Decision support for Complex Systems: a Smart Grid case

Jose Evora, Jose Juan Hernandez, and Mario Hernandez

SIANI, University of Las Palmas de Gran Canaria, Las Palmas, Spain,
jose.evora@siani.es, josejuanhernandez@siani.es, mhernandez@siani.es

Abstract. Transitioning from traditional power grids to Smart Grids involves the use of a different approach based on complex systems to analyse the demand of power grids. This analysis provides information which supports the decision making when developing new policies for Smart Grids. These policies are designed and then tested through simulations since it is not possible to test them directly in a real power grid. Simulation output data can be analysed using a Business Intelligent approach in order to find out KPI (Key Performance Indicators) which support decisions that tune policies. The way in which the results management should be dealt with is through an OLAP (On-Line Analytical Processing) approach which enhances the capability of querying data.

Keywords: power grid system, complex system, power grid, demand side management, smart grid, data exploitation, decision support system, business intelligence, information visualisation, OLAP

1 Introduction

Climate change, the liberalisation of markets and other new requirements are pushing the energy sector towards a new paradigm known as the smart grid. This paradigm is characterised by the introduction of renewable energy sources (RES) in the power grids, new technologies such as storage mechanisms, massive integration of sensors and decision makers distributed along the grid or the introduction of a communication layer for the management and control of these technologies. The smart grid paradigm is also based on the use of Demand Side Management (DSM), the objectives of which include the minimisation of the peak demand and the system operation and planning improvement [2]. The system complexity is therefore increased and new tools are needed for the analysis and design of smart grids.

Due to the introduction of DSM in the Smart Grid, it is necessary to conceive new policies in order to perform this management which looks after the efficiency of power grids. This efficiency, among other factors, is related to the efficient use of the energy available at all times, which fluctuates mainly because of RES. However, Smart Grid policies which manage power demand require an arduous analysis of individual consumers and their devices. For this reason, demand requires to be analysed in a disaggregated manner, leading to the usage of a complex system approach to represent the power grid.

Since Smart Grid policies need to be thoroughly tested before their exploitation. The procedure to test these policies is made through simulations, as it is not possible to experiment them in a real power grid. Hence, it is necessary to run complex system simulations where the power grid is represented to test the policies and thus provide feedback about them.

In figure 1, a first iteration of the life cycle of a policy design for the Smart Grid is presented. At this level, and taking into account some high-level considerations about how it should be, a policy is conceived. In order to find out whether the policy will work well or not it is needed to perform a test. To this end, Key Performance Indicators (KPI)[5] must be designed since they are required to support the decision making process which will modify the policy. These KPI are intended to make visible information which is hidden in the data provided by simulations. After this, the simulation must be designed and developed according to the test requisites. Once the simulation has been executed, results will be available. As this simulation can correspond to a big power grid where every single device is represented (complex system approach[1]), the results the simulation provides could be huge. This output must be managed in a way that enables a fast querying system so that KPI calculations can be performed and used for the decision making process. This process will involve some changes in the policy design which shall be tested afterwards when another iteration is initiated.

The complexity of a system from the point of view of Smart Grid simulations is measured in terms of the amount of entities that are in and the relationships among them which produce an emergent behaviour. Therefore, the larger the amount of elements (i.e. entities and relationships) is, the more complex the system is considered. This statement is totally transportable to the results side. The quantity of results in a complex system simulation increases proportionally to the increment of the system complexity. For example, a system that has 10 000 entities with an average of 5 state variables that have to be exported involves that, at each time step, the system will be providing 50 000 results. If the simulation is executed during 2 000 steps, the amount of results provided at the end of the simulation will be about 100 million items.

In another context of Smart Grids, hot topics are all problems related to Energy Data Management, such as the collection and exploitation for business processes of energy consumption data from smart meters installed in power grids [4]. These two examples correspond to problems related to the management of huge amounts of data.

[1] This representation could be regarded as agent-based. From our point of view, an element is considered an agent whenever it exhibits intelligence [11]. As devices have a mechanistic behaviour, we do not consider them agents. However, simulations that include intelligent elements (i.e. people switching devices in households or units that apply smart grid policies) are considered agent-based.

