Correlation among smart grid actors

Nicolas Gensollen, Monique Becker, Vincent Gauthier 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 a large penetration 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 study the formation of coalitions of these futuristic agents that both produce and consume. These coalitions, which can be seen as aggregators of the electricity market, aim at selling the aggregated surplus of production of the agents they control. We rely on real weather data sampled accross stations of a given territory in order to simulate realistic production and consumption patterns for each agent. 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. We propose here an algorithm that 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.**

I. INTRODUCTION

Designing stable power systems is a classical engineering challenge since blackouts can have catastrophic consequences. One obvious condition for stability revolves around sustaining at any time a power generation that meets the demand. If one is larger than the other, the system deviates from its synchronous state. If nothing is done to return to the synchronized equilibrium, this can lead to catastrophic cascades of failures. With traditional power plants based on fossil energies, the production can be scheduled in order to sustain the predicted consumption. Deviations from this schedule can then be supported almost on real time by using fast response power plants, interconnections with border countries, as well as regulator entities. Even if most individual consumers may not realize it, the electicity prices are not constant rates and evolve with the production/consumption conditions on the grid. It is often assumed that the granularity of these prices is meant to increase in the future as a way to pass on the production conditions on the end users. The pricing of electricity and the necessity for the grid operator to have different kinds of emergency reserves lead to an economy setting for electricity, where operations and various kind of contracts are decided on a market. Such a market environment clearly necessitates communication in order to monitor and obtain information

from the grid as well as exchange between participants. These kind of settings are already being used at the transport level of current power grid with large renewable plants in the production portfolio. Using electricity markets appears thus as a way to manage the reliability of the whole system, especially when the penetration of renewables is important.

The smart grid vision however goes further and revolutionalizes this top-down centralized architecture by assuming that bidirectional power flows are allowed. This would change completely the nature of the traditional grid since production could be located down to the very end of the distribution networks. Nodes that were simply pure loads yesterday could behave tomorow as generators or loads. On the other hand, the penetration of renewables in the production is continuously growing, and is expected to become majority in a near future. These plants completely rely on the presence and the intensity of their respective ressources (wind, sun, tiles...). Balancing production and demand in such a scenario appears much more challenging and it seems inevitable that the grid should modernize its infrastructures to sustain this transition. The key to drive safely such a system is assumed to be information, and more precisely, the capacity to measure, communicate, and analyze data on real time.

In this paper, we focus on these "*nodes*" in the distribution network that can produce and consume electricity. More precisely, we consider a set of agents that own both renewable distributed energy ressources (DER) and electrical loads. The production of an agent can be, of course, used to meet its own demand, but in cases where 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*" [1] (and will be called accordingly in this paper). The primary objective of this work is to mix the interests of the grid operator with the ones of the prosumers in a single utility function. While the latters intend to maximize their benefits with higer production contracts, the operator primary concerns are linked to the quality and stability of the injections. The coalitions that we wish to form should thus be both stable and productive as to satisfy both parts. This compromise can be difficult to obtain in real situations, especially when renewable generation and complex spatiotemporal correlation relationships come into play.

As beautiful as this distributed participatory production scenario seems to be, the constraints necessary for the electrical grid to remain in a stable state forbid any anarchic system where prosumers inject power without proper control.

Actually the architecture and control of such a distributed grid creates numerous interdisciplinary challenges. In this paper, the envisaged solution for organizing such a multi-producers system consists in extending the kind of market environement discussed in the previous section. Participating aggregators estimate a production capacity for an upcoming period of time and announce it on the market. If a contract between the grid and a participant is bought, the latter commits to injecting exactly, at any time of the contracted period, the announced amount of power, under financial penalities if it fails.

In this paper, we consider that the production of a prosumer comes from renewables. Contracts are therefore based on predictions and come with a given amount of risk due to the intermittent nature of most renewables. Using storage as buffers in order to reduce riks is a popular approach. Actually, a whole branch of the smart grid litterature is even studying the possibility to use electric vehicles as moving storage capacities for stabilizing the grid. Without storage and proper control, the over-producing state of the agents, and thus the power they inject in the grid, might be rather unstable. This is clearly unacceptable for the grid operator that cannot ensure system stability if it has to deal with numerous small unpredictable entities.

Forming coalitions/aggregations is an envisaged solution to both the number of entities sollicitating the operator, and 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, since an aggregator can be considered as a single node by the level above it. On the other hand, it is well-known that diversification of the assets is a way of minimizing risks when constructing a portfolio. One thus expect a more stable and predictable energy production for a coalition than for a single agent. Nevertheless, all coalitions are obviously not equally stable or productive, which means that special attention should be paid to the aggregation step.

Furthermore, in our settings, since the prosumers have both a consumption and a production component, each depending on location and time, the diversification of the aggregations seems not trivial. This is a central topic of the present paper : given N prosumers, what coalitions should be formed so that the compromise between expected production and variability is optimized ?

