Dealing with Uncertainty in the Smart Grid: A Learning Game Approach

Hélène Le Cadre, Jean-Sébastien Bedo

Dealing with Uncertainty in the Smart Grid: A Learning Game Approach

Hélène Le Cadre* Jean-Sébastien Bedo†

Abstract

In this article, the smart grid is modeled as a decentralized and hierarchical network, made up of three categories of agents: suppliers, generators and microgrids. To optimize their decisions concerning prices and traded power, agents need to forecast the demand of the microgrids and the fluctuating productions of the generators. The biases resulting from the decentralized learning could create imbalances between demand and supply leading to penalties for suppliers and for generators. We analytically determine prices that provide generators with a guarantee to avoid such penalties, transferring risk to the suppliers. Additionally, we prove that collaborative learning, through a grand coalition of suppliers in which information is shared and forecasts aligned on a single value, minimizes the sum of their average risk. Simulations, run for a large sample of parameter combinations using external and internal regret minimization, show that the convergence of the collaborative learning strategy is clearly faster than that resulting from distributed learning. Finally, we analyze the suppliers' individual incentives to enter into a grand coalition and the tightness of the learning algorithm's theoretical bounds.

Keywords: Distributed Learning ; Information ; Regret ; Learning Game Theory

1 Introduction

In Europe, and especially France, power networks rely heavily on nuclear-based technology. With this type of non-renewable technology, generations can be adapted by the plant operator who alternates openings and closings and optimizes the duration of the switches between modes. The objective is thus to adapt the generations so as to meet the uncertain demand level. We built a first model in [13], using two learning strategies: the first was based on tit for tat and the second, on fictitious play. For renewables, generations can only be partially controlled, for instance, by lowering the wind turbine speed [18]. Renewable integration in the power network requires deploying smart Information and Communication Technologies (ICTs) to supervise the grid operations [19]. Indeed, renewable generation is highly unpredictable since it depends

*MINES ParisTech, PSL Research University, CMA - Center for Applied Mathematics, CS 10207 rue Claude Daunesse, 06904 Sophia Antipolis, France. *Email*: helene.le_cadre@mines-paristech.fr

†Orange Labs, 38 − 40, rue du Général Leclerc, 92130 Issy-les-Moulineaux, France. *Email*: jeanse-bastien.bedo@orange.com

on uncontrolable exogenous factors like wind, sun, swell, etc. [22]. Furthermore, the new active role of end users, who can become power generators, dynamically adapt their consumption and fit into a multitude of microgrids [14], [25], [26], dramatically increases the volume of exchanged data flows. ICTs appear to be a means to retrieve the most salient information from this large amount of data and to train forecasters to provide efficient predictions regarding fluctuating generations such as renewables. These predictions will then be used as inputs to optimize the smart grid operations [2].

In practice, it is increasingly apparent that current forecasting methods cannot properly handle extreme situations corresponding to either severe weather phenomena or critical periods for power system operations. For example, forecasting methods used to predict wind power were mostly designed to provide single value forecasts to estimate the generations. Only recently, probabilistic methods have been introduced to provide estimations of the entire distribution of future generations [3]. In such methods, forecasts may take the form of either quantile estimations or density estimations [5], [9]. Learning based on regret minimization, as described in [4], belongs to this latter category. This class of method is particulary efficient [12]. It provides the forecaster with a density function that associates a weight to each possible output. The density function is updated by merging information from various experts' reports. As a result, these methods are more robust in the face of extreme events and appear particularly well suited to modeling erratic processes such as renewable productions.

In the framework of the smart grid, learning is performed in a decentralized manner since each agent primarily learns the hidden information using its own observations. Existing literature on distributed learning mostly focuses on distributed learning algorithms that are suitable for implementation in large-scale engineering systems [15], [24]. The results mainly concentrate on a specific class of games, called games of potential [27]. This class of games is of particular interest since they have inherent properties that can provide guarantees on the convergence and stability of the system. However, this framework has some limitations, the most striking of which is that it is very difficult to build a full system model from a potential game [18].

The learning game studied in this paper belongs to the category of repeated uncoupled games since one agent cannot predict the forecasts and so actions of the other agents at a given time period. To take its decision i.e., optimal prices and power orders, each agent is aware of the forecast history of all of the agents and of its utility. For finite games with generic payoffs, recent work has shown the existence of completely uncoupled learning rules i.e., rules where the agents observe only their own prediction history and their utility, leading to Pareto-optimal Nash equilibria [24]. Marden et al. exhibit a different class of learning procedures that lead to a Pareto-optimal vector of actions that do not necessarily coincide with Nash equilibria [18]. Close to the work exposed in our article, Zheng et al. propose an online algorithm that simultaneously updates the weight given to each forecaster using regularized sequential linear regression, while allowing each forecaster to be retrained based on the latest observations in an online manner [28]. The updating of the individual forecasters to accomodate the online observations relies on a gradient-descent algorithm. Expert system coordination can also be used to aggregate the set of predictors into a better global predictor. Holsapple et al. provide a method based on competition among distinct expert systems in [11].

Most collaborative mechanisms studied in literature lead to price or quality of service alignment. In addition, the group composition provides an additional state space in which information about the environment can be accumulated [20]. To our knowledge, no study has so far been made of the impact of collaboration through information sharing and forecast alignment, when prices are individually determined, on the underlying system performance. Of course, collaboration might not emerge due to the agents' natural incentives to cheat and deviate from the cooperative equilibrium and also, most frequently, due to the regulator's intervention. There are a number of well-understood reasons why regulators often do not allow horizontal collaboration: if suppliers are allowed to collaborate, they might cooperate to raise the price i.e., reduce quantity below the efficient baseline, and create market power [6]. Alternatively, suppliers might cooperate to reduce quality of service. Courts punish agreements that explicitly aim to decrease competition.

In this article, we answer the following questions:

- How will the biases, introduced by errors made by the agents in their predictions, affect the agents' average risk?
- Does collaborative learning improve the smart grid's overall performance and should it therefore be encouraged by the regulator?

The article is organized as follows. In Section 2, we introduce the economic basis of our model, the agents, their utility and their optimization program. Complete information Stackelberg game is then solved in Section 3, proceeding by backward induction. We analytically derive the optimal prices and power orders for the agents. Partial information is introduced in Section 4 where the interacting agents learn hidden individual sequences in a distributed fashion. To illustrate the theoretical results derived in the previous sections, in Section 5 we compare: firstly the time of convergence of suppliers' learning strategies under external and internal regret minimization in cooperative and non-cooperative scenarios, secondly which behaviors should emerge depending on the game parameters value and finally, analyze the tightness of the bounds derived theoretically in various scenarios.

Throughout the article, we use the notation: $x_+ \triangleq \max\{x; 0\}$ to denote the positive part of the real number x.

2 The model

A large number of agents interact in the smart grid. In this article, we model the smart grid as a three layer hierarchical network whose evolution depends on the interactions between the agents composing each layer and also, on the ability of the agents to cope with fluctuating power generations and demand [2], [19]. We detail the three categories of agents and the repeated game which captures the interactions between them in Subsection 2.1. Then, we define each agent's optimization program in Subsection 2.2.

2.1 Description of the agents

We model the smart grid through three categories of agents: microgrids, suppliers and generators. The first agent category is composed of microgrids. Microgrids generate some power demands (mostly for heating/cooling buildings and for the individual usages of their inhabitants) and also some power through non renewable sources like solar panels and wind roofings. Demand can be flattened when end users change their normal power consumption patterns in response to price variations over time [1], [16]. These online changes are called pricing demand response (DR) in literature [14]. The second agent category is composed of suppliers who buy power from several generators and resell it to the end users. The third category is composed of generators. Each generator produces and sells power to all suppliers.

We assume that each end user has a contract with one supplier only and does not churn from one supplier to another for all the duration of our study. This assumption holds good if we consider local or regional utility companies. In this sense, the set made of end users supplied by a single supplier can be seen as an individual microgrid, as defined in [14], [25], [26] and recalled in the Introduction. We denote by s_i, with i varying between 1 and n, the i-th supplier and by $\mathcal{M}_i$ the corresponding group of end users. The generators are denoted g_k with k varying between 1 and K. The generators can be associated with photovoltaic park managers, wind farm administrators, etc. In this article, we assume that the generator cannot directly influence its power generation at each time period. This assumption holds good if we look at renewable sources like a wind turbine farm without any investment in an additional wind turbine during the study period. The variation of wind intensity will impact the amount of power generated without any lever for the generator [1].

We model the interplay between all of the agents through a repeated game. At each time period t, the following game is played:

Basic Game Description

(i) Each generator g_k communicates its unitary price $\tilde{p}_k(t) > 0$ to the suppliers. The prices are fixed independently and simultaneously by each generator so as to maximize its utility.

(ii) Each supplier s_i places power orders with the generators: the quantity ordered by supplier s_i from generator g_k is denoted by $q_{ik}(t)$. Each supplier s_i communicates its unitary price $p_i(t) > 0$ to its microgrid. The power orders and prices are fixed independently and simultaneously by each supplier so as to maximize its utility.

(iii) Microgrid $\mathcal{M}_i$ generates $\nu_i^s(t)$ power units and buys $a_i(t)$ power units from supplier s_i. The quantity $a_i(t)$ is chosen so as to maximize the benefits for $\mathcal{M}_i$.

At each time period, generator g_k produces $\nu_k^g(t)$ power units. It then delivers $\alpha_{ki}(t)\nu_k^g(t)$ units to supplier s_i where $\alpha_{ki}(t) \geq 0$ denotes the proportion of its generation that generator g_k allocates to supplier s_i, with the normalization constraint: $\sum_{i=1,\ldots,n} \alpha_{ki}(t) = 1$. This proportion is defined depending on power orders received by

[1] Non-renewable generators like nuclear plants could be integrated into the grid. This would require using distributed control rules, such as those described in [13], [18].

g_k from all of the suppliers. The sum of the orders received by g_k may exceed $\nu_k^g(t)$ and so each supplier may receive less power than it initially ordered.

Penalties are incurred by both suppliers and power generators if they cannot fully satisfy their customers' demand. Supplier s_i incurs a cost $\gamma_i > 0$ per power unit missing for the supply of its microgrid, measured a posteriori. The cost scale is published on the TSO's website [29]. It is paid to the Transmission System Operator for electricity (TSO). In France, the TSO defined rules to encourage agents to become balance operators. According to these rules, a negative balance on the free market must be compensated by buying the missing power from the TSO at a unit price defined through the adjustment mechanism. This adjustment mechanism price is higher than the free market electricity unit price. The entire adjustment mechanism is implemented by the TSO, which compensates any negative balances to maintain the power network's reliability. In our article, the adjustment price is different for each supplier. The price discrimination is justified by the fact that depending on its geographic location, a negative energy balance can be easily corrected in densely interconnected areas whereas it is much more difficult in remote ones due to the high cost of electricity transmission. As a result, γ_i is higher for suppliers serving remote locations than in densely interconnected areas. Generator g_k incurs a cost $\tilde{\gamma}_i > 0$ per missing power unit in the provision of supplier s_i, measured a posteriori. This fee is paid to the regulator of the capacity market that needs to be implemented to balance supply and demand in the smart grid [30]. Indeed, to guarantee the reliability of the capacity market, it might be necessary to implement a feedback mechanism by which the regulator compensates the negative energy balances of the generators by making its own investment in capacity [30]. The costs of these investments would be recovered from the penalties imposed on the generators. These investments are spread over relatively long periods; in the short term, the regulator must appeal to the TSO whose share of imports from neighboring energy markets readjusts the level of production on suppliers' orders.

2.2 Optimization program for each agent

In this subsection, we describe the decision variables and utilities for each category of agents. The optimization program for each agent is presented using a mathematical formulation.

2.2.1 Microgrid programs

The only decision variable for microgrid $\mathcal{M}_i$ is the power that it demands from supplier s_i at time period t: $a_i(t)$. We assume that the microgrid has no lever to influence the random generation from its intermittent sources: $\nu_i^s(t)$.