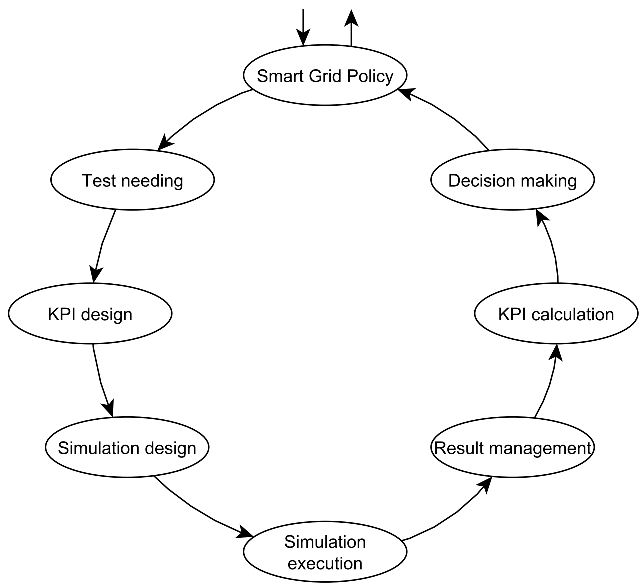

Fig. 1. Life cycle of a Smart Grid policy development

When a Smart Grid simulation is performed, the results management is one of the most important issues as they will feed the design process of the policy. The experience we have had in this field is that it is not possible to perform manually a thorough analysis on large amounts of results. When facing such amounts of data, people usually focus on some details for a certain amount of entities and then conclusions are extrapolated. It has been empirically observed that this analysis may cover a very small percentage of the result set. This implies that many other conclusions could never be found out and extracted from data remaining hidden.

At this point, the use of tools which assist result analysis must be considered in order to deal with this issue. Business intelligence (BI) [3] techniques can play an interesting role in this stage, since it is considered the set of strategies and tools that focus on administration and knowledge creation through data analysis. Among these strategies, some of them encourage the use of technologies such as OLAP (On-Line Analytical Processing) (e.g. Saiku [8]), information visualisation (e.g. Gapmind [7]) and all the data mining corpus which helps to identify and extract hidden or non-evident knowledge (e.g. Weka [9]). These three groups of technologies are especially important for this kind of decision making.

In this paper, the problem of dealing with data exploitation will be further detailed. Then, the OLAP approach to deal with this issue will be exposed. Finally, an example of a Smart Grid case will be presented where results of a simulation are managed following the OLAP approach in order to identify how it helps the decision support when designing Smart Grid policies.

2 Smart Grid simulation issues

Simulations play a crucial role in the design of Smart Grid policies since they are a way to test them before their launch. However, the output provided by the simulations must be managed in a way that allows the policy designers to make decisions. This section explains the main concerns when analysing results obtained in a Smart Grid simulation. When facing a simulation of Smart Grids based on a complex system approach, the results analysis becomes a difficult stage since the amount of entities is huge.

All systems containing a large amount of entities and relations in simulation processes provide a large amount of results. The way in which these data are normally exported is through data files. These data files are usually designed according to the data that will be managed thus avoiding the possibility of querying this data beyond what was decided to export. Therefore, whenever we deem it convenient to extract data, which was not considered to export at the design phase, a new simulation must be configured and executed.

In order to exemplify this issue, a disaggregated model of a power grid system is used. This system only consists of the demand side, which is disaggregated at the device level. It is precisely at this level where we can find a layer consisting of heterogeneous elements, since the characteristics to extract from a radiator are not the same as the ones from a television (TV). If we want to preserve all variables that are not common to every device, it will be necessary to export each device type into a different data sheet (Figure: 2). At this point, once the data exportation process has been defined, we can start thinking about querying it. The list below states some query examples and how they should be dealt with according to this data exportation structure:

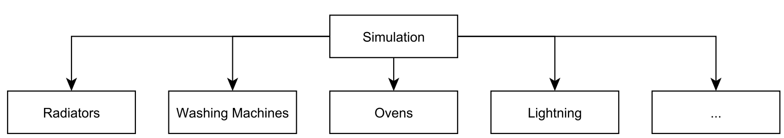

Fig. 2. Structure to export simulation results. Simulation outputs are exported to several files. Each file is related to the data produced by a device kind. For example, Radiators file contains information which regards the energy consumption of these devices

