International Energy Agency
Programme on
Energy in Buildings and Communities

Statistical analysis and prediction methods

Separate Document Volume V

Total energy use in buildings

analysis and evaluation methods

Final Report Annex 53

November 14, 2013

Stefano Corgnati, Thomas Bednar, Yi Jang, Hiroshi Yoshino,
Marco Filippi, Stoyan Danov, Natasa Djuric, Marcel Schweiker,
Christian Ghiaus, Alfonso Capozzoli, Novella Talà, Valentina Fabi

Volume V

Statistical analysis and prediction methods

Authors:	Stefano Corgnati, Thomas Bednar, Yi Jang, Hiroshi Yoshino, Marco Filippi, Stoyan Danov, Natasa Nord, Marcel Schweiker, Christian Ghiaus, Alfonso Capozzoli, Novella Talà, Valentina Fabi
With contributions from:	Tianchi Hu, Sawako Nakamura, Ayako Miura, Naomi Moroshita, Karin Schakib-Ekbatan, Simona D'Oca, Vincenzo Corrado, Fiorella Lauro, Fariborz Haghighat, Zhun Yu, Tianzhen Hong, Qingpeng Wei, He Xiao
Reviewed by:	Markus Kratz, Pierre Herant Jan Hensen, Philippe André

Contents

V-1

Statistical analysis of total energy use

Contents

1. General overview

1.1 Key points

What is covered
Assessment of potential application of statistical analysis for the prediction of total energy use in buildings and for the identification of the related most significant influencing factors. This has been covered first by an extended literature review, followed by the collection and critical analysis of activities carried out by ST-C working group with reference to: individual buildings and large building stocks up to regional/national level.
A deep connection has been established between Subtask C and Task Force on Occupant Behavior, related to the explanation of OB through statistical and probabilistic methodologies, and Subtask A, for the definition of the structure of the database ("database typologies")

Why it is important
To select a suitable methodology, the "scale" of the analysis is essential. To this aim, three main descriptors have to be considered: number of buildings to be analyzed (from an individual building to very large building stocks), number of items describing each building and time frequency of the collected time dependent parameters (annual to sub-hourly time frequency). This consideration fits with the proposal of the 3 Level Database in ST-A related to the complexity of the database ("database typologies").

The main fields of application for statistical analysis are:

- energy diagnosis, in case of individual buildings,
- energy consumption targeting and benchmarking, in case of large building stocks,
- tendency for energy policies, in case of analysis at regional/national level.

Key points learned

- The availability of suitable databases is a fundamental pre-condition to perform consistent analyses.
- Even if using statistical tools, do not forget the physical meaning of the parameters.
- Energy use can very often be described by a few main influencing factors.
- Among the influencing factors, at present only a few databases contain items related to occupant behavior description.
- Among the statistical models, regression models are mainly used for total energy use ranging from simple linear regression to complex neural network.
- Often, increasing model complexity does not increase the prediction accuracy.

Conclusions
Suitable statistical models to apply for energy use analysis have been highlighted, and recommendations about the proper application of the different models as a function of the goal of the analysis are offered. They depend on the time scale (dynamic models are for time scale of hours, static or statistical models are for time scale of months or years) and on the space scale (the variance is larger for individual buildings than for a large stock of buildings). The most important factors influencing total energy have been highlighted as well. The potentialities in using these models is very high for both individual buildings and large building stocks, but the pre-conditions are the clear

definition of the goal of the analysis and the availability of suitable data where the influencing factors required for the analysis are collected.

- It is important to the community that we assessed the potential applications of the tools to the field of total energy use.
- Benefit of the fit between building and occupant behavior of energy saving, cost saving, and thermal comfort.
- We highlighted the most important parameters and showed that models are different in terms of space and time.

1.2 Introduction

Interest in the analysis of actual building energy consumption has rapidly increased during recent years when the attention of researchers and technicians has started to move from the calculated energy demand to the real energy consumption of buildings. This crucial shift is strictly linked (not only) to ICT technologies related to energy use that were developed in the last few years.

ICT technologies are addressing a wide diffusion of energy and indoor environment long term monitoring systems, using both wired and wireless technologies. Thanks to the strong improvement in data acquisition and transmission technologies, it is now possible to have a real time picture of energy consumption and indoor environmental quality levels in buildings. But, it is fundamental to establish a clever monitoring plan in order to collect data coherently and enable the subsequent data analyses aimed at identifying the main influencing factors. To this goal, it is really important to clearly define which parameters have to be monitored, where the sensors have to be placed, how many sensors are needed and what is the suitable frequency of sampling. These decisions are strictly related to the available budget and to the results of a cost-benefits analysis. Due to these reasons, the identification of the most relevant influencing factor is also fundamental for reducing the number of parameters to be monitored.

Buildings are becoming more and more complex since different energy carriers provide energy uses for satisfying all the services of the building. The possibility of placing sensors in the building to measure all of the influencing factors related to energy use will allow for the collection of the required data. The possibility to set a platform to collect and to elaborate data allows for a huge amount of information, but when moving to the reality a plan to obtain the right interpretation of the data and information is needed. The need to analyze actual energy consumption is mainly due to the difficulties in making a realistic assessment of building energy demand when the calculation model is not suitably calibrated based on detailed knowledge of the real building behavior. In this way, a crucial role is represented by occupant behavior, especially when skipping from calculated data of energy consumption in buildings to real energy consumption. Moreover, tendencies and statistics about building energy consumptions may help to understand the actual dynamics of building energy consumption.

This change of perspective from standard calculated building energy performances to actual measured building energy uses is nowadays an important topic of research. As well known, in the first decade of this century much emphasis has been placed on the definition of indicators for characterizing the energy performance of a building. An example of the effort given by the technical and research community, together with the political bodies, is the large movement connected to the development and dissemination of the building energy certification in Europe, starting from the Energy Performance of Building Directive of 2002.

According to the definition proposed in the Directive, building energy performance has been mainly interpreted as an indicator of the building energy behavior related to “standard“ operative boundary conditions. The word “standard” highlights a crucial assumption in the calculation procedures and it is clearly explained in the picture proposed at the very start-up of IEA-ECBCS Annex 53 (see Figure 1-1). Thus, the starting point was to analyze the theoretical energy consumption, a shift toward the actual energy use is needed and all the influencing factors related to energy use should be assessed.

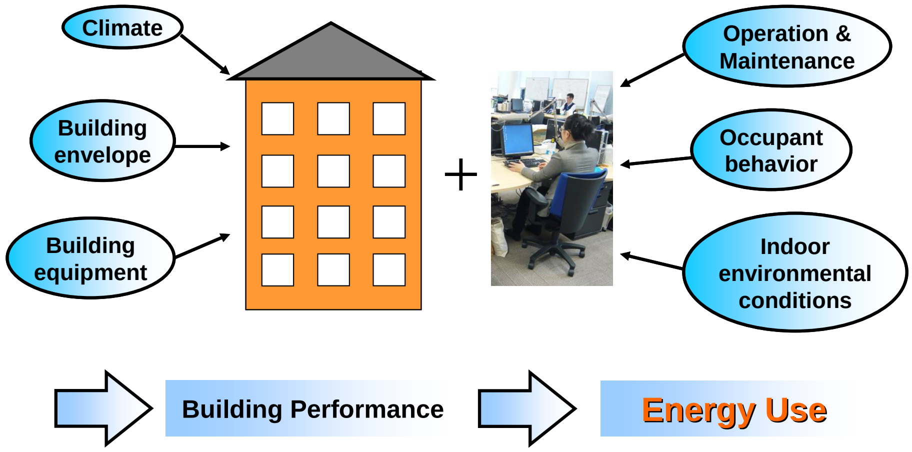


Figure 1-1: Influencing factors on total energy use in buildings (IEA-EBC Annex 53)

In Figure 1-1, the building energy consumption influencing factors are grouped into 6 main categories. The three categories listed in the left side of the picture (climate, envelope, equipment) are related to variables influencing "intrinsic" building energy performance which is calculated by fixing standard conditions for the other three categories listed in the right side of the picture (operation & maintenance, indoor environmental conditions, occupant behavior) that are specifically related to the actual building functioning. As a consequence, the building energy performance is calculated assuming that all of the analyzed building systems and functions behave ideally under the same standardized functioning/working conditions.

This approach allows a coherent comparison of the building energy performance calculated for different buildings but this energy performance is not strictly related to the actual energy consumption. When the attention moves to real energy consumption, all the six categories of influencing factors have to be taken into account to give a complete picture; moreover, a seventh category (social aspects) is also mentioned by Annex 53. As demonstrated in practice, buildings located in the same place (same climate) with the same building envelope and system characteristics, and consequently with the same value of the building energy performance index, may show high differences in the real energy consumption (see Figure 1-2) due to:

- different actual operation and maintenance,
- different actual indoor environmental quality level,
- different behavior of the occupants (ranging from energy conscious to energy unconscious).

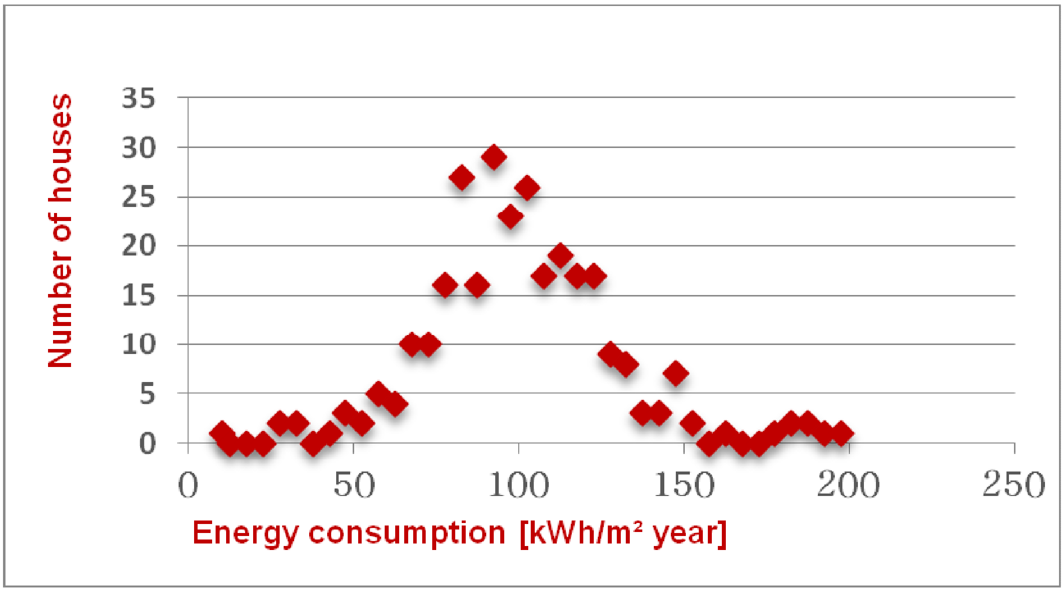


Figure 1-2: Frequency distribution of heat consumption in 290 investigated "identical" houses in Denmark, Henningsen, 1999)

The increased number of considered influencing factors, from three (related to energy performance) to six (related to energy use), as shown in Figure 1-1, highly amplified the complexity of the problem, especially because the last three categories are connected to parameters which are not deterministic and constant, but can change stochastically with time.

This consideration demonstrates the difficulties associated with realistically predicting building energy consumption using energy calculation tools such as dynamic energy simulations that are based on so called "direct (calculation) methods". In fact, it is very difficult to suitably describe the stochastic variation of the input parameters for the building energy calculations connected to the three last categories of influencing factors mentioned above.

At the same time, some important questions emerge including the following:

- Do all the factors have the same magnitude of impact on building energy consumption?
- Which are the factors showing the largest influence on building energy consumption?
- Which are the dominant factors in terms of effect on building total energy consumption?

To find an answer to these questions, it is necessary to primarily or exclusively focus of the investigations on those factors showing the greatest impact. Moreover, the identification of those factors may allow the development of prediction models based on so called "inverse modeling" techniques (this specific issue will be discussed in the following chapters).

In order to perform this kind of analysis, it is fundamental to establish a database where the information about both the energy consumption and the parameters related to the six influencing factors are collected. One of the key elements when statistical based tools are used for analysis is to clearly define to subject of the study. The diagram in Figure 1-3 shows the main investigation fields. It represents on its axes the building sample dimension and the amount of information for each building analyzed that is required to obtain suitable data for the analysis. For example, if the building sample is represented by only one building (an “individual building”) lots of detailed information should be gathered in order to have suitable information. The collective experience of the Annex 53 partners presented in Chapter 3, shows different types of information is sometimes collected, from very detailed up to the breakdown of each single final energy use. On the contrary, in the case of regional or national analysis, little information about each building in the sample is needed to provide some basic but interesting statistical analysis. Starting from these considerations, some questions arise:

- How detailed does the database need to be for each building?
- What is the format of the database?
- Which method should be used to analyze the database?

These questions are strongly connected with the goal of our analysis. This picture shows how to synthesize the purpose of the analysis.

If the goal is to analyze the individual building to show its energy consumption behavior, the benchmark is the building itself although an absolute benchmark may still prove useful (i.e. the behavior of the building from one year to the next, from one month to the next). Here, by collecting the suitable amount of information it is possible to make a detailed energy diagnosis of the examined building. In fact, this approach is typically used to make an energy diagnosis of the building and after to plan energy saving measure.

When moving to a large building stock, an initial goal is to find homogenous buildings that are grouped together in a large stock, and determine some target values, baselines and benchmarks for the building energy uses. In this specific case, where for example aggregated data for energy consumption is gathered, the idea is to make statistical analysis to provide information that can also be useful for planning energy saving actions on a national scale.

The analysis of an individual building and the national analysis of building stocks have different goals. For individual buildings the goal is to address the ability of a single building to have continuous improvements in its energy behavior. At the regional or national level the goal is to define some specific guidelines for suitable energy policies for energy retrofit actions.

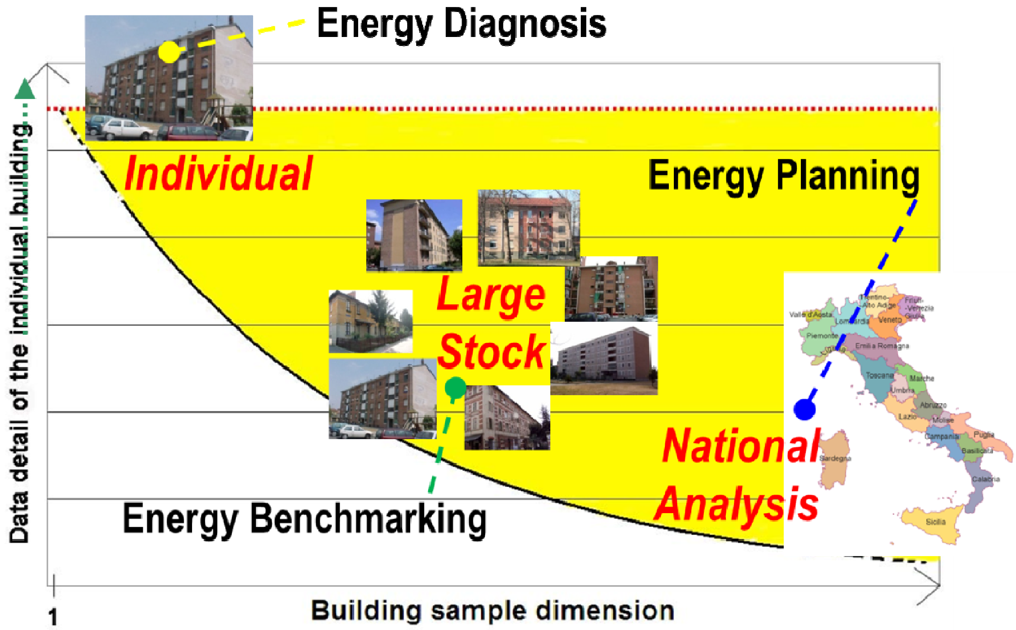


Figure 1-3: Main field of investigations related to the building sample size

Another key aspect regards not only how much data is in the database but also the frequency of the time related parameters. They typically range from hourly or sub-hourly data (detailed monitoring in one specific individual building) to yearly data (to have a general overview of the tendencies of a country).

One of the first goals of this annex is to provide a clear picture of the different databases and their possible applications.

A database is a fundamental element required before utilizing a data driven approach for statistical analysis. A fundamental aim of this data driven approach is to find relationships between influencing factors and energy use in buildings. In this annex, the influencing factors are grouped into six main categories (Figure 1-1) which are families of influencing factors. For each of them, it should be defined which items are to be collected (e.g. climate means outdoor temperature, HDD, CDD, etc.) and then verify if a relationship exists between the influencing factors and the final energy use.

As underlined in the following chapters, the database may show different characteristics according to the subject of the study (from individual buildings up to national building stocks), the categories of influencing factors considered (from climate only to all six categories), the variables collected within each specific category, as well as the frequency of the time dependent variables and consumption data collected (from annual to 15 minute intervals).

Moreover, the creation of a suitable database is the first and essential step to perform a number of statistical analyses addressed to describe the subject of the study.

As mentioned, the possible applications of statistical analysis may be divided into two big fields: to analyze individual buildings, or to focus the analysis on a large building stock. Firstly, statistics could be used to describe the object of the study (descriptive statistics) providing a clear description of the actual energy consumption and then to find out which are the dominant influencing factors in relation to the dependent variable (energy use) and independent variables (the influencing factors). When the most important influencing factors are known, statistical analysis could be used to build a prediction model.

Statistics can also be applied for creating reference buildings for a given building stock that can be implemented directly in building energy simulation tools.

Another possible application of statistics is to define “modules” meant to provide statistical inputs directly to a building energy simulation tool. For example, when dealing with occupant behavior (e.g. the adjustment of a thermostat) the problem is somewhat non deterministic, but is related to the probability of doing a certain action when some environmental parameters are present. So the input data (probability) for the direct simulation tool could be defined through a statistical approach.

According to this general scheme, the application of statistical analysis can be structured in three levels of investigations (Figure 1-4). The first level of investigation is a basic level. Since an amount of data is available, first of all tendencies related to the dataset should be identified. The use of statistical parameters (mean value, standard deviation, etc.), frequency distributions of the collected data, and etc. can provide significant information to define a clear picture of the subject of the study: it’s the use of statistics to describe. The second level of investigation utilizes statistics to find out the influencing factors that have a dominant effect on energy uses. If the most dominant influencing factors can be reduced to a limited number of parameters, it’s possible to find out the relationship between these parameters and the final energy use. Consequently, in the third level of investigation a very quick and robust prediction model can be built to provide information about the energy behavior of the building.

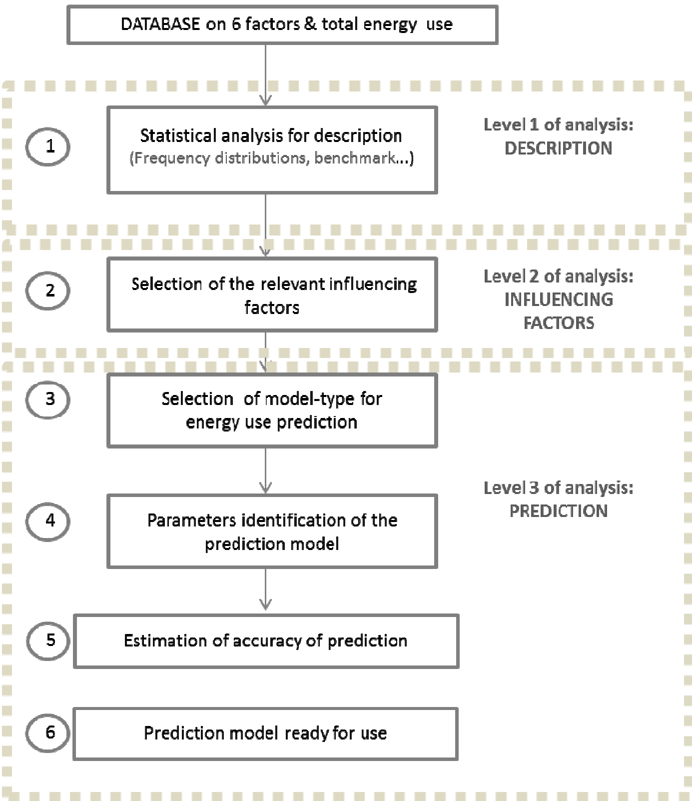


Figure 1-4: ST-C final report structure

2. Statistical analyses and prediction methods: general approaches

2.1 Database structure

The need to define a reference structure for the database is a crucial step. It is also evident from the results of a literature review focused on international research journals dealing with statistical analyses and inverse modeling approaches for prediction of building energy consumption. In order to provide homogeneous information, the analyzed papers have been organized using a specific format described below.

First of all, it is important to remark that the subject on which the analysis will focus has to be clearly defined at the very beginning of the investigation process. Forcing a subdivision into families, the subject of the statistical investigation may be divided as follows:

- Individual building; an analysis focused on one specific single building (or a group of individual buildings);
- Large building stocks: an analysis of a group of statistically representative buildings, typically showing similarities in terms of use (residential, office, school, etc.);
- Regional/national level analyses: typically statistical analyses developed from a database with a large number of buildings on a national basis.

The discussion presented here refers to residential (single and multifamily houses) and office (small and large) buildings, according to the goals of Annex 53; however, the approach can be extended to other building classes.

In general, to perform suitable analyses, the number of buildings and the (minimum) amount of information required to describe each building are related, as shown in Figure 2-.

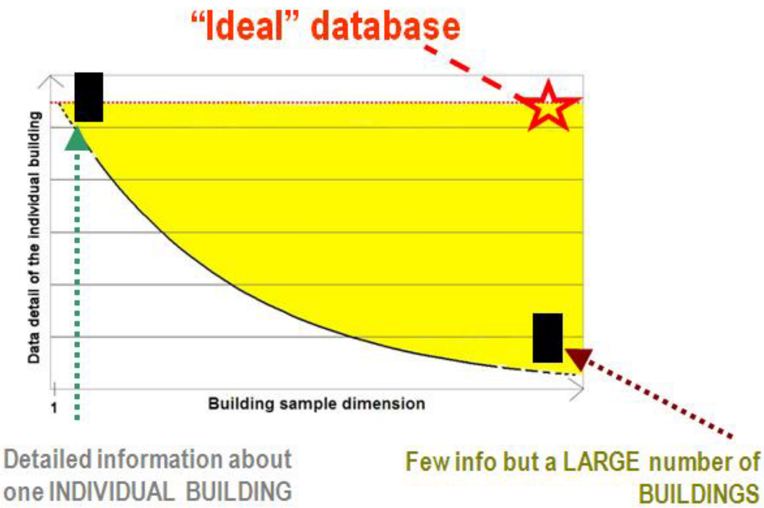


Figure 2-1: Diagram of databases information according to building sample dimension

When a single individual building is the subject of the investigation, a high number of parameters describing its energy behavior is required (one building, but with lots of information) and the analysis can be performed on each specific energy end use. On the contrary, when analyses are performed at a national level, a lot of buildings are available but described by few parameters (lots of buildings, but with less information).

When the subject of the study is chosen, the database may be identified using two main characteristics:

1. Categories of influencing factors which are collected, according to six categories previously defined,
2. Sampling frequency of the time dependent variables (consumptions and parameters belonging to the six categories of influencing factors).

Considering point 1 above, typically the database structure can refer to 3 "levels":

- Level 1 – categories of influencing factors: climate, envelope, systems,
- Level 2 – categories of influencing factors: Level 1 + control & maintenance, indoor environmental conditions,
- Level 3 - categories of influencing factors: Level 2 + occupant behavior.

The three levels may also contain information about the seventh factor (social aspects).
Considering point 2 above, typically the database structure can collect time dependent variables as follows:

- Level 1* – frequency: annual
- Level 2* – frequency: monthly
- Level 3* – frequency: hourly (or sub-hourly)

It should be noted that the time frequency of the collected data is also related to the subject of the study. For investigations at the national/regional level, annual data is typically acceptable, but for advanced analyses of individual buildings, sub-hourly data is required. As a consequence, the databases used in practice can be classified according to their reference structure and placed within a matrix as shown in Table 2-1.

Table 2-1: Database structure according to categories of influencing factors and time frequency of dependent variables.

		Categories of influencing factors		
		Level 1	*Level 2*	*Level 3*
		Climate Envelope systems	Level 1 Control&Maintenance Indoor environmental conditions	Level 2 Occupant behaviour
cy of the time depende nt variable	*Annual*			

	Monthly			
	Hourly (sub-hourly)			

The database structure is highly related to the statistical and prediction methods that can be adopted for the data analyses and elaborations, and consequently with the results obtained through the investigations performed.

2.1.1 Database structure and literature review

The database structure previously introduced is used to define a criterion for the classification of the selected and analyzed papers. In particular, each examined paper is characterized by the following items:

- authors,
- title,
- database typologies (i.e. to which kind of database reference structure it refers),
- adopted method for the data elaboration for energy consumption analysis/prediction,
- subject of the analysis,
- goal of the analysis.

This information is synthesized as presented in Table 2-2 where, for the sake of brevity, only a few of the more than 50 analyzed papers are shown. The work of cataloguing papers is still ongoing and the table is continuously updated. The literature review results, organized using the format of **エラー! 参照元が見つかりません。**, are available in electronic format.

Table 2-2: Organized structure of the literature review activity

Author	**Title**	**Influencing factors categories**	**Adopted method**	**Subject of the analysis**	**Goal of the analysis**
Merih Aydinalp, V. Ismet Ugursal, Alan S. Fung	Modeling of residential energy consumption at the national level	3+	Engineering method/conditional demand analysis method/artificial neutral network	Large building stocks/residential	Comparative assessment of the three methods
H. Farahbakhsh, V. I. Ugursal, A. S. Fung	A residential end-use energy consumption model for Canada	2	Engineering method (CREEM)	Large building stocks/residential	Forecast building energy consumption

Merih Aydinalp-Koksal, V. Ismet Ugursal	Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector	3+	Conditional demand analysis method	Large building stocks/residential	Forecast building energy consumption
Merih Aydinalp, V. Ismet Ugursal b, Alan S. Fung	Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks	3+	Artificial neutral network	Large building stocks/residential	Forecast building energy consumption
Merih Aydinalp, V. Ismet Ugursal, Alan S. Fung	Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks	3+	Artificial neutral network	Large building stocks/residential	Forecast building energy consumption

Note: in "Influencing factors categories", the symbol + means that data referring to "social aspects" was also collected.

2.2 Statistical analyses for description of the examined subject

The objective of the present chapter is to outline the main preliminary analyses used in the preparation phase of examination of the three main subjects of analysis of Annex 53, which are the following:

- Individual buildings,
- Large building stocks,
- Regional/national analyses.

Each of these subjects of study has different goals, ranging from diagnosis of the energy consumption in a specific building to development of building design guides or supporting strategic energy planning policies. Here we do not intend to describe in detail the statistical methods and theory, covered in hundreds of specialized books, but intend to outline a scope of statistical descriptors and analyses appropriate for each of the above mentioned subjects of study.

One of the main aims of this analysis is to summarize the available data sets, to facilitate the comparison between them and to account for variable dependences in the data. The resulting analysis aimed at learning and gaining insight into the populations the data represent is done by inferential statistics and is the subject of the following chapters.

The methods used in the energy consumption analysis for description of the data set include approaches from descriptive statistics and Exploratory Data Analysis (EDA). Descriptive statistics [44, 52] provides summaries about the data samples and about the observations that have been made.

Such summaries may be either quantitative (i.e. summary statistics), or visual (i.e. simple-to-understand graphs). These summaries may either form the basis of the initial description of the data as part of a more extensive statistical analysis, or they may be sufficient in and of themselves for a particular investigation. The Exploratory Data Analysis [53] found in the Engineering Statistics Handbook [21] is an approach for data analysis that employs a variety of techniques, mostly graphical, to provide insight into the data set, uncover underlying structure and distributions, extract important variables and detect anomalies.

The objectives and the statistical analyses for description of each of the subjects of examination are discussed separately in sequence. The list of analyses presented here is by no means exhaustive and is meant for orientation of the analysts. The selection of the most appropriate method depends on the specific problem.

2.2.1 **Individual buildings**

For analysis of individual buildings, normally a large amount of information is collected, comprising detailed building characteristics, climatic data, energy uses and high frequency energy consumption data ranging from several minutes to daily or monthly data intervals. Usually the objective of the study is to analyze the energy consumption, to find the main influence factors in order to alter them and achieve energy savings. Another purpose of the analysis may be to model the building and to compare the expected with the observed behavior for detecting operational faults or predict savings.

Whatever the purpose of the study is, the data set should be reduced to some representative parameters in order to obtain conclusions. Often, building analysis starts with a comparison of relevant energy consumption parameters with reference values (benchmarks) in order rapidly to situate the building in question within the range of consumption of similar buildings and to estimate its energy saving potential. In comparison, when modelling, it is necessary to find relations between variables within the data set and this is first done using exploratory analysis techniques.

When the energy consumption of an individual building is studied, the statistical analysis to describe the data set is expected to provide some of the following: a breakdown of the energy consumption by uses (heating, lightning, equipment, etc.), by energy sources (electricity, gas, etc.), by periods of use (occupied/not occupied, etc.), energy use intensities (EUI) and also other appropriate quantitative parameters that might be considered such as consumption normalized by number of users, volume, etc. The calculation of all these descriptive statistics forms part of the preliminary analysis.

To gain insight into the data set, different graphical statistical exploratory techniques can be used.

- Simple charts to visualize energy breakdowns are pie charts and bar plots.
- Frequency distributions or histograms are often used for estimation of the probability distribution of continuous variables. Frequency distributions can be plotted on an absolute or relative basis for different parameters, (e.g. outside temperature and internal temperature).
- Cumulative frequency distributions can be used to express the probability of some occupant behavior as a function of external influences, for example window opening as a function of outside temperature, or window blinds operation as a function of solar radiation level.

- Pareto charts offer the possibility to represent the factors that contribute most to a given consumption. Identifying these factors will maximize the results.
- Scatter plots can reveal the relation between two parameters in the dataset when there is an interest in analyzing dependencies. They are able to show either linear or nonlinear relationships between variables.

2.2.2 Large building stock

The objective of the analysis of large stocks of buildings is to discover common characteristics of building typologies and the main factors influencing their energy consumption. The available data is normally more reduced and with lower time frequency compared to individual buildings, but is available for a large number of buildings of similar characteristics. The results of the studies are usually used for developing design guides or recommendations and best practices aiming at the reduction of the energy consumption in new or existing buildings. Therefore, descriptive statistics are used to summarize the data set parameters and properties like the range and the distribution within the data set. This permits the identification of the most important variables and members of the set and facilitates the prioritization of the measures.

In the studies of large building stocks the energy consumption's representative parameters may be similar to those in individual buildings, but in this case the quantitative statistics for description are those characterizing the interval and the distribution: the mean, standard deviation, first quartile (Q1), median or second quartile (Q2), third quartile (Q3), as well as minimum and maximum values.

Some of the graphical exploratory and representation techniques usually used are described below.
Box plots, or box-and-whisker plots, depict groups of numerical data using five numerical parameters: smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3) and largest observation (sample maximum). Alternative forms of boxplots can be used for identification of outliers. Variations of boxplots can be found in literature [45, 16]. Histograms and cumulative frequency distributions are used to plot or estimate the probability density of the variables of interest.
A non-exhaustive summary of possible quantitative and graphical analysis techniques for description of data sets is provided in Table 2-3.

2.2.3 Regional/national analyses

The regional/national analysis of building energy consumption has the principle objective of supporting energy planning and strategic energy policies in the mid and long term. In order to perform the analyses, the energy consumption should be structured and studied by building typologies, building use and energy type. Large-scale initiatives (e.g. TABULA - Typology Approach for Building Stock Energy Assessment) aim to develop "building types" at the national level that are representative of a defined period of construction, size, etc., which permit evaluation of the potential impact of energy saving policies and regulations. Data is collected by extensive samples and usually comprises annual or monthly consumption data, climatic data, and complementary information about city/region size, rent per capita, etc.
The statistics and analysis for description of the data sets are summarized in Table 2-3.

Table 2-3: Organized structure of the literature review activity

Statistics and analyses for description of the examined subject	Subjects of study in Annex 53		
	Individual buildings	*Large building stocks*	*National / regional analyses*
Quantitative statistics for description			
Breakdown of energy consumption by uses	X		
Breakdown of energy consumption by uses (mean, median, standard deviation)		X	X
Breakdown of energy consumption by energy types	X		
Breakdown of energy consumption by energy types (mean, median, standard deviation)		X	X
Breakdown of energy consumption by periods of use (day/night, occupied/not occupied, …)	X		
Breakdown of energy consumption by periods of use (day/night, occupied/not occupied, …)(mean, median, standard deviation)		X	X
Energy Use Intensities (EUI) (energy consumption normalized by floor area, number of users, …)	X		
Energy Use Intensities (EUI) (energy consumption normalized by floor area, number of users, …) (mean, median, standard deviation)		X	X
…	…	…	…
Graphical techniques for analysis			
Bar plots, pie-charts	X	X	X
Time series plots	X		
Frequency distribution, histograms	X	X	X
Cumulative frequency distributions	X	X	X
Scatter plots	X	X	X
Pareto charts	X	X	X
…	…	…	…

2.2.4 General overview

The studies previously presented offer a general overview of the applications of statistical analysis at different scales of investigation, with different goals, ranging from diagnosis of the energy consumption in a specific building to developing of building design guides or supporting strategic energy planning policies. The time step of the collected data, useful for energy consumption predictions, is also related to the subject of the study: for investigations addressed to national/regional level annual or monthly data are typically acceptable, but for the analysis of an individual building hourly to monthly intervals are typically required.

It is therefore clear that the time step of the data available for the analysis is a function of scale of the investigation and therefore of the goal of the analysis. As a consequence, different influencing factors can be highlighted as relevant for each scale of investigation, as well as the most suitable model type to be used for the energy use prediction.

The table below (Table 2-4) summarizes the main influencing factors and the most suitable models used by different contributions presented here which were identified, based on the time step of the available data and according to the goal of the analysis identified by the scale of the investigation.

Table 2-4: Overview of main influencing factors and most suitable model used in the contributions presented

Scale of investigation	Timestep	Main influencing factor	Most suitable model
Individual buildings	Hourly energy consumption Daily energy consumption Monthly energy consumption	Building geometry Building physical characteristics Climate (indoor/outdoor) Occupancy (n° of users) Users lyfestyle	Regression analysis: Linear multivariate logistic partial least square q-q plot principal components Artificial neural network Ensembling methods
Large buildings	Monthly energy consumption Annual energy consumption	Building geometry Building physical characteristics Climate (outdoor)/location Occupancy (n° of users) Users lyfestyle Purpose of the use Heating/Cooling operation	Regression analysis: Linear multivariate logistic partial least square multiple Neural network Quantification methods
National Buildings	Monthly energy consumption Annual energy consumption	Building typology Building physical characteristics Location (Degree Days) Period of construction Heating/Cooling operation Purpose of the use	Frequency distribution Cluster analysis Hierarchical cluster techniques Monthly regression models Engineering models

3. Statistical analyses for the determination of relevant influencing factors

In the previous section, the database structure for building energy analysis is introduced. Regardless of the database level (Level 1, 2 or 3) and data frequency in the database, the goal is to develop models and define relationships among variables. During model development, variables are separated into two groups: predictor (input) and response (target or output) variables.
The influencing factors of building energy use are grouped in six main categories. Relating to the above mentioned variable classification it means that the influencing factors are predictor (input) variables, while the building energy use is a response (target or output) variable.

Sensitivity analysis can be used to find the relevant influencing factors. In fact, sensitivity analysis is the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input. Sensitivity analysis provides information such as factors that mostly contribute to the output variables; the region in the space of input factors for which the model output is either maximum or minimum or within pre-defined bounds, etc. In this way, sensitivity analysis can be very useful for the purpose of determining the influencing factors of building energy use. Further, in order to avoid overwhelming the model, sensitivity analysis can help to simplify models and find the most important input factors.
There are several possible procedures to perform uncertainty and sensitivity analysis. These procedures can be any of the following:

- Local methods, such as the simple derivative of the output with respect to an input factor,
- A sampling based sensitivity where the model is executed repeatedly for combinations of values sampled from the distribution (assumed known) of the input factors,
- Methods based on emulators (e.g. Bayesian) where the value of the output, or directly the value of the sensitivity measure of an input factor, is treated as a stochastic process,
- Screening methods where the objective is to estimate a few active factors in models with many factors (one of the most commonly used screening methods is the elementary effect method),
- Methods based on Monte Carlo filtering.

Since Subtask C in Annex 53 is dealing with data organized into databases, a sampling based sensitivity analysis is the most relevant method for defining influencing factors of building energy use. In this case, data from the databases are samples for analysis. The starting point of statistical sensitivity analysis is the generation of input-output scatter plots, which are obtained by plotting the points. An example of using the scatter plot in sensitivity analysis is shown in Figure 3-6 (from Corgnati et al. 2008 [11]) and is used here for illustration purposes only.

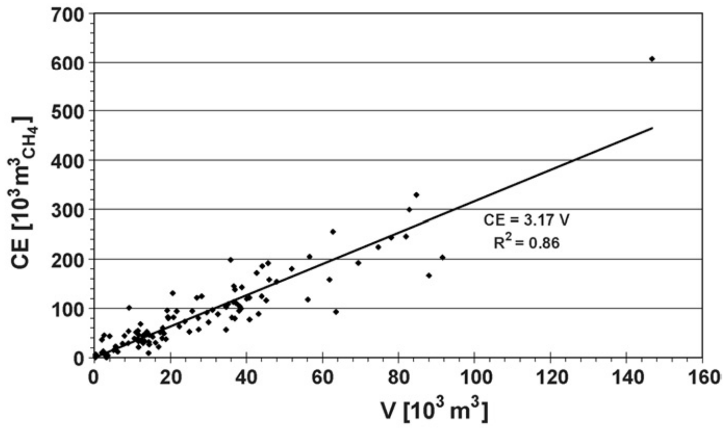


Figure 3-1: Sensitivity analysis by using the scatter plot

Finally, as shown in Figure 3- the sensitivity analysis is performed by repeating the regression analysis on data in the different scatter plots. Regression analysis assesses the importance of the input variable with respect to the uncertainty in the output component.

To perform sensitivity and regression analysis, several relations are necessary. The following relevant equations for sensitivity and regression analysis are written based on [9]. Available data in a database can be analyzed in the following way. For example, an observed database can consist of i samples. These samples are organized in the matrix as $[X_i, Y_i]$ where $i = 1, 2, \ldots, n$, X_i is the matrix of the input variables, and Y_i is the matrix of the output variables. The matrix of the input variables consists of i variables. The starting point of statistical sensitivity analysis is the generation of input-output scatter plots, which are obtained by plotting the points as shown in Figure 3-. The resulting scatter plots are then examined to find possible relations between the outputs Y_i and the inputs X_i. A more formal analysis of the input-output relationship is to perform regression analysis on a linear model between the predicted output, $Y_{predicted}$, and the input parameters X_j , of the form

$$Y_{predicted} = b_0 + \sum_{j=1}^{I} b_j \cdot X_j \quad (1)$$

In this case, sensitivity analysis implies searching for input-output relationships by using regression analysis. The regression coefficients can be used, along with other indicators computed during the regression analysis, to assess the importance of the individual input variables X_i with respect to the uncertainty in the output components Y_i. The higher the absolute value of the regression coefficient, the higher the influence on the output. Further, the calculated output Y_i in terms of the actual parameter values X_i will have the following linear form:

$$Y_i = b_0 + \sum_{j=1}^{I} b_j \cdot X_{ij} + \varepsilon_i \qquad (2)$$

where ε_i denotes the error between the calculated and predicted value of the corresponding element of the output. In order to get the best fit of the regression model, it is necessary that the sum of the squares of the deviation (shown in Equation 3) is minimized.

$$min\sum_{i=1}^{n}\left(Y_i - Y_{predicted}\right)^2 \qquad (3)$$

Finally, to gage the goodness-of-fit of the model, the coefficient of determination is used. The goodness-of-fit of the model can be calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - Y_{predicted}\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \qquad (4)$$

where $\bar{Y}_i$ is the average value of the output variables. A value of 1 indicates a perfect correlation between actual data and the regression equation; a value of 0 indicates no correlation. For the purpose of the building energy analysis, such as modeling using the utility bill data, as a rule of thumb the value of R^2 should never be less than 70 [14].

4. Selection of the "model type" for energy use prediction and identification of the prediction model parameters

4.1 Introduction

The classic approach to evaluate the building-HVAC system energy use is based on the application of a thermal model with known system structure and proprieties as well as forcing variables (forward approach). This model can be more or less complex depending on the requested result accuracy and output time step. For a tailored analysis the forward approach requires a detailed knowledge of the physical phenomena (and their relative magnitude and interactions) affecting the system behavior, and the building system operating mode. ESP-r, BLAST, DOE-2, TRNSYS and Energy Plus are the most widespread simulation codes based on forward simulation models. The application of the "forward approach" is widely discussed in ST-D of Annex 53.

A different approach for building energy analysis is based on the so called inverse or data-driven models. In this case, the input (regressor variables) and output variables (response) are known and measured and the objective is to estimate the system parameters and to describe the mathematical model. Using a data driven approach it is possible to evaluate the as-built system performance (the model parameters are calculated on the actual building energy use) allowing often a more accurate prediction of the energy consumption tendencies with respect to the forward approach.

The definition of the intended purpose of the building energy analysis is the fundamental step for the selection of the appropriate model approach. The approach must be able to match the analysis requirements with sufficient accuracy. The requirements of building energy analysis may include design optimization, energy audit, energy certification and so on. As mentioned, the different methods can be grouped into two main families, according to the goal of the analysis.

Forward approach: it is the classical presentation of any physical phenomena: it starts with the definition of the energy model. Then the collection of input variables and finally the simulation run to evaluate the output.

Data driven approach: it may be described as a bottom-up approach as it starts with the measurement of the force driven variables and of the output variables, followed by the evaluation of some building features called "system parameters" and the construction of the data driven model that will be used to assess the output for another set of force driven variables.

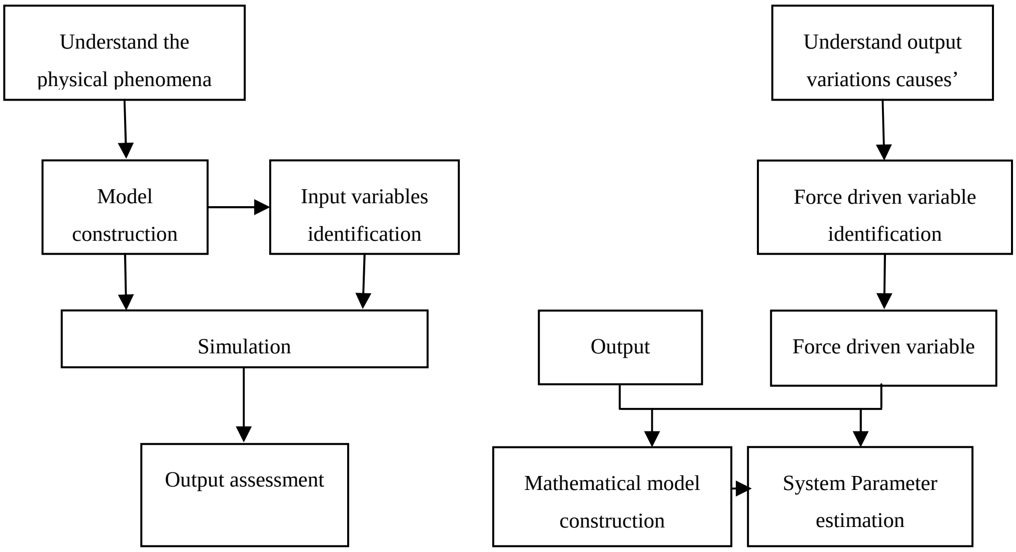


Figure 4-1: Comparison between forward approach and data driven approach

By a data driven approach (inverse modeling), an empirical analysis is carried out on the building energy behavior, and its relationship to one or more driving forces or parameters (regressor variables). This approach is referred to as a system identification and parameter identification. To develop an inverse model, it is necessary to carry out a mathematical description of the building or system, and then identify the parameters of interest using statistical analyses (estimations). The input and output variables are known and measured, and the goal is to determine a mathematical description of the relationship between the independent variables and the dependent one. In contrast to the forward approach, the data-driven approach is useful when the “system” has been built (that is, the “system” exists and works) and actual performance data are available for model development and/or identification. The model parameters are evaluated from actual building performance and working conditions, so the data driven model is fine for the evaluation of as-built system performance, allowing more accurate prediction of future system performance under specific, real boundary conditions.

The data driven modeling, based on the application of statistical tools, is adequate, among the others, for evaluating demand-side management (DSM) programs, to identify/test simple and conventional energy conservation measures in existing buildings and for baseline model development in energy conservation measurement and verification (M&V) projects.

For example, with a data driven approach, it is possible to evaluate the causes of discrepancies of actual consumption compared with design predictions and find the causes (such as anomalous weather conditions, unintended building operation, improper operation), or to verify energy savings due to a retrofit action and not to other causes (e.g., the weather).

Data-driven methods for energy-use evaluation in buildings can be classified into three categories (ASHRAE [14]):

- empirical or “black-box” approach,
- calibrated simulation approach,
- “gray-box” approach.

In the **empirical approach**, the most accepted techniques use linear or change-point linear regression to correlate energy use or peak demand as the dependent variable (output) with weather data and/or other independent variables (input). A simple, or multivariate regression model, is operated between measured energy use and the various influential parameters (e.g., climatic variables, building operation). This approach can be used with any time scale (monthly, daily, hourly sub-hourly). Single-variate, multivariate, change point, Fourier series and the artificial neural network (ANN) belong to this category. Least-squares regression is the most common regression technique to determine the coefficients of the model. Such a purely statistical approach is appropriate to evaluate demand side management programs, identify energy conservation measures in an existing building and to develop a baseline model in energy conservation measurement and verification projects while their value is limited for diagnosis and online control.

The **calibrated simulation approach** uses an existing building simulation computer program and calibrates the physical inputs to the program so that measured energy use matches closely with that predicted by the simulation program. In this way, when the subject of the study is an individual building, consistent predictions can be obtained. The calibration process (tuning) can be conducted with monthly data or data that span a few weeks over the year, but the level of accuracy decreases with the decrease of data time frequency and increase in time interval. Several difficulties prevent the use of simulation calibrated models to describe the real performance of the building: 1) arrangement of weather data used by simulation programs, 2) the methods selected to calibrate the model and 3) the selected methods to measure the input parameters required for simulation (building mass, infiltration coefficients, etc.). However the calibrated simulation approach requires a high level of user skill and knowledge in both, simulation and practical building operation, a high degree of expertise, a very large number of input parameters, and enormous amounts of computing time, as well as financial resources.

The **gray-box approach** first formulates a physical model to represent the structure or physical configuration of the building or energy system, and then identifies the representative parameters and aggregated physical parameters and characteristics by statistical analysis (Rabl and Riahle 1992).
Moreover, two primary types of inverse models are classified in the literature: steady state inverse models and dynamic inverse models. The criterion on which the classification is based is that dynamic inverse models contain time-lagged variables.

All three approaches previously described can be implemented through steady state and dynamic models. Typical single and multiple linear regressions fall under the “Black Box” steady-state models.
A model is dynamic when dependent or independent variables are explicitly expressed as functions of time. Dynamic inverse models include equivalent thermal network analysis, ARMA models, Fourier series models, machine learning, and artificial neural networks. The dynamic models are capable of taking into account dynamic effects such as thermal mass which traditionally have required the solution of a set of differential equations. The disadvantages of dynamic inverse models are that they are increasingly complex with respect to steady state models and need more detailed measurements to "tune" the model.

Table 4-1, proposed by ASHRAE, presents information that is useful for selecting an inverse model where as a function of the model (diagnostics - D, energy savings calculations - ES, design - De, and control - C), degree of difficulty in understanding and applying the model, time scale for the data used by the model (hourly - H, daily - D, monthly - M, and sub-hourly - S), calculation time, input variables used by the models (temperature - T, humidity - H, solar - S, wind - W, time - t, thermal mass - tm), and accuracy.

Table 4-1: Capabilities of different forward and data-driven modeling methods (by ASHRAE,10)

Methods	**Use[a]**	**Difficulty**	**Time Scale[b]**	**Calc. Time**	**Variables[c]**	**Accuracy**
Simple linear regression	ES	Simple	D, M	Very fast	*T*	Low
Multiple linear regression	D, ES	Simple	D, M	Fast	*T, H, S, W, t*	Medium
ASHRAE bin method and data-driven bin method	ES	Moderate	H	Fast	*T*	Medium
Change-point models	D, ES	Simple	H, D, M	Fast	*T*	Medium
ASHRAE TC 4.7 modified bin method	ES, DE	Moderate	H	Medium	*T, S, tm*	Medium
Artificial neural networks	D, ES, C	Complex	S, H	Fast	*T, H, S, W, t, tm*	High
Thermal network	D, ES, C	Complex	S, H	Fast	*T, S, tm*	High
Fourier series analysis	D, ES, C	Moderate	S, H	Medium	*T, H, S, W, t, tm*	High
ARMA model	D, ES, C	Moderate	S, H	Medium	*T, H, S, W, t, tm*	High
Modal analysis	D, ES, C	Complex	S, H	Medium	*T, H, S, W, t, tm*	High
Differential equation	D, ES, C	Complex	S, H	Fast	*T, H, S, W, t, tm*	High
Computer simulation (component-based)	D, ES, C, DE	Very complex	S, H	Slow	*T, H, S, W, t, tm*	Medium
(fixed schematic)	D, ES, DE	Very complex	H	Slow	*T, H, S, W, t, tm*	Medium
Computer emulation	D, C	Very complex	S, H	Very slow	*T, H, S, W, t, tm*	High

Notes:
[a]Use shown includes diagnostics (D), energy savings calculations (ES), design (DE), and control (C).
[b]Time scales shown are hourly (*H*), daily (*D*), monthly (*M*), and subhourly (*S*).
[c]Variables include temperature (*T*), humidity (*H*), solar (*S*), wind (*W*), time (*t*), and thermal mass (*tm*).

In Table 4-2, the methods for analyzing building energy use are classified as either forward or data-driven, and either steady-state or dynamic.

Table 4-2: Classification of analysis methods for building energy use (by ASHRAE, [14])

Data-Driven

Method	Forward	Black-Box	Simulation	Gray-Box	Comments
	Empirical or Calibrated Physical or				
		Steady-State Methods			
Simple linear regression (Kissock et al. 2002; Ruch and Claridge 1991)	—	X	—	—	One dependent parameter, one independent parameter. May have slope and y-intercept.
Multiple linear regression (Dhar 1995; Dhar et al. 1998, 1999a, 1999b; Katipamula et al. 1998; Sonderegger 1998)	—	X	—	—	One dependent parameter, multiple independent parameters.
Modified degree-day method	X	—	—	—	Based on fixed reference temperature of 18.3°C.
Variable-base degree-day method, or 3-P change point models (Fels 1986; Reddy et al. 1997; Sonderegger 1998)	X	X	—	X	Variable base reference temperatures.
Change-point models: 4-P, 5-P (Fels 1986; Kissock et al. 1992)	—	X	—	X	Uses daily or monthly utility billing data and average period temperatures.
ASHRAE bin method and data-driven bin method (Thamilseran and Haberl 1995)	X	X	—	—	Hours in temperature bin times load for that bin.
ASHRAE TC 4.7 modified bin method (Knebel 1983)	X	—	—	—	Modified bin method with cooling load factors.
Multistep parameter identification (Reddy et al. 1999)	—	—	—	X	Uses daily data to determine overall heat loss and ventilation of large buildings.
		Dynamic methods			
Thermal network (Rabl 1988; Reddy 1989; Sonderegger 1977)	X	—	—	X	Uses equivalent thermal parameters (data-driven mode).
Response factors (Kusuda 1969; Mitalas 1968; Mitalas and Stephenson 1967; Stephenson and Mitalas 1967)	X	—	—	—	Tabulated or as used in simulation programs.
Fourier analysis (Shurcliff 1984; Subbarao 1988)	X	—	X	X	Frequency domain analysis convertible to time domain.
ARMA model (Rabl 1988; Reddy 1989; Subbarao 1986)	—	—	—	X	Autoregressive moving average (ARMA) model.
PSTAR (Subbarao 1988)	X	—	X	X	Combination of ARMA and Fourier series; includes loads in time domain.
Modal analysis (Bacot et al. 1984; Rabl 1988)	X	—	—	X	Building described by diagonalized differential equation using nodes.
Differential equation (Rabl 1988)	—	—	—	X	Analytical linear differential equation.
Computer simulation: DOE-2, BLAST,	X	—	X	—	Hourly and subhourly simulation

EnergPlus (Crawley et al. 2001; Haberl and Bou-Saada 1998; Manke et al. 1996; Norford et al. 1994)					programs with system models.
Computer emulation (HVACSIM+, TRNSYS) (Clark 1985; Klein et al. 1994)	X	—	—	—	Subhourly simulation programs.
Artificial neural networks (Kreider and Haberl 1994; Kreider and Wang 1991)	—	X	—	—	Connectionist models.

In general, the construction of data driven models is based on the least-squares regression. The considered model uses the least-squares regression to determine the regression coefficients. Generalized least-squares regression seeks to estimate the model coefficients that minimize the sum of the squared error between predicted and actual observations. The matrix of dependent observations, Y, is equal to the product of the matrix of independent observations, X, and the matrix of estimated regression coefficients, β, plus an error term, E:.

$$Y = X\beta + E \qquad (5)$$

Solving for β gives:

$$\beta = (X^T X)^{-1} X^T Y \qquad (6)$$

To calculate the model residuals, the predicted values of the dependent variable, $\hat{Y}$, are computed from:

$$\hat{Y} = X\beta \qquad (7)$$

The matrix of residuals, E, is then computed from:

$$E = Y - \hat{Y} \qquad (8)$$

The root mean squared error, RMSE, is computed from:

$$\text{RMSE} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{(n-p)}} = \sqrt{\frac{Y^T Y - \beta^T X^T Y}{(n-p)}} \qquad (9)$$

where n is the number of data observations and p is the number of regression coefficients.
The root mean squared error of the model is a measure of the scatter of the data around the model.
The matrix of the standard errors of the regression coefficients, S, is computed from:

$$S = RMSE\sqrt{(X^T X)^{-1}} \qquad (10)$$

The standard error of a regression coefficient is a measure of the uncertainty of the estimate of the regression coefficient.

The squared correlation coefficient, R2, is computed from:

$$R^2 = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2} \quad (11)$$

4.2 Steady-state inverse models

The simplest steady-state inverse model regresses monthly actual energy consumption data against average outdoor air temperatures, or in general outdoor climate parameters. More robust and efficient methods include multiple linear regression (MVR), change-point linear regression, and Variable-Based Degree Day regressions. The advantage of steady-state inverse models is that their use can be automated and applied to large data sets when, for example, billed consumption data and average daily temperatures are available.

Moreover, by the multivariate approach it is possible to characterize building energy use with a few available input variables. The model should contain variables not affected by the retrofit and likely to change from pre-retrofit to post-retrofit periods. Other variables, such as changes in operating hours, occupancy levels, should be included in the model if these are not energy conservation measures (ECMs) but variables that may change during the post-retrofit period.

Proper care must be taken, however, when using MVR models to predict energy consumption. In general, the addition of independent variables to the model will always increase the strength of the correlation; however, the relative uncertainty (standard error) of each regression coefficient, and hence its predictive value, will decrease. In addition, multi-colinearity between independent variables increases the uncertainty with which the values of the regression coefficients are known. Singular Value Decomposition [4] and Principle Component Analysis [23; 50] have been shown to reduce the effects of multi-colinearity.

Claridge [58] summarized the most common methods for developing inverse models of measured energy use. The primary methods include variable-base degree-day (VBDD) models, multivariate regression (MVR) models, change-point (CP) regression models, and combination CP/VBDD/MVR regression models.

In the following, on the basis of a literature review, some numerical algorithms and equations used to find general least-squares regression for the models mentioned above are described.

