



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

Hélène Le Cadre, Jean-Sébastien Bedo

## ► To cite this version:

Hélène Le Cadre, Jean-Sébastien Bedo. Dealing with Uncertainty in the Smart Grid: A Learning Game Approach. 2013. hal-00740893v5

**HAL Id: hal-00740893**

**<https://hal.science/hal-00740893v5>**

Preprint submitted on 16 Dec 2013 (v5), last revised 15 Mar 2016 (v9)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Hélène Le Cadre<sup>†</sup>

Jean-Sébastien Bedo<sup>‡</sup>

## Abstract

In this article, the smart grid is modeled as a decentralized and hierarchical network, made of three categories of agents: producers, providers and microgrids. To optimize their decisions concerning the energy prices and the traded quantities of energy, the agents need to forecast the energy productions and the demand of the microgrids. The biases resulting from the decentralized learning might create imbalances between demand and supply, leading to penalties for the providers and for the producers. We determine analytically prices that provide to the producers a guarantee to avoid such penalties, reporting all the risk on the providers. Additionally, we prove that collaborative learning, through a grand coalition of providers where information is shared and forecasts aligned on a single value, minimizes their average risk. Simulations, run for a large sample of parameter combinations, lead us to observe that the convergence times of the collaborative learning strategy are clearly superior to times resulting from distributed learning, using external and internal regret minimization. Furthermore, a grand coalition has 98% (resp. 85%) of chances to emerge under internal (resp. external) regret minimization.

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

## 1 Introduction

In Europe, and in France especially, traditional electrical networks rely on nuclear based energies, which are non renewable energy sources. With such energies, the production level can be adapted by the plant operator who alternates openings and closings and optimizes the duration of the switches between both modes. The objective then, is to adapt the production level so as to meet the uncertain demand level. We built a first model in [12], where two learning strategies based on tit for tat and fictitious play are used to adapt the production level to meet the demand level. For renewable energies, the production level can only be partially controlled, for instance, by lowering the wind turbine speed [16]. Renewable energy integration in the electrical network requires the

---

\*The authors thank the Editor and the two anonymous referees for helpful remarks and careful proofreading.

<sup>†</sup>MINES ParisTech, Centre for Applied Mathematics, Sophia Antipolis, FRANCE Email: [helene.le\\_cadre@mines-paristech.fr](mailto:helene.le_cadre@mines-paristech.fr)

<sup>‡</sup>Orange Labs, Paris, FRANCE Email: [jeansebastien.bedo@orange.com](mailto:jeansebastien.bedo@orange.com)

deployment of smart Information and Communication Technologies (ICTs), to supervise the grid operations [17]. Indeed, renewable energy production is highly unpredictable since it depends on uncontrollable exogenous factors like wind power, sunshine level, swell intensity, etc. [20]. Furthermore, the new active role of the end users, who can become energy producers and adapt dynamically their consumption while falling into a multitude of microgrids [13], [23], [24], dramatically increase the volume of the exchanged data flows. ICTs appear as a mean to retrieve the most salient information from this big amount of data and to train forecasters to provide efficient predictions about the renewable energy production and about the microgrid energy demand. These predictions will then be used as inputs to optimize the smart grid operations [2].

In practice, it is increasingly apparent that current forecasting technology cannot properly handle extreme situations corresponding to either extreme 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 of expected productions. Only recently, probabilistic methods have been introduced to provide estimations of the entire distribution of future productions [3]. In such methods, forecasts may take the form of either quantile estimations or density estimations [5], [8]. Learning based on regret minimization, as described in [4], belongs to this latter category. This class of methods is particularly efficient [11]. It provides to the forecaster a density function which associates a weight to each possible outputs. The density function is updated by merging informations from various sources' reports. As a result, these methods are more robust to extreme events and appear as particularly well suited to model erratic processes such as renewable energy production.

In the framework of the smart grid, learning is performed in a decentralized manner since each agent primarily learns the hidden information using his own observations. The existing literature on distributed learning primarily focuses on distributed learning algorithms that are suitable for implementation in large scale engineering systems [14], [22]. The results mainly concentrate on a specific class of games, called games of potential [25]. 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, there exist some limitations to this framework. The most striking one is that it is frequently impossible to represent the interaction framework of a given system as a potential game [16]. The learning game studied in this paper belongs to the category of repeated uncoupled games since one player cannot predict the forecasts and so actions of the other agents at a given time period. To take his decision i.e., optimal prices and energy orders, each agent is aware of the history of forecasts of all the agents and of his utility. Recent work has shown that for finite games with generic payoffs there exist completely uncoupled learning rules i.e., rules where the agents observe only their own prediction history and their utility, that lead to Nash equilibria that are Pareto optimal [22]. Marden et al. exhibit a different class of learning procedures that lead to Pareto optimal vector of actions that do not necessarily coincide with Nash equilibria [16]. 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 [26]. The updating of the individual forecasters to accommodate 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 the competition among the distinct expert systems in [10].

Most collaborative mechanisms studied in the literature lead to price or quality of service alignment. Besides, the group composition provides an additional state space in which information about the environment can be accumulated [18]. To our knowledge the impact of collaboration through information sharing and forecast alignment, while prices are individually determined, on the underlying system performance has not been studied so far. Of course, collaboration might not emerge due to the agents' natural incentives to cheat and to deviate from the cooperative equilibrium and also, most frequently, due to the regulator's intervention. There is a number of well-understood reasons why regulator often does not allow horizontal collaboration: if providers are allowed to collaborate, they might cooperate to raise the price i.e., reducing quantity below the efficient baseline, and create market power [6]. Alternatively, providers might cooperate to reduce quality of service. Courts punish explicit agreements whose objectives are clearly to decrease the competition. In this article, we will answer the following questions:

- How will the biases, introduced by the errors made by the agents in their predictions, affect the agents' average risk?
- Will collaborative learning improve the smart grid wide performance and should 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. Then the complete information Stackelberg game is solved in Section 3, proceeding by backward induction. We derive analytically the optimal prices and energy orders for the agents. Partial information is introduced in Section 4 where the interacting agents learn in a distributed fashion hidden individual sequences.

To illustrate the theoretical results derived in the previous sections, we compare in Section 5: firstly, the times of convergence of providers' learning strategies under external and internal regret minimization, in cooperative and non-cooperative scenarios ; secondly, their incentives to enter a grand coalition depending on the parameter values.

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

## 2 The model

The number of agents interacting in the smart grid is large. In this article, we model the smart grid as a three layer hierarchical network which evolution depends on the interactions between the agents composing each layer and also, on the ability of the agents to cope with energy production and demand variations [2], [17]. 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 optimization program in Subsection 2.2.

## 2.1 Description of the agents

We model the smart grid through three categories of agents: the microgrids, the providers and the energy producers. The microgrids generate some energy demand, and fulfill it either by buying energy to a provider or by finding alternatives (solar panels, more efficient appliances, etc.). Furthermore, the demand can be significantly flattened through end user changes in their normal electrical consumption patterns in response to changes in the price of electricity over time [1]. These online changes are called pricing demand response (DR) in the literature [13]. Each provider buys energy to several energy producers and resell it to the end users. Each energy producer produces and sells energy to all providers.

We assume that each end user contracts with only one provider and does not churn from one provider to another during all the period of our study. This assumption holds well if we consider local or regional utility companies for example. In this sense, the set made of end users supplied by a single provider can be seen as an individual microgrid, as defined in [13], [23], [24] and recalled in the Introduction. We denote by  $(s_i)_i$ , with  $i$  varying between 1 and  $n$ , the  $i$ -th provider and by  $\mathcal{M}_i$  the corresponding group of end users. The energy producers are denoted  $(e_k)_k$  with  $k$  varying between 1 and  $K$ . The energy producers can be associated with nuclear plants, photovoltaic park managers, wind farm administrators, etc. In this article, we assume that the energy producer cannot influence directly the energy he produces at each time period. This assumption holds well if we look at renewable energy sources like a wind turbine farm without any investment in an additional wind turbine during the study period. The variation of the wind intensity will impact the amount of the produced energy without any lever for the energy producer<sup>1</sup>.