We model the microgrid benefit of consuming x power units as a quadratic function. More precisely we use the model of Fahrioglu and Alvarado [7]. The utility of consuming x power units then equals $\theta_i\left(b_0 x - \frac{1}{2} b_1 x^2\right)$ where θ_i is a positive parameter depending on the microgrid composition (households, firms, etc.) and b_0 and b_1 are generic positive parameters. We note that this model is valid only when $x < \frac{b_0}{b_1}$. Beyond this threshold, the marginal benefit of consuming an additional power unit is

zero and the total benefit of consuming power is constant and equals $\theta_i \frac{b_0^2}{2b_1}$. As a result, the net benefit for microgrid $\mathcal{M}_i$, defined as the benefit of consuming $\nu_i^s(t) + a_i(t)$ power units minus the cost of buying $a_i(t)$ power units from supplier s_i, is:

$$B_i(t) = \begin{cases} \theta_i\Big(b_0(\nu_i^s(t)+a_i(t)) - \frac{1}{2}b_1(\nu_i^s(t)+a_i(t))^2\Big) & \\ \qquad\qquad -p_i(t)a_i(t) & \text{if } \nu_i^s(t)+a_i(t) < \frac{b_0}{b_1} \\ \theta_i\Big(\frac{b_0^2}{2b_1}\Big) - p_i(t)a_i(t) & \text{if } \nu_i^s(t)+a_i(t) \geq \frac{b_0}{b_1} \end{cases} \tag{1}$$

Microgrid $\mathcal{M}_i$ chooses $a_i(t) \geq 0$ in order to maximize its net benefit[2]. Therefore, its optimization program takes the form: $\max_{a_i(t)\geq 0}\Big\{B_i(t)\Big\}$. Its decision depends on the price $p_i(t)$ fixed by supplier s_i.

2.2.2 Supplier programs

The decision variables for each supplier s_i are the unit price $p_i(t)$ and the power orders $\Big(q_{ik}(t)\Big)_k$ placed with each generator g_k.

Following our description of the interplay between the agents, supplier s_i's utility at time period t is:

$$\pi_i(t) \quad = \quad p_i(t)a_i(t) - \sum_{k=1,\ldots,K} q_{ik}(t)\tilde{p}_k(t) - \gamma_i\Big(a_i(t) - \sum_{k=1,\ldots,K} \alpha_{ki}(t)\nu_k^g(t)\Big)_+^2 \tag{2}$$

Supplier s_i chooses its unit price and its power orders in order to maximize $\pi_i(t)$, as defined in Equation (2). Its optimization program takes the form: $\max_{p_i(t)>0,(q_{ik}(t))_k\in\mathbb{R}_+^K}\Big\{\pi_i(t)\Big\}$.

2.2.3 Generator programs

The only decision variable for each generator g_k is the unit price, $\tilde{p}_k(t)$, at which it provides power to the suppliers.

The utility of generator g_k at time period t equals:

$$\tilde{\pi}_k(t) \quad = \quad \tilde{p}_k(t) \sum_{i=1,\ldots,n} q_{ik}(t) - \sum_{i=1,\ldots,n} \tilde{\gamma}_i\Big(q_{ik}(t) - \alpha_{ki}(t)\nu_k^g(t)\Big)_+$$

To define the sharing coefficients, $\Big(\alpha_{ki}(t)\Big)_i$, we consider a weighted proportional allocation of resources that allows generators to discriminate power allocation by supplier while allocating its power simultaneously among the suppliers. This framework is a generalization of the well-known proportional allocation mechanism in which the suppliers' bids coincide with their power orders weighted by their penalty coefficients $(\tilde{\gamma}_i)_i$ [23], [26]. This means that when two suppliers have identical power orders,

[2]Note that at the individual household scale, the net demand might be negative since the household can produce power through non renewable sources and use it for its own consumption.

the one with the highest penalty coefficient receives the largest share of the generator's available power. Indeed, the generator wants to minimize its overall penalty and, therefore, allocates larger shares of its power to suppliers serving isolated areas where failure of electricity supply may be critical. The choice of such a resource sharing mechanism can be justified by three factors: firstly, a small extension of Nguyen and Vojnović's work [23] shows that weighted payment auction achieves competitive transfers to generators and to the TSO compared to standard price discrimination schemes at the equilibrium; secondly, the implementation of a sequential resource allocation mechanism based on priority and without storage facilities should be avoided since some suppliers might be left without allocation at all [14]; thirdly, a sequential allocation of the power at the beginning of the time period is almost impossible because the generations are random individual sequences of which outputs are only (partially) observed at the end of the time period. This last point will be detailed in Section 4. We set:

$$\alpha_{ki}(t) \triangleq \frac{\tilde{\gamma}_i q_{ik}(t)}{C_k(t)} \tag{3}$$

where $C_k(t) = \sum_{j=1,\dots,n} \tilde{\gamma}_j q_{jk}(t)$. Using Equation (3), generator g_k's utility at time period t can be rewritten:

$$\tilde{\pi}_k(t) = \tilde{p}_k(t) \sum_{i=1,\dots,n} q_{ik}(t) - \sum_{i=1,\dots,n} \tilde{\gamma}_i q_{ik}(t) \Big(1 - \frac{\tilde{\gamma}_i}{C_k(t)} \nu_k^g(t)\Big)_+ \tag{4}$$

Generator g_k chooses its unit price so that $\tilde{\pi}_k(t)$, as defined in Equation (4), is maximized. Its optimization program takes the form: $\max_{\tilde{p}_k(t)>0} \Big\{\tilde{\pi}_k(t)\Big\}$.

3 Complete information game resolution

The game setting described in Subsection 2.1 implies that in generator-supplier relationship, generators lead and suppliers follow. Similarly, in supplier-microgrid relationships, suppliers lead and microgrids follow. Under such a setting, the game is called a Stackelberg game and, as usual, it should be solved using backward induction [14], [21].

Additionally, we assume that each generator receives at least one power order from a supplier, guaranteeing that the Stackelberg game admits non trivial solutions.

3.1 Optimization of the microgrids' decisions

Microgrid $\mathcal{M}_i$ has to choose $a_i(t)$ in order to maximize its net benefit, defined by Equation (1).

If $a_i(t) + \nu_i^s(t) \geq \frac{b_0}{b_1}$, then the derivative of $B_i(t)$ with respect to $a_i(t)$ equals $-p_i(t)$. Under this assumption, the optimal $a_i(t)$ is reached at the lower bound of the interval i.e., $a_i(t) = \frac{b_0}{b_1} - \nu_i^s(t)$.

If $a_i(t) + \nu_i^s(t) < \frac{b_0}{b_1}$, the derivative of $B_i(t)$ with respect to $a_i(t)$ equals:

$$\frac{\partial B_i(t)}{\partial a_i(t)} = \theta_i\Big(b_0 - b_1(a_i(t) + \nu_i^s(t))\Big) - p_i(t)$$

This derivative equals 0 when:

$$a_i(t) = \frac{b_0}{b_1} - \nu_i^s(t) - \frac{p_i(t)}{b_1\theta_i} \tag{5}$$

We note that for this value of $a_i(t)$ we automatically obtain $a_i(t) + \nu_i^s(t) < \frac{b_0}{b_1}$. In addition, we compute the second order derivative of $B_i(t)$ with respect to $a_i(t)$: $\frac{\partial^2 B_i(t)}{\partial^2 a_i(t)} = -\theta_i b_1 < 0$. This implies that the maximum of $B_i(t)$ is reached when $a_i(t)$ is defined by Equation (5).

To avoid a negative demand, we need to impose the following constraint: $\nu_i^s(t) < \frac{b_0}{b_1} - \frac{p_i(t)}{b_1\theta_i}$. Otherwise, the optimal demand from the microgrid is 0.

3.2 Optimization of the suppliers' decisions

To find its optimal price and power orders, supplier s_i has to replace $a_i(t)$ by its optimal value, defined in Equation (5), in $\pi_i(t)$, defined in Equation (2), and to derive the result in $p_i(t)$ and in $q_{ik}(t)$. This derivation results in two cases.

3.2.1 Case 1: power generation satisfies demand from the microgrid

This is the case when:

$$\frac{b_0}{b_1} - \nu_i^s(t) - \frac{p_i(t)}{b_1\theta_i} < \sum_{k=1,\ldots,K} \alpha_{ki}(t)\nu_k^g(t) \tag{6}$$

Then deriving the supplier's utility in $p_i(t)$ leads to: $\frac{\partial \pi_i(t)}{\partial p_i(t)} = \frac{b_0}{b_1} - \nu_i^s(t) - 2p_i(t)b_1\theta_i$. This derivative equals 0 when $p_i(t) = \frac{\theta_i\Big(b_0 - b_1\nu_i^s(t)\Big)}{2}$ which is the optimal price for supplier s_i. Then the positivity constraint for $a_i(t)$ becomes $\nu_i^s(t) < \frac{b_0}{b_1}$. In addition, the derivative of the supplier's utility in $q_{ik}(t)$ leads to: $\frac{\partial \pi_i(t)}{\partial q_{ik}(t)} = -\tilde{p}_k(t)$ which means that supplier s_i will try to minimize all of its power orders to maximize its utility. As a result, $\alpha_{ki}(t)$ will tend to be small. This implies, in turn, that supplier s_i will tend to break the inequality defining Case 1 in Inequality (6) and we will always fall on the frontier between Case 1 and Case 2. The frontier between these two cases is defined by the equation:

$$\frac{b_0}{2b_1} - \frac{\nu_i^s(t)}{2} = \sum_{k=1,\ldots,K} \alpha_{ki}(t)\nu_k^g(t) \tag{7}$$

3.2.2 Case 2: power generation does not satisfy power demand from the microgrid

This is the case when $\frac{b_0}{b_1} - \nu_i^s(t) - \frac{p_i(t)}{b_1\theta_i} \geq \sum_{k=1,\ldots,K} \alpha_{ki}(t)\nu_k^g(t)$. Then deriving supplier s_i's utility with respect to $p_i(t)$ and $q_{ik}(t)$ gives us:

$$\begin{aligned} \frac{\partial \pi_i(t)}{\partial p_i(t)} &= \frac{b_0}{b_1} - \nu_i^s(t) + \frac{\gamma_i}{b_1\theta_i} - \frac{2p_i(t)}{b_1\theta_i} \\ \frac{\partial \pi_i(t)}{\partial q_{ik}(t)} &= -\tilde{p}_k(t) + \gamma_i\nu_k^g(t)\frac{\partial \alpha_{ki}(t)}{\partial q_{ik}(t)} \end{aligned} \tag{8}$$

By using the definition of $\alpha_{ki}(t)$, given in Equation (3), we obtain:

$$\frac{\partial \alpha_{ki}(t)}{\partial q_{ik}(t)} = \tilde{\gamma}_i \frac{C_k(t) - \tilde{\gamma}_i q_{ik}(t)}{C_k(t)^2}$$

Going back to the system of Equations (8), we conclude that the derivatives equal 0 when:

$$p_i(t) = \frac{\theta_i\Big(b_0 - b_1\nu_i^s(t)\Big) + \gamma_i}{2} \tag{9}$$

$$\tilde{p}_k(t)C_k(t)^2 = \gamma_i\nu_k^g(t)\tilde{\gamma}_i\Big(C_k(t) - \tilde{\gamma}_i q_{ik}(t)\Big) \tag{10}$$

On the one side, we directly obtain the price at which the derivative of $\pi_i(t)$ equals 0 through Equation (9). We derive from this equation the positivity constraint for $a_i(t)$ which is:

$$\nu_i^s(t) < \frac{b_0}{b_1} - \frac{\gamma_i}{b_1\theta_i}$$

On the other side, Equation (10) can be rewritten as follows:

$$\tilde{\gamma}_i q_{ik}(t) = C_k(t) - \frac{\tilde{p}_k(t)C_k(t)^2}{\nu_k^g(t)\gamma_i\tilde{\gamma}_i} \tag{11}$$

If supplier s_i anticipates that the other suppliers will make the same optimization program, replicating Equation (11) for the n suppliers and adding them all together results in the following equality: $C_k(t) = nC_k(t) - \frac{\tilde{p}_k(t)C_k(t)^2}{\nu_k^g(t)} \sum_{j=1,\ldots,n} \frac{1}{\gamma_j\tilde{\gamma}_j}$ by definition of $C_k(t)$. Then, given that $C_k(t)$ is not zero because each generator g_k receives at least one power order otherwise it would be out of the game, by dividing the previous equation by $C_k(t)$ and reordering we obtain: $C_k(t) = \frac{\nu_k^g(t)}{\tilde{p}_k(t)}\frac{n-1}{\delta}$ where $\delta = \sum_{j=1,\ldots,n} \frac{1}{\gamma_j\tilde{\gamma}_j}$. By replacing $C_k(t)$ in Equation (11), we obtain the power orders for which the derivatives of $\pi_i(t)$ equal 0:

$$q_{ik}(t) = \frac{\nu_k^g(t)}{\tilde{p}_k(t)}\frac{n-1}{\delta}\frac{\beta_i}{\tilde{\gamma}_i} \tag{12}$$

in which we have introduced the notation $\beta_i = 1 - \frac{n-1}{\delta\gamma_i\tilde{\gamma}_i}$ to simplify future calculations.