In Leslie et al. [43] the results of an investigation are presented to determine which factors related to climate, occupant productivity and time-related parameters exert significant influence on energy consumption. The regression model shows that energy consumption in general depends on heating degree-days, production level, and labor force strength. Data gathered included production level by product class, heating degree-days, cooling degree-days, energy consumed by fuel type, labor force, direct and indirect man hours, etc. The best predictors among competing parameters were selected based on maximizing the adjusted multiple correlation coefficient. In general, heating degree-days and cooling degree-days are the most important parameter for predicting total energy consumption, with labor force strength and production level providing additional explanatory power.
Katipamula et al. [36] found that a multi linear regression provides better accuracy than a single variable model for modeling energy consumption. Many independent variables have been used to perform a MLR model including, cooling-degree days, heating-degree days, wind speed and direction, humidity, refrigeration type, exhaust air, supply air, average shading in winter, average shading in summer and so on. Different buildings used different independent variables, some up to ten and others as few as two. Nevertheless, MLR models based on engineering principles are difficult to develop

because they require knowledge of the HVAC system operation and how it is related to other building systems. Another disadvantage of MLR is the variables should be independent of each other, which is not the case in reality.

In Freire et al. [15] regression equations are used for predicting energy consumption by means of outdoor climate variables and HVAC systems properties in an easier and more rapid way than building energy simulation tools. The independent variables (input data) are heating, ventilation and air conditioning (HVAC) power, outdoor temperature, relative humidity and total solar radiation. The methodology for obtaining the regression equations is based, firstly, on defining a couple of linear Multiple-Input/Single-Output (MISO) models, since two main outputs are involved, (i.e. indoor temperature and relative humidity). Validation procedures have shown very good agreement between the regression equations and the simulation tool for both winter and summer periods.

Abushakra et al. [1] showed the advantage of including four driving variables in the hourly modeling of the energy use which include (1) outdoor temperature, (2) outdoor specific humidity potential, (3) lighting and receptacles, and (4) occupancy. This paper showed that the occupancy variable can be derived from the lighting and receptacles load profiles that are becoming more and more available.

In Catalina et al. [10] a development of regression models is presented to predict the monthly heating demand for the single-family residential sector in temperate climates, with the aim to be used by architects or design engineers as support tools in the very first stage of their projects in finding energy efficient solutions. All the energy prediction models were based on an extended database obtained by dynamic simulations for 16 major cities of France. The inputs for the regression models are the building shape factor, the building envelope U-value, the window to floor area ratio, the building time constant and the climate which is defined as function of the sol-air temperature and heating set-point.

About the assessment of energy savings, the most straightforward way to measure energy savings is to compare pre and post-retrofit energy use. This method implicitly assumes that the change in energy consumption between the pre-retrofit and post-retrofit periods is caused purely by the retrofit. However, the energy consumption is also influenced by other factors including weather conditions, occupancy, internal loads and building operating procedures which may change between the pre and post-retrofit periods. If these changes are not considered, savings determined by this simple method will be erroneous [33]. The most common adjustment discussed in the literature [37] is for changing weather conditions between the baseline and post-retrofit periods through the use of the data driven approach. In general, two types of measured savings, actual and normalized, can be determined. Actual savings [22; 12; 2] are calculated as the difference between the energy use predicted by the baseline model and measured post-retrofit energy use. The steps involved are:

- measure energy use and influential variables during the baseline period;
- create a mathematical model of baseline energy use as a function of influential variables;
- measure energy use and influential variables during the post-retrofit period;
- apply influential variables from the post-retrofit period to the baseline model to estimate what energy use would have been without the retrofit;
- subtract the predicted baseline energy use from the measured post-retrofit energy use to estimate savings.

Kissock [40] describes a method for calculating savings from measured data using change-point models for weather adjustment. It includes the physical basis for change-point models in commercial buildings, algorithms for change-point models, and a method to estimate the uncertainty of savings. Normalized savings [50] estimate how much energy would be saved during a ‘normalized’ year. Calculating normalized savings requires developing a statistical model of energy use as a function of influential variables for both the pre and post-retrofit periods and then driving each model with "normal" conditions to calculate the normalized annual consumption during each period.

In Haberl et al. [20] measured hourly data are used to construct a baseline model. The data can then be used to predict building consumption had the retrofit not taken place. Measured post-retrofit data are compared to predicted data to determine savings. Regression models consist of billing and/or monitored data, utilized in one-, two-, three-, four-, or five-parameter change-point models, or MLR models.

Kissock et al. [41] describe a procedure for estimating weather-adjusted retrofit savings using ambient-temperature regression models. The appropriate use of both linear and change-point models for measuring energy savings is also discussed. Ambient-Temperature is used as the single independent variable because it both eliminates problems associated with multi-colinearity problems and reduces data collection to a single easily acquired parameter.

4.3 Focuses

4.3.1 Multiple linear regression

4.3.2 Logistic regression analysis

While linear regression analyses the effect of one or more influencing variables, X_j, on a continuous outcome variable, logistic regression analysis needs to be used when the outcome variable, Y, is binary, e.g. either 0 or 1. The difference between linear and logistic regression models is thereby, that while the linear model describes the changes in the outcome variable directly, the logistic model describes the probability, p=P(Y=1), of the outcome variable being one of the two possible values. As a consequence, even though the outcome variable is binary, the probability, p, can take all values between 0 and 1.

In general, a linear relationship between the odds p/(1-p) and the logarithm of the odds, called logit, log[p/(1-p)] is assumed, i.e.

$$\text{Logit}(p) = \log[p/(1-p)] = \alpha+\beta X, \quad (12)$$

which is mathematically equivalent with

$$p = \frac{\exp(\alpha+\beta X)}{1+\exp(\alpha+\beta X)} \text{ [6]} \quad (13)$$

There are two possible applications to the statistical analysis of the total energy use: (1) as input model for occupant behavior in the forward approach, or (2) to predict the probability of binary "energetical events". The former is described within the modeling report of the Task-Force Occupant Behavior of Annex 53 and not dealt with here. Applications related to energetical events can be addressed to the analysis

- of the probability of a building occupant to belong to one of two groups, an energy-saving or energy-wasting group [46, 42],
- of factors which influence the probability of an energy usage above or below a certain threshold level, e.g. the median of a number of energy usages, or electricity consumption classes [17] or
- of the probability of a change in the energy usage, e.g. due to retrofitting measures.

Except for a few applications as listed above, the approach is not very common and linear regression analysis is much more widespread. The reason lies probably in the continuous nature of energy usage, which does not necessitate a logistic approach.

4.3.3 **Variable-Base Heating and Cooling Degree-Day Models**

In the Variable-Base Heating and Cooling Degree-Day the algorithm finds the base-temperature or balance point temperature that gives the best statistical fit between energy consumption and the number of variable-base degree-days in each energy use period.

The heating balance-point temperature is defined as the temperature at which the heat gain from internal occupants and equipment balances heat loss through the building envelope. At outdoor air temperatures above the balance-point temperature, no thermal energy is needed for space heating. Similarly, cooling energy use frequently increases as outdoor air temperature increases above some cooling balance-point temperature, below which no space cooling is necessary.

In general, the closer the outdoor air temperature is to T_{bal}, the greater is the uncertainty. The degree-day method, like any steady-state method, is defective for estimating consumption during mild weather. In fact, consumption becomes most sensitive to occupant behavior and cannot be predicted with certainty.

About variable-base heating and cooling degree-day method, IMT can find best-fit variable-base degree-day models of type:

$$Y = \beta_1 + \beta_2\, HDD(\beta_3) \quad (14)$$

$$Y = \beta_1 + \beta_2\, CDD(\beta_3) \quad (15)$$

where β_1 is the constant term, β_2 is the slope term, and $HDD(\beta_3)$ and $CDD(\beta_3)$ are the number of heating and cooling degree-days, respectively, in each energy data period calculated with base temperature β_3. The number of heating and cooling degree-days in each energy data period of n days is:

$$HDD(\beta_3) = \sum_{i=1}^{n} (\beta 3 - T_i)^+ \qquad (16)$$

$$CDD(\beta_3) = \sum_{i=1}^{n} (T_i - \beta 3)^+ \qquad (17)$$

where T_i is the average daily temperature.

4.3.4 **Change-Point Models**

There are several types of regression change point models as a function of the type of HVAC-building system analyzed:

(i) Two-Parameter Model (2-P), based on a simple linear regression of type:

$$Y = \beta_1 + \beta_2 X_1 \quad (18)$$

where β_1 and β_2 are regression coefficients, X_1 is the independent variable and Y is the dependent variable. 2-P models are appropriate for modeling building energy use that varies linearly with outdoor air temperature.

(ii) Three-Parameter Cooling and Heating Models (3-P), is characterized by a three-parameter change-point models of the type described by Kissock et al. [38]:

$$Y_c = \beta_1 + \beta_2 (X_1 - \beta_3)^+ \quad \text{and} \quad Y_h = \beta_1 + \beta_2 (X_1 - \beta_3)^- \qquad (19)$$

Where β_1 is the constant term, β_2 is the slope term, and β_3 is the change-point. The ( ) + and ( )- notations indicate that the values of the parenthetic term shall be set to zero when they are negative and positive, respectively. 3P models are appropriate for modeling building energy use that varies linearly with an independent variable over part of the range of the independent variable and remains constant over the other part. For example, 3PC models, using outside air temperature as the independent variable, are often appropriate for modeling the whole-building electricity use in residential electric air conditioning. Similarly, 3PH models, using outside air temperature as the independent variable, are often appropriate for modeling heating energy use in residences with gas or oil heating.

Also found is a combination of three-parameter and multi-variable regression models (3P-MVR), with up to four independent variables, of the type:

$$Y_c = \beta_1 + \beta_2 (X_1 - \beta_3)^+ + \beta_4 X_2 + \beta_5 X_3 + \beta_6 X_4 \qquad (20)$$
$$Y_h = \beta_1 + \beta_2 (X_1 - \beta_3)^- + \beta_4 X_2 + \beta_5 X_3 + \beta_6 X_4 \qquad (21)$$

where X_1 is typically the external temperature, and X_2, X_3 and X_4 are optional independent variables.

(iii) Four-Parameter Model (4P), is of the type described by Kissock et al. [39]:

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^- + \beta_3 (X_1 - \beta_4)^+ \qquad (22)$$

Where β_1 is the constant term, β_2 is the left slope, β_3 is the right slope and β_4 is the change point. An inverse model can also consider a combination of four-parameter multi-variable regression models (4P-MVR), with up to three independent variables, of the type:

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^- + \beta_3 (X_1 - \beta_4)^+ + \beta_5 X_2 + \beta_6 X_3 \qquad (23)$$

where X_1 is typically temperature, and X_2 and X_3 are optional independent variables.
Four-parameter models are appropriate for modeling heating and cooling energy use in variable-air-volume systems and/or in buildings with high latent loads. In addition, these models are sometimes appropriate for describing non-linear heating and cooling consumption associated with hot-deck reset schedules and economizer cycles [40].

(iv) Five-Parameter Model (5-P), described by:

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)- + \beta_3 (X_1 - \beta_5)+ \qquad (24)$$

Where β_1 is the constant term, β_2 is the left slope, β_3 is the right slope, β_4 is the left change point, and β_5 is the right change point. An inverse model can also consider a combination of five-parameter multi-variable regression models (5P-MVR), with up to two independent variables, of the type:

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^{-} + \beta_3 (X_1 - \beta_5)^{+} + \beta_6 X_2 \qquad (25)$$

where X_1 is typically temperature and X_2 is an optional independent variable. (5-P) models are appropriate for modelling energy consumption data that include both heating and cooling, such as whole-building electricity data from buildings with electric heat-pumps or both electric chillers and electric resistance heating or fan electricity consumption in variable-air-volume systems. In Table 4-3, the different model equations of the single-variate approach are described and they are also drawn in Figure 4-2.

Table 4-3: Single-variate and change point models for the heating mode (ASHRAE, [14])

Model Type	**Independent Variable(s)**	**Model equations**	**Examples**
One-parameter or constant (1-P)	None	$E=\beta_1$	Non-weather-sensitive demand
Two-parameter (2-P)	Temperature	$E=\beta_1+\beta_2(T)$	
Three-parameter (3-P)	Degree-days/ Temperature	$E=\beta_1+\beta_2(DD_{BT})$ $E=\beta_1+\beta_2(\beta_3-T)^+$ $E=\beta_1+\beta_2(T-\beta_3)^+$	Seasonal weather-sensitive use (fuel in winter, electricity in summer for cooling)
Four-parameter change point (4-P)	Temperature	$E=\beta_1+\beta_2(\beta_4-T)^+-\beta_3(T-\beta_4)^+$ $E=\beta_1+\beta_2(\beta_4-T)^++\beta_3(T-\beta_4)^+$	Energy use in commercial buildings
Five-parameter (5-P)	Degree days/ Monthly mean temperature	$E=\beta_1-\beta_2(DD_{TH})+\beta_3(DD_{TC})$ $E=\beta_1+\beta_2(\beta_4-T)^++\beta_3(T-\beta_5)^+$	Heating and cooling supplied by same meter

Note: DD denotes degree-days and T is monthly mean daily outdoor dry-bulb temperature.

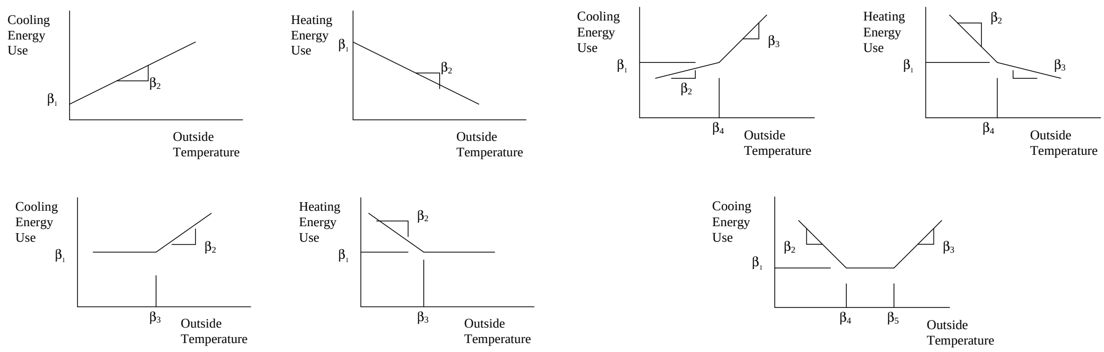


Figure 4-2: Top row-left: 2P cooling and heating models. Second row from top-left: 3P cooling and heating models. Top row-right: 4P cooling and heating models. Second row from top-right: 5P heating and cooling model.

4.4 Dynamic Inverse Models

Dynamic inverse models represent sophisticated forms of inverse models. The various existing methodologies include, among the others:

- ARMA models,
- Fourier series,
- Artificial Neural Networks – ANNs.

These models take into account the dynamic effects using the time variation of the parameters as a key aspect for understanding the energy behavior of the analyzed system. These models are used to process inter-correlated forcing functions or independent parameters.

Dynamic inverse models are capable of modeling complex systems which are dependent on more than one independent parameter, on the other hand they require more measurements to develop the model. Among the previous list of dynamic inverse models, ANNs represent the most used tool for their accuracy for modeling and forecasting and for their automated implementation in commercial software. As a consequence, a larger discussion is addressed to ANNs in this report.

4.5 **Focuses**

4.5.1 **Artificial neural network**

The ANNs are inductive models that represent an alternative approach with respect to the deductive models. In building energy modeling, the ANNs are used as surrogate of analytic computer codes to evaluate the energy flow and system performance, i.e. they are useful for forecasting and modeling.
The ANNs learn from key information patterns allowing for the discovery of complex relationships between the variables. The ANNs allow robust processing even from noisy data. On the other hand they provide a limited knowledge of process mechanisms.

It is well-known from literature that one of the most interesting features of neural models is their ability to handle even incomplete data. Several studies have shown that in some cases forecasting models for energy consumption based on neural networks are more accurate, even if more complex than those based on multiple linear regression. As briefly summarized below, ANNs take inspiration from neural systems.

Biological and artificial neurons and artificial neural network principles
Figure 4-3 shows a simplified model of the structure of the biological neuron. The main body of the cell is called the "soma" where the nucleus is placed. The cell body has fibers attached to it called dendrites that receive signals from other neurons. 'Axon' is the single long fiber which spreads from the soma and extends into fibers connecting to many other neurons at the synaptic junction. Each neuron receives stimuli from other neighboring neurons and produces output when inputs overcome the threshold limit that a neuron can hold.

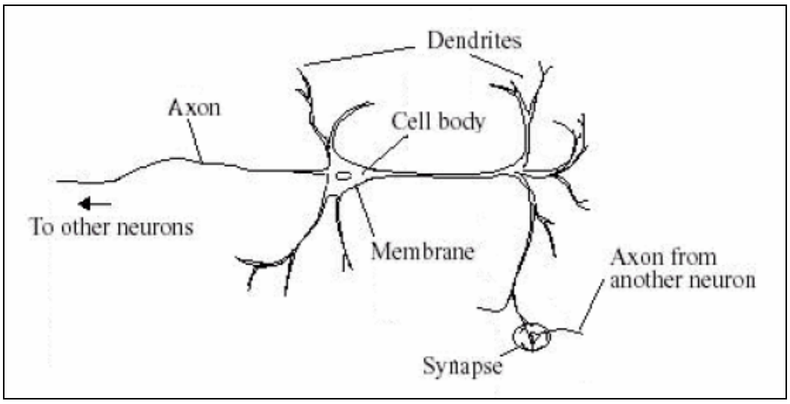


Figure 4-3: Simplified model of the structure of the biological neuron

An artificial neural network is a massively parallel distributed processor that has a large number of artificial neurons interconnected through weighted synaptic connections. Connections can be "adjusted" through a network training process based on a given pattern rather than on predefined rules. In other words, this process allows the network to learn the “rule” on which is based a physical phenomenon starting from known situations and apply it to new situations. This feature and the relative simplicity of implementation and programming encourages the application in prediction tasks. In addition, the use of a nonlinear model allows the identification of interactions between independent variables without exploiting complex models. A neural network architecture commonly used is the Multi Layer Perceptrons. Its basic structure consists of a set of units organized in layers; each element

produces its output applying an activation function to a weighted linear combination of input signals. The weights of this linear combination are those associated with connections that affect the neuron. The activation function, f_{ACT}, determines a relationship between the activation of the neuron and its output.
The equation for a single layer with one neuron can be written in the form:

$$y = f_{ACT}\left(\sum_{j=1}^{n} w_j X_j\right) \quad (26)$$

From the functional point of view the above equation describes the following model:

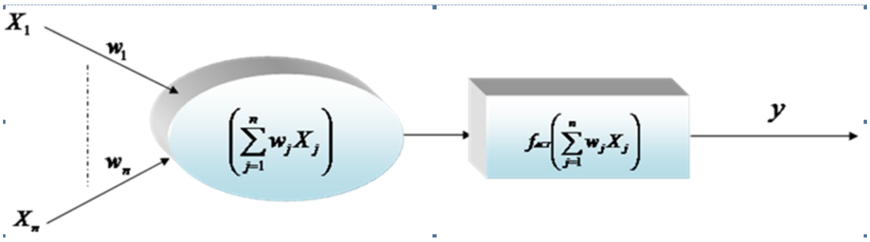


Figure 4-4: Activation function applied to summed data

In the following we describe the most used feed-forward architecture. The feed-forward neural network presents the neurons arranged in layers. All neurons in a layer are connected to all neurons in the next layers through uni-directional weighted links. The neurons of the input layer do not perform computation, but only feed input information to the neurons of the first hidden layer. The last layer represents the output from which the response of the network comes. The neuron output o_j^S is computed by passing the weighted sum (w_{ji}^S are the weights) of its inputs o_j^{S-1} by a activation function $f_{ACT,j}^S$

$$o_j^S = f_{ACT,j}^S \left[\sum_{i=1}^{n_{S-1}} w_{ji}^S \cdot o_i^{S-1}\right] \quad (27)$$

where *s* denotes the s^{th} layer and $j=1,\ldots,n_s$ with n_s the number of neurons of the s^{th} layer. The weights are obtained using the “forward propagation” and “backward propagation” algorithms.
ANNs have been used successfully in the following fields:

- classification, i.e. pattern recognition,
- forecasting, i.e. electrical and thermal load forecasting,
- optimal control, i.e. adaptive control,
- optimization, i.e. building energy management.

With reference to the application of neural networks to energy systems, ANNs have been employed to model solar water heating systems. Kalogirou et al. [24] developed an ANN to forecast useful energy produced from the system and the stored water temperature rise. The input data related to size and performance characteristics were: the collector area, storage-tank U-value, tank type, storage volume and type of system. In addition, the input data related to the weather conditions were real measures of total daily solar radiation, mean ambient air temperature and the water temperature in the storage tank at the beginning of a day. The whole data set was used to train the ANN in order to treat unusual cases. The predictions were confined within 10% and it was shown that the proposed method presents a high grade of accuracy.

In Kalogirou et al.[25], the long term performance prediction of solar domestic water heating systems was also evaluated using ANNs. The authors tested and modeled thirty thermosyphon solar domestic water heating systems following the procedures contained in the standard ISO 9459-2 at three locations in Greece. Monthly data, calculated thorough a modeling program based on standard ISO 9459-2, were used to develop two artificial neural networks.

The output of the first network was the solar energy produced from the system under suitable physical constrains (see Kalogirou et al. [26] for details) and the output of the second network was the solar energy produced from the system and the average quantity of hot water per month at demand temperatures of 35 and 40°C. The input data in both networks were geometric and performance characteristics of each system and various climatic data. In the second network, among the other input was also used the demand temperature.

The statistical coefficient of multiple determinations corresponding to the first network was equal to 0.9993 while for the second network for the two output parameters the same coefficient was 0.9848 and 0.9926, respectively. Also, the accuracy of prediction was investigated using unknown data. In the first case the coefficient was equal to 0.9913 and in the second case was equal to 0.9733 and 0.9940 for the two output parameters.

With reference to the application of neural networks to individual buildings, in Datta et al. [57] the authors proposed the use of an ANN to forecast electricity demand in a supermarket. In a supermarket the main demand categories are refrigeration systems, HVAC equipment and lighting. In addition the independent variables that affect the consumption of the refrigeration systems are building envelope; temperature and humidity and the internal environmental conditions.

The authors claim that in trying to minimize energy consumption the various energy consuming subsystems cannot be viewed in isolation but their interactions should be considered.
For this reason, the authors suggest that ANNs are flexible tools and not system specific. As a consequence ANNs can be adapted to different building types, HVAC systems and refrigeration equipment.

The Figure 4-5 below shows the prediction accuracy of the ANNs modeling. The building is a Safeway supermarket situated in Airdrie, Scotland. The ANN was used to predict electricity demand in the supermarket. The actual measured data collected from the store was used to develop feed-forward neural networks including three layers: one input, one hidden, and one output layer. Seven networks were constructed by varying the number of input variables, i.e. input nodes, n. The number of nodes of the hidden layer was varied as a function of the input nodes as (2n + 1). The backpropagation algorithm was employed to train all the networks in order to minimize the mean square error between the output of the network and the actual value.

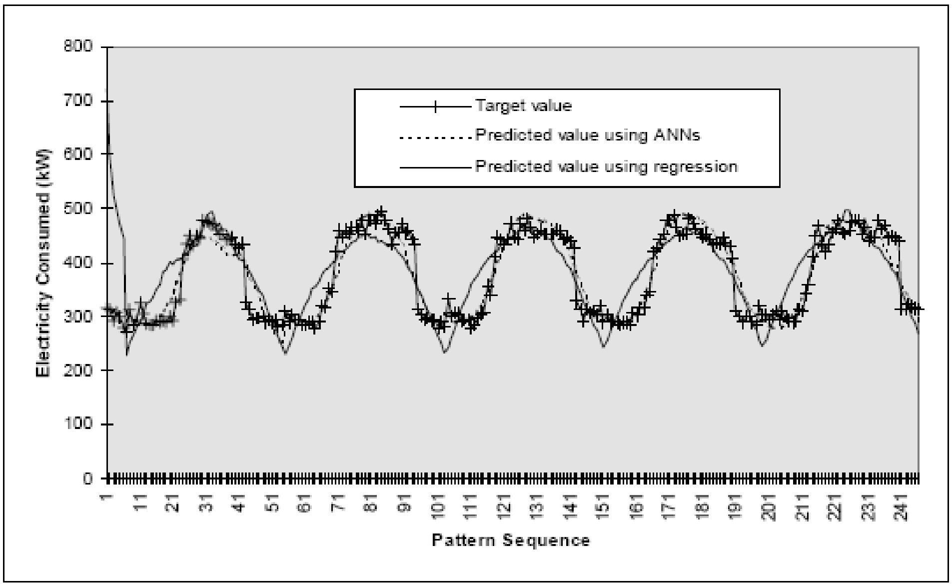


Figure 4-5: Prediction accuracy of the ANNs modeling

The following input variables were included:

- Day,
- Time,
- Temperature and relative humidity in the store,
- External air temperature and humidity,
- Total electrical power consumption of the store,
- Electrical power consumption of the refrigeration packs,
- Gas Consumption,
- Underfloor heating flow and return temperatures.

The authors noted that all the networks have a good correlation between the target and network output. The network trained with data extending on a longer period, i.e. 4 months instead of the 1 month used for the other networks, presented the lowest correlation coefficient and highest RMS error. This confirms that using a short-term data-set may be sufficient to reliably predict electrical demand in commercial stores on a half hour time scale.

The authors estimated the weight of each independent variable on the prediction accuracy of the dependent variable. They noted the time of day is the most significant independent variable. This result led to the conclusion that more detailed analysis should be conducted to determine the relative importance of each variable.

The prediction performance of ANNs and regression analysis based on the same data sets were compared using the correlation coefficient. The correlation coefficient was equal to 0.95 for the ANNs as compared to 0.79 for regression analysis.

In Kalogirou et al. [27], ANNs were used for forecasting the building's heating loads using the minimum amount of input data listed below. The data used were:

- areas of windows, walls, partitions and floors,
- type of windows and walls,
- classification on whether the space has roof or ceiling,
- design room temperature.

The network output was the heating load. More than two hundred cases were considered to develop a network capable of working with only some basic building areas and a differentiation of the various elements according to their structure. Indicative U values were used to describe different building materials. The network yielded predictions within 9%. The authors concluded that this error did not influence the sizing of the actual radiator chosen for the particular room because commercial radiators are available only with different heating loads (see Kalogirou et al. [29, 28, 30, 31, 32, 33, 34, 35] for more details).

With reference to the application of neural networks to a large building stock, Aydinalp-Koksal et al. [2] investigated the use of a CDA method and of an ANN to model the residential end-use energy consumption. In particular, the ability of the CDA model to predict and to characterize the energy end-uses was compared with those of an ANN.

Two data sources were used for the development of the CDA and ANN model: the data from the 1993 Survey of Households Energy Use (SHEU, Statistics, Canada, 1993) database and the 1993 heating and cooling degree day data for the cities in which the households in the CDA data set are located. Actually, SHEU database is the most exhaustive energy related database for the Canadian residential sector. The data were collected by conducting a mail-out survey that included several questions (376). The database was representative of the Canadian housing stock, and contains detailed information on the building construction, space heating/cooling and DHW heating equipment, household appliance and some socioeconomic characteristics of the occupants for 8767 households in Canada. At the same time, the electricity billing data and natural gas billing data of the households in the 1993 SHEU database were used to develop the models. The weather and ground-temperature data for the locations were obtained from Environment Canada (http://www.msc-smc.ec.gc.ca/cmc/index_e.html). The variables used in this study are presented below.

- Heating degree days,
- Cooling degree days,
- Ground temperature,
- Dwelling type,
- Heated living area,
- Dwelling year construction,
- Windows type,

- Door types,
- Presence of heated basements, attic or garage,
- Presence of programmable thermostat,
- Presence of heat recovery ventilation system,
- Efficiency of the boiler,
- Age of the boiler,
- Number of each appliance present in the house,
- Total number of incandescent, fluorescent, and halogen lamps,
- Central A/C unit usage,
- Window A/C unit usage,
- Average indoor temperature,
- Number of occupants,
- Presence at home,
- Dwelling ownership,
- Number of adults,
- Number of children.

The information is related to climate (HDD, CDD, local mean daily temperature, etc.), building envelope (heated living area, dwelling construction year, number of windows, etc.), building equipment (thermostat, boiler, ventilation systems, appliances, etc.), building operation and maintenance (age of the boiler, central A/C unit usage, etc.), indoor environmental quality (average indoor temperature, etc.), occupant behavior (number of occupants, presence at home, etc.) and socio-economic factors (households income, size of area of residence, etc.).

The ANN developed to model the end-use energy consumption consists of three separate models:

- NN model for appliance, lighting and space cooling (ALC) end-use energy consumption,
- NN model for space heating (SH) end-use energy consumption,
- NN model for domestic hot water (DHW) heating end-use energy consumption.

On the basis of the data gathered from different sources, the input and output of the model were defined. The input units of the networks used were gathered from the SHEU database in order to describe the construction details and usage characteristics of the houses; specifications and usage of space heating and cooling equipment appliances and lighting; socioeconomic characteristics of the occupants; and weather characteristics. Obviously, the number and choice of input units was different for each of the three networks performed, and the units were selected on the basis of their contribution on the prediction performance of the end-use network. The results of the three performed networks showed a good prediction accuracy with a very high prediction performance, but the accuracy was strictly related to the quantity of the information in the training datasets. The models were performed by isolating the effects of several socioeconomic factors on end-use energy consumption. This capability represents an interesting result, because of the impact of human behavior on the building energy consumption.

The authors highlighted that both methods could be used to model residential energy consumption, but each of them had different capabilities, advantages and disadvantages. The major advantages of the

CDA model are that it is easier to develop and to use and does not require detailed data. On the other hand, it is a regression-based model. For this reason, the database has to contain a large number of dwellings and the models do not provide much detail or flexibility. By consequence, its capability to assess the impact of energy conservation possibilities is very limited. An important point to highlight in this comparison is related to socioeconomic factors. Although it is possible to include socioeconomic parameters in the model (if such data is available in the database), the CDA model is unable to evaluate the effects of some of these parameters (dwelling ownership and size of area of residence) because of the limited number of variables included in the model, due to statistical considerations. On the other hand, ANNs are able to evaluate the effects of several socio-economic factors on end-use energy consumption, such as household income, dwelling type and ownership, number of children and adults, and area of residence.

4.5.2 **Data Mining**

Data mining is proposed as a tool to analyze measured building-related data. Data mining techniques excel at automatically analyzing huge amounts of data for useful but hidden information.

What is data mining?

In the past decade, different definitions of data mining have been given by various researchers. For example, Hand et al. [23] define data mining as "the analysis of large observational data sets to find unsuspected relationships and to summarize the data in novel ways so that data owners can fully understand and make use of the data." As defined by Cabena et al. [8], data mining is "an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large databases." Based on these statements, it can be concluded that data mining is essentially a combination of multi-disciplinary approaches. It is often used to extract hidden but useful patterns from a large volume of data and to transform the data into knowledge that could benefit further work. Data mining has been successfully applied in many scientific, medical, and application domains (e.g., banking, bioinformatics, new materials identification, fraud detection, and telecommunications). It was also identified by the MIT Technology Review (MIT Technology Review, 2001) as one of the ten emerging technologies that may change the world. Three widely accepted and implemented data mining techniques are: data classification, clustering analysis, and association rule mining.

Some basic terms and concepts in relation to data mining

Useful terminologies include the following:

- Dataset, Attribute, and Instance: a dataset is a set of data items. It is roughly equivalent to a two-dimensional (i.e. column and row) spreadsheet or database table, as shown in Figure 4-6. Each database table consists of a set of attributes (usually in different columns or fields) and stores a large set of instances (usually in rows or records). Consider an HVAC system with 100 monitored parameters. Each parameter can be considered an attribute, and a record of all these parameters in a specific time point can be considered an instance.

Attribute Instance	Attribute 1	...	Attribute m
Instance 1	x	...	x
…	...	...	...
Instance i	x	...	x
...	...	...	...
Instance j	x	...	x
...	...	...	...
Instance n	x	...	x

Figure 4-6: A schematic diagram of dataset, attribute and instance

- Target attribute and Predictor attribute: Target attribute is the attribute predicted as a function of other attributes (i.e., predictor attributes). For example, the building energy consumption is the target attribute, and could be predicted as a function of building-related parameters such as floor area and number of occupants (i.e., predictor attributes).

Data mining techniques: Data Classification and Decision Tree

Overview of Decision Tree

The decision tree method is one of the most commonly used data mining methods [47, 22]. It uses a flowchart-like tree structure to segregate a set of data into various predefined classes, thereby providing the description, categorization, and generalization of given datasets. As a logical model, a decision tree shows how the value of a target variable can be predicted by using the values of a set of predictor variables. Figure 4-7 presents a decision tree indicating whether residents turn room air conditioners (RAC) on or off in their rooms in the cooling season. For this example, assume 100 instances are used to build this decision tree, and that each instance has three attributes: outdoor air temperature, room occupancy, and the operating state of RAC.

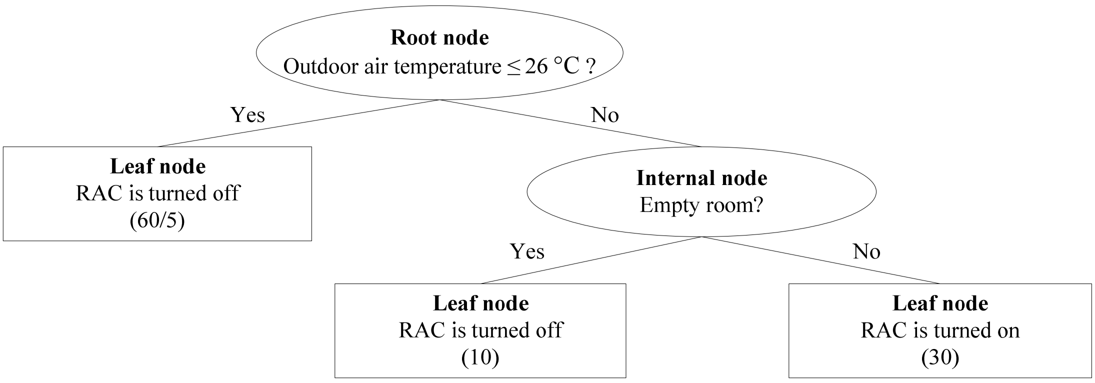


Figure 4-7: Schematic illustration of a simple hypothetical decision tree

The target variable for the above decision tree is RAC operating states, with potential states being classified as either turned on or off. The predictor variables are outdoor air temperature (≤ 26 °C or > 26°C) and room occupancy (empty or not). As shown in Figure 4-7, a decision tree consists of three kinds of nodes: root node, internal node, and leaf node. Root nodes and internal nodes denote a binary split test on an attribute while leaf nodes represent an outcome of the classification (i.e., a categorical target label). Moreover, the numbers in the parentheses at the end of each leaf node depict the number of instances in this leaf. If some leaf nodes are impure (i.e., some records are misclassified into this node), the number of misclassified instances will be given after a slash. For example, (60/5) in the left most leaf in Figure 4-7 means that among the 60 instances having an outdoor temperature lower than or equal to 26 °C that have been classified as turned off, 5 of them actually have the value turned on. By using this decision tree, the RAC operating state classification (i.e., turn on or turn off) can be predicted. For example, if the outdoor air temperature is higher than 26 °C and the room is not empty, occupants will turn the RAC on; otherwise, they will turn it off.

Decision tree generation is in general a two-step process, namely learning and classification, as shown in Figure 4-8. In the learning process, the collected data is split into two subsets, a training set and a testing set. Creation of the training and testing sets is an important part of evaluating data mining models. Usually, most of the instances in the database are arbitrarily selected for training and the remaining instances are used for testing. Note that the training and testing sets should come from the same population but should be disjointed. Then, a decision tree generation algorithm takes the training data as an input, with the corresponding output being a decision tree. Commonly used decision tree generation algorithms include ID3 [47], classification and regression trees (CART) [7], and C4.5 [48]. In the classification process, the accuracy of the obtained decision tree is first evaluated by making predictions against test data. The accuracy of a decision tree is measured by comparing the predicted target values with the true target values of the test data. If the accuracy is considered acceptable, the decision tree can be applied to a new dataset for classification and prediction; otherwise, the reason for any inaccuracies should be identified and corresponding solutions should be adopted to address these problems.

Decision Tree Generation

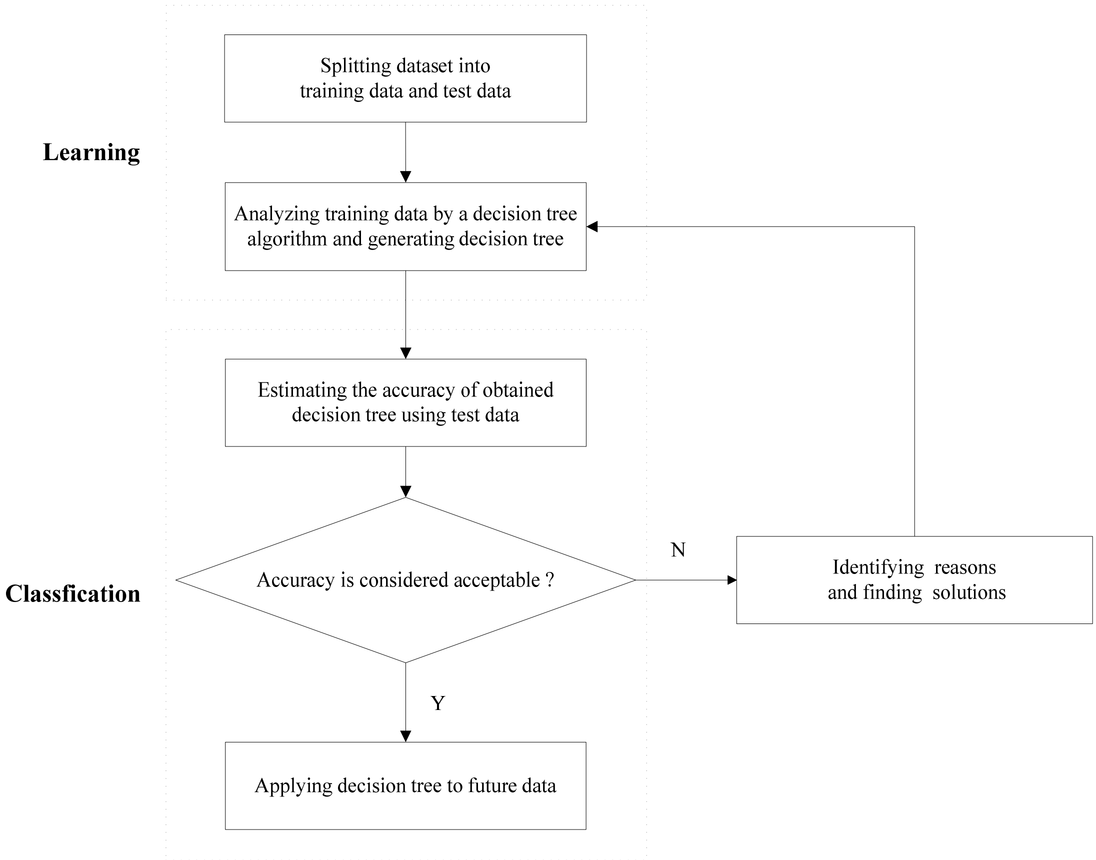


Figure 4-8: Procedure of decision tree generation

The procedure for generating a decision tree from the training data is explained as follows. Initially, all instances in the training data are grouped together into a single partition. At each iteration the algorithm chooses a predictor attribute that can "best" separate the target class values in the partition. The ability of a predictor attribute to separate the target class values is measured based on an attribute selection criterion, which can be referred to in data mining textbooks. After a predictor attribute is chosen, the algorithm splits the partition into child partitions such that each child partition contains the same value of the chosen selected attribute. The decision tree algorithm iteratively splits a partition and stops when any one of the following terminating conditions is met.

(1) All instances in a partition share the same target class value. Thus, the class label of the leaf node is the target class value.
(2) There are no more instances for a particular value of a predictor variable. In this case, a lead node is created with the majority class value in the parent partition.

With reference to the application of data classification to a large building stock, Yu et al. [54] developed a building energy demand predictive model based on the decision tree method. The model can estimate residential building energy performance indexes by modeling building energy use intensity (EUI) levels (either high or low). Its competitive advantage over other widely used modeling techniques, such as regression methods and ANN methods lies in the ability of the method to generate accurate predictive models with interpretable flowchart-like tree structures that enable users to quickly extract useful information.

One data source was used for the development of the decision-tree based model: a project entitled "Investigation on Energy Consumption of Residents All over Japan" which was carried out by the Architecture Institute of Japan from December 2002 to November 2004. For this project, field surveys on energy-related data and other relevant information were carried out in 80 residential buildings located in six different districts in Japan: Hokkaido, Tohoku, Hokuriku, Kanto, Kansai, and Kyushu. The following information was collected:

- Energy end use of all kinds of fuel used by the building at different time intervals;
- Indoor environment parameters every 15 minutes;
- Household characteristics;
- Other issues such as occupant behaviors and energy saving measures;

The following input variables were included:

- Annual average air temperature,
- House type (detached or apartment),
- Construction type (wood or non-wood),
- Floor area,
- Heat loss coefficient,
- Equivalent leakage area,
- Number of occupants,
- Space heating mode (electric or non-electric),
- Hot water supply mode (electric or non-electric),
- Kitchen equipment mode (electric or non-electric).

The output variables were building energy performance indexes (EUI levels, either high or low).
The model accuracy of predicting the EUI levels is 92%. For comparison, prediction models using regression methods and ANN methods were also developed based on the same data set. The accuracy of the obtained regression model and ANN model were 72% and 88%, respectively. However, it

should be mentioned that the decision-tree model can only predict the EUI levels while the regression model and ANN model can predict the EUI values. Moreover, a lot of useful information on building energy performance improvement can be extracted from the developed model. For example, it can automatically identify and rank significant influencing factors of building EUI. Also, the model can provide the combination of significant factors as well as the threshold values that will lead to high building energy performance. Based on such information, designers can clearly realize which parameter deserves extra attention when designing energy efficient buildings. Another advantage is that it can be utilized by users without requiring a lot of computation knowledge. The generated model, and the derived information, could greatly benefit building owners and designers; one crucial benefit is the reduction of building energy consumption.

Cluster Analysis and the K-means Algorithm

Cluster analysis is the process of grouping the observations into classes or clusters so that objects in the same cluster have a high similarity, while objects in different clusters have a low similarity. Figure 4-9 shows a clustering schema based on a hypothetical building data table. It contains various energy-related variables such as outdoor air temperature (T) and building heat loss coefficient (HLC).

The data table consists of m attributes and n instances. Each attribute represents a variable and each instance denotes a building. All the instances are grouped into w clusters. These w clusters are homogeneous internally and heterogeneous between different clusters [22]. Such internal cohesion and external separation are based upon the m attributes; it implies that these attributes have the most similar holistic effects on the building energy performance of the same cluster buildings, while the effects are significantly distinct for the buildings in different clusters.

	Instance	Attribute 1 (T)	...	Attribute m (HLC)
Cluster 1	Instance 1	x	x	x
	…	x	x	x
	Instance i	x	x	x
⋮	...	x	x	x
Cluster w	Instance j	x	x	x
	...	x	x	x
	Instance n	x	x	x

Figure 4-9: Clustering scheme

The dissimilarity between observations in the database is calculated using the distance between them in the cluster analysis. In this study, the most commonly used distance measure, Euclidean distance, was used [22]:

$$d(k,l) = \sqrt{(x_{k1} - x_{l1})^2 + (x_{k2} - x_{l2})^2 + \cdots + (x_{kn} - x_{ln})^2} \quad (28)$$

where k = (xk1, xk2, …, xkn) and l = (xl1, xl2, …, xln) are buildings. xk1, …, xkn are n parameters of k and xl1, …, xln are n parameters of l.

Commonly used clustering algorithms include K-means, K-medoids, and CLARANS [22]. In this study, we employed the K-means, along with the open-source data mining program RapidMiner [49], to perform cluster analysis due to its high efficiency and wide applicability.

The K-means algorithm is one of the simplest partition methods to solve clustering problems. Given a dataset (D) containing w objects, the K-means algorithm aims to partition these w objects into k clusters with two restraints: 1) the center of each cluster is the mean position of all objects in that cluster, 2) each object has been assigned to the cluster with the closest center. This algorithm consists of given steps: 1) Randomly select k observations from D as the initial cluster centers, 2) Calculate the distance between each remaining observation and each initially chosen center, 3) Assign each remaining observation to the cluster with the closest center, 4) Recalculate the mean values (i.e., the cluster centers) of the new clusters, and 5) Repeat Steps 2 to 4 until the algorithm converges, meaning that the cluster centers do not change.

In RapidMiner, the performance of a clustering algorithm is evaluated by the Davies Bouldin index (DBI) [13]. This index is defined as the ratio of the sum of average distance inside clusters to distance between clusters.

$$DBI = \frac{1}{n}\sum_{i=1}^{n} max_{i \neq j}\left[\frac{R_i + R_j}{M_{i,j}}\right] \qquad (29)$$

where n is the number of clusters, R_i and R_j are the average distance inside cluster i and cluster j by averaging the distance between each cluster object and the cluster center, respectively, and $M_{i,j}$ is the distance between cluster centers. The DBI is small if each cluster is comparatively dense while different clusters are far from each other. Consequently, a smaller DBI indicates better performance.
It should be mentioned that the K-means is sensitive to initial cluster centers. Therefore, different values should be tried so as to obtain the minimum sum of the distances within a cluster. At the same time, the number of clusters should be specified in advance.

With reference to the application of data clustering to a large building stock, Yu et al. [55] developed a methodology for examining the influences of occupant behavior on building energy consumption.
Various factors influence building energy consumption at the same time, leading to a lack of precision when identifying the individual effects of occupant behavior. Such effects can be shown by ‘removing’ the effects of influencing factors unrelated to occupant behavior.

The same data source as the decision-tree based model was used for the development of the clustering methodology. The methodology is realized by clustering similar buildings into various groups based on the influencing factors unrelated to occupant behavior, so that for each building in the same group these factors have similar effects on building energy consumption. Accordingly, the effects of occupant behavior can be identified accurately in these groups. The identification of building groups is the most important element of this methodology and it is achieved mainly via cluster analysis.
The following input variables were included.

- Annual mean air temperature
- Annual mean relative humidity

- Annual mean wind speed
- Annual mean global solar radiation
- House types (detached or apartment)
- Building area
- Equivalent leakage area
- Heat loss coefficient
- Number of occupants
- Space heating mode (electric or non-electric)
- Hot water supply mode (electric or non-electric)
- Kitchen equipment mode (electric or non-electric)

The output was the effects of occupant behavior, as well as behavior patterns. Particularly, the following data analysis was conducted.

- The analysis of the average annual EUI of different end-use loads for each cluster (this mainly indicates the degree to which various behavior influence the total building energy consumption).
- The analysis of the variability in annual EUI of different end-use loads for each cluster (A large variability implies that there still remains great potential for energy saving by improving occupant behavior related to the end-load uses).
- The analysis of monthly variations of average end-use loads for each cluster (this mainly indicates the effects of occupant behavior over both time and buildings).
- A reference building for each cluster is defined, and then the energy-saving potential of buildings in each cluster can be evaluated by comparison with the reference building.
- The analysis of monthly average indoor temperature of an air-conditioned room of three typical buildings.

Association Rule Mining

In data mining, association rules are often used to represent patterns of parameters that are frequently associated together. An example is given to illustrate the concept of association rules. Assume that 100 occupants live in 100 different rooms in the same building and each room has both a window and a door. Moreover, 40 occupants open the windows and 20 occupants open the doors. If 10 occupants open both the windows and doors simultaneously, it can be calculated that these 10 occupants account for 10% of all the building occupants (10/100 = 10%), and 25% of the occupants who open windows (10/40 = 25%). Then, the information that occupants who open windows also tend to open doors at the same time can be represented in the following association rule:

$$\text{open\_windows} \rightarrow \text{open\_doors}\ [support = 10\%, confidence = 25\%] \tag{30}$$

In this statement, support and confidence are employed to indicate the validity and certainty of this association rule. Different users or domain experts can set different thresholds for support and confidence according to their own requirements, in order to discover useful knowledge eventually. Accordingly, the association rule mining (ARM) can be defined as finding out association rules that satisfy the predefined minimum support and confidence from a given database.

Mathematically, support and confidence can be calculated by probability, P(X Y), and conditional probability, P(Y|X), respectively (X denotes the premise and Y denotes the consequence in the sequence). That is,

$$support(X \rightarrow Y) = P(X \cup Y) \quad (31)$$
$$confident(X \rightarrow Y) = P(Y|X) \quad (32)$$

Another concept, lift, which is similar to confidence, is commonly used to demonstrate the correlation between the occurrence of X and Y when conducting the ARM. Mathematically,

$$lift(X \rightarrow Y) = \frac{P(X \cup Y)}{P(X)P(Y)} = \frac{P(Y|X)}{P(Y)} \quad (33)$$

Particularly, a lift value greater than 1 represents a positive correlation (the higher this value is, the more likely that X coexists with Y, and there is a certain relationship between X and Y [22] while a lift value less than 1 represents a negative correlation). If the value is equal to 1, i.e. , the occurrence of X is independent of the occurrence of Y, then there is no correlation between X and Y. Commonly used ARM algorithms include the Apriori algorithm and the frequent-pattern growth (FP-growth) algorithm [22]. The specific algorithm for these methods are presented in [22].

With reference to the application of association rule mining to individual buildings, Yu et al. [56] developed a methodology for examining all associations and correlations among building operational data, thereby discovering useful knowledge about energy conservation.

One data source was used for the development of the methodology: the EV pavilion in Montreal, a complex building that mainly includes offices and wet labs,.. This building consists of two parts: the ENCS part (17 floors) and the VA part (12 floors). Both parts have their own VAV air-conditioning systems. The historical data of the air-conditioning systems in both parts were collected from December 2006 to May 2009. In total, 61 parameters of HVAC system operation were monitored at a 15-minute interval.

The input variables were the above-mentioned 61 parameters. The output was all associations and correlations among these parameters. Through analyzing these associations and correlations, we can:

- identify the energy waste in the air-conditioning system (e.g., it was found that, in the fresh air handling units, the heat added to the fresh air was first transferred to humidifier water, and then simply drained to municipal sewage. This energy waste was confirmed through the discussion with the building operator),
- detect the equipment faults (e.g., it was found that, either the fan 1 or the fan 2, or both of them, in a fresh air handling unit has a fault),
- propose low/no cost strategies for saving energy in system operation (e.g., it was found that, the existing operating strategy of extracting exhaust air from the building was to use two of three fans while the other one was turned off. Given that these three fans are identical and controlled by individual VSD, one possible energy-saving method is to use all these three fans instead of two of them).

The results obtained could help us better understand building operation and provide opportunities for energy conservation.

5. Application of the prediction model for building energy use assessment

5.1 General introduction to case studies

Thanks to the contributions of the partners participating in ST-C, a classification of the developed activities about the assessment of Total Energy Use in Buildings by "inverse (data-driven) methods" has been performed. Common questions are faced in the contributions related to the most appropriate statistical analysis method according to the database level, the fixed goals of the analysis and the dominant influencing factors. This section highlights the relationships among:

- Subject of the analysis,
- Goal of the analysis,
- Structure of the database,
- Adopted method of analysis.

The ongoing activities in Sub–Task C could be divided, according to sub-task structure, with reference to the subject of the analysis:

- Large Building Stocks,
- Individual Buildings,
- National or Regional level.

Annex 53 partners uniformly faced the two topics, focusing only in one topic or on both, for a total of 17 contributions. In particular, Austria, France, Germany, and Norway focused on the analysis of individual buildings, while Italy and Japan focused on both individual buildings and a large building stock, and Canada and Spain contributed for a large building stock. The national or regional analysis issue was faced by Italy (regional database), China (national database) and the U.S. (national database) as well. As defined in the Annex 53 project, the ST-C analysis of the partners should focus on residential or office buildings. In the following table (Table 5-1), the building typologies included in the analysis of the provided contributions are delineated.

Table 5-1: Distribution by subject of the analysis of the obtained contributions

Partner	**Individual buildings**		
	Residential	*Office*	*Other*
CETHIL, INSA de Lyon (France)			Synthetic 1 school
Karlsruhe Institute of Technology (Germany)	Synthetic and Extended 1 Multi- family house	Synthetic and Extended 1 Multi-storey office	
NTNU Trondheim (Norway)		Synthetic and Extended 1 Office building	
Polytechnic of Turin (Italy)		Synthetic 1 Office building Extended 1 Office building	

Tohoku University (Japan)	Synthetic and Extended 6 houses (single and multi-family)		
TU Wien (Austria)	Synthetic 3 Multi- family houses 8 Single- family houses	Synthetic 2 Office buildings	
Partner	**Large building stock**		
	Residential	*Office*	*Other*
CIMNE (Spain)		Synthetic and Extended 9 office buildings	
Concordia University (Canada)	Synthetic and Extended 4 contributions 80 houses (single and multi-family)		
Polytechnic of Turin (Italy)		Synthetic and Extended 4000 office buildings	
Tohoku University (Japan)	Synthetic and Extended 682 houses 80 houses (single and multi-family)	Synthetic and Extended 1121 office buildings	
Tohoku University (China houses)	Synthetic and Extended 635 houses		
Partner	**National/Regional level**		
	Residential	*Office*	*Other*
Tsinghua University (China)		Extended 4600 office buildings	
LBNL (U.S.)		Synthetic and Extended 824000 offices (CBECS database)	
Polytechnic of Turin (Italy)	Synthetic and Extended 66000 houses (Piedmont regional database)		

The goal of the analysis of the provided contributions can be synthetically divided by:

- description of subject (statistical characterization of the subject, benchmarking, etc.),
- prediction (forecasting) of the energy consumption of the subject.

Within the individual buildings contributions, a large part of the work is dedicated to the statistical characterization of the subject. The subject description through statistics is delineated with different aims: Norway uses statistical analysis for the identification of driving variables that contributed to energy use, while the analysis from Austria and Germany is particularly related to the topic of determining an accurate profile of user behavior (both in offices and in residential buildings) to represent the energy related behavior of the occupants. Even if the characterization of the occupant behavior assumed in the research takes different paths, it highlights the increasing importance of the topic.

The theme of prevision (forecasting) of the energy consumption is carried forward by both the Japanese and French groups in individual buildings. In particular, the main goal of the French forecasting analysis is heating load as a function of the outdoor temperature. In the Japanese analysis of 6 detached houses, the focus was concentrated on the prediction of the energy supply and demand in residential areas.

Finally there is a third case, where the first statistical analysis used to characterize the subject is further used to calibrate the model and forecast the building energy performance. Italy focuses on the determination of the total heat loss coefficient and the influence of solar and internal heat gains through a statistical characterization of the building. This is followed by calibration of the numerical model by comparing both expected energy need and the real measured consumption, and the expected and real aggregated parameters resulting from the first analysis.

Characterization of the sample is the most common aim within the analysis of a large building stock. In particular, work groups (Italy, Japan, Spain, Canada) focused the investigations on the understanding of the influential factors which determine the energy consumption and establishing reliable building energy demand models. Benchmarks for electrical energy uses and for total primary energy consumption for the whole building stock is a goal present in some of the investigations like in the Italian case utilizing 4000 bank branches.

Prediction is also dealt with in the large building stock, in particular with the aim of establishing building energy predictive models and goodness of fit to measurements assessment in the case of Italian bank branches and Canadian investigation on residential buildings.

Due to the huge amount of data existing in a database at a national and regional level, the main goal of the investigations are to define building typologies to estimate energy demand of a building stock and to estimate the amount of energy used for different end uses.

Table 5-2: Distribution by goal of the analysis of the obtained contributions

Partner	**Description**		**Prediction**	
	Individual Building	*Large building stock*	*Individual Building*	*Large building stock*
TU Wien (Austria)	Energy-related user behavior			
Concordia University (Canada)				To establish reliable building energy demand predictive models
CETHIL, INSA de Lyon (France)			To estimate the HVAC energy consumption	
Karlsruhe	Energy-related user			

Institute of Technology (Germany)	behavior			
Polytechnic of Turin (Italy)	Identification of the influence of solar and internal heat gains	Identification of energy consumption influential factors	To estimate the building energy performance	To establish reliable building energy demand predictive models
Tohoku University (Japan)		Identification of energy consumption influential factors	To predict the peak energy consumption	
Tohoku University (China houses)		Identification of energy consumption influential factors		
NTNU Trondheim (Norway)	Identification of energy consumption influential factors			
CIMNE (Spain)		To analyze three parameters taken as building performance indicators		
		National or regional level		
Tsinghua University (China)	To study statistical distribution characteristics of office building energy use in China			
LBNL (U.S.)	to estimate the amount of energy used for different end uses			
Polytechnic of Turin (Italy)	to estimate energy demand of a building stock			

The topic concerning the structure of the database is absolutely not trivial and very often it is faced with little rigor even though it is a crucial issue. In this section, the contributions of the partners are analyzed according to the level of database they have collected.

