Evaluating Demand Response Programs: Getting the Key Performance Indicators Right

Marilena Minou[1], George Thanos[1], Matteo Vasirani[2], Tanuja Ganu[3], Mohit Jain[3], Arne Gylling[4]
[1]Athens University of Economics and Business, Greece, [2]EPFL, Switzerland,
[3]IBM Research India, [4] LTU, Sweden

ABSTRACT
Demand Response (DR) has recently garnered great attention, with many DR programs being deployed and evaluated worldwide. They are hailed as a significant benefit enabled by the Smart Grid and an efficient method to engage consumers in managing their energy usage and reduce environmental impact and costs. But while the opportunities are great, challenges still remain to exploit the untapped potential of DR. Due to the many diverse technological and social contexts where is applied, establishing a common framework for evaluating DR programs is a rather complex but essential task in order to design more efficient and easily adopted, by utilities and users, DR programs. In this paper, we apply in practice some of, already defined in literature, Key Performance Indicators, aiming to evaluate different DR programs and we assess their applicability. In that context, we present and discuss initial results from two indicative trial sites (residential and commercial) and provide suggestions for future DR designers. Finally, we introduce the DR dashboard, a way to get an overview of a DR system and visualize the indices calculated in each trial site.

Keywords
Demand Response, Smart Grid, Key performance indicators, trial evaluation, baselines, user behaviour, socio-economic impact.

1. INTRODUCTION
The energy sector is experiencing major changes that are dictated by economical, technological and environmental factors. Due to its physical properties, electricity is not economically storable at the scale of large power systems. This means that the amount of power plant capacity available at any given moment must equal or exceed users' demand for it in real time. This is not a trivial task as both supply and demand levels could change rapidly and unexpectedly. These mismatches result in the need of reducing load using such methods as load shedding and curtailment to keep load generation and demand in constant balance.

There are various, often complementary, solutions to achieve this goal but, a promising method that also leverages the capabilities of Smart Grid infrastructures is Demand Response (DR). DR programs aim to alleviate the peak demand problem and to provide higher system reliability by altering user demand in response to power grid's supply and economic conditions. According to [1], DR alone could achieve 25-50% of the EU's 2020 targets concerning energy savings and CO2 emission reductions. Hence, DR systems are disclosed as one of the most important elements of the emerging Smart Grid networks [3][4].

DR can be applied to all categories of users (industrial, commercial, and residential) employing many different technologies and strategies to achieve shifts in demand, such as direct-load-control, incentives, prices, or a combination of these schemes. Designing a successful DR program is challenging as it depends on a myriad of factors, e.g. the type of user, generation, distribution, consumption, and demography. For example the age and lifestyle of the users might have significant influence in the design of a DR program [5].

Consequently, evaluating these programs is a rather perplexing and composite task, but it is vital, as it would also provide the means to improve them. In order to practically evaluate DR programs many factors should be addressed. For example, programs that offer incentives for participation must calculate the baseline load and measure the change from it that occurs during a DR event in order to calculate the total change in demand thus enclosing the evaluation of the baseline method into the DR program evaluation.

In this context, we exploit the Key Performance Indicators (KPIs) defined in our previous work [6], for the evaluation of the suitability and effectiveness of DR programs and practically apply and assess them in the context of two real-life trial sites (residential and commercial). Additionally, our proposed indices are easily traceable and measurable and can be utilised both from consumers and DR designers/Energy Managers (EMs), to estimate and evaluate the changes in energy consumption. For what is more, our conclusive subset of KPIs can facilitate DR program evaluation irrespectively of the magnitude of the application field and its specificities.

This work is part of the EU FP7 WATTALYST project [2], which aims to understand how consumers respond to DR signals by increasing/decreasing their demands and how their participation is influenced by external and internal factors. Another goal of the project is to understand effective methods of conveying the DR signals to the users, something affecting DR program's success.

2. MOTIVATION AND DEFINITIONS
There are various types of DR systems studied in literature [3]. Due to this variation, it becomes a challenge to establish a common framework in measuring and comparing the effectiveness of different DR programs. Evaluating their performance is an important and necessary step in the incorporation of demand resources into a well-functioning and thus sustainable electricity market. An objective evaluation would provide critical insights to the future development of DR capabilities and will help to guide technology investment priorities.