– **Querying the consumption of all devices.** This query is very likely to be required. According to our data structure, firstly we calculate the total consumption at each device type. This would involve opening as many files as device types and making the calculations to obtain the total consumption per device type. Secondly, those columns which have the aggregated value at each device type must be moved into a new sheet where the final calculation

would be performed obtaining the query result. The more device types there are, the trickier this process becomes.
- **Querying the consumption of all devices in a specific household.** This process would consist in gathering the columns belonging to all the devices contained in the household from the data files. Once they are all together in a new sheet, the query result can be obtained by adding up.
- **Querying the consumption of all devices in a specific district.** The process to obtain this query is really tricky. Firstly, all the devices belonging to a specific district must be listed. Next, all the columns which refer to the devices consumption must be gathered from the device type sheets following this list. Finally, all gathered columns can be moved to a new sheet where the query can be obtained.

Taking these examples into account, it is possible to imagine how tricky the results management of more complicated queries can get. Probably, some of these queries are easier to obtain by redefining the simulation results format and running it again. However, it would also be really tedious, and depending on the simulation kind, the results may differ from the previous simulation and in the end it would be necessary to start the result analysis from the beginning.

All these difficulties in querying the output of a simulation could involve that many other queries are not made due to the fact that they involve a strong and time consuming effort to perform them. Unfortunately, this usually leads to focus on a small subset of variables of the simulation neglecting much information and wasting too much time in performing simple queries.

The root of the problem behind the result analysis is that such results have a multi-dimensional and a multi-scale (namely temporal and spatial) nature which cannot be managed by using conventional data sheets. The example of the demand disaggregation is multi-dimensional and multi-scale. Multi-dimensional, since every data (for example, a power measure) is related to a specific device, location (household, building, district...) and time. Multi-scale, since the information can be aggregated at different time scales (per hour, per day, per month...) and at different spatial levels (device, household...).

3 OLAP

On-Line Analytical Processing (OLAP) is a solution used in BI, the aim of which is to accelerate querying large amounts of data. OLAP is based on cubes [1](Figure: 3), a multi-dimensional structure where data is stored. These cubes enable the insertion of data, namely facts, which are referred to several dimensions. For example, the measure of power taken from a washing machine can be referred to the device, the household where the device is and the time. Therefore, in this case, there would be three dimensions: devices, households and time.

The structure of an OLAP cube which addresses our problem is presented in the figure 4. Every cube consists of dimensions, measures and indicators. The list below describes every cube component.

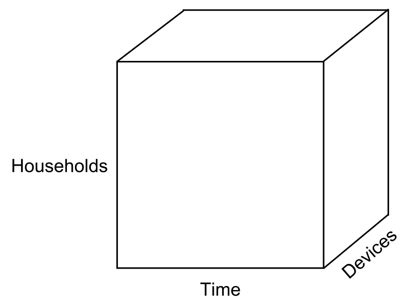

Fig. 3. An OLAP example of a cube for a power grid where every energy consumption is related to a household, device and time

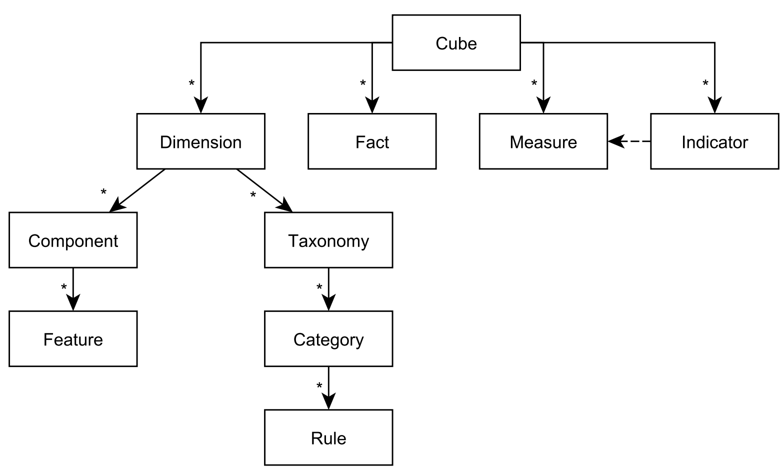

Fig. 4. An OLAP Cube structure. A cube contains dimensions, facts, measures and indicators.