We will see that variability in the aggregated productions can be quantified to a certain extent by the correlation among the agents forming the coalitions. Understanding the correlation reliationships 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 has the possibility of specifying both a maximum amount of risk and a minimum production acceptable to enter the market. Based on predictions, coalitions can decide wether or not to enter and what power quantities they are willing to provide. Of course, coalitions failing at fulfilling their obligations during the contracted period of time (because of bad forecasts, unexpected rare events, or intentional false information) will be exposed to financial penalities.

Because agents are susceptible to fail for diverse reasons, the impact on the system's stability is critical. The propensy of a system to undertake these failures is often refered to as its resilience [2]. Once formed, we also 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, with our algorithm, coalitions tend to be less impacted by random failures.

The paper is organized as follows, section II gives a brief overview of the related litterature, 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.

II. Related Work

The intermittent nature of renewables such as wind or solar power introduces new challenges in control design. On the contrary to fossil plants whose production can be scheduled in advance to meet the expected consumption, renewables by definition only produce when the ressource is present. Unfortunately, these moments do not necessarily coincide with the consumption peak hours [3]. A first approach to remedy this situation by acting on the consumption component is the use of dynamic ectricity prices sent to the end users. Together with demand side management techniques implemented on the smart meters, they consitute a first tool for shaping the consumption profiles [3]. In periods when the production is expected to be greater than the consumption, prices are scheduled low in order to encourage end users to delay their elastic loads to these periods. On the contrary, when the production is expected to be less than the consumption, high prices are scheduled to deter any delayable loads over these periods.

Another popular approach on the production side, is to combine renewable generators with storage devices, such that these are charged when there is a surplus of production, and discharged when the consumption exceeds the production. Although simple, this idea causes numerous challenges. Depending on the size of the system considered, centralized or decentralized control is desirable. In[], the authors introduce a distributed energy management system with a high penetration of renewables such that power is scheduled in a distributed fashion. In [] the optimal storage capacity problem is adressed.

There is indeed an interesting tradeoff 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.

Although there are great progresses in stabilizing the production by demand side management or storage architectures, the need for good prediction techniques is more important than ever. Understanding the statistical properties of wind or solar irradiance are already well-established areas of research. There exists indeed a very interesting litterature on techniques aiming at predicting wind or solar irradiance profiles for future periods with error margins. In this paper, we adopt a slightly different approach : we suppose that the agents have historical records of their production and consumption, and we seek aggregations of agents with high expected production and low variability.

Optimization of expected returns to risk is a traditional goal in finance. It is indeed well-known, that the more risk one is willing to take, the higher his potential gains. On the contrary, when investing exclusively on low rik assets, one should expect relatively small gains. This tradeoff is formalized in the Markowitz' portfolio theory [4]. 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 portolios that all optimize in some sense this tradeoff. 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.

This problem exhibits similarities with our objective of forming stable coalitions since our expected return can be formulated as the expected production and the risk / volatility as the variance of the production. However, it is assumed in the portfolio theory [4] that returns are normaly distributed. In this paper, we chose to use normal distributions only for explanation purposes, or for estimating some parameters. Therefore, all the results are computed based on observed distributions (sometimes far from gaussians).

One of the key point in the Markowitz theory is to consider explicitly the correlation structure between the assets since these correlation relationships impact directly the variances of the portfolios. Since the work of [5], an interresting approach consists in computing a distance metric based on the correlation coefficients in order to organize the series in a correlation graph where the weight of an edge between two series (nodes) is the metric value for these series.

Because the metric can be computed for all pairs, these graphs are complete and of little use as is. Historically, the approach used by [5] was to compute a minimum spanning tree over the correlation graph as to extract a correlation structure of the form of a hierarchical clustering. Later on, it was pointed out that, by definition, a spanning tree could not capture the underlying clustering structure hidden in the correlation graph. Another filtering approach by mean of a threshold ϵ on the edge weights solves this problem and allows one to exhibit clusters of correlated series. In the following we will refer to these filtered graphs as ϵ-graph.

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 notwithstanding being conducted and data are progressively made public [6].

In the following, we denote by $P_i(t)$ the instantaneous extra-production of agent i at time t :

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

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. As explained above, since large datasets containing this quantity over time are not yet available, 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 ressources (DER_i) and loads ($load_i$) of i :

