogenous data, a method for solving EUPHEMIA to determine the electricity price from order books is proposed in (Section III). Then, three strategies of combining an optimal decision process while learning the order book prediction model are presented in Section IV). The experimental evaluation (Section V) subsequently compares several configurations of this end-to-end predict-and-optimize model on 4 different European datasets. Afterwards, a qualitative analysis determines what information has been captured by the differentiable optimization problem using Shap values. Finally, Section VII concludes and provides future work.

II. EUPHEMIA OPTIMIZATION PROBLEM

Electricity prices in Europe are set at the continental level. Since the 1990s, the European Union (EU) has gradually opened up national electricity markets to competition in order to harmonize and liberalize the European market, and better interconnect it. The European electricity transmission network now ensures security of supply and exchanges between forty-six European zones. As electricity cannot be stored, the market is regularized to guarantee the balance between supply and demand at European scale, by prioritizing the least expensive means of production. This price harmonization promotes trade between countries. Several market places exist, such as the European stock exchange, EPEX Spot SE, on which megawatt-hours (MWh) are traded, with prices that vary by country depending on supply and demand. It is a speculative place that brings together producers and consumers, who sell and buy electricity (nuclear, renewable or fossil), for immediate or deferred deliveries. A peculiarity in this wholesale market, is that the price is set not according to the average cost of electricity production, but based on the "marginal" production cost of the last (and therefore most expensive) MWh injected into the network. This mechanism is illustrated in Figure 1.

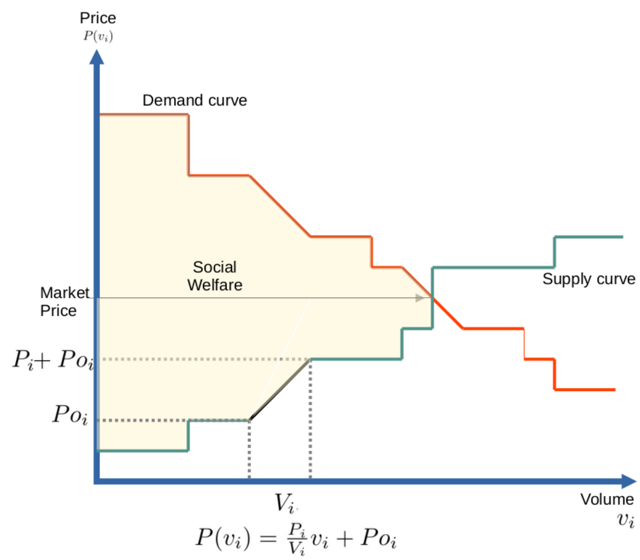

Fig. 1. Supply and demand curves.

Let us consider the supply curve. Each segment of the curve (e.g. see the black segment on the supply curve) corresponds to an order defined by a volume V_i that the seller offers at a price in the interval $[Po_i, Po_i + P_i]$ on the ordinate axis. The sell offers are sorted according to increasing prices. Similarly, the demand curve consists of bids for a given volume at a maximum price made by buyers. Demand offers are sorted in decreasing order of prices. The market price is set at the intersection of the two curves. All supply orders with prices inferior to the market price, and all demand orders superior to the market price are accepted. Power exchange members, whose orders have been accepted, trade electricity at the market price for a specific trading hour. The difference between the market price and an accepted order times the sold volume (the area in yellow) is called the *social welfare* (SW). The set of all demand and offer orders are the ORDER BOOK, denoted OB in the following. Algorithm EUPHEMIA fixes the market price by solving a Mixed-Integer Quadratic Programming problem (MIQP) whose simplified definition, using only linear orders, is given below.

*Definition 1 (*EUPHEMIA *optimization problem):* Let orders be defined by a price range $[Po_i, Po_i + P_i]$ and a volume V_i. Supply orders are defined by positive volumes and prices ($V_i > 0$ and $P_i > 0$), while demand orders are specified by negative values ($V_i < 0$ and $P_i < 0$). EUPHEMIA is defined by the following convex quadratic optimization problem over the variables $A_i \in [0,1]$, which state whether the order i is fully accepted (value 1), fully rejected (value 0), or partially accepted (other values):

$$\max_A f(A) = \max_A \sum_{i \in OB} \left(-\frac{1}{2} A_i^2 V_i P_i - A_i V_i P_{o_i} \right) \tag{1}$$

$$\text{u.c.} \sum_{i \in OB} A_i V_i = 0, \tag{2}$$

$$-A_i \leq 0, \tag{3}$$

$$A_i - 1 \leq 0 \tag{4}$$

Equation (1) is obtained as follows. To maximize the social welfare, the difference between the market price (P^*) and what is demanded as minimum payment by the supplier (P_S) must be maximized. Similarly, the difference between the maximum amount that the consumer is willing to pay (P_D) and what he actually pays (P^*) must be maximized as well. This leads to

$$\max(P^* - P_S) + (P_D - P^*) = \max P_D - P_S$$

If we denote by $v_i = A_i \times V_i$, the proportion of volume actually sold for order i, we have,

$$P(v_i) = \frac{P_i}{V_i} v_i + Po_i$$

Thus, the expected payment P_S is the result of this function applied to each possible volume:

$$P_S(v_i) = \int P(v_i)\, dv_i = \frac{1}{2} \frac{P_i}{V_i} v_i^2 + v_i\, P_{0_1} + \theta$$

with θ independent of v_i. The same result is obtained for $P_D(v_i)$. By considering that volumes V_i of demand orders are negative, and by summing over all the orders i, we obtain that $f(A)$ is proportional to $\sum_i P_D(v_i) - P_S(v_i)$.

The equation (2) is just the constraint on the energy balance to ensure that demand equals supply, and bound the decision variables A_i to belong to interval $[0,1]$.

III. Finding the optimal price from an order book

In order to improve electricity price forecasts, we want to integrate the EUPHEMIA pricing mechanism into the predictive model, as we believe that this should increase the accuracy of the predictions. Market players submit their orders before 12 a.m. each day. EUPHEMIA then calculates Day-Ahead prices that maximize *social welfare*. While EUPHEMIA is reproducible to some extent, the ORDER BOOK is not known until the prices are released. Hence, we propose to predict the ORDER BOOK using fundamental variables and use the predicted orders as input to EUPHEMIA optimization problem to get the price predictions. By doing so, we expect to obtain more accurate predictions. As the goal is to predict the market price and EUPHEMIA model is expressed in volume of electricity exchanged, we first consider the dual problem.

