Analyzing statistically the energy consumption and production patterns of European REScoop members
Results from the H2020 project REScoop Plus

Akadiadis, Charilaos; Savvakis, Nikolaos; Mamakos, Michail; Hoppe, Thomas; Coenen, Frans; Chalkiadakis, Georgios; Tsoutsos, Theocharis

Publication date
2017

Document Version
Final published version

Published in
Proceedings of the 9th International Exergy, Energy and Environment Symposium IEEES 2017

Analyzing statistically the energy consumption and production patterns of European REScoop members: Results from the H2020 project REScoop Plus

[1]Charilaos Akasiadis, [2]Nikolaos Savvakis, [1]Michail Mamakos,
[3]Thomas Hoppe, [4]Frans Coenen, [1]Georgios Chalkiadakis, [2*]Theocharis Tsoutsos

[1]Technical University of Crete, School of Electrical & Computer Engineering, Kounoupidiana Campus, Chania, 73100, Greece
[2]Technical University of Crete, School of Environmental Engineering, Kounoupidiana Campus, Chania, 73100, Greece
[3]Delft University of Technology, Dpt. of Multi-Actor Systems (MAS), Jaffalaan 5, Delft, 2628 BX, The Netherlands
[4]University of Twente, Dpt. of Gov. and Technol. for Sustainability (CSTM), P.O. Box 217, Enschede, 7500AE, The Netherlands

* E-mail: theocharis.tsoutsos@enveng.tuc.gr

Abstract

REScoops are cooperatives of renewable energy producers and/or consumers, which have begun to form in the emerging European Smart Grid. According to the Federation of REScoops in Europe (REScoop.eu), there are currently more than 2,397 REScoops, collectively having more than 650,000 members. As such, their emergence highlights the importance of producing and consuming green energy, and simultaneously puts forward principles such as energy democracy and self-consumption, assists the fight against energy poverty, and reduces Greenhouse Gas emissions. A core objective of the H2020 REScoop Plus project is to promote a better understanding and cultivate the behavioural change of the cooperatives' engagement. In order for this goal to be met, it was essential to conduct a careful logging of and statistical analysis over the energy data stored by the REScoops, as well as the current state of their engagement in energy efficiency. Such tasks are imperative in order to support the claim that REScoop engagement promotes energy sobriety; and associate consumption reduction with specific behaviours, ICT tools, and Energy Efficiency practices. In this paper, we present the logging and statistical analysis conducted in REScoop Plus, and their results. Specifically, we outline the methodology used for formulating a comprehensive format for collecting the datasets for statistical analysis, a process that involved a lengthy and fruitful interaction with REScoops and their data experts; the extracted requirements for that common format; and the detailed guidelines communicated to the REScoops in order to enable data acquisition and analysis. Finally, we present a summary of the collected data (relating to electricity and heating consumption, demographics, etc.); the detailed statistical analysis process used; and the results of this analysis.

Keywords: Renewable energy cooperatives, performance analysis, solar energy

I. Introduction

Recent research on sustainable energy and development planning indicates that a shift towards renewable energy resources and the adoption of energy conservation techniques are gaining ground towards tackling energy poverty and meeting large-scale energy efficiency (EE) (Energy Efficiency Watch, 2009). However, for EE to be attained, stakeholders, i.e. utility companies, network regulators, end-users, etc. need to get actively involved.

One of the most significant parts of the current energy market is the European renewable energy sources cooperatives (REScoops) (European Energy Mediators Group, 2015). The formation of REScoops in the European Union has begun, and they provide their members the opportunity to buy renewably generated electricity at fair prices, to democratically react with other members and co-decide the cooperative's future, and to be autonomous and independent with respect to energy. Given these features and benefits, REScoops organize events, such as meetings, conventions, etc., in order to raise their members' energy awareness. Thus, it is expected that when end-users join forces in an energy cooperative, they become more active regarding energy conservation and efficiency.

To this purpose, the REScoop Plus project (REScoop Plus, 2016) aims to gather available information and data from various European REScoops, and demonstrate that participation in such a cooperative raises energy awareness and contributes to accomplish the challenging goal of energy efficiency. Here we focus on describing the process by which this claim can be investigated from an informational (data-centered) point of view, while its behavioural aspects are going to be investigated in the near future.