- **Dimension:** it establishes a way to access the data inside the cube. Every single data is related to some elements such as when and where it happened. For example, a data of power consumption of a household would be related to the dimensions household and time.
  - **Component:** it is an element which is related to a dimension. For example, a dimension which concerns households would be filled by components which are households.
    - **Feature:** it is a property of the component. In case the components are households, a possible feature could be the number of square meters there is in each household.
  - **Taxonomy:** it is a way of categorizing a dimension. There are different ways to categorize the components inside a dimension. Each of these ways is known as taxonomy. In the example of the household dimension, a taxonomy could be the size or the orientation of the facade.
    - **Category:** it is a set of components that satisfy some specific conditions. For instance, possible categories for the size taxonomy could be small, medium or big. Therefore, each of these categories would contain a set of household components the relationship of which is having a similar size.
      - **Rule:** it establishes the condition that a component must meet in order to fall into the category that owns the rule. In the case of the small category, a possible rule could be: all the household components the feature of which *number of square meters* is below $80m^2$
- **Measure:** it provides a semantic to the data inserted in the cube, e.g. the power of the household mentioned above is just a number. However, the power measure is what provides the semantic to this number. A measure is usually related to a metric which enables the comparison among measures that are in different cubes. In this case, the metric of the power measure would be Watts.
- **Indicator:** it designates the way in which a measure or a set of measures are aggregated. For example, the power measure could be aggregated using an average (AVG) function. This way of aggregating measures is known as indicator. It is possible to have several indicators for one measure, i.e. the integral operator over the power measure would provide a second indicator over this measure which could be designated as energy indicator.
- **Fact:** it relates the measures of a cube with the dimensions. A fact indicates that a certain combination of values (measures) took place for a specific combination of elements (components). In other words, a fact can be understood as a relation of a state to a context. The state is a set of measures and the context consists of components including time. In figure 5, the state contains 20 (centigrades) and 135 (Watts) as measures. These measures are related to a context which indicates the time and household where those measures were taken.

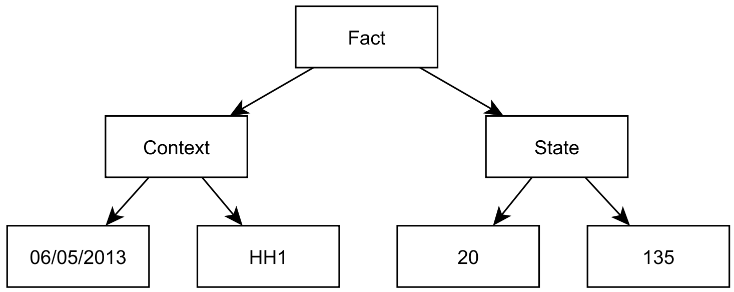

Fig. 5. A fact consists of context and state. The state is a set of measures which are related to components through the context

4 An OLAP Smart Grid example

In this section, all concepts exposed previously will be used in a practical case. Assuming that a new Smart Grid policy is to be tuned, several simulations of power grid demand will be performed. To make decisions, these simulations must focus on the power demand and the temperature at the residential sector. Therefore, the scenario for those simulations consists of several districts with households (Figure: 6). Each household contains several devices and calculates the internal temperature.

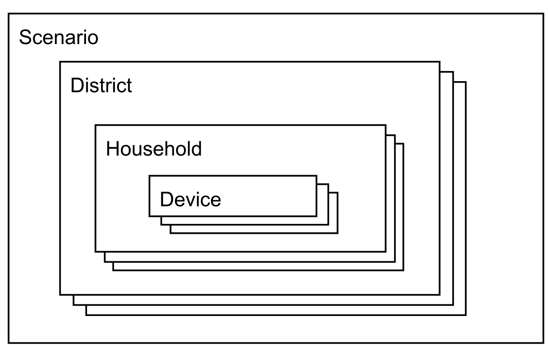

Fig. 6. Scenario composition. Several districts which contain several households and each household contains several devices.

To this end, several cubes have been designed so as to analyse the data coming from the simulation: first of all, the household cube which contains the facts regarding the temperature and, secondly, one cube per device type (TVs, Radiators and Washing Machines, among others) which contain facts about the devices. Since there are many kinds of devices in a household, in this example we are going to focus on two of them: TVs and radiators.