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

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

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

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

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

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

Where $\mathcal{F}_{WT}$ (resp. $\mathcal{F}_{PV}$) is the power curve for the windturbines (resp. photovoltaic panels). We made here the implicit assumption that all windturbines (resp. photovoltaic panels) have the same power curve. The model can be easily extended to mupliple power curves accounting for different types of generators. More details about power curves and their approximations can be found in the appendix and in [7].
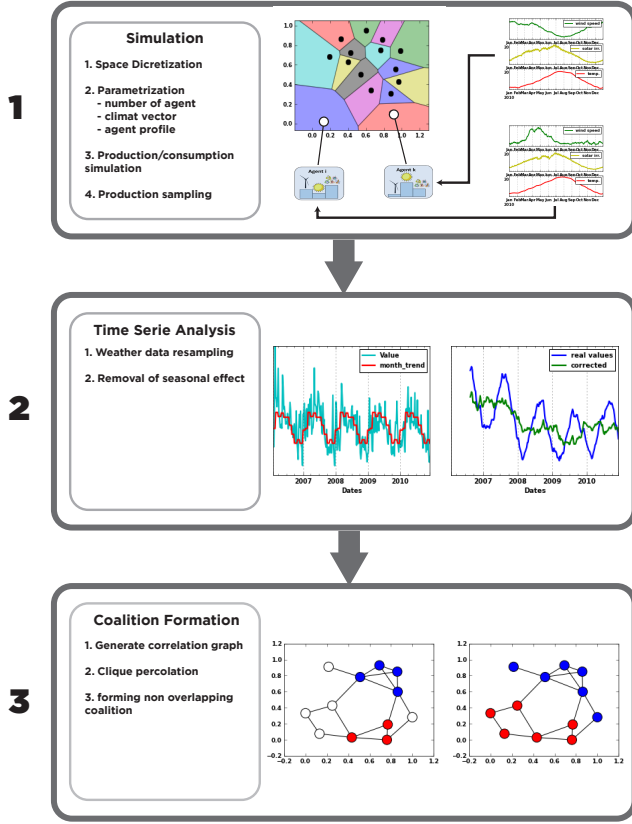
3

Figure 1: Process diagram

Weather quantities like wind speed or solar irradiance appear thus as alternative data for generating the P_i series. Fortunately, these kinds of data are easier to find, and since the developement of small personal weather stations, their geographical granularity keeps increasing. A quantity like wind speed depends both on time and location (we neglect height since this information is rarely present in the data). In the following, time is discretized into slots and space into zones. A zone is simply a portion of the considered space for which we have weather 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. The process for generating the P_i series is pictured in the first block of the process diagram (see figure 1).

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 windturbine 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 [8] starting in january 2006 and ending in december 2012, with a sampling frequency of three hours, and generate N timeseries of extra-production over this date range.

IV. NOTATIONS

This section provides most of the notations and introduces important concepts for the rest of the paper. As explained in section III, we consider a set $\mathcal{A} = \{a_1, a_2, ..., a_N\}$ of N prosumers configured randomly, and for each agent, 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 (the so-called coalitions) 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 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 penalities if it deviates. Obviously, the actual extra-production will not be constant at this value and will oscillate due to intermittencies in the production and consumption. If S always produces more than P_S^{CRCT}, it will never have to pay penalities, but 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 (see section II), 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 penalities. We can see that there is a return over risk tradeoff here, meaning that coalitions should find the right balance between announcing too low and losing some potential gains, and claiming too high and paying penalities.

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. Using simple statistics, we can write the distribution of the coalition $S = \{i, j\}$ as $P_{\{i,j\}} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$, where :

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

ρ_{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 where $\{i, j\}$ 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 penalities to the grid. The probability that $\{i, j\}$ is underproducing compared to the contract : $Pr[P_{i,j} \le P^{CRCT}]$

is thus an important indicator of the coalition's quality. A well-known result for normal distributions is that the cumulative distribution function can be written as :

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

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

The amount of risks a given coalition is willing to take depends on a lot of things, among wich its capacity to compensate for under-producing (using batteries, backup generators...), or the risk aversion of the aggregator. 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 a little bit by giving some responsabilities 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 $P_S^{CRCT\star}$ such that $Pr[P_{ij} \le P_S^{CRCT\star}] = \phi$. In the gaussian example, this translates by coalition $\{i, j\}$ announcing :

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

This is the best contract value that the coalition S can afford giving the stability policy ϕ of the grid operator. Figure 2 shows how $P_S^{CRCT\star}$ evolves according to the reliability parameter ϕ. For illustration, the range of ϕ values is shown from 0 to 1, but in practice, only small values of ϕ 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 2, coalitions with high expected productions but presenting a high unpredictability are penalized and can only afford small contracts.

In order not to overload the market with unrealistically small coalitions, the grid operator also specify a lower bound P^{MIN} on the contract values. We thus characterized a valid coalition as one satisfying the two conditions :

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

On figure 2, 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 $P_S^{CRCT\star}$ analytically. Nevertheless, such assumption is rather unrealistic in practice. In the following, we keep the same framework but release this gaussian assumption unless the contrary is specified. This assumption will indeed be convenient for computing some parameter estimates.
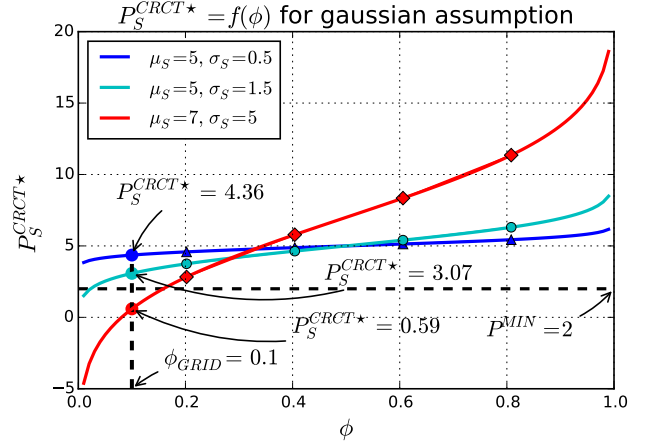