Therefore, it is important to identify DR metrics that can be used to assess the efficacy and economic performance of DR systems. Some of these might be easily quantifiable, e.g. the actual peak reduction, whereas others may not, such as users' acceptance and participation rate (dependent to the level of comfort). The choices are numerous and one of the challenges is to make the most appropriate selection and define the metrics and the success criteria of a DR program by taking into account the following potential impacts [6]: (i) Market price, (ii) different energy network architectures and infrastructures, (iii) the availability of historical or statistical data, (iv) customer impacts, (v) participation, (vi) public good impacts, (vii) the gains of involved stakeholders and the improvement of economic efficiency (social welfare).

Following the categorization in [6] and in the context of the trials deployed in [2], we have selected a set of KPIs to evaluate the applied DR programs in different contexts and requirements. The choice was based on the available information and the objectives for each trial case. The KPIs are summarized in Table 1. $KPI_{rx}, x = 1,\dots,5$ correspond to the KPIs for the residential trials, while $KPI_{cy}, y = 1,\dots,5$ refer to the KPIs for the commercial trials.

Table 1. KPIs utilised

KPIs	Description based on [6]
KPI_{r1}, KPI_{c1}	% change in total electricity consumption during the peak hours (DR event).
KPI_{r2}, KPI_{c2}	% change in total electricity consumption during the off-peak hours (DR event).
KPI_{r3}, KPI_{c3}	% difference of the real consumption from the baseline.
KPI_{r4}	% User acceptance and rate of participation.
KPI_{r5}	% User responsiveness.
KPI_{c4}	Absolute discomfort impact (1 to 5 scale).
KPI_{c5}	% discomfort level against total energy reduction constraint.

In this paper, we present results based on the aforementioned KPIs and discuss associated issues and challenges. For each trial site, we first present the problem it encounters in terms of peak energy consumption and the DR program applied in each case. Thereafter, we discuss and present briefly the KPIs used to gauge the above DR programs. Each section concludes with some initial results and evaluation from the first trial period, regarding the precision and reliability, as well as the actions that could be taken to improve the results and associated KPIs. Lastly, we introduce and present the DR dashboard (DRD), an innovative tool, which is used for the visualization of the defined KPIs.

3. BASELINE ESTIMATION

Regardless of the type of DR program employed, all require analysis to estimate the demand reduction. The estimate is the difference between what the user actually consumed and what that user would have consumed had the program not been enacted. What the user would have used is referred to as the DR baseline (or simply baseline) and is key to effective measurement and verification.

The baseline is one of the ways to determine the DR performance of a user, for example, by comparing the baseline with the metered consumption during the DR event to quantify the demand reduction provided by a user or a set of users [8]. Several methods have been used in various DR programs to compute baselines, such as HighXofY, MidXofY, exponential moving average, and linear regression [9][10]. According to several practical experiences, the DR baseline should be simple enough for all stakeholders to understand, calculate, and implement, including end-users.

3.1.1 HighXofY baseline

The baseline is the average load of the *X* highest consumption days within those *Y* days. More formally, the baseline for timeslot *t* on day *d* is:

$$b(d,t) = \frac{1}{X} \sum_{d \in High(X,Y,d)} \lambda(d,t)$$

where *λ(d,t)* is the past consumption on day *d* at timeslot *t*, and *High(X,Y,d)* is the set of *X* highest consumption days within the most recent *Y* days before the DR event. Examples of *HighXofY* baseline are PJM Economic (High4of5), NYISO (High5of10) and CAISO (High10of10) [9].

3.1.2 MidXofY baseline

The baseline is the average load of the *X* days out of a set of *Y* most recent non-DR days, from which the highest and lowest consumption days are removed. More formally:

$$b(d,t) = \frac{1}{X} \sum_{d \in Mid(X,Y,d)} \lambda(d,t)$$

where *Mid(X,Y,d)* is the set of *X* middle consumption days within the most recent *Y* days before the DR event. Example of *MidXofY* baseline is the Mid4of6 [9].

