Case Study for Energy Efficiency Measures of Buildings on an Urban Scale

Domenico Dirutigliano - Politecnico di Torino – domenico.dirutigliano@polito.it
Mark-Alexander Brüntjen – RWTH Aachen University - bruentjen@e3d.rwth-aachen.de
Christian Fliegner – RWTH Aachen University – fliegner@e3d.rwth-aachen.de
Jérôme Frisch – RWTH Aachen University - frisch@e3d.rwth-aachen.de
Vincenzo Corrado - Politecnico di Torino – vincenzo.corrado@polito.it
Christoph van Treeck – RWTH Aachen University – treeck@e3d.rwth-aachen.de

Abstract

The energy efficiency of existing buildings is one of the challenges launched by the EPBD recast. The RWTH Aachen University accepted this challenge and started the project EnEff: Campus - Roadmap aiming at reducing the specific primary energy consumption of the university campus building stock (about 300 buildings) by 50 % until 2025. For the estimation of refurbishments for this kind of big data, data mining techniques can be used like the CART method (Classification and Regression Tree). In this investigation, the method applied on the RWTH Aachen buildings stock and the estimated results will be compared to results from a simple data mining technique, called visual method. The comparison is performed by using low-order dynamic building model (LOM) performance simulation through the Modelica AixLib. The determined results of the recommendation of the CART method will be discussed and evaluated in this paper.

1. Introduction

This paper deals with work within the project “EnEff: Campus - RoadMap RWTH Aachen” (EnEff: Campus). The central aim of the project is to develop a road map for the RWTH Aachen University which leads to a cost-effective reduction of the specific primary energy consumption at RWTH Aachen University by 50 % until 2025, based on the energy consumption of 2013/14. The RWTH Aachen building stock counts about 300 buildings which differ for instance in the following characteristics: usage type, year of construction, and building structure typology. To reach the central aim of the project, a city district performance simulation is applied and a systematic approach has to be followed, by using LOM and distribution network energy performance models. The city district performance simulation needs the LOM to calculate the heating performance and demand in satisfying computation time. The parametrization of the LOM is set up by archetype buildings. Lauster’s investigations show that the used LOM leads to high accuracy compared to detailed simulation models (Lauster et al., 2014b). Concerning the usage of statistical data for enriching LOM parameters, Schiefelbein describes the generation of archetype buildings by only five input parameters: “building type, year of construction, floor height, number of floors, net floor area” (Schiefelbein et al., 2015a). As the accuracy of statistical data depends on the dataset, the parametrized LOM characterized by the five input parameters were investigated with respect to a similar building stock as the one of the RWTH Aachen Campus. The results achieved a corresponding compliance for the thermal city district simulation with respect to measurements (Lauster et al., 2014a). All things considered, Lauster showed that the LOM is suitable for city district simulation due to the accurate estimation of the heating load and energy demands (Lauster et al., 2014a).
This paper shows the possibility to identify the buildings, offering an efficient recommendation of measures for energetic retrofitting.

2. Data Mining Methods

The aim of the investigation is to apply data mining methods for the determination of efficient energetic retrofit measures on a city district scale. Data mining methods enable the examination of a large number of parameters, for instance those, which influence the energetic behaviour of a building stock, like building construction parameters, such as U-Values, transmission heat loss coefficient, average efficiency ratio of the energy supply, and ventilation rate. In this investigation, two different approaches will be compared. The first one is a visual method, which determines boundary sets in diagrams for filtering data, and the second one is the usage of the CART algorithm. The different approaches are compared by LOM building performance simulation. Therefore, the dataset is set up with the "Tool for Energy Analysis and Simulation for Efficient Retrofit", in short: TEASER (TEASER, 2016). This tool enriches a data set based on statistical approaches if information is scarce. Successively, the data set is applied to the recommended analysis.

2.1 Visual Method

There are simple methods to determine energetic building retrofit measures with a potential of energy savings, however, they do not always provide high savings. One of these simple processes to determine buildings is to set up diagrams of the building stock, as shown in Fig. 1 and 2. The visual information can be used for setting filters and estimating retrofit measures.

