

# Stability and Performance of Coalitions of Prosumers Through Diversification in the Smart Grid

Nicolas Gensollen, Vincent Gauthier, Monique Becker and Michel Marot

CNRS UMR 5157 SAMOVAR,

Telecom SudParis/Institut Mines Telecom

Email: {nicolas.gensollen, vincent.gauthier, michel.marot, monique.becker}@telecom-sudparis.eu

**Abstract**—Achieving a successful energetic transition through a smarter and greener electricity grid is a major goal for the 21st century. It is assumed that such smart grids will be characterized by bidirectional electricity flows coupled with the use of small renewable generators and a proper efficient information system. All these bricks might enable end users to take part in the grid stability by injecting power, or by shaping their consumption against financial compensation. In this paper, we propose an algorithm that forms coalitions of agents, called prosumers, that both produce and consume. It is designed to be used by aggregators that aim at selling the aggregated surplus of production of the prosumers they control. We rely on real weather data sampled across stations of a given territory in order to simulate realistic production and consumption patterns for each prosumer. This approach enables us to capture geographical correlations among the agents while preserving the diversity due to different behaviors. As aggregators are bound to the grid operator by a contract, they seek to maximize their offer while minimizing their risk. The proposed graph based algorithm takes the underlying correlation structure of the agents into account and outputs coalitions with both high productivity and low variability. We show then that the resulting diversified coalitions are able to generate higher benefits on a constrained energy market, and are more resilient to random failures of the agents.

**Keywords:** prosumers, coalition formation, correlation, stability, clique expansion, electricity market

## I. INTRODUCTION

The recent developments in the smart grid area, such as demand side management mechanisms, storage efficiency, or communication architectures to name a few, sketch a new power system. The rigid top-down centralized architecture of today's power grids is expected to evolve in more distributed and flexible systems. Bidirectional power flows, storage, and efficient use of information are some of the key principles that will enable this transition. Since production could be located down to the very end of the distribution networks, nodes that were simply pure loads yesterday could behave tomorrow as generators or loads, further complicating energy usages [5]. In this paper, we focus on these "nodes" in the distribution network that can produce and consume electricity. More precisely, we consider agents owning both renewable distributed energy resources (DER) and electrical loads. The production of an agent can be used to meet its own demand, but in cases when it is over-producing, we consider that it

has the possibility to inject, against financial compensation, its extra-production in the grid. Such an agent model is known as a "prosumer" [6]. A key point of this work is to merge the interests of the grid operator with the prosumers in a single utility function. While the prosumers intend to maximize their benefits with better contracts, the operator concerns are related to the quality and reliability of the injections.

We rely on a market representation of the interactions. Coalitions of prosumers estimate a production capacity for an upcoming period and announce it on the market. If some other entity buys the contract, the coalition has to inject exactly the contracted amount of power at any time. Obviously, deviations will cause financial penalties to the coalitions, such that respecting the contracts is paramount. Since prosumers use exclusively renewables, we are particularly concerned with the probability that the available capacity of a coalition falls below its contract value. In this case, storage or demand side management might be solutions to avoid penalties, but this probability should still be maintained low.

We focus on coalitions because it appears as a solution to both the number of prosumers communicating with the operator as well as their high variabilities. Indeed, it has been argued that using a multi-level aggregators architecture for the control could lead to better performances on the communication side [1]. On the other hand, it is known that diversification of the assets is a way of minimizing risks when building a portfolio. We thus expect more stable and predictable energy productions for coalitions than for single agents. Nevertheless, all coalitions are not equally stable or productive. Special attention should thus be paid to the aggregation step. Recall that prosumers have both consumption and production components, each depending on location and time, meaning that there are complex underlying correlations between the agents. This is a central topic of this paper : given  $N$  prosumers, what coalitions should be formed so that the compromise between expected production and variability is optimized ?

We will see that the variability in the aggregated productions can be quantified to a certain extent by the correlation among the agents forming the coalitions. Understanding the correlation relationships among the agents can thus give an indication about what coalitions to form and how much they should sell. More precisely, we use a graph representation of

the correlation structure to gain insight about the expected production to risk ratios of different coalitions. We build a framework in which the system operator specifies both the minimum production acceptable to enter the market and restrain the amount of risk he is willing to accept. We then propose a graph heuristic that uses decorrelated cliques in order to form diversified productive and stable coalitions, and we compare the results with other formation strategies (see section VII). Rather than maximizing the profit of each agent in the grid (prosumers and grid operator), our algorithm tries to form the most productive coalitions given a maximum amount of acceptable risk. In other words, it tries to maximize coalitions profits without considering individual retributions to single agents. Furthermore, the pricing strategies for both the prosumer and the grid is beyond the scope of this paper.

Because agents are susceptible to fail for diverse reasons, the propensity of the system to undertake these failures is critical [2]. We therefore investigate in this paper the resilience of the coalitions when prosumers fail. Despite the fact that losing agents is usually detrimental to the coalitions, we will see that the coalitions formed with our algorithm tend to be less impacted by random failures.

The paper is organized as follows, section II gives a brief overview of the related literature, section III clarifies how we generated realistic prosumer production traces based on weather data. In section IV, we define most of the notations and explain why correlation between prosumers is a quantity of interest for our objective. Based on the conclusions of section IV, we develop in section V a utility function quantifying how much power a coalition can announce on the market given an accepted risk level. We then develop, in section VI a greedy optimization algorithm that uses decorrelated cliques as inputs and improves their utility over a correlation-constrained environment. Finally, section VII provides some results both on performance of the method and resilience of the coalitions formed. The overall process is illustrated on figure ??.