There are two dimensions in the household (HH) cube: one measure and one indicator (Figure: 7). The Time dimension is common to all cubes and configures a standard way of categorizing the timeline. Household dimension contains the households transformed into components which are described by features. The temperature of the household is the only measure that this cube is going to store and it will be aggregated using an average criteria according to the designation of the indicator.

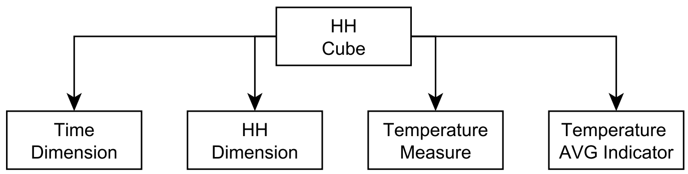

Fig. 7. Household cube. This cube contains two dimensions: time and household. Each fact will relate every temperature measure to a time and a household

The household dimension contains a taxonomy which concerns the locations (Figure: 8). This taxonomy is categorized following several levels: country, city and district. For instance, two household components have been included, both of which contain a feature which is their location using UTM coordinates. Therefore, these location features allow the dimension to identify which district each household is located in.

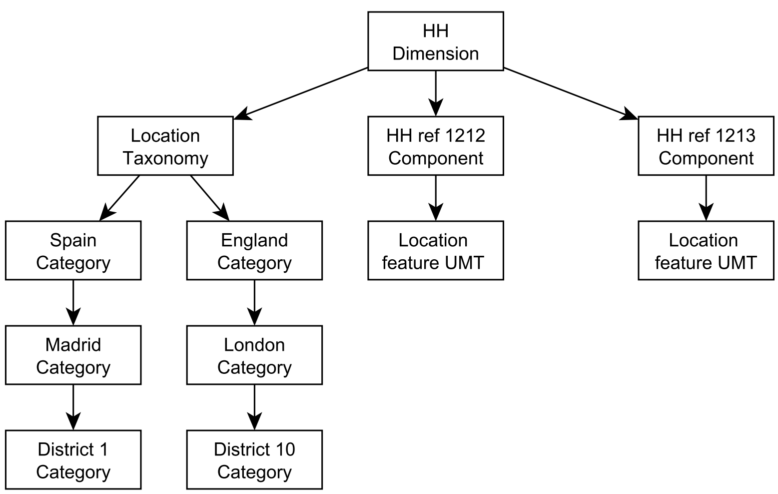

Fig. 8. Household dimension. This dimension contains a taxonomy that classifies the components in districts according to their location feature

The TV and Radiator cases are exposed in order to demonstrate why devices must be disaggregated into separated cubes. The main reason for this separation is due to the fact that both devices do not share the same features and, therefore, their classification methods are different. This separation enhances the capacity of making queries since it is possible to filter components by features that are only present in a specific kind of device.

The TV cube registers data about power consumption as well as the TV mode (off, standby and on) (Figure: 9). Every set of measures (power and mode) is related to three dimensions: time, household and TV. Time and household dimensions are exactly the same dimensions as the ones detailed above. The TV dimension contains information about the TVs in a component format. Furthermore, there are two indicators which are responsible for aggregating measures: the mode indicator, which performs a calculation that provides the percentage of TVs that are turned on, and the power indicator, which aggregates the power measures registered using an average formula.

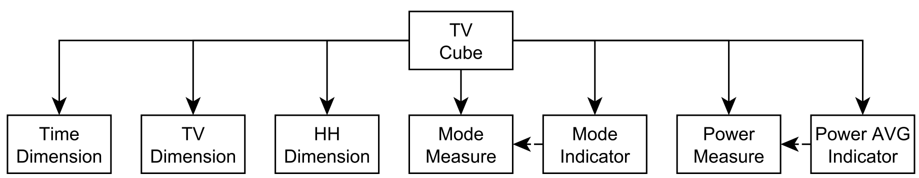

Fig. 9. TV cube. This cube contains three dimensions: time, TV and household. Therefore, each fact will relate a power and mode measures to a time and a TV which is located in a specific household

The TV dimension, like the household dimension, focuses on specific features related to TV components (Figure: 10). In this case, the possibility of filtering TVs using a technological criteria is considered relevant. Therefore, two categories have been created so as to separate LED televisions from LCD televisions. This information will allow us to compare the consumption among the different TV technologies. Hence, TV components contain the technology feature which will be used to calculate whether a TV belongs to the LED or LCD category by using the rules that are related to these categories.