The main objectives of this research were:

- To gather and statistically analyze measurements, such as historical energy consumption values of end-users, and other correlated variables, in order to prove existing reductions in consumption, and study other indicators regarding EE.
- To perform behavioural analysis of some of the REScoops members, so as to assess the impact of a number of specific EE interventions.
- To highlight the effectiveness of interventions that have been already implemented, by means of statistical analysis of the historical consumption data.

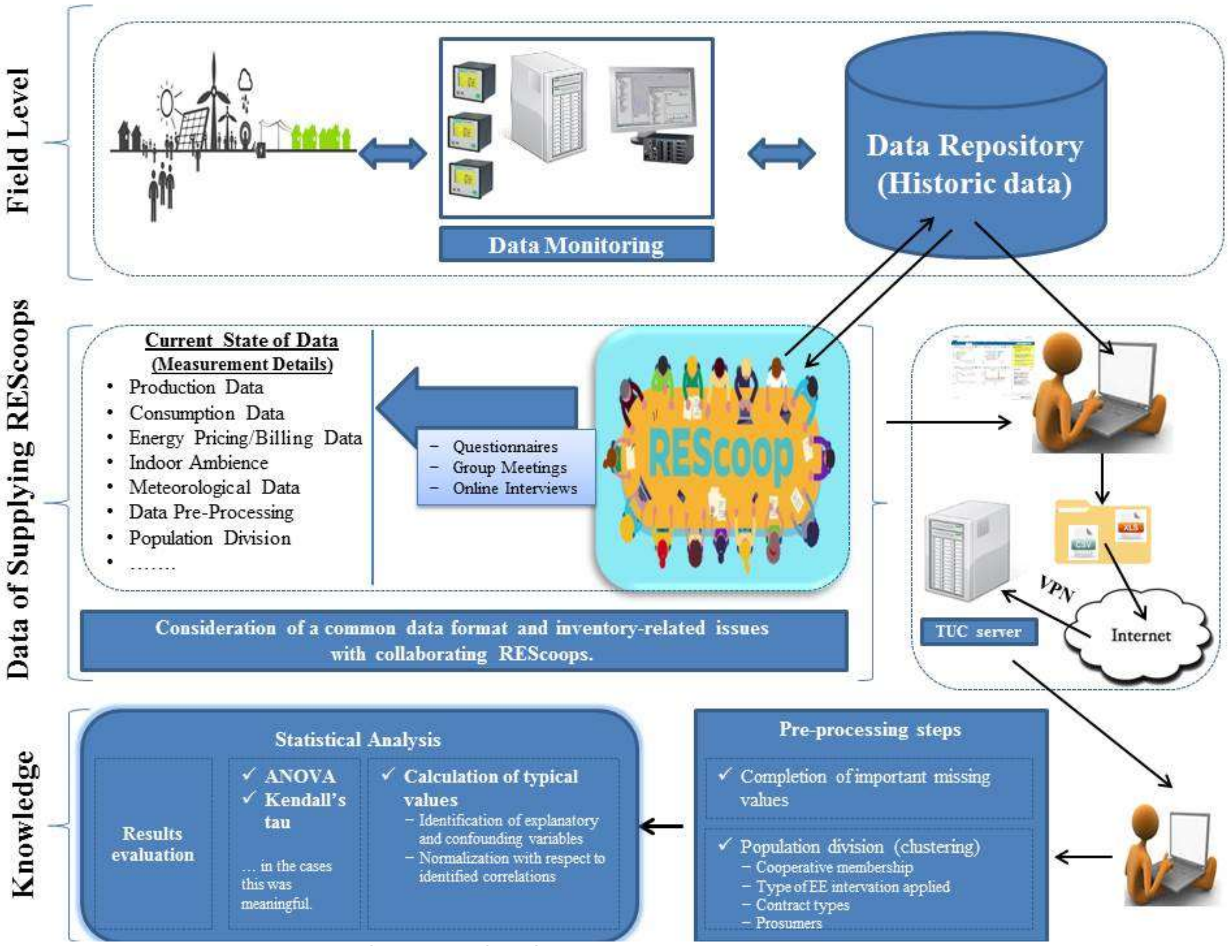

Fig. 1: Overview of REScoop Plus research methodology

- To associate energy consumption reduction with specific EE measures.