A. Dual EUPHEMIA optimization problem

To determine the dual problem of Definition 1, we introduce $\lambda \in \mathbb{R}$, $M \in \mathbb{R}^N$ and $K \in \mathbb{R}^N$ as the dual variables associated to the constraints in equations (2)-(4) and consider the Lagrangian:

$$\mathcal{L}(A, \lambda, M, K) = \sum_{i \in OB} \left(-\frac{A_i^2 V_i P_i}{2} - A_i V_i Po_i + \lambda A_i V_i - M_i A_i + K_i(Ai - 1) \right)$$

Its minimum is reached when its derivative with respect to A vanishes, that is to say when $A_i = \frac{\lambda - Po_i}{P_i} + \frac{M_i - K_i}{V_i P_i}$. At that point, the Karush–Kuhn–Tucker necessary conditions hold, especially $-M_i A_i = 0$ and $K_i(A_i - 1) = 0$, which occurs when (1) $A_i = 0$, $K_i = 0$ and $M_i \geq 0$, or (2) $A_i = 1$, $K_i \geq 0$ and $M_i = 0$, or (3) $A_i \in]0,1[$, $K_i = 0$ and $M_i = 0$. By re-injecting the expression of A_i in those conditions, we obtain the dual problem of Definition 1 defined by:

$$\min_\lambda D(\lambda) = \min_\lambda \sum_{i \in OB} D_i(\lambda) \text{ with } D_i(\lambda)$$

$$= \begin{cases} (1)\ 0, \text{ if } V_i(Po_i - \lambda) > 0 \\ (2)\ V_i(\lambda - \frac{P_i}{2} - Po_i), \text{ if } V_i(\lambda - P_i - Po_i)) > 0 \\ (3)\ \frac{V_i}{2P_i}(\lambda - Po_i)^2, \text{ if } \lambda \in [Po_i, Po_i + P_i] \end{cases}$$

The optimal dual variable, $\lambda^\star$, is the day-ahead price. Then, (1) corresponds to the situation of a fully rejected order where the optimal price λ is lower than Po_i for the supply orders ($V_i > 0$) and higher for the demand orders ($V_i < 0$). Inversely, (2) corresponds to fully accepted orders. (3) happens when the order is partially accepted (λ is in the price range and the proportion $\frac{\lambda - Po_i}{P_i}$ of volume V_i is exchanged).

B. Computing the optimal price $\lambda^\star$

The minimum of $D(\lambda)$ can be obtained by looking for the values for which the derivative of D vanishes: $\frac{\partial D}{\partial \lambda} = 0$. Each segment of the piecewise function D_i is differentiable or equals 0. Only the inflection points Po_i and $Po_i + P_i$ have to be examined by considering their limits.

Let us consider supply orders (similar results can be obtained for demand orders). Using the generic expression

$$\frac{D_i(\lambda + h) - D_i(\lambda)}{h} = \frac{2V_i}{P_i} \left(\frac{h}{2} + \lambda - Po_i \right)$$

we can compute the limits for the inflection points with $h \mapsto 0^-, h \mapsto 0^+$. We find that the limit at $\lambda = Po_i$ is 0 for both sides of h, and the limit at $\lambda = Po_i + P_i$ is V_i. Hence, D_i is differentiable for all values of λ and the derivative of D is

$$D_i'(\lambda) = \begin{cases} 0 & \text{if } V_i(Po_i - \lambda) > 0 \\ V_i & \text{if } V_i(\lambda - P_i - Po_i) > 0 \\ \frac{V_i}{P_i}(\lambda - Po_i) & \text{if } \lambda \in [Po_i, Po_i + P_i] \end{cases}$$

We can rewrite D' with the Heaviside function $H(x)$[1]:

$$D'(\lambda) = \sum \frac{x_i}{P_i} H(x_i) - \sum \frac{y_i}{P_i} H(y_i)$$

[1] $H(x) = 0$ if $x < 0$ and 1 otherwise.

with $x_i = V_i(\lambda - Po_i)$ and $y_i = V_i(\lambda - Po_i - P_i)$. This function is strictly increasing and we can use a dichotomy search to solve $D'(\lambda^\star) = 0$. Using lb, ub as the lower and upper bounds initialized at the extreme market prices fixed by EPEX, $\lambda^\star$ is computed with Algorithm 1.

Algorithm 1 Dichotomy search.

lb ← -500€/MWh
ub ← 3000€/MWh
found ←**False**
while (found = **False**) and $(ub - lb > 2 * 0.01)$ **do**
  $M \leftarrow \frac{ub+lb}{2}$
  $D_M \leftarrow D'(M)$
  found ← $D_M = 0$
  $ub \leftarrow ub - H(D_M) * (ub - M)$
  $lb \leftarrow M - H(D_M) * (M - lb)$
end while

Using heaviside function allows us to differentiate through IF statements. Analytically, it derives as the Dirac function, but numerically we have to replace it by the sigmoid to make the whole process fully differentiable as required by neural network prediction models.

IV. A DIFFERENTIABLE OPTIMIZATION APPROACH FOR EPF

We present in this section how to integrate the solution of the optimization problem into the price prediction neural network model. First, we present how the difference between the optimal price calculated by Algorithm 1 and the real price is back-propagated to adjust the parameters of the neural network.

A. Integrating the optimization process into the forward and backward passes

Figure 2 presents the model architecture. Exogenous variables X are provided to a neural network that is trained to predict ORDER BOOK. The order book is then used to determine the optimal price $\lambda^\star$ using Algorithm 1. A loss function L, between $\widehat{Y}$, the optimal price based on estimated ORDER BOOK $\widehat{OB}$, and the real price Y, evaluates the error made on the predictions. This error has to be back-propagated on the network that learns $\widehat{OB}$ from exogenous features X:

$$\frac{\partial L(\widehat{Y}, Y)}{\partial X} = \frac{\partial L}{\partial \widehat{Y}} \times \frac{\partial \widehat{Y}}{\partial \widehat{OB}} \times \frac{\partial \widehat{OB}}{\partial X}$$

In this expression, the first term is the gradient of the loss, and the third term is the standard back-propagation. The second term, $\frac{\partial \widehat{Y}}{\partial \widehat{OB}}$ is obtained by differentiating the dual problem formulation. Therefore, Algorithm 1 is implemented using PyTorch [2] and used to solve EUPHEMIA during the forward pass. It is also used to compute the derivative of the order books with respect to the optimal price $\frac{\partial \widehat{Y}}{\partial \widehat{OB}}$ during the backward pass.