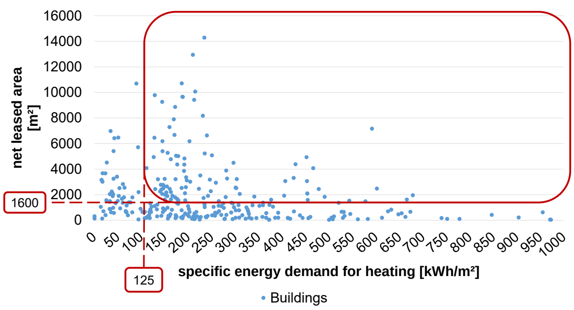

Fig. 1 – Qualitative overview of buildings with a huge net leased area and a high specific energy demand for heating

Fig. 1 is a typical illustration which helps to characterize a building stock. It is possible to extract huge buildings with a high-energy demand in the upper right part of the diagram, as highlighted by the frame. Nevertheless, there is no information about the distribution of the energy losses. Therefore, Fig. 2 illustrates a method to highlight buildings. On the abscissa in the diagram represents the average U-value [W/(m²K)] and on the ordinate the transmission heat loss coefficient [W/K] is plotted. The highlighted frame represents the buildings which have a high U-value and, due to the high H_T-value of the façade, high-energy losses are caused by transmission.

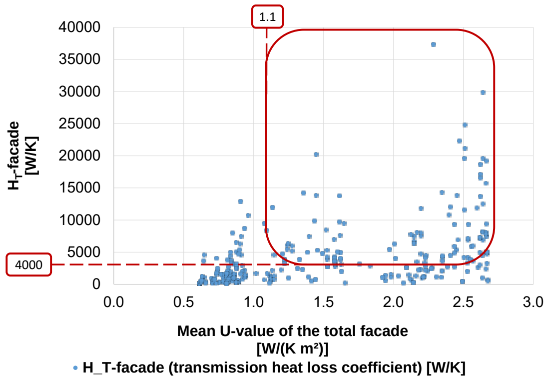

Fig. 2 – Qualitative overview of buildings with a high total facade U-Value and a big transmission heat loss coefficient H_T

These visual illustrations and analyses point out buildings, which could have a potential for retrofitting. Nevertheless, it has to be mentioned, that these are processes accounted by only four to five parameters, they hence represent a very simple way of filtering. On city district level, it is common that the energetic behaviour is influenced by more than these parameters. Thus, the data mining method will be applied to the building stock of the RWTH Aachen Campus.

2.2 Decision Tree with CART Algorithm

Data mining aims to determine models for decision making. In the energy context, the models used in this investigation represent two different types and depend on the scope: classification on the one hand and regression models on the other hand. The first kind of models try to assign a class for each observation (each line in a data set), considering information derived from a data set with classes which are already known (called learning sample). The second kind of models predict the attributes of a dataset which influence a given outcome stronger than other attributes. Common techniques of classification are decision trees. The algorithms which are frequently associated with decision trees are: ID3, C4.5, CART, CHAID, SLIQ, SPRINT. In this paper, a decision tree is chosen to show a group of rules of classification in a tree scheme and it is matched with the CART algorithm (Breiman et al., 1984). A regression tree is used to predict problems in case the response variable is numeric or continuous. This algorithm is adopted for a supervised multistage decision-making process to classify the observations in a finite number of classes. In the literature, this approach has already been tested to rank flats based on calculated normalized primary energy demands calculated with a quasi steady-state method (Capozzoli et al., 2016).

The decision tree starts with the root node which contains the complete data set and is used as learning sample. Successively, the decision tree subdivides the data set using a binary split in homogeneous subsets, considering 2k-1 ways of creating a partition of k attribute values, and gives the origin to a new node. The last nodes in the tree are called leaves and each node is labelled with the attribute's name. The tree branches show the path which respects a series of rules and classifies the samples. With a rising number of rules, the tree appears more and more complex which should be avoided to maintain the usability. For this reason, a so-called pruning can be applied. The criterion used is called Gini Index, which evaluates the degree of impurity of each node. The data are split for each node that maximizes the decrease of impurity.

Another element to characterize the tree is to evaluate the statistical performance of the model if a new dataset is used. In this investigation, a k-fold cross-validation is applied. This technique divides the dataset in equal k-parts and for each step; one part is used for the validation of the data set, while the other one is used for training the dataset.