## II. RELATED WORK

Since renewables only produce when the resource is present and because these moments do not necessarily coincide with the consumption peak hours [3], allowing large penetration of renewables in the grid appears as a difficult challenge. One possible solution consists in shifting some of the loads towards high productivity periods. Demand side management techniques can indeed mitigate the consumption peaks and waste less production [3] [4]. Since end users tend to seek a maximum utility at a minimum cost, dynamic pricing appears as a good way to give incentives to them. By carefully scheduling the prices, it is assumed that the load curve can be shaped to some extent.

Dynamic pricing is an interesting and useful tool, but it has also some limitations. It is likely that dynamic pricing will serve as a shaping mass tool while finer techniques will be needed locally. A popular approach in this direction consists in deploying storage devices in the network and using them as electricity buffers. Basically these storages would be charged when there is a surplus of production, and discharged

when the consumption exceeds the production. Although quite simple, this idea causes numerous challenges when it comes to its real implementation. As long as the system considered remains small, a centralized control of these buffers could be envisioned. However, for real large systems it is likely that more sophisticated decentralized algorithms will be necessary. In [5], the authors introduce a distributed energy management system with a high use of renewables such that power is scheduled in a distributed fashion. In [6] the optimal storage capacity problem is addressed. There is indeed an interesting trade-off between the costs of the equipments and the expected availability of power. The authors develop a framework that enables them to exhibit a Pareto front of efficient solutions.

These are only a few possibilities for balancing production and demand in the smart grid. Most of the time, these techniques will be coupled with predictions of the upcoming load curves and weather conditions. Combining all these technologies enables aggregators to quantify their expected production and the inherent risk that comes with it. The optimization of expected returns to risk is a traditional goal in finance, and a wide literature exists on this topic. It is well-known for instance, that the more risk one is willing to accept, the higher his potential gains. On the contrary, when investing exclusively on low risks assets, one should expect relatively small gains. This trade-off is formalized in the Markowitz' portfolio theory [7]. More precisely, given a set of assets for which we have some historic data of returns, the objective is to find a linear combination of these assets (the so-called portfolio) which maximizes the expected value while minimizing the variance of the portfolio's return. Markowitz's answer is a set of efficient portfolios that all optimize in some sense this trade-off. If one is able to put a number on his risk acceptance or on the target expected return, the corresponding efficient portfolio is a priori the best option. One of the most controversial assumptions in the portfolio theory is that returns are jointly normally distributed (or, at least, that the returns distribution is jointly elliptical). Some economist have pointed out the fact that this assumption might not capture well the reality of financial markets [8]. Nevertheless, one of the key point in the Markowitz theory is to consider explicitly the correlation between the assets since they impact directly the variances of the portfolios. Since the work of [9], an interesting approach consists in computing a distance metric based on the correlation coefficients in order to organize the series in a correlation graph. Nodes represent the series considered while the edges are weighted by the metric. Because the metric can be computed for all pairs, these graphs are complete and of little use as is. Historically, the approach used by [9] was to compute a minimum spanning tree as to obtain a hierarchical clustering of the series. Later on, it was pointed out that, by definition, a spanning tree could not capture the underlying clustering structure hidden in the correlation graph. In this paper, we use another classical filtering technique called  $\epsilon$ -graph [10]. It consists in selecting a threshold  $\epsilon$ , and filtering out edges with smaller weights. As we will see further in this paper, this approach has the advantage of preserving clusters of correlated series.

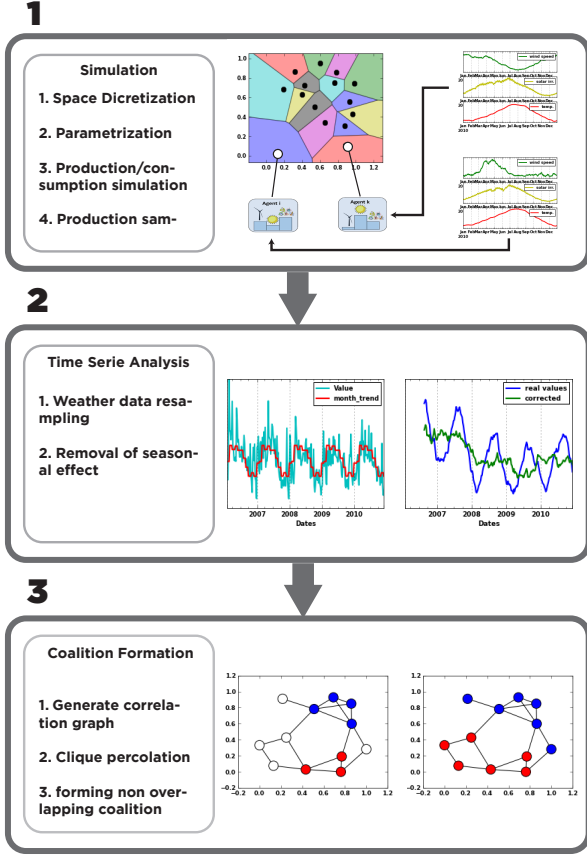


Figure 1: Process

### III. GENERATING REALISTIC PROSUMER PATTERNS

An essential component of the smart grid is the smart meter which makes the interface between the end user and the rest of the system. Smart meters coupled with sensors measure quantities of interest (like instantaneous consumption), receive informations from the grid (electricity prices for instance), and take actions accordingly (demand side management programs). Smart meters are currently and gradually deployed, and will probably provide interesting datasets to work on. Unfortunately, at the time this paper was written, production and consumption data for prosumers over a large region were not yet available to our knowledge. Some interesting experiments are nonetheless being conducted and data are progressively made public [11]. In this paper, we use weather quantities like wind speed or solar radiance as alternative data for generating realistic production and consumption series. Fortunately, these kinds of data are easier to find, and since the development of small personal weather stations, their geographical granularity keeps increasing. Since these quantities depend both on time and location, we discretize time into slots and space into zones in the following (see block 1 of fig. ??). A zone is simply a portion of the considered region of study for which we sampled data. Therefore, if prosumers  $i$  and  $j$  are positioned on the same zone, they are exposed to the same weather. Adding some intra-zone noise can easily be done though not considered in this paper. We denote by  $P_i(t)$  the instantaneous