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

---

### Basic Game Description $G(t)$

---

- (1) The energy producers  $e_k$  communicate their unitary prices  $\tilde{p}_k(t) > 0$  to the providers. The energy prices are fixed independently and simultaneously by each energy producer so as to maximize his profit.
  - (2) The providers  $s_i$  place energy quantity orders to energy producers: the quantity ordered by  $s_i$  to  $e_k$  is denoted by  $q_{ik}(t)$ . The providers  $s_i$  communicate their prices  $p_i(t) > 0$  for one energy unit to their microgrid. The orders and the energy prices are fixed independently and simultaneously by each provider so as to maximize his profit.
  - (3) Microgrid  $\mathcal{M}_i$  net demand, defined as the difference between its demand and the production of its own alternative means, reaches  $\nu_i^s(t)$  energy units for the time period. It decides to perform DR by postponing  $a_i(t)$  energy units and buys the rest  $\left(\nu_i^s(t) - a_i(t)\right)_+$  to provider  $s_i$ . The quantity of DR is chosen so as to minimize the total cost of energy for  $\mathcal{M}_i$ .
- 

<sup>1</sup>Non-renewable energy producers like nuclear plants might be integrated into the grid. It requires to use distributed control rules as the ones described in [12], [16].

At each time period, the production of energy producer  $e_k$  reaches  $\nu_k^e(t)$  energy units. He then delivers  $\alpha_{ki}(t)\nu_k^e(t)$  energy units to provider  $s_i$  where  $\alpha_{ki}(t) \geq 0$  denotes the proportion of his production that producer  $e_k$  allocates to provider  $s_i$ , with the normalization constraint:  $\sum_{i=1,\dots,n} \alpha_{ki}(t) = 1$ . This proportion is defined depending on quantity orders received by  $e_k$  from all providers. The sum of the orders received by  $e_k$  may exceed  $\nu_k^e(t)$  and so the quantity of energy received by each provider may be inferior to his quantity order.

Penalties are undertaken by both providers and energy producers if they cannot satisfy the entire demand of their customers. Provider  $s_i$  incurs a cost  $\gamma_i > 0$  per missing energy unit for his microgrid, measured a posteriori. It is paid to the transmission system operator for electricity (TSO). In France, the TSO has defined some rules to give incentives to the agents to become balance operators [27]. According to these rules, a negative energy balance on the free market must be compensated by buying the missing energy to the TSO with a unit price defined through the adjustment mechanism. This adjustment mechanism price is higher than the free market energy unit price. The whole adjustment mechanism is implemented by the TSO who compensates the negative balances to ensure the electrical network reliability. In our article, the adjustment price is different for each provider. 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 isolated ones due to the high cost of electricity transmission. As a result,  $\gamma_i$  is higher for providers serving isolated locations than over densely interconnected areas. Producer  $e_k$  incurs a cost  $\tilde{\gamma}_i > 0$  per missing energy unit for provider  $s_i$ , measured a posteriori. It is paid to the regulator of the capacity market that should be implemented to balance supply and demand in the smart grid [28]. Indeed, to guarantee the reliability of the capacity market, it might be necessary to implement a feedback mechanism where the regulator compensates the negative energy balances of the producers by investing himself in capacity [28]. The costs of these investments would be recovered from the penalties imposed to the producers. These investments are spread over relatively long periods ; in the short term, the regulator must call the TSO whose share of imports from neighboring energy markets re-balances the level of production at the providers' demand.

## 2.2 Optimization program for each agent

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

### 2.2.1 Programs of the microgrids

The only decision variable for microgrid  $\mathcal{M}_i$  is the quantity of energy that it decides to postpone through DR:  $a_i(t)$ . We assume that the microgrid has no lever to influence its random net demand, defined as the difference between its demand and the random production from its alternative sources:  $\nu_i^s(t)$ .

Flattening the energy demand by postponing or even deleting one's demand, rather than buying it to the provider, has some cost for the microgrid. More precisely, the cost of shaving  $a_i(t)$  energy units through DR, is measured through the end users' hyperbolic absolute risk aversion [15], resulting from the possibility that their postponed demand might not be fulfilled later on. As a special case of hyperbolic absolute risk aversion measure, we model the associated utility as a quadratic cost function. As a result, the total cost of energy for microgrid  $\mathcal{M}_i$  is:

$$p_i(t) \left( \nu_i^s(t) - a_i(t) \right) + c \left( a_i(t) \right) \quad (1)$$

where  $c \left( a_i(t) \right) = \frac{a_i(t)^2}{2}$ , according to the assumptions made on the end users' absolute risk aversion. This hypothesis is not restrictive and constants or more complicated cost functions can be introduced. The main advantage of this choice is that it is generic enough and allows to solve a large part of the game analytically.

Microgrid  $\mathcal{M}_i$  chooses  $a_i(t)$  in order to minimize its total energy cost. Therefore, its optimization program is of the form:  $\min_{a_i(t) \geq 0} \left\{ p_i(t) \left( \nu_i^s(t) - a_i(t) \right) + c \left( a_i(t) \right) \right\}$ . Its decision depends on the energy price  $p_i(t)$  fixed by provider  $s_i$ .

### 2.2.2 Programs of the providers

The decision variables for each provider  $s_i$  are the unit energy price  $p_i(t)$  and the energy orders  $\left( q_{ik}(t) \right)_k$  for each energy producer  $e_k$ .

Following our description of the interplay between the agents, the utility for provider  $s_i$  at time period  $t$  is:

$$\begin{aligned} \pi_i(t) = & p_i(t) \left( \nu_i^s(t) - a_i(t) \right) - \sum_{k=1, \dots, K} q_{ik}(t) \tilde{p}_k(t) - \gamma_i \left( \nu_i^s(t) - a_i(t) \right) \\ & - \sum_{k=1, \dots, K} \alpha_{ki}(t) \nu_k^e(t) \Big)_+ \end{aligned} \quad (2)$$

Provider  $s_i$  chooses his energy unit price and his energy orders toward energy producers so that  $\pi_i(t)$ , as defined in Equation (2), is maximized. His optimization program takes the form:  $\max_{p_i(t) > 0, (q_{ik}(t))_k \in \mathbb{R}_+^K} \left\{ \pi_i(t) \right\}$ .

### 2.2.3 Programs of the energy producers

The only decision variable for each energy producer  $e_k$  is the energy unit price,  $\tilde{p}_k(t)$ , that he proposes to the providers.

The utility of energy producer  $e_k$  at time period  $t$  equals:

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

To define the sharing coefficients,  $\left( \alpha_{ki}(t) \right)_i$ , we consider a weighted proportional allocation of resource that allows producers to discriminate energy allocation by providers while allocating his energy production simultaneously among the providers. This

framework is a generalization of the well-known proportional allocation mechanism where the providers' bids coincide with their energy orders weighted by their penalty coefficients  $(\tilde{\gamma}_i)_i$  [21], [24]. This means that between two providers with identical energy orders, the one having the highest penalty coefficient will receive the largest part of the producer's available energy. Indeed, the producer wants to minimize his overall penalty and, therefore, allocates larger parts of his production to providers serving isolated areas where failure of electricity supply may be critical. The choice of such a resource sharing mechanism can be justified by three points: firstly, a small extension of Nguyen and Vojnović's work [21] shows that weighted payment auction achieves competitive transfers to energy producers 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 providers might be left without allocation at all [13] ; thirdly, a sequential allocation of the energy production at the beginning of the time period is almost impossible because the productions 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:

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

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

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

Energy producer  $e_k$  chooses his energy unit price so that  $\tilde{\pi}_k(t)$ , as defined in Equation (4), is maximized. His optimization program is of the form:  $\max_{\tilde{p}_k(t) > 0} \left\{ \tilde{\pi}_k(t) \right\}$ .

### 3 Complete information game resolution

The game setting described in Subsection 2.1 implies that in the relation producers-providers, producers appear as leaders whereas providers are followers. Identically, in the relation providers-microgrids, providers appear as leaders whereas microgrids are mere followers. Under such a setting, the game is called a Stackelberg game and, as usual, it should be solved using backward induction [13], [19].

Additionally, we make the assumption that each energy producer receives at least one energy order from a provider, guaranteeing that the Stackelberg game admits non trivial solutions.



### 3.1 Optimization of the microgrids' decision

To minimize their total cost of energy, defined by Equation (1), microgrid  $\mathcal{M}_i$  has to choose  $a_i(t)$  so that the derivative of the total cost of energy equals 0 which means:

$$a_i(t) = p_i(t) \quad (5)$$

### 3.2 Optimization of the providers' decisions

To find his optimal price and energy orders, provider  $s_i$  has to replace  $a_i(t)$  by its optimal value in  $\pi_i(t)$ , defined in Equation (2), and to derivate the result in  $p_i(t)$  and in  $q_{ik}(t)$ . This derivation raises two cases.