The models are developed with Rapid Miner 7.3.001.

2.3 TEASER and Low Order Building Model

TEASER uses statistical approaches based on the IWU (Loga et al., 2005) building typology (Schiefel bein et al., 2015b). The minimum required input data consist of the following five parameters: year of construction/ year of retrofit, building height, net leased area, number of storeys, and usage type.

These parameters are the basis to estimate envelope areas for exterior walls, windows, rooftops, and basement. Furthermore, the constructions of envelope structures are parameterized. This data enrichment provides a full dataset for the “MultizoneEquipped.mo” zone model of the Aixlib library. In this investigation, TEASER is applied to set up building models of the RWTH Aachen Campus building stock and is used to highlight the differences of recommended estimated retrofit measures.

The mentioned LOM “MultizoneEquipped.mo” is an RC-Model based on the German Guideline VDI 6007-1 (Lauster et al., 2014b). Lauster modified the guideline model by adding an extra resistance representing the thermal behaviour of window elements. To keep the information content low, a minimum number of zones should be the aim of low order modelling. Therefore, only a small number of zones represent the building in the thermal building performance model.

The accuracy of the TEASER tool chain for enriching the data by the mentioned five parameters to set the lumped parameters was evaluated by Lauster (Lauster et al., 2014a), and assessed to be suitable for city district energy performance simulation.

3. Data Set

3.1 Origin of the Data Set

The building stock of the RWTH Aachen University campus consists of about 300 buildings.

For this investigation, some energy-related facts and specific particularities are of interest, for instance, the campus’ total energy demand for heating amounting to 126,000 MWh or the specific value of the energy consumption (EC) with respect to the net leased area of about 236 kWh/m² (Facility Management, 2014). The distribution of the parametersrelevant for the description of heating energy losses is illustrated in Fig. 3 and 4.

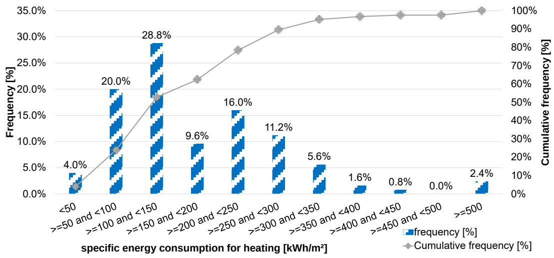

Fig. 3 – Distribution of the energy consumption of the RWTH Aachen University Campus, divided into efficiency classes

Fig. 3 shows the distribution of the EC of the RWTH Aachen building stock (only 125 consumption values are available). The following Fig. 4 describes the distribution of the estimated energy demand (ED) by applying TEASER and LOM (299 data of ED are available); the estimated average of the yearly ED for heating is about 249 kWh/m² with respect to the net leased area.

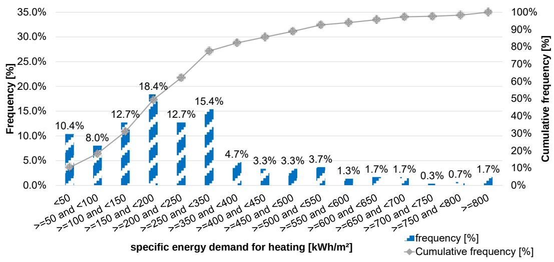

Fig. 4 – Distribution of the energy demand of the RWTH Aachen University Campus, estimated with TEASER and LOM

Fig. 3 and 4 illustrate that about 60 – 80 % of the buildings have a high specific ED for heating. This could yield to the assumption that a lot of buildings should have a high potential for energetic retrofitting. In the further reading, some characteristics of the data set are presented.

3.2 Characteristics of the Data Set

To show some important characteristics for the description of the energetic behaviour of the buildings, like U-values of the total vertical façade or the opaque facade following histograms are illustrated in Fig. 5 and 6.

The distribution of the total mean U-value of the facades is shown in Fig. 5. The figure shows that there are about 35 % of buildings with a U-value above 2.1 W/(m² K) and approximately another 20 % above 1.2 W/(m² K). Hence, the focus of the investigation is indispensable and the main goal is the determination of facade retrofit measures.