Figure 2: $P_S^{CRCT\star}$ depending on reliability parameter ϕ for gaussian distributions (see equation 8). Blue curve with triangles stands for a coalition S with an expected production of 5 units and a standard deviation of 0.5. Under a grid policy of $\phi = 0.1$, it is able to announce a contract value of $P_S^{CRCT\star} = 4.36$. 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 $P_S^{CRCT\star} = 3.07$. 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 ϕ, 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 horrizontal dashed line).

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 proportionnally to their contract values $gain(S) \propto P_S^{CRCT\star}$. More precisely, if λ is the unitary price rate for electricity, a coalition S injecting $P_S^{CRCT\star}$ in the grid during a period $[t_0, t_k]$ earns :

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

(since $P_S^{CRCT\star}$ is supposed to be a constant rate over the contracted period). Using $gain(S)$ directly as a utility function suffers a major drawback. It is indeed not a concave function of the coalitions' sizes, meaning that coalitions can grow as large as the number of agents allows it, without any counterparty.

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 ($|S|$):

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

Where parameter α controls to what extent the size of a coalition impacts its utility, and P^{MAX} is a normalizing factor. P^{MAX} can be seen as the maximum production which can be injected in the grid.

Based on $\mathcal{U}$, the marginal contribution of an agent i can be expressed as $\delta_S(i) = \mathcal{U}(S + \{i\}) - \mathcal{U}(S)$. A coalition S has thus an interest in adding an additional agent i if this marginal contribution is positive :

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

If α is set to zero, agents are added as long as they increase the contract value of the coalition. If α is greater than zero, additional agents have to increase the contract value by some factor. Since tunning the utility function with α is not necessarily convenient, we relate α to the mean sizes of the coalitions $\bar{N}$. If we wish that $\mathcal{U}(S)$ tends to form coalitions of size approximately of the order of $\bar{N}$, then :

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

In order to get an estimator for α, we solve equation 13 in a gaussian case (as in section IV). 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 α in these conditions yields :

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

Figure 3 shows how $\alpha^{\star}$ and the utility function evolves according to the mean size of the coalitions $\bar{N}$. These nice curves are of course only valid in the simplified example considered here, but they will provide some guidance when using real data.

As can be pointed out, the purpose of $\mathcal{U}$ is not a study of coalitions stability against player defection, which game theory provides a lot of tools for. This is indeed a problem on its own. The redistribution of a coalition's utility in terms of individual payoffs will therefore not be considered directly in this paper. $\mathcal{U}$ can be rather interpreted as a measure of how good a given coalition is according to our criteria since it favors coalitions with good production to risk ratios.

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. As the computation time of this quantity is not negligable, and is increasing in the sizes of the aggregations, we explain in this section the heuristic we used to compute the coalitions.
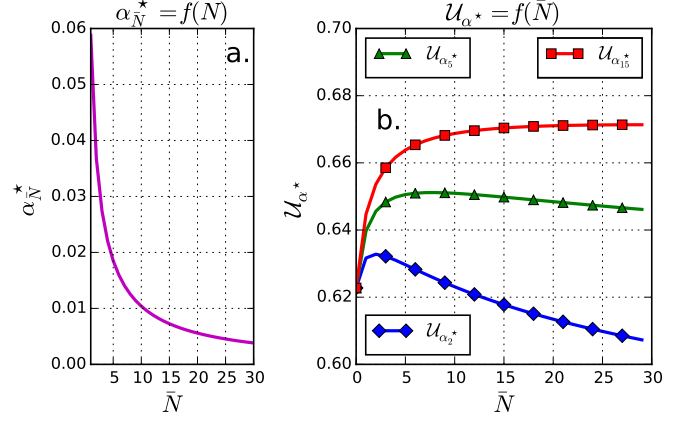

Figure 3: Gaussian mean approximation. Subplot **a** shows how the parameter α of the utility function should be chosen in function of the mean desired size of the coalitions (see equation 14). Subplot **b** dislays the corresponding utility functions for different values of α. Blue curve with diamonds favors very 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.