II. Methods and Materials

The design of a research methodology, in order to conduct a statistical analysis of energy-related data, is a complex process that requires significant technical expertise. There are many adjustments required in order to reach the optimum balance between reliability, accuracy and significance. In this section, the methodological approach and critical issues to be considered are outlined. In brief, the methodology of our research consists of the following steps, which are also depicted in Fig. 1:

- Identification of the current state of the data from the participating REScoops.
- Consideration of a common data format and inventory-related issues.
- Receiving the datasets.
  - Completion of the datasets missing values from additional third party sources.
    - This could involve using regression techniques.
- Analyze the development of the average yearly consumption for all cooperatives and separate all possible variables that might influence the decrease in consumption (climate, change in habits, etc.).
  - Population clustering
  - Identification of explanatory and confounding variables
- Validate the decrease in average consumption (if this is indeed the case). Also, potentially check whether there is a decrease in usage of conventionally produced energy (from non-renewable resources).
  - Perform hypothesis testing, following e.g., n-way ANOVA (DeGroot and Schervish, 2010);
    - Identification of degree of correlation among variables
  - Derivation of typical consumption curves for each cluster
    - Normalization with respect to identified correlations

II.1. Preparatory activities - Current state of the data from REScoops

The first step of the analysis process was to gather information regarding the cooperatives themselves, the knowledge that has been so far obtained by experience, and the techniques that have been used in the recent past. Taking into consideration that REScoops vary greatly with each other, with respect to commercial activity, internal structure, data storage methods, and legislative framework, the use of suitable information collection practices and the adoption of a common data format were deemed necessary.

Tab. 1: Overview of the data categories that each participating REScoop already stores.

	Production Data	Consumption Data	Pricing/Billing Data	Indoor Ambience	Meteorolo-gical Data	DSM-related Data	Other
Coopérnico	✓	✓	✓	-	✓	-	-
EBO Consult	✓	✓	✓	-	✓	✓	✓
Som Energia	✓	✓	✓	-	✓	✓ **	✓
SEV	✓	✓	✓ (Yearly)	-	✓	-	-
Ecopower	✓***	✓ ***	✓***	-	✓***	-	✓***
ENERCOOP	✓	✓	✓	-	✓	✓	✓
ènostra	-	✓	✓	-	✓	-	-

* Data come in monthly or higher than monthly granularity (monthly or more frequently), unless specified otherwise.
** Actually time-of-use plans data
** Some of Ecopower's members have subscribed to an additional, external monitoring service (www.energieID.be). For these members only is monthly data available; for the rest, stored data is yearly.

To this end, preparatory activities that include discussions in group meetings, online interviews, and questionnaire investigations to assess the size, state, contents of the various datasets etc., were performed. Specifically, suitably designed questionnaires were given to the REScoop experts (in two separate phases), and related feedbacks were further analyzed. The initial questionnaire contained 13 questions, and it was given to the experts in a paper form. In particular, its questions were:

1. What is the name of your cooperative?
2. How many are your members in number?
3. How many are your constituent coops/organizations in number?
4. How many are the actual people in your cooperative?
5. What is the estimated portion of which that could engage in demand-reduction actions?
6. Do you have energy production activities?
7. If yes, by what means?
8. Do you sell energy?
9. Do you provide district heating?
10. Do you apply Demand -Side Management (DSM) schemes?
11. Other ?
12. Are measurements available?
13. If yes, since when?

This questionnaire was utilised in order to get a first view of the current situation regarding each REScoop's size and activities.

Then, a detailed questionnaire (on-line) was given to REScoop experts, in order to clarify the current state of the data, as well as potential difficulties that could be faced; and to create a basis for a common data format. This questionnaire divided into four parts:

- General Information (6 questions)
- Measurements Details (13 questions)
- Demand-Side Management - related Information (7 questions)
- Final Questions (5 questions)

In the first part, they were asked for general REScoop information such as the name, the geographical location of involvement, the time since the data had been being stored, etc.; and the answers were provided in the form of text.

In the second part, specific details regarding the data were requested, that is which measurements are available relating to energy generation and consumption, pricing and billing, indoors measurements, climatic conditions, and the possibility of dividing the population into control and testing groups (REScoop Plus, 2016)). The experts were asked to pick the level of detail of each available measurement from a total of 10 different options, ranging from a second to a year. Moreover, text-box questions were considered to take additional information and clarifications of the responses.