3.1.3 Exponential moving average

The exponential moving average baseline is a weighted average of the historical load, where the weight decreases exponentially with time. This baseline is computed using all the historical consumption data. Let $D = \{d_1, d_2, \dots d_k\}$ be the set of all measured days preceding the target day *d* having the same day type as *d*. In addition, let $1 \le \tau \le k$ be a constant. We define $s(d_\tau, t)$ as the *initial* average load at time *t*:

$$s(d_\tau, t) = \frac{1}{\tau} \sum_{j=1}^{\tau} \lambda(d_j, t)$$

The exponential moving average for $\tau < j \le k$ is:

$$s(d_j, t) = \alpha \cdot s(d_{j-1}, t) + (1 - \alpha) \cdot \lambda(d_j, t)$$

where $\alpha \in [0,1]$ is the constant smoothing factor for weighting decrease. Then, we define the exponential moving average baseline on day *d* at timeslot *t* as:

$$b(d,t) = s(d_k, t)$$

An example of this baseline is the ISONE baseline, for which $\tau = 5$ and $\alpha = 0.9$.

3.1.4 Regression

The baseline of day d is computed by fitting a series of linear regressions (on for each time slot *t*) to the historical consumption data. The baseline on day *d* at timeslot *t* is computed as:

$$b(d,t) = (\boldsymbol{\theta}_t)^{\mathrm{T}} \boldsymbol{x}_t + \varepsilon_t$$

where $\boldsymbol{x}_t$ is the feature vector, $(\boldsymbol{\theta}_t)^{\mathrm{T}}$ is the (transposed) vector of regression coefficient, and ε_t is the error term. The feature vector is a vector of explanatory variables such as historical consumption, temperature, or sunrise/sunset time. The regression coefficients and the error term can be estimated with standard methods such as least squares or ridge regression.

Figure 1 summarises the performance of the different baseline methods applied to the total electricity consumption data of the SAMPOL headquarters. We used the data from June 2013 until March 2014. The Mean Absolute Error (MAE) is defined as:

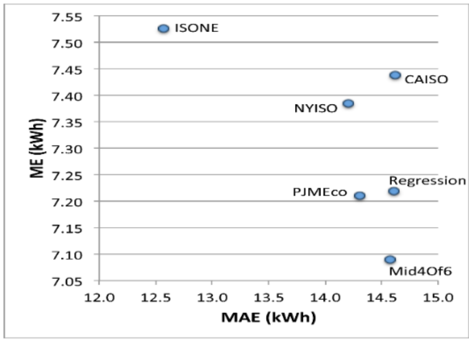

Figure 1. The performance of the different baseline methods

$$\mathbf{MAE} = \frac{1}{24N}\sum_{d=1}^{N}\sum_{t=1}^{24}|b(d,t)-\lambda(d,t)|$$

where N is the number of days for which the baseline has been computed, $b(d,t)$ is the baseline on day d at time t, and $\lambda(d,t)$ is the real consumption for the same day at the same time. The Mean Error (ME) is defined as:

$$\mathbf{ME} = \frac{1}{24N}\sum_{d=1}^{N}\sum_{t=1}^{24}b(d,t)-\lambda(d,t)$$

By keeping the sign, the Mean Error is a metric that shows whether the baseline estimation method systematically overestimates or underestimates the real consumption, representing therefore the *bias* of the estimation. Based on the results, all the methods tend to overestimate the real consumption (i.e. positive ME). Among them, ISONE turns out to be the most accurate method (i.e. lowest MAE), albeit the most positively biased, while Mid4Of6 is the least biased.

4. EVALUATING DR PROGRAMS

In this section we introduce the two trial sites, in which different DR programs were implemented. For conciseness, we have chosen to present a small number of applied DR programs that are, nonetheless, representative of the measurements we obtained from a larger set of tests.

4.1 Residential Trial Site: Luleå, Sweden

4.1.1 Introduction

The field trials in Luleå focus on energy consumption in private homes. The fine grained measurements include consumption of district heating (house level) and electrical (smart meter and selected appliances level) energy. District Heating (DH) is the major energy source used for heating in Luleå and is the lowest priced DH in Sweden, due to the availability of excess gas from the local steel work plant. In this case, the KPIs that are considered for evaluation of the applied DR program are presented in Table 2. The first two refer to the peak reduction quantification category, the next one relates to the demand variation analysis and demand reshaping category and the last two belong to the economic-related KPIs category of [6].