$N$	Number of Prosumers
$P_i(t)$	Extra-production of agent $i$ at time $t$
$P_S^{CRCT}$	Contract value of coalition $S$
$\phi \in [0, 1]$	Reliability Threshold
$P_{MIN}$	Minimum Contract value
$\alpha$	Parameter that controls the coalitions sizes
$\epsilon$	Correlation graph filtering threshold
$N_{COAL}$	Number of coalitions
$\mathcal{R}_S$	Resilience of coalition $S$

Figure 2: Main notations.

available extra-production of agent  $i$  at time  $t$  :

$$P_i(t) = P_i^P(t) - P_i^D(t) \quad (1)$$

Where  $P_i^P(t)$  represents the total production of agent  $i$  at time  $t$  and  $P_i^D(t)$  its consumption at time  $t$ . In other words,  $P_i(t)$  represents the instantaneous surplus of power that agent  $i$  is willing to sell at time  $t$ . We simulated these traces by considering separately  $P_i^P$  and  $P_i^D$ . For a prosumer  $i$ , it is possible to write both quantities as a sum over the distributed energy resources ( $DER_i$ ) and loads ( $load_i$ ) of  $i$  :

$$P_i^P(t) = \sum_{k \in DER_i} P_k(t) \quad (2)$$

$$P_i^D(t) = \sum_{k \in load_i} P_k(t) \quad (3)$$

For simplicity, in this paper we only consider wind-turbines (WT) and photovoltaic panels (PV) as possible DERs for the agents ( $DER_i = WT_i \cup PV_i$ ):

$$P_i^D(t) = \sum_{k \in WT_i} P_k(t) + \sum_{k \in PV_i} P_k(t) \quad (4)$$

We denote by  $\nu_i(t)$  and  $\Psi_i(t)$  the wind speed (in  $m.s^{-1}$ ) and the solar radiance (in  $W.m^{-2}$ ) at the location of agent  $i$  and at time  $t$ , so that :

$$P_i^P(t) = \sum_{k \in WT_i} \mathcal{F}_{WT}(\nu_i(t)) + \sum_{k \in PV_i} \mathcal{F}_{PV}(\Psi_i(t)) \quad (5)$$

Where  $\mathcal{F}_{WT}$  (resp.  $\mathcal{F}_{PV}$ ) is the power curve for the wind-turbines (resp. photovoltaic panels). We made here the implicit assumption that all wind-turbines (resp. photovoltaic panels) have the same power curve. The model can be easily extended to multiple power curves accounting for different types of generators. More details about power curves and their approximations can be found in the appendix and in [12].

Note that a prosumer  $i$  is defined by his zone  $Z_i$  as well as the sets  $DER_i$  and  $load_i$ . That is, a prosumer can be configured to represent anything from a single wind-turbine for instance ( $DER_i = \{WT_0\}$  and  $load_i = \emptyset$ ) to a pure load ( $DER_i = \emptyset$  and  $load_i = \{L_0\}$ ) through more complex combinations. In practice, we use random configurations for the agents. In the rest of the paper, we use french weather data [13] starting in January 2006 and ending in December 2012, with a sampling frequency of three hours, and generate  $N$  time-series of extra-production over this date range. Note that this

low sampling frequency might hide short time intermitencies (see [14] for more information). Studying the evolution of the correlation structure with the sampling frequency is out of the scope of this paper, but might lead to interesting results.

#### IV. NOTATIONS

This section provides most of the notations and introduces important concepts for the rest of the paper. We consider a set  $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$  of  $N$  prosumers of the distribution network. Each agent is configured randomly and we simulate its extra-production  $P_i(t)$ ,  $\forall i \in \mathcal{A}$  from 2006 to 2012. Based on these historical values, our objective is now to form groups of prosumers so that the global power production resulting from the superposition of individual's extra-productions be both sufficiently high and predictable. Let  $P_S(t) = \sum_{i \in S} P_i(t)$  be the available extra-production of coalition  $S$  at time  $t$ .

Suppose now that coalition  $S$  has to suggest a production value  $P_S^{CRCT}$  to enter the market. This means that, during the time  $S$  is on the market, it will have to inject in the grid exactly  $P_S^{CRCT}$  at any time  $t$  and will be rewarded proportionally to this amount, with penalties if it deviates. Obviously, the actual extra-production will not be constant at this value and will oscillate due to intermitencies in the production and consumption. If  $S$  has an available production always greater than  $P_S^{CRCT}$ , it is losing some gains since it could have announced a higher contract value. If the production oscillates around  $P_S^{CRCT}$ , by using batteries or demand side management techniques,  $S$  could be able to maintain its production to the contract value at any time. Nevertheless, if the oscillations are too important compared to the available storage capacity,  $S$  will probably break the contract and pay penalties. We can see that there is a return over risk trade-off here. Coalitions should thus find the right balance between announcing too low and losing some potential gains, and claiming too high and paying penalties.