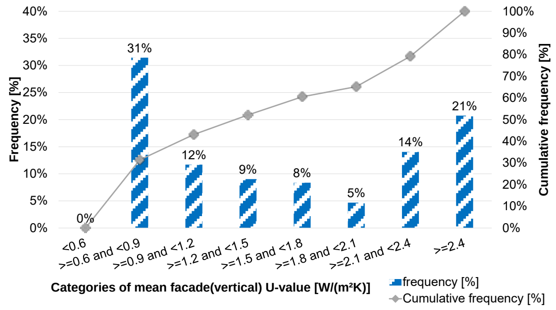

Fig. 5 – Distribution of the total mean U-value of the building facades, based on data estimated with TEASER

Fig. 6 shows the allocation of U-values from the opaque part of the facade.

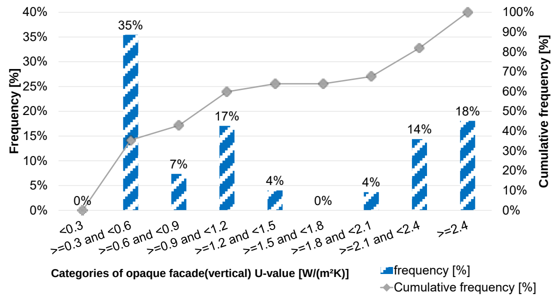

Fig. 6 – Distribution of the U-value of the opaque building facades, based on data estimated with TEASER

Furthermore, Fig. 6 illustrates that the opaque U-values are responsible for about 30 % above 2.1 W/(m² K) and approximately 42 % under 0.9 W/(m² K). This leads to the allocation of window U-values. As mentioned before, the data set which describes the campus was emulated by TEASER; therefore, only two categories of windows are available. The values are between 1.5 W/(m² K) and 2.0 W/(m² K), and above 2.8 W/(m² K).

3.3 Determination of Retrofit Measures by the Visual Method

As mentioned in section 2, the filter settings for determining the buildings which have to be retrofitted are indicated by the diagrams in Fig. 1 and 2. For this investigation two filters are set and applied on the data emulated with TEASER.

Fig. 1 shows the boundaries of the first visual method filter (VM filter setting 1): a high specific ED of 125 kWh/m² and a high total ED (considered by a high net leased area of 1600 m²). These boundaries result in a dataset of about 136 buildings. Moreover, retrofit measures would never be performed, unless they are absolutely necessary. Thus, it is indispensable to consider the U-value, which is set to 1.1 W/(m²K), as highlighted in Fig. 2. Based on to the U-value, the first retrofit recommendation is given and results in 70 buildings with a potential to retrofit the opaque facade. To determine buildings with high transmission heat losses, a high H_T-value, here 4000 W/K, is considered and leads to a second visual filter (VM filter setting 2). This results in a potential data set of about 57 buildings.

3.4 Pre-processing of the Data Set

In this phase, different strategies are considered to prepare the data for an analysis. Firstly, outliers are detected and the values are normalized. Secondly, the variables which influence attributes are selected. Finally, a data transformation is carried out.

To detect the outliers of the data set, a distance-based outlier detection algorithm is applied. Thereby, the Euclidean distance is calculated between the data points, and the ones with the greatest distance from other data points are marked as outliers.

In order to grant equal consideration of the attributes, it is necessary to normalize the data set.

After the data analyses and the review of similar studies in literature (Capozzoli et al., 2016), the following attributes are selected: aspect ratio S/V, heat transfer surface on heated volume in [m^{-1}]; U-value opaque, U-value of the vertical opaque envelope in [W/(m² K)]; H_T-value wall, the mean overall heat transfer coefficient by thermal transmission of the opaque components in [W/K]; U-value window, U-value of the vertical opaque envelope in [W/(m² K)]; H_T-value window, the mean overall heat transfer coefficient by thermal transmission of the opaque components in [W/K].

The attributes are chosen based on the information gain they can give. For this reason, it is common that the attributes of the data set are independent and only the label attribute is clearly dependent of the other attributes. In this paper, Data Sets 1 and 3 are used with all the variables showed before with the exclusion of H_T. Data Sets 2, 4, and 5 take all the previously shown variables into account. The latest data set is shown in section 6.4 to compare the results and to investigate how the information gain using H_T variables can be used, despite the correlation with the U-values.