We now have to check that the price and power orders for which the derivatives of $\pi_i(t)$ equal 0 satisfy the conditions of Case 2.

Firstly, it is easy to check that the price is positive through Equation (9) and the positivity constraint for $a_i(t)$. However, the power orders defined in Equation (12) are non-negative if, and only if, $\beta_i \geq 0$ which is equivalent to $1 \geq \frac{n-1}{\delta\gamma_i\tilde{\gamma}_i}$ which, in turn, is equivalent to:

$$\begin{aligned} & \gamma_i\tilde{\gamma}_i \geq \frac{n-1}{\delta} \\ \Leftrightarrow \quad & \frac{1}{\gamma_i\tilde{\gamma}_i} \leq \frac{1}{n-1}\sum_{j=1,\ldots,n}\frac{1}{\gamma_j\tilde{\gamma}_j} \end{aligned} \tag{13}$$

This inequality means that the penalties related to supplier s_i i.e., $\gamma_i, \tilde{\gamma}_i$, are close to the penalties related to the other suppliers i.e., $(\gamma_j, \tilde{\gamma}_j)_j$. Indeed, if all penalties are equal to γ, then $\delta = \frac{n}{\gamma^2}$ and Inequality (13) is true for all suppliers. On the contrary, if all penalties are equal to γ except for supplier s_1 which has a penalty of $\frac{\gamma}{n-1}$, then $\delta = \frac{(n-1)n}{\gamma^2}$ and Inequality (13) for supplier s_1 becomes $n \geq (n-1)^2$ which is false as soon as $n > 2$. In this case, supplier s_1 would not buy any power from the generators, and so would be out of the game.

Secondly, by replacing the power orders, defined by Equation (12), in Equation (3), we obtain:

$$\begin{aligned} \alpha_{ki}(t) &= \frac{\frac{\nu_k^g(t)}{\tilde{p}_k(t)}\frac{n-1}{\delta}\beta_i}{\sum_{j=1,\ldots,n}\frac{\nu_k^g(t)}{\tilde{p}_k(t)}\frac{n-1}{\delta}\beta_j} \\ &= \frac{\beta_i}{\sum_{j=1,\ldots,n}\beta_j} = \beta_i \end{aligned}$$

This proves that the sharing coefficient $\alpha_{ki}(t)$ depends on neither generator g_k nor time instant t. Furthermore, the above result means that the total power delivered to microgrid $\mathcal{M}_i$ is: $\sum_{k=1,\ldots,K}\alpha_{ki}(t)\nu_k^g(t) = \beta_i\sum_{k=1,\ldots,K}\nu_k^g(t)$. As a result, the price and power orders for which the derivatives of $\pi_i(t)$ equal 0 verify the inequality defining Case 2 if, and only if:

$$\frac{b_0\theta_i - \gamma_i}{2b_1\theta_i\beta_i} - \frac{\nu_i^s(t)}{2\beta_i} \geq \sum_{k=1,\ldots,K}\nu_k^g(t) \tag{14}$$

This inequality states that the total generation should not be too high. If this is not the case, then the over-supply situation would probably end up with the most expensive generator out of the game.

If Inequalities (13) and (14) are true, the optimum for supplier s_i is reached for $p_i(t)$ defined by Equation (9) and $q_{ik}(t)$ defined by Equation (12). If one of these inequalities is not true, then the optimum for supplier s_i is reached on the frontier defined by Equation (7).

3.3 Optimization of the generators' decisions

After substituting $q_{ik}(t)$ and $C_k(t)$ by the expressions found Subsection 3.2.2 in generator g_k's utility, as defined in Equation (4), we obtain:

$$\tilde{\pi}_k(t) \quad = \quad \nu_k^g(t)\frac{n-1}{\delta}\sum_{i=1,...,n}\beta_i\Big[\frac{1}{\tilde{\gamma}_i}-\tilde{\gamma}_i\Big(\frac{1}{\tilde{p}_k(t)\tilde{\gamma}_i}-\frac{\delta}{n-1}\Big)_+\Big]$$

The only part of this equation depending on $\tilde{p}_k(t)$ always has a negative impact on the utility of the generator under the assumption of fair penalties. Indeed, in that case, as raised in the previous section, we obtain: $\beta_i \geq 0$ for all suppliers $(s_i)_i$. As a result, to maximize its utility, the generator has to choose $\tilde{p}_k(t)$ such that the part depending on $\tilde{p}_k(t)$ in the above equation equals 0. This implies that the term $\frac{1}{\tilde{p}_k(t)\tilde{\gamma}_i}-\frac{\delta}{n-1}$ is below 0 for all $i = 1, ..., n$. It is equivalent to: $\tilde{p}_k(t) \geq \frac{n-1}{\delta\tilde{\gamma}_i}$. Consequently, the optimal price for the generator with fair penalties should satisfy:

$$\tilde{p}_k(t) \geq \frac{n-1}{\delta \min\limits_{i=1,...,n}\{\tilde{\gamma}_i\}}$$

4 Distributed learning game

In this section, we assume that the generations from the microgrids $\Big(\nu_i^s(t)\Big)_i$ and the productions of the generators $\Big(\nu_k^g(t)\Big)_k$ are random individual sequences. As explained in the Introduction, this means that the underlying random processes generating the sequences do not necessarily have a probabilistic structure. They can be quite erratic [4].

To guarantee optimal system-wide operation, it is fundamental that the suppliers elaborate efficient learning strategies regarding the microgrids' generation and the generators' production. The risk associated with this learning task will be measured by the supplier's loss. It will be defined in Subsection 4.1.

Suppliers should optimize their prices and quantities ordered at each time period, at which point they possess no information on either the microgrids' generation or about the generators' decentralized production. As a result, the game can be considered as having incomplete information [21]. Each supplier s_i has to forecast $\nu_i^s(t)$ and $\nu_k^g(t)$ for all $k = 1, ..., K$, at each time period, in order to optimize its decisions. The game will be repeated over a finite time horizon $0 < T < +\infty$.

To simplify, we will consider a common closed space $\mathcal{E}_g$ of possible values for the production of each generator and a common closed space $\mathcal{E}_s$ of possible values for the generation from a microgrid. $\mathcal{E}_g$,$\mathcal{E}_s \subseteq \mathbb{R}$ are considered to be of finite dimension i.e., their cardinals $|\mathcal{E}_g|$ and $|\mathcal{E}_s|$ are such that $|\mathcal{E}_g| < +\infty$ and $|\mathcal{E}_s| < +\infty$. We will denote by $f_i(X, t)$ the forecast of supplier s_i about the variable X at time period t. We will use boldface type to denote vectors. We will also use the simplifying notations:

- $\mathbf{f_i(t)} \triangleq \Big(\underbrace{f_i(\nu_i^s, t)}_{\text{forecast of } \mathcal{M}_i\text{'s generation}} , \underbrace{f_i(\nu_1^g, t), ..., f_i(\nu_K^g, t)}_{\text{forecasts of the generators' productions}}\Big)$ to denote the predictions made by supplier s_i about the generation from microgrid $\mathcal{M}_i$ and about

the production of each generator $g_k,\ k = 1, ..., K$, at time period t

- $\mathbf{f(t)} \triangleq \Big(\mathbf{f_1(t)}, ..., \mathbf{f_n(t)}\Big)$ which contains the forecasts of all of the suppliers, at time period t
- $\mathbf{f_{-i}(y, t)} = \Big(\mathbf{f_1(t)}, ..., \mathbf{f_{i-1}(t)}, \mathbf{y}, \mathbf{f_{i+1}(t)}, ..., \mathbf{f_n(t)}\Big)$ which contains the forecasts of all of the suppliers except s_i whose prediction is set equal to $\mathbf{y}$, at time period t
- $\nu(\mathbf{t}) \triangleq \Big(\underbrace{\nu_1^s(t), ..., \nu_n^s(t)}_{\text{microgrids' generations}}, \underbrace{\nu_1^g(t), ..., \nu_K^g(t)}_{\text{generators' productions}} \Big)$ which contains the generation from each microgrid $\mathcal{M}_i$ and the production of each generator $g_k,\ k = 1, ..., K$, at time period t

Under conditions of power shortage as defined in Subsection 3.2.2, we obtain the optimal price for supplier s_i by substitution of the forecasters in Equation (9) and the optimal power orders for supplier s_i by substitution in Equation (12). The optimal decisions for supplier s_i at each time period t are then: $p_i(\mathbf{f_i(t)}, t) = \frac{\gamma_i + \theta_i \Big(b_0 - b_1 f_i(\nu_i^s, t)\Big)}{2}$ and $q_{ik}(\mathbf{f_i(t)}, t) = \frac{f_i(\nu_k^g, t)}{\tilde{p}_k(t)} \frac{\beta_i}{\tilde{\gamma}_i} \frac{n-1}{\delta}$. Therefore, the demands from the microgrids are $a_i(\mathbf{f_i(t)}, \nu(\mathbf{t})) = \frac{b_0}{2b_1} - \frac{\gamma_i}{2b_1\theta_i} - \nu_i^s(t) + \frac{f_i(\nu_i^s, t)}{2}$ because the microgrids have exact knowledge of their generation, unlike the suppliers, which have to forecast theirs. In addition, since suppliers may differ in their forecasts, α_{ki} can no longer be reduced to β_i, and instead we obtain $\alpha_{ki}(\mathbf{f(t)}) = \frac{f_i(\nu_k^g, t)\beta_i}{\sum\limits_{j=1,...,n} f_j(\nu_k^g, t)\beta_j}$. As a result, the utility of supplier s_i at each time period t is:

$$\begin{aligned}
\pi_i(\mathbf{f(t)}, \nu(\mathbf{t})) \;=\; & \frac{\gamma_i + \theta_i\Big(b_0 - b_1 f_i(\nu_i^s, t)\Big)}{2}\Big(\frac{b_0}{2b_1} - \frac{\gamma_i}{2b_1\theta_i} - \nu_i^s(t) + \frac{f_i(\nu_i^s, t)}{2}\Big) \\
- \; & \frac{\beta_i}{\tilde{\gamma}_i}\frac{n-1}{\delta}\sum_{k=1,...,K} f_i(\nu_k^g, t) - \gamma_i\Big(\frac{b_0}{2b_1} - \frac{\gamma_i}{2b_1\theta_i} - \nu_i^s(t) \\
+ \; & \frac{f_i(\nu_i^s, t)}{2} - \sum_{k=1,...,K} \frac{f_i(\nu_k^g, t)\beta_i}{\sum\limits_{j=1,...,n} f_j(\nu_k^g, t)\beta_j}\nu_k^g(t)\Big)_+
\end{aligned} \tag{15}$$

The game parameters and random events (fluctuating generations and penalties) are chosen to obtain a constant power shortage, in the sense that:

- Inequality (14) transposed to the incomplete information setting will always be true i.e.:

$$\frac{b_0}{2b_1} - \frac{\gamma_i}{2b_1\theta_i} - \frac{\nu_i^s(t)}{2} \geq \sum_{k=1,...,K} \alpha_{ki}(\mathbf{f(t)})\nu_k^g(t) \tag{16}$$

- The positivity constraint on $a_i(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t}))$ holds i.e., $\frac{b_0}{2b_1} - \frac{\gamma_i}{2b_1\theta_i} \geq \nu_i^s(t) - \frac{f_i(\nu_i^s,t)}{2}$

Lemma 1. *To maximize its utility, supplier s_i should be unbiased in its forecast of the generation from microgrid $\mathcal{M}_i$ and its forecast of generator productions.*