4.1.2 The problem

Luleå Energi's DH is probably the most developed in Sweden. Within the urban area, all areas with large buildings are connected, while there are approximately 8600 private homes using DH in Luleå. Luleå Energi has four large boiler plants in reserve to replace power plant in the event of downtime, and serves as peak load plants in really cold weather. The main boiler uses the steelworks excess gas and is the best choice from both an economic and environmental perspective. But when this production is not enough extra heating generation capacity must be provided. As a result, the generation cost increases, as well as the environmental impact. The problem is mainly tracked during certain days of the week, leading mostly to two peak demand periods per week. At the peak demand time the energy supplier activates the backup system by using alternative energy production to produce electricity, in order to meet the demand. *Our objective is to reduce or shift this demand at and from the peak periods by creating and providing the right incentives to users by the means of DR mechanisms.*

4.1.3 Planning the DR events

As mentioned above, the aim is to reduce the consumption during peak demand periods. This can be achieved by exploiting the DR mechanisms, i.e. at specific periods when the demand is high, the users that participate receive event signals, i.e. text messages, notifying them about the need to alter their consumption. These messages are known as DR messages/signals and may include information about the users' real time consumption, the time of the DR event, the providers' incentives etc. However, the design of effective DR events is not apparent and we need to take into account specific variables according to the research interest. These variables are divided into categories as shown in Table 2.

The DR events are separated into different campaigns depending on which of the above variables we want to address. For example, the campaign "Context" aims to investigate how context affects the users. For this campaign we will change the Context variable [Season, Special days etc.], while we keep all other variables (User response, Communication, Incentive) fixed.

For the purposes of the project a residential Home Energy Efficiency Persuasive Interface (EEPI) was developed, which is a collection of intuitive user interfaces to be used by residential users. Users are provided with a tablet and they can access the project's residential EEPI link by using a web browser on it. Each user is provided with credentials to login into the system and view the DR messages along with real-time energy consumption information. For evaluation we also use other means to communicate to the users such as email, SMS and face-to-face meetings. This enables us to compare, which is the preferred and most effective method for DR communication.

Table 2. Campaign variables

User response	**Communication**	**Incentive**	**Context**
Levels of shifting/reduction	Phone/Display	Price level	Season
Tolerance	Environmental/Cost	Recognition	Special days
Loads	Notification time	Open reduction	Economic status
Fatigue			Awareness
			Family size

4.1.4 DR events evaluation and suggestions

In this section, we consider two tests that took place on Nov 26th 2013 (Tuesday) and Dec 1st 2013 (Sunday), in order to present indicative numerous results for the KPI_{r1} to KPI_{r4}, while results for KPI_{r5} and KPI_{r6} are presented within a campaign too.

The tests belong to the campaign "Context" as defined earlier and are representative of two different contexts: weekdays and weekends. The DR signal was sent to the users via an SMS and the message was to reduce the hot water consumption between 8:00 and 8:59. The baseline methods introduced in Section 3 have different characteristics in terms of accuracy and bias, which affect the user decision-making and participation in a DR event, as well as the demand reduction obtained by the DR provider. It has been shown that positively biased baselines tend to foster user participation and, as a side effect, benefit the DR provider. For this reason, in our case we used the NYISO baseline, which in general is a good compromise between bias and accuracy. For each test, we computed the aggregated baseline of all the users that participated to the test before (from 6 to 7), during (at 8) and after (from 9 to 10) the DR event, and we compared it with the metered consumption.

Figure 2 shows the hourly consumption and baseline of the users that participated in the test performed on Nov 26th 2013. It is possible to see that the highest consumption is expected (and metered) from 6 to 7, when users typically prepare to leave to reach their workplaces. Furthermore, there is no significant consumption change before and after the DR event. In fact, we estimated for KPI_{r2} (before and after the DR event) a small reduction of 3.8% and 0.22 % respectively. On the other hand, during the DR event, KPI_{r1} was decreased by 40.7%, indicating a strong effort from the users that accepted the DR signal, by altering their consumption.

Quite interestingly, different user behavior during the weekend is observed (Figure 3). For this test, the baseline consumption was expected to be higher at 9 and 10, given that on weekends people usually spend more time at home. The demand reduction (KPI_{r3}) before the DR event is 11.1%, while during the DR event is 33.3%. However, after the DR event a consumption increase of 47.3% has been estimated. A possible explanation is that for this test the users decided to shift the water usage after the DR event, given the absence of the time constraints that typically occurs on weekdays.