Let us illustrate the rest of the notations and concepts with a simple example. We consider only two agents  $i$  and  $j$  such that the distribution of their extra-production can be approximated by normal distributions :  $P_i \sim \mathcal{N}(\mu_i, \sigma_i)$  and  $P_j \sim \mathcal{N}(\mu_j, \sigma_j)$ . This is only for explanation purposes as it is of course rather unrealistic in real situations where the distributions are skewed. We can write the distribution of the coalition  $S = \{i, j\}$  as  $P_{\{i,j\}} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$ , where :

$$\begin{cases} \mu_{ij} = \mu_i + \mu_j \\ \sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2 + \rho_{ij}\sigma_i\sigma_j} \end{cases} \quad (6)$$

$\rho_{ij}$  being the Pearson's correlation coefficient between  $P_i$  and  $P_j$ . If the coalition  $\{i, j\}$  proposes a contract value  $P_S^{CRCT}$ , all instants when  $\{i, j\}$  will produce less than  $P_S^{CRCT}$  is critical. Indeed, in this kind of situations,  $\{i, j\}$  will either have to discharge batteries to keep up with its contract, or pay penalties to the grid. The probability that  $\{i, j\}$  is under-producing compared to the contract :  $Pr[P_{i,j} \leq P_S^{CRCT}]$  is thus an important indicator of the coalition's quality. A well-

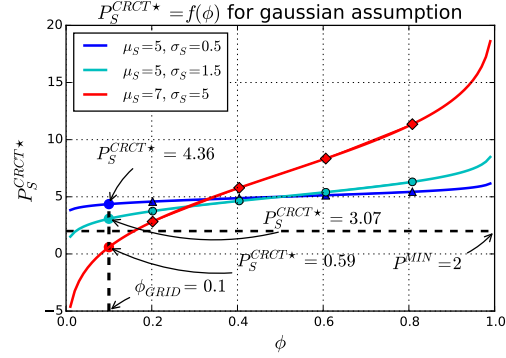


Figure 3:  $P_S^{CRCT*}$  depending on reliability parameter  $\phi$  for Gaussian distributions (see equation 8). Blue curve with triangles stands for a coalition  $S$  with an expected production of 5 units and a standard deviation of 0.5. Under a grid policy of  $\phi = 0.1$ , it is able to announce a contract value of  $P_S^{CRCT*} = 4.36$ . The same coalition in term of expected production ( $\mu = 5$ ), but with a higher variance ( $\sigma = 1.5$ , cyan curve with circles) can only afford a smaller contract value of  $P_S^{CRCT*} = 3.07$ . The red curve with diamonds stands for a coalition with a higher expected production ( $\mu = 7$ ), but with a very high unpredictability ( $\sigma = 5$ ). For low values of  $\phi$ , this coalition is thus heavily penalized and can only afford a contract of 0.59 units. Under grid policy ( $\phi = 0.1$ ,  $P^{MIN} = 2$ ), this last coalition is thus not allowed to enter the market (red dot below the horizontal dashed line).

known result for normal distributions is that the cumulative distribution function can be written as :

$$Pr[P_{ij} \leq P_S^{CRCT}] = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{P_S^{CRCT} - \mu_{ij}}{\sigma_{ij}\sqrt{2}} \right) \right] \quad (7)$$

where  $\text{erf}$  is the error function :  $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ .

The contract a given coalition is willing to take depends on its capacity to compensate for under-producing (using batteries, backup generators...), and its risk acceptance. Selecting the right contract value appears thus as an interesting problem on its own that we plan to investigate in future works. In order to keep the present paper in a reasonable length, we simplify the contract value selection problem by giving some responsibilities to a third party named the grid operator. The role of the grid operator is to constrain the market entry to coalitions able to propose both sufficiently high and sufficiently credible contract values. More formally, let  $\phi \in [0, 1]$  be the reliability threshold fixed by the grid operator as a maximum value for the probability of under-producing. The highest contract value that a coalition can propose is thus  $P_S^{CRCT*}$  such that  $Pr[P_{ij} \leq P_S^{CRCT*}] = \phi$ . In the Gaussian example, it implies that coalition  $\{i, j\}$  is announcing :

$$P_S^{CRCT*} = \mu_{ij} - \sqrt{2}\sigma_{ij}\text{erf}^{-1}(1 - 2\phi) \quad (8)$$

This is the best contract value that the coalition  $S$  can afford for a given stability policy  $\phi$  of the grid operator. Figure 3 shows how  $P_S^{CRCT*}$  evolves according to the reliability parameter  $\phi$ . For illustration, the range of  $\phi$  values is shown from 0 to 1, but in practice, only small values of  $\phi$  really make sense :  $\phi = 1$  for instance means that coalitions

can announce absolutely anything since the probability of producing less than any contract value is necessarily less than one by trivial definition of a probability. As visible on figure 3, coalitions with high expected productions but presenting a high unpredictability are penalized and can only afford small contracts. The grid operator also specifies a lower bound  $P^{MIN}$  on the contract values as not to overload the market with unrealistic small coalitions. We thus characterized a valid coalition as one satisfying the two conditions :

$$\begin{cases} Pr[P_{ij} \leq P_S^{CRCT}] \leq \phi \\ P_S^{CRCT} \geq P^{MIN} \end{cases} \quad (9)$$

On figure 3,  $P^{MIN}$  is fixed to 2 units for illustration purpose. For  $\phi = 0.1$ , only blue triangles and cyan circles coalitions are valid while red diamonds coalition is not.

The Gaussian assumption of this small example is convenient as it allows us to write  $P_S^{CRCT*}$  analytically. Nevertheless, such an assumption is rather unrealistic in practice. In the following, we keep the same framework but release this Gaussian assumption unless the contrary is specified (see eq. 14). This assumption will indeed be convenient for computing some parameter estimates.