#### 3.2.1 Case 1: the energy production fulfills the energy demand of the microgrid

It is the case when:

$$\nu_i^s(t) - p_i(t) < \sum_{k=1, \dots, K} \alpha_{ki}(t) \nu_k^e(t) \quad (6)$$

Then derivating the provider's utility in  $q_{ik}(t)$  leads to:  $\frac{\partial \pi_i(t)}{\partial q_{ik}(t)} = -\tilde{p}_k(t)$  which means that  $s_i$  will try to minimize all his energy orders to maximize his utility. As a result,  $\alpha_{ki}(t)$  will tend toward zero. This implies, in turn, that  $s_i$  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:

$$\nu_i^s(t) - p_i(t) = \sum_{k=1, \dots, K} \alpha_{ki}(t) \nu_k^e(t) \quad (7)$$

Therefore, in a complete information setting, it is possible for provider  $s_i$  to define his unit energy price so that it belongs to the frontier between both cases, avoiding the penalties from the TSO.

#### 3.2.2 Case 2: the energy production does not fulfill the energy demand of the microgrid

It is the case when  $\nu_i^s(t) - p_i(t) \geq \sum_{k=1, \dots, K} \alpha_{ki}(t) \nu_k^e(t)$ . Then derivating  $s_i$ 's utility gives us:

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

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

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

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

$$p_i(t) = \frac{\nu_i^s(t) + \gamma_i}{2} \quad (9)$$

$$\tilde{p}_k(t)C_k(t)^2 = \gamma_i\nu_k^e(t)\tilde{\gamma}_i\left(C_k(t) - \tilde{\gamma}_iq_{ik}(t)\right) \quad (10)$$

On the one side, we obtain directly the price for which the derivative of  $\pi_i(t)$  equals 0 through Equation (9). On the other side, Equation (10) can be rewritten as follows:

$$\tilde{\gamma}_iq_{ik}(t) = C_k(t) - \frac{\tilde{p}_k(t)C_k(t)^2}{\nu_k^e(t)\gamma_i\tilde{\gamma}_i} \quad (11)$$

If  $s_i$  anticipates that the other providers will make the same optimization program, replicating Equation (11) for the  $n$  providers and summing them all, results in the following equality:  $C_k(t) = nC_k(t) - \frac{\tilde{p}_k(t)C_k(t)^2}{\nu_k^e(t)} \sum_{j=1,\dots,n} \frac{1}{\gamma_j\tilde{\gamma}_j}$  by definition of  $C_k(t)$ .

Then as  $C_k(t)$  is not zero because each producer  $e_k$  receives at least one order of energy otherwise he would be out of the game, by dividing the previous equation by  $C_k(t)$  and reordering we obtain:  $C_k(t) = \frac{\nu_k^e(t)}{\tilde{p}_k(t)} \frac{n-1}{\delta}$  where  $\delta = \sum_{j=1,\dots,n} \frac{1}{\gamma_j\tilde{\gamma}_j}$ . By replacing  $C_k(t)$

in Equation (11), we obtain the energy orders for which the derivatives of  $\pi_i(t)$  equals 0:

$$q_{ik}(t) = \frac{\nu_k^e(t)}{\tilde{p}_k(t)} \frac{n-1}{\delta\tilde{\gamma}_i} \alpha(i) \quad (12)$$

where we have introduced the notation  $\alpha(i) = 1 - \frac{n-1}{\delta\gamma_i\tilde{\gamma}_i}$  to simplify future calculations.

Presently, we have to check that the price and energy 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). However, the energy orders defined in Equation (12) are non-negative if, and only if,  $1 \geq \frac{n-1}{\delta\gamma_i\tilde{\gamma}_i}$  which is equivalent to:

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

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

Secondly, by replacing the energy orders, defined by Equation (12), in Equation (3),

we obtain:

$$\begin{aligned}
\alpha_{ki}(t) &= \frac{\alpha(i)}{\sum_{j=1,\dots,n} \alpha(j)} \\
&= \frac{\alpha(i)}{\sum_{j=1,\dots,n} (1 - \frac{n-1}{\delta \gamma_j \tilde{\gamma}_j})} \text{ using } \alpha(j) \text{ definition} \\
&= \frac{\alpha(i)}{n - \frac{n-1}{\delta} \sum_{j=1,\dots,n} \frac{1}{\gamma_j \tilde{\gamma}_j}} \\
&= \frac{\alpha(i)}{n - (n-1)} \text{ using } \delta \text{ definition} \\
&= \alpha(i)
\end{aligned}$$

This proves that the sharing coefficient  $\alpha_{ki}(t)$  does depend neither on producer  $e_k$  nor on time instant  $t$ . Therefore, in the rest of the article, the sharing coefficients will be denoted  $\alpha(i)$ ,  $\forall i = 1, \dots, n$ . Furthermore, the above result means that the total energy delivered to microgrid  $\mathcal{M}_i$  is:  $\sum_{k=1,\dots,K} \alpha_{ki}(t) \nu_k^e(t) = \alpha(i) \sum_{k=1,\dots,K} \nu_k^e(t)$ . As a result, the price and energy orders for which the derivatives of  $\pi_i(t)$  equal 0 verify the inequality defining Case 2 if, and only if:

$$\nu_i^s(t) \geq \gamma_i + 2\alpha(i) \sum_{k=1,\dots,K} \nu_k^e(t) \quad (14)$$

This inequality states that the total production of energy by energy producers should not be too large compared to the demand of the microgrid. If it would not be the case, the over provisioning situation may probably get the most expensive producer out of the game.

If Inequalities (13) and (14) are true, the optimum for  $s_i$  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  $s_i$  is reached on the frontier defined by Equation (7).

### 3.3 Optimization of the energy producers' decision

After substituting  $q_{ik}(t)$  and  $C_k(t)$  by the expressions found in the previous section in energy producer  $e_k$ 's utility, as defined in Equation (4), we obtain:

$$\tilde{\pi}_k(t) = \nu_k^e(t) \frac{n-1}{\delta} \left[ \sum_{i=1,\dots,n} \frac{1}{\tilde{\gamma}_i} \left( 1 - \sum_{i=1,\dots,n} \left( \frac{\alpha(i)}{\tilde{p}_k(t)} - \frac{\tilde{p}_k(t) \tilde{\gamma}_i \delta}{n-1} \right)_+ \right) \right]$$

The only part of this equation depending on  $\tilde{p}_k(t)$  has always a negative impact on the profit of the energy producer under the assumption of fair penalties. Indeed, in that case, as raised in the previous section, we have:  $\alpha(i) \geq 0$  for all providers  $(s_i)_i$ . As a result, to maximize his profit, the energy producer has to choose  $\tilde{p}_k(t)$  such that the part

depending on  $\tilde{p}_k(t)$  in the above equation equals 0. It implies that the term  $1 - \frac{\tilde{p}_k(t)\tilde{\gamma}_i\delta}{n-1}$  is inferior to 0 for all  $i = 1, \dots, n$ . It is equivalent to:  $\tilde{p}_k(t) \geq \frac{n-1}{\delta\tilde{\gamma}_i}$ . Consequently, the optimal price for the energy producer with fair penalties should satisfy:

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

## 4 Distributed learning game

In this section, we assume that the microgrids' net energy demands,  $(\nu_i^s(t))_i$ , and the energy productions,  $(\nu_k^e(t))_k$ , 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].

In the previous section, we defined the optimal decisions for each agent at time period  $t$ . We proved that these decisions do depend neither on the microgrids net demands,  $\nu_i^s(t)$ , nor on the energy productions,  $\nu_k^e(t)$ , except for the providers. To guarantee the optimal system wide operation, it is fundamental for the providers to elaborate efficient learning strategies about the net demand of the microgrids and about the productions of the energy producers. The risk associated with this learning task will be measured by the provider's loss. It will be defined in Subsection 4.1.

Providers should optimize their prices and ordered quantities at each time period, having no information about the produced energy and the demands of the microgrids at this instant. As a result, the game can be considered as having incomplete information [19]. Each provider  $s_i$  has to forecast  $\nu_i^s(t)$  and  $\nu_k^e(t)$  for all  $k = 1, \dots, K$ , at each time period, in order to optimize his decisions. Furthermore, in a context of incomplete information, it is very likely that energy will be under provisioned to satisfy the end users' demand. The game parameters and random events (demands of the microgrids, energy productions and penalties) will be chosen so that we are always in energy shortage, in the sense that Inequality (14) will always be true, and with close values for the penalties, in the sense of Inequality (13). As a result, the optimal price for  $s_i$  is defined by Equation (9) and the optimal orders for  $s_i$  are defined by Equation (12).

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

- $\mathbf{f}_i(\mathbf{t}) \triangleq \left( \underbrace{f_i(\nu_i^s, t)}_{\text{forecast of } \mathcal{M}_i \text{'s net demand}}, \underbrace{f_i(\nu_1^e, t), \dots, f_i(\nu_K^e, t)}_{\text{forecasts of the productions}} \right)$  to denote the predictions made by provider  $s_i$  about microgrid  $\mathcal{M}_i$ 's net demand and about the production of each energy producer  $e_k$ ,  $k = 1, \dots, K$ .
- $\mathbf{f}(\mathbf{t}) \triangleq (\mathbf{f}_1(\mathbf{t}), \dots, \mathbf{f}_n(\mathbf{t}))$  which contains the forecasts of all the providers.