[2]https://pytorch.org/docs/stable/index.html

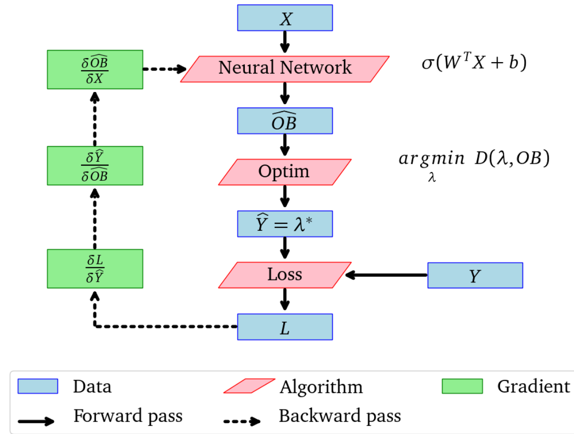

Fig. 2. Differentiable optimization: Predict the order book variables $\widehat{OB}$, find optimal prices given the order book, and back-propagate the errors on the prediction model.

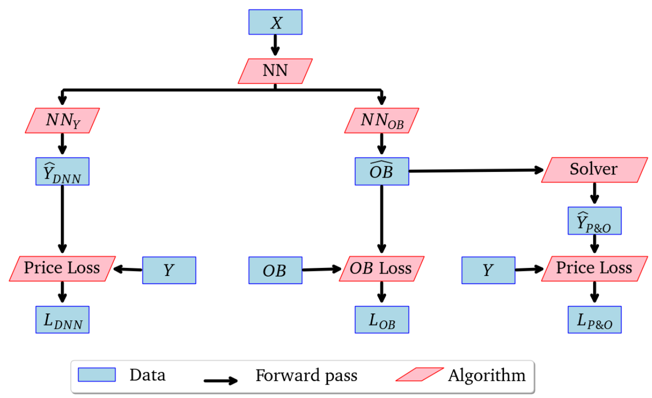

Fig. 3. Three models architecture combined in one. The α branch directly forecasts the prices from the exogenous variables. The γ branch forecasts the OB and the β branch solves the EUPHEMIA problem. The final loss is $L_f = \alpha L_{DNN} + \gamma L_{OB} + \beta L_{P\&O}$.

B. Varying the impact of optimization problem on model learning

To be able to vary and evaluate the impact of the integration of the optimization process within the predictive model, we propose to consider three scenarios (see Figure 3). The first one is a traditional Deep Neural Network (DNN) used to solve the EPF problem. Without considering order books, the DNN directly predicts prices form exogenous variables and computes the loss L_{DNN} with respect to the real prices. The second approach is a DNN that uses the exogenous variables to predict *order books*. The loss function L_{OB} evaluates the difference between the predicted order books and the true ones. The last scenario is the predict-and-optimize approach that uses order books predictions to solve EUPHEMIA, then computes the loss $L_{P\&O}$ between solved prices and real prices.

As these three models share a common neural network NN, it is possible to combine them. We propose to evaluate different linear combinations of the losses, and hence the networks, in a multi-task setting. We define the general loss as a linear combination of the three losses: $L_f = \alpha L_{DNN} + \gamma L_{OB} + \beta L_{P\&O}$. Note that we chose a relative loss for L_{DNN}, L_{OB} and $L_{P\&O}$ so that they produces values in the same range. The final price predictions are computed over the price prediction $\widehat{Y}_{DNN}$ as well the optimized price $\widehat{Y}_{P\&O}$ over predicted order books:

$$\widehat{Y} = \frac{\alpha \widehat{Y}_{DNN} + \beta \widehat{Y}_{P\&O}}{\alpha + \beta}$$

V. Experiments

In this section we describe a series of experiments with multiple objectives. First, we want to compare our ORDER BOOK model and differentiable EUPHEMIA method to the baseline (direct neural network) and determine if they improve the performances. We then analyze more precisely the impact of parameter β to understand its links with the model performances. Lastly, we seek to link the effects of differentiable optimization to the input variables. For this, we perform a contribution analysis using Shap values to identify which features have been put forward by adding differentiable optimization.

A. Datasets

We consider the EPF problem on the European market where the data is available free of charge[3]. We forecast the prices of 4 countries: France (FR), Germany (DE), Belgium (BE) and the Netherlands (NL). As predictive variables, we use the consumption forecasts, the generation forecasts, the renewable generation forecasts and the current prices of nine European countries: France, Germany, Belgium, the Netherlands, Austria, Italy, Spain, Switzerland and England. To those variables, we add the reference gas price, as well as the date indicators (day, day of week, week, month) that are circularly encoded. Given a cyclic feature x in the domain $\mathcal{C}$ (with cardinality α), this function f returns two numeric values:

$$f : \mathcal{C} \mapsto \mathbb{R}^2$$
$$x \mapsto (\sin(\frac{2\pi x}{\alpha}), \cos(\frac{2\pi x}{\alpha}))$$

Hence, each day is described by $9 + 36 \times 24$ predictive features and the targets to be predicted are the 24 hourly prices for each country. The variables can be grouped into families: domestic variables (variable of the country being predicted), foreign variables (variable of another country), gas price and date. The Swiss and English prices are also considered separately because they are available at 11am and can be used in the training set. Our dataset spans from 01/01/2016 to 31/12/2019. We use the last year (2019) as test set, to account for the prices seasonality.

Additionally, it is possible to include order books from the previous day as predictive features. They are made available by the Epex exchange against a fee. Due to their high dimension (hundreds of orders per hour) and variable size, we first need to define a procedure to map the order books onto a smaller fixed-size representation, while keeping the same day-ahead price and matched volume. For a desired dimension of size $n = n_S + n_D + 4$, we select the n_S supply orders and n_D demand orders closest to the intersection price. The orders not selected are summarized by 4 fictitious orders, for the supply and demand orders that are far before or far after the intersection. These fictitious orders are defined by the following values:

Variable	Supply	Demand
V	$\sum_i V_i$	$\sum_i V_i$
Po	$\min_i Po_i$	$\max_i Po_i$
P	$\min_i (Po_i + P_i) - Po$	$\max_i (Po_i + P_i) - Po$

Graphically, this replaces the step curve of the portion of unselected orders by a line, as displayed in the left-hand side of Figure 4.