In the third part, the main goal was to extract information related to additional demand reduction measures (REScoop Plus, 2016). Feedback was requested, in the form of lists with checkboxes, and additional text-fields in case further explanations were needed.

In the last part, additional information on available data for demographics, past participation in EE actions, and the existence of past surveys was requested.

The identification of the current state of the data, which is depicted in Table 1, also enabled us to conclude on a common data format that can be adopted by all partners in order to ease the data collection and analysis activities, and to render comparisons among the various REScoops information and practices realistic and meaningful (REScoop Plus, 2016).

II.2. Adoption of a Common Data Format

The identification of a common structure allowed the setting of a list of requirements to be followed, in order to establish a similar starting point for the analysis. To this end, REScoops were asked to incorporate all available information, despite the fact that specific parts of some datasets were missed out. This was crucial, since submitting all the requested measurements makes the statistical analysis more sound and realistic. Five different files were requested:

1. Electricity Consumption related data
2. Heating Consumption related data
3. Energy Production related data
4. Demographics related data
5. README file (.txt format) explaining the column titles, measurement units, the group codes of (1) and (2), and additional required information

It should be also noted that the requested measurement frequency was considered as that of monthly measurements, because (a) it is currently the most widely adopted, (b) can provide enough intuitions regarding energy consumption behaviour, and (c) is not enough to be considered as privacy violating.

II.3. Receiving the Dataset

One of the major challenges associated with receiving the datasets was that at least some of the partners were very reluctant to share their data, due to their confidentiality concerns. To address these concerns and to render data submissions secure and private, the TUC virtual private network (VPN) was utilized. By means of this computer networking solution, it was ensured for specified connections to use accounts and credentials, which were kept safe, and only communicated to REScoop experts via direct messages. In particular, the VPN allowed connectivity to protected servers on which submitted data were stored and kept during the project's time horizon.

II.4. Data Analysis

In our approach, we aim to utilize two well-known approaches, i.e., statistical hypothesis testing, and analysis of variance (ANOVA). We must note that it is quite probable that certain pre-processing steps will be required, such as filling in missing values via regression techniques and normalizing samples with respect to correlated variables. Also, a very important step before we perform the analysis is the division of end-users into groups (or clusters), according to the similarity of their data. For this purpose, we ask certain information in the dataset, such as demographics, meteorological region, and control and testing groups divisions. In case we do not have sufficient data for manual group divisions, certain techniques will be incorporated (e.g. k-means clustering (Hartigan and Wong, 1979).

As soon as the division in clusters is complete, the next step is to identify the explanatory variables, which are variables that are not independent and have effects on the energy consumption. Another important variable type is the confounding variables. Confounding variables have effects on both independent and dependent variables, thus the latter might appear as correlated, without actually having a direct correlation. Confounding variables cannot be discovered easily, however, once they are, the adjustment procedure removes the biases that they induce in the dataset.

The main method that we plan to use for our analysis is the Analysis of Variance (ANOVA). According to this procedure, multiple subject groups are compared against each other with respect to average values and variances of certain measures (DeGroot and Schervish, 2010). For our purposes, the subject members are going to be divided into control and testing groups, i.e., some are going to apply EE interventions, while others are not. ANOVA is going to be used to assess the effectiveness of each EE intervention type when applied to different member groups. In some detail, we can use ANOVA to test how different EE interventions and incentives (factors) influence consumption for each energy end-user (observations). By filling the ANOVA table with the observations corresponding to the various factors, we will find out the correlations between these variables, and discover if the factors are additive. Importantly, also, note that if information on non-REScoop members is available, we can use ANOVA to compare the performance of groups of REScoop members with that of non-REScoop ones.

However, the exact methodology of analysis strongly depends on the actual data that REScoops will offer, thus the final decision regarding the statistical analysis method is subject to changes until the actual datasets are obtained. More specifically, ANOVA assumes that the data come from normal distributions, which, is not always the case when it comes to electricity consumption data. Alternative methods for hypothesis testing are the Student's t-tests, and the calculation of p-values (DeGroot and Schervish, 2010). Student's t-test can be used to determine whether two datasets vary significantly from each other, while the p-value indicates how probable is to observe "data-extremes"; for instance, when the p-value is less than (the most commonly used value of) 0.05, the null-hypothesis (which states that there is no statistically significant difference) is rejected.