The database structure referred firstly to the detail of time disaggregation of the energy consumption and of the time dependent influencing factors:

- Level 1 – Annual energy consumption,
- Level 2 – Monthly energy consumption,
- Level 3 – Daily/hourly energy consumption.

The acceptable minimum level depends on the goal and on the subject of the analysis, but typically:

- For analyses on large building stocks, level 1 was acceptable for the investigations,
- For analyses on individual buildings, level 2 is considered as the minimum level.

When the study focuses on very large building stocks, useful analyses can be performed even if little information for each single building is available (annual energy consumption and some influencing parameters) but for a wide number of buildings; when the study focuses on individual buildings, the amount of required information increases, at least because the data about energy consumption (and the corresponding influencing factors) have to be collected at a monthly level.

In particular, in cases where the main aim of the investigations is related to the building occupants, the database contains:

- Building geometries and qualities,
- Climate (indoor and outdoor) information,
- Occupancies and lifestyles.

Such database characteristics are summarized in Table 5-3 for the provided contributions.

Table 5-3: Distribution by database level of the obtained contributions

Partner	**Residential**		**Office**	
	Individual Building	*Large building stock*	*Individual Building*	*Large building stock*
TU Wien (Austria)	Level 3		Level 3	
Concordia University (Canada)		Level 3		
CETHIL, INSA de Lyon (France)	Level 3			
Karlsruhe Institute of Technology (Germany)	Level 3			
Polytechnic of Turin (Italy)			Level 3	Level 1
Tohoku University (Japan)	Level 3	Level 1		
Tohoku University (China)		Level 1		
NTNU Trondheim (Norway)			3	
CIMNE (Spain)				1
National or regional level				
Tsinghua University			Level 1	

(China)		
LBNL (U.S.)		Level 1
Polytechnic of Turin (Italy)	Level 1	

The analysis methods adopted in the contributions depends on the goal and the subject of the investigations. Generally, the description of the subject is dealt with using regression techniques.

In the contributions on individual buildings, different types of regression analysis (linear, multivariate, logistic, partial least square) are used. Norway used partial least square regression to establish the variables with the greatest effect on energy use in buildings. The Austrian and German investigations applied multivariate and logistic regression types of analysis to identify a series of user profiles. The Italian investigation with linear regression and the French regression analysis based on a quartile – quartile plot are aimed at determining correlations between heating energy need and external temperature.

Within a large building stock, the identification of possible influential factors on energy consumption is mainly dealt with using regression techniques. Quantification method 1 is also used to analyze qualitative factors by Japanese investigations of both residential and office buildings.

Table 5-4: Analysis methods used in the contributions for description of the subject

Partner	**Residential**	**Office**
TU Wien (Austria)	Regression analysis	Regression analysis
Karlsruhe Institute of technology (Germany)	Multivariate linear and logistic regression analysis Model optimization through AIC and Nagelkerke's R^2	Graphical analysis
Polytechnic of Turin (Italy)		Linear regression
Tohoku University (Japan)	Quantification method 1	Multiple regression
Tohoku University (China houses)	Quantification method 1	
NTNU Trondheim (Norway)		Partial least squares regression
CIMNE (Spain)		Linear regression
National or regional level		
Tsinghua University (China)	Frequency distribution Cluster analysis	
LBNL (U.S.)		Regressions techniques
Polytechnic of Turin (Italy)	Hierarchical clustering techniques	

Prediction methods within the provided contributions are dealt with using multiple regression analysis and cluster analysis, tree structure, association rule mining and neural networks. Multiple regression analysis was chosen to identify a mathematical model able to forecast energy consumption in buildings with a set of already-known individual variables by using linear functions, while the method of neural networks (a basic data mining technique) is used to analyze the non-linear relationship between energy consumption and individual variables. Decision tree structure is a data mining technique using both numerical and categorical variables with interpretable flow-chart tree structures that enables users to quickly extract useful information.

Table 5-5: Analysis methods used in the contributions for prediction

Partner	Residential	Office	Other
CETHIL, INSA de Lyon (France)			Regression based on quartile - quartile plot
Concordia University (Canada)	Tree method technique Cluster analysis Association rules		
Polytechnic of Turin (Italy)		Multiple linear regression t-test F-test Chi-squared test VIF Neural network	
Tohoku University (Japan)	Multiple regression analysis Neural network		

According to the goal of Annex 53 “Total energy use in buildings: analysis and evaluation methods”, the contributions that took into account the total energy use in buildings are highlighted in the following table.

Table 5-6: Contributions taking into account the total energy use in buildings

Partner	Total	Heating	Cooling	DHW	Electricity	Lighting	Other
Concordia University (Canada)	X						
CETHIL, INSA de Lyon (France)		X	X				
Polytechnic of Turin (Italy)		X	X		X		
Tohoku University (Japan)							
*a)**	X	X	X	X			

*b)***					X		
*c)****	X						
*d)*****	X	X	X	X	X	X	
Tohoku University (China houses)	X	X					
NTNU Trondheim (Norway)		X			X		
Polytechnic of Turin (Italy) - regional							
LBNL (U.S.) - national	X	X	X	X	X	X	X (oil fuel, Natural gas)

a)* houses in Sendai b)** peak electricity in 6 detached houses c)*** office buildings d)**** 80 households

5.2 Individual buildings

Focusing on single buildings, statistical methods can be used for different purposes dealing with the total energy use. In the following figure (Figure 5-1) the different stages are presented till a prediction model for the total energy use is well defined and validated. For the description of the influencing factors, the analyses of the relevant influencing factors, the parameter identification for the prediction model and the estimation of the accuracy of the prediction statistical methods can be used.

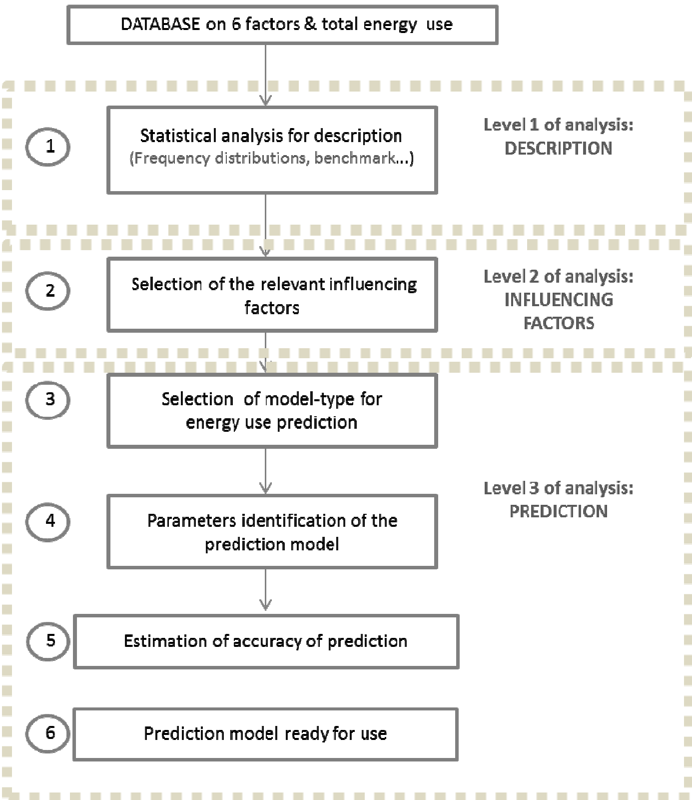


Figure 5-1: Stages of development a model for prediction the total energy use in buildings

<u>STAGE 1: DESCRIPTION</u>

For description of the climate, the building, the operation and maintenance and the occupants typically descriptive statistics (Average Values, Standard deviations, Distributions) are used.
Some example parameters for each category are listed below.

- CLIMATE: Distribution of Hourly Mean Outdoor Temperature in January,
- BUILDING: Distribution of Thermal conductivity of an insulation material,
- BUILDING: Average living space area per person,
- OPERATION: Average value and standard deviation of indoor temperature in sleeping rooms,
- MAINTENANCE: Distribution of lifetime of a lighting bulb,
- MAINTENANCE: Average lifetime of a glazing system,
- OCCUPANTS: Average value and standard deviation for occupation during a weekday,
- OCCUPANTS: Average opening time of windows.

Also, the energy use can be investigated with descriptive statistics. In Experience 1, Yoshino et al. analyzed the energy consumption in 6 detached houses out of a field survey of 80 houses. The next figure presents the frequency distributions of peak load electricity for different time spans as a histogram and as a cumulative distribution.

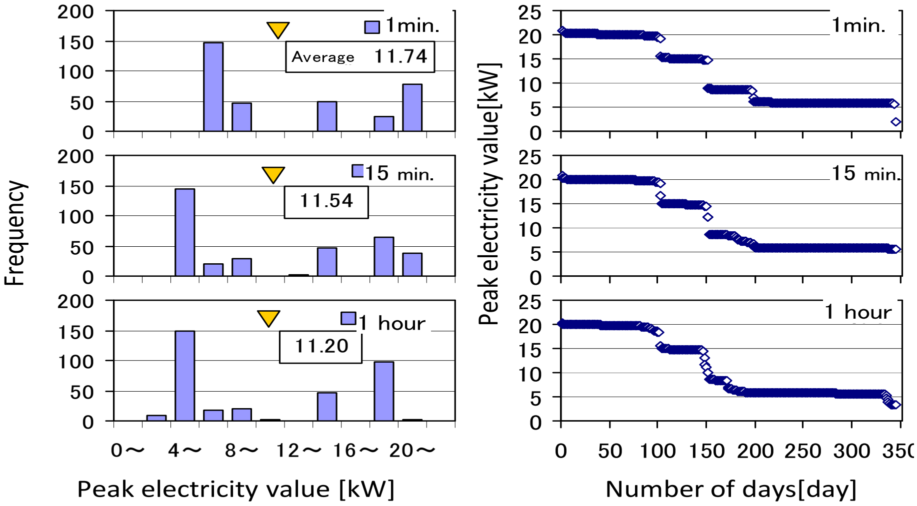


Figure 5-2: Frequency distribution of peak value of a year (TOHOKU 07) [Experience 1].

A very interesting application of statistical methods is to find correlations between energy consumption and parameters describing the objects. In the next figure from Experience 7, an example from Austria is presented where the consumption of hot and cold water and electricity for household equipment is correlated with the number of persons in the household.

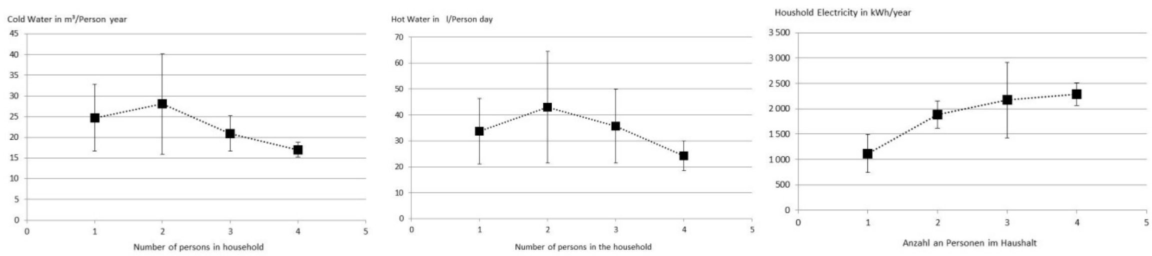


Figure 5-3: Analyses of measured consumption of cold water/hot water and electricity for household equipment in a multifamily building in Vienna (Year of Measurement and Questionnaire: 2011, Number of Household 44) [Experience 7]

<u>STAGE2: SELECTION OF RELEVANT FACTORS</u>

In stage 2 of the development of prediction models for single buildings statistical methods are used to identify the important parameters. In Experience 2, the data of a Building Energy Management System from a real office building is analyzed with a multivariable regression method looking for the important parameters that govern the heating energy use, the electricity consumption and the fan energy use. In that experience a partial least squares regression (PLSR) and principal components regression (PCR) are used to model a response variable when there are a large number of predictor variables, and those predictors are highly correlated or even collinear. Both methods construct new predictor variables, known as principal components (PCs), as linear combinations of the original predictor variables. In the following figure the importance of an original variable is presented for the four most important PCs by showing the PLS weights.

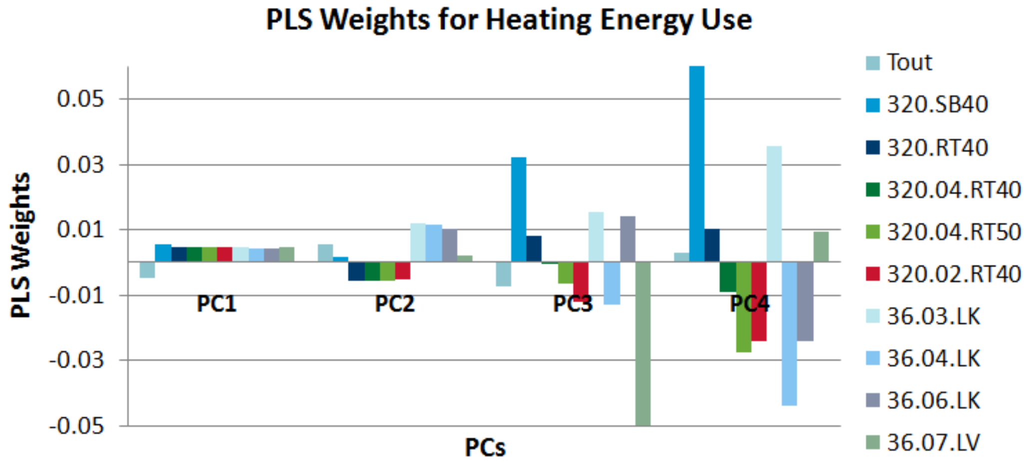


Figure 5-4: PLS weights for heating use of the four most important principal components [Experience 2]

By using procedures for model scaling and finding driving variables based on PLS weights, it was found that the most important variables for the heating energy use are outdoor temperature, control parameters and temperatures in the substation, and some of ventilation parameters. These ventilation parameters were related to the AHUs that were mostly in use.

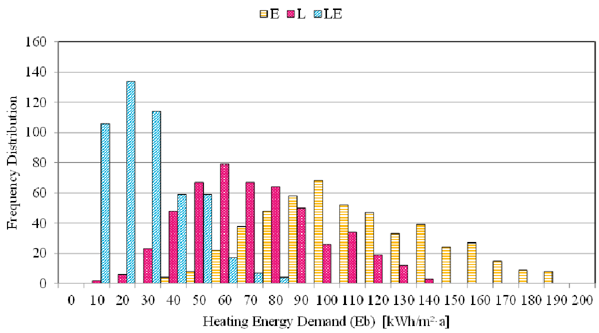
More focused on occupant behavior modeling, in Experience 3 a regression analysis was chosen in order to analyze the influence of physical and individual factors on the frequency of AC-unit usage for cooling and heating as well as the chosen set-point temperature. The importance of one factor compared to the others has been calculated as the product of the absolute value of the coefficient (the respective value of ß) and the range of the variable. This product was called the “importance value”.
In Experience 4 the main focus was to identify energy-related user behavior patterns for window control and the usage of sun protection devices in relation to outdoor in indoor climate. Logistic regression has been applied as the methodological approach to explore patterns of user behavior. The method allows predicting the outcome of a binary dependent variable by modeling the probability of an event such as window-opening (‘yes’ or ‘no’).

<u>STAGE 3 – SELECTION OF MODELING TYPE FOR ENERGY USE PREDICTION</u>

In Experience 5 several classical methods to predict the energy consumption of a real office building in Rome, Italy are compared to the modeling method using Artificial Neural Networks Ensembling (ANNE). The results show that the proposed ANNE approach provides a remarkable improvement with respect to the best classical method (using the average load profiles).

<u>STAGE 4&5: IDENTIFICATION OF PARAMETERS AND ANALYZING THE PREDICTION ACCURACY</u>

In Experience 8 based on a probabilistic occupant model the heating energy demand of a single family building with three different types of building envelopes has been calculated with a detailed building simulation. From this database of 1000 "virtual families", in three types of buildings the most important occupant related parameters (average indoor temperature, average internal loads and average outdoor air exchange) have been identified by minimization of the difference between the average of the ensemble and the simplified calculation with the average values. The parameters have been identified using 2 of the three types of buildings and the accuracy has been analyzed using the third type of building.



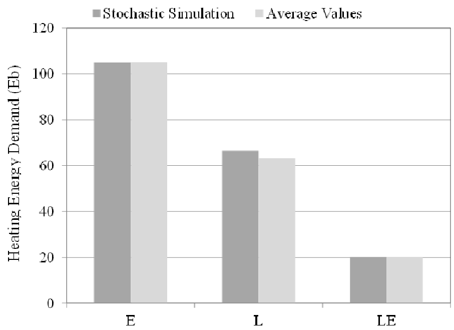


Figure 5-5: Frequency distribution of heating energy use calculated with a full building simulation and an probabilistic occupant model (left) with three different building envelopes(E=Existing building from 1970, L=Low energy house, LE=Lowest energy house). Analyzing the accuracy of the simplified model using the parameters identified with the buildings E and LE. The comparison of the case with Low energy house shows the accuracy of the model. [Experience 8]

References (contributions to the Annex 53)

Extended contributions

- Experience 1: Hiroshi Yoshino, Ayako Miura , Survey of the peak electricity in residential buildings - Analysis based on survey of energy consumption for 80 houses in Japan
- Experience 2: Djuric, N. and V. Novakovic , Application of the prediction model for building energy use assessment, Analysis of BEMS database by using multivariable regression analysis
- Experience 3: Schweiker, M. and Shukuya, M., Modelling of Occupant Behaviour
- Experience 4: Fatma Zehra Çakıcı, Karin Schakib-Ekbatan, Marcel Schweiker
- Experience 5: Lauro, F., Building energy consumption modeling with neural Ensembling approaches
- Experience 6: Stoyan Danov, Jordi Carbonell, Jordi Cipriano. Building energy performance evaluation using daily consumption data in nine individual office buildings in Spain.

Synthetic contributions

- Experience 7: Naomi Morishita, Thomas Bednar, Evaluation of a Low Energy Multifamily building in Vienna, Austria
- Experience 8: Thomas Bednar, Naomi Morishita, Kerstin Seif – Development of statistical analysis for total energy use in individual buildings

- Experience 9: Bednar, Naomi Morishita, Kerstin Seif – Development of statistical analysis for total energy use in small office buildings

5.3 Large Building Stock

Statistical analysis of a large stock of buildings represent methods used to estimate the energy consumption and/or the peak demand of a building at a level of detail that is suited to apply to a number of buildings that is statistically significant (usually more than tens of buildings). The principle of the approach is to project the experimental data on a basis. The methods depend on the type of basis which is typically defined by its dimension and its components.

One type of projection is on categories (Experiences 1, 2, 3, 4, 5, 6). For example, Hu and Yoshino (Experience 4) consider the climate zones, the area of the building, the type of the heating system and its operation, as well as the number of people in the household and their annual income. In another study, Yoshino (Experience 5) considers, besides the categories mentioned before, the weather, indicated by the cooling and heating degree days and the indoor temperature during the heating and the cooling season. The resulting models are regression models using different variants: multi regression, neural networks, and quantification methods (Experiences 1, 2, 5).

Categorizing reduces the variance of the predicted results. The physical explanation of the result is embedded in the categories. Usually, these approaches do not differentiate between the inputs (e.g. weather), the parameters (e.g. floor area, total heat loss coefficient) and the outputs (e.g. indoor air temperature) of a physical (or direct) model. The results indicate the influence of each category given by the weighting coefficient in the model.

This kind of approach, which uses less data (in fact the data available), is very effective in practice. It allows the prediction of energy consumption with an expected variance for real buildings by using data which are available mainly on monthly and/or annual bills.

Comparison between categories needs a criterion which "normalizes" the consumption in order to negate the effect of parameters specific to a given building. For example, Corgnati et al. (Experience 6) propose and demonstrate the application of an indicator that normalizes the data as a function of the heated volume and the climate, described by the degree days of the site.

The second class of projection is on parameters of physical models. The main idea in this approach is to consider a physical model based on the heat balance and to identify the parameters of this model which increase the fit between the predicted results and the measurements. One of the most common approaches is to use the load curve, which expresses the dependence of the heating (or cooling) consumption on the outdoor temperature. This "thermal signature" of the building can be used together with the distribution of degree-days or degree-hours in order to estimate the energy consumption (e.g. the bin method). Basically, the building signature is obtained by regression. Robust regression may be used to improve the prediction in case of perturbation such as the usage of the building (Experience 7). The advantage of this approach is that the thermal behavior of the building, the comfort and the climate are decoupled. A variant of this method is to use the free-running temperature, which allows the estimation of the energy savings for cooling by using free-cooling by ventilation (Experience 8).

Refinements of the thermal signature or the load curve method are proposed (Experiences 8, 9). Ghiaus (Experience 8) demonstrated the equivalence between the load curve and the free running temperature. By using the free-running temperature, the whole range of building operation (heating, ventilation and cooling) is described by a single concept.

Normally, thermal signature is a static method. However, the heat balance may be written taking in account the accumulation. By doing so, Danov et al. (Experience 7) obtained a dynamic model which can estimate the influence of the thermal mass of the building on the energy consumption. Solar gains may be also included in the thermal signature, reducing the variance of the energy estimation (Experience 7).

References (contributions to the Annex 53)

Extended contribution

- Experience 1: Sawako Nakamura, Hiroshi Yoshino, Ayako Miura. Statistical analysis for energy consumption of office buildings in Japan
- Experience 2: Sawako Nakamura, Hiroshi Yoshino, Ayako Miura. Statistical analysis for energy consumption in residential buildings in Sendai
- Experience 3: Hiroshi Yoshino, Ayako Miura. Survey of the peak electricity in residential buildings (see Experience 1 in Individual Buildings for details)
- Experience 4: Tianchi Hu, Hiroshi Yoshino. Statistical analysis on energy consumption of residential buildings in China
- Experience 5: Hiroshi Yoshino. Field Survey and Statistical Analyses on Energy Consumptions in the Residential Buildings in Japan
- Experience 6: Stefano Paolo Corgnati, Federica Ariaudo, Marco Filippi. Heating consumption assessment and forecast of existing buildings: investigation on Italian school buildings
- Experience 7: Stoyan Danov, Jordi Carbonell, Jordi Cipriano. Building energy performance evaluation using daily consumption data

Synthetic contributions

- Experience 7: Cristian Ghiaus. Experimental estimation of building energy performance by robust regression
- Experience 8: Cristian Ghiaus. Equivalence between the load curve and the free-running temperature in energy estimating methods
- Experience 10: Zhun Yu, Fariborz Haghighat. Mining Hidden Patterns from Real Measured Data to Improve Building Energy Performance

5.4 National/Regional analyses

Statistical analysis of national building energy consumption is aimed at defining a general overview of the energy end use due to the construction sector, at a national level. Actually, the knowledge of national building energy use has remained under investigation, due to a lack of information regarding the overall characteristics. With the aim of building strong national databases, national agencies and institutions (CBECS in the U.S., MOHURD in China, TABULA in Europe) have gathered real energy use data and physical characteristics on the national building stock. Specifically, China has collected data of government office buildings and large-scaled commercial buildings [Experience 1], the U.S. has built a national sample database of commercial buildings [Experience 2], whereas the European countries have collected data characterizing the national residential building stock [Experience 4].

The subject of the task is to collect and subsequently to elaborate data characterizing national building stocks in order to offer a realistic interpretation of typical building energy consumption. Different approaches have been tested and used for statistical analysis of the databases.

Wei, Xiao and Jiang [Experience 1] adopted two statistical research methods: boxplot and a key statistical parameter of energy use data and frequency distribution analysis. Both these two approaches have been presented as effective and suitable for future analysis and international comparison. Database characteristics have been gathered based on regional government website releases and included Gross Floor Area (GFA) as well as annual electricity consumption (excluding district heating) of 4600 office buildings. Cluster analysis showed that the average national stock electric consumption is 107 kWhe/m2a for private office buildings and 67.6 kWhe/m2a for government office buildings.

Hong and Wang [Experience 2] analyzed utility bills (monthly energy use for electricity and natural gas) of the CBECS U.S. sample survey and broke them down into energy end use for the national commercial building stock. Statistical regressions and engineering modeling approaches were used to estimate national end use based on consumption data. Average energy consumption for the commercial buildings in the U.S. - taken from monthly regression models of 1518 gathered buildings - is 292.6 kWh/m2, where the single largest part (35.3%) is due to space heating.

The European project TABULA (Typology Approach for Building Stock Energy Assessment) [51] presented by Talà et al. [Experience 4] aimed to create a homogeneous database for European residential building typologies. The research tested three statistical methods with the final goal of estimating the energy consumption of residential building stocks to subsequently predict the impact of potential energy efficiency measures of benchmark models at the national level (based on a singular evaluation for each European country participating the project). These methodologies shoot for the enhancement of the potential impact of energy saving measures and carbon dioxide reduction, by means of the selection of the most adequate energy retrofitting strategies and interventions in existing buildings [Experiences 3,4]. Model calculations aimed at estimating the energy saving potential of national residential building stocks using the Energy Balance Method were developed by four countries (Denmark, Germany, Italy and Czech Republic) representative of the main European climatic regions. This was accomplished using the national EPBD asset rating method [Experience 3]. Moreover, the same modeling method (EBM) can possibly be extended to the energy performance assessment of the whole national building stock.

For each country, two levels of building retrofit were considered: (a) standard refurbishment, applying standard national measures and (b) advanced refurbishment, applying the best national technologies available [Experience 3]. Specifically, the Italian database contained records for more than 66.000 houses rated across the Piedmont region and also gathered information on physical characteristics and calculated energy requirements of single houses. On the basis of three independent variables elaborated by means of statistical analysis (location, age, form of the building), a total of 84 building types (archetypes) representative of the Italian residential building stock were generated [Experience 4].

All these kind of approaches, which use statistical analysis of national building stock samples, are very effective. As a matter of fact, average predictions of energy consumption at the national level are available. Public existing building energy use has existed for a long time at a micro-perspective [Experience 2] due to a lack of shared definitions and outdated information [Experience 3]. Nonetheless the development and the statistical analysis of strong national energy-use datasets, could be one element towards a more robust estimation of the overall energy consumption of the national building stocks.

References (contributions to the Annex 53)

Extended contributions

- Experience 1: Qingpeng Wei, He Xiao, Yi Jiang. National Database of Office Building Energy Use in China
- Experience 2: Tianzhen Hong, Liping Wang. The U.S. Commercial Buildings Energy Consumption Survey (CBECS)
- Experience 3: Cristina Becchio, Stefano P. Corgnati, Ilaria Ballarini and Vinecenzo Corrado. Energy saving potentialities by retrofitting the European residential sector.
- Experience 4: Novella Talà. National/Regional level, Single & Multi family houses.

6. General conclusion and perspective

In this report, the assessment of potential application of statistical analysis for the prediction total energy use in buildings and for the identification of the related most significant influencing factors is dealt with. This has been covered first by an extended literature review, followed by the collection and critical analysis of experiences carried out by ST-C working group.
First, it emerged one of the key passage when statistical based tools are used for the analysis is to clearly define to subject of the study. As mentioned, the possible applications of statistical analysis may be divided into two big fields: to analyze individual buildings, or to focus the analysis on large building stock up to national or regional analysis.
The collection of the Annex 53 partners' experiences, showed different types of information, very detailed up to the breakdown of each single final energy use. On the contrary, in the case of regional or national analysis, few information about each building of the sample are needed to provide some basic but interesting statistical analysis. Starting from these considerations, it arises that to select a suitable methodology, the "scale" of the analysis is essential. To this aim, three main descriptors have to be considered: number of buildings to be analyzed (from an individual to very large building stocks), number of items describing each buildings and time frequency of the collected time dependent parameters (annual to sub-hourly time frequency).
The possible application of statistical analysis depends on the aim of the analysis itself. Statistics could be used to describe the object of the study (descriptive statistics) to provide a clear description of the actual energy consumption; then to find out which are the dominant influencing factors, to put in relation the dependent variable (energy use) and independent variables (the influencing factors). When the most important influencing factors are known, statistical analysis could be used to build up a prediction model. Statistics can also be applied for creating reference buildings representative of a building stock, that can be implemented in direct building energy simulation tools.
Another possible application of statistics is to define "modules" meaning to provide statistical inputs for a direct building energy simulation tool. For example, dealing with occupant behavior, the action on adjusting the thermostat, it's something that is not deterministic, but it's related to the probability of doing a certain action when some environmental parameters are present. So the input data (probability) for direct simulation tool could be defined through a statistical approach.
According to this general scheme, the application of statistical analysis can be structured in three levels of investigations.
The first level is a basic level: since an amount of data is available first of all tendencies related to the dataset should be clarified. The use of statistical parameters (mean value, standard deviation, ...), of frequency distributions of the collected data, etc. can provide significant information to define a clear picture of the subject of the study. It's the use of statistics to describe.
The second level is to use statistics to find out the dominant influencing factors on energy uses. If the most dominant influencing factors can be identified reduced to a limited number of parameters, it's possible to find out the relationship between these parameters and the final energy use and then a very quick and robust prediction model could be built to provide information about the energy behavior of the building (level three).
The adopted analysis methods in the contributions depend on the goal and the subject of investigations. In the contributions on individual buildings, different types of regression analysis (linear, multivariate, logistic, partial least square) are used. Generally, the description of the subject is dealt with regression techniques. Within large building stock arrived experiences, the identification of possible influential

factors on energy consumption is mainly dealt with regression techniques. Quantification method is also used to analyze qualitative factors.
Prediction methods (within the arrived contributions) are dealt with both multiple regression analysis and cluster analysis, tree structure, association rule mining and neural network.
Multiple regression analysis was chosen to identify a mathematical model able to forecast energy consumption in buildings with a set of already-known individual variables by using linear functions, while the method of neural network (a basic data mining technique) is used to analyze the non-linear relationship between energy consumption and individual variables. Decision tree structure is a data mining techniques using both numerical and categorical variables with interpretable flow-chart tree structures that enables users to quickly extract useful information.
As well known, a significant problem to face is the difference between the predicted building energy demand and its actual energy consumption. A big impact on this difference is due to the lack of knowledge about the real functioning of the building during its day by day life. A fundamental aspect for a better description of the building real functioning is to investigate, to highlight and to express those factors related to the actual functioning. Annex 53 “Total Energy Use in Buildings -Analysis and evaluation methods” had as its ultimate outcome a better understanding and a strengthening of the knowledge for robust prediction of total energy usage in buildings, hence enabling the assessment of energy-saving measures, policies and techniques. On this base, Annex 53 focused on the influence of occupant behavior on building energy consumption, with the purpose to bring the occupants behaviors into the building energy field so as to conduct the building energy works (research, practice, policy, etc) more closed with the real world.
Annex 53 clearly defined the approach to describe occupant behavior quantitatively in the field of building energy performance, by setting up probabilistic models for predicting occupant behavior in dwellings and office buildings. A further perspective of the statistical analysis is represented by new methodologies (modeling approaches) and techniques (monitoring hardware and software platforms) for analyzing real building total energy use and for investigating the factors influencing occupant behavior in buildings, are therefore available for further investigations and for specific applications as in energy auditing and smart metering of HVAC systems. Combinations of deterministic and probabilistic behavioral models can define all the possible interactions between users and building controls (such as occupancy presence, heating and cooling set point adjustments, use of lighting and equipment) and they can be implemented simultaneously in building models in energy simulation tools, in order to obtain a complete evaluation of users’ influence in modifying the building energy performance. Such models should also help in identifying wastes in environmental control, as, for example, excessive air renovation, too accurate humidity control, etc.

7. References

[1] B. Abushakra, D. Claridge, Accounting for the Occupancy Variable in Inverse Building Energy Baselining Models, ESL-IC-01-07-29.

[2] M. Aydinalp-Koksal, V. Ismet Ugursal. Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. Applied Energy 85 (2008) 271–296.

[3] American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Measurement of Energy and Demend Savings, ASHRAE Guideline 14-2002.

[4] Anderson, D., Electrical Usage Predictors Based on the Singular Value Decomposition Algorithm", M.S. Thesis, Civil, Environmental and Architectural Engineering Department, 1990, University of Colorado at Boulder.

[5] K. J. Baker, R. M. Rylatt, Improving the prediction of UK domestic energy-demand using annual consumption-data, in Applied Energy 85 (2008) 475–482.

[6] R. Bender, A. Ziegler, S. Lange, 2007, Logistische Regression, Deutsche Medizinische Wochenschrift; 132, pp. 33-35 (in German).

[7] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). Classification and Regression Trees, Wadsworth, Inc., California.

[8] Cabena P, Hadjinian P, Stadler R, Verhees J, Zanasi A. Discovering data mining: from concept to implementation. Upper Saddle River, NJ: Prentice Hall; 1998.

[9] Cacuci, D.G., M. Ionescu-Bujor, and I.M. Navon, Sensitivity and uncertainty analysis, 2003, Boca Raton: Chapman & Hall/CRC Press. b. <1-2>.

[10] T. Catalina, J. Virgone, E. Blanco, Development and validation of regression models to predict monthly heating demand for residential buildings, in Energy and Buildings 40 (2008) 1825–1832.

[11] Corgnati, S.P., V. Corrado, and M. Filippi, A method for heating consumption assessment in existing buildings: A field survey concerning 120 Italian schools. Energy and Buildings, 2008. 40(5): p. 801-809.

[12] Cowan, J. and S. Schiller., Measuring' Energy Savings for Modernization Projects, ASHRAE Journal, August, pp. 60-62, 1997.

[13] Davies, D.L., Bouldin, D.W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2, 224.

[14] Energy Estimating And Modeling Methods in 2009 ASHRAE Handbook—Fundamentals (SI), cap. 32.

[15] R. Z. Freire, G. H.C. Oliveira, N.Mendes, Development of regression equations for predicting energy and hygrothermal performance of buildings, in Energy and Buildings 40 (2008) 810–820.

[16] M.Frigge, D. C. Hoaglin, B. Iglewicz (February 1989). "Some Implementations of the Boxplot". The American Statistician 43 (1): 50–54.

[17] M. Fuks and E. Salazar, 2008, Applying models for ordinal logistic regression to the analysis of household electricity consumption classes in Rio de Janeiro, Brazil, Energy Economics, Vol. 30(4), pp. 1672-1692.

[18] J. S. Haberl, C. Culp, D. E. Claridge, ASHRAE Guideline 14-2002 for Measurement of Energy and Demend Savings: how to determine what was really savede by the retrofit, Energy Systems Laboratory, Texas A&M University, ESL-IC-10/05-50.

[19] J. S. Haberl, S.Cho, Literature Review Of Uncertainty Of Analysis Methods (Inverse Model Toolkit), Report to the Texas Commission on Environmental Quality, Energy Systems Laboratory, Texas Engineering Experiment Station, Texas A&M University System, October 2004.

[20] J. S. Haberl, Thamilseran, Reddy, Claridge, O'Neal, and Turner, Baseline Calculations For Measurement And Verification Of Energy And Demand Savings In A Revolving Loan Program In Texas, ASHRAE Transactions Vol. 104, Pt. 2, pp. 841-858, 1998.

[21] Engineering Statistics Handbook, http://www.itl.nist.gov/div898/handbook.

[22] Han J, Kamber M. Data Mining Concepts and Techniques, second ed. San Francisco: Elsevier Inc; 2006.

[23] Hand D, Mannila H, Smyth P. Principles of data mining. Cambridge, MA: MIT press; 2001.

[24] Kalogirou SA. Applications of artificial neural-networks for Energy systems. Applied Energy 67 (2000) 17-35.

[25] Kalogirou SA, Neocleous CC, Schizas CN. A comparative study of methods for estimating intercept factor of parabolic trough collectors. In: Proc. of the Int. Conf. EANN'96, London (UK) 1996. p. 5±8.

[26] Kalogirou SA, Neocleous CC, Schizas CN. Artificial neural networks in modelling the heat-up response of a solar steam generation plant. In: Proc. of the Int. Conf. EANN'96, London (UK), 1996.

[27] Kalogirou SA, Neocleous CC, Schizas CN, Artificial neural networks for the estimation of the performance of a parabolic trough collector steam generation system. In: Proc. of the Int. Conf. EANN'97, Stockholm (Sweden), 1997. p. 227±32.

[28] Kalogirou S, Panteliou S, Dentsoras A. Modelling of solar domestic water heating systems using artificial neural networks. Solar Energy 1999;65(6):335±42.

[29] Kalogirou S, Panteliou S, Dentsoras A. Artificial neural networks used for the performance prediction of a thermosyphon solar water heater. Renewable Energy 1999;18(1):87±99.

[30] Kalogirou S, Panteliou S. Thermosyphon solar domestic water heating systems long-term performance prediction using artificial neural networks. Solar Energy Journal, in press.

[31] Kalogirou S. Forced circulation solar domestic water heating systems long-term performance prediction using artificial neural networks. Applied Energy 2000;66(1):63±74.

[32] Kalogirou S, Panteliou S. Dynamic system testing method and artificial neural networks for solarwater heater long-term performance prediction. In: Proceedings of the European Symposium on Intelligent Techniques ESIT'99 on CD-ROM, Crete (Greece) 1999.

[33] Kalogirou SA, Neocleous CC, Schizas CN. Heating load estimation using artificial neural networks. In: Proc. of CLIMA 2000 Conf., Brussels (Belgium) 1997.

[34] Kalogirou S, Eftekhari M, Pinnock D. Prediction of air flow in a single-sided naturally ventilated testroom using artificial neural networks. In: Proceedings of Indoor Air'99, vol. 2, The 8th International Conference on Indoor Air Quality and Climate, Edinburgh, (Scotland), 1999. p. 975±80.

[35] Kalogirou S, Bojic M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy 2000;25(5):479±91.

[36] Katipamula, S., T. Reddy and D. Claridge. Multivariate Regression Modeling, ASME Journal of Solar Energy Engineering, Vol. 120, No. 3, pp. 177-184, 1998.

[37] J. K. Kissock, Jeff S. Haberl, David E. Claridge, Development of a Toolkit for Calculating Linear, Change-point Linear and Multiple-Linear Inverse Building Energy Analysis Models, Final Report for ASHRAE Research Project 1050-RP, November 2002.

[38] J. K. Kissock, Jeff S. Haberl, David E. Claridge, Inverse Modeling Toolkit: Numerical Algorithms, KC-03-2-1 (RP-1050), ASHRAE 2003.

[39] J. K . Kissock, , Claridge, D.E., Haberl, J.S. and Reddy, T.A., Measuring Retrofit Savings For the Texas LoanSTAR Program: Preliminary Methodology and Results", Solar Engineering, 1992: Proceedings of the ASME-JSES-SSME International Solar Energy Conference, Lahaina, HI, April.

[40] J.K. Kissock, A Methodology to Measure Energy Savings in Commercial Buildings, Ph.D. Dissertation, Mechanical Engineering Department, Texas A&M University, College Station, TX, December, 1993.

[41] Kissock, Reddy and Claridge, Ambient-Temperature Regression Analysis for Estimating Retrofit Savings in Commercial Buildings", ASME Journal of Solar Energy Engineering, Vol. 120, No. 3, pp. 168-176, 1998.

[42] I. Kostakis and E. Sardianou, 2012, Which factors affect the willingness of tourists to pay for renewable energy?, Renewable Energy, Vol. 38(1), pp. 169-172.

[43] Leslie, N., G. Aveta and B. Sllwinski, Regression Based Process Energy Analysis System, ASHRAE Transactions, Vol. 92, Pt. 1A., pp, 93-102, 1986.

[44] Mann, Prem S. (1995) Introductory Statistics, 2nd Edition, Wiley. ISBN 0-471-31009-3

[45] Robert McGill, John W. Tukey, Wayne A. Larsen (February 1978). "Variations of Box Plots". The American Statistician 32 (1): 12–16.

[46] I.E. Nikolaou, H. Vitouladitis, K.P. Tsagarakis, 2012, The willingness of hoteliers to adopt proactive management practices to face energy issues, Renewable and Sustainable Energy Reviews, Vol. 16(5), pp. 2988-2993.

[47] Quinlan JR. Induction of decision trees, Machine Learning 1986; 1: 81–106.

[48] Quinlan, J.R. C4.5 Programs for Machine Learning, San Mateo: Morgan Kaufmann; 1993.

[49] RapidMiner: http://rapid-i.com/content/view/181/190/

[50] Ruch, D. and Claridge, D., NAC for Linear and Change-Point Energy Models, Proceedings of the 1992 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, August, pp. 3.263 - 3.273, 1992.Stoecker, W.F., Design of thermal systems, 1989, New York: McGraw-Hill. xvi, 565 s.

[51] TABULA - Typology Approach for Building Stock Energy Assessment. IEE project (www.building-typology.eu)

[52] Trochim, William M. K. (2006). "Descriptive statistics". Research Methods Knowledge Base. http://www.socialresearchmethods.net/kb/statdesc.php Retrieved 14 March 2011.

[53] Tukey, John Wilder (1977). Exploratory Data Analysis. Addison-Wesley. ISBN 0-201-07616-0.

[54] Yu, Z., Haghighat, F., Fung, B.C.M., Yoshino, H. (2010). A decision tree method for building energy demand modeling, Energy and Buildings, 42, (10), 1637-1646.

[55] Yu, Z., Fung, B.C.M., Haghighat, F., Yoshino, H., Morofsky., E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and Buildings, 43, (6), 1409-1417.

[56] Yu, Z., Haghighat, F., Fung, B.C.M., Zhou, L. (2012). A novel methodology for knowledge discovery through mining all associations between building operational data. Energy and Buildings, 47, 430-440.

[57] Datta D, Tassou SA, Marriot D. Application of neural networks for the prediction of the energy consumption in a supermarket. In: Proc. Of CLIMA 2000 Conf., Brussels (Belgium) 1997.p.98.

[58] Claridge, D. 1998. "A Perspective On Methods For Analysis Of Measured Energy Data From Commercial Buildings", ASME Journal of Solar Energy Engineering, Vol. 120, No. 3, pp. 150-155.

V-2

Statistical analysis models and applications

Contents

1. Statistical analysis of individual buildings

1.1 Introduction

Focusing on single buildings statistical methods can be used for different purposes dealing with the total energy use. In the following figure the different stages are presented till a prediction model for the total energy use is well defined and validated. For the description of the influencing factors, the analyses of the relevant influencing factors, the parameter identification for the prediction model and the estimation of the accuracy of the prediction statistical methods can be used.

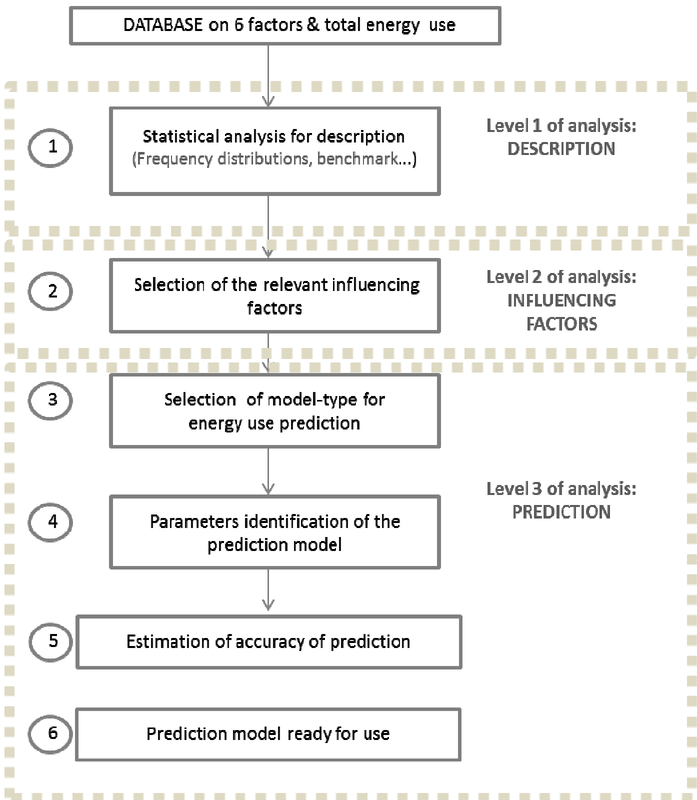


Figure 2-1: "Stages of development a model for prediction the total energy use in buildings"

<u>STAGE1: DESCRIPTION</u>

For description of the climate, the building, the operation and maintenance and the occupants typically descriptive statistics (Average Values, Standard deviations, Distributions) is used.

Examples are:

CLIMATE: Distribution of Hourly Mean Outdoor Temperature in January
BUILDING: Distribution of Thermal conductivity of an insulation material
BUILDING: Average living space area per person
OPERATION: Average value and standard deviation of indoor temperature in sleeping rooms
MAINTENANCE: Distribution of lifetime of a lighting bulb
MAINTENANCE: Average lifetime of a glazing system
OCCUPANTS: Average value and standard deviation for occupation during a weekday
OCCUPANTS: Average opening time of windows

Also the energy use can be investigated with descriptive statistics. In Experience1Yoshino et. al. analyzed the energy consumption in 6 detached houses out of a field survey in 80 houses. The next figure presents the frequency distributions of peak load electricity for different time spans as a histogram and as a cumulative distribution

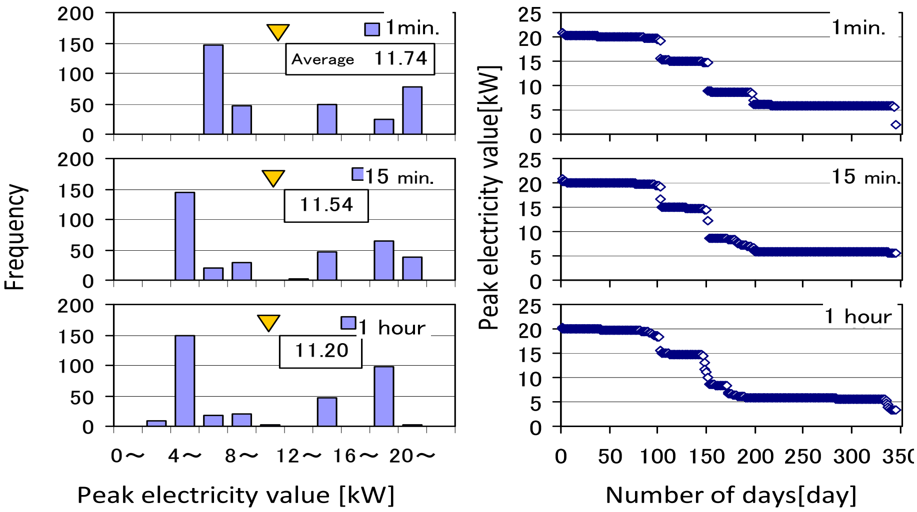


Figure 2-2: “Frequency distribution of peak value of a year (TOHOKU 07) [Experience 1].”

A very interesting application of statistical methods is to find correlations between energy consumption and parameters describing the objects. In the next figure from Experience 7 an example from Austria is presented, where the consumption of hot and cold water and electricity for household equipment is correlated with the number of persons in the household.

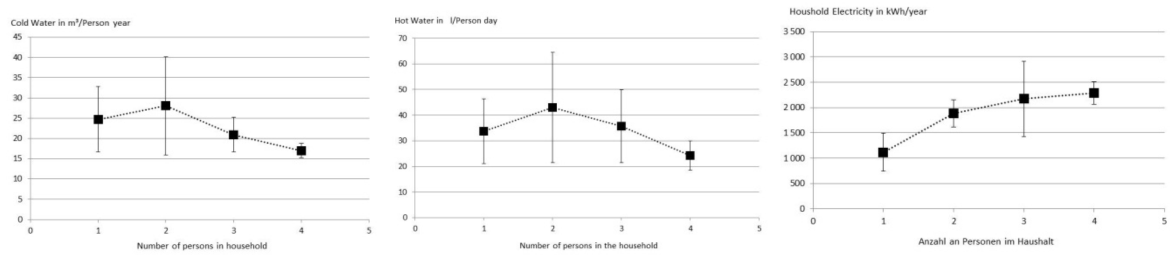


Figure 2-3: “Analyses of measured consumption of cold water/hot water and electricity for household equipment in a Multifamily building in Vienna (Year of Measurement and Questionnaire: 2011, Number of Household 44) [Experience 7]”

<u>STAGE2: SELECTION OF RELEVANT FACTORS</u>

In stage 2 of the development of prediction models for single buildings statistical methods are used to identify the important parameters. In Experience 2 the data of a Building Energy Management Systems of a real office building is analyzed with a multivariable regression method looking for the important parameters that govern the heating energy use, the electricity consumption and the fan energy use. In that experience a partial least squares regression (PLSR) and principal components regression (PCR) are used to model a response variable when there are a large number of predictor variables, and those predictors are highly correlated or even collinear. Both methods construct new predictor variables, known as principal components (PCs), as linear combinations of the original predictor variables. In the following figure the importance of an original variable is presented for the four most important PCs by showing the PLS weights.

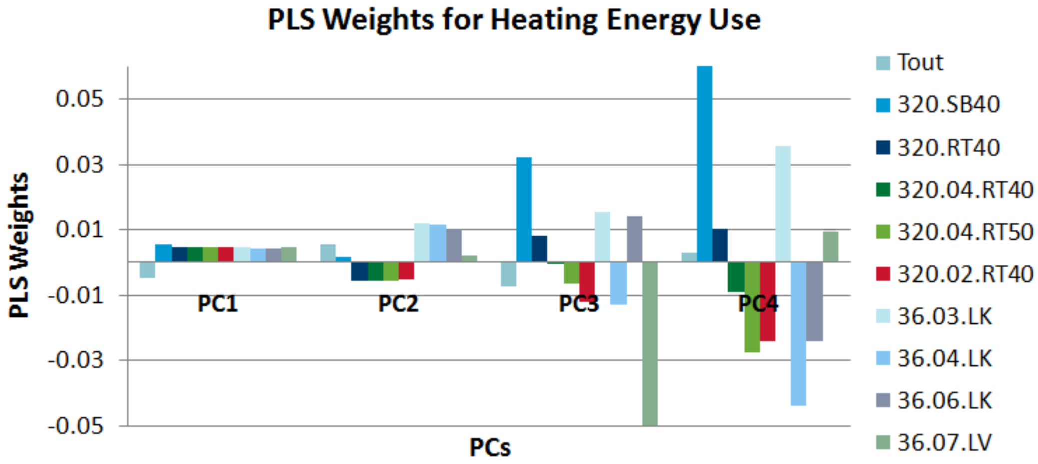


Figure 2-4: "PLS weights for heating use of the four most important principal components [Experience 2]"

By using procedures based on PLS weights for model scaling and finding driving variables, it was found that the most important variables of the heating energy use are outdoor temperature, control parameters and temperatures in the substation, and some of ventilation parameters. These ventilation parameters were related to the AHUs that were mostly in use.

More focused on Occupant behavior modelling in experience 3 a regression analyses was chosen in order to analyze the influence of physical and individual factors on the frequency of AC-unit usage for cooling and heating as well as the chosen set-point temperature. The importance of one factor compared to the others has been calculated as the product of the absolute value of the coefficient (the respective value of ß) and the range of the variable. This product was called the "importance value".

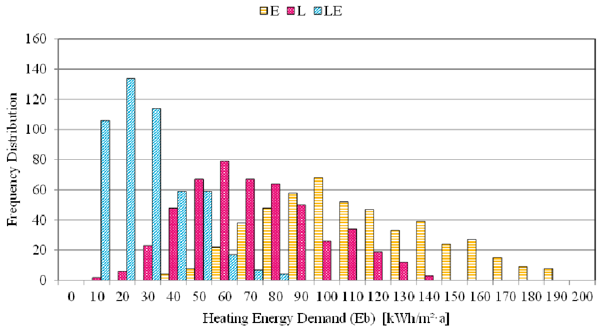
In experience 4 the main focus was to identify energy-related user behavior patterns for window control and the usage of sun protection devices in relation to outdoor in indoor climate. As a methodological approach to explore patterns of user behavior logistic regression has been applied. The method allows predicting the outcome of a binary dependent variable by modeling the probability of an event such as window-opening ('yes' or 'no').

STAGE 3 – SELECTION OF MODELING TYPE FOR ENERGY USE PREDICTION

In Experience 5 several classical methods to predict the energy consumption of a real office building in Rome, Italy are compared to the modeling method using Artificial Neural Networks Ensembling (ANNE). The results show that the proposed ANNE approach can get a remarkable improvement with respect to the best classical method (using the average load profiles).

STAGE 4&5: IDENTIFICATION OF PARAMETERS AND ANALYZING THE PREDICTION ACCURACY

In Experience 8 based on a probabilistic occupant model the heating energy demand of single family buildings with three different types of building envelopes have been calculated with a detailed building simulation. From this database of 1000 "virtual families" in three types of buildings the most important occupant related parameters (Average indoor temperature, average internal loads and average outdoor air exchange) have been identified by minimization the difference between the average of the ensemble and the simplified calculation with the average values. The parameters have been identified using 2 of the three types of buildings and the accuracy has been analyzed using the third type of building.



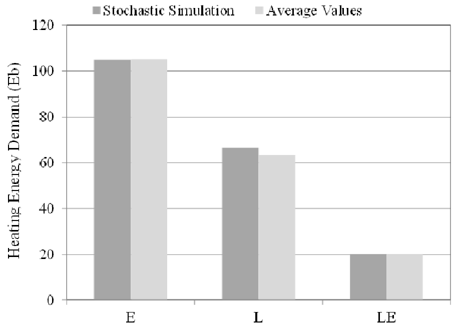


Figure 2-5: "Frequency distribution of heating energy use calculated with a full building simulation and an probabilistic occupant model (left) with three different building envelopes(E=Existing building from 1970, L=Low energy house, LE=Lowest energy house). Analyzing the accuracy of the simplified model using the parameters identified with the buildings E and LE. The comparison of the case with Low energy house shows the accuracy of the model. [Experience8]"

References (contributions to the Annex 53)

Extended contributions

Experience 1: Hiroshi Yoshino, Ayako Miura , Survey of the peak electricity in residential buildings - Analysis based on survey of energy consumption for 80 houses in Japan
Experience 2: Djuric, N. and V. Novakovic , Application of the prediction model for building energy use assessment, Analysis of BEMS database by using multivariable regression analysis
Experience3: Schweiker, M. and Shukuya, M., Modeling of Occupant Behavior
Experience 4: Fatma Zehra Çakıcı, Karin Schakib-Ekbatan, Marcel Schweiker
Experience 5: Lauro, F., Building energy consumption modeling with neural Ensembling approaches
Experience 6: Stoyan Danov, Jordi Carbonell, Jordi Cipriano. Building energy performance evaluation using daily consumption data in nine individual office buildings in Spain.

Synthetic contributions

Experience 7: Naomi Morishita, Thomas Bednar, Evaluation of a Low Energy Multifamily building in Vienna, Austria
Experience 8: Thomas Bednar, Naomi Morishita, Kerstin Seif – Development of statistical analysis for total energy use in individual buildings
Experience 9: Bednar, Naomi Morishita, Kerstin Seif – Development of statistical analysis for total energy use in small office buildings

1.2 Experience 1: Survey of the peak electricity in residential buildings (analysis based on survey of energy consumption for 80 houses in Japan)

(Hiroshi Yoshino, Ayako Miura)

1.2.1 Introduction

The annual peak electricity consumption in Japan has been increasing due to the changes in the energy demand structure. Although a peak demand is known to occur on the hottest day of the year, the biggest consumption has a tendency of reaching higher prominence value. As the annual electricity supply is determined on the basis of last year's annual peak electricity demand, the boost of peak electricity is a huge problem. Consequently, the smoothing down the peak electricity load is a principal goal from the viewpoint of maintaining a stable electricity supply, lowering costs and conserving environment. Despite the fact that peak smoothing has occurred in the residential sector (e.g. use of night time power for thermal storage devices), the biggest demand of the day in residential sector occurs from 8:00pm to 10:00pm. It is important to grasp the peak energy consumption in order to predict the energy supply and demand in the residential sector. In this paper, frequency distribution, day-load-rate and outbreak time of peak electricity analysis was carried out on 6 detached houses based on field measurements of the energy consumption for 80 houses in Japan. This report was authored corresponding to Ref [1].