Proof of Lemma 1. Firstly, since the conditions are a constant power shortage, we have $\frac{\partial \pi_i(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t}))}{\partial f_i(\nu_i^s,t)} = -\frac{\theta_i b_1}{2} a_i(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t})) + \frac{p_i(\mathbf{f_i}(\mathbf{t}),t)}{2} - \frac{\gamma_i}{2} = \frac{\theta_i b_1}{2}\Big(\nu_i^s(t) - f_i(\nu_i^s,t)\Big)$. This derivative equals 0 if, and only if, $f_i(\nu_i^s,t) = \nu_i^s(t)$. In addition, $\frac{\partial^2 \pi_i(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t}))}{\partial f_i(\nu_i^s,t)^2} = -\frac{\theta_i b_1}{2} < 0$ which means that π_i is concave in $f_i(\nu_i^s,t)$ so its maximum is reached when $f_i(\nu_i^s,t) = \nu_i^s(t)$. Secondly, we have:

$$\frac{\partial \pi_i(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t}))}{\partial f_i(\nu_k^g,t)} = -\frac{\beta_i}{\tilde{\gamma}_i}\frac{n-1}{\delta} + \gamma_i \nu_k^g(t)\frac{\beta_i C_k(\mathbf{f}(\mathbf{t})) - \beta_i^2 f_i(\nu_k^g,t)}{C_k(\mathbf{f}(\mathbf{t}))^2}$$

where $C_k(\mathbf{f}(\mathbf{t})) \triangleq \sum_{j=1,\dots,n} f_j(\nu_k^g,t)\beta_j$. This derivative equals 0 if, and only if:

$$\beta_i f_i(\nu_k^g,t) = C_k(\mathbf{f}(\mathbf{t})) - \frac{1}{\gamma_i\tilde{\gamma}_i}\frac{n-1}{\delta}\frac{C_k(\mathbf{f}(\mathbf{t}))^2}{\nu_k^g(t)} \tag{17}$$

By summing this condition for all suppliers, we obtain: $C_k(\mathbf{f}(\mathbf{t})) = nC_k(\mathbf{f}(\mathbf{t})) - \frac{n-1}{\delta}\frac{C_k(\mathbf{f}(\mathbf{t}))^2}{\nu_k^g(t)}\sum_{j=1,\dots,n}\frac{1}{\gamma_j\tilde{\gamma}_j}$ which is equivalent to: $C_k(\mathbf{f}(\mathbf{t})) = \frac{C_k(\mathbf{f}(\mathbf{t}))^2}{\nu_k^g(t)}$ by definition of δ, which means $C_k(\mathbf{f}(\mathbf{t})) = \nu_k^g(t)$. By replacement in Equation (17), we obtain that the derivative equals 0 if, and only if: $f_i(\nu_k^g,t) = \nu_k^g(t)$. In addition, $\frac{\partial^2 \pi_i(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t}))}{\partial f_i(\nu_k^g,t)^2} = -2\gamma_i\beta_i\nu_k^g(t)\frac{\sum_{j\neq i}\beta_j f_j(\nu_k^g,t)}{C_k(\mathbf{f}(\mathbf{t}))^3} < 0$ which means that π_i is concave in $f_i(\nu_k^g,t)$ so its maximum is reached when $f_i(\nu_k^g,t) = \nu_k^g(t)$. □

4.1 Learning risk measure definition and first observations

As already mentioned, the supplier's risk, associated with the learning task, will be measured by its resulting loss. We have chosen a loss function representing the lack of profit compared to a case in which the supplier produces perfect forecasts of power demands and generations. More precisely, for any supplier $s_i, i = 1, ..., n$, its loss is defined as:

$$l_i\Big(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})\Big) = \Big(\pi_i^0(t) - \pi_i(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t}))\Big)$$

where $\pi_i^0(t)$ corresponds to supplier s_i's utility evaluated in $f_i(\nu_i^s,t) = \nu_i^s(t)$ and $f_i(\nu_k^g,t) = \nu_k^g(t)$ for any $k = 1, ..., K$.

We now upper bound supplier s_i's loss as the sum of a loss function depending exclusively on the supplier's predictions, $l_i^{(1)}$, and on another function, $l_i^{(2)}$, which relies

on the disagreements between all the suppliers' predictions. The notion of disagreement needs to be carefully explained. To that purpose, we introduce:

$$d_{ij}^k(t) \triangleq f_i(\nu_k^g, t) - f_j(\nu_k^g, t), \forall i, j = 1, ..., n, j \neq i, \forall k = 1, ..., K$$

as a measure of the disagreement between supplier s_i and supplier s_j, $i \neq j$, in the prediction of generator g_k's power production, at time period t.

Proposition 2. *Supplier s_i's loss can be upper bounded by the sum of two functions: the first, $l_i^{(1)}$, depends exclusively on its forecasts and the second, $l_i^{(2)}$, depends on its disagreement with the other suppliers' predictions:*

$$\begin{aligned} l_i\Big(\mathbf{f}(\mathbf{t})\nu(\mathbf{t})\Big) \quad \leq \quad & \underbrace{l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t})\Big)}_{\text{loss caused exclusively by supplier } s_i\text{'s forecasts}} \\ + \quad & \underbrace{l_i^{(2)}\Big((d_{ij}^k(t))_{j,k}, \nu(\mathbf{t})\Big)}_{\text{loss caused by the disagreements of supplier } s_i \text{ with the others' forecasts}} \qquad \forall i = 1, ..., n \end{aligned}$$

with

$$l_i^{(1)}\Big(f_i(t), \nu(t)\Big) = b_1\theta_i \frac{\Big(f_i(\nu_i^s, t) - \nu_i^s(t)\Big)^2}{4} - \frac{\beta_i}{\tilde{\gamma}_i}\frac{n-1}{\delta} \sum_{k=1,...,K} \Big(\nu_k^g(t) - f_i(\nu_k^g, t)\Big)$$

and

$$l_i^{(2)}\Big((d_{ij}^k(t))_{j,k}, \nu(\mathbf{t})\Big) = \gamma_i\beta_i \sum_{k=1,...,K} \Big(\frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\sum_{j=1,...,n} \beta_j \xi_1(d_{ij}^k(t))}\Big)$$

where $\xi_1(x) = 1 - \frac{x}{\min\{\mathcal{E}_g\}}\mathbf{1}_{x<0} - \frac{x}{\max\{\mathcal{E}_g\}}\mathbf{1}_{x\geq 0}$.

Proof of Proposition 2. The proof can be found in Appendix A.

4.2 Optimal learning strategies for each supplier

In this context of incomplete information on the power generations from the microgrids and on the fluctuating renewable productions, we test two regret criteria to build the suppliers' learning strategy, $d_t(.)$: external and internal regret minimization [4]. Both regret minimization algorithms give rise to an optimized learning strategy [4] i.e., a density function defined over the space $\mathcal{E}_s \times \mathcal{E}_g^K$. As explained in the Introduction, regret minimization is more robust to extreme events as it provides a density function over the prediction set. Other learning rules based on different regret criteria exist such as regret-matching [10] and regret-testing [8]. However, they offer no guarantee on the convergence of the algorithm and require longer times to reach an equilibrium when it exists.

The external regret over the sequence of time periods $1, ..., T$, is the difference between the observed cumulative loss and the cumulative loss of the best constant prediction i.e., pure strategy. To be more precise, for supplier s_i, it takes the form:

$$R_i(T) \triangleq \sum_{t=1}^{T} l_i\Big(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})\Big) - \min_{\mathbf{y} \in \mathcal{E}_s \times \mathcal{E}_g^K} \sum_{t=1}^{T} l_i\Big(\mathbf{f}_{-\mathbf{i}}(\mathbf{y}, \mathbf{t}), \nu(\mathbf{t})\Big)$$

We will consider that the learning strategy of supplier s_i is optimal if asymptotically its external regret remains in $o(T)$ where T is the number of time periods that have been played. This means that with probability 1: $\limsup_{T \to +\infty} \frac{1}{T} \sum_{t=1}^{T} R_i(t) = 0$. Forecasters satisfying these inequalities are said to be Hannan consistent [4].

By definition, a strategy $d_t(.)$ has a small internal regret if for every couple of predictions $\mathbf{y}, \mathbf{y}' \in \mathcal{E}_s \times \mathcal{E}_g^K$, the forecaster does not regret not having chosen prediction $\mathbf{y}'$ for time periods for which it chose prediction $\mathbf{y}$:

$$RI_i(T) \triangleq \max_{\mathbf{y}, \mathbf{y}' \in \mathcal{E}_s \times \mathcal{E}_g^K} \sum_{t=1,...,T} d_t(\mathbf{y}) \Big[l_i\Big(\mathbf{f}_{-\mathbf{i}}(\mathbf{y}, \mathbf{t}), \nu(\mathbf{t})\Big) - l_i\Big(\mathbf{f}_{-\mathbf{i}}(\mathbf{y}', \mathbf{t}), \nu(\mathbf{t}))\Big)\Big]$$

In a repeated game, this regret criterium ensures that the joint empirical frequencies of play converge with the set of correlated equilibria [4] whereas there is no guarantee that the product of the marginal empirical frequencies of play will converge with the Nash equilibria, under external regret minimization in a general game. Internal regret minimization will be used exclusively in the simulations in Subsection 5.3 as a benchmark to compare external regret minimization properties.

In the following lemma, we prove that it is possible to construct learning strategies for the suppliers that asymptotically minimize their external regret.

Lemma 3. *A Hannan consistent learning strategy exists for each supplier s_i.*

Proof of Lemma 3. In our case setting, at the end of each time period, supplier s_i knows[3] the power demand from microgrid $\mathcal{M}_i$ and it can infer its generation, $\nu_i^s(t)$, from Equation (5). Supplier s_i also knows the power that has been delivered to it by each generator g_k, from which it can infer the power that could have been delivered to it if it had ordered a different quantity $q_{ik}(t)$, all other suppliers ordering the same power quantities, using Equation (3). As a result, supplier s_i can calculate its loss for all of its possible actions. In [4], Cesa-Bianchi and Lugosi proved that a Hannan consistent learning strategy always exists if the agent can compute its loss for each possible action at the end of each time period. □

We now introduce lower and upper bounds on the disagreements between supplier s_i and the other suppliers regarding the predictions of the generations: $\underline{D}_{ss}(i) \triangleq \min_{t=1,...,T} \min_{j \neq i,k} d_{ij}^k(t)$ and $\overline{D}_{ss}(i) \triangleq \max_{t=1,...,T} \max_{j \neq i,k} d_{ij}^k(t)$. They contain the extreme disagreement values between the suppliers, about the estimated generations.

[3] In the smart grid, the monitoring is performed through communicating meters deployed at the end user level [2], [19].

Lemma 4. *If supplier s_i plays according to a Hannan consistent strategy then an upper bound exists for the external regret associated with supplier s_i's loss caused by its own predictions, $l_i^{(1)}$, which depends only on the extreme disagreement values between the suppliers regarding the estimated generations, $\underline{D}_{ss}(i)$ and $\overline{D}_{ss}(i)$. More precisely:*

$$\begin{aligned}\limsup_{T\to+\infty} \quad & \frac{1}{T}\Big[\sum_{t=1}^{T} l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}),\nu(\mathbf{t})\Big) - \min_{\mathbf{y}\in\mathcal{E}_s\times\mathcal{E}_g^K}\Big(\sum_{t=1}^{T} l_i^{(1)}\Big(\mathbf{y},\nu(\mathbf{t})\Big)\Big)\Big] \\ \leq \quad & \psi_i\Big(\underline{D}_{ss}(i),\overline{D}_{ss}(i)\Big)\sum_{k=1,\dots,K}\nu_k^g(t)\end{aligned}$$

where the function ψ_i from $\mathbb{R}^2$ to $\mathbb{R}$ is defined by:

$$\psi_i(x,y) = \gamma_i\beta_i\Big(\frac{1}{\xi_2(y)} - \frac{1}{\xi_1(x)} + \frac{\max\{\mathcal{E}_g\}}{\min\{\mathcal{E}_g\}} - \frac{\min\{\mathcal{E}_g\}}{\max\{\mathcal{E}_g\}}\Big)$$

with $\xi_2(x) = 1 - \frac{x}{\min\{\mathcal{E}_g\}}\mathbf{1}_{x\geq 0} - \frac{x}{\max\{\mathcal{E}_g\}}\mathbf{1}_{x<0}$.

Proof of Lemma 4. The proof can be found in Appendix B.

The aim of the next subsections will be to derive bounds for suppliers' losses under cooperative and non-cooperative scenarios.

4.3 Analysis of the upper bounds of the sum of suppliers' loss functions

We express the TSO's loss as the opposite of the sum of all of the suppliers' losses. This coincides with the balance price that the TSO would have to pay to ensure the reliability of the power network:

$$l\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \triangleq \sum_{i=1,\dots,n}\Big(\pi_i(t) - \pi_i^0(t)\Big)$$

It is also possible to consider that the suppliers play against Nature [4] which exhibits its worst behavior towards suppliers when setting the random individual sequences. Similarly to the suppliers, the TSO will try to keep its external regret $R(t)$ in $o(T)$.