Data transformation introduces criteria to label each building according to the "high", "medium" or "low" category. These labels are necessary, as the classification tree is based on a categorical response variable. Each "high" category starts from the median value to the maximum value of energy performance. The thresholds between the categories "high-medium" and "medium"-"low" of the ED data set are 241.05 kWh/m² and 50.00 kWh/m², respectively. The thresholds between the categories of the EC data set are similar with 177.84 kWh/m² and 74.00 kWh/m², respectively. The threshold limit of the "low" category applying ED comes from the energy efficiency class of EnEv2014 (BMWi, 2014). The threshold limit of the "low" category applying EC is based on a similar percentage of buildings as in the "low" category applying ED. The percentage of buildings in the categories "high", "medium", "low" with the ED data set is 36 %, 54 % and 10 %. The categories with the EC data set have the following percentages 41 %, 50 % and 9 %.

4. Limitations

In the following, some boundary conditions shall be mentioned. The applied low order models used for this investigation are supplied by the AixLib library version "The Modelica _Annex60_ library". This library is currently still under development and, furthermore, TEASER enriches the parameter sets for LOM and uses statistical approaches.

5. Decision Trees

In this investigation, two different approaches are evaluated and a set of buildings of the RWTH Aachen building stock with opaque facades should be retrofitted. Two different decision trees are evaluated with CART algorithms; the first one is determined with the input of the specific EC for heating with Data Set 1. It is illustrated in Fig. 7.

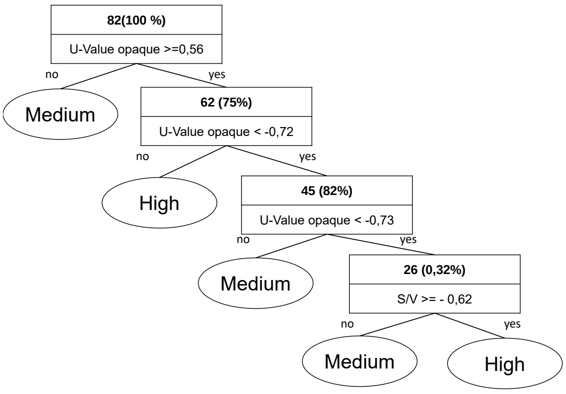

Fig. 7 – Decision tree, based on EC with Data Set_1

After the pre-processing of Data Set 1 with EC, the decision tree classifies 82 buildings.

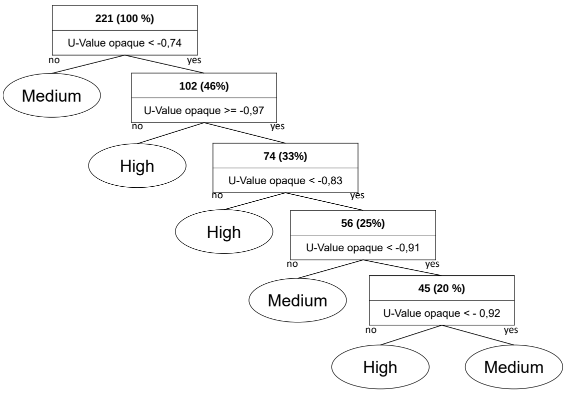

Fig. 8 – Decision tree, based on ED with Data Set 3.

The decision tree of Data Set 3 (normalized data) with ED is shown in Fig. 8. It classifies 221 buildings, thus, more buildings than the first one.

6. Supervised Classification Process

6.1 Analysis of the Classification Tree Split Attributes

The first attribute enables us to split the data in the Root Node, representing the one with the most influence on the energy consumption or demand.

The decision tree based on EC has 5 leaf nodes and a tree size of 9. The main attribute is if the U-value of the opaque facade is bigger or equal than 0.56. Furthermore, in this decision tree, there aren't any attributes about transparent components.

The decision tree, based on ED, has 6 leaf nodes and the size of 11. The main attribute of this decision tree is if the U-value of the opaque facade is smaller than 0.7 (with normalized data).

6.2 Classification Accuracy

The training records which are correctly classified by the decision tree based on EC are about 66 % of all buildings. The accuracy of the same model is 52.97 % with 5 k- folders of the cross validation.