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 for this mapping :

$$\begin{cases} d_{ij}^1 = 1 - \rho_{ij}^2, \\ d_{ij}^2 = \rho_{ij}^2 = 1 - d_{ij}^1 \end{cases} \quad (15)$$

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 ϵ-graphs of G_1 and G_2 such that only meaningful edges remain. Selecting the right filter ϵ is thus an important point since it affects the landscape

search for the coalition formation. Unfortunately, there seems to be no clear consensus in the litterature 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 thus select our optimal threshold for G_2 as :

$$\epsilon^\star = min_{\epsilon \in [0,1]} \{\epsilon \ s.t. \ |\Theta_k(G_2^\epsilon)| \geq N_{COAL}\} \quad (16)$$

Where G_2^ϵ is the de-correlation graph G_2 filtered by ϵ, and $\Theta_k(G)$ is the set of non overlapping cliques of size k in a given graph G. In other words we select $\epsilon^\star$ as the smallest threshold posible such that the filtered de-correlation graph contains at least N_{COAL} non overlapping cliques of size k. The existence of $\epsilon^\star$ as defined in equation 16 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 as de-correlated starting seeds

Correlation can be seen as cosine of angles in L^2, hence even if there is no strict transivity relation for correlation, there is, to a certain extent, some partial notion of it. More precisely, if a, b, and c are three items such that $\rho_{ab} > \delta$ and $\rho_{bc} > \delta$, then we know, by the cosine addition formula[1], that $\rho_{ac} > 2\delta^2 - 1$. That is, if a and b are strongly correlated ($\rho_{ab} > 0.9$) and b and c are also strongly correlated ($\rho_{bc} > 0.9$), then there is a high probability for a and c of being strongly correlated ($\rho_{ac} > 0.62$).

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^ϵ is much higher than the one of G_2^ϵ. This can be seen as another formulation of [9] 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^ϵ 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^ϵ, which is a complete subgraph of G_2^ϵ. This is indeed a structure of interest for our purpose. 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 ϵ.

Figure 4 shows the distributions of the utility values for cliques of size 3 (triangles) in $G_2^{\epsilon^\star}$ 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 $G_2^{\epsilon^\star}$ as coalitions seems therefore appealing. Nevertheless, the quality of the results seems to
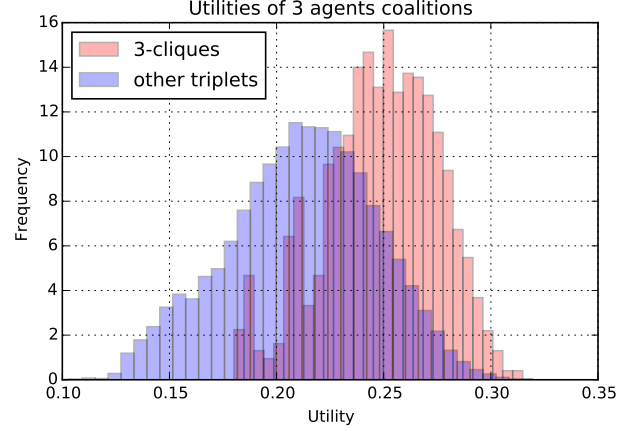
[1] $cos(a + b) = cos(a)cos(b) - sin(a)sin(b)$

Figure 4: Histograms of utility values for coalitions of size 3. Red bars stands for cliques in the decorrelation graph, and blue bars represents all the other possible triplets. As visible, cliques tend to exhibit higher utilities than randomly selected coalitions.

decrease as the sizes of the cliques increase. Indeed, the larger the desired cliques, the more dense $G_2^{\epsilon^\star}$ becomes (see equation 16). There is a point where cliques results 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 a higher unstability. This can be quantified by the marginal benefit of 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}, ϕ),
- **The "user" :** Number of desired coalitions N_{COAL} and size of starting cliques k.

The first steps consists in computing the de-correlation graph G_2 as well as the optimal threshold $\epsilon^\star$. Cliques of size k in $G_2^{\epsilon^\star}$ are considered as coalition seeds. The next step is a local greedy improvement over the landscape represented by $G_2^{\epsilon^\star}$. Cliques add alternatively the node $i^\star$ 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. This addition occurs only if $i^\star$ is not already involved in another coalition, and if $\delta_{clique}(i^\star) \geq 0$, meaning that utilities are increasing. The algorithm stops when all nodes are distributed in a coalition or when the global utility stops increasing. See the details in algorithm 1 in the appendix.

7

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. Algorithm 2 uses such a sampling and returns the highest utility coalition structure sampled. This algorithm will be refered to as "*random*" in the following.
- **Correlated :** This is the complete opposite of our algorithm. It uses the correlation graph G_1 and performs a community detection. The resulting coalitions have thus very high internal correlations. We thus expect this algorithm to perform very bad compared to the others. See algorithm 3.

Before running the algorithms, we need to calibrate the utility function by choosing the value of the α 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, neither the communication network nor the electrical grid are explicitly considered. Thus, we do not have any technical constraints on coalition sizes even if we designed the utitility such that these could be taken into account. We select the desired size as being $\lfloor N/N_{COAL} \rfloor$ (where $\lfloor . \rfloor$ means floor). Figure 5 shows how the mean utility of a coalition evolves with its size when the optimum size is set to 40 agents. Using equation 14 to estimate α based on the mean quantities and gaussian approximations seems to give acceptable results for the utility function behavior.

Figure 6 displays the evolution of the global utility and the number of involved agents during the course of the greedy algorithm 1. The transition from unvalid to valid coalitions is clearly visible on the blue diamond curve and occurs between iteration 10 and 15. After this transition, coalition's utilities improve slowly up to a maximum point.

