APPENDIX

A. Algorithms

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

Figure 1 displays the evolution of the global utility and the number of involved agents during the course of the greedy algorithm 1. The transition from an invalid to a 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.

B. Notations

Figure 2 presents all the notations used in the article.

C. Net production series

Data were collected from [1] (data for the United States at [2]). The variables used in the simulation are :

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

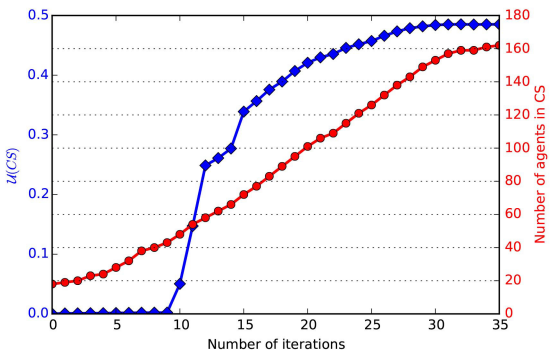

Figure 1: Evolution of the global utility $\mathcal{U}(CS)$ (blue diamond curve, left axis) and the number of agents involved in the coalitions (red circle curve, right axis) during the greedy optimization of algorithm 1

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 wind-turbines and solar arrays these functions are well studied and approximations have been proposed [3] [4], [5]. 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 literature [3].

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

N	Number of Prosumers
$P_i^P(t)$	Available production of agent i at time t
$P_i^D(t)$	Consumption of agent i at time t
$P_i(t) = P_i^P(t) - P_i^D(t)$	Available extra-production of agent i at time t
WT_i	Wind turbine set of agent i
PV_i	Solar array set of agent i
$DER_i = WT_i \cup PV_i$	DER of agent i
$\nu_i(t)$	wind speed (in ms^{-2}) at agent i's location and time t
$\Psi_i(t)$	solar irradiance (in Wm^{-2}) at agent i's location and time t
$\mathcal{F}_{WT}$	power curve for the wind turbines
$\mathcal{F}_{PV}$	power curve for the solar arrays
P_S^{CRCT}	Contract value of coalition S
$\phi \in [0, 1]$	Reliability Threshold
P^{MIN}	Minimum Contract value
α	Parameter that controls the coalitions sizes
ϵ	Correlation graph filtering threshold
N_{COAL}	Number of coalitions
$\mathcal{R}_S$	Resilience of coalition S
λ	unitary price rate for electricity
$\delta_S(i)$	marginal utility of agent i in coalition S
ρ_{ij}	Pearson's correlation coefficient between P_i and P_j
$\Theta_k(G)$	set of non overlapping cliques of size k in graph G
$\mathcal{R}_S$	resilience of coalition S

Figure 2: Notations

Helios model described in [4], [5]. 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 [4], [5] for more details) :

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

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 :

$$\mathcal{F}_{PV}(\Psi real(t)) = S_{PV}\Psi real(t)e_{PV} \quad (2)$$

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 literature. Models can be basically divided into 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 :

$$P_i^D(t) = \mathcal{F}_i^{heat}(\tau(t), t) + \mathcal{F}_i^{elec}(t) \quad (3)$$

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 thermodynamic laws $\mathcal{F}_i^{heat}(\tau(t), t)$ can be approximated by :

$$\mathcal{F}_i^{heat}(\tau(t), t) = \frac{B_i}{R_i}\left[T_i - \tau(t)\right] \quad (4)$$

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 :

$$\mathcal{F}_i^{elec}(t) = \Omega_i(\omega_i(t) + \epsilon) \quad (5)$$

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 distinguished with this kind of model.

D. Resilience algorithm

The following algorithm (see algorithm 4) explains the process through which the resilience of a coalition structure is explored. For a given coalition structure, we consider random failures of physical components of the agents. This results in the agent being unable to support its coalition. Therefore, the agent is simply removed from the coalition, which impacts the coalition's production. Eventually, after some number of agent failures, some coalitions might not be sufficiently stable and productive to fulfill the market's constraints. In this case, the coalition fails, and is removed from the market.

Data: $CS = \{S_1, S_2, ..., S_k\}$ Coalition structure ;
$pool = \cup_{k \in N_{COAL}} S_k$;
while $pool \neq \emptyset$ **do**
  Select randomly agent i from pool;
  i fails $\Rightarrow$ Remove i from its coalition :
  $S_k \leftarrow S_k - \{i\}$;
  $pool = pool - \{i\}$;
  if $Pr\left[P_{S_k} < P^{MIN}\right] > \phi$ **then**
    S_k fails $\Rightarrow$ Remove S_k from the market
  end
end

Algorithm 4: Random failures algorithm

E. Utility function

We have the following utility function (see main article) that attributes a real value to any coalition S depending on the best contract value that S can propose given the grid policy (P^{MIN}, ϕ) :

$$U(S) = \frac{1}{|S|^{\alpha}} \frac{P_S^{CRCT\star}}{P^{MAX}} \tag{6}$$

U depends on the specific agents within the coalition S because $P_S^{CRCT\star}$ is directly impacted by the correlations between the agents. U is also sensitive to the size of the coalition S through the term $\frac{1}{|S|^{\alpha}}$. The purpose of this term is to introduce in the utility function the following tradeoff : as a coalition grows, its contract value $P_S^{CRCT\star}$ tends to grow as well but the cost of coordinating all the agents also becomes larger. The parameter α is used in order to control to what extend the size of the coalition impacts its utility.