The training records of the decision tree based on ED are about 78 % of the whole considered buildings, and the model accuracy about 72.28 % with 10 k-folders of the cross validation. The accuracy of the whole classification of about 70-80 % (Gao et al. 2010; Yu et al., 2010) is considered acceptable. A lower number of buildings influences the model based on energy consumption negatively. The model of classification based on ED is recommended to evaluate retrofitting measures for higher accuracy.

6.3 Evaluation of Retrofit Actions

The decision trees visualize the main attributes which classify buildings and influence the energy consumption or demand. In the upper part of the tree, close to the Root Node, there are attributes that classify most of the buildings.

Each node could consider a retrofit action. In this study, for each building the following retrofit measures are considered: retrofitting of only transparent components or retrofitting of only opaque components. The retrofit actions are applied only in leaf nodes.

An attribute doesn’t necessarily give the possibility of a refurbishment, such as in the case of the S/V-ratio (last node of the Fig. 7). The retrofit actions are applied on all the buildings with the characteristics indicated by the attributes in the leaf nodes, including the buildings not classified by the decision tree. The excluded buildings from retrofit actions belong to the “low” categories (both with EC and ED). These are not considered, because priority is given to the buildings which are classified as “high” and “medium”.

6.4 Results

With the visual approach and the CART approach, four different recommendations of a data set are evaluated. These contain the buildings with a high potential for retrofitting the opaque facade. With the visual method, the following filters are used: filter setting 1 with spec. energy demand > 125 kWh/m² AND net leased area > 1600 m² AND U-value > 1.1 W/(m²K); filter setting 2 with spec. energy demand > 125 kWh/m² AND net leased area > 1600 m² AND U-value > 1.1 W/(m²K) AND H_T-value > 4000 W/K. The visual method filter recommends about 70 buildings, whereas the second filter selects about 57. The first CART method based on EC recommends 73 buildings and the second, based on ED, recommends 268 buildings. All the results of retrofit actions are analysed by applying a building performance simulation using LOM. The results are illustrated in Fig. 9. They show that with the CART method, it is possible to save more energy while refurbishing more buildings. In order to investigate possible recommendations, an analysis is conducted with the Data Set 2 (with H_T and U values). Fig. 9 shows that both recommendations of the visual methods yield an energy saving percentage of about 7 % and 8.25 %. This means that the influence of 13 buildings which are retrofitted in addition to the second data set offer no great advantages.

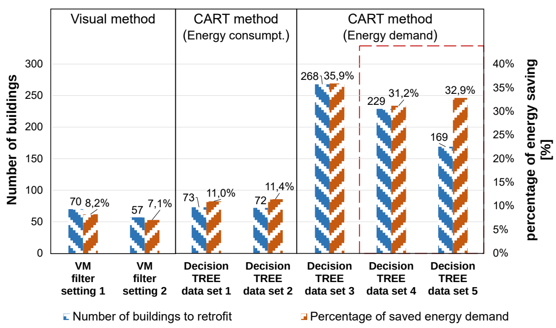

Fig. 9 – Comparison of different evaluated recommendations by the two different methods: visual method and CART method

Data Sets 1 and 2 do not show large differences. It is possible to see that Data Set 3 has the most energy saving potential but also the highest number of buildings to refurbish. Data Sets 4 and 5 consider the same data set but with different retrofit actions. In Data Set 4, the buildings are refurbished following the attributes of the decision tree and, therefore, with actions on opaque and transparent components. In Data Set 5 all the buildings are object of only opaque retrofit actions.

The estimated recommendation by applying Data Set 2 of the CART method results in energy savings of about 11.4 % and with Data Set 4 of about 31 %. However, Data Set 5 saves about 33 % with fewer buildings, as in the case of Data Set 4. This proves that also retrofit actions should be analysed. Concerning Data Set 5, it has to be mentioned that more than half of the building stock is recommended to be retrofitted. This value seems to be high, but in contrast to a theoretical investigation of retrofitting all 299 campus buildings, the percentage of energy savings was calculated to be 36 %.

Hence, the CART method identifies the main influencing parameter, as it recommends 169 buildings to be optimized (57 % of the building stock) and offers a reduction of about 33 % energy savings, which is very close to the theoretical investigation. However, the accuracy of the decision tree of ED with Data Sets 4 and 5 is lower than the decision tree of ED with Data Set 3.