- $\mathbf{f}_{-i}(\mathbf{y}, \mathbf{t}) = \left( \mathbf{f}_1(\mathbf{t}), \dots, \mathbf{f}_{i-1}(\mathbf{t}), \mathbf{y}, \mathbf{f}_{i+1}(\mathbf{t}), \dots, \mathbf{f}_n(\mathbf{t}) \right)$  which contains the forecasts of all the providers except  $s_i$  of which prediction is set equal to  $\mathbf{y}$ .
- $\nu(\mathbf{t}) \triangleq \left( \underbrace{\nu_1^s(t), \dots, \nu_n^s(t)}_{\text{microgrids' net demands}}, \underbrace{\nu_1^e(t), \dots, \nu_K^e(t)}_{\text{productions}} \right)$  which contains the microgrids' net demands and the production of each energy producer  $e_k$ ,  $k = 1, \dots, K$ .

By substitution of the forecasters in the Stackelberg game solution at equilibrium as obtained in Section 3, we infer the optimal decisions for provider  $s_i$  at each time period  $t$ :  $p_i(t) = \frac{f_i(\nu_i^s, t) + \gamma_i}{2}$  and  $q_{ik}(t) = \frac{f_i(\nu_k^e, t)}{\tilde{p}_k(t)} \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta}$ . As a result, the utility of provider  $s_i$  at each time period  $t$  is:

$$\begin{aligned} \pi_i(t) &= \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \right) - \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} \sum_{k=1, \dots, K} f_i(\nu_k^e, t) \\ &\quad - \gamma_i \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{f_i(\nu_k^e, t) \alpha(i)}{\sum_{j=1, \dots, n} f_j(\nu_k^e, t) \alpha(j)} \nu_k^e(t) \right) \end{aligned} \quad (15)$$

The game will be repeated over a finite time horizon  $0 < T < +\infty$ .

#### 4.1 Learning risk measure definition and first observations

As already mentioned, the provider's risk, associated with the learning task, will be measured by his resulting loss. We have chosen a loss function representing the missing profit compared to the case where the provider would have made perfect forecasts of energy demands and energy productions. More precisely, for any provider  $s_i$ ,  $i = 1, \dots, n$ , his loss is defined as:

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

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

We now upper bound provider  $s_i$ 's loss as the sum of a loss function depending exclusively on the provider's predictions,  $l_i^{(1)}$ , and on another function,  $l_i^{(2)}$ , which relies on the disagreements between all the providers' predictions. The notion of disagreement needs to be carefully explained. To that purpose, we introduce:

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

as a measure of the disagreement between provider  $s_i$  and provider  $s_j$ ,  $i \neq j$ , in the prediction of producer  $e_k$ 's energy production, at time period  $t$ . By extension,  $d_{ij}^k(\mathbf{y}, t)$  contains the disagreement between provider  $s_i$  and all the other providers when  $s_i$  makes the prediction  $\mathbf{y}$  at time period  $t$  without any change in the predictions of the other providers.

**Proposition 1.** *Provider  $s_i$ 's loss can be upper-bounded by the sum of two functions: the first one,  $l_i^{(1)}$ , depending exclusively on his forecasts and the second one,  $l_i^{(2)}$ , which depends on his disagreement with the other providers' predictions:*

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

Proof of Proposition 1. The proof can be found in Appendix A.

We now demonstrate functional properties for provider  $s_i$ 's upper bounds which lead us to the following observations concerning the provider's strategic learning behavior.

**Corollary 2.** *To minimize his loss, provider  $s_i$  should:*

- *Have no bias in his forecast of microgrid  $\mathcal{M}_i$ 's demand*
- *Make his disagreements with the other providers' forecasts close to zero*

Proof of Corollary 2. We use the analytical expressions of  $l_i^{(1)}$  and  $l_i^{(2)}$  detailed in Appendix A. The derivative of  $l_i^{(1)}$  in  $f_i(\nu_i^s, t)$  equals 0 when  $f_i(\nu_i^s, t) = \nu_i^s(t)$ . Since  $l_i^{(1)}$  is a second order polynomial in  $f_i(\nu_i^s, t)$  with a positive first order coefficient ( $\frac{1}{4}$ ), the minimum of  $l_i^{(1)}$  is reached in  $\nu_i^s(t)$ . As a result, to reduce his loss,  $s_i$  has incentives to choose  $f_i(\nu_i^s, t) = \nu_i^s(t)$ . Furthermore, since the derivative of  $l_i^{(2)}$  in  $d_{ij}^k(t)$  is always positive,  $l_i^{(2)}$  increases when the disagreement with the other providers  $d_{ij}^k(t)$  increases.  $\square$

## 4.2 Optimal learning strategies for each provider

In this context of incomplete information on the microgrids' net demands and on the energy productions, we test two regret criteria to build the providers' learning strategy,  $d_t(\cdot)$ : 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  $\mathcal{E}_e \times \mathcal{E}_s^K$ . 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 [9] and regret-testing [7]. However, they do not offer any guarantees on the convergence of the algorithms and require larger times to reach an equilibrium when it exists.

The external regret over the sequence of time periods  $1, \dots, T$ , 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 provider  $s_i$ , it takes the form:

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

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

By definition, a strategy  $d_t(\cdot)$  has a small internal regret if for every couple of predictions  $\mathbf{y}, \mathbf{y}' \in \mathcal{E}_s \times \mathcal{E}_e^K$ , the forecaster does not regret of not having chosen prediction  $\mathbf{y}'$  each time instant he has chosen prediction  $\mathbf{y}$ :

$$\mathfrak{R}_i(T) \triangleq \max_{\mathbf{y}, \mathbf{y}' \in \mathcal{E}_s \times \mathcal{E}_e^K} \sum_{t=1, \dots, T} d_t(\mathbf{y}) \left[ l_i(\mathbf{f}_{-i}(\mathbf{y}, \mathbf{t}), \nu(\mathbf{t})) - l_i(\mathbf{f}_{-i}(\mathbf{y}', \mathbf{t}), \nu(\mathbf{t})) \right]$$

In a repeated game, this regret criterium ensures convergence of the joint empirical frequencies of play to the set of correlated equilibria [4] whereas there is no guarantees on the convergence of the product of the marginal empirical frequencies of play to Nash equilibria, under external regret minimization in a general game. Internal regret minimization will be used exclusively in the simulations in Subsection 5.2 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 providers which minimize their external regret asymptotically.

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

**Proof of Lemma 3.** In our case setting, at the end of each time period, provider  $s_i$  knows the quantity of demand postponed by the microgrid  $\mathcal{M}_i$  i.e., its  $\text{DR}^2$  as defined in Equation (5), and he can infer  $\nu_i^s(t)$  from that quantity using Equation (9).  $s_i$  also knows the energy which has been delivered by each energy producer  $e_k$  to him. He can infer from that the energy which could have been delivered to him, if he had ordered a different quantity  $q_{ik}(t)$ , all other providers ordering the same energy quantities, using Equation (3). As a result,  $s_i$  can calculate his loss for all his possible actions. In [4], it is proved that a Hannan consistent learning strategy always exists when the player can compute his loss for each possible action at the end of each time period.  $\square$

We now introduce lower and upper bounds on the disagreements between provider  $s_i$  and the other providers about the predictions of the energy productions:  $\underline{D}_{ss}(i) \triangleq \min_{t=1, \dots, T} \min_{j \neq i, k} d_{ij}^k(t)$  and  $\overline{D}_{ss}(i) \triangleq \max_{t=1, \dots, T} \max_{j \neq i, k} d_{ij}^k(t)$ . They contain the extreme disagreement values between the providers, about the estimated energy productions. For the sake of simplicity, we introduce the following notation:  $g(x) = 1 - \frac{\max\{x; 0\}}{\max\{\mathcal{E}_e\}} - \frac{\min\{x; 0\}}{\min\{\mathcal{E}_e\}}, \forall x \in \mathbb{R}$ . It will be used in the following lemma proof.

**Lemma 4.** *If provider  $s_i$  plays according to a Hannan consistent strategy for his loss upper bound then, there exists an upper bound for the external regret associated with  $s_i$ 's loss caused by his own predictions,  $l_i^{(1)}$ , which depends only on the extreme disagreement values between the providers about the estimated energy productions,  $\underline{D}_{ss}(i)$  and  $\overline{D}_{ss}(i)$ .*

<sup>2</sup>In the smart grid, the monitoring is performed through communicating meters deployed at the end user level [2], [17].