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
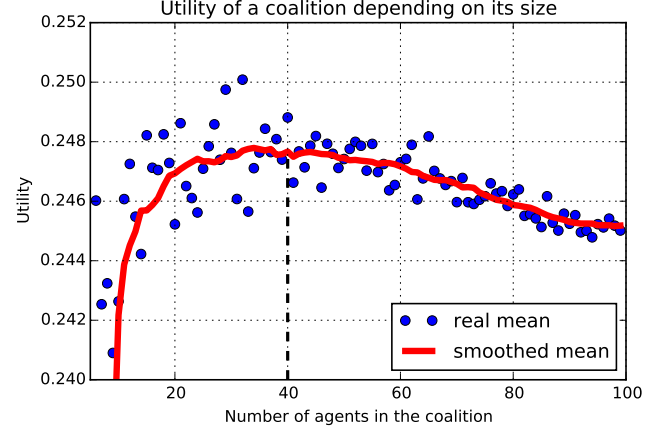

Figure 5: Utility of random coalitions depending on their size. Blue dots shows real mean utility values and the thick red curve its smoothed version by applying a Savitzky-Golay filter. On this plot the α parameter of the utility function was selected according to equation 14 in order to favor 40 agents coalitions.
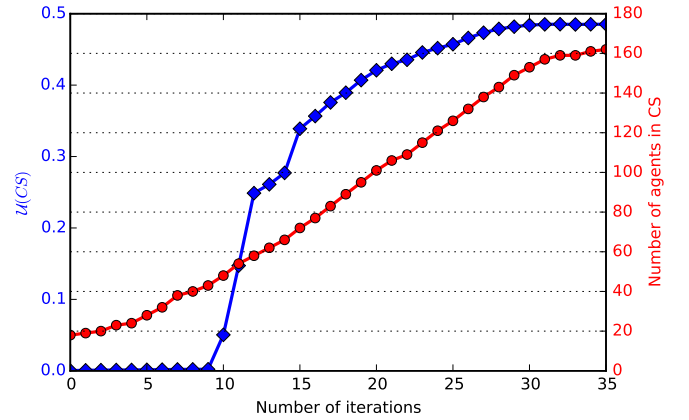

Figure 6: Evolution of the global utility $\mathcal{U}(CS)$ (blue diamond curve, left axis) and the number of agents involved in the coalitions (red circle curve, right axis) during the greedy optimization of algorithm 1

the marker size is also proportional to the coalition size. We can see that the utility function results in approximatively balanced 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.

On figure 7, 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 readalibility, but are usually quite large since
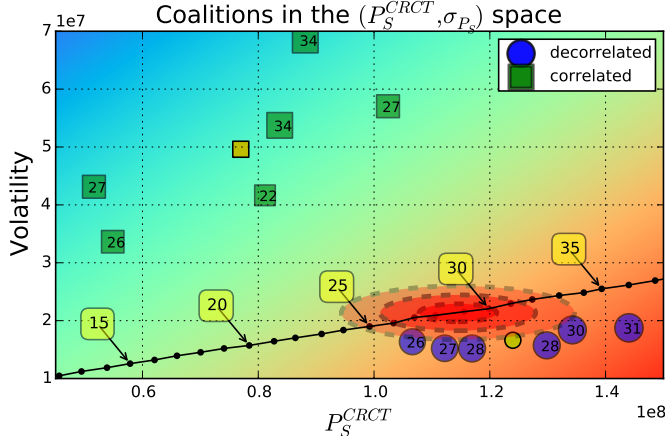
8

Figure 7: Coalitions formed in the (contract value, volatility) space. The color map indicates qualities of portions of the plane. The closer to red the better (high contract values with small volatility). On the opposite, blue areas shows 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 stands 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 increase (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. As we can also draw random coalition structures, we show the distribution of their gravity centers by the red ellipses (center is the mean, and each ellipse add 1 standard deviation). For comparison with our algorithm where the utility function favors balanced coalitions, we also constrained loosely the sizes of the coalitions in the random process.

this sampling only take the size as a constraint. Moreover, note that only a subset of the points are labeled with the size they represents as not to overload the figure. 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.

On figure 7, the results of the random coalition structure sampling is 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 can be roughly described as its ability to perform its tasks when subject to failures of its components. Therefore, the notion of resilience we will use in the following can be seen as the
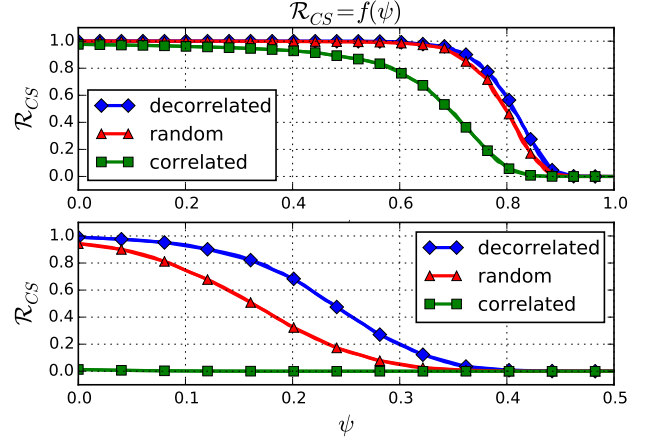