7. Conclusion

In this paper, two different data mining approaches, namely the visual method and the CART method are analysed and evaluated. The result of each data mining concept is a list of buildings. These buildings are recommended to be retrofitted concerning their opaque vertical facade. The first method depends on human interpretation and is subjective, whereas the latter method is based on a statistical data mining process, which is more objective. The main differences between the methods are the handling (time), reliability, and accuracy. Thus, for a quick recommendation to estimate data sets, the visual method could be considered. But if the input data are reliable and recommendations should be dependable, the CART method should be preferred.

In further investigations, the estimated recommendations of combined retrofit measures will be analysed and discussed.

References

Breiman, L., J.H. Friedman, R.A. Olshen, C.J. Stone. 1984. *Classification Regression Trees*. Chapman and Hall/CRC.

Capozzoli, A., G. Serale, M.S. Piscitelli, D. Grassi. 2016. "Data mining for energy analysis of a large data set of flats". In: *Proceedings of the Institution of Civil Engineers - Engineering Sustainability* 170: 3-18. doi: 10.1680/jensu.15.00051.

Facility Management. 2014. "Energiebericht 2013". RWTH Aachen University. Accessed January 04. 2017. http://www.rwth-aachen.de/cms/root/Die-RWTH/Einrichtungen/Verwaltung/Dezernate/~pvm/Facility-Management/

Federal Ministry for Economic Affairs and Energy (BMWi). 2014. "Energieeinsparverordnung (EnEV 2014)". 2014. Accessed on November 18th 2013. http://www.bmwi.de/Redaktion/DE/Downloads/E/envkv-entwurf-lesefassung-aenderung.pdf?__blob=publicationFile&v=1

Gao, Y., E. Tumwesigye, B. Cahill, K. Menzel. 2010. "Using data mining in optimisation of building energy consumption and thermal comfort management". *Software Engineering and Data Mining*: 434-439.

Lauster, M., J. Teichmann, M. Fuchs, R. Streblow, D. Mueller. 2014a. "Low order thermal network models for dynamic simulations of buildings on city district scale". *Building and Environment* 10: 223-231. doi: 10.1016/j.buildenv.2013.12.016

Lauster, M., M.A. Brüntjen, H. Leppmann, M. Fuchs, J. Teichmann, R. Streblow, C. van Treeck, D. Müller. 2014b. "Improving a low order building model for urban scale application". In: *Proceedings of BAUSIM 2014, German –Austrian of IBPSA Conference.*

Loga, T., N. Diefenbach, J. Knissel, R. Born. 2005. "Entwicklung eines vereinfachten statistisch abgesicherten Verfahrens zur Erhebung von Gebäudedaten für die Erstellung des Energieprofils von Gebäuden". Online resource. Accessed on November 25th 2016. http://www.iwu.de/fileadmin/user_upload/dateien/energie/werkzeuge/iwu-kurzverfahren_energieprofil-endbericht.pdf

Schiefelbein, J., A. Javadi, M. Lauster, P. Remmen, R. Streblow, D. Müller. 2015a. "Development of a city information model to support data management and analysis of building energy systems within complex city districts". In: *International Conference Future Buildings & Districts Sustainability from Nano to Urban Scale Lausanne* 2: 949-954. doi:10.5075/epfl-cisbat2015-949-954.

Schiefelbein, J., M. Diekerhof, A. Javadi, G. Bode, R. Streblow, D. Müller, A. Monti. 2015b. "Development of a tool chain for complex city district energy system modeling and simulation". In: *Proceedings of Building Simulation 2015*. Hyderabad, India: IBPSA.

TEASER. 2016. Energy Efficient Buildings and Indoor Climate, TEASER – "Tool for Energy Analysis and Simulation for Efficient Retrofit" released. Online resource. Accessed February 4th 2017. https://www.ebc.eonerc.rwth-aachen.de/go/id/gips/?lidx=1#aaaaaaaaaaagipu

Yu, Z., F. Haghighat, B.C.M. Fung, H. Yoshino. 2010. "A decision tree method for building energy demand modelling". *Energy and Buildings* 42: 1637–1646. doi:10.1016/j.enbuild.2010.04.006.