B. Models' implementation

We elaborate on the models introduced in Figure 3. As losses L_{DNN}, $L_{P\&O}$, L_{OB}, we use the Symmetric Mean Absolute Percentage Error $SMAPE(Y, \widehat{Y}) = \frac{200}{n} \sum_i \frac{|Y_i - \widehat{Y}_i|}{|Y_i| + |\widehat{Y}_i|}\%$ that accounts for the market's volatility. The NN part is a dense layer with 873 inputs and 888 outputs, followed by batch normalization, dropout and ReLU activation layers. The NN_Y part is a dense layer with 888 inputs and 24 outpus. The NN_{OB} part is detailed on Figure 5. This network receives as input X the output of the shared neural network NN. The input data is first reshaped to hourly granularity, and then sent through a dense layer with 37 inputs and 37 outputs, followed by batch normalization, dropout and ReLU activation layers. This is followed by 6 distinct dense layers with 37 inputs and 20 outputs that forecast the components of the supply and demand sides.

We establish two baseline models and four models to test, by setting α, β or γ to 0. We define these models in Table I where the first two are common DNN models from the literature (see [1]) since they do not use the differentiable EUPHEMIA solver ($\beta = 0$). The last 4 models are the novel models ($\beta > 0$).

C. Results

a) Configurations: We compare the values between the predicted $\hat{Y}$ and the real Y target variable. We use standard measures as $MAE(Y, \hat{Y})$ (the average of the absolute difference between the values over the target variables), $DAE(Y, \hat{Y})$, that first computes the average price of a given day and then compares it to the true average price, $RMAE(Y, \hat{Y})$ that compares the MAE of the predictions with the MAE of a naive forecaster (that forecasts the last day's prices), $SMAPE(Y, \hat{Y})$ (the symmetric mean absolute percentage error over the target variables). To check the statistical significance of the results, we use the Diebold & Mariano (DM) test [26] that compares two models M_1 and M_2. The null hypothesis H_0 is that $MAE(M_1) > MAE(M_2)$, i.e. the

[3]https://transparency.entsoe.eu

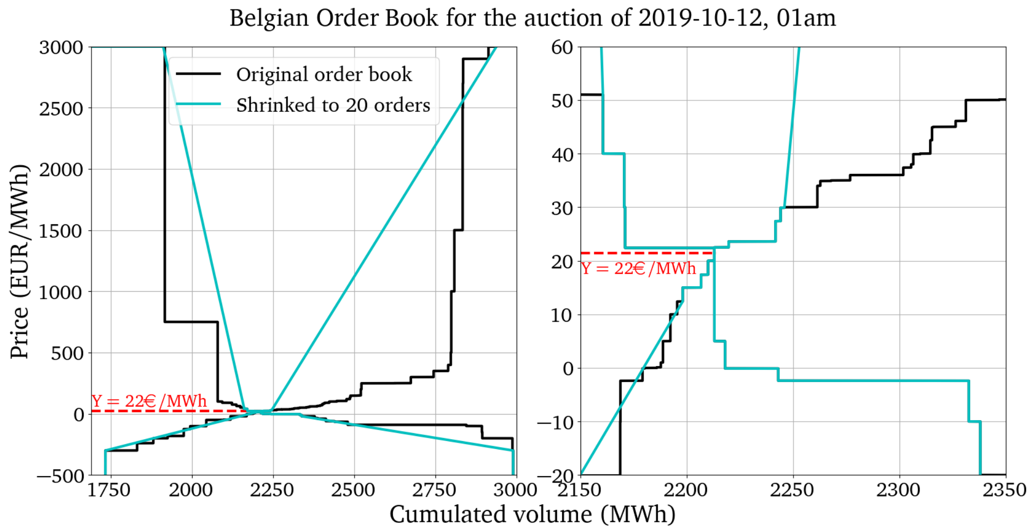

Fig. 4. The order book of the 2019/10/12, 1am auction in Belgium (black), and its reduced version to 20 orders (blue). Around the intersection, the real and reduced order books are the same (right). Orders far from the intersection are replaced by a straight line. Both real and reduced order books cover the same volume range and solve to the same price (left).

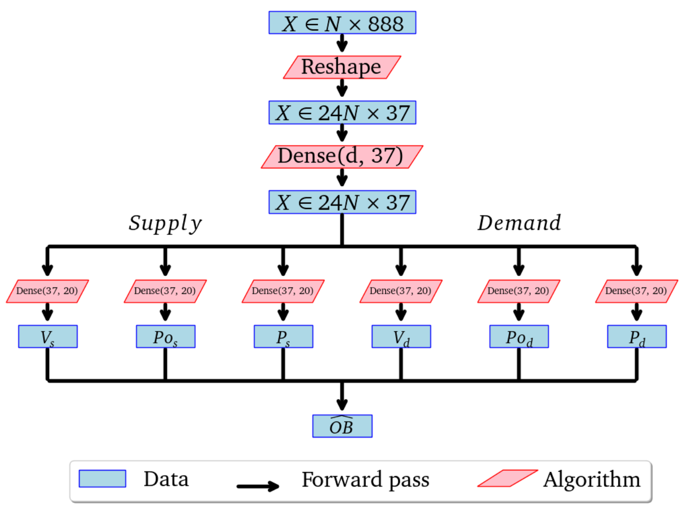

Fig. 5. Implementation details of the NN_{OB} network. The input X, coming from the common NN part, is reshaped to hourly granularity and goes through a dense layer. Then, it is fed to 6 different linear layers to forecast supply and demand components of the order book.

Model	α	γ	β	Description
DNN_Y	1	0	0	Predict $\widehat{Y}$ using a standard DNN.
$\text{DNN}_{Y,OB}$	$\frac{1}{2}$	$\frac{1}{2}$	0	Predict $\widehat{OB}$ and $\widehat{Y}$ without differentiable optimization.
P&O	0	0	1	Pure predict-and-optimize model: there is no loss on $\widehat{OB}$, and $\widehat{Y}$ is only obtained through the solving the optimization problem.
$\text{P\&O} + \text{DNN}_{OB}$	0	$\frac{1}{2}$	$\frac{1}{2}$	A loss is applied to $\widehat{OB}$, but $\widehat{Y}$ is only obtained through solving.
$\text{P\&O} + \text{DNN}_Y$	$\frac{1}{2}$	0	$\frac{1}{2}$	No $\widehat{OB}$ forecast loss. $\widehat{Y}$ is obtained directly, but also by solving the optimization problem.
$\text{P\&O} + \text{DNN}_{Y,OB}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	The three models are combined.