We define $\tilde{l}_g$ as the sum of the suppliers' losses exclusively caused by their own predictions:

$$\tilde{l}_g\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \triangleq \sum_{i=1,\dots,n} l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}),\nu(\mathbf{t})\Big) \tag{18}$$

We let F_s be the set of all of the predictors (i.e., discrete density function set or alternatively, randomized prediction set) for each supplier and F_m the set of all of the predictors for the TSO. The value of the game, in which the suppliers exclusively consider the losses caused by their own predictions as utilities, is defined as follows:

$$\tilde{V}_g \triangleq \underbrace{\min_{\otimes_{i=1,\dots,n} d(\mathbf{f_i})\in F_s^n}}_{\text{suppliers' predictors}} \underbrace{\max_{d(\nu)\in F_m}}_{\text{TSO's predictor}} \mathbb{E}\Big[\tilde{l}_g(X,Y)|X\sim\otimes_{i=1,\dots,n} d(\mathbf{f_i}), Y\sim d(\nu)\Big]$$

where $\tilde{l}_g$ is defined in Equation (18).

Theorem 5. *Assume that all suppliers play according to Hannan consistent strategies for their loss upper bound then when T is large enough ($T \to +\infty$):*

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\tilde{l}_g\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \\ \leq\ & \tilde{V}_g + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1,\dots,n}\psi_i\Big(\underline{D}_{ss}(i),\overline{D}_{ss}(i)\Big)\sum_{k=1,\dots,K}\nu_k^g(t) \end{aligned}$$

Proof of Theorem 5. The proof can be found in Appendix C.

Corollary 6. *Assume that the TSO plays according to a Hannan consistent strategy for its loss upper bound. Then when T is large enough ($T \to +\infty$):*

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{l}_g\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \geq \tilde{V}_g - \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1,\dots,n}\psi_i\Big(\underline{D}_{ss}(i),\overline{D}_{ss}(i)\Big)\sum_{k=1,\dots,K}\nu_k^g(t)$$

Proof of Corollary 6. Applying Theorem 5 to the TSO i.e, by symmetry, considering that the TSO's loss upper bound is the opposite of the sum over i of supplier s_i's loss upper bounds, and using von Neuman-Morgenstern's Minimax Theorem [21] for $\tilde{V}_g$, we derive the proposed inequality.

□

We let:

$$l_g\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \triangleq \sum_{i=1,\dots,n} l_i\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \tag{19}$$

be the sum of the suppliers' losses. Using the definitions established in Equations (18) and (19), we derive the following inequality:

$$l_g\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) \leq \tilde{l}_g\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) + \sum_{i=1,\dots,n} l_i^{(2)}\Big((d_{ij}^k(t))_{j=1,\dots,n,k=1,\dots,K},\nu(\mathbf{t})\Big)$$

By substitution in Theorem 5, we obtain the following result:

Corollary 7. *If all suppliers play according to a Hannan consistent strategy for their loss upper bounds then, when T is large enough ($T \to +\infty$), their average loss cannot be greater than:*

$$\begin{aligned} \tilde{V}_g \ &+\ \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1,\dots,n}\psi_i\Big(\underline{D}_{ss}(i),\overline{D}_{ss}(i)\Big)\sum_{k=1,\dots,K}\nu_k^g(t) \\ &+\ \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1,\dots,n} l_i^{(2)}\Big((d_{ij}^k(t))_{j=1,\dots,n,k=1,\dots,K},\nu(\mathbf{t})\Big) \end{aligned}$$

whatever strategy is chosen by the TSO.

4.4 Collaborative learning strategy

Collaboration takes place within coalitions. In cooperative Game Theory literature, a coalition is a group of agents who have incentives to collaborate by sharing resource access, information, etc., in the hope of increasing their revenue, knowledge, social welfare (in case of altruism), etc., compared to a case where they behave non-cooperatively [21], [26]. Adapted to our learning context, we define coalitions of agents as follows:

Definition 8.
- *A coalition of suppliers is a group of suppliers which collaborate to learn the hidden productions of the generators* $\left(\nu_k^g(t)\right)_k$.
- *The grand coalition contains all the suppliers involved in the learning task i.e.,* $\{s_i\}_{i=1,\dots,n}$.
- *Cooperation takes place within the coalition when its members share their information and align their predictions to a common value.*

Shared information concerns only the power productions of the generators. Indeed, each supplier independently predicts the generation from its microgrid and has no impact on the other suppliers.

At this stage, the objective is to identify conditions on the disagreement levels between the suppliers regarding the forecasted power productions such that the term at the right of $\tilde{V}_g$ defined in Corollary 7, remains as small as possible. Indeed, the smaller the term defined in Corollary 7, the smaller the upper bound of the sum of the agents' losses will be.

Such a strategy would minimize $\psi_i\Big(\underline{D}_{ss}(i), \overline{D}_{ss}(i)\Big)$ and $l_i^{(2)}\Big((d_{ij}^k(t))_{j,k}, \nu(\mathbf{t})\Big)$ at any time period. This implies that $\underline{D}_{ss}(i) = \overline{D}_{ss}(i) = 0$ and $d_{ij}^k(t) = 0, \forall i, j, k, \forall t$.

This means that suppliers can decrease the upper bound of their average loss by coordinating their predictions about the power productions $\left(\nu_k^g(t)\right)_k$, at any time period t. Suppliers therefore have an incentive to form a grand coalition because it might enable them to decrease their total loss.

Proposition 9. *If the suppliers cooperate through a grand coalition and play Hannan consistent strategies, the suppliers' average loss over time interval* $[1;T]$ *when* T *is large enough (*$T \to +\infty$*) cannot be larger than:* $\tilde{V}_g + \sum_{i=1,\dots,n} \gamma_i\beta_i\Big(2\frac{\max\{\mathcal{E}_g\}}{\min\{\mathcal{E}_g\}} - \frac{\min\{\mathcal{E}_g\}}{\max\{\mathcal{E}_g\}} - 1\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1,\dots,K}\nu_k^g(t)\Big)$

Proof of Proposition 9. By substitution in $l_i^{(2)}$, as introduced in Proposition 2, since $d_{ij}^k(t) = 0$ for all i, j, k, we have:

$$l_i^{(2)}\Big((d_{ij}^k)_{j,k}(t), \nu(\mathbf{t})\Big) = \gamma_i\beta_i \sum_{k=1,\dots,K}\Big(\frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \nu_k^g(t)\Big) \tag{20}$$

This depends on the supplier index (i) and on time period (t), and not on the suppliers' forecasts. In addition, $\psi_i(0,0) = \gamma_i\beta_i\Big(\frac{\max\{\mathcal{E}_g\}}{\min\{\mathcal{E}_g\}} - \frac{\min\{\mathcal{E}_g\}}{\max\{\mathcal{E}_g\}}\Big)$. As a result, applying Corollary 7, we obtain that the suppliers' average loss over time interval $[1;T]$ when T is large enough ($T \to +\infty$) cannot be larger than:

$$\tilde{V}_g + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1,\ldots,n}\gamma_i\beta_i\Big(\frac{\max\{\mathcal{E}_g\}}{\min\{\mathcal{E}_g\}} - \frac{\min\{\mathcal{E}_g\}}{\max\{\mathcal{E}_g\}}\Big)\sum_{k=1,\ldots,K}\nu_k^g(t)$$
$$+ \quad \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1,\ldots,n}\gamma_i\beta_i\sum_{k=1,\ldots,K}\Big(\frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \nu_k^g(t)\Big) \tag{21}$$

Then the proposition statement is straightforward. □

5 Simulations

The aim of this section is to explain how the economic model of the hierarchical network, described in Section 2, can be applied in practice to take decisions in an uncertain context and then to check that the results derived analytically in Section 4 hold, for a given smart grid structure.

The rest of the section is organized as follows: Subsection 5.1 deals with payoff function estimation for each forecast, Subsection 5.2 elaborates on the update of mixed strategies for each forecast and in the last part we discuss the numerical illustrations that we have obtained, for a large sample of parameters, considering non-cooperative and cooperative scenarios.

5.1 Payoff functions

At each time period, each supplier must make $K+1$ forecasts: one for its microgrid power generation and one to evaluate the fluctuating production of each of the K generators. As a result, each supplier should define a randomized strategy on the space $\mathcal{E}_s \times \mathcal{E}_g^K$. We recall that a randomized strategy is the standard terminology used in Game Theory for a discrete density function defined over the considered set [21]. The size of the set grows very fast with K and, as a result, each probability in the randomized strategy of forecasts is very low, with the effect that errors are rounding off during computation. In order to overcome this issue, we decided to divide the suppliers into smaller entities, with each making only one forecast at each time period, and to consider that these entities are uncoupled. This trick results in $K+1$ randomized strategies in the space of forecasts $\mathcal{E}_s \times \mathcal{E}_g^K$ for each supplier.

For a given forecast X, we derive the payoffs for each value $x \in \mathcal{E}$ ($\mathcal{E} = \mathcal{E}_s$ for power generation from the microgrid and $\mathcal{E} = \mathcal{E}_g$ for power productions from the generators) of the forecast at each time period t by using the utilities of the suppliers and retaining only those terms that depend on forecast X. This is summarized in the following definition:

Definition 10. *The payoff function associated with forecast X, evaluated in $x \in \mathcal{E}$, coincides with the utility of supplier s_i restricted to its terms depending solely on forecast X and evaluated in x.*

For the forecasts of power generation from microgrid $\mathcal{M}_i$, supplier s_i's payoff takes the form:

$$H_{f_i(\nu_i^s)}(x,t) = \frac{\gamma_i + \theta_i(b_0 - b_1 x)}{2}\Big(\frac{b_0}{2b_1} - \frac{\gamma_i}{2b_1\theta_i} - \nu_i^s(t) + \frac{x}{2}\Big) - \gamma_i\frac{x}{2}$$

Concerning the forecasts of generator g_k's power production, supplier s_i's payoff takes the form:

$$H_{f_i(\nu_k^g)}(x,t) = -\frac{\beta_i}{\tilde{\gamma}_i}\frac{n-1}{\delta}x + \gamma_i\frac{\beta_i x}{\sum\limits_{j=1,\ldots,n,j\neq i}\beta_j f_j(\nu_k^g,t) + \beta_i x}\nu_k^g(t)$$

As already stated in Section 4, we will also consider that the TSO is non oblivious and tries to maximize the sum of the suppliers' losses. As for the suppliers, we uncouple $\nu_i^s(t)$ and $\nu_k^g(t)$ to improve the computation. More precisely the TSO's payoffs are:

$$H_{\nu_i^s}(x,t) = \Big(\frac{\theta_i(b_0 - b_1 f_i(\nu_i^s,t)) - \gamma_i}{2}\Big)x$$

and

$$H_{\nu_k^g}(x,t) = -\frac{\sum\limits_{i=1,\ldots,n}\gamma_i\beta_i f_i(\nu_k^g,t)}{\sum\limits_{j=1,\ldots,n}\beta_j f_j(\nu_k^g,t)}x$$

It is very straightforward to adapt the repeated learning game and payoffs when considering that the suppliers integrate a grand coalition. The grand coalition payoffs take the following forms:

$$\begin{aligned} H_{f_C(\nu_i^s)}(x,t) &= H_{f_i(\nu_i^s)}(x,t) \\ H_{f_C(\nu_k^g)}(x,t) &= -\sum_{i=1,\ldots,n}\frac{\beta_i}{\tilde{\gamma}_i}\frac{n-1}{\delta}x + \sum_{i=1,\ldots,n}\gamma_i\beta_i \end{aligned}$$

The TSO's payoff $H_{\nu_i^s}(x,t)$ is unchanged whereas payoff $H_{\nu_k^g}(x,t)$ becomes:

$$H_{\nu_k^g}(x,t) = -\sum_{i=1,\ldots,n}\gamma_i\beta_i x$$

5.2 Updates of forecasting strategies

We consider two types of update for the forecasting randomized strategies based on the exponential forecaster for signed games: one based on external regret and the other

based on internal regret. We assume that the game considered in this article is a signed game because the range of values of payoff function H_X might include a neighborhood of 0.

We let: $\vartheta_t \triangleq \sum_{s=1}^{t} Var\Big(H_X(X_s,s))\Big) = \sum_{s=1}^{t} \mathbb{E}\Big[\Big(H_X(X_s,s)) - \mathbb{E}[H_X(X_s,s)]\Big)^2\Big]$ be the sum of the variances associated with the random variable $H_X(X_t,t)$ under the mixed strategy $d_t(X)$ which is defined over space $\mathcal{E}$. Using the exponential forecaster for signed games with external regret [4] means that the mixed strategy is updated according to the algorithm described below.