Knowing what value to use for α is not trivial and is important since it influences the shape of the utility function. For some range of values, the utility might not even be a concave function of the coalition size. Given a situation with N agents and a desired number of coalitions N_{COAL}, the algorithm 1 will start with N_{COAL} seeds and increase their sizes as long as the utilities are improved. The order through which coalitions incorporate new prosumers is randomized but leads generally to relatively balanced coalitions. We denote by $\bar{N}$ the desired approximate final size of the coalitions, i.e we target a utility function that is concave in the coalition size with a maximum at $\bar{N}$ (in the article, $\bar{N}$ is fixed at $\frac{N}{N_{COAL}}$). We propose to obtain such utility function through tunning α. Ideally, we seek to express α as a function of $\bar{N}$ and the agents power distributions and correlations.

In the general case, neither the power distributions nor the correlation structure have special forms which makes obtaining an analytical expression for α difficult. Therefore we use two simplifying assumptions in order to have an estimate for α. First, we approximate the power distributions by normal distributions, such that the optimal contract value can be written in the following way :

$$P_S^{CRCT\star} = \mu_S - \sqrt{2}\sigma_S erf^{-1}(1 - 2\phi) \tag{7}$$

With :

$$\begin{cases} \mu_S = \sum_{i \in S} \mu_i \\ \sigma_S = \sqrt{\sum_{i \in S} \sigma_i^2 + 2 \sum_i \sum_j \rho_{ij} \sigma_i \sigma_j} \end{cases} \tag{8}$$

Therefore, we can write the utility function as :

$$U(S) = \frac{1}{|S|^{\alpha} P^{MAX}} \left[\mu_S - \sqrt{2}\sigma_S erf^{-1}(1 - 2\phi)\right] \tag{9}$$

As μ_S and σ_S depends on the agents in S (and consequently on the size of S), we make a second simplifying assumption :

$$\begin{cases} \forall i \in \mathcal{A}\ \mu_i = \bar{\mu} = \frac{1}{N} \sum_{i \in \mathcal{A}} \mu_i \\ \forall i \in \mathcal{A}\ \sigma_i = \bar{\sigma} = \frac{1}{N} \sum_{i \in \mathcal{A}} \sigma_i \\ \forall i, j \in \mathcal{A}^2\ \rho_{ij} = \bar{\rho} = \frac{2}{N(N-1)} \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{A}} \rho_{ij} \end{cases} \tag{10}$$

That is, any agent is replaced by a mean approximation. In these conditions, we have :

$$\begin{cases} \mu_S = |S|\bar{\mu} \\ \sigma_S = \bar{\sigma}\sqrt{|S|\left[1 + \bar{\rho}(|S| - 1)\right]} \end{cases} \tag{11}$$

And the utility function becomes :

$$U(S) = \frac{\bar{\mu}}{|S|^{\alpha - 1} P^{MAX}} - \frac{\sqrt{(2)}\bar{\sigma}}{|S|^{\alpha} P^{MAX}} \sqrt{|S| + \bar{\rho}|S|(|S| - 1)} \tag{12}$$

$\bar{\mu},\ \bar{\sigma},\ \bar{\rho}, and\ P^{MAX}$ are all constants in these very particular conditions, such that U is a function of the coalition size with a parameter α. Depending on α, the utility function can have very different shapes.
We select $\alpha^{\star}$ such that :

$$\begin{cases} \left|\frac{\partial U}{\partial |S|}\right|_{|S| = \bar{N}} = 0 \\ \left|\frac{\partial^2 U}{\partial |S|^2}\right|_{\alpha = \alpha^{\star}} \leq 0 \end{cases} \tag{13}$$

Solving these equations yields the following expression :

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

Since $\alpha^{\star}(\bar{N})$ is obtained for a simplified case derived from the real situation, it should be considered as a soft constraint on the coalitions sizes.

Using such an approach on 100 prosumers examples with $N_{COAL} = 4$ and $\bar{N} = 25$ results in coalitions which sizes are comprised between 17 and 34 agents.

REFERENCES

[1] "Infoclimat." [Online]. Available: http://www.infoclimat.fr

[2] "National Centers for Environmental Information." [Online]. Available: http://www.ncdc.noaa.gov/

[3] 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.

[4] C. Piedallu and J.-C. Gégout, "Multiscale computation of solar radiation for predictive vegetation modelling," *Annals of Forest Science*, vol. 64, no. 8, pp. 899–909, 2007.

[5] ——, "Efficient assessment of topographic solar radiation to improve plant distribution models," *Agricultural and Forest Meteorology*, vol. 148, no. 11, pp. 1696 – 1706, 2008.