1.2.2 Database characteristic

- Number of Buildings: 6 detached houses (distributed in six different districts)
- House type: Multi-family (minimum of 2 residents to maximum of 6)
- Analysed period: January 2003 to December 2003 (Actual field measurement was carried out from November 2002 to March 2005)
- Contents: Energy data, temperature and humidity data, building structure and kinds of appliances
- Interval:
  - Electricity: every minute,
  - Kerosene: every five minutes,
  - Gas: every 15 minutes,
  - Temperature and Relative humidity: every 15 minutes basis
- Online database: Energy, temperature and humidity data (available in database: Energy Consumption in 80 Residential Buildings in Japan)

1.2.3 Method

In order to clarify the relationship between peak electricity demand and residential appliances, and the regional characteristics of peak electricity demand, some analyses were conducted on factors which may affect the residential peak energy demand: 1) Time of peak electricity occurrence, 2) Relationship between appliance use and peak energy consumption and 3) Regional characteristics. As for the time variation of the peak value, integrated value of 15 minutes basis has been converted into hourly consumption amounts (15 minutes X4), and the average hourly electricity consumption value is indicated. The biggest electricity consumption of the integrated value per minute is given as the peak value of the day. Furthermore, peak electricity consumption occurrence time, frequency distribution and the ratio of daily load factor were investigated..

1.2.4 **Results and discussion**

Outline of the investigated houses

Table 2-1 shows basic information of the investigated houses. These 6 detached houses were extracted from an analysis of field measurements on energy consumption for 80 houses in Japan. Three houses are all-electric houses and other 3 houses use combined energy sources were selected, since they have enough information and measurement data for the analysis.

Table 2-1: Basic information of investigated houses

Category	Name of detached house	Location	Built year	Gross floor area [m²]	Structure	Q-value [W/m²·K]	Leakage area [cm²/m²]	Energy sourse by each use				Number of family	Air-conditioning Equipments	
								Heating	Cooling	H/W supply	Cooking		Heating	Cooling
All-electric house	Hokkaido 01	Sapporo	1999	147.4	Wood	1.40	0.50	Electricity	Electricity	Electricity	Electricity	6	Panel heater with Hot water	–
	Tohoku 07	Morioka	2000	140.0	Wood	1.01	0.70	Electricity	Electricity	Electricity	Electricity	4	Thermal strage heater	A/C
	Kyushu 06	Maebara	2001	145.7	Wood	2.5	3.00	Electricity	Electricity	Electricity	Electricity	4	A/C, oil heater	A/C
Combined energy sourse housing	Hokkaido 07	Sapporo	1999	240.0	RC+Wood	1.44	0,79	Kerosene	Electricity	Kerosene	Electricity	4	Fan heater	–
	Hokuriku 03	Niigata	2002	117.0	Wood	2.18	0.95	Electricity	Electricity	Gas	Gas	4	A/C, Elctric carpet	A/C
	Kyuushuu 04	Fukuoka	2001	158.9	Wood	2.3	4.00	Electricity	Electricity	Gas	Gas	2	A/C	A/C

Investigation of peak electricity consumption variation and the occurrence time in winter

a) All-electric house

Figure 2-6 shows the time variation of electricity consumption and time of the peak occurs in Hokkaido 01, which is located in one of the coldest part of Japan. Space heating is used all day other than 4:00pm to 6:00pm. Energy consumption is relatively high from 11:30pm to 5:00pm, when a hot water supplier is put into action. Start-up electricity of the hot water supplier causes the peak value to frequently occur between 11:00 pm to 1:00am. Another peak shown around 7:00am is caused by reheating the hot water. Each peak value in Hokkaido 01 stays constant around 12kW.

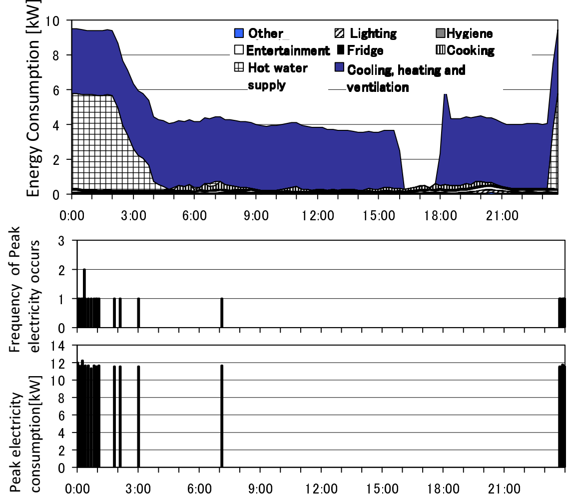


Figure 2-6: "Time variation of electricity consumption and distribution of occurrence time of peak electricity consumption (Hokkaido 01)"

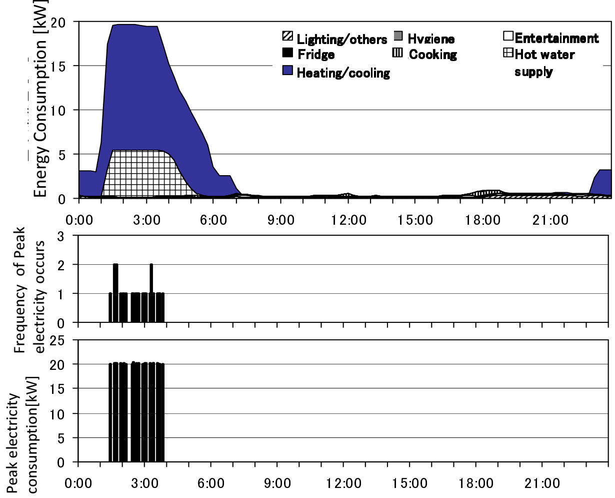


Figure 2-7: "Time variation of electricity consumption and distribution of occurrence time of peak electricity consumption (Tohoku 07)"

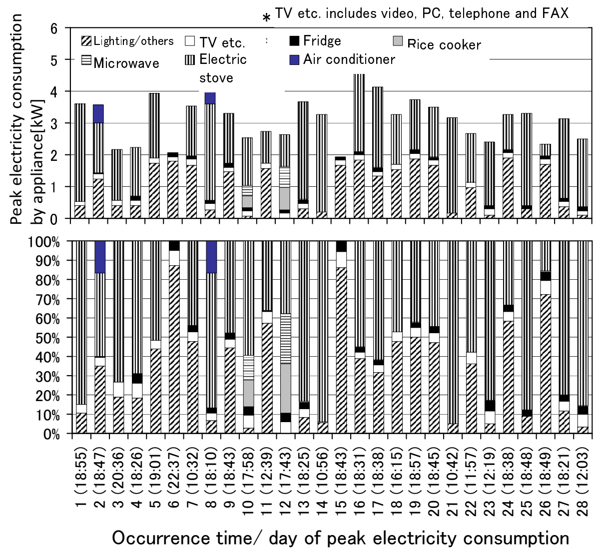


Figure 2-8: "Time variation of electricity consumption and distribution of occurrence time of peak electricity consumption except for night time electricity (Tohoku 07)"

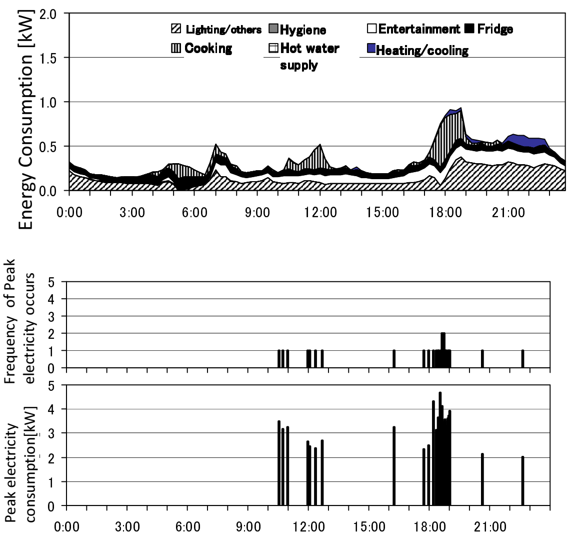


Figure 2-9: "Electricity consumption of each appliance and the percentage (Tohoku 07)"

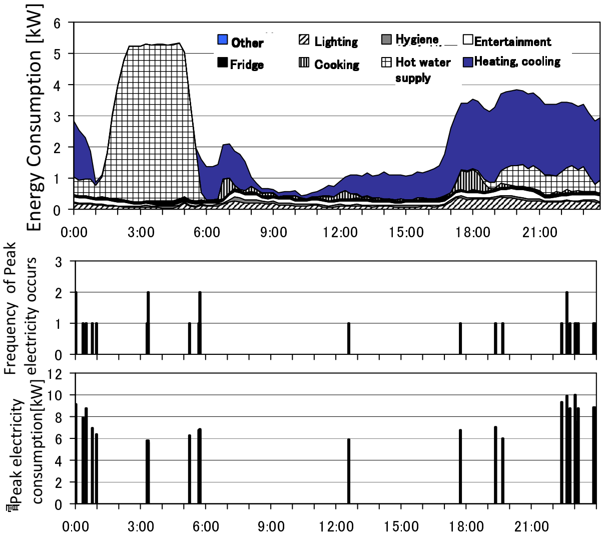


Figure 2-10: "Time variation of electricity consumption and distribution of occurrence time of peak electricity consumption except for night time electricity (Kyushu 06)"

The energy consumption characteristics of other all-electric-house, Tohoku 07 are indicated in Figure 2-7. Tohoku 07 uses a thermal storage electric heater that uses night time electricity. From 1:30am to 4:00am, about 20kW of peak electricity occurs due to the simultaneous operations of the hot water supplier and thermal storage heater. Almost all of the peak electricity consumption amounts are caused by the use of the thermal storage heater. In order to understand the peak energy consumption excluding night time electricity, another analysis was carried out on Tohoku 07.The result is shown in Figure 2-8. Peak values decline when cooking devices are used both in daytime and night time. Peak values are highest around 7:00 pm and the value is 4kW. Figure 2-9 indicates the peak electricity consumption of each appliance used and the percentage used by each appliance. Electric stove and lighting consume more energy than other appliances, as the monthly average ratio indicates: 56.7% for stove, 31.7% for lighting and others and then 4.7% for TV.

The electricity consumption of Kyushu 09, which is located in the warmest region of the three all-electric houses is shown in Figure 2-10. An air conditioner and oil heater are used as heating resources. The occurrence time of peak electricity is distributed widely compared to other all- electric houses. The peak value found at night occurs when the space heating and hot water supplier are operated at the same time. Other peak values are presented when residents use cooking devices, with the peak values distributed around 6 to 10kW.

In those three all-electric houses, thermal storage devices such as hot water suppliers have a large influence on the peak electricity consumption. One distinguishing aspect of energy consumption in all-electric-houses are that equalization of peak electricity occurs due to the concentration of the peak in night time, while general households consume the largest amount of energy around 8:00pm to 10:00pm, when family members tend to spend time together. However, the investigation with the exception of night time consumption shows that the peak occurred around 7:00pm because of electric cooking devices.

b) **Combined energy source houses**

The electricity consumption of a detached house, Hokkaido 07, is shown in Figure 2-11. This house uses kerosene for space heating and the hot water supply. Electricity is used for all other purposes. The main appliance for heating is a kerosene fan heater and an electric fan heater is operated when required. In Hokuriku 03, there are three peaks occurring in the morning, daytime and night as shown in Figure 2-12, but the largest peak occurs around 7:00am. Regarding the effect of microwave and TV usage on producing the peak value, is shown in Figure 2-13. Several appliances have almost the same ratio contrary to Tohoku 07 (all-electric house) of which peak by the electricity was accounted for electricity cooking devices. Figure 2-14 shows the data for Kyushu 04. Gas is used for the hot water supply and cooking, while electricity is used for air conditioning. The peak values have a wide distribution widely and occur frequently around 10:00pm. In addition, there is another peak occurrence around 7:00 when residents operate an air conditioner. In houses which have a variety of energy sources, the occurrence time of peak electricity consumption varies widely in one day, when cooking devices and other appliances are operated simultaneously.

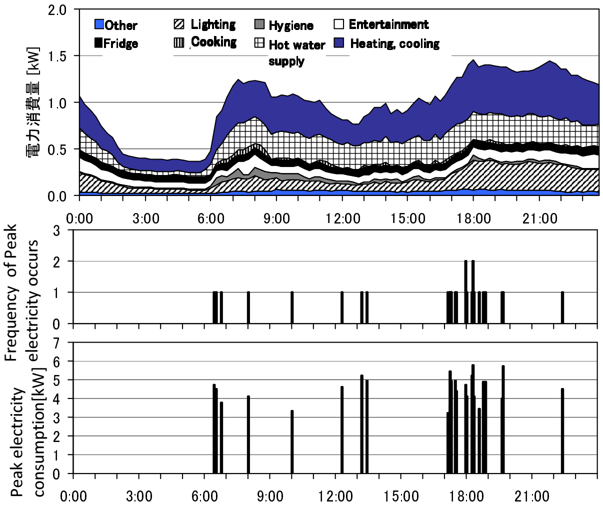


Figure 2-11: "Time variation of electricity consumption and distribution of occurrence time of peak electricity consumption except for night time electricity (Hokkaido 07)"

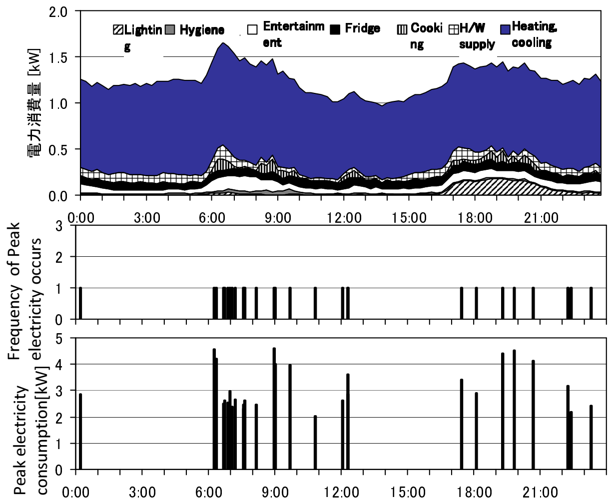


Figure 2-12: "Time variation of electricity consumption and distribution of occurrence time of peak electricity consumption except for night time electricity (Hokuriku 03)"

Investigation on long-term fluctuation of peak electricity

a) All-electric houses

Figure 2-13 shows the integrated electricity consumption in a day and the long term fluctuation in Hokkaido 01. Since the house uses space heating all the day, energy consumption in winter is high (10~12kW). As the season changes into summer, the peak value of energy consumption gradually goes down. The same data for Tohoku 07 is shown in Figure 2-14. The peak electricity consumption in winter is remarkably high (peak of 20kW). From May to October, when energy consumption for heating does not occur, peak electricity is around 7kW which is the same as other houses. In Figure 2-15 (Kyushu 06), the energy consumption for heating varies drastically in one day. Despite of cooling energy consumption in summer, there was no effect on the peak value.

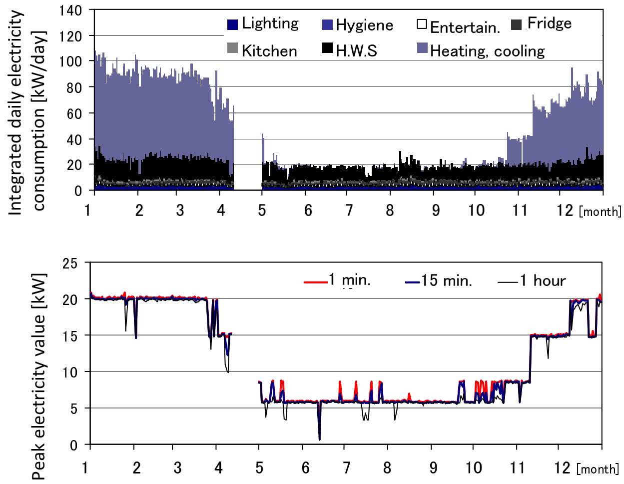


Figure 2-13: "Integrated electricity consumption in a day and long term fluctuation (Hokkaido 01) "

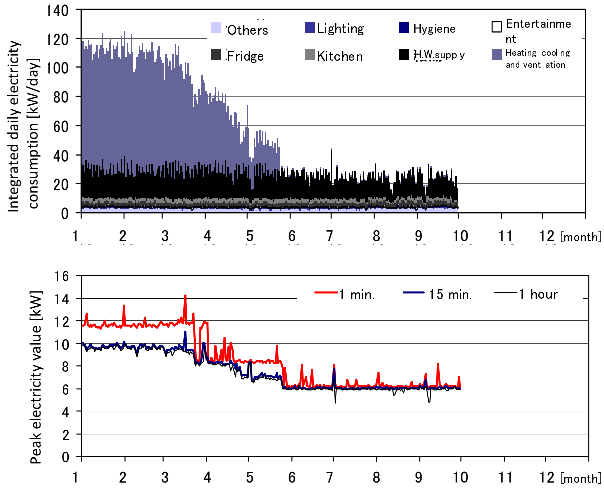


Figure 2-14: "Integrated electricity consumption in a day and long term fluctuation (Tohoku 07)"

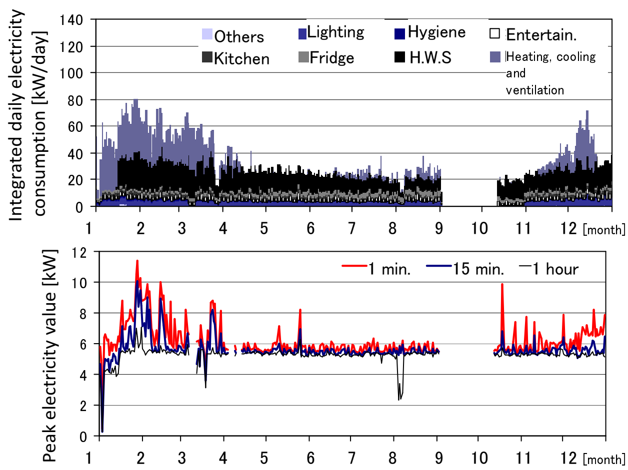


Figure 2-15: "Integrated electricity consumption in a day and long term fluctuation (Kyushu 06)"

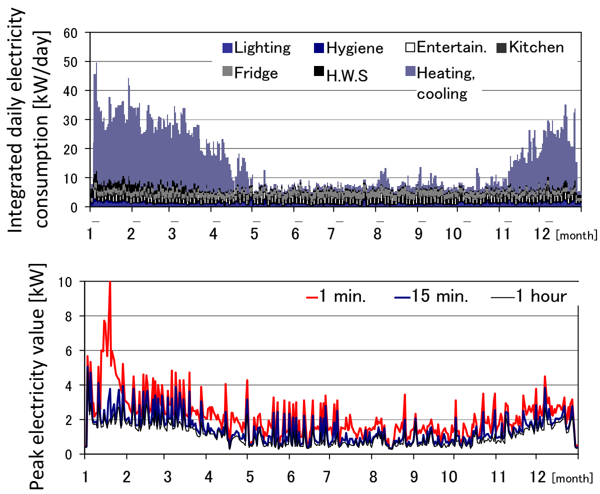


Figure 2-16: "Integrated electricity consumption in a day and long term fluctuation (Hokkaido 07) "

In all-electric houses, the peak energy consumption is very sensitive to the season. The use of electricity for heat may have a large effect on it. Especially in houses in colder parts of Japan, such as Hokkaido and Tohoku, where seasonal effects clearly lead to high heating energy consumption.

b) Combined energy source houses

Figure 2-16 indicates the integrated electricity consumption in a day and the long term fluctuation in Hokkaido 07. Daily peak electricity consumption fluctuates within one day, and there was no difference between each season. Figure 2-17 shows the data for Hokuriku 03. The winter energy consumption is relatively high, because this can be attributable to using air conditioners and electrically heated carpet for space heating. Japanese people often use electrically heated carpet as partial heating, and occupants sit on it directly. Despite peak electricity increasing in winter, daily variation is larger than that of the seasonal difference. The data for Kyushu 04 is shown in Figure 2-18.

Space heating is operated from December to April, while space cooling is operated from August to September. The peak value is around 3~4kW as space conditioning devices are used. In combined energy source houses, the variation of peak electricity consumption varies in a day is more significant than the seasonal variation.

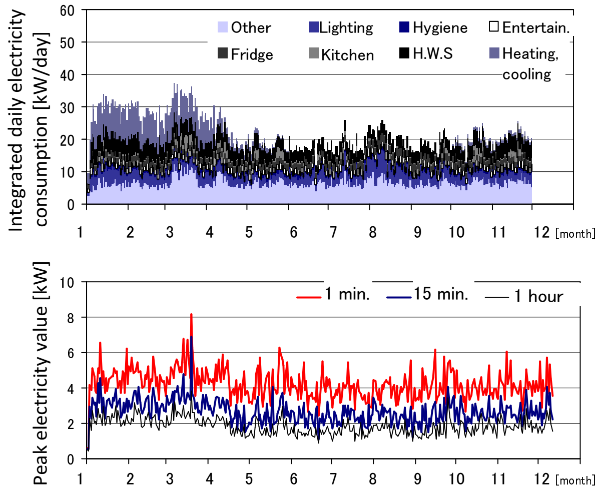


Figure 2-17: "Integrated electricity consumption in a day and long term fluctuation (Hokuriku 03)"

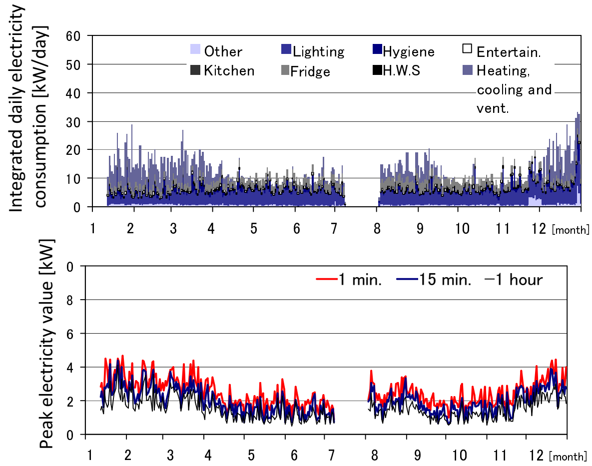


Figure 2-18: "Integrated electricity consumption in a day and long term fluctuation (Kyushu 04)"

Frequency distribution of peak electricity consumption of a year

Frequency distribution of the effect of varying the time interval on peak electricity consumption in a year was analyzed.

a) All-electric houses

Figure 2-19 indicates the frequency distribution of peak electricity in Tohoku 07. As the time period gets longer, the average value decreases and the distribution are gets narrower. In addition, there is a large difference between the most frequent value and the average. This is simply because the peak electricity consumption varies depending on the season.

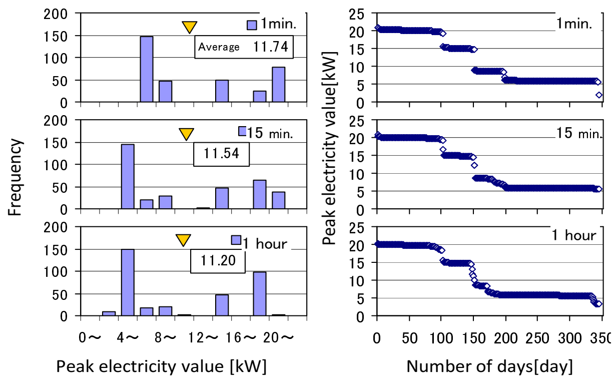


Figure 2-19: "Frequency distribution of peak value of a year (Tohoku 07)"

b) Combined energy source houses

Figure 2-20 indicates the result of Kyushu 04 which uses electricity, gas and kerosene. Likewise, for the all-electric houses, the distribution range gets narrower as the time period increases.

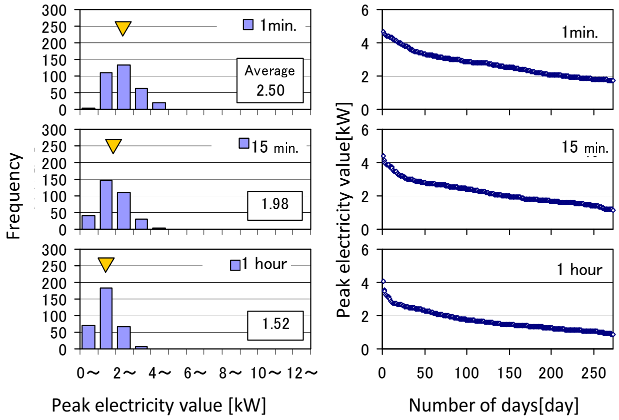


Figure 2-20: "Frequency distribution of peak value of a year (Kyushu 04) "

Long- term variation of daily load factor ratio

Daily load factor ratio means the ratio of the average electricity consumption per day and the maximum electric power. The formula for this computation is indicated as:

$$Daily\ load\ factor\ ratio = \frac{Average\ electricity}{Maximum\ electric\ power} \times 100 \qquad (1)$$

No matter how small the peak electricity value occurs, the supply side is required to cover the demanded electricity and provide a power generation scale to balance the request. As the peak value increase in the residential sector, the more the real demand will differ from the supply.

a) All-electric houses

Figure 2-21 indicates the variation of day load factor ratio in all-electric houses. Day load factor varies depends on the season and it decreases 10% to 20% as the season changes from winter to summer. Especially in Hokkaido 01 and Kyushu 06, the difference among each season is the biggest. The largest differences between two seasons were 36.5% and 41.1%, respectively. The lowest ratio of daily load factor of 8.9% occurred on September 9th in Hokkaido 01 and 8.7% on August 6th in Kyushu 06. In Tohoku 07, the value was below 23%, which is relatively small for a whole year.

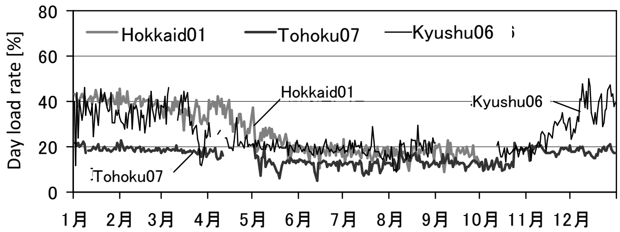


Figure 2-21: "Long- term variation of daily load factor in all-electric houses"

b) Combined energy source houses

Figure 2-22 indicates the variation of day load factor ratio in combined energy source houses. In Hokkaido 07, except for the New Year when residents are away from home and almost no energy consumption was shown, a day load factor ratio of around 20% was indicated throughout a whole year. In contrast, the ratio in Hokuriku 03 and Kyushu 04 varies widely in a day with difference of 56% and 43.7%. The lowest ratio of daily load factor of 9.7% occurred on May 20th in Hokuriku 03 and 6.7% on August 3rd in Kyushu 04.

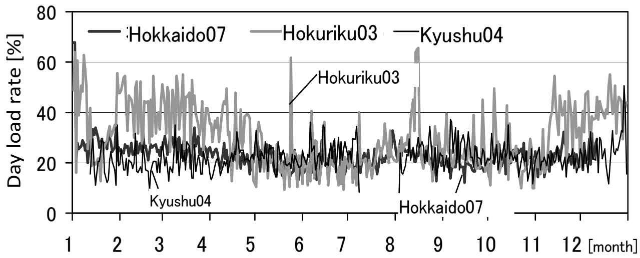


Figure 2-22: "Long- term variation of daily load factor in combined energy source houses"

1.2.5 **Conclusion**

(1) In all-electric-houses, the occurrence times of peak electricity in winter are concentrating at midnight due to the operation of hot water supplier and thermal storage heater.

(2) In combined energy source houses, the peak electricity consumption are distributed widely day by day while in all-electric houses it varies depending on the season. Therefore, the performance advantage of appliances is important in order to make the energy load flat.

(3) Use of electricity for heating has a great influence on increasing energy consumption. When excluding night-time electricity consumption, peak electricity was found when occupant use electric stove in all-electric house. Therefore, a strategy for increasing the efficiency of electric cooking devices is required in order to smooth the peak electricity consumption. For the long-term variation of annual peak electricity frequency, there was a seasonal difference in all-electric houses especially in houses which use electric space heating. On the other hand, there was a huge difference within one day in the combined energy source houses.

(4) Frequency distribution of peak electricity consumption a year is spread out in all-electric house because of the seasonal difference of electricity consumption..

(5) Ratio of the daily load factor decreases as the season changes from winter to summer in all-electric houses. In combined energy source houses a huge variation was found in a day.

Reference:

[1] Hiroshi Yoshino, Shuzo Murakami, Shin-ichi Akabayashi, Kazuaki Bogaki, Toshihiko Tanaka, Hirofumi Hayama, Akihito Ozaki, Hanako Sugawara: SURVEY OF THE PEAK ELECTRIC IN RESIDENTIAL BUILDINGS : Analysis of the data from survey of energy consumption for 80 houses in Japan, AIJ. Journal of environmental engineering(610), pp.99-106, 2006

1.3 Experience2: Application of the prediction model for building energy use assessment. Analysis of BEMS database by using multivariable regression analysis

1.3.1 Introduction

Principal component regression (PCR) and partial least squares regression (PLSR) present application of principal component analysis (PCA) in linear models. PCA can be used for different purposes. Some of applications are establishing relationships among data in databases and fault detection and diagnosis. Due to wide application of PCA, there have been developed software based on PCA (The Unscrambler® X [1]) and specialized calculation tools within MATLAB. PCR and PLSR were implemented to analyze building energy management system (BEMS) data and to relate them to the building energy use.

The idea was to encourage a smart use of BEMS data for energy use analysis. Databases of 76 and 41 variables, which included occupancy level, control signals, and water and air temperatures, were used to explain heating, electricity, and fan energy use. Variable contributions to the principal components (PCs) were used to simplify model and found the most important variables. This way, energy use was defined indirectly by using available variables in BEMS. The approach was tested on a low energy office building located in Trondheim, Norway. The suggested approach could be used by building operators to identify opportunities for decreasing energy use and for energy use estimation when data were lost due to data transmission problems.

In this study analyzed databases for prediction of heating energy use and electricity use were developed with BEMS data. Even though BEMS data are different in nature, temperature, control signals, pressures, etc., they can be correlated. For example in office buildings at the mid of day, outdoor temperature is usually higher, equipment is ON, occupancy level is higher in office building, while during the night time situation is usually opposite. In addition, BEMS data are correlated to time. Therefore, it could be useful to decouple data and establish new variables that would be uncorrelated. These new uncorrelated variables should be used to define building energy use. PCA may also be used to analyze time series, if variables of time are included as predictor variables.

1.3.2 Aim of the analysis

The aim of the study was to identify driving variables that contributed to energy use in low energy office building by integrating BEMS and energy use data.

1.3.3 Database structure

Available data from BEMS were used as predictor variables, while the heating energy use, the total electricity use, and the electricity for fans were used as target variables. BEMS data that were used for the predictor variable databases are given in Table 2-2. To effectively present variables, in Table 2-2 variables for only the first air handling unit (AHU) 36.01 are presented, because variables for the seven rest AHUs were the same except that their values were different in operation depending on use.

Table 2-2: Database description predictor variables

Variable name	**Description**	**Value range**	**Application**
Day	Day of week	1 for working, 2 for nonworking	H^*,E^*,F^*
Hour	Hour	0 – 1,	H^*,E^*,F^*
Tout	Outdoor temperature	-20 – 30 ^{o}C	H^*,E^*,F^*
Tin_R4031	Indoor temperature in the 4th floor office	18 – 23 ^{o}C	H^*,E^*,F^*

Tin_R4010	Indoor temperature in the 4th floor office		H*,E*,F*
Tin_R4099	Indoor temperature in the 4th floor office		H*,E*,F*
OCC_R4031	Occupancy level in the 4th floor office	0.5 (not occupied), 1 – 1.5 (bypass), 3 (occupied)	H*,E*,F*
OCC_R4010	Occupancy level in the 4th floor office		H*,E*,F*
OCC_R4099	Occupancy level in the 4th floor office		H*,E*,F*
320.SB40	Valve position in the main branch	0 – 100 %	H*
320.RT40	Supply temperature in the main branch	30 – 70 ^{o}C	H*
320.RT50	Return temperature in the main branch	30 – 60 ^{o}C	H*
320.02.RT40	Supply temperature in floor heating	20 – 35 ^{o}C	H*
320.02.RT50	Return temperature in floor heating	20 – 30 ^{o}C	H*
320.03.SB40	Valve position for snow melting	0 – 100 %	H*
320.03.RT40	Supply temperature for snow melting	20 – 35 ^{o}C	H*
320.03.RT50	Return temperature for snow melting	15 – 25 ^{o}C	H*
320.04.SB40	Valve position in the radiator branch	0 – 100 %	H*
320.04.RT40	Supply temperature in the radiator branch	30 – 70 ^{o}C	H*
320.04.RT50	Return temperature in the radiator branch	25 – 55 ^{o}C	H*
36.01.LK	Valve position at heating/cooling coil	0 – 100 %	H*
36.01.LV	Valve position at heating coil	0 – 100 %	H*
36.01.RT55	Return temp. after LV AHU	20 – 50 ^{o}C	H*
36.01.LX01	Input signal for recovery wheel	0 – 100 %	H*,E*,F*
36.01.JV40	Input signal for supply fan	0 – 100 %	H*,E*,F*
36.01.JV50	Input signal for exhaust fan	0 – 100 %	H*,E*,F*
36.01.RT40	Supply air temperature	16 – 24 ^{o}C	H*,E*,F*

H – heating energy use, E – electricity use, F – fan electricity use

Monitoring of the energy use in the energy service database was on hourly basis. Therefore, data from BEMS in Table 2-2 were calculated as hourly mean values. The column "Application" in Table 2-2 shows for prediction of which energy use the variables were used.

1.3.4 **Relevant influencing factors**

Much data can be measured via BEMS. In general, it can be assumed that all these data contribute to some extent to the building energy use. If the entire BEMS database is used in the PCR and PLSR to establish the energy use model, then all the BEMS data would be related to the building energy use. However, the BEMS data can be correlated and then a data redundancy could appear. To extract the most important data that contribute mostly to the building energy use, it could be beneficial to scale the models obtained by using PCR and PLSR. Model scaling from a model based on the database to a model based on 10 variables was performed based on the predictor variable contribution to PCs. Model scaling in this way was used to identify driving variables.

1.3.5 **Model type**

Partial least squares regression (PLSR) and principal components regression (PCR) are both methods to model a response variable when there are a large number of predictor variables, and those predictors are highly correlated or even collinear. Both methods construct new predictor variables, known as principal components (PCs), as linear combinations of the original predictor variables. Detail theory

behind the application and use of the PCA is explained in [2]. Important relations for this analysis are explained in [3].
To decrease number of variables and find the most influencing, the PLS weights and PC loadings were used. The PLS weights are the linear combinations of the original variables that define the PCs in the PLSR. Actually, they describe how strongly each component in the PLSR depends on the original variables. Similarly, the PC loadings describe how strongly each component in the PCR depends on the original variables.

1.3.6 **Results and discussion**

Use of the entire databases to calculate target variables could be demanding and requires specific computer programs to perform calculation. Smaller database of predictor variables with several variables could be very simple for practical use and presentation of influential parameters on the building energy use. To introduce approach for decreasing number of variables gradually, method effectiveness are presented first. Method effectiveness was estimated by using model accuracy. Accuracies for the heating energy use model for both regression methods and different amount of data are presented in Figure 2-23. In Figure 2-23, accuracy is presented by coefficient of variation of the root mean squared error (CV(RMSE)), which was estimated by using 10-fold cross validation.

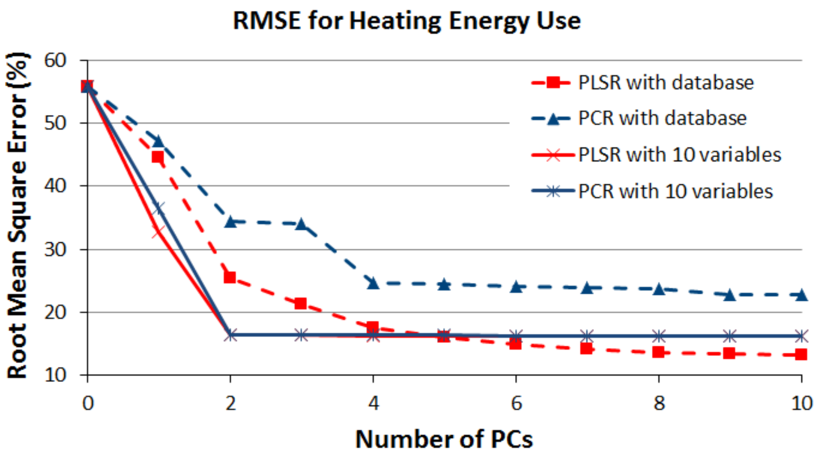


Figure 2-23: Accuracy of the heating energy use model for different amount of data and different methods

In Figure 2-23, the model accuracies are presented as function of number of used PCs to model the target variable. Models are declared to be calibrated if they produce CV(RMSE) within ±30% when using hourly data [4]. This means that the models developed by using the entire database in Figure 2-1 have acceptable accuracy when four PCs were used, while the models described with 10 variables produced acceptable accuracy already with two PCs. In Figure 2-23, it is also possible to notice that the 10 variable models had faster improvement then the models based on the database. This faster improvement of the simpler models indicated that there were redundancy and mutual correlation among variables in the database. In Figure 2-24, when comparing the regression methods, it is possible to notice that PLS regression has faster improvement than PCR, either by using the entire database or 10 variables. This result was expected, since in the PLS regression method PCs are obtained to directly

reflect the relationship between the predictor and the response [2]. In PCR method, the PCs explain only variation in the predictor variables, with no regard to the target variables.
In Figure 2-24 it was shown that the models with 10 variables could achieve acceptable accuracy and even faster improvement due to decreased redundancy. These 10 variables were chosen based on PLS weights.
To simplify models based on the entire databases and find the most influencing variables, values of PLS weights on the first four PCs were used. The first four PCs were used because of the results related to the percent of the variance explained in the predictor variables. 97 % of the model variance was explained in the first four PCs for the heating energy use model as shown in Figure 2-25. 99 % of the model variance was explained in the first four PCs for the electricity use model as shown in Figure 2-25. Therefore, the first four PCs were assumed to be enough for the analysis on the important variables.

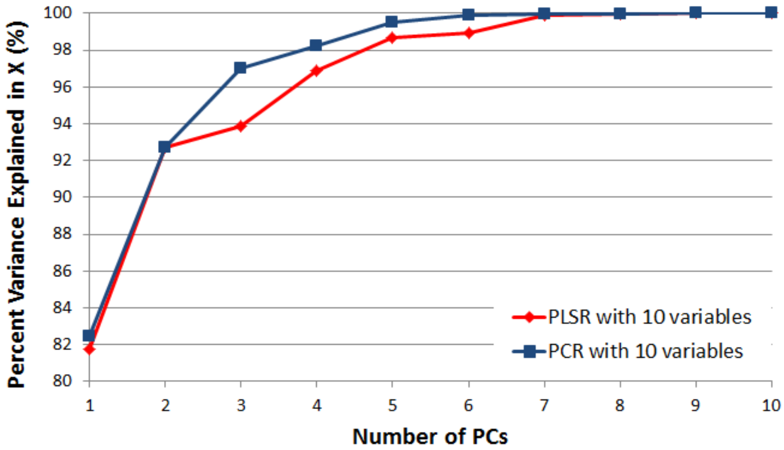


Figure 2-24: Percent variance for heating energy use with 10 variables

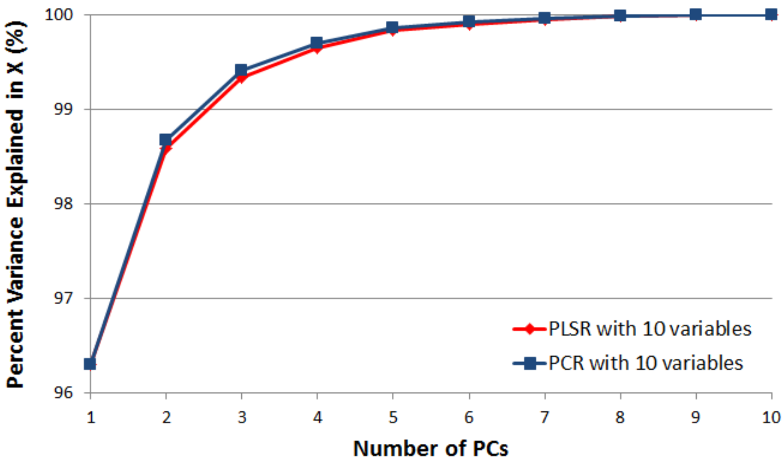


Figure 2-25: Percent variance of electricity use model with 10 variables

Procedure of defining driving variables consists of two parts. First, matrixes of the original variables defined in Table 2-2 were normalized. Afterwards, the first 10 variables that had highest contribution

to the first four PCs, were chosen as the model important or driving variables. The same procedure was repeated for each target variable, heating energy, electricity, and fan electricity use.
PLS weights on the first four PCs for the heating energy use, the total electricity use, and the fan electricity use are displayed in Figure 2-26, Figure 2-27, and Figure 2-28, respectively. If a variable has high contribution to PCs, then PLS weights have higher values, and consequently it could be concluded that that one contributes to the target variable. In Figure 2-27, Figure 2-28, and Figure 2-29 driving variables for energy use in November are shown.

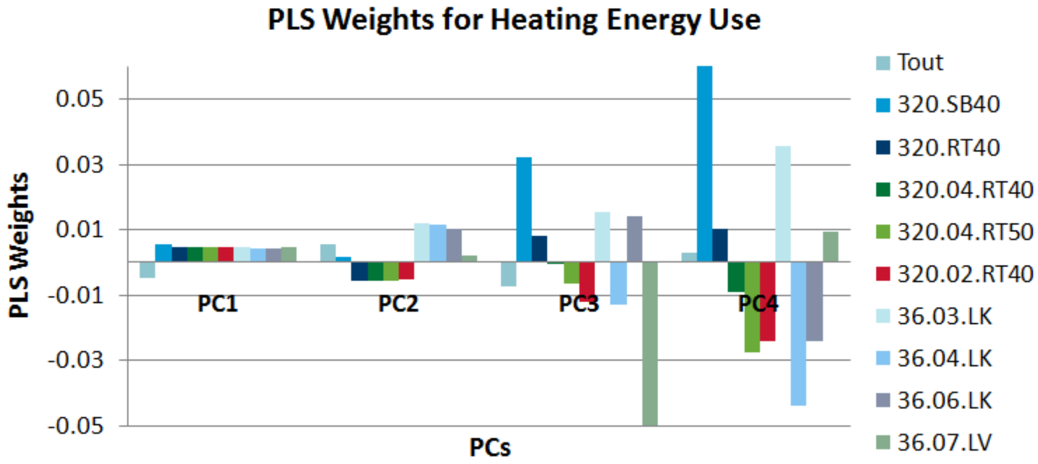


Figure 2-26: PLS weights of 10 important variables for heating energy use model

By using procedure for model scaling and finding driving variables based on PLS weights, it was found that the most important variables of the heating energy use are outdoor temperature, control parameters and temperatures in the substation, and some of ventilation parameters. These ventilation parameters were related to the AHUs that were mostly in use. In Figure 2-27, PLS weights of the different variables on the first and second PC had quite similar values, while on the third and fourth PLS weights were different. Therefore, the values of PLS weights on the third and fourth PC could explain variable importance. Based on that, it is possible to conclude that the heating energy use was influenced by the operation parameters rather than by the outdoor temperature.

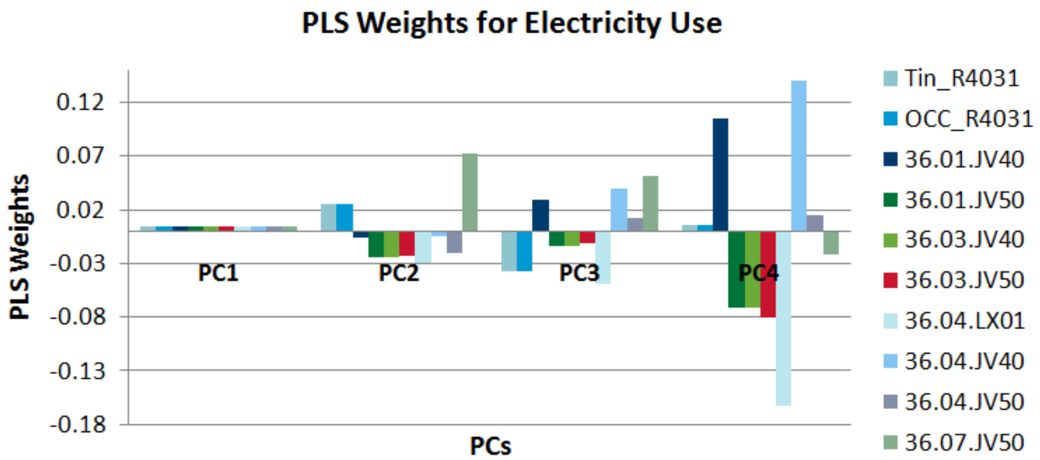


Figure 2-27: PLS weights of 10 important variables for electricity use model

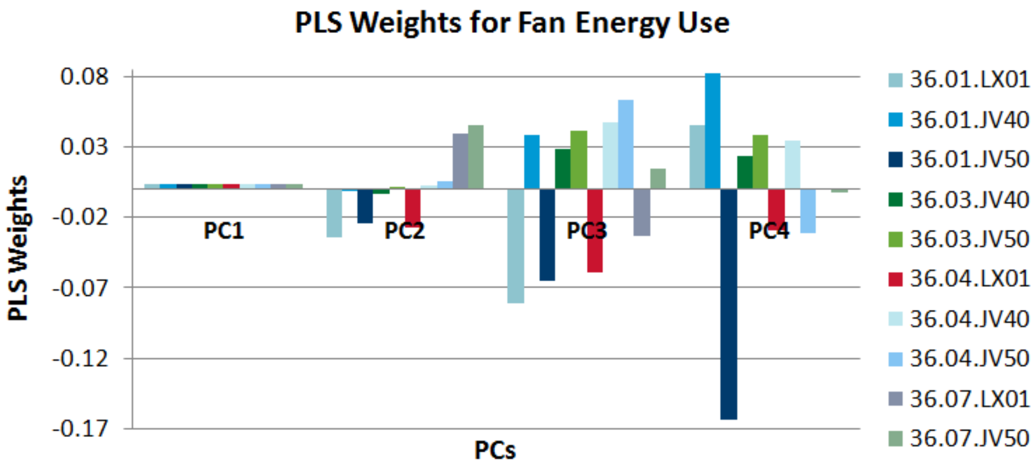


Figure 2-28: PLS weights of 10 important variables for fan electricity use model

Results in Figure 2-28 show variables that could explain the total electricity use in November. Among 10 variables, occupancy level, indoor temperature, and parameters of the fourth AHU were variables that could explain the electricity use well. The fourth AHU was supplying the most typical occupied part of the building. Results in Figure 2-28 shows that occupancy level could be included in the electricity use model. In this model, occupancy level and indoor temperature had significant contribution already on the second and third PC. Negative values of PLS weights of the occupancy level and indoor temperature should not imply directly negative influence on the target variable, because the original variable matrix was normalized. In the case of normalized original variable matrix, values of PLS weights should indicate variable importance to the model.

1.3.7 **Conclusions**

The PCA application in linear models was implemented to utilize BEMS data for energy use estimation and identifying of driving variables of energy use. The idea was to relate building information with the building energy use. PCR and PLSR were used to relate BEMS data to the building energy use. BEMS data were used as original predictor variables, while the heating energy use, the electricity use, and the fan electricity use were target variables. To simplify models and find the most influencing variables, values of PLS weights on the first four PCs were used. The suggested approach was tested on the low energy office building, located in Trondheim. The analysis showed that it is possible to utilize PCR and PLSR to analyze BEMS database. The methods are robust and it is possible to perform different analysis. In this analysis, these methods were implemented only to relate BEMS data to the building energy use and to identify important variables. The important variables were identified by scaling the models. In this case, the model scaling meant that the models based on the entire BEMS database were scaled to models with fewer variables.
The results showed that the heating energy use in the low energy office building was influenced by the operation parameters rather than by the outdoor temperature. The total electricity use could be explained by using occupancy level, indoor temperature, and some of the AHU electrical signals. The AHU electricity use could be explained by using the input electrical signals of supply and exhaust fans. However, results indicated that the regression models should be updated on monthly level. All the

simplified regression models with 10 variables had acceptable accuracy. This indicated that driving variables obtained by using suggested approach could be used to explain building energy use.
Further, this approach could indicate possible reasons of change in building energy use. Current results showed that PLS regression method was more accurate in recovering the heating energy use, while PCR was more accurate in recovering the electricity use. The results showed that important variables were different for different months in the case of heating energy use. The total electricity and fan electricity use could be defined with the same variables in different months. The total electricity use could be defined by using occupancy level and input fan signals.

1.3.8 **References**

[1] CAMO, The Unscrambler® X.

[2] Jackson, J.E., A user's guide to principal components2003, Hoboken, New Jersey: John Wiley. 569 s.

[3] Djuric, N. and V. Novakovic, Identifying important variables of energy use in low energy office building by using multivariate analysis. Energy and Buildings, 2012. 45: p. 91-98.

[4] Measurement of Energy and Demand Savings. Vol. Guideline 14 2002: ASHRAE.

1.4 Experience 3: Modeling Occupant Behavior in multifamily houses (Marcel Schweiker)

1.4.1 Subject of the work

This work is dealing with the frequency of usage and the set-point temperature selection of AC-units for cooling and heating in an international student dormitory in the Tokyo area. Both parts are directly related to the energy used within the building. The setting of the measurement of 15m^2 rooms with multiple possibilities to adjust the indoor conditions, described in detail in [1], allows the observation of the individual occupant's behavior in a laboratory-like setting.

Table 2-3: Data on building

Type of building	International student dormitory
Dimension	320 rooms of 15 m^2 each
Location	Tokyo, Japan
Thermal characteristics	Low level of insulation and single-glazing windows with aluminum frame
Type of observed spaces	Single-room apartment
Year of construction	1989
No. of floors	5
Windows, orientation	One window each (either E, S, W)
Window opening	sliding
Shading devices	Fixed overhang
Sources of heat gains	Fridge, computer, lights, occupancy
Activity, sex and age of occupants	Students (M or F, 20-35)
Origin of occupants	27 different countries from 5 continents

1.4.2 Building characteristics

The building was opened in 1989 and is a 5-storied building with 320 identical single rooms (see Figure 2-29). The construction is made of concrete with little thermal insulation and single glazed windows. The single rooms of 15m2 each including the bathroom are oriented to east, south or west. Each room has one door facing to the corridor and one window with curtain on the opposite side, and is equipped with one air-conditioning unit for heating and cooling. The residents are free to use electrical fans or other measures to keep their rooms as comfortable as possible without using the air-conditioning unit.

Figure 2-29: View into one of the rooms.

1.4.3 Aim of the work

In reality, the occupant behavior is expected to be influenced by quite a large number of factors, both external and internal. In order to apply the important findings related to occupant behavior in disciplines such as social sciences and neural sciences to the empirical models based on the measured data within the field of building sciences, this work describes an empirical determination of the external and internal factors and their constants for the case of air-conditioning usage in a residential setting under Japanese climatic conditions. Knowing such model and the influence of each factor on the behaviors will be useful on one hand to design buildings, especially residential buildings, supporting the occupants towards a less exergy consuming lifestyle with sufficient well-being and on the other hand to help support the occupants in the existing buildings towards also the less exergy consuming lifestyle.

1.4.4 Database characteristics

Table 2-4: Database characteristics

Number of buildings	1	
Period of measurement	29.06.2007 - 13.08.2007	11.01.2008 - 11.02.2008
Duration (days)	46	32
Number of observed spaces	39	34
Number of observed spaces with window sensors	19	24
	Items	Interval
IF1. Climate	Outdoor air temperature, humidity, wind speed, solar radiance	2 minute
IF2. Building envelope	Not in database	
IF3. Building service & Systems		
IF4. Operation & Maintenance		
IF5. Indoor environmental quality	Indoor air temperature, humidity	2 minute
IF6. Occupants' activities and behavior	State of AC-unit (on/off/set-point temperature)[1)]	Event

	Window state (open/closed)	
IF7. Social and economical aspects	Occupants personal preferences, individual characteristics (age, sex, height, weight), personal background (country of origin, sleeping habits during childhood, ...)	

1) not measured, but derived from data

The data base used for the present investigation comes from two measurements in above described building. The two measurements were conducted around the hottest weeks of Japanese summer in 2007 from 29 June to 13 August and the coldest weeks in winter, early 2008 from 11 January to 11 February. Figure 2-30 shows the frequency of prevailing outdoor air temperatures both in summer and winter.

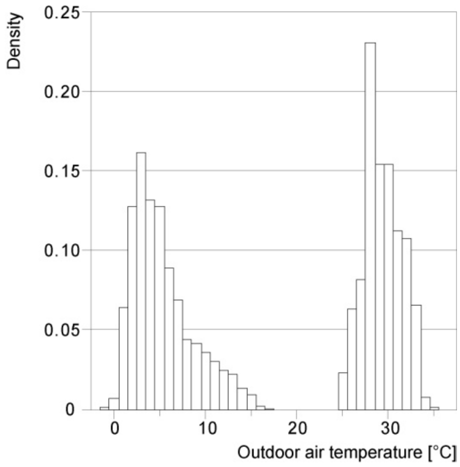


Figure 2-30: Distribution of outdoor air temperature during the summer measurement in 2007 and that during winter early 2008.

For the physical measurements, one wireless temperature and relative-humidity sensor and another wireless sensor logging the times the window was opened or closed were installed in each of the 39 observed rooms. As well as this, additional sensors were installed to measure the outdoor air temperature, relative humidity, solar irradiance and wind speed (wind speed was only measured for the summer case). The occupancy could not be explicitly recorded during the daytime, but the students stated the nights they were not sleeping in their rooms as well as continued absence for more than a day. The measurement took place in all observed rooms at the same time with a logging interval of 2 minutes. This garnered 30325 data sets for each room for the six-week summer measurement, and 22080 data sets for the four-week winter measurement. These physical measurements were accompanied by an introductory written questionnaire survey and a personal interview, which included questions about the students' current and past cooling and heating behavior, their thermal background, lifestyle, preferences, knowledge of passive heating and cooling strategies and personal evaluation of the effectiveness of those strategies.

1.4.5 Method/Methods applied for the data analysis

The model development consisted of three steps. The models derived in each steps were called standard, advanced and final models for winter and summer season, respectively. For the standard

models, the use of the AC-unit for cooling and heating in relation to the mean outdoor air temperature was analyzed using the logit model. The analysis was done for the whole day period of 24hours, and then for three different periods within the day, namely daytime (8am to 6pm), evening (6pm to 12am) and night-time (12am to 8am). The parameters of the model were calculated with the statistical software package R as a function of the mean outdoor air temperature, based on the two-minute data points of the respective period of time.
For the advanced models, factors other than the outdoor air temperature during the night in question, were analyzed.

The general form of a logistic model is expressed as follows:

$$p_{log} = \frac{1}{1+\alpha e^{-(\beta_1 T_{oav} + \sum_{j=2}^{i} \beta_j x_j)}} \quad (1)$$

where, p_{log} is the probability that the AC-unit is used, T_{oav} is the mean outdoor temperature of the respective night, x_j are additional factors and α, β_1 as well as β_j are constants to be determined by statistical analysis. The simplest model, which was used for the standard model, is the one with a single parameter of mean outdoor temperature only, namely the model with $\beta_j = 0$ except for j=1. For the advanced models, those factors having a coefficient signaling an above-average linear relationship with the AC-unit usage were taken as the starting point for the selection process of a rational statistical model.
As the decision criterion during the selection process, Akaikes Information Criterion (AIC) [2] was used to determine whether the alternative model is better than the current model. The AIC is calculated by taking into account the fit of the model compared with the data together with the number of variables used in the model. The lowest AIC-value is supposed to be calculated for the model which best describes the measured data with the minimum number of variables necessary. When comparing two models, only the absolute difference Δ_{AIC} between their AIC values should be evaluated and not the absolute values themselves. In order to define one model as being better than the other, Δ_{AIC} should be greater than 2 [3]. Because of quite a few possible combinations of variables to define a model, the "stepAIC" function within the software R, which automatically selects the model with the smallest value of AIC, was used [4]. In addition, Nagelkerke's R^2 index was calculated, which was adapted to mimic the R^2 analysis for logistic regression [5] to have a second index when comparing the different models.