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

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

### 4.3 Analysis of the sum of providers loss functions upper bounds

We express the TSO's loss as the opposite of the sum of all the providers' losses. It coincides with the balance price that he should pay to ensure the electrical network reliability:

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

It is also possible to consider that the providers play against Nature [4] who adopts the worst behavior with regards to the providers, when setting the random individual sequences.

Similarly to the providers, the TSO will try to keep his external regret  $\mathfrak{R}(t)$  in  $o(T)$ .

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

$$\tilde{l}_g(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) = \sum_{i=1, \dots, n} l_i^{(1)}(\mathbf{f}_i(\mathbf{t}), \nu(\mathbf{t})) \quad (16)$$

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

$$\tilde{V}_g \triangleq \underbrace{\min_{\otimes_{i=1, \dots, n} d(\mathbf{f}_i) \in F_s^n}}_{\text{providers' predictors}} \underbrace{\max_{d(\nu) \in F_m}}_{\text{TSO's predictor}} \tilde{l}_g^E(\otimes_{i=1, \dots, n} d(\mathbf{f}_i), d(\nu))$$

where  $\tilde{l}_g^E$  represents the expectation of function  $\tilde{l}_g$  defined in Equation (16).

To simplify the analytical derivation of the following theorem, which is detailed in the Appendix, we define the function  $\psi$  from  $\mathbb{R}^2$  to  $\mathbb{R}$  such that:

$$\psi(\underline{D}_{ss}(i), \overline{D}_{ss}(i)) = \gamma_i \alpha(i) \sum_{k=1, \dots, K} \nu_k^e(t) \left( \frac{1}{g(\underline{D}_{ss}(i))} - \frac{1}{g(\overline{D}_{ss}(i))} \right) \quad (17)$$

**Theorem 5.** Assume that all providers play according to Hannan consistent strategies for their loss upper bound then:

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

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 his loss upper bound. Then:*

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

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  $s_i$ 's loss upper bounds, and using von Neuman-Morgenstern's minimax theorem [19] for  $\tilde{V}_g$ , we derive the proposed inequality.  $\square$

We let:

$$l_g(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) = \sum_{i=1, \dots, n} l_i(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) \quad (18)$$

be the sum of the providers' losses. Using the definitions settled in Equations (16) and (18), we derive the following inequality:

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

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

**Corollary 7.** *If all the providers play according to a Hannan consistent strategy for their loss upper bounds then, their average loss cannot be larger than:*

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

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 the case where they behave non-cooperatively [19], [24]. Adapted to our learning context, we define coalitions of agents as follows:

**Definition 8.** • *A coalition of providers is a group of providers who collaborate to learn the hidden energy productions  $(\nu_k^e(t))_k$ .*

- The grand coalition contains all the providers 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 on a common value.

Shared information concerns only energy productions. Indeed, each provider predicts independently his microgrid's net demand and it has no impact on the other providers.

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

Such a strategy would satisfy the following relations, at any time period  $t$ :

$$\begin{aligned} \psi\left(\underline{D}_{ss}(i), \overline{D}_{ss}(i)\right) \sum_{k=1,\dots,K} \nu_k^e(t) &= 0, \quad \forall i = 1, \dots, n \\ \Leftrightarrow \underline{D}_{ss}(i) &= \overline{D}_{ss}(i), \quad \forall i = 1, \dots, n \end{aligned}$$

It means that providers can decrease the upper bound of their average loss by coordinating their predictions about the produced energies  $\left(\nu_k^e(t)\right)_k$ , at any time period  $t$ . Providers therefore have incentives to form a grand coalition because it might enable them to decrease their total loss.

**Proposition 9.** *If the providers cooperate through a grand coalition and play Hannan consistent strategies, the system average loss over time interval  $[1; T]$  cannot be larger*

$$\text{than: } \tilde{V}_g - \sum_{i=1,\dots,n} \gamma_i \alpha(i) \frac{1}{T} \sum_{t=1}^T \sum_{k=1,\dots,K} \nu_k^e(t).$$

Proof of Proposition 9. By substitution in the second part of the loss upper bounds, as introduced in Proposition 1 and detailed analytically in Appendix A, we obtain:

$$l_i^{(2)}\left((d_{ij}^k(t))_{j,k}, \nu(t)\right)|_{d_{ij}^k(t)=0, \forall j,k} = -\gamma_i \alpha(i) \sum_{k=1,\dots,K} \nu_k^e(t)$$

It depends only on the provider index ( $i$ ) and on time period  $t$ , but not on the providers' forecasts. Furthermore, by definition:

$$\begin{aligned} l_g(t) &= \sum_{i=1,\dots,n} l_i(t) \\ &\leq \underbrace{\sum_{i=1,\dots,n} l_i^{(1)}(t)}_{\tilde{l}_g(t)} + \sum_{i=1,\dots,n} l_i^{(2)}(t) = \tilde{l}_g(t) - \gamma_i \alpha(i) \sum_{k=1,\dots,K} \nu_k^e(t) \end{aligned}$$

Taking the average of these values over time interval  $[1; T]$ , the proposition statement is straightforward.  $\square$

## 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 we discuss in the last part 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 provider has  $K + 1$  forecasts to do: one for his microgrid net demand and one to evaluate the production of each of the  $K$  energy producers. As a result, each provider should define a randomized strategy on the space  $\mathcal{E}_s \times \mathcal{E}_e^K$ . We recall that a randomized strategy is the classical terminology used in game theory to name a discrete density function defined over the considered set [19]. The size of the set grows very fast with  $K$  and, as a result, each probability in the randomized strategy of forecasts, is very small, which leads to rounding errors during computation. In order to overcome this issue, we have decided to cut the providers in smaller entities, each of them 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  $\mathcal{E}_s \times \mathcal{E}_e^K$  for each provider.

For a given forecast  $X$ , we derive the payoffs for each value  $x \in \mathcal{E}$  ( $\mathcal{E} = \mathcal{E}_s$  for net energy demands and  $\mathcal{E} = \mathcal{E}_e$  for energy productions) of the forecast at each time period  $t$  by using the utilities of the providers and keeping only the terms depending on forecast  $X$ . This is summarized in the following definition:

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

For the forecasts of microgrid  $\mathcal{M}_i$ 's net demand, provider  $s_i$ 's payoff takes the form:

$$\begin{aligned} H_{f_i(\nu_i^s)}(x, t) &= \frac{x + \gamma_i}{2} \left( \nu_i^s(t) - \frac{x + \gamma_i}{2} \right) - \gamma_i \left( \nu_i^s(t) - \frac{x + \gamma_i}{2} \right) \\ &\quad - \sum_{k=1, \dots, K} \frac{\alpha(i) f_i(\nu_k^e, t)}{\sum_{j=1, \dots, n} \alpha(j) f_j(\nu_k^e, t)} \nu_k^e(t) \Big)_+ \end{aligned}$$

Concerning the forecasts of energy producer  $e_k$ 's production, provider  $s_i$ 's payoff takes

the form:

$$H_{f_i(\nu_k^e)}(x, t) = -\frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} x - \gamma_i \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \right. \\ \left. - \sum_{l \neq k} \frac{\alpha(i) f_i(\nu_l^e, t)}{\sum_{j=1, \dots, n} \alpha(j) f_j(\nu_l^e, t)} \nu_l^e(t) - \frac{\alpha(i) x}{\sum_{j \neq i} \alpha(j) f_j(\nu_k^e, t) + \alpha(i) x} \nu_k^e(t) \right) +$$

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

$$H_{\nu_i^s}(x, t) = \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \left( \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - x \right) \\ + \gamma_i \left( x - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{\alpha(i) f_i(\nu_k^e, t)}{\sum_{j=1, \dots, n} \alpha(j) f_j(\nu_k^e, t)} \nu_k^e(t) \right) +$$

and

$$H_{\nu_k^e}(x, t) = \sum_{i=1, \dots, n} \gamma_i \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - \sum_{l \neq k} \frac{\alpha(i) f_i(\nu_l^e, t)}{\sum_{j=1, \dots, n} \alpha(j) f_j(\nu_l^e, t)} \right. \\ \left. \nu_l^e(t) - \frac{\alpha(i) f_i(\nu_k^e, t)}{\sum_{j=1, \dots, n} \alpha(j) f_j(\nu_k^e, t)} x \right) +$$

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

$$H_{f_C(\nu_i^s)}(x, t) = H_{f_i(\nu_i^s)}(x, t) \\ H_{f_C(\nu_k^e)}(x, t) = - \sum_{i=1, \dots, n} \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} x$$