TABLE I

THE DIFFERENT MODELS DERIVED FROM OUR ARCHITECTURE. MODELS DNN_Y AND $\text{DNN}_{Y,OB}$ ARE THE BASELINES THAT DO NOT USE THE DIFFERENTIABLE OPTIMIZATION.

first model is less efficient than the second. We can reject H_0 and conclude that M_1 outperforms M_2 if the resulting p-value is lower than a fixed threshold of 0.05. The source code and the free data are made available.[4]

The metrics computed on the test period are displayed in Table II. The p-values of the Diebold & Mariano test are displayed in Figure 6. Our observations are the following:

- On the French market, dominated by nuclear energy production whose marginal cost is independent of other variables, the order books are more difficult to predict and less significant. Consequently, the metrics of the baseline models DNN_Y and $\text{DNN}_{Y,OB}$ are significantly better : the order book modeling plus differentiable optimization models fail to improve performances.
- On the German, Dutch and Belgian datasets, models $\text{P\&O} + \text{DNN}_Y$ and $\text{P\&O} + \text{DNN}_{Y,OB}$ outperform other models with statistical significance. These markets are characterized by an important gas or coal-fired electricity generation, whose marginal costs are tight to the commodity prices. Such pricing produces more relevant order books and hence, the models using differentiable optimization are more adapted to these markets.
- Models P&O and $\text{P\&O} + \text{DNN}_{OB}$ that do not use direct forecast ($\alpha = 0$) are often outperformed by others on every datasets. Models using the differentiable optimization are better trained when the loss is also computed using direct prediction. This suggests a deeper analysis of the α and β parameters.

[4]CODE: https://github.com/Leonardbcm/MOB, DATA: https://rb.gy/v9ui3

Country	Model	MAE	DAE	RMAE	SMAPE
BE	DNN_Y	7.74	5.79	0.941	21.27
	$\text{DNN}_{Y,OB}$	9.63	4.17	1.17	26.48
	P&O	7.27	4.37	0.884	19.73
	$\text{P\&O} + \text{DNN}_{OB}$	19.85	19.49	2.425	42.19
	$\text{P\&O} + \text{DNN}_Y$	6.85	4.32	0.832	20.35
	$\text{P\&O+DNN}_{Y,OB}$	**6.28**	**3.44**	**0.763**	**17.28**
DE	DNN_Y	7.28	6.67	0.778	29.83
	$\text{DNN}_{Y,OB}$	8.87	6.52	0.946	30.36
	P&O	9.01	6.96	0.958	29.87
	$\text{P\&O} + \text{DNN}_{OB}$	9.24	7.1	0.983	31.24
	$\text{P\&O} + \text{DNN}_Y$	6.99	5.15	0.745	25.97
	$\text{P\&O+DNN}_{Y,OB}$	**6.91**	**4.53**	**0.735**	**25.53**
FR	DNN_Y	**4.54**	3.06	**0.653**	15.5
	$\text{DNN}_{Y,OB}$	5.11	**3.03**	0.734	**15.21**
	P&O	6.47	4.8	0.93	20.31
	$\text{P\&O} + \text{DNN}_{OB}$	5.92	3.5	0.849	18.25
	$\text{P\&O} + \text{DNN}_Y$	5.3	3.22	0.759	16.2
	$\text{P\&O+DNN}_{Y,OB}$	5.79	3.87	0.831	19.51
NL	DNN_Y	6.32	4.43	1.057	18.84
	$\text{DNN}_{Y,OB}$	5.77	3.81	0.965	15.5
	P&O	6.53	3.96	1.092	16.47
	$\text{P\&O} + \text{DNN}_{OB}$	10.97	10.34	1.838	25.18
	$\text{P\&O} + \text{DNN}_Y$	**5.22**	**3.49**	**0.874**	**13.4**
	$\text{P\&O+DNN}_{Y,OB}$	5.79	4.47	0.968	14.41

TABLE II
METRICS OBTAINED ON THE TEST PERIOD FOR DIFFERENT MODEL CONFIGURATIONS. BOLD VALUES INDICATE THE BEST METRIC AMONG ALL MODEL CONFIGURATIONS FOR A GIVEN DATASET.

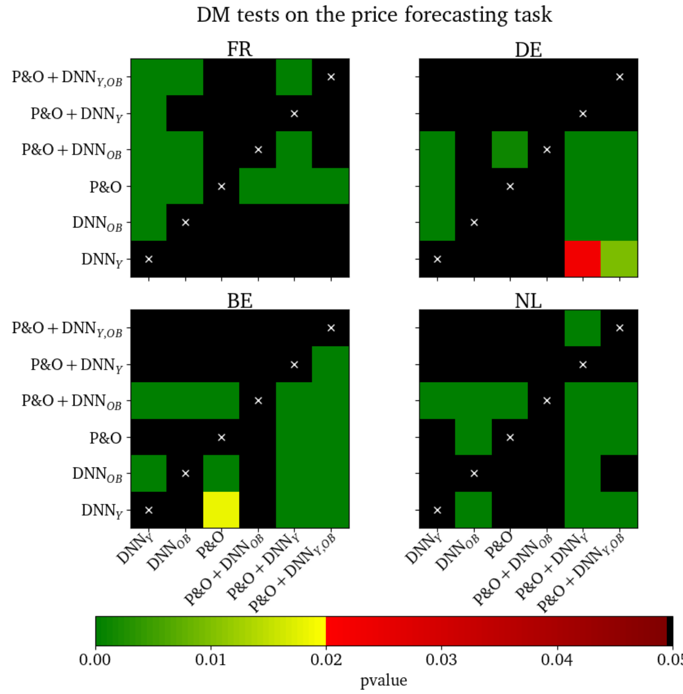

Fig. 6. p-values of the Diebold & Mariano test for the price forecast task. Colored squares in (i, j) indicate that the forecasts of model i are significantly more accurate than forecasts of model j. Green columns indicate that the corresponding models are significantly better than every other. Black lines indicate that the model on the y-axis' forecasts are significantly worse than every other.

b) Varying the β parameter: In this experiment, we focused on the Belgian dataset. For this dataset, we have seen that model DNN_Y ($\alpha = 1$) is outperformed by model P&O ($\beta = 1$), but both models are outperformed by model $\text{P\&O} + \text{DNN}_Y$ ($\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$) and $\text{P\&O} + \text{DNN}_{Y,OB}$ ($\alpha = \frac{1}{3}$ and $\beta = \frac{1}{3}$). Our purpose is to find even more adequate values for α and β parameters. To this aim, we start from the configuration of DNN_Y ($\alpha = 1$) and increase β by steps of 5%, while decreasing α by steps of 5%. Results are displayed in Figure 7. It is clear that not considering the optimization problem during training ($\beta = 0$) does not yield the best results. However, considering only the predict-and-optimize part ($\beta = 1$) is not an adequate solution either. We see in Fig. 7 that adding the optimization loss even with a very small weight in the global loss (5%) increases all the considered metrics.