External Regret Learning Algorithm: Updating of the Exponential Forecaster

---

Initialization. For $t=0$, we set: $w_0(x) = \frac{1}{|\mathcal{E}|},\ \forall x \in \mathcal{E}$.

Step 1 *to* T. The updating rules are the following:

$$\begin{aligned}
d_t(x) &= \frac{w_t(x)}{\sum_{x\in\mathcal{E}} w_t(x)},\ \forall x \in \mathcal{E} \\
w_{t+1}(x) &= \exp\Big(\eta_{t+1}\sum_{s=1}^{t} H_X(x,s)\Big) \\
&= w_t(x)^{\frac{\eta_{t+1}}{\eta_t}} \exp\Big(\eta_{t+1} H_X(x,t)\Big),\ \forall x \in \mathcal{E} \\
\eta_{t+1} &= \min\Big\{\frac{1}{2\max\{|H_X|\}}; \sqrt{\frac{2(\sqrt{2}-1)}{e-2}}\sqrt{\frac{\ln|\mathcal{E}|}{\vartheta_t}}\Big\} \\
\vartheta_{t+1} &= \vartheta_t + Var\Big(H_X(X_{t+1},t+1)\Big)
\end{aligned}$$

For the internal regret, the definition of which was introduced in Subsection 4.2, the updating rules are similar but with $d_t(.) = \sum_{y\neq y'} d_t^{y\to y'}(.)\Delta_{(y,y')}(t)$ where $d_t^{y\to y'}(.)$ is the modified forecasting strategy obtained when the forecaster predicts y' each time it would have predicted y and

$$\Delta_{(y,y')}(t) \triangleq \frac{\exp\Big(\eta_t \sum_{s=1}^{t-1}\sum_{x\in\mathcal{E}} d_s^{y\to y'}(x) H_X(x,s)\Big)}{\sum_{z\neq z'} \exp\Big(\eta_t \sum_{s=1}^{t-1}\sum_{x\in\mathcal{E}} d_s^{z\to z'}(x) H_X(x,s)\Big)}$$

We note that, if we take the notation $w_{(y,y')}(t) \triangleq \exp\Big(\eta_t \sum_{s=1}^{t-1}\sum_{x\in\mathcal{E}} d_s^{y\to y'}(x) H_X(x,s)\Big)$, then:

$$\begin{aligned}
\Delta_{(y,y')}(t) &= \frac{w_{(y,y')}(t)}{\sum_{z\neq z'} w_{(z,z')}(t)} \\
w_{(y,y')}(t) &= w_{(y,y')}(t-1)^{\frac{\eta_t}{\eta_{t-1}}} \exp\Big(\eta_t \sum_{x\in\mathcal{E}} d_{t-1}^{y\to y'}(x) H_X(x,t-1)\Big)
\end{aligned}$$

5.3 Numerical illustrations

Convergence times and emerging behaviors: We consider two suppliers and two generators ($n = 2, K = 2$). We compare the cumulative payoff of each agent (supplier or grand coalition) to the cumulative payoff of the same agent in a case where it has forecasted the best value at each time period in terms of payoffs. More precisely, we compute for each agent a (the supplier s_1, s_2 or the grand coalition C), the following performance metric:

$$R_a(T) \triangleq \frac{1}{T} \sum_{s=1}^{T} \sum_{X \in F_a} \Big(H_X(X_s, s) - \max_{x \in \mathcal{E}} (H_X(x, s)) \Big)$$

where F_a is the generic set of forecasts made by agent a. Then, we measure the convergence of the learning algorithm through the convergence of this performance metric. That is to say, we consider that convergence is reached when the variation of the performance metric, $\frac{R_a(T)-R_a(T-1)}{R_a(T-1)}$, is less than 10^{-2}.

We let $T^*_{s_i}, i = 1, 2$ (resp. T^*_C) be the number of time steps needed for the regret-based algorithm for supplier s_i (resp. the grand coalition) to converge. According to these notations, supplier s_i has incentives to cooperate if, and only if, $T^*_{s_i} \geq T^*_C$. Depending on the position of T^*_C with respect to $\min\{T^*_{s_1}; T^*_{s_2}\}$ and $\max\{T^*_{s_1}; T^*_{s_2}\}$ we identify three emerging behaviors:

- Both suppliers have incentives to cooperate if, and only if, $\min\{T^*_{s_1}; T^*_{s_2}\} \geq T^*_C$.
- The suppliers have no incentive to cooperate if, and only if, $\max\{T^*_{s_1}; T^*_{s_2}\} < T^*_C$
- The smart grid is unstable (one supplier having an incentive to cooperate and not the other) if, and only if, $\min\{T^*_{s_1}; T^*_{s_2}\} < T^*_C \leq \max\{T^*_{s_1}; T^*_{s_2}\}$

In our simulations, we calculated the convergence times of learning algorithms for a wide range of combinations of penalty coefficients (γ_1,γ_2,$\tilde{\gamma}_1$,$\tilde{\gamma}_2$). More precisely, in Figures 1 and 2 (a), we make the assumption that: $\gamma_1 = 2\gamma_2 = 2\gamma \in [0; 1]$ while in (b): $\gamma = \gamma_1 = \gamma_2 \in [0; 1]$. For each figure, we use 1000 combinations corresponding to 10 values equally distributed between 0.1 and 1 for γ, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$. We can easily check that all these penalty coefficient combinations satisfy Equation (13). In addition, we chose $\mathcal{E}_s = [5, 8]$ and $\mathcal{E}_g = [1, 2]$ so that Equation (16), i.e. energy shortage, is always true.

In Figure 1, the learning strategies of the suppliers and the grand coalition are based on external regret minimization while in Figure 2, they rely on internal regret minimization. In the top of Figures 1 and 2 (a) and (b), we plot the histograms of the maximum of $T^*_{s_1}, T^*_{s_2}$ (resp. T^*_C) left (resp. right) for all of the combinations of penalty coefficients (γ,$\tilde{\gamma}_1$,$\tilde{\gamma}_2$). The bin heights of each histogram are determined by the number of penalty coefficients that have the same convergence time. The algorithms are run for $T = 100$ time periods. At the bottom of Figures 1 and 2 (a) and (b), we plot the ratio of the maximum (resp. minimum) of $T^*_{s_1}$ and $T^*_{s_2}$ over T^*_C, left (resp. right), for all penalty coefficient combinations $(\gamma, \tilde{\gamma}_1, \tilde{\gamma}_2)$. From the top figures in both cases we observe that, for a far larger number of penalty coefficients, the learning algorithm

convergence times are shorter under cooperative scenarios than under non-cooperative scenarios. Furthermore, by comparison of Figures 1 and 2 top, the convergence times are shorter for learning algorithms based on internal regret minimization than for learning algorithms based on external regret minimization, though convergence occurs under both regret criteria.

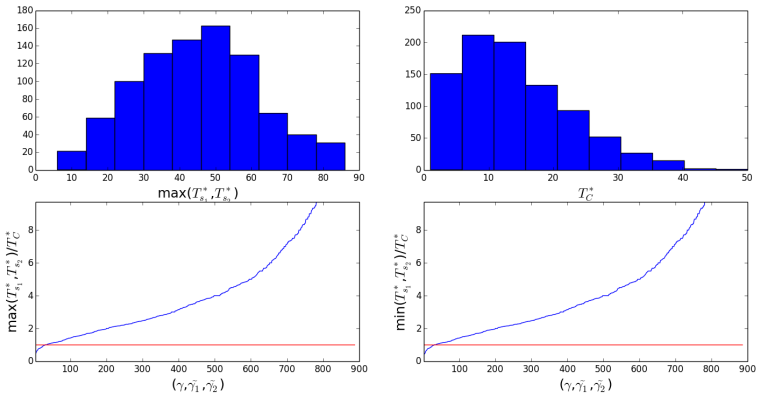
Regarding the potential for a grand coalition to emerge, we infer from Figure 1 (a) and (b) (resp. 2 (a) and (b)) bottom, that for 97% of the penalty coefficient combinations at least one supplier has incentives to cooperate and that for 95% (resp. 94.5%) of the combinations of penalty coefficients both suppliers have incentives to cooperate, using external regret minimization (resp. internal regret minimization) as criterion.

In terms of scalability, the complexity of our learning algorithm is in $O(nK)$. An interesting property is that it can be easily parallelized (one agent corresponding to one core) due to the fact that the weight updating rules are specific to each supplier (i.e., they do not depend on the other suppliers' forecasts).

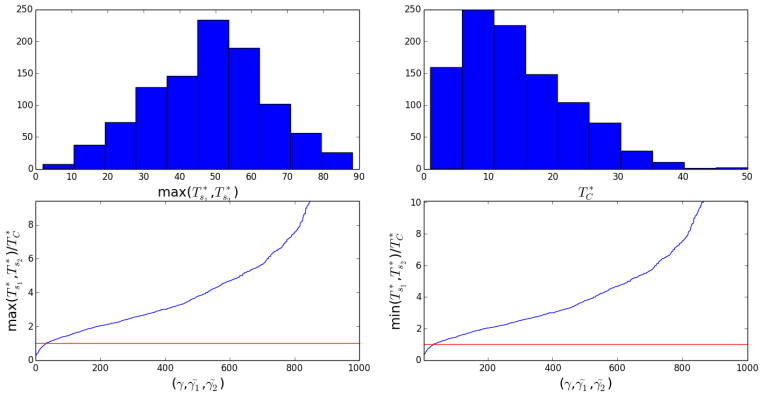
Tightness of convergence bounds: To measure performance, we chose to compute two upper bounds: firstly, the upper bound derived in Corollary 7, where the suppliers perform distributed learning non-cooperatively. This is called **BOUND SELFISH**. Secondly, the upper bound derived in Proposition 9, where the suppliers enter a grand coalition and align their forecasts of the generators' productions. This is called **BOUND COALITION**. Both upper bounds are matched with the sum of the suppliers' average loss: $\frac{1}{T}\sum_{t=1}^{T} l_g(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t}))$. The latter is computed under external (**AV. LOSS**$_{ext,.}$) and internal regret minimization (**AV. LOSS**$_{int,.}$), assuming that either the suppliers' learning process is non-cooperative or cooperative. These performance measures are averaged over 1000 combinations of parameters $(\gamma, \tilde{\gamma_1}, \tilde{\gamma_2})$, each parameter takes one of the 10 values equally distributed between 0.1 and 1. In Table 1, we compute **BOUND SELFISH**, **AV. LOSS**$_{ext,selfish}$ and **AV. LOSS**$_{int,selfish}$ for a fixed number of values of $\frac{\gamma_1}{\gamma_2}$. In Table 2, we compute **BOUND COALITION**, **AV. LOSS**$_{ext,coal}$ and **AV. LOSS**$_{int,coal}$ for a fixed number of values of $\frac{\gamma_1}{\gamma_2}$. According to both tables, we observe that **BOUND COALITION** is stricly smaller than **BOUND SELFISH** and that **AV. LOSS**$_{.,coal}$ is stricly smaller than **AV. LOSS**$_{.,selfish}$ whichever learning scenario (i.e., either non-cooperative or cooperative) is chosen by the suppliers.