Both tests had a considerable participation rate (KPI_{r4}) in the DR event, 60% and 70%, respectively altered their consumption, while the users' responsiveness (KPI_{r5}) was quite high, i.e. 70% and 90% respectively answered by "ACCEPT" or "REJECT". Figure 4 shows how KPI_{r4} and KPI_{r5} evolve within the campaign. Only KPI_{r5}= 42% of the users responded to the DR message, while KPI_{r4}=24% accepted to enroll the action proposed.

During the campaign, the users were notified for the existence of a DR message through the display, except of the two tests described above, in which users were alerted via SMS only. The results show (Figure 4) that via SMS more users were triggered to engage themselves in the DR event, when comparing with the rest of the days. This means that an improvement of the display is essential, so that it becomes more attractive and easy to use. According to users' feedback, the display should be part of their normal behaviour if this is the way for the DR designer/energy manager to interact with them.

4.2 Commercial Trial Site: SAMPOL, Palma de Mallorca

4.2.1 Introduction

SAMPOL's headquarters are located in the city of Palma, capital of the Balearic Islands (Spain). It is a large building including different business units with 150 employees. Most electricity consumption is due to lighting, air conditioning/heating (HVAC), computers and other appliances. During the trials, consumption measurements are carried out in 3 different areas of the building. In each area, electricity consumption is monitored separately for lighting, HVAC and appliances.

The KPIs exploited in this case are presented in Table 2. The first two refer to the peak reduction quantification category, the next two relate to the demand variation analysis and demand reshaping category and the last belong to the economic related KPIs category.

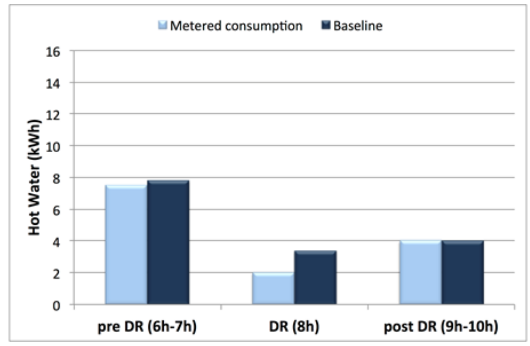

Figure 2. Luleå trial site, Nov 26th 2013 (Tuesday)

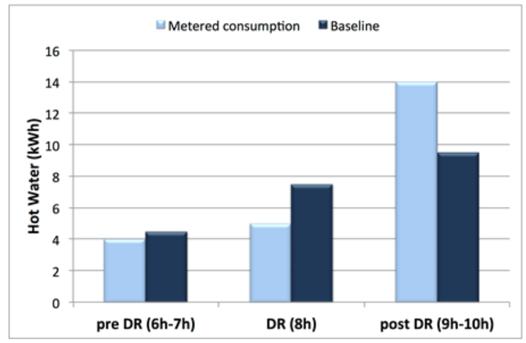

Figure 3. Luleå trial site, Dec 1st 2013 (Sunday)

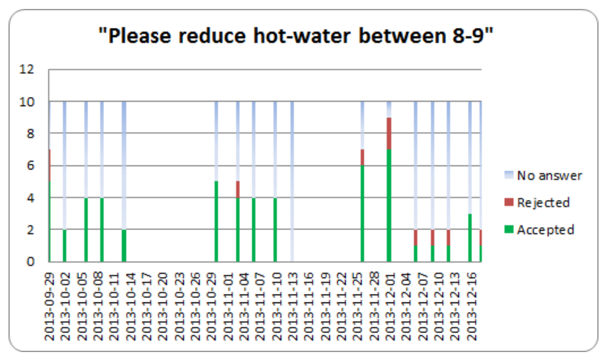

Figure 4. KPI_{r4} and KPI_{r5} for campaign "Context"

4.2.2 The problem

The problem in the Mallorca site resides on peak load reduction. The trials aim to evaluate DR programs, designed for this purpose. Furthermore, we aim to investigate the way users choose to attain the required energy reduction, e.g. what appliances they choose to modify in relation with external factors and also in regard to the change of their comfort zone. *In the Mallorca case the DR program designer aims at re-shaping the demand curve in two ways, by reducing demand and/or shifting it from the peak hours to the non-peak hours.*

Similarly to the Lulea trial, the DR events are separated into different campaigns depending on the objective we are targeting and composed from several use cases. For example, the campaign "Act" includes all the events, in which the users are not informed about their existence. A campaign aims to investigate the consequences of the event actions in the comfort perception of the users. The selected appliances for DR are the HVAC, lighting and total energy of the building, with HVAC being the most important source of energy consumption in the building. HVAC and lighting are monitored separately and are also the most susceptible to be shifted or altered. Appliances such as computers or other working equipment are unlikely to be affected, something that differentiates this setting compared to the Lulea one.