For the final models, having the same form as shown in eq. (1), the advanced models were amended with individual factors in order to show their influence on the behavior of AC-unit usage at night-time following the findings of [6]. Due to the fact that none of the students grew up in Japan, it was possible to include a variable related to different climate groups referring to their region of origin. This was called "thermal background". Using the climate map of Koeppen [7], the students were sorted into four groups: hot and dry, hot and humid, moderate, and cold climates. There was no student from a polar climate region.
The "stepAIC" function was used again in order to reduce the number of variables implemented into the logit model for predicting the percentage of AC-unit usage in relation to external and individual factors.

1.4.6 Results

Influence of season and time of day

Figure 2-31 shows the relationship between the average outdoor air temperature and the percentage of persons using the AC-units for cooling. Each plot represents the average of all students for each of the 43 measured days in summer. The lines represent the best-fit of these plots and are similar to those presented in the foregoing studies for window opening behavior [8,9]. Compared to the curve presented by Nicol and Humphreys [10], where the value of temperature resulting in half of the persons using AC-units (T_{50}) is around 29°C, we get a similar value for the case of daytime, but a much lower value of T_{50}, 23 to 26°C, which in turn means a more frequent use of the AC-unit for all other periods of the day.

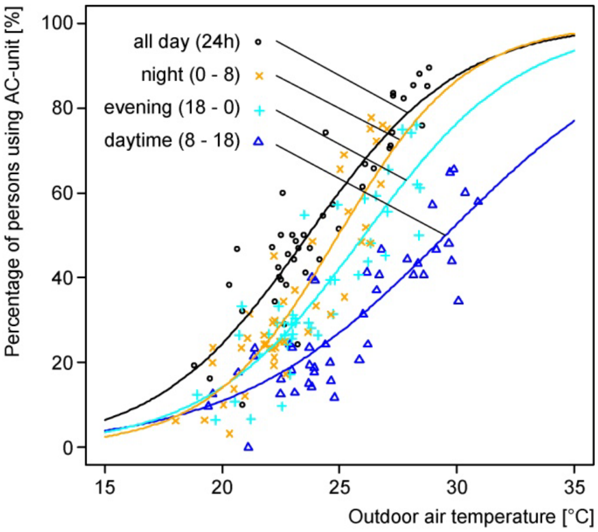


Figure 2-31: Relationship between outdoor air temperature and percentage of persons using AC-units for the case of cooling in summer time. The outdoor air temperature is the value for respective period as shown inside the graph. The thin dotted lines along with the bold lines in between show the limits of the 5% confidence interval of the respective line.

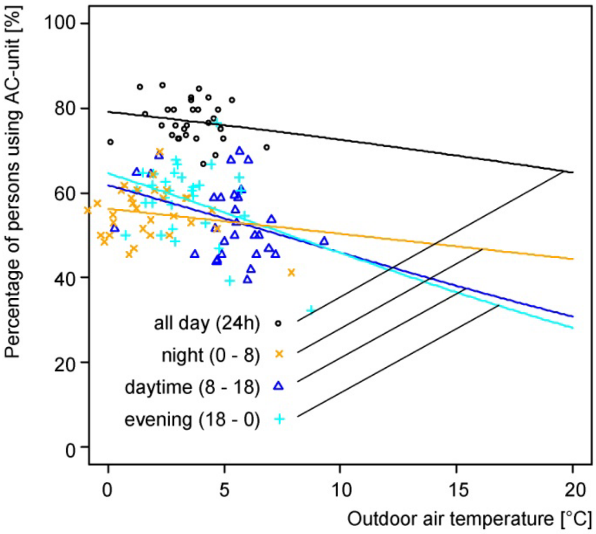


Figure 2-32: Relationship between outdoor air temperature and percentage of persons using AC-units for the case of heating in wintertime. The thin dotted lines along with the bold lines in between show the limits of the 5% confidence interval of the respective line

Looking at the difference between the times of the day, it is clearly visible that the AC-unit is used more during night-time than daytime. This can be explained by the lifestyle of the students, who have to leave their rooms in order to go to university or work. Due to the circumstances of this study that the real occupancy was known only for the night-time, the lines for other periods must be too low if we assume an occupancy rate of 100%. For the prediction of a similar type of building, these lines can be used as combination of occupancy and behavior. The analyses of all other models to be described later were based only on the night-time period, where the true occupancy rate was known.

Figure 2-32 shows the logit lines for the heating case. While the logit curve presented by Nicol [11] for heating with general heating devices approaches 100% around a value of mean outdoor

temperature of 2°C, the curves shown here are much flatter. Especially the flatness of the curve for the night period shows that around 40% of people sleep without any heating system regardless of the outdoor temperature and adjust themselves with additional layers of clothes and blankets. This statement can be partly supported by the answers given to the questionnaire survey. The question whether one prefers to sleep in a room a) heated with an AC-unit, b) heated with other means than an AC-unit or c) not heated at all, was answered by more than 33% of the students with c). This number is a bit lower than the 40% obtained from the measured data, but still in the same range.
By comparing both figures it can be seen that the percentage of AC-unit usage is much more dependent on the outdoor conditions during summer time than winter time. The most obvious reason is the greater possibility of clothing adjustment during wintertime than in summer time, although clothing adjustment was not part of our survey. Further reasons are the special conditions within this dormitory, which has in general a higher indoor temperature than other buildings due to the high density of people, lights and electronic appliances. The latter reason also applies to the comparison with the logit lines presented by Nicol and Humphreys [12].

Influence of foregoing nights

The following results were derived from a closer look at the decision to sleep with the AC-unit on or off. This was in particular interesting to investigate, as it is a decision made before knowing how the thermal conditions will change during the sleeping period and results in a different amount of exergy consumption for cooling and heating during each period.
According to the statistical analysis of various external factors and their influence on the AC-unit usage, the best prediction was achieved by considering the mean outdoor air temperature of the night in question T_{oav}, the mean outdoor temperature of the first and third night before, $T_{oav\text{-}n1}$, $T_{oav\text{-}n3}$, and only for the winter case the mean outdoor temperature during daytime just before the night in question $T_{oav\text{-}dt}$. Table 2-5 shows the corresponding values of constants (α, $ß_1$-$ß_4$) and the comparison between the standard models and the advanced models. The advanced model for the summer case shows a better fit to the measured data; the difference in AIC value, Δ_{AIC}, is 17.9 and also the R^2-index is a little bit higher, even though it is still very low. Despite a high variation in relative humidity during the measurement period between 46% and 99%, the tested factors concerning humidity, namely the mean relative humidity of the respective night, the evening and daytime before as well as the night beforehand, did not show any linear or other correlations with the usage of the AC-unit.

Table 2-5: Comparison of standard and advanced models to predict the percentage of persons using their AC-units during night-time in summer and winter

		α	$β_1$ [1)]	$ß_2$	$ß_3$	$ß_4$	AIC	R^2
summer	standard	10915±0.67	0.374±0.03	–	–	–	1571.8	0.184
	advanced	71663±0.82	0.245±0.42	0.120±0.04	0.0914±0.03	–	1553.9 [2)]	0.203
winter	standard	0.888±0.09	-0.0286±0.04	–	–	–	1250.1	0.001
	advanced	0.657±0.20	0.0147±0.04	-0.0221±0.04	-0.0382±0.03	-0.0469±0.04	1252.1 [3)]	0.007

1) All of βj are related to the mean outdoor temperature: β1 for respective night; β2 for one night before; β3 for three nights before; β4 for daytime

2) ΔAIC, for summer: |1571.8 - 1553.9 | = 17.9

3) ΔAIC, for winter: |1250.1 - 1252.1 | = 2

In the case of winter, the advanced model has a little higher AIC value compared to the standard model; it is not possible to conclude which model describes the data better, since the difference in AIC value between standard and advanced, 2, is too small. The R^2-index is also extremely small. For simplicity, $ß_1$ to $ß_4$ were combined into one variable, the running mean outdoor temperature.

Influence of individual factors

The final models include the temperature related factors and several individual factors. The variables and parameters of the final models for summer and winter can be found in [13]. A separate discussion of each factor is given in Appendix B of [14]. In order to be able to judge how much one factor in the model influences AC-unit usage behavior, the importance of one factor compared to the others can be calculated as the product of the absolute value of the coefficient (the respective value of ß) and the range of the variable. This product was called the "importance value".

Comparing the AIC values of the models, the difference in AIC values Δ_{AIC} between the advanced models and the final models as well as between the standard models and the final models are exceeding 300 in the case of summer and 400 in the case of winter. As stated above, a value of 2 is necessary to be able to declare one model as better than another, therefore the result shows that the advanced model leads to a much better prediction of the individual occupant behavior by considering various individual factors. This is also supported by the R^2-indices that are both in a range between medium and large for the final models.

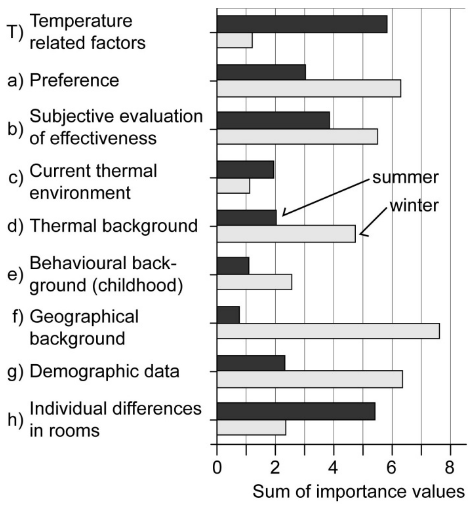


Figure 2-33: Comparison of the sum of importance values for groups of variables

The results of summing up the importance values of the respective factors shown are presented in Figure 2-33. Quite different results for summer and winter can be observed. In summer, the most important groups are "T) temperature related factors" and "h) individual differences in rooms", while on the other hand, the least important are "e) behavioral background" and "f) cultural background". In winter the most important are "f) cultural background" and "g) demographic data", while the least important are "T) temperature related factors" ($ß_3$-$ß_4$) and "c) current thermal environment". This shows that the external factors have a strong influence on the behavior in summer, but a very small in winter. Similar findings are present in the design of the new European Standard EN 15251, where there is no further change in the comfort limits below an outdoor temperature of 10°C.

The sum of importance values of “T) temperature related factors“ together with “h) individual differences in rooms” and that of all other factors turned out to be 11.2 and 14.6 for summer and 4.0 and 34.4 for winter, respectively. This shows that factors generating from experience, attitude and origin affect the reference level as much as the external conditions in summer and even much more in winter.

Application to set-point temperature

Analyzing the data from the field measurement it was also found that there is a rather huge difference between individuals in the indoor air temperature preferred and chosen. To analyze these differences, a linear multiple regression analysis was done for the same factors that appeared in the final models described above.

One would normally define the desired room air temperature to be the temperature chosen by the occupant and displayed on the remote control (also called the set-point temperature). However, the set-point temperature was neither observed within the measurement nor asked in the surveys. Even though the indoor air temperature is especially in the first minutes after switching on the AC-unit different from the set-point temperature, it is reasonable to assume the room-air temperature being similar to the set-point temperature for the steady-state case, which assumes that there was no change in the status of the AC-unit. Additionally, one can say that the indoor air temperature, not the set-point temperature, must be the temperature judged by the occupant as being comfortable, because otherwise they would have changed the status of the AC-unit with the remote control.

Therefore, the decision was made to use the indoor air temperature instead and to analyze the data from the student rooms according to the average of the maximum air temperatures during one period of the AC-unit usage. This led to the determination of the constants for the linear formula of the indoor air temperature as a function of the same factors as present in the final models:

$$\theta_{ind} = a + b_1 \theta_{mean} + b_2 x_2 + \cdots + b_j x_j \qquad (2)$$

The values for the coefficients can be found in [13]. With the result of linear regression analyses, a direct interpretation of the coefficients can be done. In such a way, it can be seen, that the choice of set-point temperature is influenced by the outdoor conditions together with individual preferences and characteristics. For each degree rise in the outdoor temperature, the set-point temperature increases by 0.43°C, while for example those who stated to prefer sleeping with an AC-unit on, maintain it at more than 2°C lower than those who prefer not to have it switched on during nighttime. In winter, the outdoor conditions have a neglectable influence on the choice for set-point temperature; each degree rise in the running mean outdoor temperature leads to an increase of the set-point temperature of only 0.00012°C. On the other hand, the influence of individual preferences increases; the before mentioned factor related to the preferences leads in this case to an increase in the set-point temperature of 4.8°C. It must be easy to imagine, that such difference leads to corresponding variations in the energy usage as well.

Effect on energy usage

The obtained regression models can be used directly for simplified steady-state calculations of the energy usage for AC-unit usage. Please refer to [14] for a detailed description of the calculation procedure.

Based on the regression models, one can assume two types of occupants, one preferring to sleep with an AC-unit switched on, the other not. As found by the approach presented above, the one who does not like to sleep with an AC-unit is still using it when necessary. However, the usage has a lower frequency and in summer a higher and in winter a lower set-point temperature. The choice towards a lower indoor room air temperature in winter leads thereby to a reduction of the energy usage by 25% to 30%. The combined effect of indoor air temperature choice and percentage of AC-unit usage leads to a reduction of 40% to over 90% and strongly depends on the outdoor air temperature. The reduction is in winter larger with higher outdoor temperature and in summer with lower outdoor temperature.

1.4.7 **Discussion**

Regression analyses were chosen in order to analyze the influence of physical and individual factors on the frequency of AC-unit usage for cooling and heating as well as the chosen set-point temperature. This method allows a straight-forwarded combination of factors in a continuous form (e.g. temperature levels) with those being in a binary form (e.g. gender) into one combined model. Limitations are in the individual modulation of each variable. I.e. a variable can either influence on the outcome in a positive manner over its full range or in a negative manner, but e.g. not in a U-shaped manner. To overcome this limitation, a variable could be included not only in its first power, but also in its second, third, … power. One of the main potential of this application is, that the outcome of the regression analysis (an equation to calculate a dependent variable based on individual variables together with their coefficients) can be easily implemented in advanced simulation tools, such as IDA-ICE or TRNSYS.
Using the AIC-value for model selection assures models, which have a good fit, while being as simple as possible. Nevertheless, such procedure is valid for nested models only, so that it can be used as done in this case to decide whether a variable should be included into a model or not, but not to compare two distinctive models. Calculating the importance values for each variable in the model, the magnitude of the influence on the outcome variable (here frequency of AC-unit usage) can be compared between all variables. The importance values appeared to be easy in its application, but strong in its interpretation. It could be therefore used in order to evaluate directly the influence of various factors on the energy usage; and not only as done here directly over the occupant behavior.
In conclusion, the methods applied here to models for AC-unit usage, could be used to derive models for energy usage, including a well established procedure to decide on the acceptance or rejection of single factors and statements with respect to the magnitude each factor has on the outcome.

1.4.8 **References**

[1] Schweiker, M. and Shukuya, M. (2009), Comparison of Theoretical and Statistical Models of Air-Conditioning-Unit Usage Behaviour in a Residential Setting under Japanese Climatic Conditions; Building and Environment 44: p. 2137-49.

[2] Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, BN (Ed.). *Proceedings of the Second International Symposium on Information Theory*; 1973: 267-81.

[3] Burnham KP, Anderson DR. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research 2004;33(2): 261-304.

[4] Venables WN, Ripley BD. *Modern Applied Statistics with S*. Fourth edition (Ed.). Springer; 2002.

[5] Nagelkerke NJD. A Note on a General Definition of the Coefficient of Determination. Biometrika 1991;78: 691-2.

[6] Schweiker M, Shukuya M. Investigation on the Relationship Between Occupants' Individual Difference and Air-Conditioning Usage During Nighttime in Summer. Journal of Environmental Engineering 2008;73(633): 1275-82.

[7] Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World Map of the Koeppen-Geiger climate classification updated. Meteorologische Zeitung 2006;15: 259-63.

[8] Haldi F, Robinson D. On the Behaviour and Adaptation of Office Occupants. Building and Environment 2008;43: 2163-77.

[9] Rijal HB, Tuohy P, Humphreys MA, Nicol JF, Samuel A, Clarke J. Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings. Energy and Buildings 2007;39(7): 823-36.

[10] Nicol FJ, Humphreys MA. A Stochastic Approach to Thermal Comfort – Occupant Behavior and Energy Use in Buildings. ASHRAE Transactions 2004;110 Part II: 554-68.

[11] Nicol JF. Characterising Occupant Behaviour in Buildings: Towards a Stochastic Model of Occupant Use of Windows, Lights, Blinds heaters and fans. In: *Proceedings of the Seventh International IBPSA Conference, Rio 2*.International Building Performance Simulation Association; 2001: 1073-8.

[12] Nicol FJ, Humphreys MA. A Stochastic Approach to Thermal Comfort – Occupant Behavior and Energy Use in Buildings. ASHRAE Transactions 2004;110 Part II: 554-68.

[13] Schweiker, M. (2010), Occupant behaviour and the related reference levels for heating and cooling – Analysis of the factors causing individual differences together with the evaluation of their effect on the exergy consumption within the residential built environment, http://www.buildup.eu/publications/9079, Dissertation.

[14] Schweiker, M. and Shukuya, M. (2010), Comparative Effects of Building Envelope Improvements and Occupant Behavioural Changes on the Exergy Consumption for Heating and Cooling. Energy Policy, 38(6), 2976-2986

1.5 Experience 4: Analysis of the occupant behavior in relation to the energy concept of an office building

(Fatma Zehra Çakıcı, Karin Schakib-Ekbatan, Marcel Schweiker

1.5.1 Introduction

Office buildings represent an important part of our living environment. With respect to energy consumption, occupant satisfaction and behavior are worthwhile issues in the context of sustainable office buildings with innovative technologies and materials [11],[6],[8]. The experiences show that there is often a gap between the predicted energy consumption based on simulation and the consumption during the day-to-day operation once the building is in use. Within the complex bundle of aspects such as design, construction, maintenance and occupant expectations the occupant behavior might not fit with the energy concept and cause counterproductive behavior. The “Desire for Control” [2] over ambient environmental conditions such as temperature or indoor air quality however has a strong impact on the well-being of the employees [10]. The understanding of the relationship between building and user behavior therefore plays an important role in the consideration of the energy consumption. Thus, the focus of this work is to explore patterns of energy-related behavior such as window-opening at the workplace, which is the most favorite taken adaptive opportunity [1]. Perennially gathered data for outdoor and indoor climate as well as occupant behavior from a monitoring project in a German office building with a passive cooling concept are analyzed [5]. In contrast to the other works, which are using complex statistical analyses in order to analyze the reasons for the discrepancy between predicted and monitored energy consumption, this work follows a different strategy. First, simple statistics of occupant behavior are shown to analyze differences due to the orientation of the room and the occurrence of not optimal behaviors. Second, logistic regression analyzes are presented in order to analyze the influence of indoor and outdoor conditions.

1.5.2 Subject of the Work

The Ostarkade is an extension of the building complex of the KfW Bankengruppe based in Frankfurt am Main, the largest city in the German state of Hessen. Frankfurt is located in central Germany with a temperate-oceanic climate with relatively cold winters and warm summers. The building is naturally ventilated and cooled in summer with a nighttime ventilation concept. Above the two-level underground car park, the building has five floors of mainly offices and meeting rooms, hosting about 350 employees. On the fourth floor, there are group offices for exchange traders, while the north-west part of the building is used for special purposes. Offices and apartments are grouped around an atrium with glazed roof. This allows natural lighting for the traffic zones which lead into the atrium. To make enough space available for a large conference hall on the ground floor, the atrium begins on the first floor.

Table 2-6: Building characteristics

Name of the building	The KfW Ostarkade
Type of building	Multi-storey office building
Dimension	17402 m^2 (8585 m^2 heated)
No. of Employee	~350 employees
Location	Frankfurt, Germany
Thermal characteristics	Low energy standard of building envelope (U-values walls 0.24 to 0.5 W/m2K, windows 1.5 W/

	m2K))
Type of observed spaces	Office rooms
Year of construction	2002
No. of floors	2-level underground car park + 4 office floors + 1 floor apartments on top
Windows, orientation	Mostly E and W
Window opening	Tilt-and turn
Shading devices	External sun protection (automatic + occupant driven mode)

The construction of the multi-storey office building was completed in 2002. The building shows an ambitious integral planning concept. From the beginning, the KfW Bankengruppe outlined very strict design criteria in terms of energy efficiency since the group is Germany's largest development bank with funding programs for energy efficiency in existing and new private buildings. Therefore, concepts for saving water and energy optimization for heating, cooling, ventilation and lighting were developed for the Ostarkade building.

Figure 2-34: Exterior view of the KfW Ostarkade building from the South-East

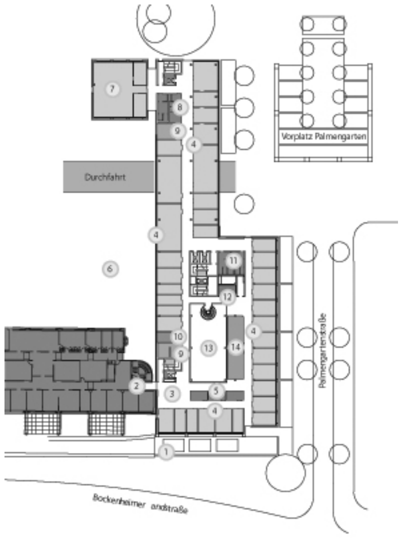


Figure 2-35: Ground floor plan of the building

The building's structural system is a reinforced concrete construction. . The façade is insulated externally and has a U-value between 0.24 and 0.5 W/m²K. The south facing façade of the building is a double-skin façade for noise reduction reasons. The roofs have a foam-glass insulation and are equipped with roof greening to a large extent. The compactness of the building and the high insulation standard minimize its transmission heat losses. With an average U-value of 0.54 W/m²K (façade including windows), the building exceeds the requirements of German Energy Saving Standards of the year 2002 by approx. 30%. In the standard offices the concrete ceiling is directly exposed and merely surfaced with a thin layer of plaster. This thermal mass increases the building's thermal inertia and was essential for the passive cooling concept with night ventilation. Pipes, cables and ducts are laid in elevated floors.

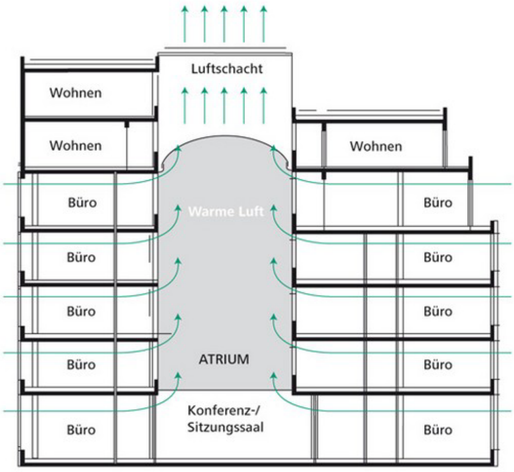


Figure 2-36: Natural ventilation via the atrium is used for night ventilation

Figure 2-37: The atrium lights the traffic areas and enables natural ventilation

A moderate percentage of glazing, an exterior automatic shading system and the use of solar control glass reduce the solar heat gains from outside. High windows, light-diverting blinds and the proximity of workplaces to the window enable very good use of daylight, thereby reducing the amount of electricity required for artificial lighting. The insulating glazing that is used has a less than 40% of total solar energy transmittance, a U-value of 1.5 W/m²K and 70% light transmittance.

Figure 2-38: Two-part sun protection enables glare-free use of daylight

Figure 2-39: Double-skin façade facing the very busy road

The building combines a high energy standard with high occupant comfort. The natural ventilation of the offices and the night ventilation cooling concept have proven good performance in practice. Even in 2003, the hottest summer in Germany within the last hundred years, comfortable conditions were maintained in the offices without mechanical ventilation or active cooling. This required appropriate manual ventilation with windows and the correct use of the sun protection. Excluding the energy consumption for the building control equipment and IT as well as for other specific technical services, the annual primary energy consumption was less than 100 kWh/m² in the third year of monitoring which is close to the predicted value of 107 kWh/m². Thus the energy consumption is well below that of conventional office buildings and in this case shows a good fit between predicted and measured values.

Figure 2-40: Offices with operable windows and sun protection, allowing natural ventilation and natural lighting

Window control options

The building management system (BMS) which automatically operates the top lights in the façade and the top lights facing the corridor, provides intermittent flush ventilation for 10 minutes in the morning before working hours. After that the occupants have the following control options: Windows can be opened completely and also tilted by hand. The top lights can be opened as well by the occupants (see Figure 2-41). The control device provides an additional button for intermittent ventilation.

Figure 2-41: Control device and its functions

1.5.3 Aim of the work

The main focus is to identify energy-related occupant behavior patterns for window control and the usage of sun protection devices in relation to outdoor in indoor climate. Factors like season, location of the office or time of the day are considered as well. The objective of this analysis is to evaluate how much the occupants interact with their building in a manner suitable to the building concept with natural ventilation.

1.5.4 Database characteristics

The analyses for the room sample (see chapter 4.2) are based on a scientific monitoring in the context of the research program 'SolarBau' 2001 to 2005 (funded by the German Ministry for Economics and Technology BMWi) and was continued on behalf of the KfW until 2011.

Table 2-7: Summary of database characteristics

Number of buildings	1	
Period of measurement	Since April, 2003 with varying intensity	
Number of observed spaces with window sensors	~35	
Number of observed spaces with CO_2-concentration	5	
	Items	Interval
IF1. Climate	Outdoor air temperature, humidity, wind speed, solar radiance	10 min
IF2. Building envelope	Not in database	
IF3. Building service & Systems		10 min
IF4. Operation & Maintenance	Monitoring of heating, cooling, lighting and ventilation system, and related energy flows	10 min
IF5. Indoor environmental quality	Indoor operative temperature, humidity, (CO_2)	10 min
IF6. Occupants' activities and behavior	Window state (open/closed) Presence State of sun protection (open/closed) Usage of lighting equipment	Event
IF7. Social and economical aspects	none	

1.5.5 Measurements

The KfW building has been monitored, and data for 35 offices has been collected since April 2003. In the database there are mainly three types of data which can be grouped with respect to outdoor conditions, indoor conditions, and actions of occupants or events. Among these data, some have driver effect, and some others are driven. Drivers are changes in indoor and outdoor conditions while driven ones are defined as behavior in response to these changes. Table 2-8 summarizes the monitored data for all conditions as follows.

Table 2-8: Monitored data

Outdoor	**Indoor**	**Behaviour**
Solar radiation [W/m2]	Room air temperature [°C]	Occupancy [0/1]*
Rain – amount [l/m2]	Surface temperature [°C]	Window contact [0/1 ; Reed contacts]*
Rain – event [yes/no]	Ceiling slab temperature [°C]	Top light control [0/1 ; Reed contacts]*
Light intensity– horizontal [lx]	CO_2 concentration [ppm]	Sun protection [% of closure: 0% = open to 100% = closed]
Light intensity - South [lx]		Electricity consumption [kWh]
Light intensity - East [lx]		
Light intensity - North [lx]		
Light intensity - West [lx]		
Outdoor temperature [°C]		
Wind – velocity [m/s]		
Wind – direction [°]		
CO2 content in air [ppm]		
Outdoor humidity [%rH]		

*for analyzes aiming at duration in terms of daily means, data were transformed from 10 minute intervals to minutes

A weather station is located on the top of the building, providing data regarding the outdoor conditions for all offices, such as temperature. However, the microclimate on the façades can differ, e.g. depending on the intensity and direction of wind. Location and types of the offices vary as described in the following. The office rooms can be grouped in four types; standard offices, traders' offices, large offices and others with a special function in use. Traders offices are special offices, facing south, with before mentioned double skin façade and specially designed ceiling panels for acoustic performance and mechanical cooling. They are excluded from the analysis. Large offices which are located in four directions of the building are 2-3 times larger than standard office spaces and one of them is located at the corner with 2-side windows, which enable cross-ventilation. Other special offices include open-plan offices and meeting rooms, which differ from each other in façade type (single/double skin), air conditioning and differentiation in function in use. All of them are also excluded from this analysis. Standard offices all have the same size (~20m2), facing mostly east and west (one is facing south). They have one fixed and two operable windows, internal top light windows above the doors (to allow for night ventilation through the atrium) and sun protection elements (operated both manually and automatically). They are occupied by one or two persons.

Room Sample

Due to the diversity of office usage in the building and their different it was decided to only analyze standard offices in terms of occupant behavior. Since the building was completed and started to be monitored in 2003, the starting year for the analyses was selected as 2004. The analysis period was determined as 3 years, from January 1st, 2004 to December 31st, 2009.

In the database, besides the data of outdoor conditions, there are several data indicating changes, behaviors and events in the building. Among indoor conditions data, five data is available for all offices, which are presence of the occupant(s), window contact, top window control, room air temperature and use of sun protection, while others are not available for all offices, including CO_2 concentration, surface temperature, component temperature and electricity consumption. A summary of the variables for monitored standard offices is shown in Table 2-9. For the analyses presented in chapter 6, we concentrate on those parameters, which are available for all 16 rooms of the sample; therefore CO_2-concentration and surface temperature are not included.

Table 2-9: Variables for monitored standard offices

	Room ID	**Occupancy**	**Window control**	**Top window control**	**Room air temperature**	**CO_2-Concentration**	**Surface temperature**
East	E01	●	●	●	●		
	E02	●	●	●	●		
	E03	●	●	●	●		•
	E04	●	●	●	●	•	•
	E05	●	●	●	●		
	E06	●	●	●	●		•
	E07	●	●	●	●	•	•
	E08	●	●	●	●		

	E09	●	●	●	●		
	E10	●	●	●	●		
	E11	●	●	●	●		
West	W01	●	●	●	●		•
	W02	●	●	●	●		•
	W03	●	●	●	●		
	W04	●	●	●	●		
	W05	●	●	●	●	•	•
	W06	●	●	●	●		

1.5.6 **Methods**

Extraction of the Data

Monitored data was compiled with MoniSoft . Required data for the analyses have been extracted from the software. The data obtained were processed and implemented into SPSS

Analysis Method

A variety of statistics is applied to explore patterns of user behavior, which are described shortly below.

a) **t-tests.** To explore group differences with respect to *duration of window opening*, *season* or *orientation of the office,* t-tests were applied. The t-test assesses whether the means of two groups are *statistically* different from each other [4], e.g. east versus west oriented rooms.

b) **Logistic regression.** The method allows predicting the outcome of a binary dependent variable by modeling the probability of an event such as *window opening* (*yes* or *no*). The analyses are based on one or more predictor variables such as *outdoor temperature* or *operative temperature* [3].

c) **Calculation of optimal *duration of window opening*.** To analyze if the building concept and the user behavior concerning window opening fit, two approaches were chosen.

1. Based on the paper of van Paassen, Liem & Groninger [9] the necessary respectively optimal ventilation duration for the offices was calculated including factors such as the opening angle, effective ventilation opening, height and width of the windows which results in an effective air change for each window. Taking into account the number of persons in the room and the specific fresh air volume (10 liter per person and second) the optimal duration of ventilation was calculated, revealing that 18 minutes of ventilation per hour would be adequate. Taking 8 hours of work as a basis and taking into account the automated ventilation in the morning, 7x18 minutes were taken as a basis, resulting in a maximum of 126 minutes per day.
2. Recommendations for optimal ventilation in offices are widely spread in the context of building science, occupational health and safety and comfort. Looking at German websites and literature the recommendations for an optimal window opening on a working day vary: e.g. 3 minutes every hour (8x3 minutes = 24 minutes) or 5 to 10 minutes every 2 hours (5x10 minutes = 50 minutes)[12, 13, 14]. Taking into account the automated ventilation before working hours, the calculation for the optimal window opening duration was calculated as follows: 4x10 minutes (=40 minutes per day).

Difference between outdoor and indoor temperature.
Regarding adequate window opening and energy consumption as well as comfort the difference between outdoor and indoor temperature might play an important role. For the analyses this difference was calculated. To separate the summer from the winter season, data for window opening behavior where taken for which the *outdoor temperature is higher* than the *indoor temperature.*

1.5.7 Results

Table 2-10 shows descriptive values for parameters regarding orientation of the offices and season based on data from 2004 to 2009. Differences of data concern daily means of indoor temperature, occupancy as well as opening of window and top light. As expected there are statistically significant differences between winter and summer mean outdoor temperatures, and no differences between east and west outdoor temperatures. Generally indoor temperatures are about 2 degrees Celsius higher in summer, and offices oriented to the east show slightly higher temperatures in summer. Mean time duration for occupancy is comparable between east and west offices. Generally duration of window and top light opening is higher in summer season, with longer duration for the top lights in comparison to the windows. For both seasons duration for window opening is significantly higher in west offices, while for the top lights it is the other way round.

Table 2-10: Differences between parameters regarding orientation and season

Ø daily mean values; sample: 17 rooms							
time periode	**01.01.2004 – 31.12.2009[a]**						
orientation	**east**			**p[b]**	**west**		
season	**winter**	**p[c]**	**summer**		**winter**	**p[c]**	**summer**
outdoor temperature (°C)	4,9	***	21,9	=	5,1	***	22,1
indoor temperature (°C)	21,6	***	23,9	>>>	21,6	***	23,5
occupancy (minutes)	456	n.s.	456	>>>	442	n.s.	445
window opening (minutes)	27	***	261	<<<	63	***	338
toplight opening (minutes)	40	***	457	>>	13	***	438

Note.
a. missing data for 2008; the data showed nearly no variance in sun protection usage (sun protection is predominantly closed in the offices), therefore this aspect is not included in the analyses.
b. Statistical test for differences: t-test (means) with not significant n.s., $p < .05$*; $p < .01$**; $p < .001$***. Differences between east and west are indicated with n.s. =, $p < .05$< or >; $p < .01$<< or >>; $p < .001$<<< or >>>.
c. Differences between winter and summer are indicated with not significant n.s., $p < .05$*; $p < .01$**; $p < .001$***.

Influencing factors on window opening
The following figures (Figures 2-42 to 2-49) show the percentage of open windows in relationship to operative temperature as well as to outdoor temperature differentiated for season and orientation.

Black dots represent the percentage of open windows per 10 minutes interval, bars represent the 95% confidence interval and the black lines show the probability of the logistic regression models.

a) Winter

Comparing Figures 2-42 and 2-45, the variability of the proportion of open windows with regard to outdoor temperatures is much higher for the west facing windows compared to the east facing windows. The occupants of the east facing rooms are opening their windows much less, the T_{50} (temperature at which 50% of windows are open) is far beyond the scale for the east orientation, while it is 18°C for the west orientation.
In contrast to all presented results in the literature, the opening probability related to the operative temperature is negative for both orientations, i.e. higher indoor temperatures are related to lower opening proportions.

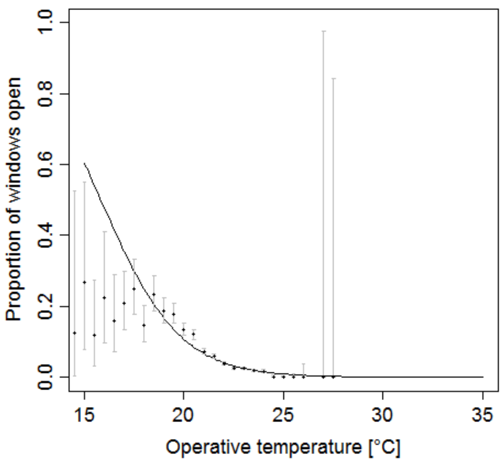


Figure 2-42: Influence of operative temperature on window opening in offices oriented to the east (winter)

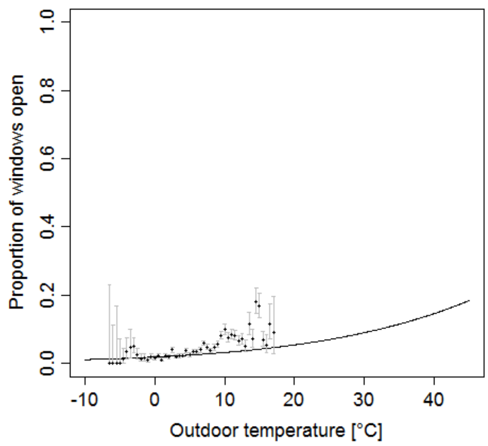


Figure 2-43 Influence of outdoor temperature on window opening in offices oriented to the east (winter)

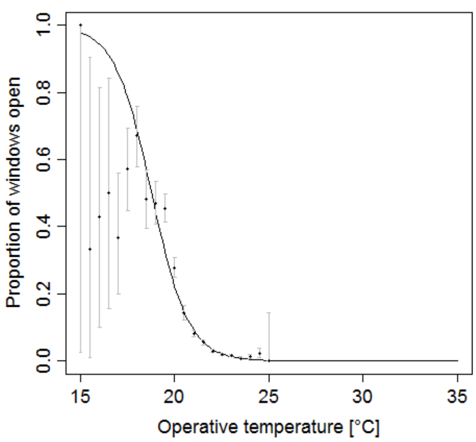


Figure 2-44: Influence of operative temperature

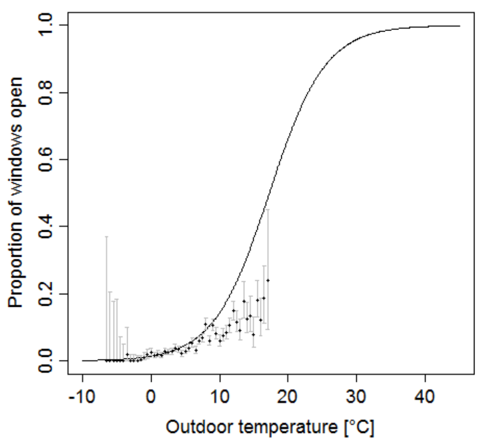


Figure 2-45: Influence of outdoor temperature on

on window opening in offices oriented to the west (winter)

window opening in offices oriented to the west (winter)

Figure 2-43 and 2-45 show the increase of the probabability for window opening with increasing outdoor temperatures. as well as slight increase for operative temperatures in offices which are oriented to the west.

b) Summer

The data for the summer season (Figures 2-46 to 2-49) show very similar trends for east and west facing offices with respect to the relationship between outdoor temperature and opening proportion. Besides the multinominal regression model presented here showing an increase towards higher temperatures, the data points show that there is a peak opening proportion in both façades around 27-28°C with decreasing opening proportion above this point. The opening proportion compared to the operative temperature also shows similar trends for east and west orientation. The T50 is or both cases around 30°C.

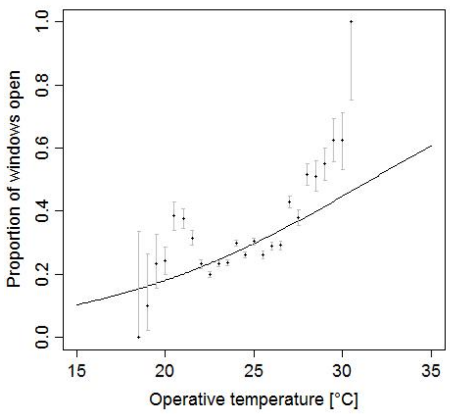


Figure 2-46: Influence of operative temperature on window opening in offices oriented to the east (summer)

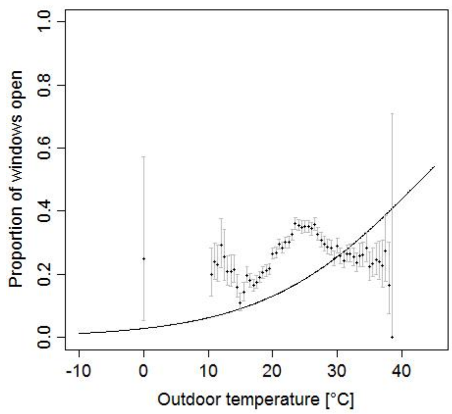
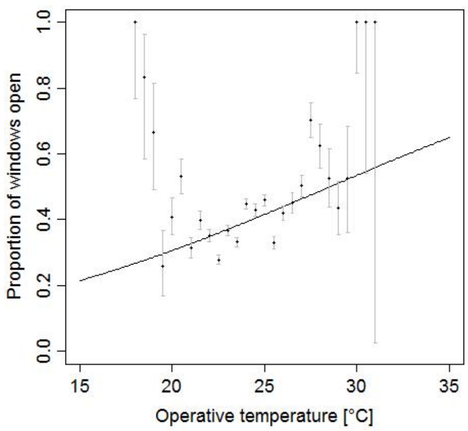


Figure 2-47: Influence of outdoor temperature on window opening in offices oriented to the east (summer)



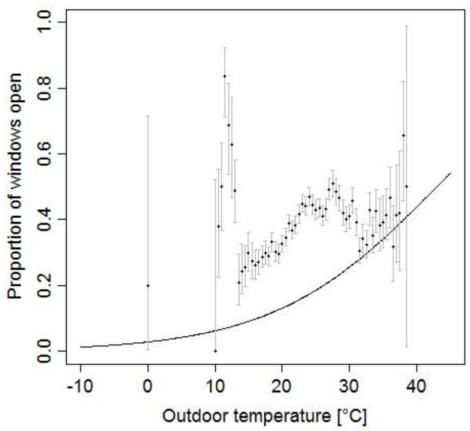


Figure 2-48: Influence of operative temperature on window opening in offices oriented to the east (summer)

Figure 2-49: Influence of outdoor temperature on window opening in offices oriented to the east (summer)

In general, these results confirm that operative temperature as well as outdoor temperature have an influence on window opening. While in summer there are only minor differences between the orientation, offices oriented to the west rooms have a higher probability for window opening in winter.

Fit between occupant behavior and the energy concept of the building

a) Winter

The results show that the occupants' behavior regarding window opening fits for over 80% with the energy concept of the building in offices oriented to the east and over 90% in offices oriented to the west according to the calculation of an optimal opening duration of a maximum of 126 minutes per day (see 5.2).

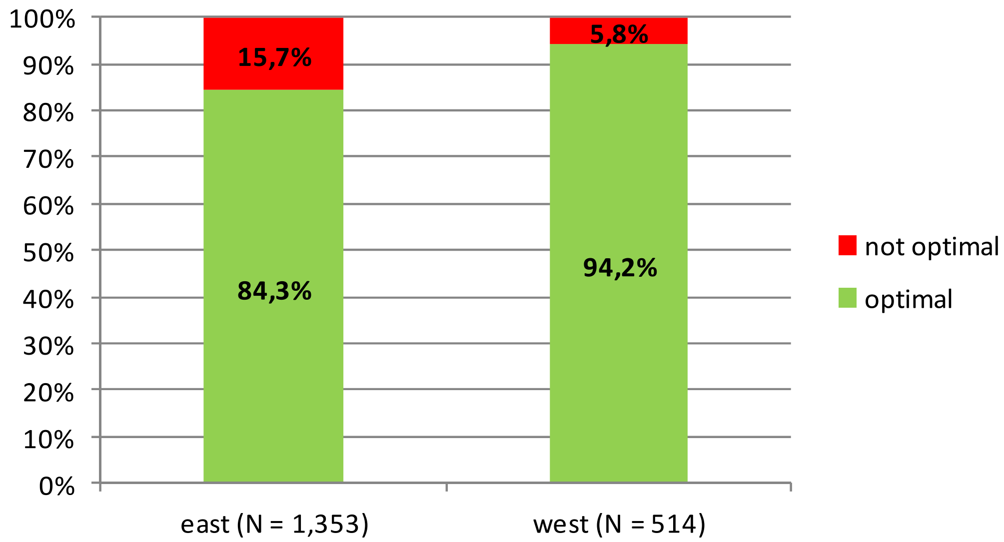


Figure 2-50: Percentages of optimal and not-optimal window opening (max. 127 minutes) based on calculations according to [9].

When taking common recommendations (see 5.2) regarding window opening in offices, which is much stricter, the percentage of optimal opening duration decreases.

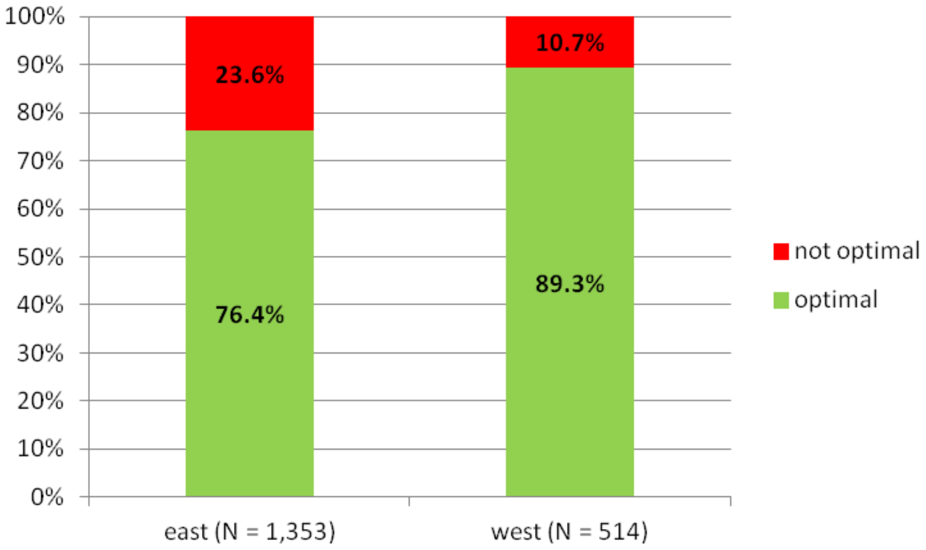


Figure 2-51: Percentages of optimal and not-optimal window opening (max. 40 minutes) based on common recommendations.

b) Summer

While extended opening times in winter time directly lead to a higher energy consumption, this is not the case in summer. However, an open window at conditions with higher outdoor temperatures than indoor temperatures increases indoor temperatures (which might decrease thermal comfort) and necessitates a higher need for nighttime ventilation in order to lower temperatures to comfortable values until the next morning. Such extra cooling demand could lead to a higher demand of night ventilation which then comes along with an additional electricity demand for fans if forced ventilation has to be used.

Figure 2-52 and Figure 2-53 shows the mean values of indoor and outdoor temperature in the course of daily hours for the whole summer periods. Around noon, the outdoor temperatures start being above the indoor temperatures.

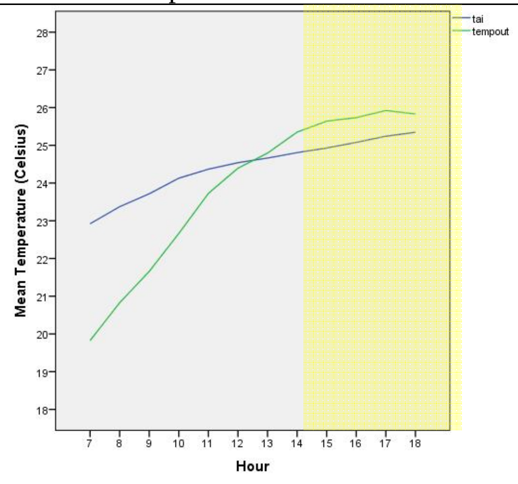
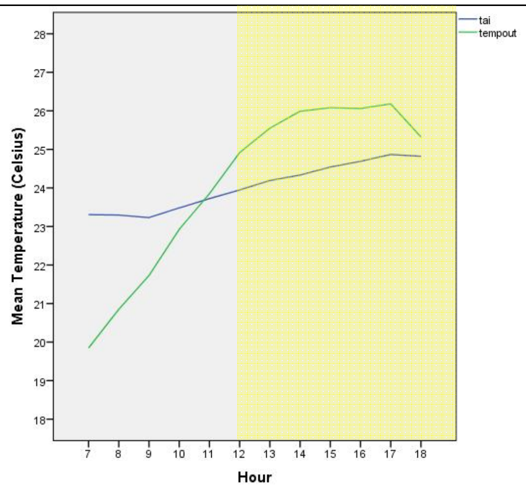
	

Figure 2-52: Differences between outdoor and indoor temperatures for summer working hours (east) and open windows	*Figure 2-53: Differences between outdoor and indoor temperatures for summer working hours (west) and open windows*

For the following analyses topligt opening is excluded, because a distinction between the automated opening and the manually opening through the occupants cannot be made. In addition to looking at the development of outdoor and indoor temperatures the distribution of open and closed window states are analysed (Figure 2-54 and Figure 2-55). As a tendency, the percentage of open windows in the east offices is slightly lower than in the west offices. In the east offices there is small increase in the first working hours and a constant increase during the afternoon hours with a maximum of 40% open windows at the end of the working day. In the west offices there is a peak with 44% of open windows, decreasing percentages until noon and slightly increasing percentages during the afternoon hours and a decrease at the end of the working day.

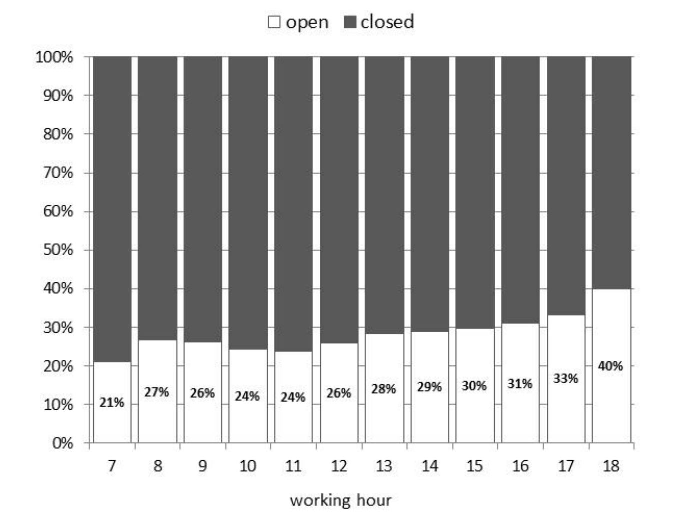


Figure 2-54: Distribution of open windows during the day (east)

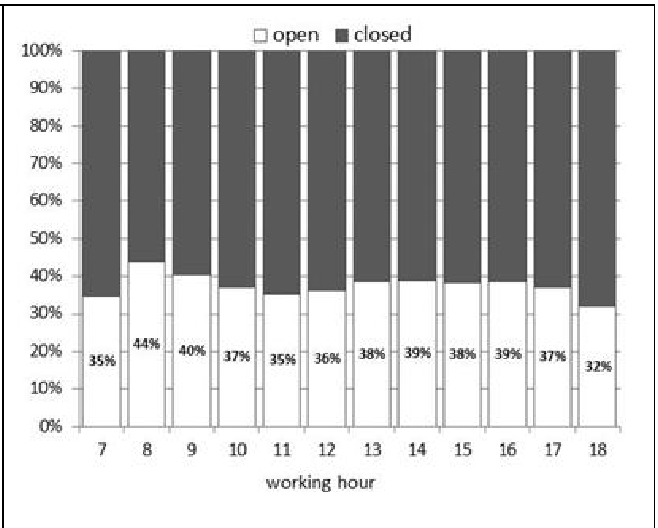


Figure 2-55: Distribution of open windows during the day (east)

Looking at the differences between outdoor and indoor temperatures (Figure 2-56), the graph shows that around 35% of the windows are open, when the indoor temperature is more than 12 degrees Celsius higher than the outdoor temperature. This may be due to opening the window in the morning when the occupants start working, which would fit to the results in Figure 2-54 and 2-55. There is a peak of almost 40% when outdoor temperature and indoor temperature are very close. The percentage of open windows decreases constantly when the outdoor temperature is getting higher than the indoor temperature.

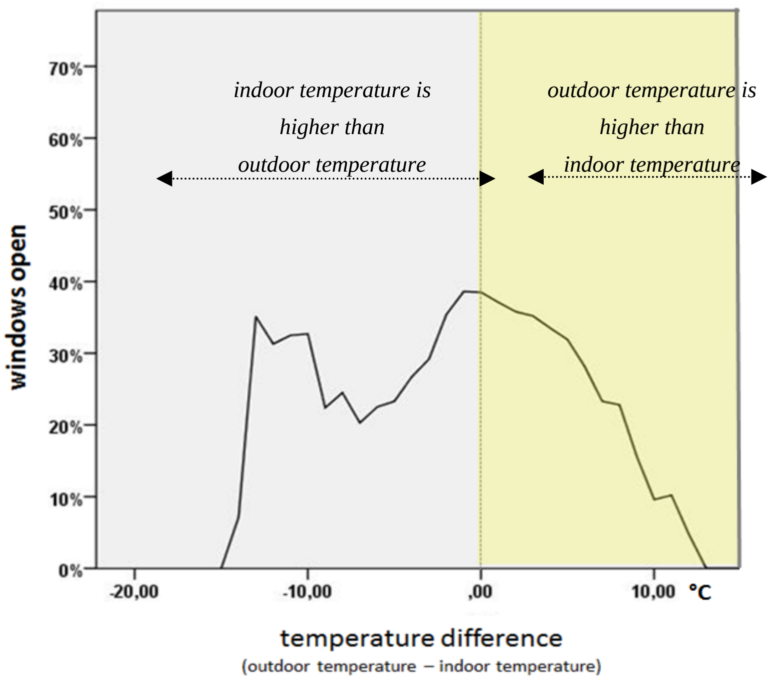


Figure 2-56: Percentage of open windows in summer regarding temperature differences in the offices

1.5.8 **Discussion and conclusions**

Regression analyses showed that outdoor temperatures as well as indoor temperatures turned out to be influencing factors on window opening which is in accordance to the literature [3], [5], [8]. This is similar to the effect shown by [7] for the Japanese data. The authors argue that for the Japanese data the reason was the availybility of an AC-unit for the occupants. Thus, the results of this data show the same results even without an AC-unit available for occupants.

Concerning the important issue about the match of the building concept and the occupants' behavior, the analyses revealed that for the KfW-building relatively long window opening times are necessary respectively adequate and that more than 80% of the occupants practice optimal window opening. However, there is potential for optimization, especially with respect to offices oriented to east. It can be assumed that shorter window opening duration would lead to less comfortable conditions in the KfW-offices, compared to often recommended window opening times for offices. A positive result is the fact that there is a clear tendency to keep windows closed when outdoor temperatures are higher than indoor temperatures.

A reason for lower percentages of open windows in offices oriented to the east might be found in the surrounding of the building. East offices are facing a street, while the offices on the west side are oriented to an inner courtyard. Noise coming from the street might result in the tendency keeping the windows closed in offices of the east side. Additionally, along the east side there are high trees, giving shade in the afternoon hours compared to the inner courtyard facing the west offices. As a result, opening the window during the afternoon working hours might provide fresh air for the east offices, but higher temperatures coming from the heated courtyard without trees for the west offices. This might explain the tendency to keep windows closed during the afternoon hours for the west offices.

A variety of restrictions to the findings has to be stressed. Although the database of the monitoring provides a relatively long time period resulting in a considerably number of data, the interpretation of findings is limited when important information is not available and cannot be taken into account. This is for instance true for information on CO_2 in the offices (not to mention VOC emission) or subjective ratings of the occupants concerning comfort. Despite strong efforts permission for surveys could not be obtained for the KfW-building. Subjective data in terms of direct feedback of the occupants are a relevant part when it comes to a comprehensive understanding and interpretation of behavior profiles based on information coming from sensors. With respect to the data higher indoor temperatures coincide with lower opening proportions. However, it is not clear, which is the depending and which the independent variable here; the assumption is that the higher operative temperatures are a result of less window opening and not that higher temperatures lead to less opening behavior. This has to be investigated in more detail.
Another restriction is that the data does not allow a distinction between automated opening and manually handling of the top lights. Although the database seemed to be very large, with respect to specific statistical analyses the sample shrunk when comparable room samples had to be built (e.g. office type, orientation).

In conclusion, the findings show that behavior profiles of window opening give helpful hints regarding the interaction between building and occupants. The behavior might be counterproductive with respect to the energy concept. While it is obvious that prolonged window opening in winter is linked to higher energy consumption, the analysis of the summer situation for a naturally ventilated building with night ventilation is more complex. In the summer case, prolonged window opening at high outdoor temperatures does not lead directly to a higher energy demand, because no cooling system exists. However, the increased heat gain leads to a higher demand for night ventilation, which in some cases is facilitated by an electrical fan. In such a way, the auxiliary energy demand can be increased. Nevertheless, for the building analyzed in a German climatic context, such effect was found to be small due to most people closing their windows when outdoor temperatures are higher than indoor temperatures.