Another, potentially more appropriate test which is non-parametric -i.e. does not make any assumptions regarding the distributions the data originate from - is the calculation of the Kendall's tau correlation coefficient. This test checks the correlation of two vectors and the result is a value between -1 and 1. If the result is 1, then the perfect correlation is detected; 0 indicates no correlation at all, and -1, perfect negative correlation. In our case, the first vector is the average consumption of each customer before EE intervention application, and the second the average consumption after application. Equipped a priori with the knowledge that each measurement in each of the two vectors corresponds to the same customer, we know that the samples are expected to be correlated.

Thus, in case the results show that there is no correlation, we can conclude that there exists behavioural change in consumption as a result of the EE intervention. To guarantee that the consumption is influenced only by the EE intervention and not from other known influential factors, instead of testing plain kWh measurements, we use normalized measures, such as kWh per Degree Days (DD), kWh per square meter of the buildings surface, and kWh per number of residents.

During the next few months, the aforementioned and other candidates for conducting hypothesis testing will be thoroughly examined to determine which are best for our needs. Our final choices will be reported in forthcoming deliverables.

To summarize, the statistical analysis technique that we will be using will most probably consist of the steps below.

a) Divide samples into similarity groups with respect to (wrt.) factors such as contract type, type of EE intervention applied, etc.
b) For each group:
    a. Calculate normalized average consumption for each individual consumer before and after EE intervention application
    b. Calculate average reductions in the normalized average consumption as a result of the EE intervention
    c. Test statistical significance with appropriate techniques from the literature (according to the assumptions that the available data fulfil)
c) Draw conclusions regarding the effectiveness of each EE intervention.

III. Results and Discussion

In this section, we present the results from the data gathering process, as well as some preliminary results[1] from two REScoop cases, that of Danish district heating cooperatives managed by EBO[2], and of a large Belgian electricity cooperative, ECOPOWER[3]. First, we summarize the population from which the gathered data originates from and provide intuitions regarding their contract types, and the different regions the customers are placed. Next, we briefly discuss the energy efficiency (EE) interventions that each REScoop has already applied, and, finally, we present the results from the statistical analysis of the reductions in consumption that the EE interventions induced.

III.1. Summary of Submitted Data by Two Cooperative Cases

Regarding the EBO case, the submitted dataset included monthly heating consumption values from 300 residential customers, which are cooperative members, for the period of 5/2012-9/2016. Additionally, EBO responded to past yearly consumption values of the members before joining EBO, and also with a dataset from a non-cooperative company, containing monthly data samples that indicate the consumption of 1,000 non-cooperative members. Most EBO member measurements were accompanied by the respective demographics regarding the buildings' surface in square meters, the number of residents, additional building characteristics, as well as meteorological data, e.g. minimum, maximum, and average temperature values, and heating degree days.

The EE intervention that EBO has applied to their members is termed as "Technical Support". The Technical support EE intervention includes technical inspections by experts and suggestions for equipment or insulation upgrades, etc. As numerical results illustrate, this particular EE intervention was effective in reducing consumption when applied to the cooperative members.

As far as ECOPOWER is concerned, consumption measurements were gathered on a yearly basis. The number of customers with measurements was significantly larger, totalling to 33,596 customers. The reported period regarded the years 2011-2015, and the number of residents for each building was included. Now, although ECOPOWER does not gather monthly data itself, some of their customers have subscribed to EnergieID, a consumption monitoring and analysis software service, which includes the gathering of monthly consumption data available, and also of the buildings surface in square meters, which is used as a normalizer for the consumption of these members. EnergieID service itself is the EE intervention that is analyzed for its effectiveness in reducing electricity consumption.

In particular, ECOPOWER customers come in three different contract types, which is an influencing factor in electricity demand. The contract types are three: Residential contract customers, Commercial contract, and Social contract, which is a special category that, by governmental regulation, receives the lowest rate for consumption existing in the market. As the data indicates, Residential customers have the lowest average consumption, with the Social contract following with 21% higher consumption than the Residential, and the most consuming being the Commercial contracts, with 63% higher consumption than the Residential contract, on average. The dataset population percentages of the different contract types are shown in Fig. 2.