4.2.3 DR events evaluations and suggestions

In this section, in order to illustrate the use of the KPIs, we present the results of one of the tests that have been performed at the SAMPOL headquarters. The test took place on Oct 15th 2013 (a Tuesday). The recipient of the DR message was the EM, who was asked to reduce electricity usage by 12kWh between 10h and 14h. In order to fulfil the objective of this test, the EM decided to disconnect the HVAC system of 4 departments. Figure 5 shows the energy consumption of the HVAC system in one of the affected department (Energy Department). One can see, during the DR event, how KPI_{c1} significantly decreases by 80% of the expected baseline consumption, i.e. 19.1kWh. A significant rebound effect caused by the action taken by the EM to respond to the DR message is also visible. After the DR event (from 14h to 20h), the consumption (KPI_{c2}) increased by nearly 80%, i.e. 12.3kWh (almost the same amount saved during the DR event).

The actions taken by the EM also impacted headquarters total electricity consumption (see Figure 6) as there was a reduction of total electricity usage of 38.6 kWh, which represents KPI_{c3}=13.6% of the baseline consumption. Although the EM accomplished (and exceeded) the DR objective, it is worth noticing an increase of the consumption (KPI_{c2}) of 41.6 kWh after the DR event, which is more than the amount saved during the DR event.

Considering the users' side, which in this case are the employees of the company, after the DR event they were asked to answer a questionnaire about their comfort perception on the event. Apparently, the percentage of comfort varies with the different time zones of evaluation, i.e. between 8h to 11h, users were experiencing 81.82% of comfort with the inside temperature, while during the DR event (11h to 14h) their comfort rate was decreased to 68.18%. Thus, KPI_{c4} reaches the value of 3.4, which shows that the discomfort level was acceptable. The values are scaled down in the range of 5, where 1 shows that the room temperature is very high and 5 shows that the room temperature is very good. Regarding that in this case the EM was asked to accomplish a specific target reduction, KPI_{c5} is estimated to -35.6% and shows that the discomfort caused by the DR event was lower than the energy saved. This means that the users could afford working without the operation of the HVAC during those specific hours.

However, the increase in the energy consumption after the DR event that is accompanied with an increase of the comfort rate of the users to 91.67%, indicates that extreme manual actions such as shutting down loads are very prone to rebound effects. In fact, these actions are reactive short-term responses and they do not take into account the medium term effects. For this reason, decision support systems and visual dashboards that help the recipients of a DR messages to select the best action to respond to the DR message are likely to be very important for a successful DR program.

5. KPIs VISUALIZATION

Based on the inputs obtained from DR Designers and EMs, we designed DR Dashboard, an interactive web application that DR designers and EMs can use to get an overview of a DR system deployed for residential and/or commercial consumers. Figure 7 shows a screenshot of the DR message KPI page, which can be accessed by clicking the 'Analyze' icon on the top left corner of the screen. This page is divided into three parts: A) *Select Search Parameter*: providing a list of search parameters for DR designers to choose from, for the search query (left half of the page, Figure 7). B) *Time, Event and Weather Selector*: provides a way to select time, and show/hide weather and event charts (top right of the page, Figure 7). C) *Search Result*: provides the analysis output using charts and tabular data (left half of the page, Figure 7).