Acknowledgements
This research was supported by the German Federal Ministry of Economics and Technology (BMWi) with the project ID 03ET1035B.

1.5.9 **References**

[1] Barlow, S. & Fiala D. (2007). Occupant comfort in UK offices - How adaptive comfort theories might influence future low energy office refurbishment strategies. *Energy and Buildings,* 39, 837–846.

[2] Burger, J. M. (1992), *Desire for Control. Personality, Social, and Clinic Perspectives*. New York: Plenum.

[3] Finney, D.J. (1964). *Probit analysis: a statistical treatment of the sigmoid response curve, 2nd. Ed.*. London: Cambridge University Press.

[3] Haldi, F. & Robinson D. (2008). On the behaviour and adaptation of office occupants. *Building and Environment*, 43, 2163–2177.

[4] http://www.socialresearchmethods.net/kb/stat_t.php

[5] Kleber, M., Wagner, A. Results of Monitoring a Naturally Ventilated and Passively Cooled Office Building in Frankfurt, Germany, EPIC 2006 AIVC Conference, Lyon, 20.-22. November 2006

[5] Nicol, J. F. (2001). Characterising occupant behaviour in buildings: towards a stochastic model of occupant use of windows, lights, blinds heaters and fans. *Proceedings of the seventh international IBPSA conference, Rio 2. International Building Performance Simulation Association* (pp. 1073–1078).

[6] Schakib-Ekbatan, K. & Wagner, A. (2011). Occupant satisfaction as an indicator for the socio-cultural dimension of sustainable office buildings. In: Proceedings CISBAT, École Polytechnique Fédérale de Lausanne [CD].

[7] Schweiker, M., Haldi, F., Shukuya, M. and Robinson, D. (2012). Verification of stochastic models of window opening behaviour for residential buildings. *Journal of Building Performance Simulation*, Special Issue: Modelling Occupants' Presence and Behaviour – Part II, Vol. 5 (1), p. 55-74, 2012

[8] Schweiker, M. & Shukuya, M. (2009). Comparison of Theoretical and Statistical Models of Air-Conditioning-Unit Usage Behaviour in a Residential Setting under Japanese Climatic Conditions, *Building and Environment*, 44, 2137-49.

[1] Toftum, J., Andersen, R., V. & Jensen, K.L. (2009). Occupant performance and building energy consumption with different philosophies of determining acceptable thermal conditions, Building and Environment, 44(10), 2009–2016.

[9] Van Paassen, A. H. C., Liem, S. H. & Groninger, B. P. (1998). Control of night cooling with natural ventilation: sensitivity analysis of control strategies and vent openings. *Proccedings of the 19TH ANNUAL AIVC CONFERENCE. NatVent. OSLO, NORWAY* (pp. 438-447).

[10] Wagner, A., Gossauer, E., Moosmann, C., Gropp, T. & Leonhart, R. (2007). Thermal comfort and workplace occupant satisfaction - Results of field studies in German low energy office buildings. *Energy and Building*, 39, 758-769.

[11] Wagner, A. & Schakib-Ekbatan, K. (2011). User satisfaction as a measure of workplace quality in the office. In C. Schittich (Hrsg.), *Work Environments: Design in Physical Space, Mobility, Communication*, 54-57. Basel: Birkhäuser Architektur.

[12] Hansmeier, N. & Matthies, E. (2007). Energiebewusste RUB – Richtig Heizen *und Lüften. Ergebnisse einer umweltpsychologischen Intervention zur Förderung energieeffizienten Verhaltens an der Ruhr-Universität Bochum*. Unveröffentlichter Projektbericht

[13] http://www.arbeitssicherheit.de/de/html/nachrichten/anzeigen/683/Richtig-L%C3%BCften/

[14] http://publikationen.dguv.de/dguv/pdf/10002/bgi7004.pdf

1.6 Experience 5: Building energy consumption modeling with neural Ensembling approaches

(F. Lauro)

In the proposed work the aim is modeling building energy consumption. Several classical methods are compared to the latest Artificial Intelligence modeling technique: Artificial Neural Networks Ensembling (ANNE). Therefore, in this study is shown how the ANNE was built. Experimentation has been carried out over three months data sets coming from a real office building located in the ENEA 'Casaccia' Research Centre in Rome. Experimental results show that the proposed ANNE approach can get a remarkable improvement with respect to the best classical method (the statistical one).

1.6.1 Introduction

Building energy consumption represents about 30%-40% of the global energy consumption [1] and it is the cause of about 40% of CO_2 emissions [2]. Therefore, the study of building energy demand has got in the recent years a remarkable relevance [3] in order to improve the management of existing buildings and the design of the new ones. In this context, having reliable energy estimations, and thus accurate models, is the key for energy efficiency with remarkable economic and environmental advantages.

In this scenario, at present there are three different approaches [4] for modeling energy consumption in buildings: Statistic Modeling (SM), Simulation Programs (SP) and Intelligent Computer Systems (ICS). The first one, known also as inverse modeling [5], is based on the building behavior. With this approach a priori hypothesis about the model structure are made and the internal model parameters are tuned up through statistical analysis methods and the most popular techniques are linear regression and multivariate analysis [4]. Therefore the structure of the models is pretty straightforward but sophisticate statistical analysis methods are needed. The second approach, known also as direct modeling [5], starts from the physical description of the building which feeds a simulation program. Such methods need very accurate ambient information, a highly detailed building description and information about the occupants behavior. All this information makes this approach computationally very expensive. The last approach, known also Artificial Intelligence (IA) approach [4], is based on Expert Systems (ES) and Artificial Neural Networks (ANN). ES are computer systems [6] that emulates the decision-making ability of a human expert. ES are designed to solve complex problems by reasoning about knowledge, like an expert, and it has a unique structure, different from traditional programs, which is divided into two parts, one fixed, independent of the expert system: the inference engine, and one variable: the knowledge base. ANN [7,8] are a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. An ANN consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

The application of these methods depends on their own characteristics [4]. SM is mainly used in energy modeling of clusters of buildings [9,10] or in the design of areas where different end users are present. SP is mostly applied as energy estimation tools of single buildings [11], in design and retrofit interventions. ICS are somehow in between the two because can be applied to single as well as

building clusters [12, 13], moreover ICS are very effective in diagnosis, automation, control and management optimization.
Therefore, in this paper is described a novel ICS for energy consumption modeling aimed at making diagnosis systems and controlling a complex building.

1.6.2 **Methods**

In this paragraph the modeling techniques compared in the experimentation are shortly described.

Naïve Model
In order to perform a meaningful comparison for the forecasting, a naive model should be introduced in order to quantify the improvement given by more intelligent and complex forecasting techniques. For seasonal data a naïve model might be defined as:

$$x_t = x_{t-s} \tag{1}$$

with S the appropriate seasonality period. This model gives a prediction at time t presenting the value observed exactly a period of S steps before. For this work, the value of S is 24 * 7 = 168 which corresponds to a week, given that the sampling considered is hourly.

1.6.3 **Statistical Model**

One the simplest and most widely used models is to build an average weekly distribution of the consumption sampled hourly. Thus, from the data for each day the average consumption is computed hour by hour in order to get an average distribution made of 24 * 7 = 168 points.

Artificial Neural Networks
Artificial Neural Networks (ANN) [7,8] are computational models which try to simulate some properties of biological neural networks in order to solve complex modeling problems of non-linear systems. An ANN is an interconnected group of artificial neurons (called also nodes) that uses a mathematical or computational model for information processing based on a connectionistic approach to computation. In more practical terms ANN are non-linear data modeling or decision making tools which can be used to model complex relationships between inputs and outputs or to find patterns in data. ANN are referred also as black-box or data-driven models and they are mainly used when analytical or transparent models cannot be applied. Building such models needs several stages as input analysis and training through algorithms which minimize the error between the real values to be modeled and the ANN output. ANN demonstrated their effectiveness in modeling many real-world applications.
Once modeling an ANN model, we must take into account three basic components. First, the synapses of the biological neuron are modeled as weights. Let's remember that the synapse of the biological neuron is the one which interconnects the neural network and gives the strength of the connection. For an artificial neuron, the weight is a number, and represents the synapse. A negative weight reflects an inhibitory connection, while positive values designate excitatory connections. The following components of the model represent the actual activity of the neuron cell. All inputs are summed altogether and modified by the weights. This activity is referred as a linear combination. Finally, an activation function controls the amplitude of the output. Mathematically, this process is described in Figure 2-57.

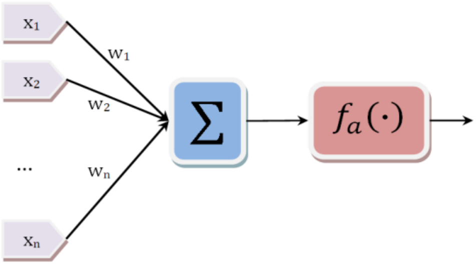


Figure 2-57: Artificial neuron model

From this model the activity of the neuron can be shown to be:

$y = f_a(\Sigma w_i x_i - \theta)$ (3)

where θ is a threshold called BIAS (Basic Input Activation System) which identifies the sensitivity of the neuron to respond to the external inputs. The most common function used to model f_a are the hyperbolic tangent, the sigmoid and the linear function.

Therefore each unit performs a relatively simple job: receive input from neighbors or external sources and use this to compute an output signal which is propagated to other units. Apart from this processing, a second task is the adjustment of the weights. The system is inherently parallel in the sense that many units can carry out their computations at the same time. Within neural systems it is useful to distinguish three types of units: input units which receive data from outside the neural network, output units which send data out of the neural network, and hidden units whose input and output signals remain within the network.

The way units are connected defines the network topology or architecture. In the past years many of them have been studied and the most widely used is the feed-forward one. In this network structure neurons are grouped into layers. There exists at least two layers, the input and the output one, which are those gathering the corresponding input and output variables. This basic structure is also known as perceptron [14]. Moreover, in order to let the model cope with non-linear problems, it is possible to add one or more intermediate layers, known as hidden layers. These models are also known as multi-layer perceptrons (MLP)[15].

The flow of data from input to output units is strictly in one direction (forward). The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers.

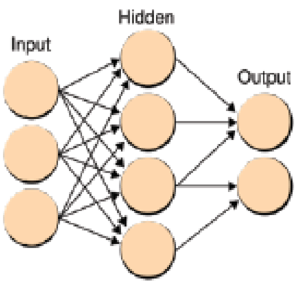


Figure 2-58: Feed-forward neural network topology

A neural network has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of

the connections exist but the most used way is to 'train' the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule.
Several different training algorithms for feedforward networks use the gradient of the performance function to determine how to iteratively adjust the weights to minimize performance. The gradient is determined using a technique called backpropagation [16, 17], which involves performing computations backward through the network. The simplest implementation of backpropagation learning updates the network weights and biases in the direction in which the performance function decreases most rapidly, the negative of the gradient. One iteration of this algorithm can be written:

$$x_{k+1} = x_k - \alpha_k \frac{\partial E_k}{\partial x_k}$$

where x_k is a vector of current weights and biases, $\frac{\partial E_k}{\partial x_k}$ is the current gradient (E_k is the current error between the network outputs and the target outputs), and α_k is the learning rate.
There are two different ways in which this gradient descent algorithm can be implemented: incremental mode and batch mode. In incremental mode, the gradient is computed and the weights are updated after each input is applied to the network. In batch mode, all the inputs are applied to the network before the weights are updated.
The backpropagation training algorithms are often too slow for practical problems. The Levenberg – Marquardt is a high – performance algorithm that can converge from ten to one hundred times faster than the backpropagation algorithm discussed previously. The Levenberg – Marquardt algorithm [18, 19] is an approximation to Newton’s method: it was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feedforward networks), then the Hessian matrix can be approximated as:

$$H = J^T J$$

and the gradient can be computed as:

$$J^T E_k$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and E_k is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique that is much less complex than computing the Hessian matrix. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - \left[J^T J + \mu I \right]^{-1} J^T E_k$$

When the scalar μ is zero, this is just Newton’s method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton’s method is faster and more accurate near an error minimum, so the aim is to shift toward Newton’s method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm.

Ensembling Methods
The term ‘ensemble’ describes a group of learning machines that work together on the same task, in the case of ANN they are trained on some data, run together and their outputs are combined as a single

one. The goal is obtain better predictive performances than performances that could be obtained from any of the constituent models.

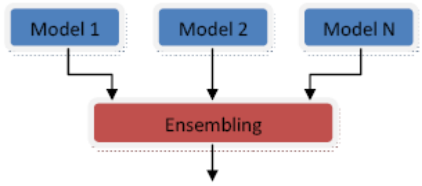


Figure 2-59: Ensembling

In the last years several ensembling methods have been carried out [20, 21, 22]. The ensemble methods can be divided into two categories: generative and non generative.

The non generative methods seek to combine in the best way the outputs of the machines. The first one, also known as Basic Ensemble Method (BEM), is the simplest way to combine M neural networks as an arithmetic mean of their outputs y_i. This method can improve the global performance [23, 24] although it does not takes into account that some models can be more accurate than others. This method has the advantage to be very easy to apply. A direct BEM extension is the Generalised Ensemble Method (GEM) [23, 24] in which the outputs of the single models are combined in a weighted average where the weights have to be properly set, sometimes after an expensive tuning process.

The generative methods spawn new sets of learner from the original one, so as to create differences between them that can improve the overall performance. For example, Bagging and Adaboost generative methods use the bootstrapping resampling technique [25] that allows to generate different sets for machines training. In Bagging (Bootstrap AGGregatING) [26], the bootstrapping technique on a database consists in the extraction with replacement of its elements to create several new training set. The probability of extraction of each example is equal to that of the other. The basic algorithm consists in creating models for each training set and then in combining the various estimates on the test set through an average operation (Figure 2-60). The name Adaboost [27, 28] derives from the fact that the ensemble provides the bootstrap adaptive: it possesses the ability to adapt to the difficulty characteristics of the training set. The central idea is to extract a random number of examples from the training set, then assign a higher probability of extraction for the examples more difficult to learn. Initially, a first machine is trained with a training set constructed by random selection with equal probability for all examples. After that, the extraction probabilities are updated for the next training set by increasing probability of the original set worst learned examples. It generates a new training set and a new machine is trained and so on.

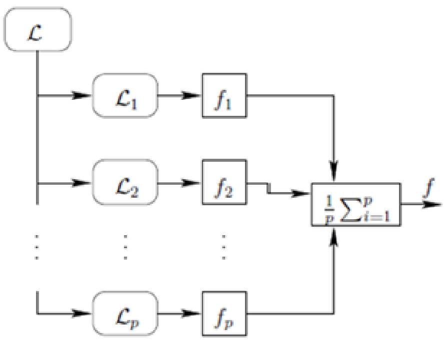


Figure 2-60: Bagging algorithm block diagram

1.6.4 **Experimentation**

In this paragraph the methods presented in the previous section are tested and compared.
The test case has concerned the energy consumption modeling of a real office building (building 'C59') located in the ENEA 'Casaccia' Research Centre (Rome, Italy). Modeling refers to three energy consumption types of the building: lighting, electromotive force (emf) and conditioning. Experimentation has been carried out over three different data set, each one made approximately of 3 months of hourly measurement (from September to November 2009 for lighting and emf, from June to August 2009 for conditioning). Each sample consists of measurements like month, day of the month (1-31), day of the week (1-7), time, working day (true/false), occupancy, solar radiation, outdoor temperature, sunset time, used as inputs of the neural models, and consumption, used as the target (output).
The data sets have been split in two parts : training (approximately 10 weeks) and validation (approximately 3 weeks, one for each month) and the reported results refer to the last one.
The ANN applied for the data analysis are feed-forward MLP with several inputs dependent on the particular type of consumption (Tab. 15), 1 hidden layer consisting of 10 neurons, 1 output (lighting, emf or conditioning consumption), hyperbolic tangent as activation function for the hidden neurons and linear for the output. The artificial neural networks ensembling is built according to BEM.

Table 2-11: Best input combinations for each consumption type

	NUMBER OF INPUTS	INPUT VARIABLES
LIGHTING	7	Month, day of month, day of the week, time, occupancy, global radiation, outdoor temperature (working days only)
EMF	11	Month, day of month, day of the week, time, working day, occupancy, global radiation, diffuse radiation, direct radiation, outdoor temperature, sunset time
CONDITIONING	9	Month, day of month, day of the week, time, occupancy, global radiation, diffuse radiation, direct radiation, outdoor temperature

Training has been performed with MATLAB (ver. R2010a) through the Levenberg-Marquardt algorithm stopping after 1000 iterations. The reported results are averaged over 10 different runs (with standard deviation in brackets) and the ensemble is therefore made by the same 10 models.
Performance has been measured according to the Mean Absolute Error (MAE) and the Maximum Absolute Error (MAX) (Tab. 16):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{5}$$

$$MAX = \max\left\{\left|y_i - \hat{y}_i\right|\right\}_{i=1}^{N} \tag{6}$$

where y_i is the real output, $\hat{y}_i$ is the estimated output and N is the data set real values size.

Table 2-12: Experimental results (testing).

		Naive	Statistical	ANN	BEM
LIGHTING	MAE (kW)	2.95	0.97	1.22 (±0.10)	0.95
	MAX (kW)	4.90	4.74	5.53	4.02
EMF	MAE (kW)	1.51	1.38	1.01 (±0.21)	0.68
	MAX (kW)	12.40	7.50	8.15	4.49

CONDITIONING	MAE (kW)	4.76	4.06	3.45 (±0.34)	2.95
	MAX (kW)	14.84	13.00	15.55	8.69

Experimental results show that lighting and emf consumptions are more easily modeled than the conditioning one (Figure 2-61, 2-62, 2-63): for this energy consumption type, further data, in addition to those available, are necessary (i.e., building internal temperature).
We can see that the proposed method (BEM) clearly outperforms all the others. The reason for that is that the proposed approach is capable to provide reliable estimations when out of standard conditions because it takes into account several input features (as occupancy) which affect the energy consumptions.
Moreover, it is interesting to point out that statistical modeling performs pretty well, even better than the single neural models (ANN). These get a remarkable accuracy, and an error slightly lower than the statistical model, only as an ensemble.
The accuracy achieved by the proposed model is such that it can be applied for intelligent monitoring, diagnostic systems and optimal control in order to reduce energy consumptions.

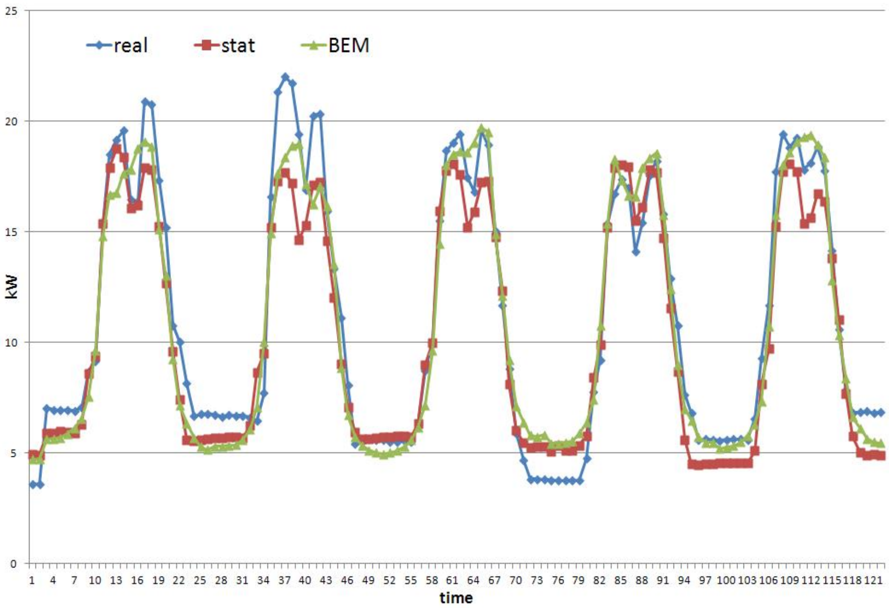


Figure 2-61: Lighting modeling comparison (testing results).

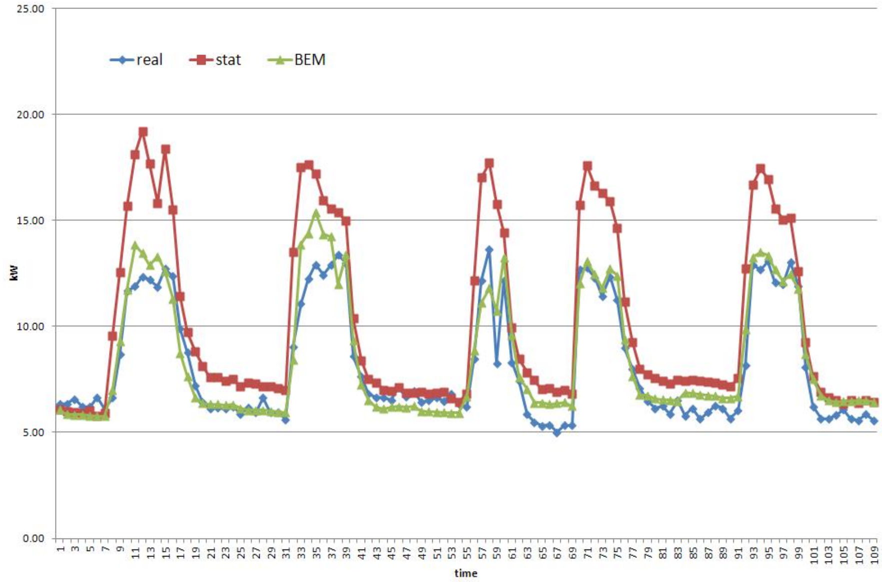


Figure 2-62: Emf modeling comparison (testing results).

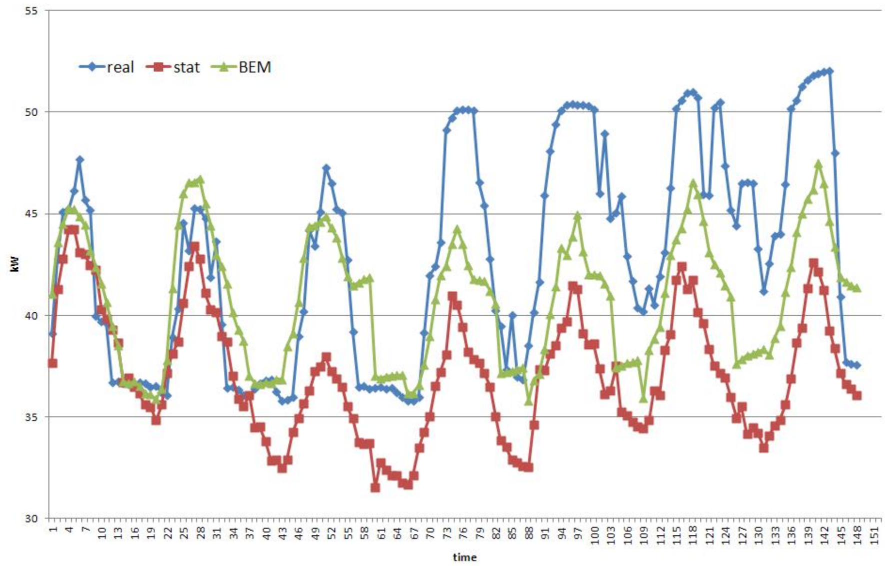


Figure 2-63: Conditioning modeling comparison (testing results).

1.6.5 **Conclusions**

In this work we proposed a new approach aimed at modeling building lighting, emf and conditioning energy consumptions. The basic idea is to build a new model based on neural networks ensembling. Experimentation has been carried out over three months data sets coming from a real office building located in the ENEA 'Casaccia' Research Centre and experimental results show that the proposed method can get a remarkable improvement with respect to the best classical method.

The reason for that is that the neural ensembling model is capable to provide reliable estimations when out of standard conditions because it takes into account several input features (as occupancy) which affect the energy consumptions.

The accuracy of the proposed model is such that it can be applied for intelligent monitoring, diagnostic systems and optimal control in order to reduce energy consumptions.
As future work we are going to apply the same approach to model other building energy consumptions as thermal flows. Moreover, we are going to try applying more sophisticated ensembling methods and also try neural - statistical hybrid models.

1.6.6 **References**

[1] United Nations Environment Programme, Buildings Can Play Key Role In Combating Climate Change (2007).

http://www.unep.org/Documents.Multilingual/Default.asp?DocumentID=502&ArticleID=5545&l=en.

[2] Yudelson J., Greening Existing Buildings, Green Source/McGraw-Hill, New York (2010).

[3] European Union, Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings. Official Journal of the European Communities (2003). EPJ Web of Conferences

[4] Pedersen L., Use of different methodologies for thermal load and energy estimations in buildings including meteorological and sociological input parameters. Renewable and Sustainable Energy Reviews, Elsevier, 11, 998-1007 (2007).

[5] Rabl A., Rialhe A., Energy signature models for commercial buildings: test with measured data and interpretation. Energy and Buildings, Elsevier, 19, 143-154 (1992).

[6] Jackson, Peter, Introduction To Expert Systems (3 ed.), Addison Wesley, p. 2, ISBN 978-0-201-87686-4 (1998).

[7] Arbib M.A., The Handbook of Brain Theory and Neural Networks, The MIT Press, Cambridge (MA) (1995).

[8] Haykin S., Neural Networks, a comprehensive foundation (2nd edition), Prentice Hall, New Jersey (1999).

[9] Caldera M., Corgnati S. P., Filippi M., Energy demand for space heating trough a statistical approach: application to residential buildings. Energy and Buildings, Elsevier, 40, 1972–1983 (2008).

[10] Rauhala K., A simple computer model for estimating the energy consumption of residential buildings in different microclimatic conditions in cold regions. Energy and Buildings, Elsevier, 15–16, 561–569 (1991).

[11] Farahbakhsh H., Ugursal V. I., Fung A. S., A residential end-use energy consumption model for Canada. International Journal of Energy Research, 22, 1133-1143 (1998).

[12] Mihalakakou G., Santamouris M., Tsangrassoulis A., On the energy consumption in residential buildings. Energy and Buildings, Elsevier, 34, 727-736 (2002).

[13] Tso G. K. F., Yau K. K. W., Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy, Elsevier, 32, 1761-1768 (2007).

[14] Rosenblatt, Frank, The Perceptron--a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory (1957).

[15] Rosenblatt, Frank. x. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC (1961)

[16] NeuralWare. Neural Computing – A technology Handbook for Professional II/PLUS and NeuralWorks Explorer. (1996).

[17] Jain A. K., Mao J. and Mohiuddin K. M. Artificial Neural Networks: A Tutorial. *Computer*, vol. 29 no. 3 (1996) 31 - 44.

[18] Hagan M. T. and Menhaj M. B. Training feedforward networks with the Marquardt algorithm. IEEE Transactions on neural networks, vol. 5, no. 6, pg. 989 – 993, November 1994.

[19] Demuth H., Beale M. and Hagan M. Neural Network Toolbox 6 User's Guide. The MathWorks, Inc. March 2010.

[20] Krogh A., Vedelsby J., Neural network ensembles, cross validation and active learning, in Tesauro G., Touretzky D. S., Leen T. K., editors, Advances in Neural Information Processing Systems, 7, 231-238, MIT Press (1995).

[21] Liu Y., Yao X., Ensemble learning via negative correlation. Neural Networks, 12(10), 1399-1404 (1999).

[22] Breiman L., Combining Predictors, in Sharkey A.J.C., (ed.): Combining Artificial Neural Nets – Ensemble and Modular Multi-net Systems, Springer, Berlin, 31-50 (1999).

[23] Perrone M. P. e Cooper L. N.. When networks disagree: ensemble methods for hybrid neural networks. Neural Networks for Speech and Image Processing by R.J. Mammone, ed. Chapman-Hall (1993).

[24] Bishop C. M.. Neural Networks for Pattern Recognition. Oxford University Press, 364 – 369 (1995).

[25] Valentini G. and Masulli F. Ensembles of learning machines. In Tagliaferri R. and Marinaro M., editors, *Neural Nets Wirn Vietri-02*, LNCS 2486 (2002) 3 - 19 Heidelberg (Germany). Series *Lecture Notes in Computer Sciences*, Springer-Verlag.

[26] Kohavi R. and Bauer E. An empirical comparison of voting classification algorithms: Bagging, boosting and variants. Machine Learning, 36 , 105 – 142 (1999).

[27] Drucker H. Improving regressors using boosting techniques. In Douglas H. Fisher, ed., ICML,Morgan Kaufmann 107 – 115 (1997).

[28] Avnimelech R. and Intrator N. Boosting regression estimators. Neural Computation, 11 491 –513 (1999).

1.7 Experience 6: Office Buildings of municipality of Livorno Ferraris

(Vincenzo Corrado et al. – Polytechnic of Turin)

1.7.1 Subject of the work

The subject of work is the analysis of the energy and environmental performance of a small office building in Italy, having a gross floor area of 1100 m^2 and a net floor area of 750 m^2. According to the Italian Building Energy Code the building is located in the climatic zone "E" (HDD between 2100 and 3000).

1.7.2 Aim of the work

In order to improve the energy efficiency in existing buildings and to design appropriate energy-saving measures, it is important to split the effect of the building features (envelope, energy systems) and the effect of the human behavior factors (heating/ventilating, system control).
The work is based on a survey on users behavior, on a long-term energy and environmental monitoring, on the application of inverse models based on linear regression and on the run of a tailored numerical simulations.

1.7.3 Database characteristics

The following building features are fully described:

- geometry (surfaces, orientation, external context);
- envelope (layers, material properties, windows);
- building services typologies and efficiencies (DHW, space heating, ventilation, lighting, common appliances).

Occupancy schedules and equipment use are fully described. The boiler energy supply has been continuously monitored on a two-year period by means of a direct method (supply and return temperatures, and water flow rate). Also an indirect method has been used to evaluate water flow rates. The following indoor and outdoor environmental quantities were also monitored (with a 15 minutes time step):

- air temperature;
- air relative humidity;
- CO2 concentration.

The building has been simulated through a numerical simulation code (EnergyPlus), using the real occupancy schedules and real environmental profiles (tailored energy rating).

1.7.4 Building data

The case study is public office "Palazzo Ciocca", built in 1860, situated in Livorno Ferraris (VC), Italy. Livorno Ferraris is a town located in North-Western Italy in Piedmont region. the number of degree days is 2549 °C·d.
The building floor has an area of 130 m^2 with a ceiling height of 3.5 m. The simulation and the monitoring were performed at the second floor.
The building structure is in bearing solid brick masonry. On the floor below the monitored zone are heated rooms (through autonomous heating systems) and also used for offices and similar, while locals on the upper floor are used for archives and use the heating system is generally off.

Figure 2-64: Satellite view

Figure 2-65: Palazzo Ciocca – portion subject to monitoring

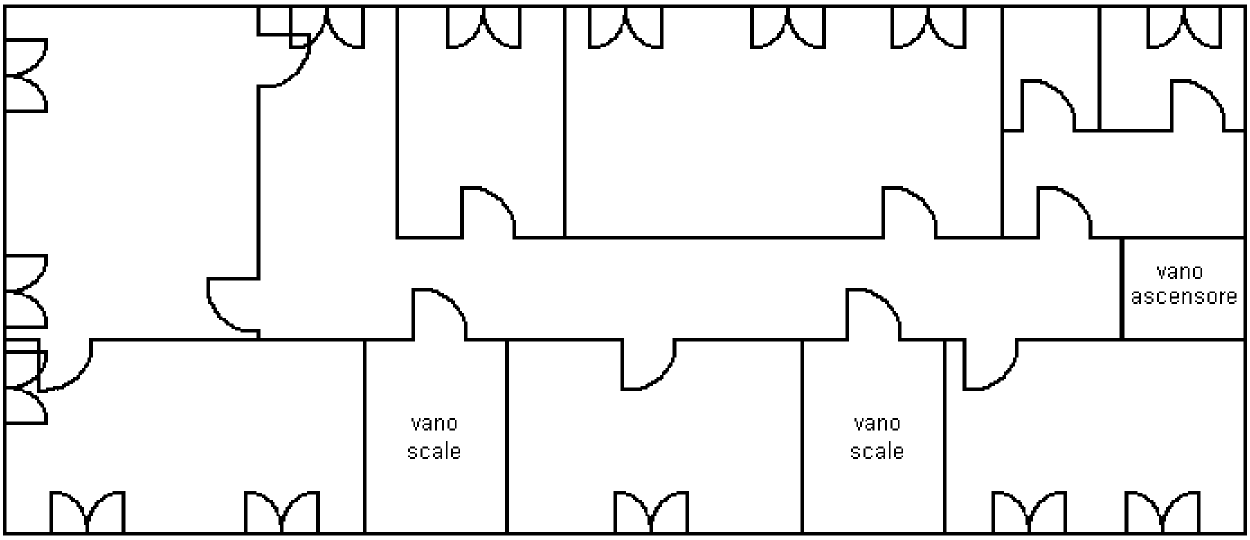


Figure 2-66: Plan of the 2nd floor above ground

Envelope

The thermo physical properties of building envelope layers are presented in Table 2-13. The glazed building façade represents 20% of the total envelope area. Windows are 4/6/4 uncoated and air filled with an U-value of 3.15 W/(m^2K) and a SHGC of 0.74. Windows are shaded from outside by blinds with reflectance (0.2) and transmittance (0.7).

Table 2-13: Envelope compositions and thermo physical properties of materials

	Layers (outer/inner)	**Thickness (cm)**	**Density (kg/m^3)**	**Specific heat (J/kg.K)**	**Thermal conductivity (W/m.K)**
External wall	Brick	72	1800	840	0.72
Ceiling/floor	Flooring screed	5	1200	1000	0.41
	Brick	20	1700	840	0.56
	Concrete	20	2000	1000	1.13

Heating system

The heating system is centralized type equipped with a cast iron heat generator of which the main characteristics are provided below:

- Type: non-condensing
- Rated power at firebox: 33,72 kW
- Output power: 29,09 kW
- Efficiency at full power at firebox: 91%
- Feed: natural gas
- Burner: atmospheric
- Year of installation: 1992
- Condition: Intact insulation coat.

The heating distribution system is zone type (floor collector) and the pipes are embedded in the floor. The circulation pump has the following features:

Table 2-14: Heating system features

Power supply	220 V
	0.66/0.5 A
P1 (max)	132/99 W
P2 (min)	49/22 W
Giri	2000/2400 m^{-1}

A programmable control (timer) is installed (Figure 2-67) and a zone thermostat (Figure 2-68) is placed in room 9, closet o the "I" measurement point.

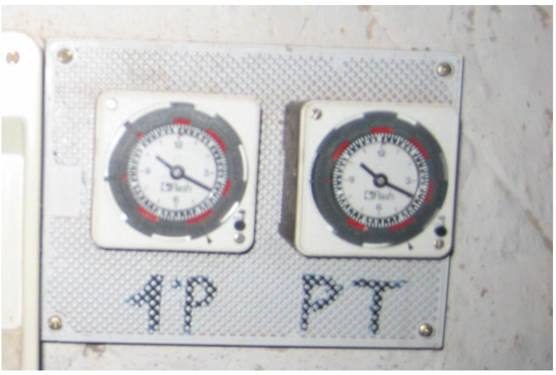


Figure 2-67: Timer

Figure 2-68: Zone thermostat

The terminals are cast iron radiators in columns. All the radiators except i, m, p, q and r are cavities below window sills. In Figure 2-69 the position of the radiators id identified.

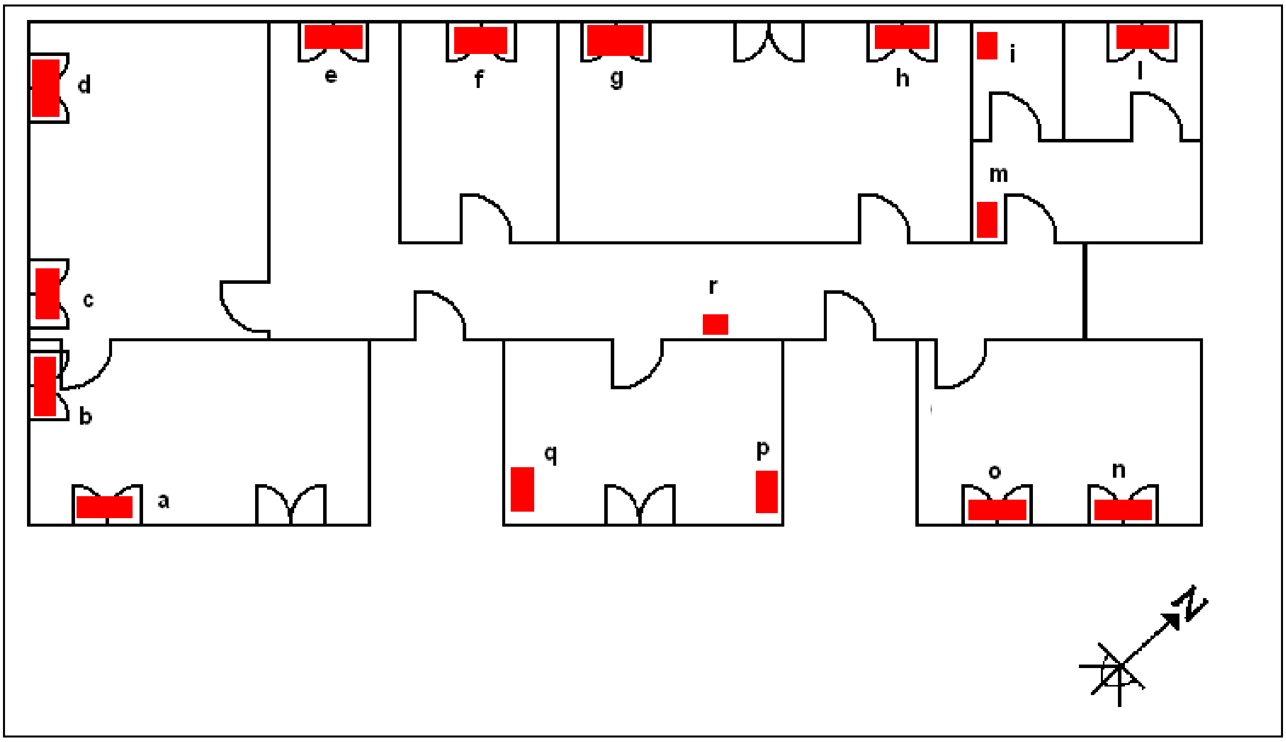


Figure 2-69: Position of the radiators

Table 2-15: Characteristics of the radiators in the monitored rooms

	Id	Elements	Columns	Height [mm]
	a	12	4	600
	b	10	4	600
	c	6	4	600
	d	14	4	600
	e	8	4	600
	f	8	4	600
	g	10	4	600
	h	10	4	600
	i	3	3	840
	l	7	4	600
	m	6	3	840
	n	13	4	600
	o	13	4	600
	p	10	3	840
	q	10	3	840
	r	4	3	840

Building use
The building office is occupied from 9 to 18 h during weekdays. The internal heat sources are 15 W/m^2 and the infiltration is fixed to 0.7 h^{-1}. A natural ventilation rate is scheduled with a nominal rate of 3 h^{-1}. The shading device is scheduled with a solar set-point of 120 W/m^2.

Monitoring
The following quantities were monitored:

- external temperature
- external relative humidity
- external CO_2 concentration
- internal temperature
- internal relative humidity
- internal CO_2 concentration
- thermal energy delivered to the system

The environmental monitoring has been carried out both by means of wireless sensors and by means of traditional data-loggers.

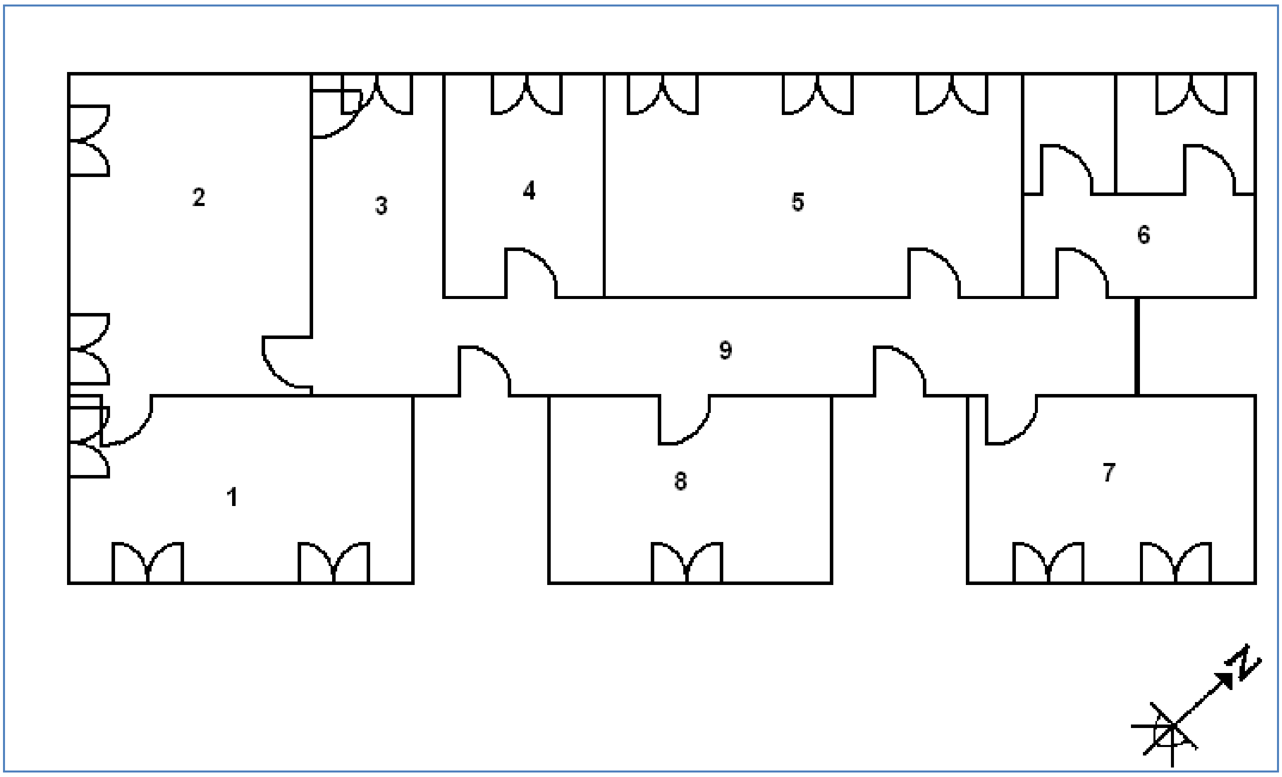


Figure 2-70: Identification of the rooms

Hereafter are reported, for each room, the geometrical features, the occupancy schedule and the type of climatic control.

Table 2-16: Occupancy characteristics of the monitored rooms

Room	Occupancy schedule	Equipment	Notes

1	1 person from Monday to Friday – time 8-14 and 15-18	1 PC, 1 printer	
2	2 persons from Monday to Friday – time 8-14 and 15-18	2 PCs, 2 printers	The doors of separation of room 2 from room 3 are open for most of the occupancy period
3	Occasionally occupied	2 printers	The doors of separation of room 2 from room 3 are open for most of the occupancy period
4	Occasionally occupied	No equipment	The doors of separation of room 4 from room 9 are open for most of the occupancy period
5	2 persons from Monday to Friday – time 8-14 and 15-18	2 PCs, 2 printers	
6	See notes	1 small refrigerator, 1 microwave oven, 1 automatic bar machine, 1 distributor of beverages, 1 boiler for DHW production	Bathrooms and service room
7	1 person from Monday to Friday – time 8-14 and 15-18	2 PCs, 2 printers	
8	Occasionally occupied	No equipment	

1.7.5 Method/Methods applied for the data analysis

The method for data analysis includes two steps:

- simplified analysis based on a simplified inverse model;
- calibration of the numerical detailed model.

The first step applies linear regression based on weekly and daily data for determining correlations between heating energy need and average external temperature. Such regression allows to determine the total heat loss coefficient and the influence of solar and internal heat gains.

By knowing solar radiation data and occupancy schedules, it is also possible to split the effects of solar gains and internal gains.

Besides, the analysis of internal temperature drop due to thermostat set-back or switch-off of the heating plant allows to determine the effective thermal capacity of the building.

In the second step the numerical model has been calibrated by comparing both expected energy need and the real measured consumption, and the expected and real aggregated parameters springing up from the first analysis.

The aim is to built a data-driven model and to evaluate energy saving. The construction of inverse models is based on the following assumptions:

- Dependent variables: energy consumption for heating and cooling, obtained with the detailed simulation tool EnergyPlus;
- Independent variables: external air temperature and sol-air temperature.

The procedure of inverse model construction is based on the least-squares regression method (Kissock et al. 2003). This approach estimates model coefficients, β, that minimize the sum of the squared error, E, between predicted, $\hat{Y}$, and actual observations, Y, following this equation:

$$Y = X \cdot \beta + E \qquad (1)$$

The root mean squared error, RMSE, is to be minimized, computed as:

$$RMSE = \sqrt{\frac{\sum (Y - \hat{Y})^2}{(n - p)}} \tag{2}$$

where n is the number of data observations, p is the number of regression coefficients. The root mean squared error of the model is a measure of the scatter of the data around the model.

The second task is to evaluate the energy savings following the realization of several retrofit actions. In the case-study, data driven approach is carried out with simulated data because as measurement operation is still on progress.

The method is based on the following steps:

- Assessment of energy consumption for heating and cooling in day time intervals during the pre and post retrofit periods using the dynamic tool EnergyPlus;
- Construction of data-driven models;
- Calculation of the energy saving for each retrofit action.

Simulation

The energy evaluation of the building is carried out with EnergyPlus dynamic tool for six different combinations of climate data and building data characteristics: (i) pre-retrofit simulation performed with climate and building data in the pre-retrofit period (ii-v) post-retrofit simulations with climate and building data in the post-retrofit period and (vi) final simulation with post-retrofit climate data and pre-retrofit building data. The pre and post-retrofit climatic data have been modelled with two different cities in the same climatic zone.

Table 2-17: Description of Retrofitting actions

Retrofitting actions	Energy aspect	Load characterization	Force driven
Infiltration reduction $Q_{pre\text{-}retro}$= 0.7 h^{-1} $Q_{post\text{-}retro}$= 0.2 h^{-1}	Heat ventilation losses	Instantaneous load	External temperature
Decreasing window solar transmittance $SHGC_{pre\text{-}retro}$ = 0.74 $SHGC_{post\text{-}retro}$ = 0.39	Heat sources	Delay load	Solar irradiation
Adding external envelope insulation $Uop_{pre\text{-}retro}$ = 0.89 W/m^{2}K $Uop_{post\text{-}retro}$ = 0.28 W/m^{2}K	Heat transmission losses	Delay load	External temperature and solar irradiation
Double window with argon gas filling $Uw_{pre\text{-}retro}$ = 3.15 W/m^{2}K $Uw_{post\text{-}retro}$ = 2.55 W/m^{2}K	Heat transmission losses	Quasi-instantaneous load	External temperature

1.7.6 **Results**

A better understanding of the building heat balance and of the influence of users have been achieved by combining an inverse analysis based on energy and environmental monitoring and a calibrated direct tailored modeling.

A few aggregated parameters have been defined to describe the building thermal behavior: global heat transfer coefficient, thermal capacity, solar effective area, user parameters.
Table 2-18 shows the regressions coefficients and uncertainty parameters of the data driven model using the (2-P), (3-P), (4-P) and (5-P) models for the heating and cooling mode and respectively using sol-air temperature and air temperature as independent variables. Generally speaking, the squared correlation coefficient, R^2, has a good value (more than 0.6). It can be noted that the dry air temperature is more appropriate than the sol-air temperature in the case of cooling data driven model. At the opposite, the sol-air temperature is more suitable during the heating mode. Besides, table IV illustrates that the balance temperature T_{bal} has a low value: in the heating mode around 12.5 °C and in the cooling mode around 20°C, this is due to the fact the conditioning system runs in intermittent mode.

Table 2-18: Uncertainty parameters and regression coefficients for models analyzed

Independent variable		Model	Uncertainty Parameters			Regression coefficients				
			R^2	RMSE	CV-RMSE	1 [kWh]	2 [kWh/°C]	3	4 [°C]	5 [kWh]
Sol-air temperature	Cooling	2P	0.646	10.839	42.71%	-167.88	4.15			
		3PC	0.648	10.809	42.60%	3.32	4.30	23.76 [°C]		
		4P	0.657	10.758	42.39%	34.01	4.89	2.39 [kWh/°C]	30.14	
	Heating	2P	0.678	16.201	32.83%	160.65	-4.40			
		3PH	0.701	15.592	31.60%	6.53	-5.15	15.56 [°C]		
		4P	0.702	15.65	31.72%	14.22	-5.31	-1.33 [kWh/°C]	13.78	
		5P	0.744	4.061	34.60%	4.76	-5.31	3.87 [kWh/°C]	15.56	23.33
Dry-bulb temperature	Cooling	2P	0.784	8.473	33.39%	-224.41	6.29			
		3PC	0.829	7.537	29.71%	5.09	7.88	19.71 [°C]		
		4P	0.83	7.566	29.82%	6.72	1.07	7.87 [kWh/°C]	19.91	
	Heating	2P	0.634	17.263	34.99%	173.73	-5.44			
		3PH	0.642	17.067	34.59%	5.71	-5.80	12.56 [°C]		
		4P	0.642	17.146	34.75%	7.51	-5.81	-0.65 [kWh/°C]	12.22	
		5P	0.733	14.372	35.36%	3.21	-5.79	7.19 [kWh/°C]	13.01	19.05

In figures 2-71, 2-72, 2-73 the building energy needs for heating and for cooling are presented as a function of the driving forces T and T_{sa} as well as the fit by baseline equation for the regression model analyzed.

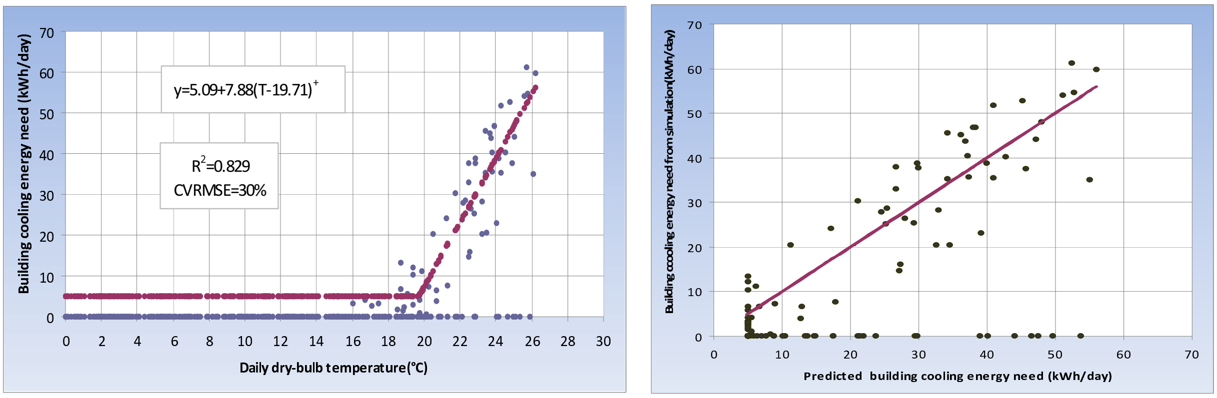


Figure 2-71: Left: Building cooling energy need as a function of dry-bulb air temperature and fit by baseline equation (3Pc) Right: Simulated vs predicted building cooling energy need over the pre-retrofit period.

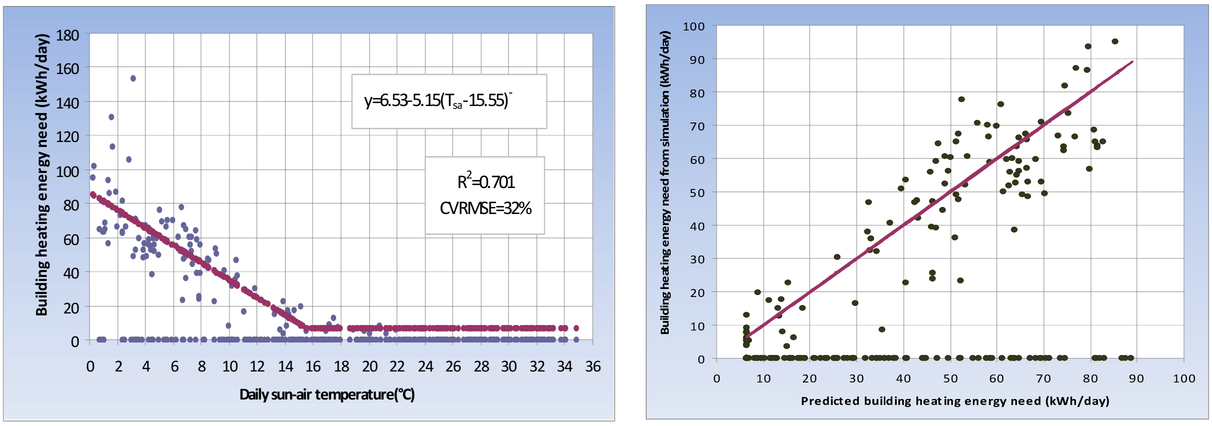


Figure 2-72: Left: Building heating energy need as a function of soil-air temperature and fit by baseline equation (3Ph) Right: Simulated vs predicted building heating energy need over the pre-retrofit period.

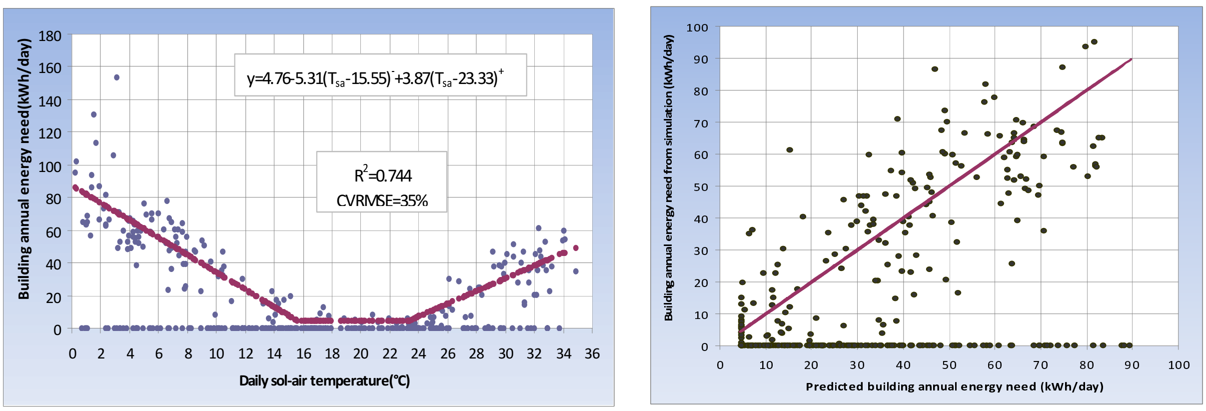


Figure 2-73: Left: Building annual energy need as a function of sol-air temperature and fit by baseline equation (5P) Right: Simulated vs predicted building annual energy need over the pre-retrofit period.

1.7.7 References

[1] Capozzoli A.; Corrado V.; Mechri H.E. (2009). A critical review on forward and data-driven methods for building energy analysis. In: Systems, Energy and Built Environment. Towards a Sustainable Comfort. Proceedings of the 47th AICARR International Conference, Tivoli, 8-9 October 2009, Milano: AICARR, p. 885-896, ISBN 978-88-95620-53-4.

[2] Claridge D. (1998). A perspective on methods for analysis of measured energy data from commercial buildings. ASME Journal of Solar Energy Engineering. Vol. 120, pp. 150-155.

[3] Corrado V.; Mechri H.E. (2009). Uncertainty and Sensitivity Analysis for Building Energy Rating. Journal of Building Physics, vol. 33, no. 2, p. 125-156, ISSN: 1744-2591, October 2009, doi: 10.1177/1744259109104884.

[4] Corrado V.; Mechri H.E. (2009). Uncertainty analysis on simplified and detailed calculation models for heating and cooling energy assessment of buildings. In: BAYAZIT, MANIOĞLU, ORAL & YILMAZ. Energy Efficiency and New Approaches. Proceedings of the 4th International Building Physics Conference, Istanbul, 15-18 Giugno 2009, ISTANBUL: Istanbul Technical University (Turkey), p. 1007-1013, ISBN: 978-975-561-350-5.