## V. UTILITY FUNCTION

In this section, we use the notions of contract values and valid coalitions developed in section IV in order to design a proper utility function. The contract basically indicates the rate at which a coalition has to inject power in the grid. It seems then natural that coalitions are remunerated proportionally to their contract values  $gain(S) \propto P_S^{CRCT*}$ . More precisely, if  $\lambda$  is the unitary price rate for electricity, a coalition  $S$  injecting  $P_S^{CRCT*}$  in the grid during a period  $[t_0, t_k]$  earns :

$$gain(S) = \int_{t_0}^{t_k} \lambda P_S^{CRCT*} dt = P_S^{CRCT*} \int_{t_0}^{t_k} \lambda dt \quad (10)$$

(since  $P_S^{CRCT*}$  is supposed to be a constant rate over the contracted period). Unfortunately  $gain(S)$  is not a concave function of the coalitions' sizes, meaning that coalitions can grow as large as the number of agents allows it, without any counterbalance effects. Such a model, that virtually allows infinitely large coalitions and contract values, is in practice not realistic. There are indeed costs (communication costs for instance) that increase with the coalitions sizes. We take this observation into account by rescaling the utility of a coalition  $S$  by its size in term of number of agents ( $|S|$ ):

$$U(S) = \begin{cases} \frac{1}{|S|^\alpha} \frac{P_S^{CRCT*}}{P^{MAX}}, & \text{if } S \text{ is valid,} \\ 0, & \text{if } S \text{ is not valid} \end{cases} \quad (11)$$

where parameter  $\alpha$  controls to what extent the size of a coalition impacts its utility, and  $P^{MAX}$  is a normalizing factor. Based on  $U$ , the marginal contribution of an agent  $i$  can be expressed as  $\delta_S(i) = U(S + \{i\}) - U(S)$ . A coalition  $S$  has thus an interest in adding an additional agent  $i$  if :

$$\delta_S(i) \geq 0 \Leftrightarrow P_{S+\{i\}}^{CRCT*} \geq P_S^{CRCT*} \left( \frac{|S|+1}{|S|} \right)^\alpha \quad (12)$$

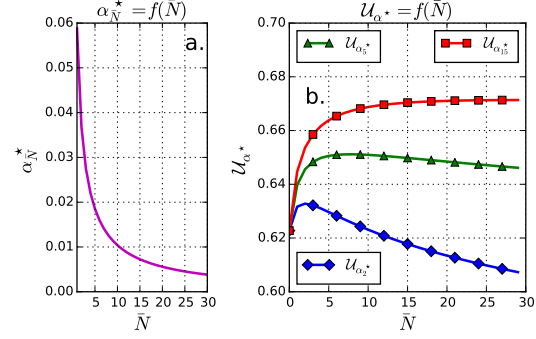


Figure 4: Gaussian mean approximation. Subplot **a** shows how the parameter  $\alpha$  of the utility function should be chosen in function of the mean size of the coalitions (see equation 14). Subplot **b** displays the corresponding utility functions for different values of  $\alpha$ . Blue curve with diamonds favors small coalitions of 2 agents while the green one with triangles favors 5 agents coalitions. Finally, the red curve with squares has an optimal size of 15 agents.

If  $\alpha$  is set to zero, agents are added as long as they increase the contract value of the coalition. If  $\alpha$  is greater than zero, additional agents have to increase the contract value by some factor. We relate now  $\alpha$  to the mean sizes of the coalitions  $\bar{N}$  by solving the following equation in the gaussian case (as in section IV) :

$$\left[ \frac{\partial U}{\partial |S|} \right]_{|S|=\bar{N}} = 0 \quad (13)$$

Furthermore, since considering all the possible interactions between agents is analytically intractable, we use here a mean approximation. Any quantity  $x$  that varies over the agent set is thus simplified in its mean value  $\bar{x}$ . Solving equation 13 for  $\alpha$  in these conditions leads to:

$$\alpha_N^* = \frac{0.7\bar{\sigma}(\bar{\rho}-1)\text{erf}^{-1}(2\phi-1)}{\bar{\mu}\sqrt{\bar{N}(\bar{\rho}\bar{N}-\bar{\rho}+1)} + 1.4\bar{\sigma}\text{erf}^{-1}(2\phi-1)(\bar{\rho}\bar{N}-\bar{\rho}+1)} \quad (14)$$

Figure 4 shows how  $\alpha^*$  and the utility function evolves according to the mean size of the coalitions  $\bar{N}$ . These curves are only valid in the simplified Gaussian example considered here. Nonetheless, they will provide some guidance when using real data. As can be pointed out, the purpose of  $U$  is not to study coalitions stability against player defection which could be done through game theory, nor to redistribute the coalition's utility in terms of individual payoffs.  $U$  is rather designed as a quality measure for the coalitions.

## VI. COALITION FORMATION

Section IV explained how contract values for the coalitions are computed, and in section V we related this quantity to the utility and gains of a coalition. Since the computation time of this quantity is not negligible, we derive in this section the heuristic we used to form the coalitions structure.

### A. Representing the correlation structure

As seen in section IV, the variance of the aggregated production impacts directly the contract values, and depends on



the covariances between the agents productions. We argue here that, by having some representation of the correlation structure between the agents, the search landscape for high utility coalitions could be reduced, such that good coalitions are more likely to be found quickly. Usually, this correlation structure is formalized with a covariance matrix or a correlation matrix that contains all the correlation coefficients between the agents :  $M = (\rho_{ij})_{\forall i,j \in \mathcal{A}^2}$ . By using a metric to map this matrix in a weighted adjacency matrix (see section II), it is possible to obtain a graph representation of the correlation relationships between the agents. In the following, we use two opposite distance metrics  $d_{ij}^1 = 1 - \rho_{ij}^2$  and  $d_{ij}^2 = \rho_{ij}^2 = 1 - d_{ij}^1$ .