[1] Note that, due to privacy concerns, absolute consumption values of the separate customer groups are omitted from publications.
[2] http://www.ebo.dk/
[3] http://www.ecopower.com

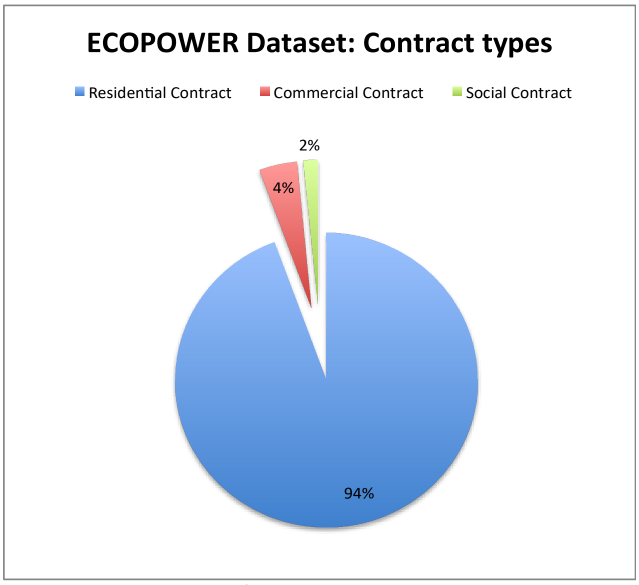

Fig. 2: Percentages of each contract type in the submitted dataset by ECOPOWER.

III.2. Assessing the Impact of EE Interventions in Consumption Reductions

As we mentioned in the previous subsection, each of the two cooperatives that are analyzed here, have implemented specific EE interventions: EBO applied Technical Support to their customers that is, a series of contacts and inspections by energy efficiency technical experts; while ECOPOWER applied EnergieID, an energy consumption monitoring and analysis software tool. Here, we validate the differences (reductions) induced by these two specific EE intervention measures.

In the case of EBO's Technical Support, results show that the treated customers achieved 20% reductions in their monthly kWh normalized by heating degree days' consumption on average, after receiving the intervention. This indicates that this particular EE intervention can be proven a valuable tool for other REScoops that also deal with district heating. The results of statistical significance are shown in Table 2. The p-value being less than 0.05, indicates that there is a significant difference between the sample distributions of the consumption measurements before, and after applying this particular EE intervention. In addition, Kendall's τ value being less than 1 means that the two vectors containing the values before and after the EE intervention are not correlated, thus a significant change in consumption behaviour has occurred. Furthermore, the obtained results show that the customers who became cooperative members reduced their consumption (in kWh/m^2) by 19.92%, as compared to the time before their membership. Receiving technical support led to a reduction of 21.42% in average heating energy consumption in kWh/ (m^2 * HDD). Moreover, the "technical support" EE intervention led to reduced CO_2 emissions for the members who received it - by 274.13 kg on average.

Tab. 2: Significance tests for the monthly average reductions achieved in kWh/HDD after applying the Technical Support EE intervention

Significance test	Result
p-value	*0.00017*
Kendall's τ	*0.414*

Moving forward to the next EE intervention, the energy monitoring and management software service, EnergieID, results have shown that on average of Residential and Social contract types for which the number of residents is available[4], treated customers achieve 11.97% reductions in their kWh/No. of Residents consumption values, indicating that such software services can be proven quite useful for such a cause. The p-value of the statistical test 2.403e-07 is very small, far lower than the 0.05 significance threshold, indicating that there is indeed a difference between the sample distributions before and after using EnergieID. The lower than 1 value of Kendall's τ=0.573 indicates that the vectors containing the values before and after the intervention application are not correlated, thus we can conclude that there exists behavioural change in electricity consumption after subscribing to this software service. The significance tests for each of the three different contract types are shown in Table 3.

Tab. 3: Significance tests for the yearly average reductions achieved in kWh/No. of Residents after applying the EnergieID EE intervention

Contract Type	Significance test	Result
Residential	p-value	*2.105e-07*
	Kendall's τ	*0.572*
Social	p-value	*0.256*
	Kendall's τ	*0.578*

Note that all null-hypotheses are rejected, indicating that there is behaviour change, apart from the Social contract type's p-value. This is mainly due to the fact that there is only a small sample of measurements from customers that both have Social contract type and subscribed to EnergieID (21 customers, in contrast to the 1,764 samples from the Residential contract type), and the samples are not described by a normal distribution.