Figure 8: Resilience of the coalition structures when nodes fail randomly (see equation 18) for $P^{MIN} = 10MW$ (top subplot) and $P^{MIN} = 80MW$ (bottom subplot)

ability of the coalition structures to inject stable power in the grid when node failures occur. According to our model, the grid operator specified two thresholds (P^{MIN} and ϕ) such that the power injected by every coalition is constrained : $P_S^{CRCT} \in [P^{MIN}, P_S^{CRCT\star}]$. 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 :

$$\mathcal{R}_S = Pr[P_S >= P^{MIN}] = 1 - Pr[P_S < P^{MIN}] \quad (17)$$

And we extend this measure to the coalition stuctures :

$$\mathcal{R}_{CS} = \prod_{S \in CS} \left(1 - Pr[P_S < P^{MIN}]\right) \quad (18)$$

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 evolve according to ψ. On the top subplot, P^{MIN} was volontarily 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 de-correlated coalitions seem to achieve a more resilient production on the market in the sense that they are able to 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. As the correlations between the prosumers that

form these coalitions impact directly their volatilities, we 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" graph of the agents and makes local improvements offers a good compromise between speed and quality of the results. We compare 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 they tend to have globally better production over volatility ratios, these coalitions will tend to use less storage and spill less energy than more unstable ones. We plan to study these benefits for the control of the aggregations in future works.

Because in real situations, agents are prone to failure, resilience is also an important criterium for the quality of the aggregations. We therefore studied how the coalitions are able to remain on the market when their agents fail randomly one by one. We showed that, in this situation, the coalitions resulting from our algorithm better withstand losses of agents.

REFERENCES

[1] A. J. D. Rathnayaka *et al.*, "Prosumer management in socio-technical smart grid," in *Proceedings of the CUBE International Information Technology Conference*, 2012, pp. 483–489.
[2] S. Pahwa, C. Scoglio, and N. Schulz, "Topological Analysis and Mitigation Strategies for Cascading Failures in Power Grid Networks."
[3] M. Milligan and B. Kirby, "Utilizing Load Response for Wind and Solar Integration and Power System Reliability," no. July, 2010.
[4] H. M. Markowitz, "Portfolio Selection: Efficient Diversification of Investments," pp. 217–234, 1959. [Online]. Available: http://www.informaworld.com/openurl?genre=article&doi=10.1080/00137917908965215&magic=crossref
[5] R. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.
[6] [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/
[7] 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. [Online]. Available: http://dx.doi.org/10.1016/j.rser.2013.10.030
[8] [Online]. Available: http://www.infoclimat.fr
[9] 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.
[10] [Online]. Available: http://www.ncdc.noaa.gov/
[11] G. Dans and F. Science, "Rayonnement solaire," vol. 360, 2007.

APPENDIX

A. Algorithms

B. Net production series

Data were collected from [8] (similar data can be found at [10] for the United States). The variables used in the simulation are :

- Average wind speed (in $m.s^{-1}$)
- Nebulosity (integer in $[0, 8]$)
- Temperature (in degree celsius)

Data: P_i series,
Grid policy (P^{MIN}, ϕ),
Desired number of coalitions N_{COAL},
size of starting cliques k
Result: $CS = \{S_1, ..., S_{N_{COAL}}\}$
Compute $G_2^{\epsilon^\star}$;
Find the N_{COAL} cliques in $G_2^{\epsilon^\star}$;
while $\mathcal{U}(CS)$ *is improving* **do**
    for *each clique* **do**
        Find $i^\star$;
        if $\delta_{clique}(i^\star) \geq 0$ **then**
            $clique \leftarrow clique \cup \{i^\star\}$;
        end
        if $\exists j \in clique, \ s.t \ \delta_{clique}(j) < 0$ **then**
            $clique \leftarrow clique - \{j\}$;
        end
    end
end

Algorithm 1: Local greedy optimization algorithm

Data: Agent set $\mathcal{A}$,
Desired number of coalitions N_{COAL},
Maximum number of iterations $Loop^{max}$
Result: $CS = \{S_1, ..., S_{N_{COAL}}\}$
$N^{loop} \leftarrow 0$;
$CS^\star \leftarrow \emptyset$;
while $N^{loop} < Loop^{max}$ **do**
    $CS \leftarrow SelectRandomCS()$;
    if $\mathcal{U}(CS) > \mathcal{U}(CS^\star)$ **then**
        $CS^\star \leftarrow CS$;
    end
    $N^{loop} \leftarrow N^{loop} + 1$;
end
return $CS^\star$