[5] Kissock J.; Kelly; Haberl Jeff S.; Claridge David E. (2003). Inverse Modeling Toolkit: Numerical Algorithms, KC-03-2-1 (RP-1050).

[6] Tiberiu C.; Virgone J., Blanco E. 2008. Development and validation of regression models to predict monthly heating demand for residential buildings. Energy and Buildings Vol 40, pp. 182

1.8 Experience 7: Evaluation of a Low Energy Multifamily building in Vienna, Austria (Synthetic contribution)

(Thomas Bednar, Kerstin Seif, Naomi Morishita. Wien University of Technology)

1.8.1 Subject of the Work

Three Austrian apartment buildings (Kammelweg D, Utendorfergasse, Dreherstraße) studied at a high level of detail, Level C, complex, according to the "3 Level Database" definition.

1.8.2 Aim of the Work

The aims of this research are:

- to determine an accurate profile of the current typical Austrian household considering direct and indirect influences on user behavior and energy demand;
- to verify if the profile of a single average household is satisfactorily accurate to represent the energy-related behavior of the Austrian population;
- to see how energy-related user behavior has changed over time;
- to establish what the most essential influencers are on energy-related user behavior.

1.8.3 Database Characteristics

Two groups of data are to be generated for the database: building geometries and qualities and occupancies and lifestyles.
The database contains the results of a regression analysis of key influencers of energy-related user behaviors. The parameters representing the range of possible user behaviors are selected for the database using the results of previous research into the behavior trends, and five building energy efficiency groups:

- Existing buildings
- Low Energy buildings
- Low Energy buildings with renewable energy systems (PV panels, solar hot water panels, etc.)
- Passive House buildings
- Plus Energy buildings

The case study apartment buildings located in different parts of Vienna are used as the basis for the scenario combinations. The apartment buildings are analyzed in detail at Level C, complex.
The interplay of both groups of parameters are combined using random regression analyses to form a library of potential combinations of building standard, occupancy, and lifestyle.

1.8.4 Data Analysis Methodology

The first group of data will be comprised of the building performance characteristics of five building efficiencies as listed above.
The second group of data will be based upon the literature research, where parameters influencing energy-related behaviors will be identified and sub-categorized. The dominant occupancy schedules will also be included in this portion of the database.

The two data groups will be combined using regression analysis to establish a series of user profiles, and compared to determine if a dominant profile results.
The dominant parameters, as determined by the regression analyses, will then be used as the input parameters for whole building simulation models of the apartment buildings.

1.8.5 Expected Results

The dominant characteristics of energy-related user behavior will be graphically compared to ascertain if a single profile dominates each building typology. If a consistent profile dominates all typologies, this profile can be defined as an accurate user profile for the typical Austrian household to be used as the standard profile for energy certificate calculations and whole building simulations.

1.9 Experience 8: Development of statistical analysis for total energy use in individual buildings (Synthetic contribution)

(Thomas Bednar, Kerstin Seif, Naomi Morishita. Wien University of Technology)

1.9.1 Subject of the Work

Eight Austrian single family homes between 100 m² and 220 m² studied at a high level of detail, Level C, complex, according to the “3 Level Database” definition.

1.9.2 Aim of the Work

The aims of this research are,
to determine an accurate profile of the current typical Austrian household considering direct and indirect influences on user behavior and energy demand;
to verify if the profile of a single average household is satisfactorily accurate to represent the energy-related behavior of the Austrian population;
to see how energy-related user behavior has changed over time;
to establish what the most essential influencers are on energy-related user behavior.

1.9.3 Database Characteristics

Two groups of data are to be generated for the database:

- building geometries and qualities
- occupancies and lifestyles.

The database contains the results of a regression analysis of key influencers of energy-related user behaviours. The parameters representing the range of possible user behaviors are selected for the database using the results of previous research into the behavior trends, and five building energy efficiency groups:

- Existing buildings
- Low Energy buildings
- Low Energy buildings with renewable energy systems (PV panels, solar hot water panels, etc.)
- Passive House buildings
- Plus Energy buildings

Eight case study single family homes located in different parts of Austria are used as the basis for the scenario combinations. The homes are analyzed in detail at Level C, complex.
The interplay of both groups of parameters are combined using random regression analyses to form a library of potential combinations of building standard, occupancy, and lifestyle.

1.9.4 Data Analysis Methodology

The first group of data will be comprised of the building performance characteristics of five building efficiencies as listed above.

The second group of data will be based upon the literature research, where parameters influencing energy-related behaviors will be identified and sub-categorized. The dominant occupancy schedules will also be included in this portion of the database.
The two data groups will be combined using regression analysis to establish a series of user profiles, and compared to determine if a dominant profile results.
The dominant parameters, as determined by the regression analyses, will then be used as the input parameters for whole building simulation models of the eight homes.

1.9.5 **Expected Results**

The dominant characteristics of energy-related user behavior will be graphically compared to ascertain if a single profile dominates each building construction typology. If a consistent profile dominates all typologies, this profile can be defined as an accurate user profile for the typical Austrian household to be used as the standard profile for energy certificate calculations and whole building simulations.

1.10 Experience 9: Development of statistical analysis for total energy use in small office buildings (Synthetic contribution)

(Thomas Bednar, Kerstin Seif, Naomi Morishita. Wien University of Technology)

1.10.1 Subject of the Work

Two small Austrian office buildings (Getreidemarkt, BH Melk) studied at a high level of detail, Level C, complex, according to the “3 Level Database” definition.

1.10.2 Aim of the Work

The aims of this research are,
to determine an accurate profile of the current typical Austrian office worker considering direct and indirect influences on user behavior and energy demand;
to verify if the profile of a single average office is satisfactorily accurate to represent the energy-related behavior of the Austrian population;
to see how energy-related user behavior has changed over time;
to establish what the most essential influencers are on energy-related user behavior.

1.10.3 Database Characteristics

Two groups of data are to be generated for the database: building geometries and qualities and users and lifestyles.

The database contains the results of a regression analysis of key influencers of energy-related user behaviours. The parameters representing the range of possible user behaviors are selected for the database using the results of previous research into the behavior trends, and five building energy efficiency groups:

- Existing buildings
- Low Energy buildings
- Low Energy buildings with renewable energy systems (PV panels, solar hot water panels, etc.)
- Passive House buildings
- Plus Energy buildings

Two case study office buildings located in different parts of Austria are used as the basis for the scenario combinations. The buildings are analysed in detail at Level C, complex.
The interplay of both groups of parameters are combined using random regression analyses to form a library of potential combinations of building standard, occupancy, and lifestyle.

1.10.4 Data Analysis Methodology

The first group of data will be comprised of the building performance characteristics of five building efficiencies as listed above.
The second group of data will be based upon the literature research, where parameters influencing energy-related behaviours will be identified and sub-categorized. The dominant occupancy schedules will also be included in this portion of the database.

The two data groups will be combined using regression analysis to establish a series of user profiles, and compared to determine if a dominant profile results.

The dominant parameters, as determined by the regression analyses, will then be used as the input parameters for whole building simulation models.

1.10.5 **Expected Results**

The dominant characteristics of energy-related user behavior will be graphically compared to ascertain if a single profile dominates each building typology. If a consistent profile dominates all typologies, this profile can be defined as an accurate user profile for the typical Austrian office building to be used as the standard profile for energy certificate calculations and whole building simulations.

2. Statistical analysis of large building stock

2.1 Introduction

Statistical analysis of a large buildings stock represent methods used to estimate the energy consumption and/or the peak demand of a building at a level of detail that is suited to apply to a number of buildings that is statistical significant (usually more than tens of buildings). The principle of the approach is to project the experimental data on a basis. The methods depend on the type of basis: its dimension and its components.

One type of projection is on categories (Experiences 1, 2, 3, 4, 5, 6). For example, Hu and Yoshino (Experience 4) consider the climate zones, the area of the building, the type of the heating system and its operation, as well as the number of people in the household and their annual income. In another study, Yoshino (Experience 5) considers, besides the categories mentioned before, the weather, indicated by the cooling and heating degree days and the indoor temperature during the heating and the cooling season. The model resultes are regression models in different variants: multi regression, neural networks, quantification methods (Experiences 1, 2, 5).

Categorizing reduces the variance of the predicted results. The physical explanation of the result is embedded in the categories. Usually, these approaches do not differentiate between the inputs (e.g. weather), the parameters (e.g. floor area, total heat loss coefficient) and the outputs (e.g. indoor air temperature) of a physical (or direct) model. The results indicate the influence of each category given by the weighting coefficient in the model.

This kind of approach, which uses less data (in fact the data available), is very effective in practice. It allows the prediction of energy consumption with an expected variance for real buildings by using data which are available mainly on monthly and/or annual bills.

Comparison between categories needs a criterion which “normalizes” the consumption in order to negate the effect of parameters specific to a given building. For example, Corgnati et al. (Experience 6) propose and demonstrate the application of an indicator that normalizes the data as a function of the heated volume and the climate, described by the degree days of the site.

The second class of projection is on parameters of physical models. The main idea in this approach is to consider a physical model based on heat balance and to identify the parameters of this model which increase the fit between the predicted results and the measurements. One of the most common approaches is to use the load curve, which expresses the dependence of the heating (or cooling) consumption on the outdoor temperature. This “thermal signature” of the building can be used together with the distribution of degree-days or degree-hours in order to estimate the energy consumption (e.g. the bin method). Basically, the building signature is obtained by regression. Robust regression may be used to improve the prediction in case of perturbation such as the usage of the building (Experience 7). The advantage of this approach is that the thermal behavior of the building, the comfort and the climate are decoupled.

A variant of this method is to use the free-running temperature, which allows the estimation of the energy savings for cooling by using free-cooling by ventilation (Experience 8).

Refinements of the thermal signature or the load curve method are proposed (Experiences 8, 9). Ghiaus (Experience 8) demonstrated the equivalence between the load curve and the free running temperature. By using the free-running temperature, the whole range of building operation (heating, ventilation and cooling) is described by a single concept.

Normally, thermal signature is a static method. However, the heat balance may be written taking in account the accumulation. By doing so, Danov et al. (Experience 7) obtained a dynamic model which can estimate the influence of the thermal mass of the building on the energy consumption. Solar gains may be also included in the thermal signature, reducing the variance of the energy estimation (Experience 7).

References (contributions to the Annex 53)

Extended contribution

Experience 1: Sawako Nakamura, Hiroshi Yoshino, Ayako Miura. Statistical analysis for energy consumption of office buildings in Japan

Experience 2: Sawako Nakamura, Hiroshi Yoshino, Ayako Miura. Statistical analysis for energy consumption in residential buildings in Sendai

Experience 3: Hiroshi Yoshino, Ayako Miura. Survey of the peak electricity in residential buildings (see Experience 1 in Individual Buildings for details)

Experience 4: Tianchi Hu, Hiroshi Yoshino. Statistical analysis on energy consumption of residential buildings in China

Experience 5: Hiroshi Yoshino. Field Survey and Statistical Analyses on Energy Consumptions in the Residential Buildings in Japan

Experience 6: Stefano Paolo Corgnati, Federica Ariaudo, Marco Filippi. Heating consumption assessment and forecast of existing buildings: investigation on Italian school buildings

Experience 7: Stoyan Danov, Jordi Carbonell, Jordi Cipriano. Building energy performance evaluation using daily consumption data

Synthetic contributions

Experience 8: Cristian Ghiaus. Experimental estimation of building energy performance by robust regression

Experience 9: Cristian Ghiaus. Equivalence between the load curve and the free-running temperature in energy estimating methods

Experience 10: Zhun Yu, Fariborz Haghighat. Mining Hidden Patterns from Real Measured Data to Improve Building Energy Performance

2.2 Experience 1: Statistical analysis for energy consumption of office buildings in Japan (Sawako Nakamura, Hiroshi Yoshino, Ayako Miura)

2.2.1 Introduction

A detailed database is important for an effective measure to reduce CO_2 emissions from non-residential building sector. Therefore, a national wide project DECC (Data-base for Energy Consumption of Commercial Buildings) was founded in 2007. The goal of the project was to understand the actual conditions for energy consumption of non-residential buildings in Japan. This report is authored corresponding to Reference[1].

2.2.2 Aim of the analysis

Goals of the survey are:
Understanding basic information and introduction of energy-saving measures of buildings
Understanding the average energy consumption per unit of floor area by building usage.
In order to identify the influential factor on the energy consumption, the multiple regression analysis was done.

2.2.3 Database characteristics

Number of buildings: 1128 office buildings(distributed in 8 different districts)
Period: April 1st, 2007 to March 31st, 2008
Questionnaire survey
Contents: Building information including location, floor area, annual energy consumption data, and energy saving measures.
Online database: unavailable

2.2.4 Research method

Questionnaire survey
Questionnaire survey was carried out in office buildings located in eight districts in Japan. The survey was conducted from April 1st, 2007 to March 31st, 2008. The questionnaire sheets ware sent to the building owners or building managements. Investigation contents are shown in Table 2-19. Questionnaire has 2 parts: building characteristic and energy consumption. Figure 2-74 shows the number of the valid data by district. We obtained 1128 valid data in Japan.

Table 19: Investigation contents

Building characteristic	Location,floor area,building area,parking area,storey,completed year,office hours,airconditioning period, etc…
Energy Consumption	Annual/monthly consumptions for electricity,city gas,LPG(Liquefied Petroleum Gas),heavy oil,kerosene,DHC(District Heating and Cooling) and others

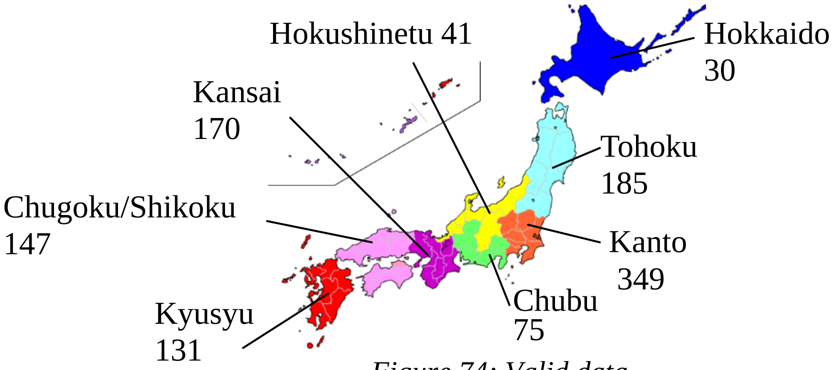


Figure 74: Valid data

2.2.5 **Results**

Building information

Figure 2-75 shows building scales. In Kanto, large scale buildings (more than 10000 m^2) accounted for about 60%, since there were lots of high-rise buildings in Tokyo area. On the other hand, in Tohoku and Chugoku/Shikoku, percentage of small scale buildings (up to 2000 m^2) was high. Figure 2-76 explains combination of energy sources. The combination of electricity and city gas were the most common energy source in Japan. While in Hokkaido and Tohoku, the percentage of buildings using oil such as kerosene, heavy oil were relatively high. The reasonable explanation could be that they have long and cold winters, so oil was more practical to be used.

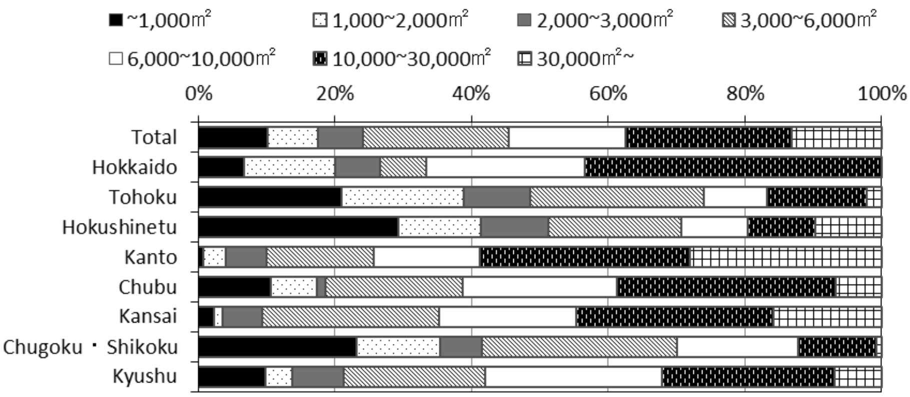


Figure 75: Building scales

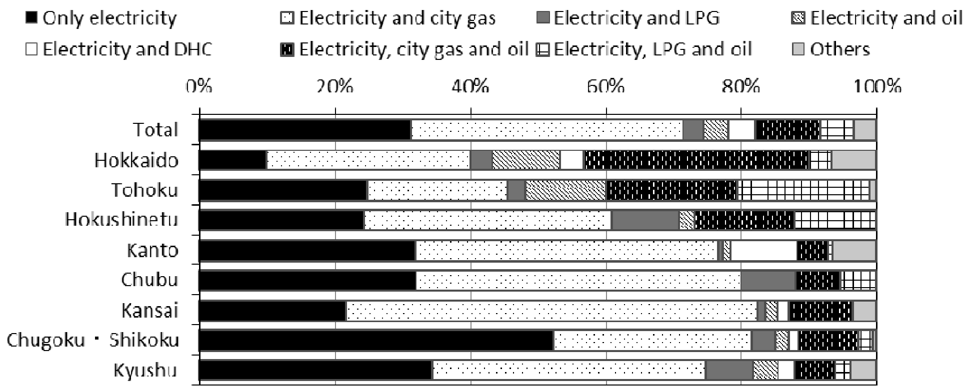


Figure 2-76: Energy source

Operation of heating and cooling

Figure 2-77 illustrates the cooling period. In Hokkaido and Tohoku, cooling period were shorter than the others. Some of the buildings in other regions were using cooling systems throughout the whole year. In Kanto, cooling period was the longest, but Kanto is not located in southern part in Japan. The explanation could be that there are many large scale buildings with many workers and office equipment such as computers and photocopiers, so that the office temperature tends to rise due to internal heat gain. Figure 2-78 shows heating period of each region. Generally in Japan most of the buildings used heating system from December to March. Heating periods of the buildings in Hokkaido were the longest which is from November to April.

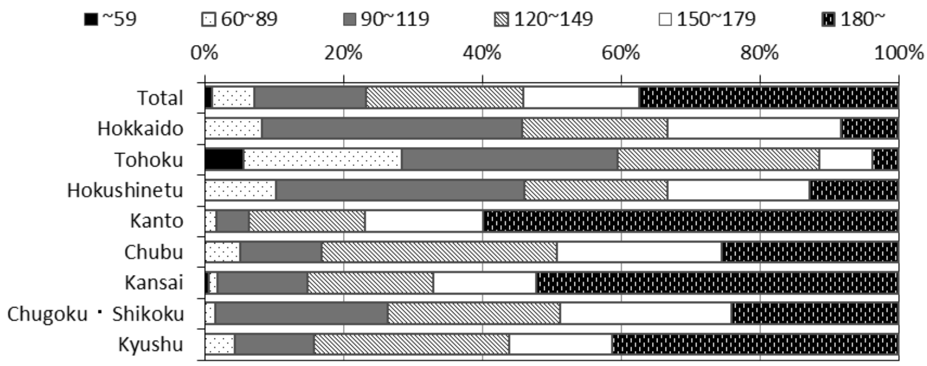


Figure 2-77: Cooling period

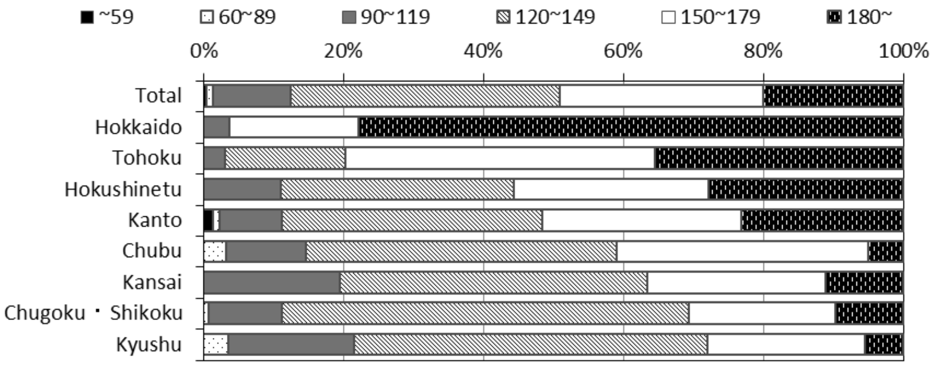


Figure 2-78: Heating period

Energy consumption

Figure 2-79 shows the relations between annual primary energy consumption and floor area. The correlation between energy consumption and floor area was strong. Figure 2-80 shows annual energy consumption per square meter by region. The average of energy consumption per square meter was 1738[MJ/ m^2]. In Kanto, annual primary energy consumption per square meter was higher than the others. On the other hand, in Tohoku and Hokushinetu, it was lower.

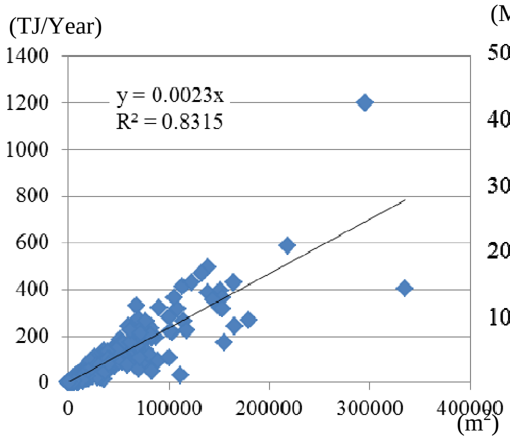


Figure 79: Relations between annual primary energy consumption and floor area

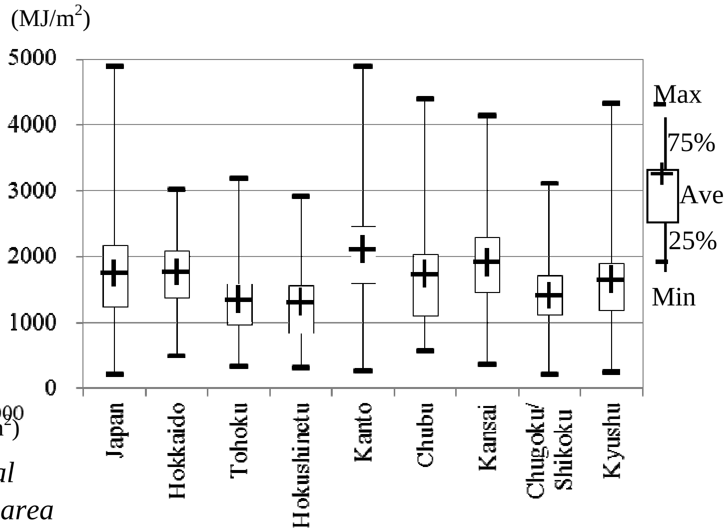


Figure 80: Annual energy consumption per square meter

Multiple regression analysis

In order to understand the influential factors which determine the energy consumption, multiple regression analysis was carried out by SPSS statistic 18. In this analysis, backward selection method was used. By using this method, influential factors with low impact were removed. Annual total energy consumption and annual energy consumption per square meter were set as the dependent variable. Some factors such as floor area, completed years, cooling period, heating period were selected as independent variables. To avoid multicollinearity, storey and cooling degree day were not used. Table 2-20 shows result of multiple regression analysis for total energy consumption. The coefficient of determination was 0.843. Cooling period, density of users, weekly business hours and completed years were removed by backward selection method. The standardized partial coefficient of floor area was 0.901. For total energy consumption, floor area was the largest impact factor. Table 2-21 shows the result of multiple regression analysis for energy consumption per square meter. The coefficient of determination was 0.255. Completed years and heating period were removed by backward selection method. Cooling period (0.264), density of users (0.194), floor area (0.143) were relatively high in standardized partial regression coefficient. These factors have certain effect on energy consumption per square meter. If you look at heating degree day, you can find the partial regression coefficient have a minus sign. It means that energy consumption in building which is located in region with cold winter is smaller than that in other regions.

Table 2-20: Result of multiple regression analysis for total energy consumption

Independent variables	Unit	Partial regression coefficient	Standardized partial regression coefficient	P-value
Floor area	m^2	2.389	0.901	0.000
Heating period	days/year	154.190	0.077	0.000
Heating degree day	℃・day	-5.743	-0.042	0.005
Cooling period	days/year	-	-	-
Density of users	person/m^2	-	-	-
Weekly business hours	hours/week	-	-	-
Completed years	Year	-	-	-
Constant	-	-15532.665		0.001

Table 2-21: Result of multiple regression analysis for energy consumption per square meter

Independent variables	Unit	Partial regression coefficient	Standardized partial regression coefficient	P-value
Cooling period	days/year	2.784	0.264	0.000
Density of users	person/m^2	6631.338	0.194	0.000
Floor area	m^2	0.004	0.143	0.000
Weekly business hours	hours/week	3.280	0.119	0.000
Heating degree day	℃・day	-0.087	-0.065	0.043
Completed years	Year	-	-	-
Heating period	days/year	-	-	-
Constant	-	914.074		0.000

2.2.6 **Conclusions**

In this study, the outline of Data-base for Energy Consumption of Commercial Building was shown. In order to understand the actual usage condition of office building, the investigation was held in 2008, where 1128 valid data were obtained. In Kanto, cooling period was longer than the others, because there were many buildings with more internal heat from high density of workers in them. The national averages of the annual primary energy consumption per square meter was 1738[MJ/m^2・Year]. In Kanto, the average annual energy consumption per square meter was higher than the other regions. The explanation could be that there are many high-rise buildings with many carrier devices and long cooling period. On multiple regression analysis for total energy consumption, the coefficient of determination was high, because total energy consumption and floor area have a strong linear relationship. On the other hand, on multiple regression analysis for energy consumption per square meter, the coefficient of determination for was not so high. However, cooling period, density of users and floor area have some effects on energy consumption per square meter. From the results, it's important to reduce internal heat load for energy saving in office buildings.

Acknowledgments

The examining board for Data-base for Energy Consumption of Commercial Buildings, or DECC (Chairperson: Shuzo Murakami, Chief Executive of Building Research Institute) is seated in Japan Sustainable Building Consortium (before 2008, in Institute for Building Environment and Energy Conservation). The board is supported by Ministry of Land, Infrastructure and Transport. This study was carried out as a part of the project. The authors would like to thank all the people who involved in this study.

2.2.7 Reference

[1] Nakamura, H. Yoshino, S. Murakami, K. Bogaki, K. Matsunawa, S. Kametani, H. Takaguchi, H. Hanzawa, M. Okumiya, Y. Asano, Y. Shimoda, S. Murakawa, T. Watanabe: Statistical Analysis for Energy Consumption of Office Buildings in Japan, Proceedings of the 7th International Symposium on Heating, Ventilating and Air Conditioning (ISHVAC 2011), pp.529-534, Shanghai, China, 2011.11

2.3 Experience 2: Statistical analysis for energy consumption of residential buildings in Sendai

(Sawako Nakamura, Hiroshi Yoshino, Ayako Miura)

2.3.1 Introduction and aim of the analysis

The energy consumption of residential sector has been increasing significantly. Therefore, it is necessary to analyze how much energy is used by various sources in order to reduce energy consumption in residential buildings. In this study, questionnaire survey has been distributed to clarify actual pattern of yearly energy consumption in Sendai. Sendai is one of the cities in Tohoku region. Regarding climate condition, Sendai is cold and snowy in winter, but is hot with high humidity in summer.

Quantification method 1 was conducted based on the results of the measurements, so as to find out the influential factors on residential energy consumptions in Sendai city.

This report was authored corresponding to a paper Ref [1].

2.3.2 Database characteristics

-Number of buildings: 1274 houses

-Period: October, 2007 to March, 2009

- Questionnaire survey

-Contents: number of household, floor area, energy consumptions, housing structure, lifestyles and energy saving consciousness and so on.

-Online database: unavailable

2.3.3 Method

In order to understand the influential factors which determine the energy consumption, quantification method 1 was carried out by SPSS statistic 18. Quantification method 1 analyzes qualitative factors. Annual energy consumption for cooling, heating and hot water supply were set as the dependent variable one by one. Some factors such as number of household, floor area, completed years, house type, and occupant behavior were selected as independent variables.

2.3.4 Results

Figure 2-81 shows the basic information of the households. In terms of house type, detached houses accounted for about 57% of the total houses and apartment houses accounted for about 43%. As for detached houses, wood construction accounted for about 80% of the total. As for apartment houses, RC (reinforced concrete) was the most common construction, which accounted for about 70%, on the other hand, wood construction accounted for about 17% of the total. The average floor area of a detached house was 131.1 m^2 and that of an apartment house was 55.9 m^2. The average of number of family members was about 2.7 people.

Figure 2-82 shows the frequency distribution of the total energy consumption. The energy consumption was converted by using the energy conversion value for each heat source, which are Electricity:3.6MJ/kWh, City gas:45.0MJ/m^3, LPG:100.5MJ/m^3, Kerosene:36.7MJ/l. The annual average energy consumption was 40.8GJ/household, and the standard deviation was 27.7GJ. In this investigation, since the response came from many kinds of families and houses, energy consumption varied greatly between each household. There are two peaks, the first peak around 15 GJ/household is mostly from single-person households and the second peak around 30GJ/household is mainly from the households with two or more people.

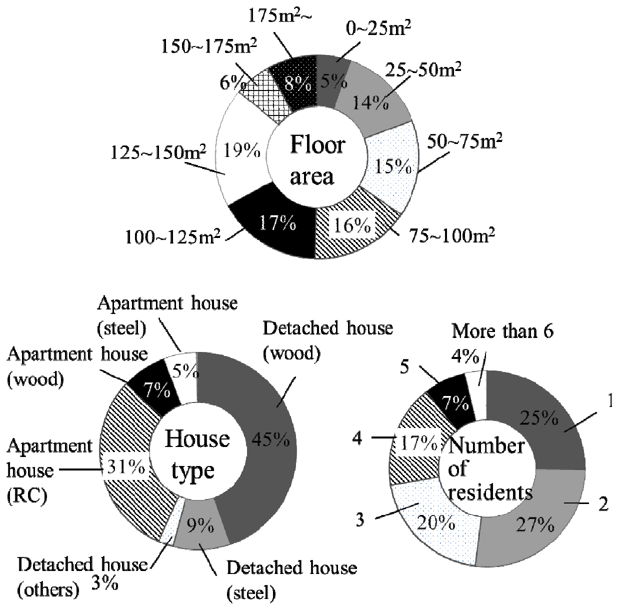


Figure 2-81: Basic information of the households

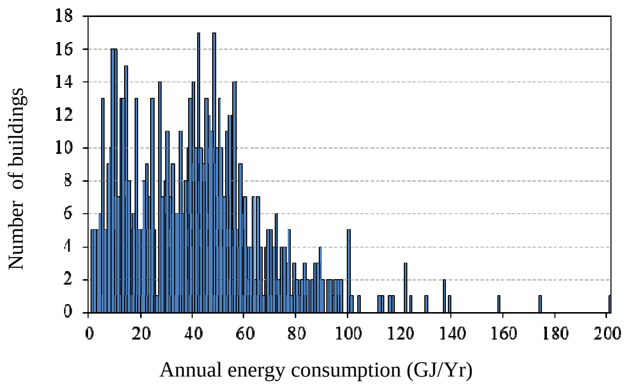


Figure 2-82: Annual total energy consumption of all buildings

Figure 2-83 shows energy consumption according to the house type. Average annual energy consumption of a detached house and an apartment house are 57.0GJ and 27.3GJ, respectively. Normally, detached houses have larger floor area than apartment houses, and there is also greater heat loss. Therefore, energy consumption of a detached house is considered to be bigger than an apartment house. It can be found that kerosene consumption of a detached house is much bigger than that of an apartment house. This also points to the possibility that the ownership ratio of the heating apparatus using kerosene is higher in detached houses.

Figure 2-84 shows energy consumption according to the number of occupants. In the figure, DH means detached house, AH means apartment house, and the number after DH or AH means the number of occupants. Energy consumption increases as the number of family members increases. As for energy consumption per person, energy consumption tends to decrease as the number of family members increases. However, single-person households consume less energy because the occupants go out for a long time, and they tend not to fill the bathtub with hot water, but only use the shower.

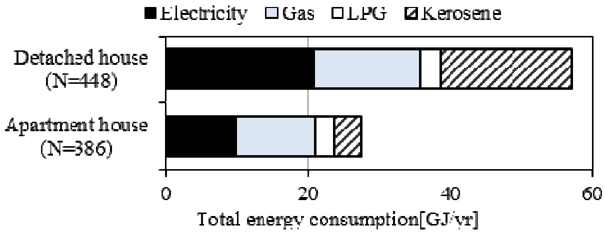


Figure 2-83: Energy consumption by house type

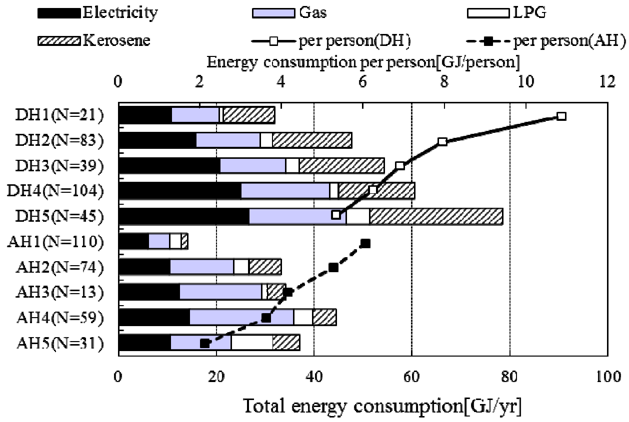


Figure 2-84: Energy consumption by number of occupants

Figure 2-85 shows the boxplot of energy consumption for each end use. The annual average energy consumption for hot water supply was 15.2GJ/household, space heating was 12.5GJ/household, space cooling was 0.3GJ/household and other was 13.8GJ/household. The highest energy consumption came from hot water and space heating required the second most amount of energy. Energy consumption of space heating varies greatly compared to the categories “hot water supply” and “other”. This may be because the airtightness and insulation of the house influence the consumption, in addition to space heating usage condition. The energy consumption for space cooling was much smaller than those of hot water supply and space heating.

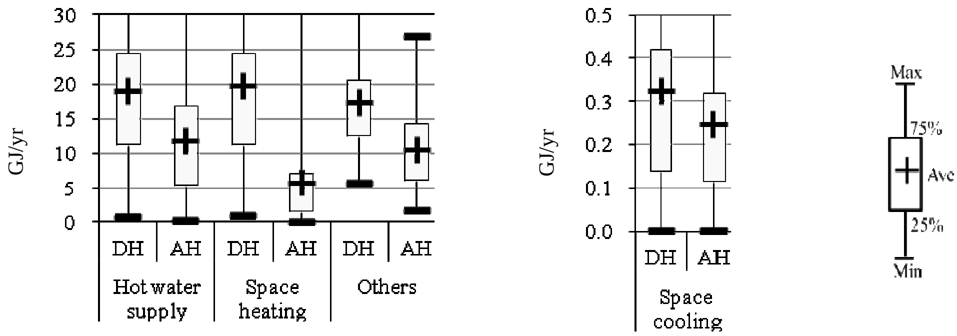


Figure 2-85: Energy consumption for each end use

Figure 2-86 shows the relationship between the energy consumption of space heating and usage condition of air-conditioner. The energy consumption of space heating increases as the utilization frequency of the air-conditioner is higher, and it was low when the air-conditioner is turned off frequently.

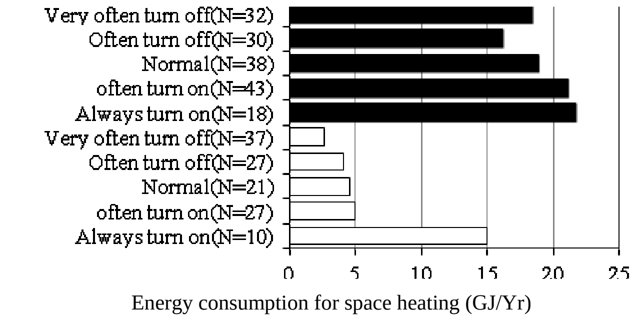


Figure 2-86: Relationship between energy consumption for space heating and usage condition of air-conditioner

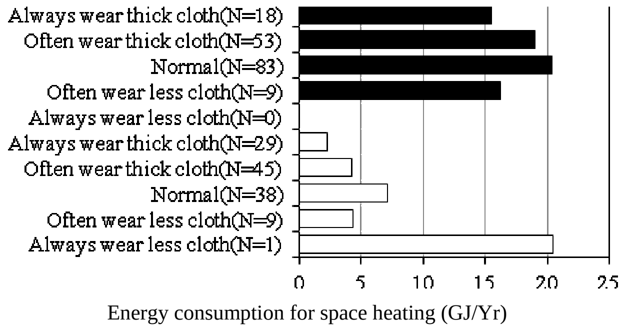


Figure 2-87: Relationship between energy consumption of space heating and clothes

Figure 2-87 shows the relationship between energy consumption and clothes. When occupants always wear many layers, energy consumption of space heating was small, however, when occupants always wear less clothes, energy consumption was higher.

Figure 2-88 shows the relationship between energy consumption of space cooling and the utilization frequency of the air-conditioner. When the number of days that space cooling is used increases, the energy consumption also increases.

Figure 2-89 shows the relationship between the energy consumption of space cooling and usage condition of air-conditioner. The energy consumption of space cooling directly correlates with the utilization frequency of the air-conditioner, and therefore the energy consumption resulting from space cooling was low when the air-conditioner is turned off frequently.

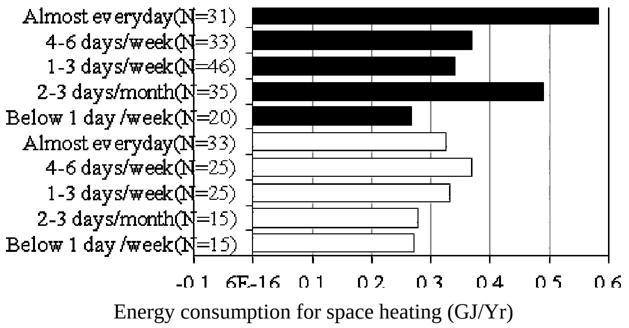


Figure 2-88: Relationship between energy consumption of space cooling and usage frequency of air-conditioner

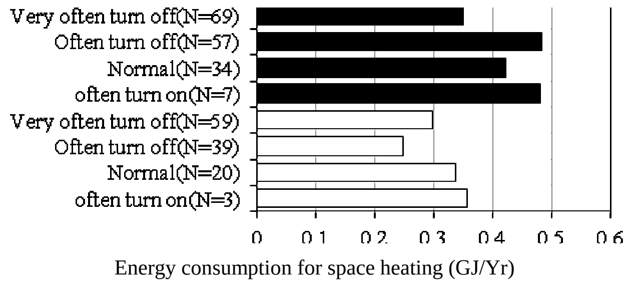
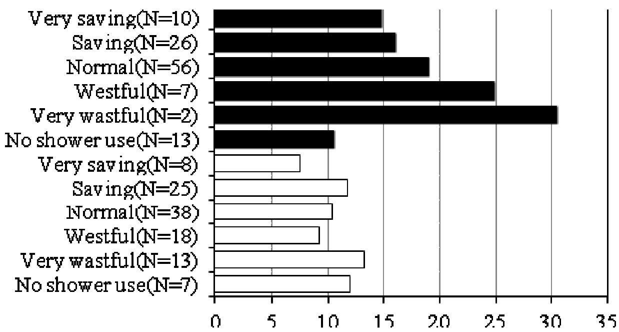
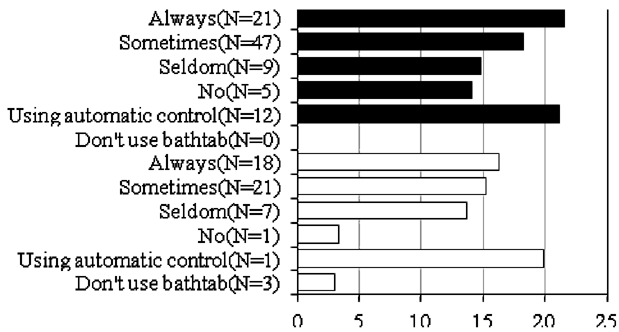


Figure 2-89: Relationship between energy consumption of space cooling and usage condition of air-conditioner





Energy consumption for space heating (GJ/Yr)

Figure 2-90: Relationship between energy consumption of hot water supply and energy saving action taking shower

Energy consumption for space heating (GJ/Yr)

Figure 2-91: Relationship between energy consumption of hot water supply and additional heating of bath water

Figure 2-90 shows the relationship between the energy consumption for hot water supply and saving water when taking shower. When occupants try to save water, energy consumption of hot water supply tends to decrease, and when occupants use a lot of water, energy consumption increases.

Figure 2-91 shows the relationship between energy consumption for hot water supply and additional heating of bath water. As the frequency of additional heating increases, energy consumption clearly increases. If the occupants do not fill bathtub with hot water, energy consumption of hot water is small. If the occupants use an automatic control to set the temperature of the hot water in the bath, energy consumption of hot water is especially high. The energy consumption of space heating is set as dependent variable.
Figure 2-92 shows the category weight of each item of energy consumption for space heating. The coefficient is 0.436 and the constant is -1.74. The house type and the number of occupants have large impacts on space heating. In detached houses, the perimeter area is much larger than that of apartment houses, so heat loss and energy consumption are both higher. Moreover, the earlier the house was built, the higher the energy consumption. One reasonable explanation could be that the quality of the insulation and airtightness in old houses are low, so energy consumption is higher. As for the usage condition of space heating, the household that answered, "Often keep on" and "Always keep on" consumed high amounts of energy. Furthermore, if the occupants wear many layers, energy consumption is lower than others. From these results, energy consumption can be reduced by certain occupant behaviors.

Energy consumption of space cooling is set as a dependent variable. Here, since a question is asked about air-conditioner (except for an electric fan) usage, the household that do not use an air-conditioner are removed from analysis. Figure 2-93 shows the regression coefficient. The coefficient of determination is 0.120 and the constant is 0.093. The utilization frequency and the number of occupants have a large impact on energy consumption. In addition, the energy consumption of space cooling increases as number of occupants increases. Moreover, energy consumption increases as the number of days an air-conditioner is used increases. The year built and turning off air-conditioner frequently do not influence energy consumption. The energy consumption of hot water supply is set as a dependent variable.

Figure 2-94 shows the regression coefficient. The coefficient is 0.396 and the constant is -1.9. The number of occupants has a high influence on energy consumed by the hot water supply. Energy consumption of hot water supply clearly increases as number of occupants increases. The second largest factor is trying to save water when taking a shower. If the occupants try to save water, energy consumption decreases. The third largest factor is reheating bath water. Energy consumption increases as the frequency of reheating bath water increases. Moreover, if occupants control the water temperature using an automatic control, energy consumption is the largest. When the occupants use an automatic control, the bath water reheats frequently without occupants being aware. When the occupants intentionally try to save water, they can reduce their energy consumption.

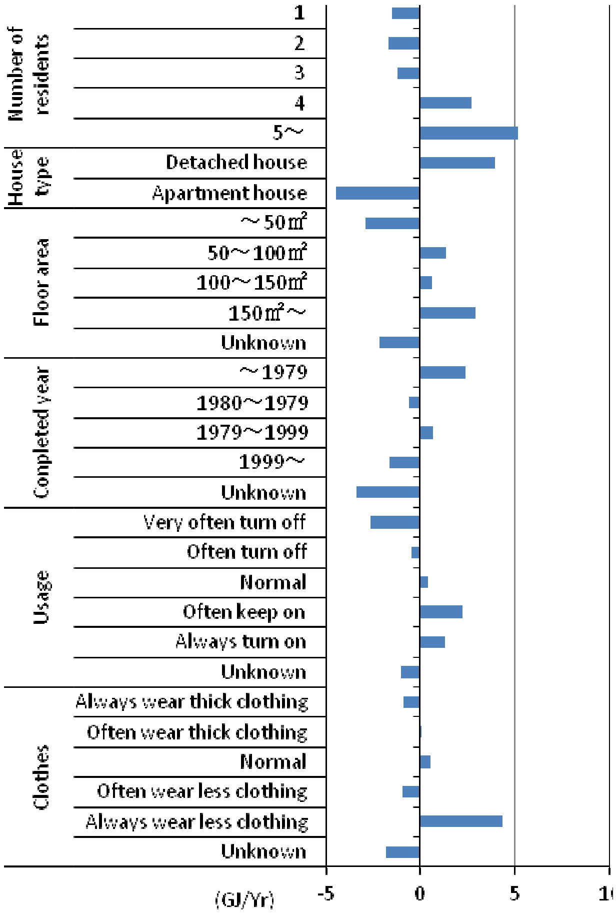


Figure 2-92: Category weight of space heating

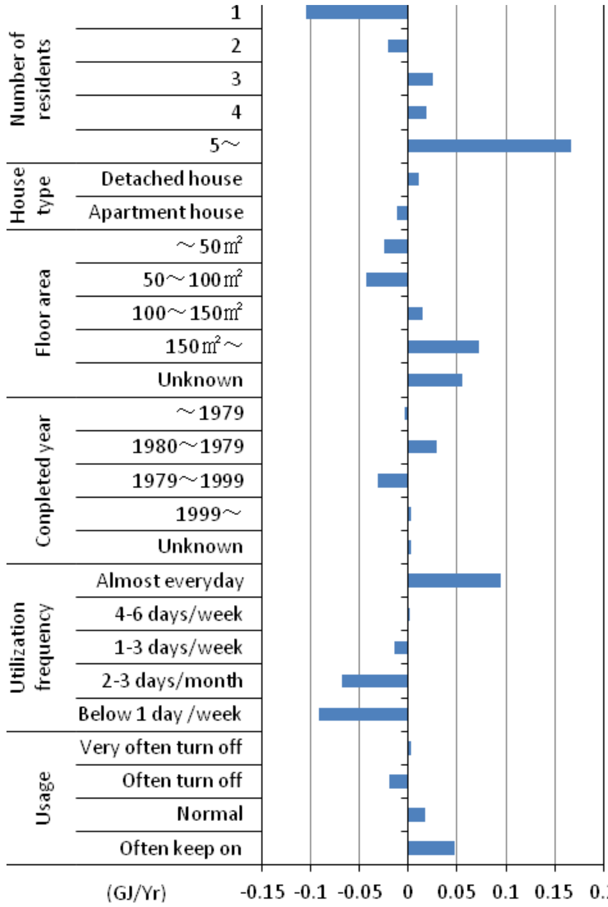


Figure 2-93: Category weight of space cooling

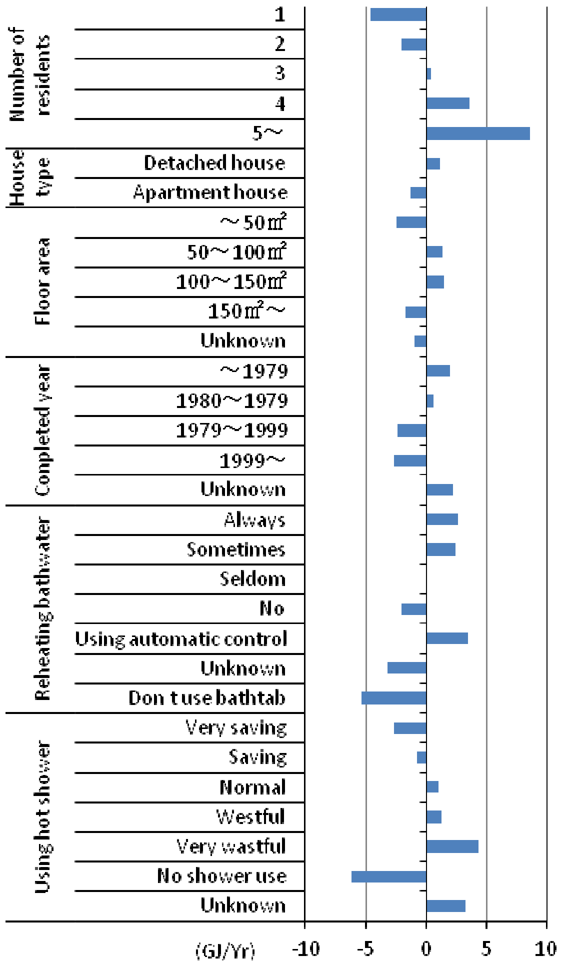


Figure 2-94: Category weight of hot water supply

2.3.5 Conclusions

The survey was conducted to clarify actual energy consumption from October 2008- March 2009 in Sendai. A total of 1274 households responded to the survey. The results are as follows:

The average annual energy consumption in Sendai was about 41GJ/household. The average energy consumption of the detached houses was larger than that of the apartment houses.

Energy consumption increased as the number of occupants increased. As for energy consumption per person, energy consumption tends to decrease as the number of family members increases.

Energy consumption of hot water supply was the largest end use. The annual average was 16.8 GJ/household, followed by energy consumption of other and space heating. The energy consumption of space cooling was much smaller than other end uses.

There was a clear difference in levels of energy consumption between the households that tried to save energy and ones that did not tried to save.

The actual largest end use of energy consume is hot water supply, but most households think space heating is the biggest. Few occupants responded that they could reduce the energy consumed by the hot water supply. This may be because they do not know that the energy consumption of hot water supply is large.

Most people practiced energy saving actions. On the other hand, the energy saving actions related to the bath were not carried out very often.

From the results of multiple regression analysis, the number of occupants and year built have a large impact on the energy consumption of space heating. For the energy consumption of space cooling, the utilization frequency and the number of occupants have a large impact. For the energy consumption of hot water supply, the number of occupants and utilization behavior related to the shower and reheating bath water have a large impact.

According to the analysis, it was found that occupant behavior had a large impact on energy consumption. There is potential to reduce energy consumption in the residential building sector by changing occupant behavior. It is important for occupants to understand how much energy they use for each end use and find the best way to reduce their energy consumption.

2.3.6 **Reference**

[1] Hiroshi Yoshino, Sawako Nakamura, Sayuri Nishiya: Statistical Analysis on Relationships Between Energy Consumption and Energhy Saving Consciousness in Residential Buildings in Sendai City, Proceedings of IBPC5 conference, pp. 907-912, 2012.

2.4 Experience 4: Statistical analysis on energy consumption of residential buildings in China

(Tianchi Hu, Hiroshi Yoshino)

2.4.1 Introduction and the aim of the analysis

The energy consumption of residential sector has been increasing significantly over past three decades. Therefore, it is necessary to analyze how much energy is used by various sources and the factors influencing energy so as to reduce energy consumption in residential buildings. In this study, a large-scale questionnaire survey has been conducted to clarity actual condition of yearly energy consumption in the urban areas of Harbin, Urumqi, Dalian, Beijing, Maanshan, Shanghai, Chongqing, Changsha, Guangzhou and Kunming. The investigated city of Beijing was divided into Beijing(A) where households use district space heating system, and Beijing(B) where households use domestic space heating.

Quantification Theory I is used based on the results of measurement and questionnaire, so as to find out the important factors that influence the energy consumptions in ten cities.

2.4.2 Outline of the survey

Location of investigated cities

The survey was conducted in the urban areas of Harbin, Urumqi, Dalian, Beijing, Maanshan, Shanghai, Chongqing, Changsha, Guangzhou during 2007/10-2008/9, and Kunming during 2008/10-2009/9. Figure 2-95 shows the location of these investigated cities. The investigated cities are all major cities in China and are distributed in the five zones [1].

Figure 2-95: Location of investigated cities.

2.4.3 Investigation method

This study was done by using questionnaire survey. The questionnaire were distributed as well as collected through the cooperative researchers in local universities. Total number of 1004 families living in urban areas of the ten cities was selected by the researchers. Table 22 lists the investigation date, and number of distributed questionnaire, feedback and meter readings. Each family was asked to answer a questionnaire for the summer and winter seasons, including basic information related to the characteristics of their building, heating & cooling periods, daily operation time, usage of heating & cooling appliances, number of occupants, annual income and thermal sensation. In addition, monthly consumptions of electricity & gas of each family in a year was collected by the meter readings. In this investigation, consumption of energy by central district heating is not included in Harbin, Urumqi, Dalian and Beijing. The investigated city of Beijing was divided into Beijing(A) where households use district space heating, and Beijing(B) where households use domestic space heating.

Data processing for deleting and selection

The families are taken as samples, and all the items in the questionnaire are taken as variables. The methods of the processing for the missing data refer to the method in the research by Chen et al. [2]. The numbers of feedbacks of questionnaire and meter readings are the valid sample quantities.

2.4.4 Results of the investigation

Building characteristics

67% in Chongqing, 52% in Changsha and 48% in Guangzhou are more than 120m^2, as shown in Figure 2-96. The average floor area in Chongqing is 130.6 m^2 (the largest among the ten cities). And in Guangzhou, Changsha and Kunming, the average floor area is about 105 m^2. In Dalian, the average floor area of residences is 57 m^2, and 41.5% of residences are below 60 m^2.

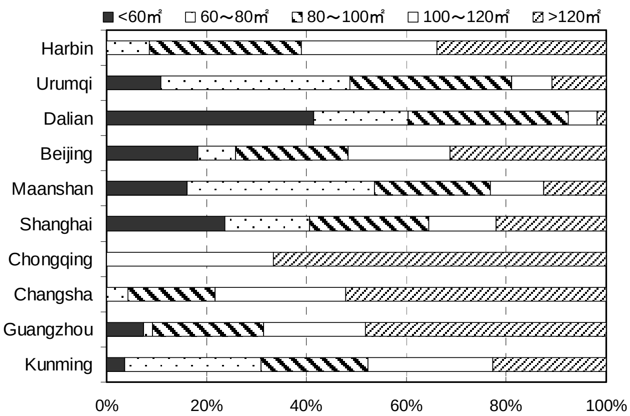


Figure 2-96: Floor area of residences

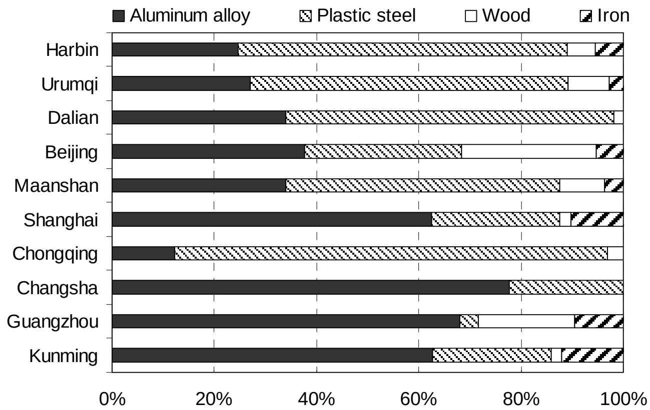


Figure 2-97: Material of window frame

Figure 2-97 shows the material of window frames used in the ten investigated cities. More than 60% of the housing units used aluminum alloy as window frame material in Kunming, Guangzhou, Changsha and Shanghai. Plastic steel window frames are popular in Chongqing, Maanshan, Dalian, Urumqi and Harbin. In Beijing, belongs to the cold zone of China, plastic steel window frame and wooden window frame are normally adopted because of their heat transfer coefficients are low.

Housing appliances

The residences in Harbin, Urumqi, Dalian and Beijing were equipped with central heating systems. In Beijing, besides 41% of households used central heating, 32% used individual heating, and 24% used both of them. Households in Changsha, Chongqing, Shanghai and Maanshan used individual space heating units. The possession rate of heating appliances in Guangzhou and Kunming is very small, especially in Kunming, only 2% of households have space heaters shown as Figure 2-98.