The reductions in kWh/No. of Residents measurements for each of the two different contract types of ECOPOWER are shown in Fig. 3.

[4] To compare reductions for the Commercial contract case, additional data on buildings surface in square meters must be obtained.

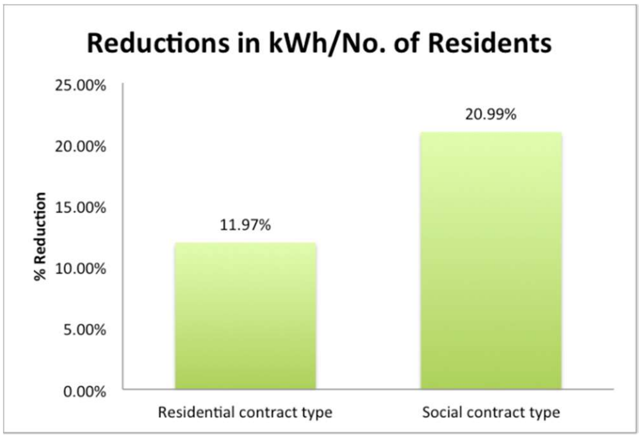

Fig. 3: Reductions in average yearly consumption after EE application for each contract type in the submitted dataset by ECOPOWER.

As we can observe, subscribing to the EnergieID service leads to significant reduction in the yearly average measurements for both contract types, with the Residential type customers being a little more affected in changing electricity consumption habits. The ECOPOWER clients who became cooperative members produced 235.12 fewer kg of CO_2 per year.

IV. Conclusions and Future Work

Gathering REScoops data and analyzing the energy consumption of their members is a highly complex task, but one which is nevertheless required in order to assess the impact of a variety of EE interventions. Such interventions are key for driving changes in the consumption behaviour of European citizens, in order to reach the much coveted near-zero emissions targets. REScoop Plus organizes the data gathering efforts of REScoops across Europe; and provides a methodology for assessing the effectiveness of any given EE intervention.

In this paper, we presented the exact steps of the REScoop Plus data gathering process; the main pillars of the common data format (set after a systematic research and deliberations process); and our plan for, and initial results of, the statistical analysis intended. We also presented the state of data currently being collected by each of the collaborating REScoops. To illustrate the analysis methodology, we evaluated two distinct EE interventions, namely, Technical Support and EnergieID, which have been applied by two specific REScoops in the recent past. Results of these two EE interventions demonstrate that they manage to induce almost 20% reduction in the normalized consumption of the REScoops customers that were applied to. Moreover, becoming a cooperative member has led to a reduction of CO_2 footprint by about 274.13 kg for EBO customers and by about 235.12 for ECOPOWER customers.

We conclude this paper by reporting that we already have additional preliminary analysis results for a number of participating cooperatives, which are very encouraging (though we did not present them here for the sake of brevity). Ongoing work includes analyzing in depth the consumption partners of all REScoops participating in the project; and evaluating a number of EE interventions. The results of our analysis will allow us to build a recommendation toolkit, which will be offered to existing and newly formed RES cooperatives, so as to help their members adopt behaviours that are truly energy efficient and thus environmentally friendly.

Acknowledgements

The research methodology and results presented in this paper are part of the H2020 - REScoop Plus Project. This project has received funding from the European Union’s Horizon-2020 research and innovation programme under grant agreement No. 696084.

Nomenclature

EE : Energy Efficiency
RES : Renewable Energy Sources
ICT : Information and Communication technology

References

DeGroot M.H., Schervish M.J., Probability and Statistics (4th Edition), Addison-Wesley, ISBN 0-321-50046-6 (2010).

Energy Efficiency Watch Project, Evaluation of National Energy Efficiency Action Plans (2009).

European Energy Mediators Group, Working Group Consumers as Energy Market Actors (2015).

Hartigan J.A., Wong M.A., Algorithm AS 136: A k-means clustering algorithm, Journal of the Royal Statistical Society, Series C (Applied Statistics) 28.1, 100-108 (1979).

REScoop Plus project, https://rescoop.eu/european-project/rescoop-plus, last accessed on 18/2/2017

RESCOOP PLUS project, Zero-point report on data of Supplying REScoops (2016)