c) Contribution Analysis: We now perform a contribution analysis using SHAP [27]. Our aim is to determine which features have been prioritized by adding the differentiable optimization problem, and how the weight of the differentiation in the loss affects features contribution. To this aim, we compute the difference of contribution between a variable with $\beta > 0$ and its contribution when $\beta = 0$. For each value of β, we compute 1000 SHAP values on the test set. For a clearer analysis, we regroup the contributions by families of variable and we display them on Fig. 8. Colored squares quantifies the variation of contribution between $\beta = 0$ and $\beta > 0$. We observe almost no variation on the domestic features (first 4 columns), with only a slight increases of the price contribution at the expense of the generation forecast (gen). Changes for Foreign features are more pronounced. The contribution of Foreign Prices (F. price) are increased for smaller values of β while higher values favor the Swiss and English prices (CH price, UK price). We also note important decrease of the contribution of the Foreign generation forecast (F. ren gen) for all values of β, and of the contribution of the Foreign generation and consumption forecast (F. gen) and (F. conso) for higher values of β.

d) Discussion: In the scope of our study, the order book modeling and differentiation of the optimization problem lead to significantly better results on the datasets where the energy mix is eclectic. For the Belgian dataset, the analysis of the β hyper-parameter shows that the best model is a combination of the traditional approach where prices are directly forecast from exogenous features $\beta = 0$ and the differentiable optimization approach ($\beta > 0$). This experiment suggests that the optimal value of β could be found using hyper-parameter search methods. The variation in the contribution also reveals that adding differentiable optimization to the model guides it to a more refined representation of the variables. The Domestic and Foreign consumption, generation and renewable generation forecasts are less regarded, at the profit of prices. Indeed, while those forecasts are to an extent correlated to the prices, this correlation does not hold with the order books whose complex structure is more price-dependent. This is especially true for Swiss and English prices that are set at 11am and thus available for market players to form their order books.

To illustrate this finding, we provide an extract of the test

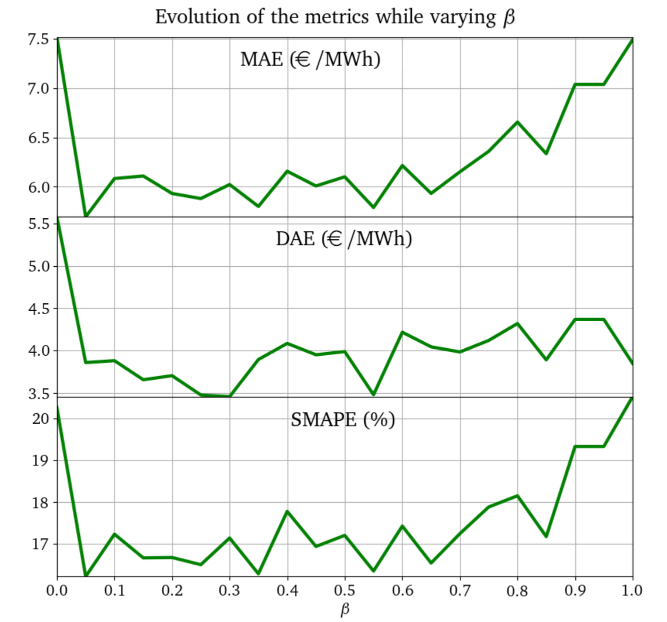

Fig. 7. Quality of the obtained predictions according to β on the Belgian dataset: MAE (top), DAE (middle) and SMAPE (bottom).

set predictions in Figure 9, where the Belgian day-ahead prices from 10/10/2019 to 12/10/2019 are displayed on the top diagram along with the forecasts of different models ($\beta = 0$ in red, $\beta = \frac{1}{2}$ in yellow and $\beta = 1$ in green). On the bottom diagram, the market conditions are displayed (forecast generation and forecast residual load), expressed as % of deviation from their normal values. Focusing on the 1am auction of the 12/10/2019, we remark that the model without differentiable optimization ($\beta = 0$), given that the residual load is 40% below the normal, greatly underestimates the price. The real sensibility of the price to consumption variations, given by the order book displayed in Figure 4, cannot be captured by the model without differentiable optimization, while the other two models with $\beta = \frac{1}{2}$ and $\beta = 1$ predict the price accurately.

VI. RELATED WORK

While many ML models have been proposed for EPF [6], [11]–[21], only two papers have tried to include the price-fixing algorithm. First, Schürch and Wagner [28] aim to extract features from the ORDER BOOK before feeding them to a ML algorithm. Then, Tschora et al. [29] model energy flow between countries by solving an optimization problem before forecasting prices using Graph Neural Networks. Those two approaches sequentially apply a transformation to the data, and then learn a Neural Network for prediction. In this current work, we try to link both steps so that the data transformation is optimal with respect to the prediction task. In this context, we aim at developing a decision-focused framework that forecasts ORDER BOOKS then used to solve the EUPHEMIA algorithm.

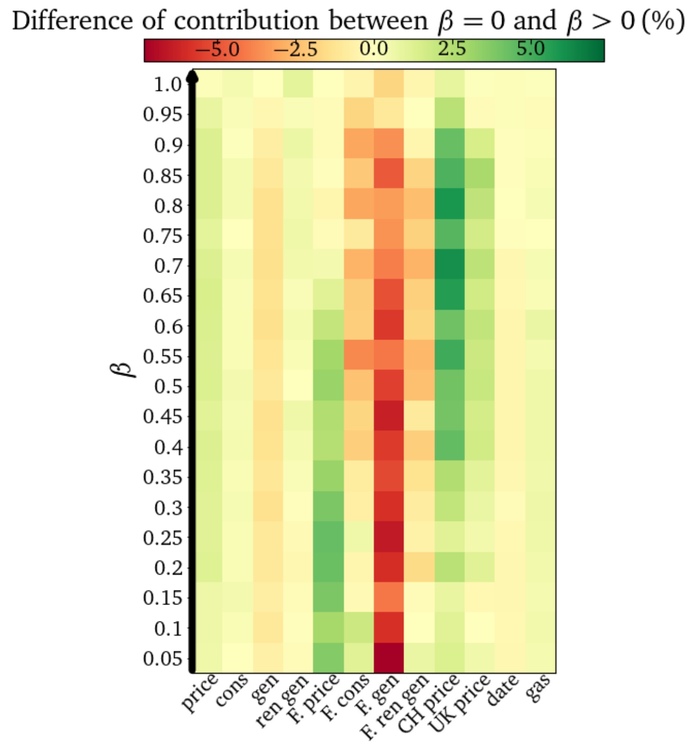

Fig. 8. Variation of the feature contributions while increasing β on the Belgian dataset. A green square at coordinate (i, j) indicates that the model trained with $\beta = i$ increase the contribution of feature j compared to the model with $\beta = 0$. A red square indicates the opposite. On the x-axis, the features are regrouped by category. For instance, F. CONS is the sum of contribution of all Foreign Consumption Forecasts (for Belgium, it is the French, German, Dutch, Spanish, Italian, English, Austrian and Swiss prices).