The radiator cube stores measures related to both the power consumption and the thermostat level (Figure: 11). These measures are related to three dimensions, as in the case of the TV cube. In this case, apart from time and household dimensions, a new dimension has been designed: radiator dimension. This dimension contains components that represent radiators and their features. In addition, there are two indicators which aggregate the measures. On the one hand, the thermostat indicator aggregates the measures stored using a gradient function which shows big changes in the thermostat level in short periods of time. On the other hand, the power indicator aggregates the power measures using an average formula like in the TV cube.

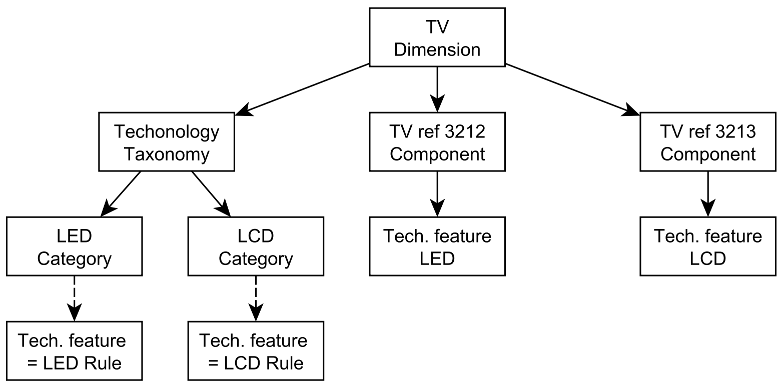

Fig. 10. TV dimension. In this case, the presented dimension has a taxonomy which classifies TV components according to their technology

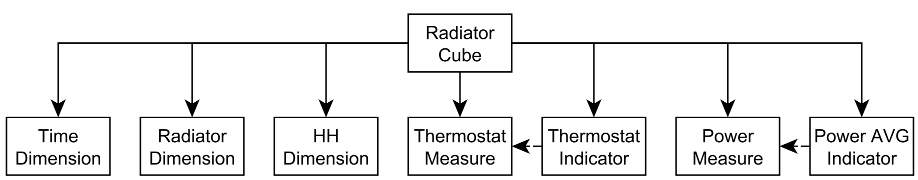

Fig. 11. Radiator cube. This cube contains three dimensions, as in the case of the TV cube. However, the measures are different since, in this case, the thermostat level of the radiator is stored. As it can be observed, devices are heterogeneous. This is the reason why devices have been separated according to a type criteria

The radiator dimension focuses on specific features which concern radiator components (Figure: 12). Since radiators are usually considered big consumers, a taxonomy to classify them into two groups has been designed. Indeed, this taxonomy will allow us to find out the amount of radiator components which are in what we consider a small consumer category (under 1kW installed power) or a big consumer category (over 1kW). Two components belong to this dimension and contain the feature installed power which is used to perform the classification in the installed power taxonomy.

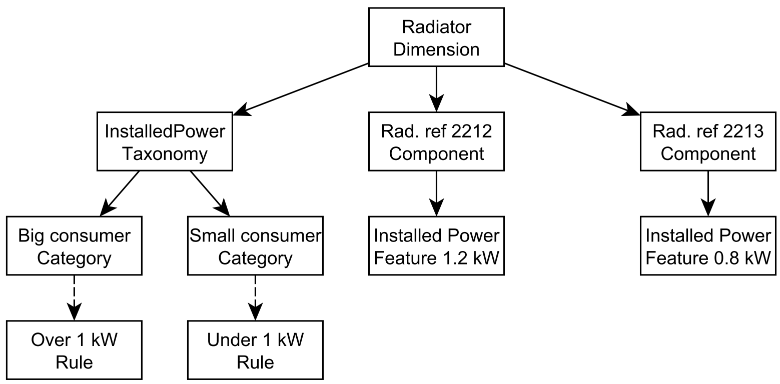

Fig. 12. Radiator dimension. This dimension is intended to store radiator components and classify them according to a criteria based on the installed power. Radiators are thus categorised into small or big consumers according to this feature

4.1 N-Level indicators and Data mining

So far, some mechanisms which allow us to extract information based on the measures have been presented: indicators. These indicators are regarded as first level indicators since they are just based on measures. For this reason, it is possible to define, from first level indicators, a second level of indicators, which are computations carried out based on previous level indicators. This idea can be extended to the concept of N-Level indicators. Introducing this concept, data mining[6] procedures can be used in order to find out patterns.