6 Conclusion

In this article, we study a model of decentralized renewable generations in which generators, suppliers and microgrids are organized into a hierarchical network. Renewable generations are modeled by random individual sequences which need not have a probabilistic structure. This extraordinarily general demand and supply structure is capable of taking into account exogenous events. As a result, it is more robust to extreme events and appears particularly suitable for modeling fairly erratic processes such as renewables. We analytically determine the optimal prices that enable generators to avoid the

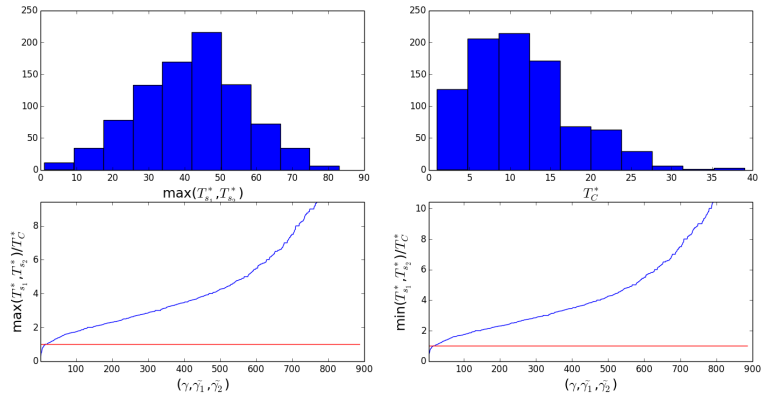


(a)

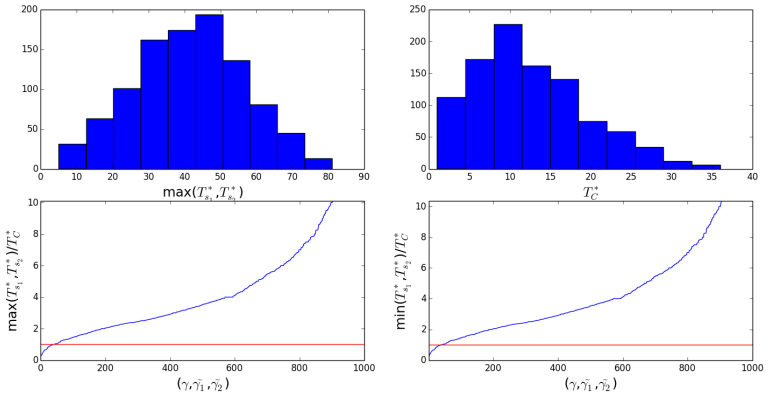


(b)

Figure 1: Convergence times and incentives to collaborate under *external regret* minimization. In (a), we have: $\gamma_1 = 2\gamma_2 = 2\gamma \in [0; 1]$ and in (b), we have: $\gamma = \gamma_1 = \gamma_2 \in [0; 1]$. At the top of each subfigure, we plot the histograms of the maximum of $T^*_{s_1}, T^*_{s_2}$ (resp. T^*_C) left (resp. right) for all the combinations of penalty coefficients and for a maximum number of time periods $T = 100$. At the bottom left (resp. right), we plot the ratio of the maximum (resp. minimum) of $T^*_{s_1}, T^*_{s_2}$ over T^*_C as a function of all of the penalty coefficient combinations.



(a)



(b)

Figure 2: Convergence times and incentives to collaborate under *internal regret* minimization. In (a), we have: $\gamma_1 = 2\gamma_2 = 2\gamma \in [0; 1]$ and in (b), we have: $\gamma = \gamma_1 = \gamma_2 \in [0; 1]$. At the top of each subfigure, we plot the histograms of the maximum of $T^*_{s_1}, T^*_{s_2}$ (resp. T^*_C) left (resp. right) for all of the penalty coefficient combinations and for a maximum number of time periods $T = 100$. At the bottom left (resp. right), we plot the ratio of the maximum (resp. minimum) of $T^*_{s_1}, T^*_{s_2}$ over T^*_C as a function of all the penalty coefficient combinations.

$\frac{\gamma_1}{\gamma_2}$	**BOUND SELFISH**	**AV. LOSS**$_{ext,selfish}$	**AV. LOSS**$_{int,selfish}$
0.2	$1.539.10^{-2}$	$-8.152.10^{-3}$	$-2.204.10^{-2}$
0.5	0.00	$-6.611.10^{-3}$	$-1.738.10^{-2}$
0.7	$1.634.10^{-2}$	$-7.150.10^{-3}$	$-1.604.10^{-2}$
1.0	$3.668.10^{-2}$	$-7.630.10^{-3}$	$-1.624.10^{-2}$
1.5	$6.627.10^{-3}$	$-4.395.10^{-3}$	$-1.719.10^{-2}$
2.0	0.0	$-3.124.10^{-3}$	$-1.552.10^{-2}$
2.5	$3.371.10^{-2}$	$-2.853.10^{-3}$	$-1.233.10^{-2}$
3.0	0.00	$-1.755.10^{-3}$	$-9.433.10^{-3}$
3.5	0.00	$-1.600.10^{-3}$	$-1.060.10^{-2}$
4.0	0.00	$-7.738.10^{-4}$	$-6.675.10^{-3}$

Table 1: Comparison of the upper bound derived in Corollary 7 with the suppliers' average loss in cases where the latter learn the generators' power productions non-cooperatively through external and internal regret minimization.

$\frac{\gamma_1}{\gamma_2}$	**BOUND COALITION**	**AV. LOSS**$_{ext,coal}$	**AV. LOSS**$_{int,coal}$
0.2	$5.579.10^{-3}$	$-4.853.10^{-2}$	$-5.991.10^{-2}$
0.5	0.00	$-5.694.10^{-2}$	$-6.490.10^{-2}$
0.7	$5.584.10^{-3}$	$-6.047.10^{-2}$	$-6.896.10^{-2}$
1.0	$1.297.10^{-2}$	$-5.114.10^{-2}$	$-7.176.10^{-2}$
1.5	$2.441.10^{-3}$	$-5.512.10^{-2}$	$-6.884.10^{-2}$
2.0	0.0	$-5.445.10^{-2}$	$-6.923.10^{-2}$
2.5	$1.181.10^{-2}$	$-4.258.10^{-2}$	$-5.213.10^{-2}$
3.0	0.00	$-3.940.10^{-2}$	$-4.371.10^{-2}$
3.5	0.00	$-2.839.10^{-2}$	$-3.757.10^{-2}$
4.0	0.00	$-2.384.10^{-2}$	$-2.740.10^{-2}$

Table 2: Comparison of the upper bound derived in Proposition 9 with the suppliers' average loss in cases where the latter enter a grand coalition and align their forecasts of the generators' power productions under external and internal regret minimization.

penalties that the balance operators threaten to apply when suppliers' orders are not entirely satisfied. All the risk is then transferred to the suppliers. Additionally, we prove that the latter can minimize their average risk by sharing information and aligning their forecasts. Finally, numerical simulations, run on a large sample of parameter combinations, lead us to observe that the convergence times in collaborative learning are clearly lower than times resulting from decentralized learning and that they are lower for learning algorithms based on internal regret minimization than for external regret minimization, though convergence occurs under both criteria. The tightness of convergence bound under collaborative learning is shown to be clearly better than for distributed learning.

An area of improvement is the design of the penalties paid to the electricity Transmission System Operator, which compensates negative energy balances. Is it possible

to design more generic mechanisms? Could rules be adapted to guarantee open markets and avoid speculation, such as capacity retention or under-investment in the means of production?

Appendix

Appendix A: Proof of Proposition 2

Following a number of calculations, we obtain:

$$\begin{aligned} & a_i(\nu(\mathbf{t}),\nu(\mathbf{t}))p_i(\nu(\mathbf{t}),t) - a_i(\mathbf{f_i}(\mathbf{t}),\nu(\mathbf{t}))p_i(\mathbf{f_i}(\mathbf{t}),t) \\ = \quad & \frac{f_i(\nu_i^s,t)-\nu_i^s(t)}{4}\Big(b_1\theta_i(f_i(\nu_i^s,t)-\nu_i^s(t)) - 2\gamma_i\Big) \end{aligned}$$

and

$$a_i(\nu(\mathbf{t}),\nu(\mathbf{t})) - a_i(\mathbf{f_i}(\mathbf{t}),\nu(\mathbf{t})) = \frac{\nu_i^s(t)-f_i(\nu_i^s,t)}{2}$$

Given supplier s_i's loss due to power shortage conditions, we have:

$$l_i\Big(\mathbf{f}(\mathbf{t}),\nu(\mathbf{t})\Big) = l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}),\nu(\mathbf{t})\Big) + \gamma_i\beta_i \sum_{k=1,\ldots,K} \frac{(\nu_k^g(t)-f_i(\nu_k^g,t))\nu_k^g(t)}{C_k(\mathbf{f}(\mathbf{t}))} \tag{22}$$

where:

$$l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}),\nu(\mathbf{t})\Big) = b_1\theta_i\frac{(f_i(\nu_i^s,t)-\nu_i^s(t))^2}{4} - \frac{\beta_i}{\tilde{\gamma}_i}\frac{n-1}{\delta}\sum_{k=1,\ldots,K}\Big(\nu_k^g(t)-f_i(\nu_k^g,t)\Big)$$

In addition, we observe that:

$$\begin{aligned} \Big(\frac{f_i(\nu_k^g,t)}{\sum_{j=1,\ldots,n}\beta_j f_j(\nu_k^g,t)}\Big)^{-1} \quad &= \quad \frac{\sum_{j=1,\ldots,n}\beta_j\Big(f_i(\nu_k^g,t)-d_{ij}^k(t)\Big)}{f_i(\nu_k^g,t)} \\ &= \quad \sum_{j=1,\ldots,n}\beta_j\Big(1-\frac{d_{ij}^k(t)}{f_i(\nu_k^g,t)}\Big) \\ &\leq \quad \sum_{j=1,\ldots,n}\beta_j\xi_1\Big(d_{ij}^k(t)\Big) \end{aligned}$$

where $\xi_1(x) = 1 - \frac{x}{\min\{\mathcal{E}_g\}}\mathbf{1}_{x<0} - \frac{x}{\max\{\mathcal{E}_g\}}\mathbf{1}_{x\geq 0}$. To obtain the inequality above, we followed the logic that since $\mathcal{E}_g$ is a close subset of $\mathbb{R}$, the forecasters are upper and lower bounded i.e., $\min\{\mathcal{E}_g\} \leq f_i(\nu_k^g,t) \leq \max\{\mathcal{E}_g\}$. Furthermore, $f_j(\nu_k^g,t) \geq \min\{\mathcal{E}_g\}$ means that:

$\frac{1}{\sum_{j=1,\dots,n} \beta_j f_j(\nu_k^g, t)} \leq \frac{1}{\min\{\mathcal{E}_g\}}$ since $\sum_{j=1,\dots,n} \beta_j = 1$. As a result, we have:

$$\begin{aligned} & l_i\Big(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})\Big) - l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t})\Big) \\ \leq \quad & \gamma_i \beta_i \sum_{k=1,\dots,K} \Big(\frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\sum_{j=1,\dots,n} \beta_j \xi_1(d_{ij}^k(t))} \Big) \end{aligned}$$

We then introduce the notation:

$$l_i^{(2)}\Big((d_{ij}^k)_{j,k}(t), \nu(\mathbf{t})\Big) = \gamma_i \beta_i \sum_{k=1,\dots,K} \Big(\frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\sum_{j=1,\dots,n} \beta_j \xi_1(d_{ij}^k(t))} \Big)$$

We obtain an upper bound of $l_i\Big(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})\Big)$ as the sum of two parts: the first, $l_i^{(1)}$, depends exclusively on its predictions and the second, $l_i^{(2)}$, depends on its interactions with the others' predictions.

□

Appendix B: Proof of Lemma 4

Suppose that supplier s_i plays according to a Hannan consistent strategy. Taking into account Equation (22) defining the loss of supplier s_i, this means that:

$$\begin{aligned} & \limsup_{T \to +\infty} \frac{1}{T} \Big[\sum_{t=1}^{T} \Big(l_i^{(1)}(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t})) + \zeta(f_i(t), \mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) \Big) \\ - \quad & \min_{y \in \mathcal{E}_s \times \mathcal{E}_g^K} \Big(\sum_{t=1}^{T} \Big(l_i^{(1)}(y, \nu(\mathbf{t})) + \zeta(y, \mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) \Big) \Big) \Big] \leq 0 \end{aligned} \tag{23}$$

where $\zeta(y, \mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) = \gamma_i \beta_i \sum_{k=1,\dots,K} \frac{\nu_k^g(t)^2}{C_k(\mathbf{f_{-i}}(\mathbf{y}, \mathbf{t}))} - \frac{y_{k+1} \nu_k^g(t)}{C_k(\mathbf{f_{-i}}(\mathbf{y}, \mathbf{t}))}$.

Let $d_{ij}^k(\mathbf{y}, t)$ denote the disagreement between supplier s_i and supplier s_j when supplier s_i makes the prediction $\mathbf{y}$ at time period t without any change in the predictions of the other suppliers. Following the same approach as in Appendix A, we obtain for all $\mathbf{y} \in \mathcal{E}_s \times \mathcal{E}_g^K$:

$$\frac{1}{\sum_{j=1,\dots,n} \beta_j \xi_1\Big(d_{ij}^k(\mathbf{y}, t)\Big)} \leq \frac{y_{k+1}}{C_k(\mathbf{f_{-i}}(\mathbf{y}, \mathbf{t}))} \leq \frac{1}{\sum_{j=1,\dots,n} \beta_j \xi_2\Big(d_{ij}^k(\mathbf{y}, t)\Big)}$$

where $\xi_2(x) = 1 - \frac{x}{\min\{\mathcal{E}_g\}} \mathbf{1}_{x \geq 0} - \frac{x}{\max\{\mathcal{E}_g\}} \mathbf{1}_{x < 0}$.