Whereas, the TSO's payoffs become:

$$H_{\nu_i^s}(x, t) = \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \left( \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - x \right) + \gamma_i \left( x - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \right. \\ \left. - \alpha(i) \sum_{k=1, \dots, K} \nu_k^e(t) \right) +$$

and

$$H_{\nu_k^e}(x, t) = \sum_{i=1, \dots, n} \gamma_i \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - \alpha(i) \left( \sum_{l \neq k} \nu_l^e(t) + x \right) \right) +$$

## 5.2 Updates of forecasting strategies

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

$$\text{We let: } \vartheta_t \triangleq \sum_{s=1}^t \text{Var}\left(H_X(X_s, s)\right) = \sum_{s=1}^t \mathbb{E}\left[\left(H_X(X_s, s) - \mathbb{E}[H_X(X_s, s)]\right)^2\right]$$

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 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:

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


---

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

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

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

then:

$$\Delta_{(y,y')}(t) = \frac{w_{(y,y')}(t)}{\sum_{z \neq z'} w_{(z,z')}(t)}$$

$$w_{(y,y')}(t) = w_{(y,y')}(t-1)^{\frac{\eta t}{\eta t-1}} \exp\left(\eta_t \sum_{x \in \mathcal{E}} d_{t-1}^{y \rightarrow y'}(x) H_X(x, t-1)\right)$$

We see that we need to compute the maximum of the absolute value of the payoff function  $|H_X|$  for all forecasts  $X$  to run a simulation of the game. This maximum is reached for  $x = \min\{\mathcal{E}\}$  or  $x = \max\{\mathcal{E}\}$  for all payoff functions except for  $H_{f_i(\nu_i^s)}$  because their derivative with respect to  $x$  is never equal to 0. For  $H_{f_i(\nu_i^s)}$ , the derivative equals 0 if, and only if,  $f_i(\nu_i^s, t) = \nu_i^s(t)$ , so the maximum of  $|H_X|$  is reached either for  $x = \min\{\mathcal{E}\}$  or  $x = \max\{\mathcal{E}\}$  or  $x = \nu_i^s(t)$ .

### 5.3 Numerical illustrations

The code for the numerical illustrations is written in `Python`. We consider two providers and two producers. We compare the cumulative payoff of each player (provider or grand coalition) to the cumulative payoff of the same player who would have forecasted the best value at each time period in terms of payoffs. More precisely, we compute for each actor  $a$  (the provider  $s_1$ ,  $s_2$  or the grand coalition  $C$ ), the following performance metric:

$$\mathfrak{R}_a(T) \triangleq \frac{1}{T} \sum_{s=1}^T \sum_{X \in F_a} \left( H_X(X_s, s) - \max_{x \in \mathcal{E}} (H_X(x, s)) \right)$$

where  $F_a$  is the generic set of forecasts made by the actor  $a$ . 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,  $\frac{\mathfrak{R}_a(T) - \mathfrak{R}_a(T-1)}{\mathfrak{R}_a(T-1)}$ , is smaller than  $10^{-2}$ .

We let  $T_{s_i}^*$ ,  $i = 1, 2$  and  $T_C^*$  be the number of time steps which are necessary for the convergence of the regret based algorithms for provider  $s_i$  and for the grand coalition, in non-cooperative and cooperative scenarios respectively. According to these notations, provider  $s_i$  has incentives to cooperate if, and only if,  $T_{s_i}^* \geq T_C^*$ . Both providers have incentives to cooperate if, and only if,  $\min\{T_{s_1}^*; T_{s_2}^*\} \geq T_C^*$ .

In our simulations, we have calculated the convergence times of learning algorithms for a wide range of combinations of penalty coefficients  $(\gamma_1, \gamma_2, \tilde{\gamma}_1, \tilde{\gamma}_2)$ . More precisely, in Figures 1 and 2 (a), we make the assumption that:  $\gamma_2 = 2\gamma_1 = 2\tilde{\gamma} \in [0; 1]$  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  $\gamma_2$ ,  $\tilde{\gamma}_1$  and  $\tilde{\gamma}_2$ . We can check easily that all these combinations of penalty coefficients satisfy Equation (13). In addition, we have chosen  $\mathcal{E}_s = [5, 8]$  and  $\mathcal{E}_e = [1, 2]$  so that Equation (14), i.e. energy shortage, is always true.

In Figure 1, the providers and the grand coalition's learning strategies 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 the combinations of penalty coefficients  $(\gamma, \tilde{\gamma}_1, \tilde{\gamma}_2)$ . The height of each histogram is determined by the number of penalty coefficients having the same convergence time. The class  $T_{s_i}^* = 30$  (resp.  $T_C^* = 30$ ) contains combinations of penalty coefficients that did not converge for  $T \leq 30$ . 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 the combinations of penalty coefficients  $(\gamma, \tilde{\gamma}_1, \tilde{\gamma}_2)$ . According to the top of both figures, we observe that, for a far larger number of penalty coefficients, the learning algorithm convergence times are smaller in cooperative scenarios than in non-cooperative scenarios. Furthermore, by comparison of Figures 1 and 2 top, the convergence times are smaller for learning algorithms based on internal regret minimization than for learning algorithms based on external regret minimization.

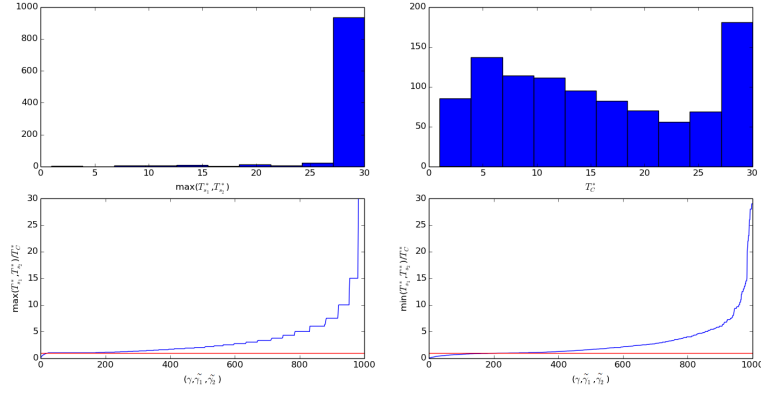
On the opportunity of a grand coalition to emerge, we infer from Figure 1 (a) and (b) (resp. 2 (a) and (b)) bottom, that for 99% of the combinations of penalty coefficients at least one provider has incentives to cooperate and that for 85% (resp. 98.5%) of the combinations of penalty coefficients both providers have incentives to cooperate, using external regret minimization (resp. internal regret minimization) as criterion.

Judging by the numerical illustrations, we conclude that:

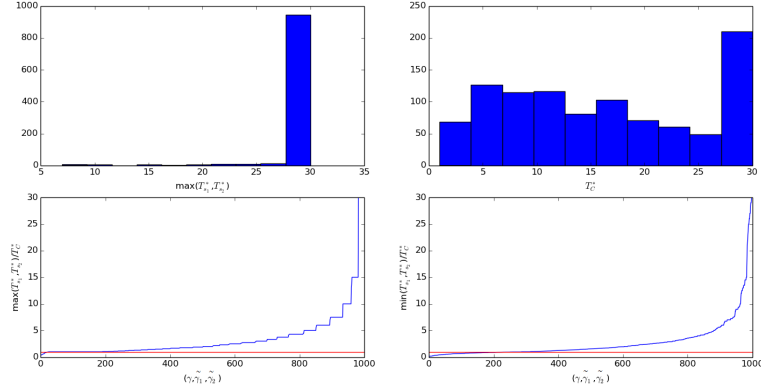
- For all the combinations of penalty coefficients, convergence times are smaller for cooperative scenarios than for non-cooperative scenarios and under internal regret minimization than under external regret minimization
- A grand coalition has 85% (resp. 98.5%) of chances to emerge under external (resp. internal) regret minimization

## 6 Conclusion

In this article, we studied a model of decentralized renewable energy production in which producers, providers and microgrids are organized in a hierarchical network. Renewable energy productions were modeled by random individual sequences which need not to have a probabilistic structure. This extraordinarily general demand and supply structure allows to take into account exogenous events. As a result, it is more robust to extreme events and appears as particularly well suited to model quite erratic processes such as renewable energy production. We determined analytically the energy prices enabling the producers to avoid the penalties that the balance operators threaten to apply in case where the providers' orders would not entirely be satisfied. All the risk was then reported on the providers. Additionally, we proved that these latter can minimize their average risk by sharing information and aligning their forecasts. Finally, numerical simulations, run on a large sample of parameter combinations, led us to observe that the convergence times of collaborative learning strategy are clearly superior to times resulting from decentralized learning, using external and internal regret minimization. Furthermore, a grand coalition has 85% (resp. 98.5%) of chances to emerge under external (resp. internal) regret minimization.