Clearly,  $d^1$  (resp.  $d^2$ ) maps two correlated series as close points (resp. distant) while two uncorrelated series are distant (resp. close). These metrics enable us to compute a correlation graph  $G_1 = (\mathcal{A}, E_1)$  and a "de-correlation" graph  $G_2 = (\mathcal{A}, E_2)$ . For any  $i$  and  $j$ , the weight of the edge  $e_{ij}$  is  $d_{ij}^1$  in  $G_1$  and  $d_{ij}^2$  in  $G_2$ . In both cases, we want to keep only the edges which weights are located in the lower tail of the distance distributions. In other words, we want to compute the  $\epsilon$ -graphs of  $G_1$  and  $G_2$  such that only meaningful edges remain. Selecting the filter  $\epsilon$  is an important point affecting the landscape search for the coalition formation. Unfortunately, there seems to be no clear consensus in the literature on how to select such a threshold. We will see later in this section that cliques in  $G_2$  are potential seeds for the coalitions. Since we want to generate  $N_{COAL}$  coalitions, we need at least  $N_{COAL}$  cliques of a given size to start. Besides, since we consider coalitions as disjoint, the starting cliques should be non overlapping. We select our optimal threshold for  $G_2$  as :

$$\epsilon^* = \min_{\epsilon \in [0,1]} \{ \epsilon \text{ s.t. } |\Theta_k(G_2^\epsilon)| \geq N_{COAL} \} \quad (15)$$

where  $G_2^\epsilon$  is the de-correlation graph  $G_2$  filtered by  $\epsilon$ , and  $\Theta_k(G)$  is the set of non overlapping cliques of size  $k$  in a given graph  $G$ . In other words we select  $\epsilon^*$  as the smallest threshold possible such that the filtered de-correlation graph contains at least  $N_{COAL}$  non overlapping cliques of size  $k$ . In practice, since finding cliques requires exponential time, we use triangles [15] ( $k = 3$  in eq. 15) rather than cliques as soon as the number of nodes is not small. Note also that the existence of  $\epsilon^*$  as defined in equation 15 is not guaranteed. The users has indeed to provide consistent values of  $N_{COAL}$  or  $k$  compared to the size of the agent population  $\mathcal{A}$ .

### B. Cliques

In [10] the structural roles of weak and strong links on financial correlation graphs is investigated. The author shows that strong links, accounting for strong correlation relationships, are responsible for the clustering, while weak links provide the connectivity between clusters. Indeed, if we consider three items, say  $a$ ,  $b$ , and  $c$  such that  $a$  and  $b$  are strongly correlated and  $b$  and  $c$  are also strongly correlated, then it is likely that  $a$  and  $c$  are also strongly correlated. It can be easily shown using the cosine addition formula<sup>1</sup>, that if  $\rho_{ab} > x$  and  $\rho_{bc} > x$  with

<sup>1</sup>  $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$

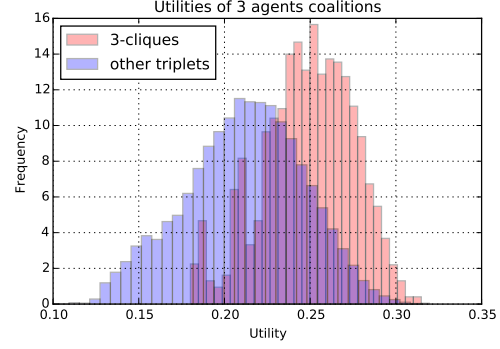


Figure 5: Histograms of utility values for coalitions of size 3. Red bars stand for cliques in the decorrelation graph, and blue bars for all the other possible triplets. Cliques tend to exhibit higher utilities than random coalitions.

$x > 0$ , then  $\rho_{ac} > 2x^2 - 1$ ). Correlation graphs capture this weak transitivity notion through clusters of correlated series.

Nevertheless de-correlation seems like a more complex concept than correlation in the sense that there is not even a partial notion of transitivity when it comes to it. Therefore, the clustering coefficients of  $G_1^\epsilon$  is much higher than the one of  $G_2^\epsilon$ . This can be seen as another formulation of [10] on the structural roles of weak and strong links on financial correlation graphs. Strong links, accounting for strong correlation relationships, are responsible for the clustering, while weak links provide the connectivity between clusters. Searching for clusters in  $G_2^\epsilon$  and hoping that this strategy will provide a nice coalition structure of internally uncorrelated coalitions seems thus pointless. Consider now a clique in  $G_2^\epsilon$ , which is a complete subgraph of  $G_2^\epsilon$ . Since there is a link for every pairs of nodes, we know, by construction, that a clique has a mean correlation and a maximum correlation less than  $\epsilon$ .

Figure 5 shows the distributions of the utility values for cliques of size 3 (triangles) in  $G_2^{\epsilon^*}$  and for all the other possible triplets of agents. It is clearly visible that cliques tend to exhibit higher utilities because of their de-correlation property. Choosing cliques in  $G_2^{\epsilon^*}$  as coalitions seems therefore appealing. Nevertheless, the quality of the results seems to decrease as the sizes of the cliques increase. Indeed, the larger the desired cliques, the more dense  $G_2^{\epsilon^*}$  becomes (see equation 15). There is a point where cliques result more from noisy edges than true de-correlation, which decreases the quality of the results. Directly mapping cliques to coalitions by this de-correlation oriented approach is thus not sufficient. It is indeed possible that adding agents to these cliques has the combined effect of increasing the expected production while decreasing its stability. The question revolves around measuring the benefits of this production surplus compared to the disadvantage of having coalition with high volatility. This can be quantified by the marginal benefit in equation 12.

### C. Algorithm

The algorithm takes inputs from :

- **The agents** : historical series of available productions  $P_i$ ,
- **The grid operator** : market entrance policy  $(P^{MIN}, \phi)$ ,

- **The "user"** : Number of desired coalitions  $N_{COAL}$  and size of starting cliques  $k$ .

The first steps consist in computing the de-correlation graph  $G_2$  as well as the optimal threshold  $\epsilon^*$ . Cliques of size  $k$  in  $G_2^{\epsilon^*}$  are considered as coalition seeds. The next step is a local greedy improvement over the landscape represented by  $G_2^{\epsilon^*}$ . Cliques add alternatively the node  $i^*$  in their neighborhood that yields the best marginal benefit  $MAX_{i \in N(clique)} \delta_{clique}(i)$  where  $N(clique)$  is the neighborhood of a given clique. See algorithms section in the appendix.