An example of this is presented in figure 13. In this case, a data miner has been designed in order to identify consumption habits which concern radiators. Using the thermostat indicator, which calculates the gradient based on the thermostat level measures, this data miner is able to identify habits throughout time. Therefore, common patterns of radiator usage could be identified and used to feed the Smart Grid policy design.

Moving onto low level details, this miner queries, for each radiator, its thermostat indicator throughout time. Based on this indicator, it uses techniques to extract habit patterns. These habits can be used by the policy in order to exert a more personalised control over the demand which enhances customer quality of service.

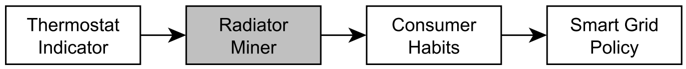

Fig. 13. Radiator data miner which intends to identify consumer habits

The list below presents two other cases where miners can be used in order to improve the design of a smart grid policy:

- Based on the technology feature of TV components, a miner can calculate the average time to amortise a TV based on a low-consumption technology by comparing them to the consumption of other technological kinds. According to these results, a smart grid policy could subsidise the purchase of TVs with a lower consumption. This kind of policy applies to other device kinds such as fridges or washing machines, among others.
- In the household cube, a miner can correlate the temperature and energy consumption of a household with its isolation features, supposing the information is available. Based on this correlation, the improvement of household isolation could be proposed.

4.2 Information visualisation

Information visualisation is the use of visual representations of data which step up the human cognition[10]. This is an important stage when analysing data since a proper visualisation may reveal information that could not be possible to extract using other visual representations. Figure 14 presents different visual representations which are discussed in the paragraph below.

Time (h)	Temp. (ºC)	Power(kW)
0	14	0,7
4	13	0,5
8	16	0,9
12	19	1,2
16	18	0,9
20	17	1,5
24	15	0,9

a)

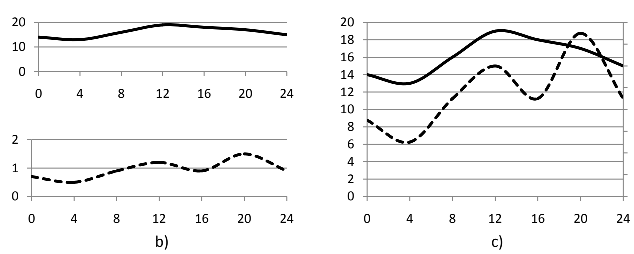

Fig. 14. Visual representations of both temperature (temp) and power

Part a in figure 14 presents the data in a table format. From there, it is cognitively difficult to extract temperature or power trends throughout the day, especially when there are many rows. In order to deal with this issue, both series can be represented in separated charts to enable the extraction of trends (part b). These trends can be extracted at each series but relational effects among them are neglected. However, part c represents both individual trends and relational effects among them. It is possible to extrapolate the relation among high temperatures and high consumption at noon.

The example presented in the section is a simple case which intends to provide insights of what is known as information visualisation. In this case, the representation required to show the information correctly is too evident. However, there are cases in which finding out the proper way to represent the information requires a deeper study.

4.3 Decision making

Using this approach to perform the simulation analysis enhances the capability of making decisions. The way in which the data is structured facilitates the interaction. From now on, queries can be as complex as needed in order to find out interesting conclusions which feed the decision making at the Smart Grid policy design. This structure is to be consumed using information visualisation patterns which could reveal interesting information that cannot be detected simply by analysing numbers.