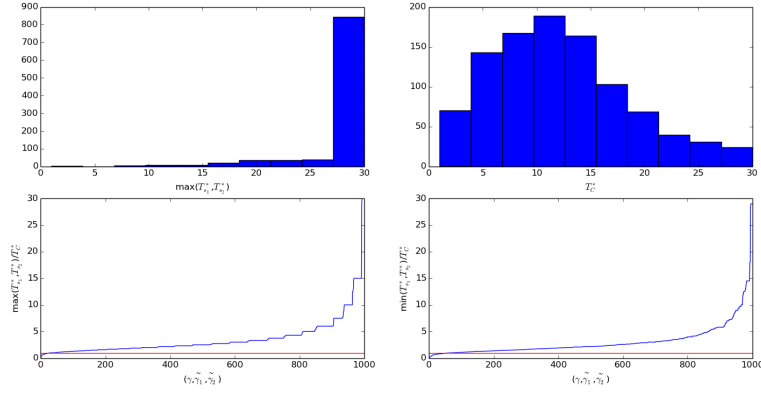
(a)



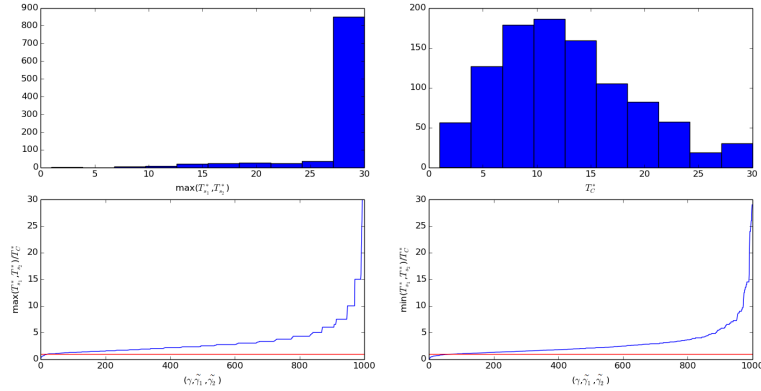
(b)

Figure 1: Convergence times and incentives to collaborate under *external regret* minimization. In (a), we have:  $\gamma_2 = 2\gamma_1 = 2\gamma \in [0; 1]$  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 steps  $T = 30$ . 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 combinations of penalty coefficients.





(a)



(b)

Figure 2: Convergence times and incentives to collaborate under *internal regret* minimization. In (a), we have:  $\gamma_2 = 2\gamma_1 = 2\gamma \in [0; 1]$  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 steps  $T = 30$ . 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 combinations of penalty coefficients.

An area of improvement concerns the design of the penalties paid to the electricity transmission operator who compensates the negative energy balances. Is it possible to design more generic mechanisms? Could rules be adapted to guarantee the market opening and avoid speculations, like capacity retention or under investment in the means of production?

## Appendix

### Appendix A: Proof of Proposition 1

Using Equation (15), we infer the analytical expression of provider  $s_i$ 's utility when his forecasts are perfectly aligned with microgrid  $\mathcal{M}_i$ 's net demand and with the production of each energy producer:

$$\begin{aligned}\pi_i^0(t) &= \frac{\nu_i^s(t) + \gamma_i}{2} \left( \nu_i^s(t) - \frac{\nu_i^s(t) + \gamma_i}{2} \right) - \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} \sum_{k=1, \dots, K} \nu_k^e(t) \\ &\quad - \gamma_i \left( \nu_i^s(t) - \frac{\nu_i^s(t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{\nu_k^e(t) \alpha(i)}{\sum_{j=1, \dots, n} f_j(\nu_k^e, t) \alpha(j)} \nu_k^e(t) \right)_+\end{aligned}$$

By definition of provider  $s_i$ 's loss and thanks to Equation (15), we have:

$$\begin{aligned}l_i(f(t), \nu(t)) &= \left( \pi_i^0(t) - \pi_i(t) \right) \\ &= \frac{\nu_i^s(t) + \gamma_i}{2} \left( \nu_i^s(t) - \frac{\nu_i^s(t) + \gamma_i}{2} \right) - \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} \sum_{k=1, \dots, N} \nu_k^e(t) \\ &\quad - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \right) + \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} \sum_{k=1, \dots, K} f_i(\nu_k^e, t) \\ &\quad - \gamma_i \left[ \left( \nu_i^s(t) - \frac{\nu_i^s(t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{\nu_k^e(t) \alpha(i)}{\sum_{j=1, \dots, n} f_j(\nu_k^e, t) \alpha(j)} \nu_k^e(t) \right)_+ \right. \\ &\quad \left. - \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{f_i(\nu_k^e, t) \alpha(i)}{\sum_{j=1, \dots, n} f_j(\nu_k^e, t) \alpha(j)} \nu_k^e(t) \right)_+ \right]\end{aligned}$$

A first approach might consist at assuming that:

$$\left( \nu_i^s(t) - \frac{\nu_i^s(t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{\nu_k^e(t) \alpha(i)}{\sum_{j=1, \dots, n} \nu_k^e(t) \alpha(j)} \nu_k^e(t) \right)_+ = 0 \text{ since in a case of complete information, provider } s_i \text{ might choose his price } p_i(t) \text{ so that Equation (7) holds}$$

i.e., that no penalty applies. On the contrary, in case of incomplete information, the provider might be in energy shortage. This implies that:  $\left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} - \sum_{k=1, \dots, K} \frac{f_i(\nu_k^e, t) \alpha(i)}{\sum_{j=1, \dots, n} f_j(\nu_k^e, t) \alpha(j)} \nu_k^e(t) \right)_+ > 0$

$\sum_{k=1,\dots,K} \frac{f_i(\nu_k^e, t) \alpha(i)}{\sum_{j=1,\dots,n} f_j(\nu_k^e, t) \alpha(j)} \nu_k^e(t) \Big)_+ > 0$ . In the setting of this first approach, it is

now possible to break  $l_i(f(t), \nu(t))$  into two parts: the first one,  $l_i^{(1)}$ , depending exclusively on his predictions and the second one,  $l_i^{(2)}$ , depending on his interactions with the others' predictions. Therefore, we let:

$$\begin{aligned} l_i^{(1)}(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) &= \frac{\nu_i^s(t) + \gamma_i}{2} \left( \nu_i^s(t) - \frac{\nu_i^s(t) + \gamma_i}{2} \right) - \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} \sum_{k=1,\dots,K} \nu_k^e(t) \\ &\quad - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \right) \\ &\quad + \frac{\alpha(i)}{\tilde{\gamma}_i} \frac{n-1}{\delta} \sum_{k=1,\dots,K} f_i(\nu_k^e(t)) + \gamma_i \left( \nu_i^s(t) - \frac{f_i(\nu_i^s, t) + \gamma_i}{2} \right) \end{aligned}$$

The second part takes the form:

$$\begin{aligned} l_i(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) - l_i^{(1)}(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) &= -\gamma_i \alpha(i) \sum_{k=1,\dots,K} \frac{f_i(\nu_k^e, t) \nu_k^e(t)}{\sum_{j=1,\dots,n} f_j(\nu_k^e, t) \alpha(j)} \\ &\leq -\gamma_i \alpha(i) \sum_{k=1,\dots,K} \nu_k^e(t) \left[ \sum_{j=1,\dots,n} \alpha(j) \left( 1 - \frac{d_{ij}^k(t)}{\min\{\mathcal{E}_e\}} \mathbf{1}_{d_{ij}^k(t) < 0} \right. \right. \\ &\quad \left. \left. - \frac{d_{ij}^k(t)}{\max\{\mathcal{E}_e\}} \mathbf{1}_{d_{ij}^k(t) \geq 0} \right) \right]^{-1} = l_i^{(2)}((d_{ij}^k(t))_{j=1,\dots,n, k=1,\dots,K}, \nu(\mathbf{t})) \end{aligned}$$

Indeed, we observed that:

$$\begin{aligned} \left( \frac{f_i(\nu_k^e, t)}{\sum_{j=1,\dots,n} \alpha(j) f_j(\nu_k^e, t)} \right)^{-1} &= \frac{\sum_{j=1,\dots,n} \alpha(j) (f_i(\nu_k^e, t) - d_{ij}^k(t))}{f_i(\nu_k^e, t)} \\ &= \sum_{j=1,\dots,n} \alpha(j) \left( 1 - \frac{d_{ij}^k(t)}{f_i(\nu_k^e, t)} \right) \\ &\leq \sum_{j=1,\dots,n} \alpha(j) \left( 1 - \frac{d_{ij}^k(t)}{\min\{\mathcal{E}_e\}} \mathbf{1}_{d_{ij}^k(t) < 0} - \frac{d_{ij}^k(t)}{\max\{\mathcal{E}_e\}} \mathbf{1}_{d_{ij}^k(t) \geq 0} \right) \end{aligned}$$

since,  $\mathcal{E}_e$  being a close subset of  $\mathbb{R}$ , the forecasters are upper and lower bounded i.e.,  $\min\{\mathcal{E}_e\} \leq f_i(\nu_k^e, t) \leq \max\{\mathcal{E}_e\}$ .