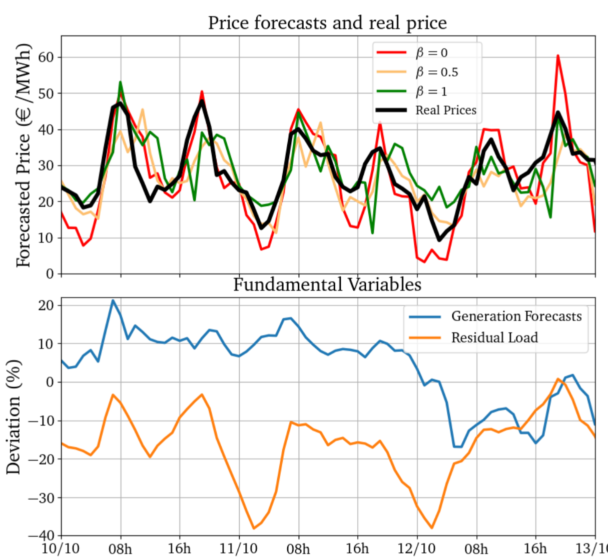

Fig. 9. Day-Ahead prices on the Belgian dataset from 10/10/2019 to 12/10/2019 (black line) and their predicted values for various β (Top). Fundamental variables of the Belgian market for the same period, expressed as percentage of deviation from the average (Bottom).

Real-world decision-making problems often require the integration of machine learning and combinatorial optimization.

In such scenarios, combinatorial optimization problems are solved to arrive at decisions that maximize or minimize an objective function. However, it is common for certain parameters of the optimization problem to be unknown but can be estimated from other feature attributes using historical data. A common solution to tackle this problem is to adopt a **predict-then-optimize** strategy. Such an approach sequentially involves a machine learning model to provide point estimates for uncertain parameters, followed by solving the optimization problem using these predictions. However, this type of methods assumes independent parameter errors and neglects the interplay between these errors and their impact on the combinatorial optimization problem.

To overcome this issue, **decision-focused learning** [30] – also known as predict-and-optimize[5]– was introduced. Decision-focused learning has gained significant attention in the field of machine learning and optimization. Decision-focused learning involves optimization problems where the optimization parameters are only partially defined. A notable advancement in this area is the introduction of the differentiable optimization layer [33]. This layer calculates gradients by differentiating the Karush-Kuhn-Tucker (KKT) optimality conditions of a quadratic program. The solver comes with a cubic complexity. Amos and Kolter [33] recommend not exceeding 1000 hidden dimensions which is largely exceeded by our problem. In comparison, our solution based on dichotomy search has a complexity in $\mathcal{O}\left(log_2(|OB|)\right)$, which make the problem tractable. The so-called Optnet approach [33] is not applicable to linear programming. To address this limitation, Wilder et al. [34] propose incorporating a small quadratic regularizer in the objective function. Mandi and Guns [35] introduce a log-barrier regularization term and compute gradients using the homogeneous self-dual embedding of the linear programming. In [36], the authors focus on mixed integer linear programmings and convert them to linear programming by adding cutting planes to the root linear programming node. Recent works have also investigated the integration of learning to rank into decision-focused learning [25], [37]. To the best of our knowledge, our paper is the first decision-focused learning method to predict the electricity price, taking into account the EUPHEMIA optimization through differentiable optimization.

VII. Conclusion

In this paper, we addressed the problem of Electricity Price Forecasting (EPF) by combining a prediction model with an optimization task based on the EUPHEMIA algorithm. We aimed to demonstrate the benefits of the end-to-end predict-and-optimize approach for this problem. Our research hypothesis was that tightly coupling the EUPHEMIA optimization task with order book prediction from exogenous data would improve the accuracy of the model predictions. To achieve this, we first formalized the EUPHEMIA optimization problem and proposed a method for solving it to determine electricity prices from order books. We then explored different ways of integrating the optimal decision process while learning the order book prediction model. By directly minimizing the error related to the downstream decision-making task, we aimed to improve the overall performance of the EPF model. In the experimental evaluation, we compared several configurations of the end-to-end decision-focused model using four different European datasets. The results demonstrated the effectiveness of our approach, showing improvements in the accuracy of the predictions compared to traditional methods for the datasets for which the exogenous variables make it possible to finely estimate the order books. The tight coupling of the EUPHEMIA optimization task with the order book prediction task then allows the model to capture more relevant information and make more accurate price predictions. Furthermore, we conducted a qualitative analysis using Shap values to understand the factors and features that influenced the predictions and decisions made by the predict-and-optimize model. This analysis provided insights into the captured information and shed light on the relationships between the input features and the final decision-making process. Our study highlights the benefits of the end-to-end predict-and-optimize approach for Electricity Price Forecasting. By integrating the EUPHEMIA optimization task with order book prediction, we achieved improved accuracy in price forecasting, which has significant implications for the efficient management of electricity production and consumption. Our findings pave the way for further research and development in this area, including exploring other optimization algorithms and refining the prediction models to enhance the overall performance of EPF systems.

[5] In [31], a Dagstuhl seminar report, the group of experts recommend to use decision-focused learning instead of predict-and-optimize [24], [25] or predict+optimize [32] because of some confusion around the terminology.

References

[1] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, “Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark,” *Applied Energy*, vol. 293, p. 116983, 2021.

[2] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *International Journal of Forecasting*, vol. 30, 10 2014.

[3] L. Amabile, D. Bresch-Pietri, G. El Hajje, S. Labbé, and N. Petit, “Optimizing the self-consumption of residential photovoltaic energy and quantification of the impact of production forecast uncertainties,” *Advances in Applied Energy*, vol. 2, p. 100020, 2021.

[4] J. Mei, Y. Zuo, C. H. Lee, X. Wang, and J. L. Kirtley, “Stochastic optimization of multi-energy system operation considering hydrogen-based vehicle applications,” *Advances in Applied Energy*, vol. 2, p. 100031, 2021.