In the previous example, important information can be extracted to be used in the Smart Grid policy design. The list below summarises some of the most relevant information:

- **Differences among districts:** Using the household dimension, it is possible to find differences among the districts located in the same city or, even, among cities. These differences can be noted in the way in which power is consumed, the devices are used or the temperature in the households. All of this could help in the design of a policy which provides enough flexibility in order to deal with these differences without losing efficiency when applied.
- **TV case:** it is possible to compare the differences in the consumption related to the TV technology. However, the TV dimension can be designed to take into account other aspects such as labelling and size. For instance, the labelling taxonomy could give information about whether it is worth promoting a Smart Grid policy which would subsidise the purchase of new high efficiency TVs.
- **Radiators:** using N-Level indicators allows us to identify consumption patterns that can be used to design more efficient Smart Grid policies which take into account customer usage. In other words, those patterns may be identified in order to build an intelligent control. On the one hand, this control could take note of the customer timetable in order to look after the quality of service. On the other hand, this control could take into account the grid state in order to reduce or increase consumption dynamically.

5 Conclusions and outlook

Transitioning from classical power grids to Smart Grids conveys a huge set of decisions to make. Among others, some of the most important are related to the management of demand. Therefore, an important analysis on the demand in a disaggregated manner is needed. This disaggregation involves the understanding of the power grid as a complex system.

Since consumption management policies in the context of the Smart Grids are not possible to experiment directly on the infrastructure, it is necessary to simulate them in order to make decisions. The complexity of the power grid system when it is disaggregated is so high that the results that the simulations return are huge. At this level, we have found out that the way in which those results are handled is crucial for making decisions.

Using an OLAP approach has been really helpful as much important information hidden among the data results was discovered. Information visualisation studies have definitely been useful so as to detect relational effects among variables. These effects can be further studied so that modifications on the Smart Grid policy take them into account. Another important strategy to extract information which is hidden or not evident is data mining. Thanks to data mining, consumption patterns can be identified and used to modify the Smart Grid policy in order to take care of the quality of service.

OLAP solutions can be used in many other environments since they are meant to facilitate decision supporting at management positions. We have identified heterogeneous environments such as metrics to program code, product selling and public information systems. In other simulation environments, this method of data analysis can be really interesting, e.g. a set of simulations which run different configurations using the same scenario. Indeed, using an OLAP solution would make it possible to compare all of these configurations with each other.

6 Acknowledgment

This work has been partially supported by Agencia Canaria de Investigación, Innovación y Sociedad de la Información of Canary Islands Autonomic Government thanks to the PhD grant funding assigned to José Évora with reference TESIS20100095 and also the project "Framework para la Simulación de la Gestión de Mercado y Técnica de Redes Eléctricas Insulares basado en Agentes Inteligentes. Caso de la Red Eléctrica de Gran Canaria", with reference SolSub200801000137.

References

[1] Chaudhuri, S., Dayal, U.: An Overview of Data Warehousing and OLAP Technology. ACM Sigmod Record 26(1) (1997)

[2] Gabaldon, A., Molina, A., Roldan, C., Fuentes, J., Gomez, E., Ramirez-Rosado, I., Lara, P., Dominguez, J., Garcia-Garrido, E., Tarancon, E.: Assessment and simulation of demand-side management potential in urban power distribution networks. In: Power Tech Conference Proceedings, 2003 IEEE Bologna. vol. 4. IEEE (2003), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1304784

[3] Luhn, H.P.: A business intelligence system. IBM Journal of Research and Development 2(4), 314–319 (Oct)

[4] Pedersen, T.B., Lehner, W., Hackenbroich, G.: Report on the First International Workshop on Energy Management Data. Sigmod Record 42(1) (2013)

[5] Peterson, E.T.: The big book of Key Performance Indicator (2006), http://www.webanalyticsdemystified.com/

[6] Rajamaran, A., Keskovec, J., Ullman, J.D.: Mining of Massive Datasets. Cambridge University Press (2011)

[7] Rosling, H., Rosling, O., Rosling, A.: Gapminder, http://www.gapminder.org/

[8] Saiku: Saiku, http://analytical-labs.com/

[9] The university of Waikato: Weka, http://www.cs.waikato.ac.nz/ml/weka/

[10] Ware, C.: Information Visualization: Perception for Design. Morgan Kaufmann Publishers Inc (2012)

[11] Wooldridge, M.: An Introduction to MultiAgent Systems - Second Edition. John Wiley Sons (2009)