A second approach where energy shortage occurs both under complete and incomplete information is possible. The statements of the proposition remain unchanged.  $\square$

## Appendix B: Proof of Lemma 4

Suppose that provider  $s_i$  plays according to a Hannan consistent strategy according to his loss upper bound i.e.,  $l_i^{(1)}(\mathbf{f}_i(\mathbf{t}), \nu(\mathbf{t})) + l_i^{(2)}((d_{ij}^k)_{j,k}(t), \nu(\mathbf{t}))$ . This means that:

$$\limsup_{T \rightarrow +\infty} \frac{1}{T} \left[ \sum_{t=1}^T \left( l_i^{(1)}(\mathbf{f}_i(\mathbf{t}), \nu(\mathbf{t})) + l_i^{(2)}((d_{ij}^k(t))_{j,k}, \nu(\mathbf{t})) \right) \right. \\ \left. - \min_{y \in \mathcal{E}_s \times \mathcal{E}_c^K} \left( \sum_{t=1}^T \left( l_i^{(1)}(y, \nu(\mathbf{t})) + l_i^{(2)}((d_{ij}^k(y, t))_{j,k}, \nu(\mathbf{t})) \right) \right) \right] \leq 0 \quad (19)$$

In Section 4, we have introduced lower and upper bounds on the disagreements between provider  $s_i$  and the other providers about the forecasts of the energy productions:  $\underline{D}_{ss}(i)$  and  $\overline{D}_{ss}(i)$ . According to Corollary 2,  $l_i^{(2)}$  being increasing in  $d_{ij}^k(t)$ , it is possible to provide lower and upper bounds for the function by evaluating it in  $\underline{D}_{ss}(i)$  and  $\overline{D}_{ss}(i)$  respectively. The lower bound is:  $b_l(i, t) = -\frac{\gamma_i \alpha(i)}{g(\underline{D}_{ss}(i))} \sum_{k=1, \dots, K} \nu_k^e(t)$ . Whereas, the upper bound takes the form:  $b_u(i, t) = -\frac{\gamma_i \alpha(i)}{g(\overline{D}_{ss}(i))} \sum_{k=1, \dots, K} \nu_k^e(t)$ .

If Inequality (19) is checked, then the following inequality holds:

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

□

## Appendix C: Proof of Theorem 5

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

$$\limsup_{T \rightarrow +\infty} \frac{1}{T} \left[ \sum_{t=1}^T l_i^{(1)}(\mathbf{f}_i(\mathbf{t}), \nu(\mathbf{t})) - \min_{y_i \in \mathcal{E}} \sum_{t=1}^T l_i^{(1)}(y_i, \nu(\mathbf{t})) \right] \\ \leq \frac{1}{T} \psi(\underline{D}_{ss}(i), \overline{D}_{ss}(i)) \sum_{t=1}^T \sum_{k=1, \dots, K} \nu_k^e(t) \quad (20)$$

Summing Inequality (20) over all  $i = 1, \dots, n$ , the external regret evaluated in the sum of the providers' losses  $(l_i^{(1)})_{i=1, \dots, n}$ , becomes:

$$\limsup_{T \rightarrow +\infty} \frac{1}{T} \left[ \sum_{t=1}^T \tilde{l}_g(\mathbf{f}(\mathbf{t}), \nu(\mathbf{t})) - \min_{\mathbf{f}} \sum_{t=1}^T \tilde{l}_g(\mathbf{f}, \nu(\mathbf{t})) \right] \\ \leq \frac{1}{T} \sum_{i=1, \dots, n} \psi(\underline{D}_{ss}(i), \overline{D}_{ss}(i)) \sum_{t=1}^T \sum_{k=1, \dots, K} \nu_k^e(t)$$

In addition:  $\min_{\mathbf{f}} \frac{1}{T} \sum_{t=1}^T \tilde{l}_g(\mathbf{f}, \nu(\mathbf{t})) = \min_{\otimes_i d(\mathbf{f}_i) \in F_s^n} \frac{1}{T} \sum_{t=1}^T \tilde{l}_g^E(\otimes_i d(\mathbf{f}_i), \nu(\mathbf{t}))$  where  $\tilde{l}_g^E$  represents the expectation of function  $\tilde{l}_g$ .

We assume that each provider makes his forecasts independently of the other providers. Then  $\tilde{l}_g^E(\cdot, \nu(\mathbf{t}))$  is linear in  $\otimes_{i=1, \dots, n} d(\mathbf{f}_i)$ . As a result, its minimum over the simplex of probability vectors is reached in one of the corners of the simplex.

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  $\mathbf{z} \in \mathcal{E}_s^n \times \mathcal{E}_e^K$ . We obtain:

$$\begin{aligned} \min_{\otimes_i d(\mathbf{f}_i) \in F_s^n} \frac{1}{T} \sum_{t=1}^T \tilde{l}_g^E(\otimes_i d(\mathbf{f}_i), \nu(\mathbf{t})) &= \min_{\otimes_i d(\mathbf{f}_i) \in F_s^n} \sum_{\mathbf{z} \in \mathcal{E}_s^n \times \mathcal{E}_e^K} d_T(\mathbf{z}) \tilde{l}_g^E(\otimes_i d(\mathbf{f}_i), \mathbf{z}) \\ &= \min_{\otimes_i d(\mathbf{f}_i) \in F_s^n} \tilde{l}_g^E(\otimes_i d(\mathbf{f}_i), d_T) \\ &\leq \underbrace{\max_{d(\nu) \in F_m} \min_{\otimes_i d(\mathbf{f}_i) \in F_s^n} \tilde{l}_g^E(\otimes_i d(\mathbf{f}_i), d(\nu))}_{\hat{V}_g} \end{aligned}$$

□

## 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, 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, Issue 2, pp.866 – –876, 2013
- [3] Bossavy A., Girard R., Kariniotakis G., Forecasting ramps of wind power production with numerical weather prediction ensembles, Wind Energy, vol.16, Issue 1, published online feb. 28, 2012
- [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, Issue 6, 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] Foster D. P., Peyton H., Learning, Hypothesis testing, and Nash equilibrium, Games and Economic Behavior, vol.45, pp.73 – –96, 2003

- [8] 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
- [9] Hart S., Mas-Colell A., A Simple Adaptive Procedure Leading to Correlated Equilibrium, *Econometrica*, vol.68, pp.1127 – 1150, 2000
- [10] 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
- [11] 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, EEEV 2012
- [12] 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 2012
- [13] Le Cadre H., Mercier D., Is Energy Storage an Economic Opportunity for the Eco-Neighborhood?, *NETNOMICS: Economic Research and Electronic Networking*, vol.13, Issue 3, pp.191 – 216, 2012
- [14] Li N., Marden J. R., Decoupling coupled constraints through utility design, Working Paper, University of Colorado, Department of ECEE, 2011
- [15] Ljungqvist L., Sargent T. J., recursive Macroeconomic Theory, MIT Press, 2-nd Edition, 2007
- [16] 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
- [17] 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, Issue 3, pp.320 – 331, 2010
- [18] Morales A. J., On the Role of the Group Composition for Achieving Optimality, *Annals of Operations Research*, vol.137, pp.387 – 397, 2005
- [19] Myerson R., Game Theory: An Analysis of Conflict, Harvard University press, 2006
- [20] Nair J., Adlakha S., Wierman A., Energy Procurement Strategies in the Presence of Intermittent Sources, Working Paper, California Institute of Technology, 2013
- [21] Nguyen T., Vojnović M., Weighted Proportional Allocation, in proc. of ACM Sigmetrics 2011
- [22] Pradelski B. R., Young H. P., Learning efficient Nash equilibria in distributed systems, Working Paper, University of Oxford, Department of Economics, 2010

- [23] Saad W., Han Z., Debbah M., Hjørungnes 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
- [24] Saad W., Han Z., Poor V. H., Coalitional Game Theory for Cooperative Micro-Grid Distribution Networks, in *proc. of the 2-nd IEEE International Workshop on Smart Grid Communications*, 2011
- [25] Shapley L. S., Stochastic games, in *proc. of the National Academy of Sciences of the United States of America*, vol.39, pp.1095–1100, 1953
- [26] 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
- [27] Becoming a upstream-downstream balance operator: RTE rules, <http://clients.rte-france.com/>, 2012
- [28] NOME law, <http://www.cre.fr/dossiers/la-loi-nome>, 2012