## VII. RESULTS

The algorithm presented in the previous section is supposed to generate a given number of coalitions that have good utilities. As it comprises mainly of a greedy optimization based on local improvements, there is no guarantee that the algorithm finds the global optimum. Since there is, to our knowledge, no state of the art algorithm that aggregates uncorrelated agents in an optimum way (see section II for related problems), we compare the results with :

- **Random sampling of coalitions** : Coalitions are formed randomly without any other constraint that the desired size. This enables us to have an idea about the distributions of utility values for coalitions of a given size.
- **Random sampling of coalition structures** : Coalition structures are sampled randomly by shuffling and random divisions of the agents. This algorithm (see appendices) uses such a sampling and returns the highest utility coalition structure sampled. We refer to this algorithm as "random".
- **Correlated** : Opposite version of our algorithm. It performs a community detection on the correlation graph  $G_1$ . The resulting coalitions have thus very high internal correlations. We thus expect this algorithm to perform very bad compared to the others. See appendices.

Before running the algorithms, we need to calibrate the utility function by choosing the value of the  $\alpha$  parameter. Recall that the purpose of this parameter is to take into account some constraints on the coalition's sizes if needed. In this paper, we do not have any technical constraints on coalition sizes though we designed the utility such that these could be taken into account. We select the desired size as being  $\lfloor N/N_{COAL} \rfloor$  (where  $\lfloor \cdot \rfloor$  means floor). Figure 6 shows how the mean utility of a coalition evolves with its size when the optimum size is set to 40 agents.

Figure 7 shows the coalitions formed with the considered algorithms in the contract value / volatility space. The color map in the background indicates regions where we expect high utilities (red) and the ones where we expect very poor utility values (blue). The bottom right corner, with high contract values and low volatilities, is therefore the region where we wish to form our coalitions. A single coalition is represented by a marker and the color and shape of a marker indicates by which algorithm the coalition has been formed. Besides, the sizes of the coalitions are indicated on the markers, and the marker size is also proportional to the coalition size. We can see that the utility function results in approximatively balanced coalitions.

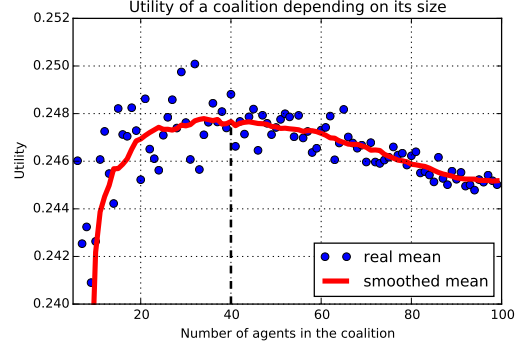


Figure 6: Utility of random coalitions depending on their size. Blue dots show real mean utility values and the thick red curve its smoothed version by applying a Savitzky-Golay filter. On this plot the  $\alpha$  parameter of the utility function was selected according to eq. 14 in order to favor 40 agents coalitions.

Small yellow markers indicates the gravity centers of their respective coalition structures. The coalitions of correlated agents (green squares) are clearly of poor quality according to our criteria since they can only afford small production contracts, and with a very high volatility. The decorrelated coalitions (blue dots) are closer to the bottom right corner indicating a much better quality in term of productivity over volatility ratio. The black dotted line indicates the mean values for the random coalitions sampling technique. Each small dot stands for the mean position of all sampled coalitions of this given size. Variances are not indicated for readability, but are usually quite large since this sampling only takes the size as a constraint. We can see that as coalitions get larger, they tend to increase on average their contract values, but at the price of a higher volatility. The results of the random coalition structure sampling are shown with the red ellipses that represent the distribution of the gravity centers of the sampled structures. Since the center of the ellipses stands for the mean and each ellipse adds one standard deviation, more than 99% of the sampled gravity centers are within the largest ellipse. The small yellow dot below the ellipses indicates the gravity center of our solution. It is thus visible that our greedy graph based algorithm is able to find a quite good coalition structure in terms of volatility and contract values.

A key point for the coalitions, besides stability and productivity, is their resilience. The resilience of a system refers to its ability to perform its tasks when subject to failures of its components. We consider here the case of random failures of the power electronics of some agents that has the consequence of preventing them from participating in the market. Therefore, the notion of resilience we will use in the following can be seen as the ability of the coalition structures to inject stable power in the grid when some of its internal agents are removed. According to our model, the grid operator specified two thresholds ( $P^{MIN}$  and  $\phi$ ) such that the power injected by every coalition is constrained :  $P_S^{CRCT} \in [P^{MIN}, P_S^{CRCT*}]$ . As long as a coalition can propose a contract value higher than  $P^{MIN}$ , it is valid and allowed to enter the energy market. We define the resilience of a coalition  $S$  as the probability that  $S$  produces more than the  $P^{MIN}$  threshold

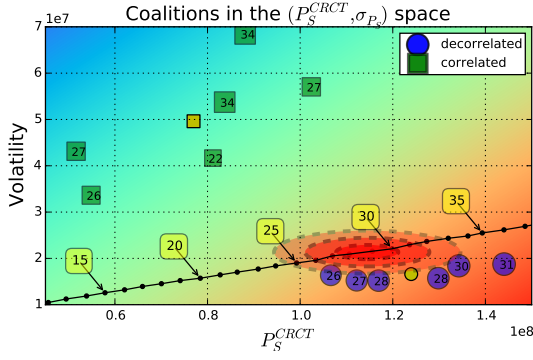


Figure 7: Coalitions formed in the (contract value, volatility) space. The color map indicates the underlying quality. The closer to red the better (high contract values with small volatility). On the opposite, blue areas show poor quality (small contract values with high volatility). Blue dots stand for the decorrelated coalitions that we formed while green squares show correlated coalitions. The smaller yellow markers stand for the gravity centers of the coalition structures. The black dotted line shows how contract values and volatility evolve when the size of the coalitions increases (a subset of the points are labeled by the size of the coalition they represent). Each point is the average over  $10^5$  unconstrained draws of a random coalition. We also show the distribution of the gravity centers of random coalition structures with red ellipses (center is the mean, and each ellipse corresponds to one standard deviation).