[5] Z. Tan, J. Zhang, J. Wang, and J. Xu, “Day-ahead electricity price forecasting using wavelet transform combined with arima and garch models,” *Applied Energy*, vol. 87, no. 11, pp. 3606 – 3610, 2010.

[6] J. Che and J. Wang, “Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling,” *Energy Conversion and Management*, vol. 51, p. 1911–1917, 10 2010.

[7] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models,” *IEEE transactions on power systems*, vol. 20, no. 2, pp. 1035–1042, 2005.

[8] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” *IEEE transactions on power systems*, vol. 18, no. 3, pp. 1014–1020, 2003.

[9] J. Crespo Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner, "Forecasting electricity spot-prices using linear univariate time-series models," *Applied Energy*, vol. 77, pp. 87–106, 01 2004.

[10] R. Weron and A. Misiorek, "Forecasting spot electricity prices with time series models," *Proceedings of the European Electricity Market EEM-05 Conference*, 05 2005.

[11] D. C. Sansom, T. Downs, and T. K. Saha, "Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants," *Journal of Electrical & Electronics Engineering, Australia*, vol. 22, no. 3, pp. 227–233, 2003.

[12] A. Shiri, M. Afshar, A. Rahimi-Kian, and B. Maham, "Electricity price forecasting using support vector machines by considering oil and natural gas price impacts," in *mybook*, 08 2015, pp. 1–5.

[13] J. Zhao, Z. Dong, Z. Xu, and K. Wong, "A statistical approach for interval forecasting of the electricity price," *Power Systems, IEEE Transactions on*, vol. 23, pp. 267 – 276, 06 2008.

[14] J. Mei, D. He, R. Harley, T. Habetler, and G. Qu, "A random forest method for real-time price forecasting in New York electricity market," in *2014 IEEE PES General Meeting| Conference & Exposition*. IEEE, 2014, pp. 1–5.

[15] A. R. Miralles, J. R. Dorronsoro, and J. Diaz, "Day-ahead price forecasting for the Spanish electricity market," Instituto de Ingenieria del Conocimiento and Universidad Autonoma de Madrid, Tech. Rep., 2019, https://www.ijimai.org/journal/node/2300.

[16] R. Pino, J. Parreno, A. Gomez, and P. Priore, "Forecasting next-day price of electricity in the spanish energy market using artificial neural networks," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 1, pp. 53 – 62, 2008.

[17] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network based models," *Applied Energy*, vol. 172, pp. 132 – 151, 2016.

[18] H. Mosbah and M. El-Hawary, "Hourly electricity price forecasting for the next month using multilayer neural network," *Canadian Journal of Electrical and Computer Engineering*, vol. 39, 09 2015.

[19] D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner, "Extended forecast methods for day-ahead electricity prices applying artificial neural networks," *Applied Energy*, vol. SCI, pp. 218–, 01 2016.

[20] S. Anbazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using recurrent neural network," *IEEE Systems Journal*, vol. 7, no. 4, pp. 866–872, 2013.

[21] Z. Khan, S. Fareed, M. Anwar, A. Naeem, H. Gul, A. Arif, and N. Javaid, "Short term electricity price forecasting through convolutional neural network (cnn)," in *BOOK*, 02 2020.

[22] L. Tschora, E. Pierre, M. Plantevit, and C. Robardet, "Electricity price forecasting on the day-ahead market using machine learning," *Applied Energy*, vol. 313, p. 118752, 2022.

[23] PCR, "Euphemia public description," Price Coupling of Region, Tech. Rep., 2016. [Online]. Available: https://www.nordpoolgroup.com/globalassets/download-center/single-day-ahead-coupling/euphemia-public-description.pdf

[24] O. El Balghiti, A. N. Elmachtoub, P. Grigas, and A. Tewari, "Generalization bounds in the predict-then-optimize framework," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[25] J. Mandi, V. Bucarey, M. M. K. Tchomba, and T. Guns, "Decision-focused learning: Through the lens of learning to rank," in *International Conference on Machine Learning, ICML*, 2022, pp. 14 935–14 947.

[26] F. Diebold and R. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 20, pp. 134–44, 02 1992.

[27] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *NEURIPS*, 2017, pp. 4765–4774. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[28] S. Schnürch and A. Wagner, "Machine learning on epex order books: Insights and forecasts," *arXiv preprint arXiv:1906.06248*, 2019.

[29] L. Tschora, E. Pierre, M. Plantevit, and C. Robardet, "Forecasting electricity prices: An optimize then predict-based approach," in *Advances in Intelligent Data Analysis, IDA 2023*, ser. Lecture Notes in Computer Science, vol. 13876. Springer, 2023, pp. 446–458.

[30] J. Kotary, F. Fioretto, P. Van Hentenryck, and B. Wilder, "End-to-end constrained optimization learning: A survey," *arXiv preprint arXiv:2103.16378*, 2021.

[31] S. Teso, L. Bliek, A. Borghesi, M. Lombardi, N. Yorke-Smith, T. Guns, and A. Passerini, "Machine learning for combinatorial optimisation of partially-specified problems: Regret minimisation as a unifying lens," *arXiv preprint arXiv:2205.10157*, 2022.

[32] E. Demirovic, P. J. Stuckey, T. Guns, J. Bailey, C. Leckie, K. Ramamohanarao, and J. Chan, "Dynamic programming for predict+optimise," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020, pp. 1444–1451.

[33] B. Amos and J. Z. Kolter, "Optnet: Differentiable optimization as a layer in neural networks," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ser. ICML'17. JMLR.org, 2017, p. 136–145.

[34] B. Wilder, B. Dilkina, and M. Tambe, "Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization," in *The Thirty-Third AAAI Conference on Artificial Intelligence*, 2019, pp. 1658–1665.

[35] J. Mandi and T. Guns, "Interior point solving for LP-based prediction+optimisation," in *Conference on Neural Information Processing Systems, NeurIPS*, 2020.

[36] A. M. Ferber, B. Wilder, B. Dilkina, and M. Tambe, "MIPaaL: Mixed Integer Program as a Layer," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, 2020, pp. 1504–1511.

[37] E. Demirovic, P. J. Stuckey, J. Bailey, J. Chan, C. Leckie, K. Ramamohanarao, and T. Guns, "Predict+optimise with ranking objectives: Exhaustively learning linear functions," in *International Joint Conference on Artificial Intelligence, IJCAI*, 2019, pp. 1078–1085.

ACKNOWLEDGMENT

This research has partially been funded by the ANRT (French National Association for Research and Technology).