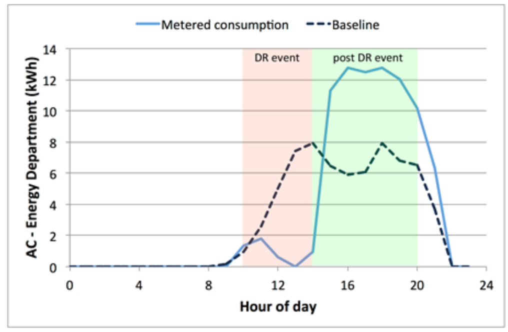

Figure 5. Sampol trial site, Oct 10th 2013, AC energy consumption of the Energy Department during the test

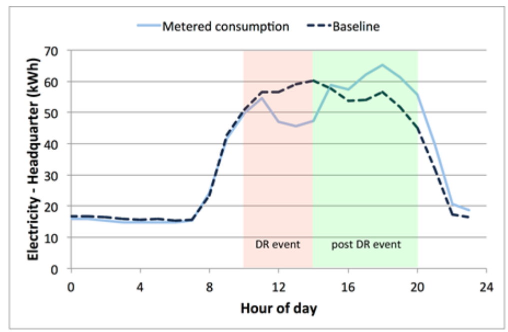

Figure 6. Sampol trial site, Oct 10th 2013, Total energy consumption of the Energy Department during the test

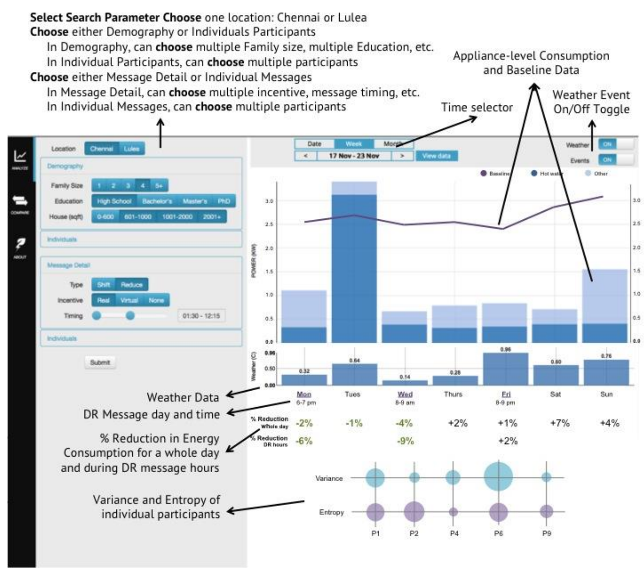

Figure 7. DR Dashboard interface: DR message KPI page

5.1 Select Search Parameters

DR designer can select from a list of Search parameters to get the desired analysis results. The designer can choose trial site *location*, choose either *Demography* or *Individuals Participants* (*e.g.*, in Figure 7, *Demography* is selected), and can choose either *Message Detail* or *Individual Messages* (e.g. in Figure 7, *Message Detail*). Further, in *Demography*, the designer can choose multiple Family size (from 1, 2, 3, 4, and 5+), multiple Education (from High School, Bachelor's, Master's, and PhD), multiple House size (from 0-600 sqft, 601-1000 sqft, 1001-2000 sqft, and 2001+ sqft). In *Individual Participants*, designer can choose multiple participants/users. Similarly in *Individual Messages* can choose multiple messages, while in Message Detail, can choose multiple Types of message (from shift-related and reduce-related message), multiple Incentive (from real, virtual, and no incentive), and Message Timing (a start time and end time). After selecting the required Search parameter, DR designer can submit the search query by clicking the 'Submit' button.

5.2 Time, Event and Weather Selector

Furthermore, the DR Designer can select a time period for which the query is made, using the *Time Selector* at the top of the page. Data can be obtained over different time periods, such as months (on a day-basis), weeks (on a day-basis), and days (on an hour-basis). Date, Week, and Month can be selected by clicking on the top three buttons on the page (in Figure 7, e.g. 'week' is currently selected), and the left and right arrow buttons need to be used to navigate and select a particular date, start-date and end-date for a week, or name of a month, as required. After selecting the time, clicking on the 'View data' button updates the search result.

5.3 Search Result

Figure 7 (right half) shows the result section, which can be divided in three parts. (i) *Power Consumption and Baseline Data*: Stacked bar graph with overlaying baseline as line graph, shows appliance-level power consumption and baseline data for the selected time period, as per the search query. In addition, individual appliances and baseline can be selected and deselected using the *Appliance Selector* at the top of the graph. (ii) *Weather and Event Data*: A bar graph showing the weather data of the selected location, during the selected time period (Figure 7). This graph is only visible when the *Weather toggle button* is in ON state. Similarly a graph showing the event data of the selected location, during the selected time period (hidden in Figure 7, as the Event Toggle button is OFF). (iii) *KPI Data*: A tabular layout showing KPI related data such as percentage reduction in energy consumption compared to baseline at a day level, and also during DR message hours (Figure 7). Moreover the legend highlights the days with DR messages along with the time of the DR event. At an individual participant level, variance and entropy is shown using a bubble chart.