:  $\mathcal{R}_S = Pr[P_S \geq P^{MIN}] = 1 - Pr[P_S < P^{MIN}]$ . And we extend this measure to the coalition structures :

$$\mathcal{R}_{CS} = \prod_{S \in CS} (1 - Pr[P_S < P^{MIN}]) \quad (16)$$

We consider that prosumers fail randomly, and we denote by  $\psi \in [0, 1]$  the fraction of agents that failed. Figure 8 exhibits how the resilience of the coalition structures evolves according to  $\psi$  (see appendix section *Resilience algorithm* for more details). On the top subplot,  $P^{MIN}$  was voluntarily selected relatively low such that the resiliences of the three structures fit on the same figure. When the  $P^{MIN}$  requirement increases, the differences between the algorithms also increase as visible on the bottom subplot of figure 8. The decorrelated coalitions achieve a more resilient production on the market in the sense that they sustain a higher fraction of node failures.

## VIII. CONCLUSION

In this paper we studied how aggregations of prosumers could be authorized to sell their surplus of production to the grid operator. By relying on the past values of the agents, we constrained the market entry to both sufficiently productive and stable coalitions. The power that a coalition is able to propose on the market is therefore related to production and stability. The correlations between the prosumers impact directly the volatility of the coalitions, which led us to seek uncorrelated aggregations of agents. We used a graph representation of the correlation relationships between the agents as a reduced landscape for the coalition formation. A greedy algorithm that starts with cliques of the "de-correlation"

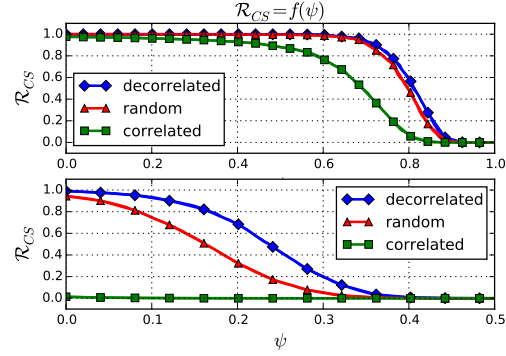


Figure 8: Resilience of the coalition structures when nodes fail randomly (see equation 16) for  $P^{MIN} = 10 MW$  (top subplot) and  $P^{MIN} = 80 MW$  (bottom subplot)

graph of the agents and makes local improvements offers a good compromise between speed and quality of the results. We compared these results with random samplings, and an opposite strategy that clusters correlated agents together. We showed that the coalitions resulting from our algorithm are able to provide more power to the grid with a lower volatility. Because of their better production over volatility ratios, these coalitions will tend to use less storage and waste less energy than more unstable coalitions. Because in real situations, agents are prone to failure, resilience is also an important criterion for the quality of the aggregations. We therefore studied how the coalitions remain on the market when their agents fail randomly. We showed that, in this situation, the coalitions resulting from our algorithm better withstand losses of agents.

## REFERENCES

- [1] E. Negeri and N. Baken, "Architecting the Smart Grid As a Hierarchy," *Proceedings of the 1st International Conference on Smart Grids and Green IT Systems*, pp. 73–78, 2012.
- [2] S. Pahwa, C. Scoglio, S. Das, and N. Schulz, "Load-shedding strategies for preventing cascading failures in power grid," *Electric Power Components and Systems*, vol. 41, no. 9, pp. 879–895, 2013.
- [3] M. Milligan and B. Kirby, "Utilizing Load Response for Wind and Solar Integration and Power System Reliability," in *Proc. of WindPower*, 2010.
- [4] T. Logenthiran, D. Srinivasan, and T. Z. Shun, "Demand side management in smart grid using heuristic optimization," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1244–1252, 2012.
- [5] Y. Zhang, N. Gatsis, and G. B. Giannakis, "Robust energy management for microgrids with high-penetration renewables," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 944–953, 2013.
- [6] M. Shadmand and R. Balog, "Multi-objective optimization and design of photovoltaic-wind hybrid system for community smart dc microgrid," *Smart Grid, IEEE Transactions on*, vol. 5, no. 5, pp. 2635–2643, Sept 2014.
- [7] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.
- [8] R. Chicheportiche and J.-P. Bouchaud, "The joint distribution of stock returns is not elliptical," *International Journal of Theoretical and Applied Finance*, vol. 15, no. 03, p. 1250019, 2012.
- [9] R. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.
- [10] A. Garas *et al.*, "The structural role of weak and strong links in a financial market network," *The European Physical Journal B*, vol. 63, no. 2, pp. 265–271, 2008.
- [11] "Irish Social Science Data Archive." [Online]. Available: <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>



- [12] M. Lydia, S. S. Kumar, a. I. Selvakumar, and G. E. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 452–460, 2014.
- [13] "Infoclimat." [Online]. Available: <http://www.infoclimat.fr>
- [14] M. Anvari, G. Lohmann, M. R. R. Tabar, M. Wächter, P. Milan, D. Heinemann, J. Peinke, and E. Lorenz, "Suppressing the non-Gaussian statistics of Renewable Power from Wind and Solar," 2015.
- [15] T. Schank and D. Wagner, "Finding, Counting and Listing all Triangles in Large Graphs, An Experimental Study," *Framework*, vol. 001907, pp. 3–6, 2001.