Algorithm 2: Random algorithm

Data: P_i series,
Desired number of coalitions N_{COAL},
search step size $\beta << 1$
Result: $CS = \{S_1, ..., S_{N_{COAL}}\}$
$\epsilon \leftarrow 1$;
$CS \leftarrow \emptyset$;
while $|CS| < N_{COAL}$ **do**
    Compute G_1^ϵ ;
    $CS \leftarrow computeClusters(G_1^\epsilon)$;
    if $|CS| = N_{COAL}$ **then**
        return CS;
    end
    else
        $\epsilon \leftarrow \epsilon - \beta$;
    end
end

Algorithm 3: Correlated algorithm

1) Wind power curve: Power curves are functions that, for a given type of generator, map some input quantity to the output power produced. For windturbines and solar arrays these functions are well studied and approximations have been proposed [7] [11]. For the wind turbines, the power curve can be specified by 4 values :

- Cut-in-speed : The wind speed at which the turbine first starts to rotate and generates power.
- Rated-output-power : The maximum power that the turbine can generate.
- Rated-output-speed : The wind speed at which the turbine attains its rated output power.
- Cut-out-speed : The speed at which the turbine is turned off as not to damage the rotor.

The most interesting part is the increase of output power when the wind speed is in the cut-in-speed rated-output-speed range. Even if sometimes a simple linear model is used, the increase has been shown to be non linear and some more complex exponential fit can be found in the litterature [7].

2) Solar power curve: The input quantity desired for our power curve model for solar arrays is a radiance in $W.m^{-2}$, which can be difficult to find in weather station available data. As we mainly collected nebulosity series, we used the Helios model described in [11]. This model enabled us to compute perfect (clear blue sky situation) solar radiances at some specific locations on earth and at given timestamps. As nebulosity is a measure of the sky cloudiness, we can use the nebulosity series as degradation factors on the clear blue sky model (see [11] for more details) :

$$\begin{cases} \Psi real(t) = \Psi_{perfect}(t)\eta(t) \\ \eta(t) = 1 - 0.75\left(\frac{N(t)}{8}\right)^{3.4} \end{cases} \quad (19)$$

Where $\Psi_{perfect}(t)$ and $\Psi real(t)$ are respectively the clear blue sky and real radiances at time t, $\eta(t)$ is the degradation factor at time t, and $N(t)$ is the nebulosity index at time t.

Once we have input data in the forms of radiances, we compute the production of a solar array with the following simplified power curve :

$$\mathcal{F}_{PV}(\Psi real(t)) = S_{PV}\Psi real(t)e_{PV} \quad (20)$$

Where S_{PV} is the surface of the array, and e_{PV} is its efficiency. The very simple form of this power curve is due to some simplifications in order not to overload the model. For instance, it does not take into account angles and orientations degradations. These could be incorporated if needed by changing the power curve in the simulations.

3) Consumption: Modeling electric consumption has already been widely tackled in the litterature. Models can be basically divided in two main categories : Top-down and bottom-up approaches. Top-down techniques take aggregated consumption data as inputs and try to estimate individual consumption patterns while bottom-up methods use a fine modeling of users consumptions as to obtain realistic aggregated consumption curves. In this paper, we used a bottom-up model since the end user, or relatively small aggregations of end users, are in our interest. The main objective was to capture both daily patterns and seasonal variations of the consumptions. We assumed an additive model where the consumption of an agent is the sum of a seasonal heating term that depends on the outside temperature and an electronic consumption term that only depends on the hour of the day. By denoting $\tau(t)$ the outside temperature at timestamp t, we can express the consumption $P_i^D(t)$ of agent i at time t :

$$P_i^D(t) = \mathcal{F}_i^{heat}(\tau(t),t) + \mathcal{F}_i^{elec}(t) \quad (21)$$

Where $\mathcal{F}_i^{heat}(\tau(t),t)$ is the power curve that maps the temperature to a heating consumption, and $\mathcal{F}_i^{elec}(t)$ computes the consumption of agent i (other than heating) at a given hour of the day. In the simulation, all agents have a desired inside temperature T_i, supposed to be a constant for simplification. By using termodynamic laws $\mathcal{F}_i^{heat}(\tau(t),t)$ can be approximated by :

$$\mathcal{F}_i^{heat}(\tau(t),t) = \frac{B_i}{R_i}[T_i - \tau(t)] \quad (22)$$

Where B_i is the surface of thermal exchanges for agent i and R_i is their thermal resistance.

We denote by Ω_i the maximum consumption possible for agent i, which is basically the sum of all its appliances powers. We also denote by $\omega_i(t) = \{\omega_i(t_0),...,\omega_i(t_{24})\}$ the vector of the average fraction of Ω_i used for each hour. We can therefore write :

$$\mathcal{F}_i^{elec}(t) = \Omega_i(\omega_i(t) + \epsilon) \quad (23)$$

Where ϵ is a noise term. The vector $\omega_i(t)$ enables us to easily differentiate agent consumption behaviors. Business or residential areas for instance can be easily distinguised with this kind of model.