Similarly, we have:

$$\frac{1}{\max\{\mathcal{E}_g\}} \leq \frac{1}{C_k(\mathbf{f}_{-\mathbf{i}}(\mathbf{y},\mathbf{t}))} \leq \frac{1}{\min\{\mathcal{E}_g\}}$$

As a result, we obtain for all $\mathbf{y} \in \mathcal{E}_s \times \mathcal{E}_g^K$:

$$\gamma_i\beta_i \sum_{k=1,\ldots,K} \frac{\nu_k^g(t)^2}{\max\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\sum_{j=1,\ldots,n} \beta_j \xi_2(d_{ij}^k(\mathbf{y},t))} \leq \zeta(y, \mathbf{f}(\mathbf{t}), \nu(\mathbf{t}))$$

$$\leq \quad \gamma_i\beta_i \sum_{k=1,\ldots,K} \frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\sum_{j=1,\ldots,n} \beta_j \xi_1(d_{ij}^k(\mathbf{y},t))}$$

In Section 4, we introduced lower and upper bounds on the disagreements between supplier s_i and the other suppliers regarding the forecasts of the generations: $\underline{D}_{ss}(i)$ and $\overline{D}_{ss}(i)$. Since ξ_1 and ξ_2 are decreasing in x (they are linear functions by parties with negative coefficients), we have $\xi_1(\overline{D}_{ss}(i)) \leq \xi_1(x) \leq \xi_1(\underline{D}_{ss}(i))$ and $\xi_2(\overline{D}_{ss}(i)) \leq \xi_2(x) \leq \xi_2(\underline{D}_{ss}(i))$ for any disagreement value x.

Then, since $\sum_{j=1,\ldots,n} \beta_j = 1$, the lower bound of $\zeta(y, \mathbf{f}(\mathbf{t}), \nu(\mathbf{t}))$ is: $b_l(i,t) = \gamma_i\beta_i \sum_{k=1,\ldots,K} \frac{\nu_k^g(t)^2}{\max\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\xi_2(\overline{D}_{ss}(i))}$. Whereas, the upper bound takes the form: $b_u(i,t) = \gamma_i\beta_i \sum_{k=1,\ldots,K} \frac{\nu_k^g(t)^2}{\min\{\mathcal{E}_g\}} - \frac{\nu_k^g(t)}{\xi_1(\underline{D}_{ss}(i))}$. If Inequality (23) is checked, then the following inequality holds:

$\limsup_{T\to+\infty} \frac{1}{T}\Big[\sum_{t=1}^{T} l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t})\Big) + \sum_{t=1}^{T} b_l(i,t) - \min_{\mathbf{y}\in\mathcal{E}_s\times\mathcal{E}_g^K} \Big(\sum_{t=1}^{T} l_i^{(1)}\Big(\mathbf{y}, \nu(\mathbf{t})\Big) - \sum_{t=1}^{T} b_u(i,t)\Big)\Big] \leq 0$. This last inequality provides an upper bound for the external regret associated with supplier s_i's partial loss.

□

Appendix C: Proof of Theorem 5

With the proposed expression of ψ_i, the upper bound of the external regret evaluated in supplier s_i's loss, $l_i^{(1)}$, becomes:

$$\limsup_{T\to+\infty} \frac{1}{T}\Big[\sum_{t=1}^{T} l_i^{(1)}\Big(\mathbf{f_i}(\mathbf{t}), \nu(\mathbf{t})\Big) - \min_{y_i\in\mathcal{E}} \sum_{t=1}^{T} l_i^{(1)}\Big(y_i, \nu(\mathbf{t})\Big)\Big]$$

$$\leq \quad \limsup_{T\to+\infty} \frac{1}{T}\psi_i\Big(\underline{D}_{ss}(i), \overline{D}_{ss}(i)\Big) \sum_{t=1}^{T} \sum_{k=1,\ldots,K} \nu_k^g(t) \qquad (24)$$

Summing Inequality (24) over all $i = 1, ..., n$, the external regret evaluated in the sum of the suppliers' losses $(l_i^{(1)})_{i=1,...,n}$, becomes:

$$\lim \sup_{T\to+\infty} \frac{1}{T}\Big[\sum_{t=1}^{T} \tilde{l}_g\Big(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})\Big) - \min_{\mathbf{f}} \sum_{t=1}^{T} \tilde{l}_g\Big(\mathbf{f}, \nu(\mathbf{t})\Big)\Big]$$
$$\leq \lim \sup_{T\to+\infty} \frac{1}{T} \sum_{i=1,...,n} \psi_i\Big(\underline{D}_{ss}(i), \overline{D}_{ss}(i)\Big) \sum_{t=1}^{T} \sum_{k=1,...,K} \nu_k^g(t)$$

We assume that each supplier makes its forecasts independently of the other suppliers. Then $\tilde{l}_g\Big(X, \nu(\mathbf{t})\Big)$ is linear in X. As a result, its minimum over the simplex of probability vectors is reached in one of the corners of the simplex. Consequently:

$$\min_{\mathbf{f}} \frac{1}{T} \sum_{t=1}^{T} \tilde{l}_g\Big(\mathbf{f}, \nu(\mathbf{t})\Big)$$
$$= \min_{\otimes_i d(\mathbf{f_i}) \in F_s^n} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Big[\tilde{l}_g(X, Y) | X \sim \otimes_i d(\mathbf{f_i}), Y = \nu(\mathbf{t})\Big]$$

Let: $d_T(\mathbf{z}) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{\nu(\mathbf{t})=\mathbf{z}\}}$ be the marginal empirical frequency of play evaluated in prediction $\mathbf{z} \in \mathcal{E}_s^n \times \mathcal{E}_g^K$. We obtain:

$$\min_{\otimes_i d(\mathbf{f_i}) \in F_s^n} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Big[\tilde{l}_g(X, Y) | X \sim \otimes_i d(\mathbf{f_i}), Y = \nu(\mathbf{t})\Big]$$
$$= \min_{\otimes_i d(\mathbf{f_i}) \in F_s^n} \sum_{\mathbf{z} \in \mathcal{E}_s^n \times \mathcal{E}_g^K} d_T(\mathbf{z}) \mathbb{E}\Big[\tilde{l}_g(X, Y) | X \sim \otimes_i d(\mathbf{f_i}), Y = z\Big]$$
$$= \min_{\otimes_i d(\mathbf{f_i}) \in F_s^n} \mathbb{E}\Big(\tilde{l}_g(X, Y) | X \sim \otimes_i d(\mathbf{f_i}), Y \sim d_T\Big)$$
$$\leq \underbrace{\max_{d(\nu) \in F_m} \min_{\otimes_i d(\mathbf{f_i}) \in F_s^n} \mathbb{E}\Big[\tilde{l}_g(X, Y) | X \sim \otimes_i d(\mathbf{f_i}), Y \sim d(\nu)\Big]}_{\tilde{V}_g}$$

□

References

[1] Albadi M. H., El-Saadany E. F., Demand Response in the Electricity Markets: An Overview, in *proc. IEEE Power Engineering Society General Meeting*, pp. $1 - -5$, 2007

[2] Atzeni I., Ordonez L. G., Scutari G., Palomar D. P., Fonollosa J. R., Demande-Side Management via Distributed Energy Generation and Storage Optimization, *IEEE Transactions on Smart Grid*, vol. 4, pp. $866 - -876$, 2013

[3] Bossavy A., Girard R., Kariniotakis G., Forecasting Ramps of Wind Power Production with Numerical Weather Prediciton Ensembles, *Wind Energy*, vol. 16, pp. $51 - -63$, 2013

[4] Cesa-Bianchi N., Lugosi G., Prediction, Learning, And Games, *Cambridge University Press*, 2006

[5] Costa A., Crespo A., Navarro J., Lizcano G., Madsen H., Feitosa E., A Review on the Young History of the Wind Power Short-Term Prediction, *Renewable and Sustainable Energy Reviews*, vol. 12, pp. $1725 - -1744$, 2008

[6] Creti A., Fabra A., Supply Security and Short-Run Capacity Markets for Electricity, *Energy Economics*, vol. 29, pp. $259 - -276$, 2007

[7] Fahrioglu M., Alvarado F. L., Designing Cost Effective Demand Management Contracts using Game Theory, in *proc. IEEE Power Engineering Society Winter Meeting*, vol. 1, pp. $427 - -432$, 1999

[8] Foster D. P., Peyton H., Learning, Hypothesis testing, and Nash equilibrium, *Games and Economic Behavior*, vol. 45, pp. $73 - -96$, 2003

[9] Giebel G., Kariniotakis G., Brownsword R., The State of the Art on Short-Term Wind Power Prediction: A Literature Overview, *Working Paper*, ANEMOS EU Project, 2003

[10] Hart S., Mas-Colell A., A Simple Adaptive Procedure Leading to Correlated Equilibrium, *Econometrica*, vol. 68, pp. $1127 - -1150$, 2000

[11] Holsapple C. W., Lee A., Otto J., A Machine Learning Method for Multi-Expert Decision Support, *Annals of Operations Research*, vol. 75, pp. $171 - -188$, 1997

[12] Le Cadre H., Auliac C., Energy Demand Prediction in a Charge Station: A Comparison of Statistical Learning Approaches, in *proc. of the* 2*-nd European Electric Vehicle Congress (EEVC)*, 2012

[13] Le Cadre H., Bedo J.-S., Distributed Learning in Hierarchical Networks, in *proc. of the* 6*-th International Conference on Performance evaluation Methodologies and Tools (ValueTools)*, pp. $188 - -197$, 2012

[14] Le Cadre H., Mercier D., Is Energy Storage an Economic Opportunity for the Eco-Neighborhood?, *NETNOMICS: Economic Research and Electronic Networking*, vol. 13, pp. $191 - -216$, 2012

[15] Li N., Marden J. R., Decoupling Coupled Constraints through Utility Design, *IEEE Transactions on Automatic Control*, pp. $2289 - -2294$, Online $08/2014$

[16] Li N., Chen L., Low S. H., Optimal Demand Response Based on Utility Maximization in Power Networks, in *proc. IEEE Power and Energy Society General Meeting*, pp. $1 - -8$, 2011

[17] Ljungqvist L., Sargent T. J., Recursive Macroeconomic Theory, *MIT Press*, 2-nd Edition, 2007

[18] Marden J. R., Young H. P., Pao L. Y., Achieving Pareto Optimality Through Distributed Learning, *Discussion Paper Series*, University of Oxford, Department of Economics, 2011

[19] Mohsenian-Rad A.-H., Wong V. W. S., Jatskevich J., Schober R., Leon-Garcia A., Autonomous Demand-Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid, *IEEE Transactions on Smart Grid*, vol. 1, pp. $320 - -331$, 2010

[20] Morales A. J., On the Role of the Group Composition for Achieving Optimality, *Annals of Operations Research*, vol.137, pp. $387 - -397$, 2005

[21] Myerson R., Game Theory: An Analysis of Conflict, *Harvard University press*, 2006

[22] Nair J., Adlakha S., Wierman A., Energy Procurement Strategies in the Presence of Intermittent Sources, *Working Paper*, California Institute of Technology, 2013

[23] Nguyen T., Vojnović M., Weighted Proportional Allocation, in *proc. of ACM Sigmetrics*, 2011

[24] Pradelski B. R., Young H. P., Learning efficient Nash Equilibria in distributed systems, *Working Paper*, University of Oxford, Department of Economics, 2010

[25] Saad W., Han Z., Debbah M., Hjorrungnes A., Başar T., Coalitional Game Theory for Communication Networks: A Tutorial, *IEEE Signal Processing Magazine*, Special Issue on Game Theory, vol. 26, pp. $77 - -97$, 2009

[26] Saad W., Han Z., Poor V. H., Coalitional Game Theory for Cooperative Micro-Grid Distribution Networks, in *proc.* 2*-nd IEEE International Workshop on Smart Grid Communications*, 2011

[27] Shapley L. S., Stochastic Games, in *proc. National Academy of Sciences of the United States of America*, vol. 39, pp. $1095 - -1100$, 1953

[28] Zheng H., Kulkarni S. R., Poor H. V., A Sequential Predictor Retraining Algorithm and its Application to Market Prediction, *Annals of Operations Research*, vol. 208, pp. $209 - -225$, 2013

[29] Rules for Balance Operators, http://www.cre.fr/operateurs/responsables-d-equilibre, Online $09/2014$

[30] Nome law, http://www.cre.ft/dossiers/la-loi-nome, Online $08/2014$