Currently, DR Dashboard is being actively used by the researchers to analyse the data. In future, we would evaluate the usability and effectiveness of this dashboard with DR Designers and EMs.

6. CONCLUSIONS AND FURTHER WORK

KPIs are essential in revealing the efficacy of a DR system in terms of flattening the peak load, economic sustainability, altering consumer behaviour, etc. In this work, we presented a practical evaluation of different types of KPIs, based on [6], and in the context of two trial sites with different requirements and specificities. The data collected from appliance level sensors and smart meters along with users' feedback were used to calculate the KPIs and assess the DR programs.

The proposed KPIs are justifiably well-defined and adaptable to different environments and contexts. The results indicate that the KPIs for peak reduction quantification, together with the appropriate baseline methods, are the easiest way of detecting changes in energy consumption. Percentages are more appealing to the perception of users and these metrics can be easily captured in graphics, as shown in Figure 7. They can be used as a stand-alone core of metrics or even in combination with other KPIs. In the latter context, the KPIs that are relevant to users' acceptance and responsiveness, as well as to the rate of their participation are very important, because they determine not only whether a DR program is successful or not, but also what actions led to this success. Additionally, we paid particular attention to the economic-related KPIs, for which the methods of calculating users' comfort level are of great importance. We found that using a rating scale is a good approach, as it is a comprehensible way of the users to define and express their comfort or discomfort. A combination of these metrics is essential, in order to guarantee the success of a DR program and to identify the actions for further improvement.

Our obtained results attest to the validity of the above conclusions. In the residential trials, initial tests resulted in considerable reductions in energy consumption during DR events as well as a significant participation rate. In addition, the users were more engaged to responding to the DR message after they were alerted by an SMS, which indicates the need of improvement of the communication display to be more user friendly and interactive. As for the commercial trial site, the results indicated that by taking abrupt actions, the rebound effect is strongly reinforced. Visual dashboards such as the one presented in this paper constitute valuable tools for a DR designer in diminishing these side effects.

Finally, several new campaigns are planned to be tested in both trial sites in the future. The results will be used to evaluate additional parameters that influence user participation as well as the impact of the KPIs in order to gain better insights about the design, sustainability and evolution of future DR systems.

7. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Program (FP7/2007-2013) under grant agreement n° 288322.27 (WATTALYST). Finally, we would like to thank all the partners of WATTALYST (Modelling and Analysing Demand Response Systems) project [2] who have contributed with their analysis and discussions in the production of these results.

8. REFERENCES

[1] Demand Response: a decisive breakthrough for Europe, Capgemini, VaasaETT and Enerdata, June 2008

[2] WATTALYST project http://www.wattalyst.org/

[3] Benefits of demand response in electricity markets and recommendations for achieving them. Technical report, U.S. Dept. of Energy, 2006.

[4] van Deventer, J., Gustafsson, J., and Delsing, J., Controlling district heating load through prices, Systems Conference (SysCon), 2011 IEEE International, 461-465.

[5] D. Kirschen. Demand-side view of electricity market. IEEE Transactions on Power Systems, 18(2):520–527, May 2003.

[6] Thanos, G., Minou, M., Ganu, T., Arya, V., Chakraborty, D., van Deventer, J., & Stamoulis, G. D.. Evaluating demand response programs by means of key performance indicators. In Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on (pp. 1-6). IEEE.

[7] IEA, 2003. The Power to Choose—DR in Liberalised Electricity Markets. International Energy Agency, 2003.

[8] S. Schnittger and S. Beare, Economic implications of the proposed demand response mechanism. Australian Energy Market Commission, Tech. Rep., 2012

[9] PJM Empirical Analysis of Demand Response Baseline Methods. White Paper, KEMA, Inc., Apr. 2011.

[10] The Demand Response Baseline. White Paper, EnerNOC, 201