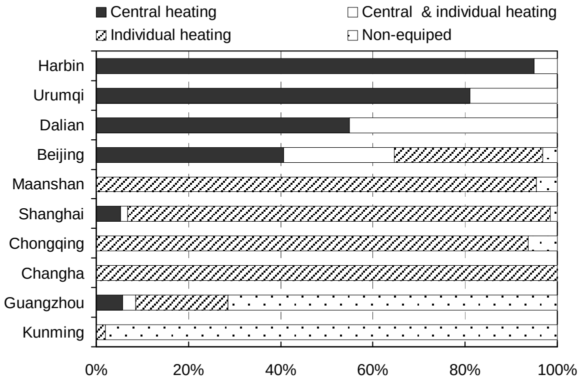


Figure 2- 98: Types of heating appliances

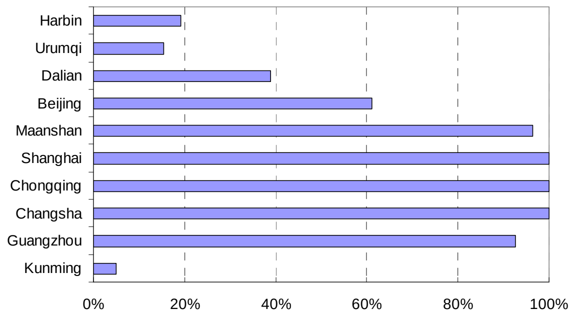


Figure 2- 99: Possession of air-conditioner

Figure 2-99 shows the possession of air conditioner. Each household in Changsha, Chongqring and Shanghai, and more than 95% of households in Maanshan and Guangzhou had installed air-conditioners. However, very few households had air-conditioner in Kunming, Urumqi and Harbin, as the climate is moderate in these cities in summer, especially in Kunming, only 5% of homes had air-conditioners. The possession rate of electric fan of each household in Kunming, Urumqi and Harbin was smaller than that of other seven cities, especially in Kunming, only 14% of households had a fan. The possession rate was highest in Guangzhou, and 50% of the families owned three fans.

Considering the water heater, 9.6%, 6.1% and 29.5% of residences in Harbin, in Beijing and in Changsha were respectively equipped with central hot water supply systems, while residences in the other cities were equipped with individual water heater. Figure 2-100 shows the energy sources of individual water heater. The gas water heater is the most popular type in Guangzhou, Changsha, Chongqing and Shanghai, while the electrical water heater is common in Dalian and Urumqi. It is worth mentioning that many households in Kunming, Maanshan, Beijing and Urumqi were equipped with water heater of solar energy, especially in Kunming with large percentage above 55%.

Family characteristics

Regarding the number of family members, the three people is the main type among investigated families, the size of families in Harbin, Urumqi, Dalian, Beijing and Changsha was commonly 3-person. The average family size in Maanshan, Shanghai and Chongqing was below 2.7 people, while the average Guangzhou family was 3.4 people.

Figure 2-101 shows the annual income of the family. Households in Chongqing had the highest income, there were 17% of the households having annual income above 150,000 RMB, and 25% of the families had the annual income between 50,000 and 100,000 RMB. The annual income of households in Kunming ranked second. The households with annual income above 150,000 RMB, between 100,000 and 150,000 RMB, and between 50,000 and 100,000 RMB, were 9%, 4% and 37% respectively. The annual income between 30,000 and 50,000 RMB is common in Changsha. The families in Urumqi had the lowest annual income, and 82% of the families earned less than 30,000 RMB in a year.

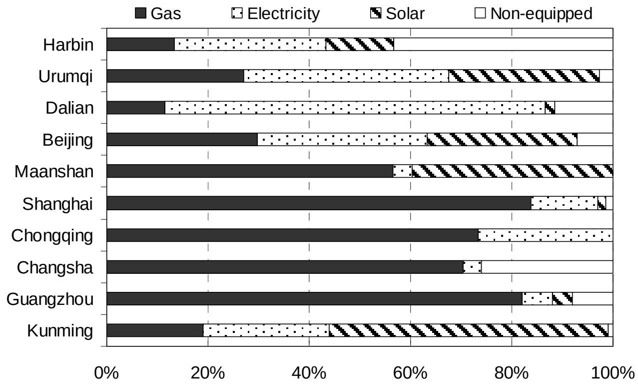


Figure 2-100: Energy sources of individual water heaters

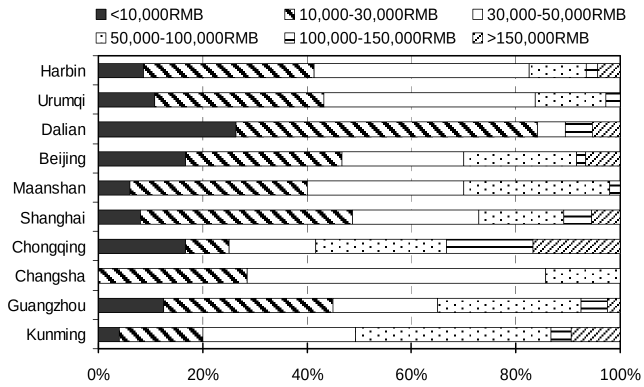


Figure 2-101: Annual income of families

Operation of heating and cooling

Figure 2-102 illuminates the operation period of heating in a year among all the investigated families. The heating period in Harbin, Urumqi, Dalian and Beijing was longer than other investigated cities. The households in Harbin and Urumqi used space heating from October to April of next year, while the households in Dalian and Beijing used space heating from November to March of next year. In Guangzhou and Kunming, the percentage of households using heating was very low with the average of 15% from October to March of next year and 5% from December to March of next year respectively, since these two cities have generally warmer winter than the other cities. Heating period in very cold and cold climate zones is normally longer than that the other zones.

In Harbin, Urumqi, and Dalian, the heating system was operated throughout each day, as shown in Figure 2-103. In the other seven cities, there is one peak in the evening hours from 18:00 to 22:00. The daily using time in very cold and cold climate zones was longer than that the other zones.

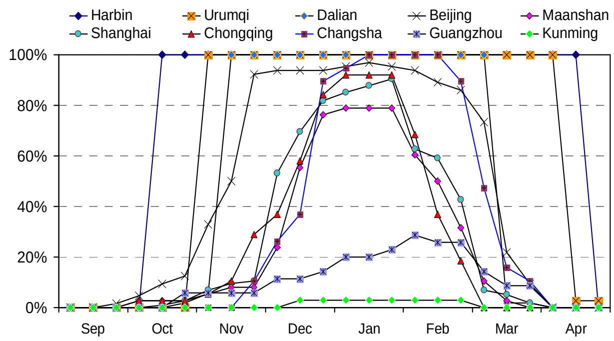


Figure 2-102: Operation period of heating in winter

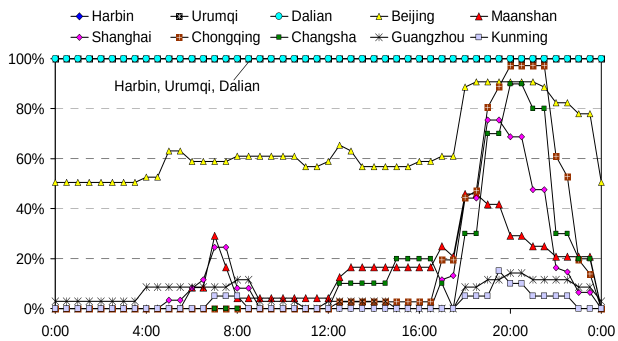


Figure 2-103: Daily operation of heating in winter

Figure 2-104 shows the operation period of air-conditioning in summer among all the investigated cities. Each city used air-conditioning from June to October, except Kunming where less than 30% of households used air-conditioning between July and August.

Regarding the daily operation of the air-conditioning, a very few residences in Harbin, Urumqi and Dalian used air-conditioning due to the good weather of the summer in these cities. The hours of peak air-conditioning usage in Beijing was from 18:00 to 22:00 utilized by 30% of the households, and in

Maanshan the peak was around 20:00 used by 90% of families, as shown in Figure 2-105. The daily using time of air-conditioning in Chongqing was the longest among all ten cities.

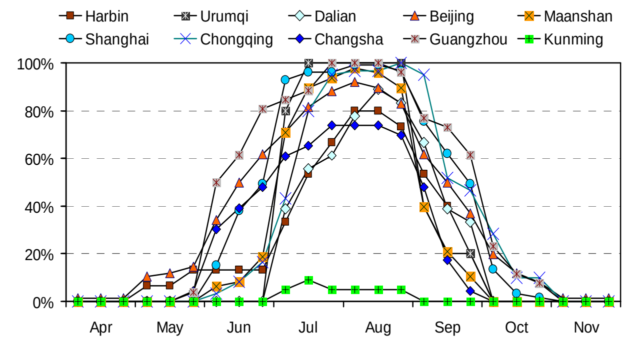


Figure 2-104: Operation period of AC (summer)

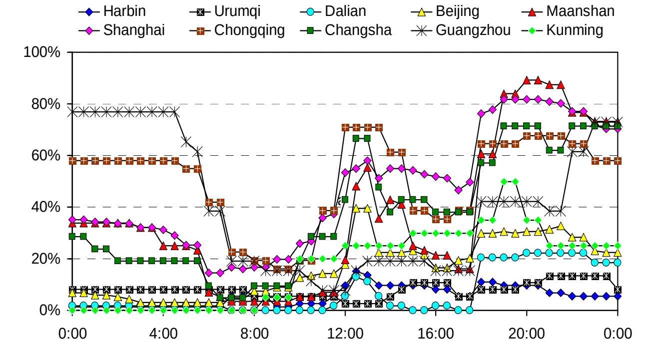


Figure 2-105: Daily operation of AC (summer)

Energy consumption

a) Monthly energy consumption

Electricity and the different kind of gas are converted to calorific values [2-3]. Figure 2-106 shows the annual energy consumptions of ten cities with the total number of respondents shown next to the city name. The annual energy consumption in Chongqing reached 19.2 GJ which was the largest consumer among the ten investigated cities, cooling accounted for 1.4 GJ which was also the highest among all the cities, and cooking accounted for 7.1 GJ. In Beijing(A), households who use district heating consumed 12.5 GJ of energy. On the other hand, Beijing(B) households who use domestic heating consumed 15.6 GJ which was the second largest, heating used 3.1 GJ accounting for 20% of the annual energy use. From the results of this survey, it was found that Guangzhou was the third largest energy consumer with the total annual energy consumption reaching 15.2 GJ, and cooking accounted for 7.2 GJ. It can be seen that energy consumption of cooking in Guangzhou and Chongqing are the first and second largest consumers respectively. This is mostly due to the fact that Chinese people in these two cities enjoy cooking more comparing to other cities.

Table 2-22: End use of energy consumption

Type	Energy source	End use
Type 1	Electricity	Cooling & heating, lighting, others
	Gas	Cooking, hot water
Type 2	Electricity	Cooling & heating, lighting, hot water, others
	Gas	Cooking

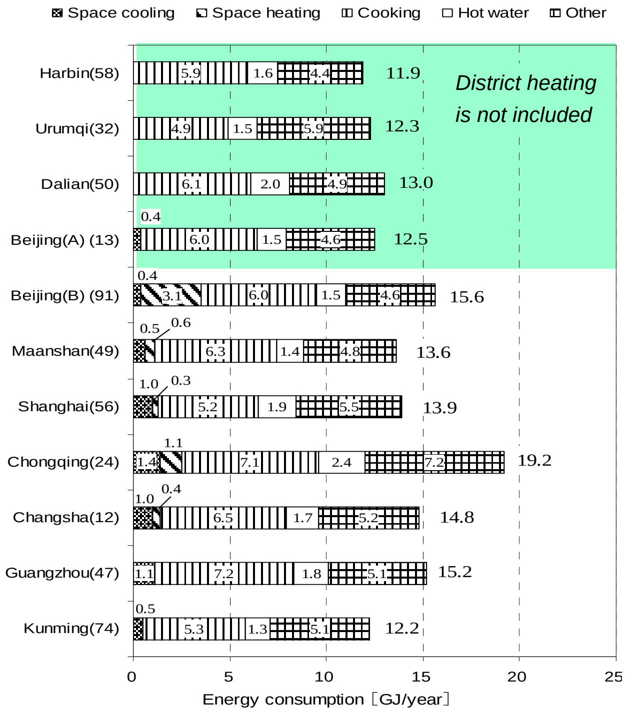


Figure 2-106: Energy consumption in a year

b) Analysis on the influence factors of energy consumption

The partial correlation coefficient of Quantification Theory I is usually used as an important index to evaluate contribution extents of independent variables to the dependent variable. The significance test is taken to judge what extent the partial correlation coefficient would be, then the factors will have effect on residential energy consumption. The significance probability is bigger, the partial correlation coefficient is less, and the factor affects residential energy consumption is less. It is assumed that if the significance probability is less than 0.05, the factor has influence on energy consumption [4]. The category weight value of each variable is used to analyze the influence extent of all the categories of qualitative variables and quantitative variables on the dependent variable. The larger the value is, the more the energy is used.

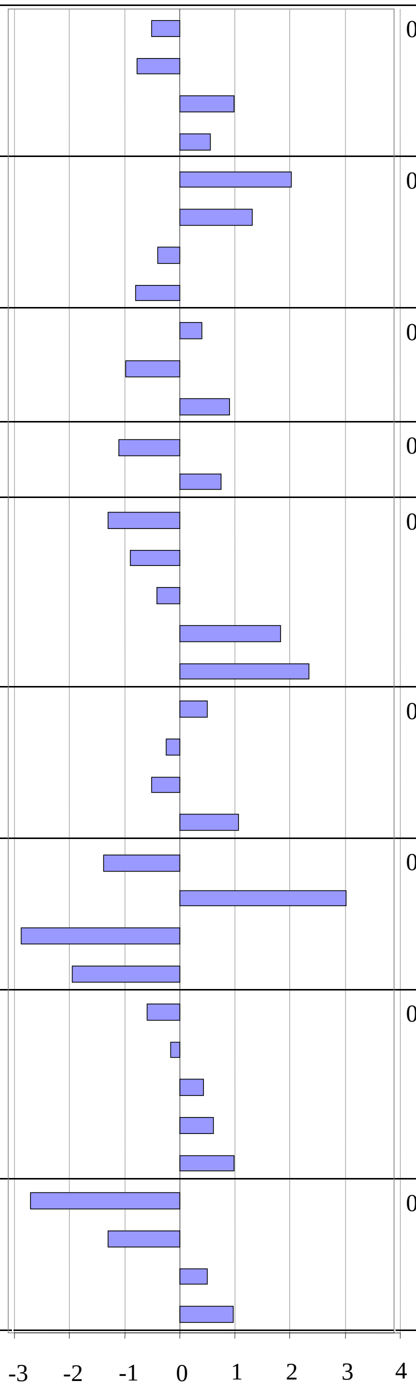
The qualitative and quantitative independent variables used in this analytic model refer to building unit characteristics, family characteristics and housing appliances. The annual energy consumption amount of each family sample is taken as the dependent variable. Software of SPSS (Statistical Program for Social Sciences) is used for calculation. Ten cities are classified into two main groups, Group 1 and Group 2, based on the type of space heating. Since energy consumption data of district heating is not included in this survey, Group 1 includes Harbin, Urumqi, Dalian and Beijing(A) where households use district heating, and Group 2 includes Beijing(B), Maanshan, Shanghai, Chongqing, Changsha, Guangzhou and Kunming where households use domestic heating.

c) Analysis on the influence factors of energy consumption in Group 1

Table 2-23 shows the results of the influence factors on annual energy consumption of Group 1. It can be seen that the type of water heater and the number of family members are the two important influence factors on annual energy consumption. The water heater type is the most important factor influencing annual energy consumption based on its values of the partial correlation coefficient and the significance probability. The value of category weight can be used to judge the influence extent of categories of qualitative variables, and larger value indicates that this category leads to larger energy use. Thus, water heater of electricity has the largest category weight value, indicated that the households using electricity-driven water heaters in these cities consumed the most energy. The households using solar energy, on the other hand, consumed the least energy. Regarding the construction year, the older buildings consumed larger energy since new buildings has better energy saving performance. Considering window frame, wood and plastic steel window saved more energy than the other types. The buildings using wood as window frame consumed the lesser energy. As for the quantitative variables, the larger the floor area is, the larger the number of family members is, the more annual income is, the more HDD (Heating degree-day, indoor temperature set as 18 ºC) is, the more the energy is used.

Table 2-23: Influence factors on annual energy consumption of Group 1.

Influence factors	Categories	Sample	Categories weight	Partial correlation coefficient	Significance probability
Location	Harbin	58		0.148	0.330
	Urumqi	36			
	Dalian	53			
	Beijing(A)	13			
Construction year	Before 1980	16		0.296	0.007
	80s	28			
	90s	67			
	After 2000	49			
Building structure	Brick concrete	100		0.202	0.040
	RC	49			
	Other	11			
HDD (ºC·day)	≤3000	64		0.186	0.275
	>3000	96			
Floor area	< 60 m^2	28		0.217	0.026
	60 - 80 m^2	38			
	80 - 100 m^2	50			
	100 - 120 m^2	23			
	> 120 m^2	21			
Window frame	Aluminum alloy	38		0.084	0.776
	Plastic steel	89			
	Wood	18			
	Iron	15			
Energy source of water heater	Gas	31		0.429	0.001
	Electricity	74			
	Solar energy	22			
	Other	33			
Annual income (Yuan)	< 10000	13		0.102	0.807
	10000 - 30000	14			
	30000 - 50000	37			
	50000 - 100000	61			
	> 100000	35			
Number of family members	1	11		0.324	0.002
	2	35			
	3	86			
	≥ 4	28			



d) Analysis on the influence factors of energy consumption in Group 2

Table 2-24: Influence factors on annual energy consumption of Group 2.

Influence factors	Categories	Sample	Categories weight	Partial correlation coefficient	Significance probability
Location	Beijing(B)	90		0.404	0.001
	Maanshan	56			
	Shanghai	123			
	Chongqing	56			
	Changsha	26			
	Guangzhou	50			
	Kunming	74			
Construction year	Before 1970	17		0.224	0.002
	70s	25			
	80s	86			
	90s	180			
	After 2000	167			
Building structure	Brick concrete	220		0.217	0.003
	RC	221			
	Other	34			
HDD (ºC·day)	<1000	124		0.209	0.004
	1000 - 2000	235			
	> 2000	116			
CDD (ºC·day)	< 100	220		0.223	0.002
	100 - 200	205			
	> 200	50			
Floor area	< 60 m^2	63		0.226	0.002
	60 - 80 m^2	103			
	80 - 100 m^2	96			
	100 - 120 m^2	83			
	> 120 m^2	130			
Window frame	Aluminium alloy	231		0.203	0.004
	Plastic steel	163			
	Wood	46			
	Iron	36			
Energy source of water heater	Gas	266		0.215	0.003
	Electricity	83			
	Solar energy	87			
	Other	39			
Annual income (Yuan)	< 10000	45		0.201	0.004
	10000 - 30000	119			
	30000 - 50000	145			
	50000 - 100000	121			
	> 100000	45			
Number of family member	1	47		0.192	0.005
	2	103			
	3	224			
	4	73			
	≥ 5	28			

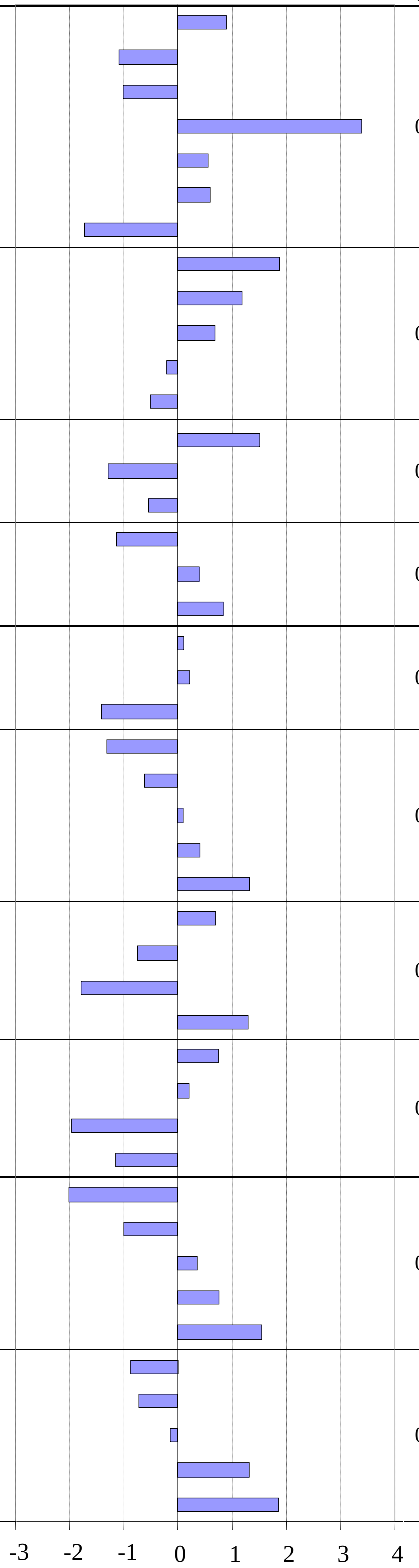


Table 2-24 shows the results of the influence factors on annual energy consumption of Group 2. Annual energy consumption results from the integrated influence factors in different aspects. City location, construction year, building structure, floor area and CDD (Cooling degree-day, indoor temperature set as 26 ºC) are important factors influencing the annual energy consumption; city location is the most important influence factor among these factors. The results showed that Chongqing had the largest category weight value, while Beijing (B) Guangzhou and Changsha ranked after Chongqing, while Kunming had the value at the smallest in seven in cities. This change trend of value in these cities is also consistent with the magnitudes of energy use amounts in these cities. Considering the building structure, the buildings constructed by reinforced concrete are distinctly less energy efficiency than the other structures. Regarding the window frame, the buildings used wood as window frame consumed the least energy since the material has the smallest heat transfer coefficient. Solar energy water heater helps to save energy, while the families used gas water heaters consumed the most energy. Based on the saturation of water heaters in these cities, the prevalence of solar water heaters in Kunming and Maanshan can lead to the low energy use in some extent in the two cities. Considering CDD, category of 100-200 used more energy than the others.
For the other quantitative variables, the larger the floor area is, the larger the number of family members is, the more annual income is, the more HDD is, the more the energy is used.

2.4.5 **Conclusions**

This study investigated the urban residential energy consumption under actual conditions by the way of questionnaire survey on Chinese families that reside at various climate zones. The following are the key findings of this study:
Most of the investigated buildings are built by brick concrete except in Guangzhou and Kunming where almost 70% of the buildings use reinforced concrete as building structure. As for floor area, 67% of the residences in Chongqing are more than 120 m^2, which is the largest.
Family size with 3-person is common. However, the average family size in Guangzhou is 3.4 people. Households in Chongqing has the highest income, 17% of the households have the annual income above 150,000 RMB, and 25% of the families have the annual income between 50,000 and 100,000 RMB.
Almost every household in Harbin, Urumqi, Dalian and Beijing equipped with central heating system. Regarding the operation of heating, the heating periods in Harbin, Urumqi, Dalian and Beijing are longer than the other cities, all residences in Harbin, Urumqi and Dalian use heating in every day. On the other hand almost all households in Maanshan, Shanghai, Chongqing, Changsha and Guangzhou have cooling equipment such as air-conditioners and fans.
Annual energy consumption in Chongqing reaches 19.2GJ which is the largest among the ten investigated cities, and cooking accounts for 7.1 GJ. In Beijing (B) where households use domestic heating consumes 15.6 GJ which it is the second largest, the heating use 3.1 GJ accounting for 20% in the total.
In Group 1, the water heater type is the most important influence factor, and the number of family members is the second most important influence factors on annual energy consumption. In Group 2, city location, construction year, building structure, floor area and CDD are important factors influencing annual energy consumption. The city location is the most important factor influencing annual energy consumption.

2.4.6 References

[1] Department of Construction of the PRC, Thermal design code for civil buildings GB50176-93, China Plan Press, 1993.

[2] G. Yuan, Designing Handbook of Gas in Buildings, China Architecture and Building press, Beijing, 2001.

[3] Y. Xiang, Common-use Data Handbook in Gas Thermodynamic Engineering, China Architecture and Building Press, Beijing, 2000.

[4] W. Zhang, Statistical Application, Advanced Education Press, Beijing, 2004.

2.5 Experience 5: Field Survey and Statistical Analyses on Energy Consumptions in the Residential Buildings in Japan

(Hiroshi YOSHINO, Graduate School of Engineering, Tohoku University, Japan)
(Ting SHI, Building Research Center, Vanke, China)

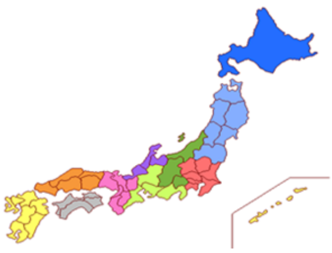
2.5.1 Introduction

Field surveys of energy consumptions have been carried out in eighty residential buildings which are located in six different districts of Japan, with the purpose to understand the energy consumptions by end users in residential buildings and then to build up a national-level database. In addition, statistical analyses were conducted so as to find out the influential factors on residential energy consumptions in Japan. Finally, the database and the statistical analyses results will be of help as evidences for the house makers and equipment manufacturers to do environment-friendly developments, as well as for the residents to select environment-friendly household appliances to install into their houses.

2.5.2 Outline of the investigation

Selection of the households

The two-year survey on energy consumptions was conducted in eighty households located in six different districts of Japan (Figure 2-107) from December 2002 to November 2004. The six districts are Hokkaido, Tohoku, Hokuriku, Kanto, Kansai and Kyusyu (including Okinawa). Fifteen households were selected in Kanto District, and thirteen households were selected in each of the rest five districts. Two types of residential buildings, namely detached house and apartment, respectively, were under investigated. In each district, nine households live in detached houses and the others live in apartments. Households which met the following essential conditions and/or optional conditions were selected in the survey.



	District	Number
	Hokkaido	13
	Tohoku	13
	Hokuriku	13
	Kanto	15
	Kansai	13
	Kyushu・Okinawa	13
	Total	**80**

Figure 2-107: Location

Essential conditions

Among the nine detached houses in each district, at least four of them are wood constructed with floor area around 100~150m^2 and meet the local new energy saving standard.

Three or four family members: husband, wife and one to two children.

Optional conditions

The houses which can represent the district characteristics (e.g., well insulated and air-tight houses in Hokkaido) are selected.

Give priorities to the households which can cooperate in the long-term investigation.

Investigation items and methods

Table 2-25 shows the investigation items and methods. Special instruments as shown in Figure 2-108 (a, b, c) were used to measure the consumptions of electricity, gas and kerosene, while data loggers with temperature sensors as shown in Figure 2-108 (d) were used to measure indoor air temperatures. The electricity measuring systems record electricity consumption in one minute (Wh) and the peak value (W). The kerosene measuring system has a flow meter inside, which record the kerosene volume (l) in five minutes by the pulse logger. The temperature and humidity data loggers, which were set in the air-conditioned and non-air-conditioned rooms at the height of 1.1m above floor, record the data every 15 minutes. In addition, questionnaire and hearing surveys with the cooperation of the occupants have been carried out to know their lifestyles.

Table 2-25: Investigation items and methods

Item		Method
Field Measurements	Electricity	Measured by instrument every one minute
	Gas	Measured by instrument every five minute
	Kerosene	Measured by instrument every five minute
	Temperature	Measured by data loggers with temperature sensors every 15 minutes
Questionnaire Survey		Life style, Concious on environment, Annual income, utilization of equipment…
Hearing		Basic information of the building, Family members…

Figure 2-108: Measuring instruments (a, b, c and d from left to right: electricity, gas, kerosene and air temperature)

Energy indicator

Basically, energy indicator in the database is final (secondary) energy consumption. Table 2-26 shows the energy conversion coefficients for different kinds of energy sources [12]. Energy end users were classified as shown in Table 2-27. The classification has three levels. There are ten categories in the first level, which are 1) total energy consumption, 2) heating, cooling and ventilating (HVAC), 3) domestic how water (DHW), 4) lighting, 5) kitchen, 6) refrigerator, 7) entertainment and information, 8) housework and sanitary, 9)others, and 10) generation (e.g. photovoltaic power generation). The ten categories were classified into more detailed, which is the second level. For example, hot water supply is divided into hot water supply for bathroom, for kitchen and others. Finally, the third level is the energy consumption for individual household electrical appliance. In addition, rated electricity consumption and stand-by power of the electrical appliances were divided if possible.

Table 2-26: Conversion coefficients of energy sources

Energy Source	Conversion Coefficient
Electricity	3.6 MJ/kWh
Kerosene	36.7 MJ/L
LPG	50.2 MJ/Nm^3
City Gas (4A~7C)	20.4 MJ/Nm^3
City Gas (12A~13A)	45.9 MJ/Nm^3

Table 2-27: Energy end users

First level	Second level	Third level	Energy source
1) Total energy consumption	Household	Purchased power	○Secondary energy
		Power selling	○Secondary energy
		City gas	○Secondary energy
		Kerosene	○Secondary energy
		Wood, etc	
2) Heating, cooling and ventilating	Cooling (summer)	Air-conditioner	○Electricity
		Gas air-conditioner	Gas
		Electrical fan	Electricity
		Dehumidifier	Electricity
		Others (heat exchanger, etc)	Electricity
	Heating (winter)	Air-conditioner	○Electricity
		Gas air-conditioner	Gas
		Kerosene air-conditioner	kerosene
		Electrical floor heating	○Electricity
		Gas floor heating	Gas
		Kerosene floor heating	kerosene
		Kotatsu (Japanese electrical heating table)	Electricity
		Electrical carpet	Electricity
		Gas fan heater	Gas
		Kerosene fan heater	kerosene
		Kerosene stove	kerosene
		Electrical heater (panel heater, etc)	Electricity
		Heat storage heater	Electricity
		Humidifier	Electricity
		Electrical blanket	Electricity
		Others	Electricity
	Ventilation (exclude range hood)	24hours ventilation system	○Electricity
		Local ventilation	Electricity
		Air cleaner	Electricity
3) Hot water supply	Hot water supply (bathroom)	Electrical water heater	○Electricity
		Gas water heater	Gas
		Kerosene water heater	kerosene
		Others	Electricity
	Hot water supply (kitchen)	Electrical water heater	Electricity
		Gas water heater	Gas
		Kerosene water heater	kerosene
		Others	Electricity
	Hot water supply (others)	Electrical water heater	Electricity
		Gas water heater	Gas
		Kerosene water heater	kerosene
		Others	Electricity
4) Lighting	Lighting	Lighting	Electricity
		Table lamp	Electricity
5) Kithchen	Cooking	Electrical cooker (IH and 200V equipment)	○Electricity
		Microwave	Electricity
		Electrical oven	Electricity
		Gas oven	Gas
		Rice cooker	Electricity
		Gas rice cooker	Gas
		Pot	Electricity
		Table stove/plate	Electricity
		Toaster	Electricity
		Coffee maker	Electricity
		Juicer/blender	Electricity
		Home bakery	Electricity
		Gas cooker	Gas
		Refrigerator	○Electricity
	Others	Range hood	Electricity
		Dishwasher	Electricity
		Gas dishwasher	Gas
		Water filter	Electricity
		Rice mill	Electricity
6) Entertainment and information	Entertainment	Television	Electricity
		Vedio	Electricity
		DVD player	Electricity
		Audio/radio casette	Electricity
		Game	Electricity
		BS/CS turner	Electricity
		Tuner	Electricity
		CATV terminal	Electricity
		CATV booster	Electricity
		Wireless lan (main and extension)	Electricity
		Electrical piano	Electricity
	Information	Computer	Electricity
		Telephone and fax	Electricity
		Telephone extension	Electricity
		Entry phone	Electricity
		TV door phone	Electricity
		Telephone security unit	Electricity
		Telephone battery charger	Electricity
		Cellphone battery charger	Electricity
		Shredder	Electricity
	Security	Home security	Electricity
7) Housework and Sanitary	Housework	Washing machine	Electricity
		Gas washing machine	Gas
		Cloth dryer	Electricity
		Gas cloth dryer	Electricity
		Iron	Electricity
		Vacuum	Electricity
		Sewing machine	Electricity
		Futon dryer	Electricity
		Trousers press	Electricity
	Sanitary	Warm-water cleaning toilet seat	Electricity
		Dryer	Electricity
		Bathroom heating (dryer)	Electricity
		Gas bathroom heating (dryer)	Gas
		Electrial shaver	Electricity
		Electrical tooth brush	Electricity
		Inhaler	Electricity
		Electrical mosquito swatter	Electricity
		Electrical spetic tank	Electricity
	Medical care	Medical machine	Electricity
8) Others	Others	Electrical shutter	Electricity
		Tank	Electricity
		Unclear iterms	Electricity
9) Generation	Generation	Photovaltaic power generation	Electricity
		Solar water heater	Electricity

the items with "○" are necessary items

2.5.3 Investigation results

Annual energy consumption

Figure 2-109 and Figure 2-110 show the annual energy consumptions for the 80 households during one year from December 2002 to November 2003 and from December 2003 to November 2004, respectively. The blanks in the figures indicate that measurements were not carried out in such households. The results show that energy consumption varied widely in different households. The maximum value was about twice of the minimum value even in the same district. In general, annual energy consumption decreased when the household goes southward. In Hokkaido, Tohoku and Hokuriku, where have cold winters, annual energy consumptions were larger than other districts. Besides, energy consumptions for HVAC and DHW accounted for a large ratio (about 80% of the total energy consumption) in these districts. In addition, energy consumption for HVAC and DHW also varied widely in different households. The maximum value was about three or four times of the minimum value. The differences of energy consumptions for HVAC and DHW are considered to result in the differences of total energy consumption in residential buildings.

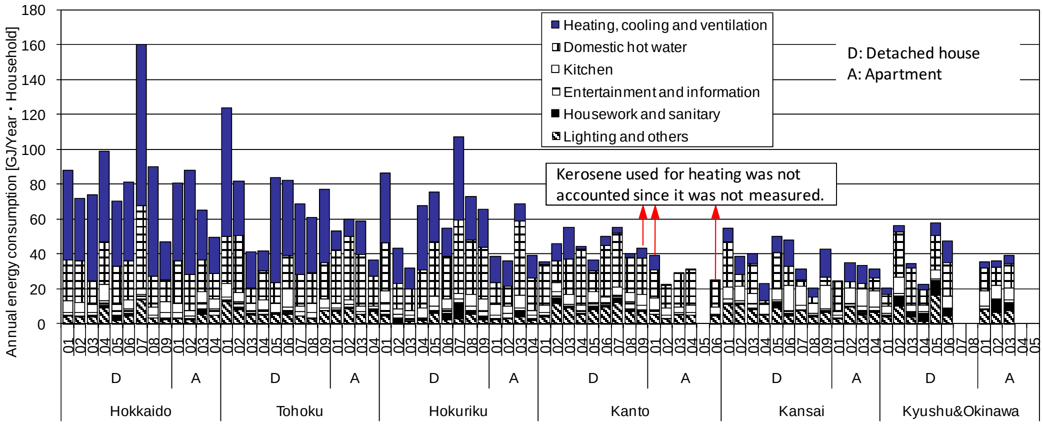


Figure 2-109: Annual energy consumptions for the 80 households (Dec. 2002 ~ Nov. 2003)

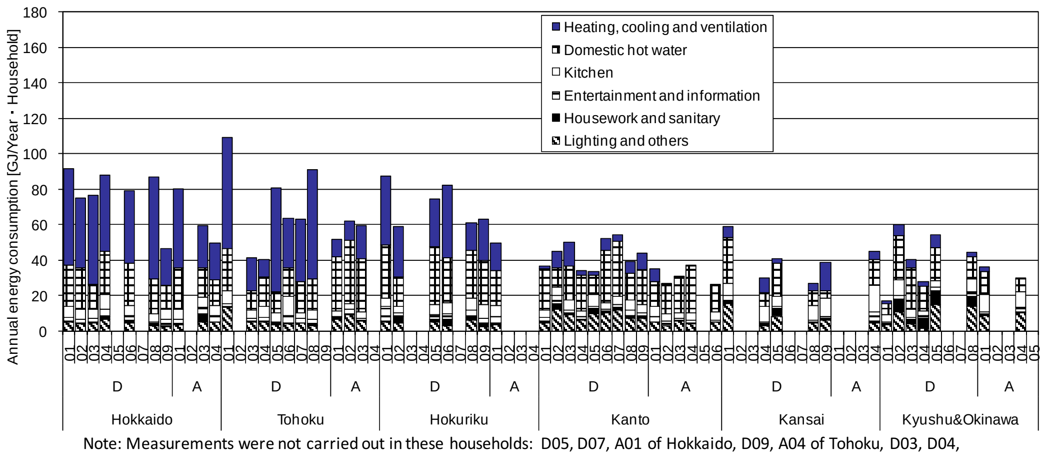


Figure 2-110: Annual energy consumptions for the 80 households (Dec. 2003 ~ Nov. 2004)

Comparison between the first year and the second year

Figure 2-111 shows the comparison of the annual energy consumptions (Left: Energy consumption for HVAC; Right: Total Energy Consumption) between the first year (Dec. 2002 ~ Nov. 2003) and the second year (Dec. 2003 ~ Nov. 2004). The results show that, although energy consumptions in these households had some differences between the first year and the second year, the differences were not obvious. This indicates that residents have a stable lifestyle and they do not like to change much. On the other hand, households in Tohoku district cooperated to live an energy-saving lifestyle (e.g., shortening heating/cooling period, appropriating setting temperature, reducing heating/cooling space, unplugging electrical appliances when not using, etc.) in the second year. Annual energy consumption in such households became smaller compared to the first year, and the reductions in two of the households were very obvious. Such results indicate that lifestyle of the residents is an important factor to save energy in residential buildings.

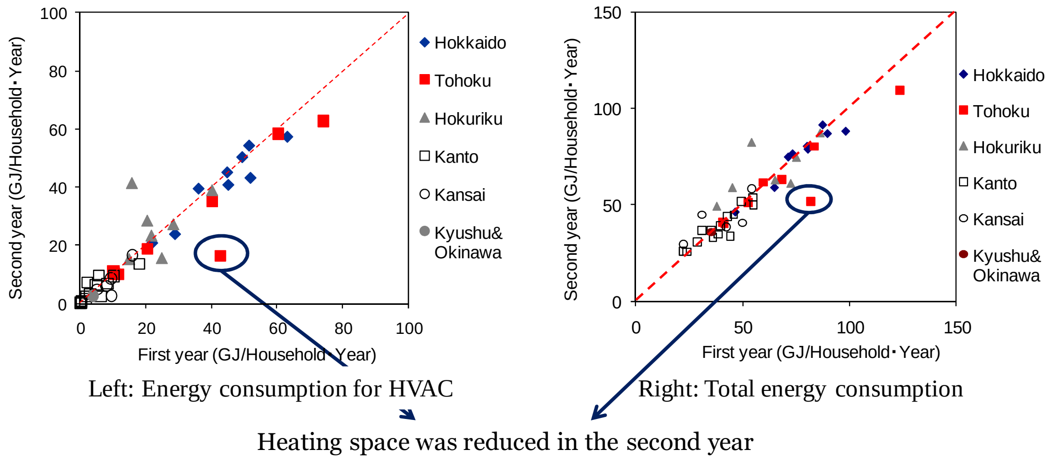


Figure 2-111: Comparison between the first year and the second year (Left: Energy consumption for HVAC; Right: Total Energy Consumption)

2.5.4 Statistical analyses

Introduction

The investigation results showed that energy consumptions varied widely in Japanese residential buildings. In order to find out what factors and how they influence on residential energy consumptions, statistical analyses were carried out with the results of 72 households with full energy consumption information. In the statistical analyses, the dependant variables are the energy consumptions for different end uses, and the explanatory variables are the considerable factors as shown in Table 2-28. When selecting the explanatory variables, the following matters should be well examined so as to ensure the precision of the analyses.

- Do Not select any useless factor
- Do Not miss any useful factor
- Do Not select highly correlated factors so as to avoid the problem of multi co-linearity.

With these considerations, explanatory variables which can be gained from the investigations were used in the analyses, while those factors exceeded the investigation items were not considered.

Three kinds of statistical methods were used in the analyses, which are multiple regression analysis, neural network, and quantification method. The models of multiple regression analysis and neural network are used to predict the relationships between dependent variable and explanatory variables in linear and non-linear way, respectively. On the other hand, quantification method is used to analyze the influences of qualitative factors, which are described by texts, e.g. district, building type, etc.

Table 2-28: Statistical methods

NO.	Factor	Unit	Multiple Regression Analysis	Neural Network	Quantification Mothod Ⅰ
1	District	-	-	-	○
2	Type of the Building	-	-	-	○
3	Age of the Building	[Years]	○	○	○
4	Floor Area	[m²]	○	○	○
5	Coefficient of Heat Loss	[W/m²K]	○	○	○
6	Equivalent Area of Interstice	[cm²/m²]	○	○	○
7	$CDD_{22\text{-}24}$	[℃・Day]	○	○	○
8	$HDD_{14\text{-}14}$	[℃・Day]	○	○	○
9	Living Room Temperature in Summer	[℃]	○	-	○
10	Living Room Temperature in Winter	[℃]	○	-	○
11	Number of Family members	[Person]	○	○	○
12	Number of Electrical Household	[Piece]	○	-	○

2.5.5 **Result**

Multiple regression analyses

The dependent variable is annual energy consumption (GJ/Year), and the explanatory variables are shown in Table 2-29. Floor area had the largest standardized partial regression coefficient, followed by heating degree day and living room temperature in winter, which indicated that floor area, heating degree day and living room temperature in winter had greater effect on annual energy consumption, compared to other factors. Figure 2-112 shows the relationship between predicted values and observed values. The coefficient of determination was 0.72, which indicated that energy consumptions in 72% of the households can be explained by the multiple regression analyses.

Table 2-29: Results of multiple regression analyses

Factor	Partial Regression Coefficient	Standardized Partial Regression Coefficient
Building Age [Year]	-0.99	-0.18
Floor Area [m²]	0.36	0.48
Coefficient of Heat Loss [W/m²・K]	2.71	0.12
Equivalent Area of Interstice [cm²/m²]	0.35	0.03
Family Members [Persopn]	3.57	0.14
$CDD_{22\text{-}24}$ [℃・Day]	-0.03	-0.10
$HDD_{14\text{-}14}$ [℃・Day]	0.01	0.38
Living room Temperature in Summer [℃]	1.15	0.09
Living room Temperature in Winter [℃]	2.49	0.27
Electrical Household Appliances [Piece]	0.08	0.03
Constent	-103.79	

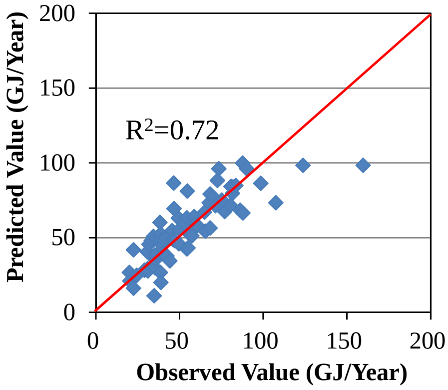


Figure 2-112: Relationship between the predicted values and observed values

Neural network

Figure 2-113 shows the input and output layers in neural network analyses. In neural network, the data will be divided into training data and validating data, where training data is used to analyze how explanatory variables influence dependent variable, while validating data is used to validate the analyses. In this research, one household was randomly selected as validating data from each district and the rest 66 households were used as training data. Table 2-30 shows the comparisons between training data and validating data. Compared to the training data, the validating data had smaller floor area, equivalent area of interstice, but higher living room temperatures and annual energy consumption.

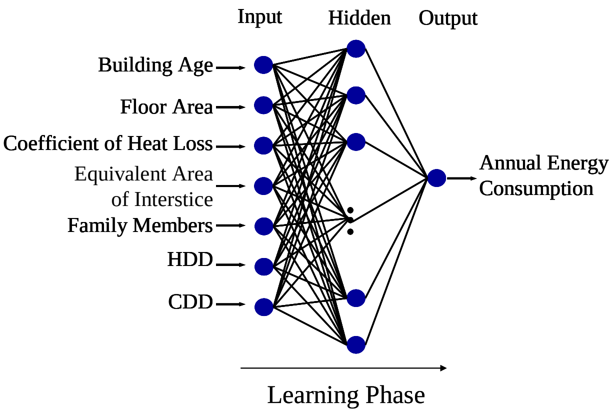
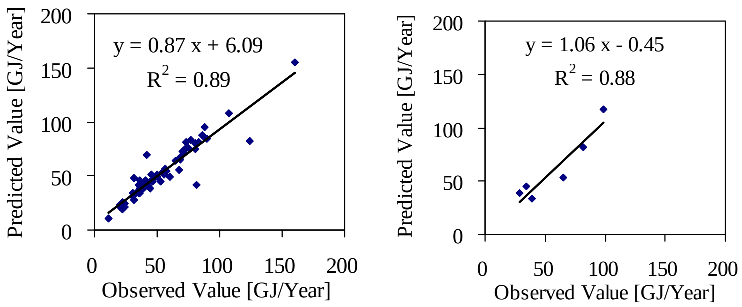


Figure 2-113: Input and output layers in neural network

Table 2-30: Training data and validating data

NO.	Factor	Training Data (66Households)			Validating Data (6Households)		
		MAX	MIN	AVE	MAX	MIN	AVE
No.1	District	—	—	—	—	—	—
No.2	Building Type	—	—	—	—	—	—
No.3	Building Age [Years]	35.0	1.0	5.5	13.0	1.0	5.7
No.4	Floor Area [m^2]	240.0	46.0	122.9	159.0	70.0	108.2
No.5	Coefficient of Heat Loss [W/m^2·K]	7.7	0.6	2.4	3.2	1.3	2.2
No.6	Equivalent Area of Interstice [cm^2/m^2]	13.3	0.2	2.8	2.9	0.3	1.4
No.7	Family Members [Person]	6.0	2.0	3.6	5.0	2.0	3.5
No.8	CDD[D_{22-24}]	625.5	0.0	118.1	245.3	0.0	107.0
No.9	HDD [D_{14-14}]	3220.7	5.4	1525.5	3220.7	623.2	1597.6
No.10	Living room Temperature in Winter [℃]	24.9	11.1	18.5	23.5	17.3	19.6
No.11	Living room Temperature in Summer [℃]	30.9	21.9	27.4	28.7	26.8	28.0
No.12	Electrical Household Appliances [Pieces]	68.0	11.0	33.9	51.0	21.0	34.7
No.13	Annual Energy Consumption [GJ/Year]	159.8	11.2	52.8	123.8	29.0	65.1

The Figure on the left hand side of Figure 2-114 shows the relationship between predicted values and observed values in neural network model with training data, while the right hand side of Figure 2-114 shows precision of the neural network analyses with validating data. The coefficient of determination in the validating model was 0.88, which indicated that the model in fact correct in 88% of the times.



Neural Network Model with Training Data

Validating NNW Model using Validating Data

Figure 2-114: Results of neural network.

Figure 2-115 shows the importance of input layers to annual energy consumptions. It indicated that heating degree day, family members and coefficient of heat loss had more importance to annual energy consumptions in Japanese residential buildings compared to other factors.

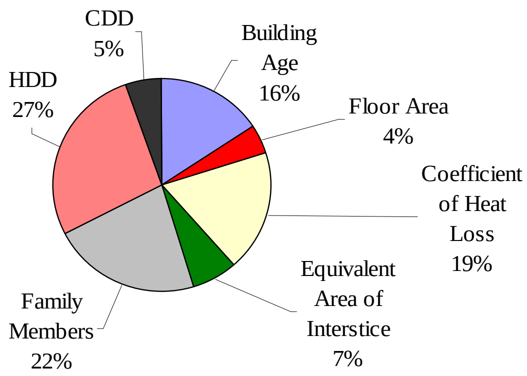


Figure 2-115: Importance of input layers to annual energy consumptions

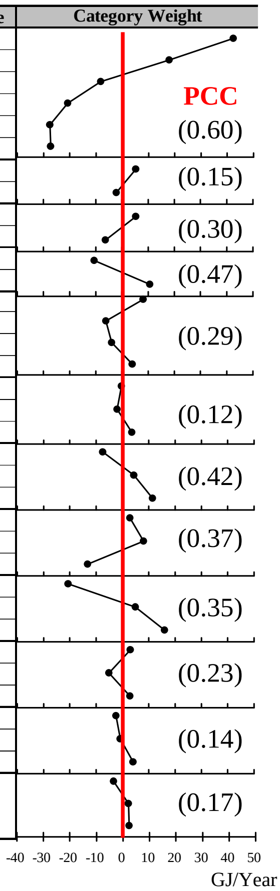
Quantification method I

In the quantification method type I, dummy variables are used and analyzed by multiple regression method. Factors were divided into several categories and the category weights were calculated by minus the average value.

Table 2-31 shows the results when the dependent variable is annual total energy consumption. Numbers on the right hand side were Partial Correlation Coefficients of the factors (PCC). The partial correlation coefficients of district, floor area, and family members were larger than other factors, which indicated that these three factors had greater effect on annual energy consumption compared to others. This table also shows that energy consumption decreased as the household goes southward. The difference between Hokkaido and Kyushu was about 70GJ/Year. Households located in Hokkaido with floor area larger than 120m^2, and have more than four people in their home consumed most energy per year.

Table 2-31: Quantification method I: (Dependant Variable: annual total energy consumption)

Item	Category	Sample	Category Weight (PCC)
District	Hokkaido	13	(0.60)
	Tohoku	13	
	Hokuriku	10	
	Kanto	10	
	Kansai	9	
	Kyushu	9	
Building Type	House	46	(0.15)
	Apartment	18	
Building Age (Year)	~3	35	(0.30)
	3~	29	
Floor Area (m^2)	~120	32	(0.47)
	120~	32	
Coefficient of Heat Loss (W/m^2・K)	~1.5	13	(0.29)
	1.5~2	18	
	2~2.5	16	
	2.5~	17	
Equivalent Area of Interstice (cm^2/m^2)	~1	28	(0.12)
	1~3	21	
	3~	15	
Family Members (Person)	~4	29	(0.42)
	4	26	
	4~	9	
CDD (℃・Day)	~100	42	(0.37)
	100~200	8	
	200~	14	
HDD (℃・Day)	~1000	11	(0.35)
	1000~2000	36	
	2000~	17	
Living room Temperature in Summer (℃)	~27	19	(0.23)
	27~29	23	
	29~	22	
Living room Temperature in Winter (℃)	~17	23	(0.14)
	17~20	22	
	20~	19	
Electrical Household Appliances (Piece)	~25	25	(0.17)
	25~30	17	
	30~	22	



2.5.6 Conclusions and discussions

The national database of a two-year investigation in the residential buildings in Japan shows the following results.

Energy consumptions of the households in Hokkaido district are the largest, followed by Tohoku district and Hokuriku district. These three districts have long and cold winter, heating and hot water supply are the largest energy user there.

Annual energy consumption did not change much between the first year and the second year, due to the residents are willing to have a stable lifestyle. However, in the households where the residents cooperated to live an energy-saving lifestyle in the second year, energy consumptions were smaller than in the first year. Lifestyle is an important factor that influences energy consumption in residential buildings.

Besides air-conditioner, hot water supplier and refrigerator, television is the largest energy user among the household electrical appliances in most of the households. In some households, electromagnetic cooker, warm-water cleaning toilet seat, washing machine (with dryer function) or dishwashing machine consumed more energy than other appliances. Energy saving priorities should be given to these appliances.

Statistical analyses using three different statistical methods have been carried out to understand how the factors influence energy consumptions in Japanese residential buildings. Multiple regression method was used to predict energy consumption in residential buildings with a set of already-known individual variables (floor area, HDD, CDD, etc.), by using linear functions, while the method of neural network was used to analyze the non-linear relationship between energy consumption and individual variables. On the other hand, quantification method analyzed the influence of qualitative variables (e.g. district, building type, etc.) by introducing dummy variables. Due to the different analysis approaches of the three different statistical methods, the influences of individual variables were different in three methods. However, heating degree day (HDD) has been clarified as an important factor that influences annual energy consumption by all the three statistical methods. Besides that, district was the most important qualitative factor, and the considerable reason is because different districts reflect different heating periods and heating areas.

However, the set of individual variables used in statistical analyses should be further discussed. For example, factors related to human behaviors (e.g. operating schedule of individual heating and cooling equipment, setting temperature of hot water used for bath and/or shower, utilization of natural energy (e.g. photovoltaic system, natural ventilation), etc.) should be taken in to account. Researches and analyze methods focus on the influence of human behaviors on residential energy consumption should be developed in the soon future.

Acknowledgement

This part comes from Doctoral Thesis (obtained from Tohoku University) by Xie Jingchao at Beijing University of Technology. Data for the research was provided by Research Committee on Investigation on Energy Consumption of Residential Buildings (2001-203) and Committee on Energy Consumption of Residential and Countermeasures for Global Warming (2004-2005) of the Architectural Institute of Japan. The analysis used the data base of Cd-Rom titled “Energy consumption for residential buildings in Japan. (Shuzo Murakami, Shin-ichi Akabayashi, Takashi Inoue, Hiroshi Yoshino, Ken-ichi Hasegawa, Kazuhiro Yuasa, Toshiharu Ikaga: Energy consumption for residential buildings in Japan, Architectural Institute of Japan, Maruzen Corp., 2006.)

2.6 Experience 6: Heating consumption assessment and forecast of existing buildings: investigation on Italian school buildings

(Stefano Paolo Corgnati, Federica Ariaudo, Marco Filippi)

2.6.1 Introduction

The study here presented is part of the long-term analysis of actual energy consumption, aimed at monitoring, analyzing and forecasting energy consumption of a school building stock. Results can be also useful for the definition of priorities in building energy upgrading.
The provisional models for energy forecasting should be robust and simple, based on easily collectible data, in order to show with reasonable precision the building stock energy demand tendencies.
With reference to Figure 2-116 inspired by the outlines of IEA- ECBCS ANNEX 53, the investigation here presented can be referred to the box related to large building stocks, where seasonal consumptions are available and analyzed.

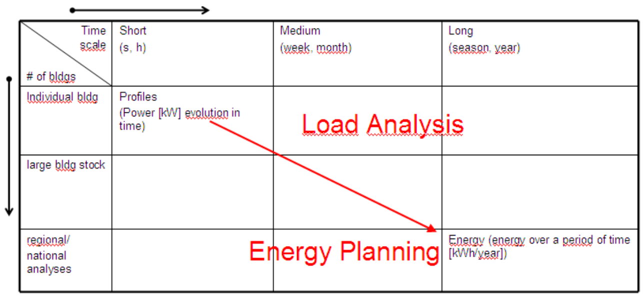


Figure 2-116. Building sample dimension vs Time frequency of energy consumptions (from hourly data to annual data)

The analysis is firstly carried out in order to define a specific heating energy consumption indicator for each school building. A simplified model, defined and proposed by a previous research work (*Corgnati et al., A method for heating consumption assessment in existing buildings: a field survey concerning 120 Italian schools, Energy and Buildings, Vol. 40 pp. 801-809, 2008*), is applied and then verified for the energy consumption forecast.
In particular, the investigation is carried out through the following steps:

1. measured data collection and analysis of heating consumptions for each single building of the investigated stock;
2. verification of the heat generation efficiency for each analysed heating plant;
3. definition of actual and conventional occupancy hours of the buildings;
4. definition of the energy performance indicator for each single building
5. energy consumption forecast for each building and for the building stock.

Moreover, the validity of the performance index proposed in the former research activity is verified, and the results of the previous investigation campaign are compared with the ones obtained from the new data set consequently to energy retrofit actions on the buildings.

2.6.2 **Database characteristics**

A building sample of 103 schools is investigated, located in the Province of Torino. Diagram in Figure 2-117 shows the frequency distribution of the building sample on the basis of the gross heated volume.

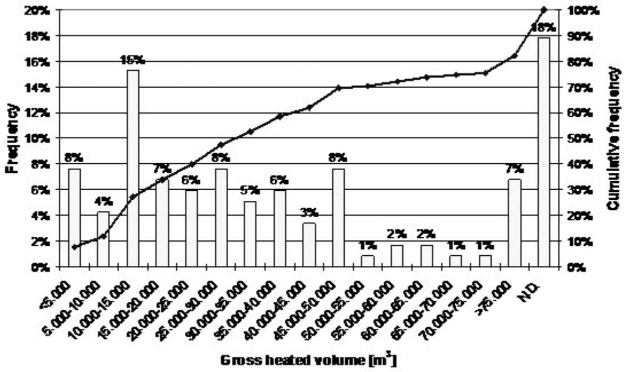


Figure 2-117. Frequency distribution of the building sample on the basis of the gross heated volume

The diagram of Figure 2-117 shows that the sample is divided within an extremely wide interval, with heated volumes ranging from under 5000 m^3 to over 75000 m^3.
Moreover, Figure 2-118 shows the frequency distribution of the building sample on the basis of the Italian conventional heating Degree Days (in accordance to the Italian Law D.P.R. 412/93), calculated with an indoor temperature of 20 °C. All the involved buildings are located within the climatic zones E (Degree Days between 2100 and 3000) and F (Degree Days over 3000), that are the two zones coldest climatic zones in Italy.

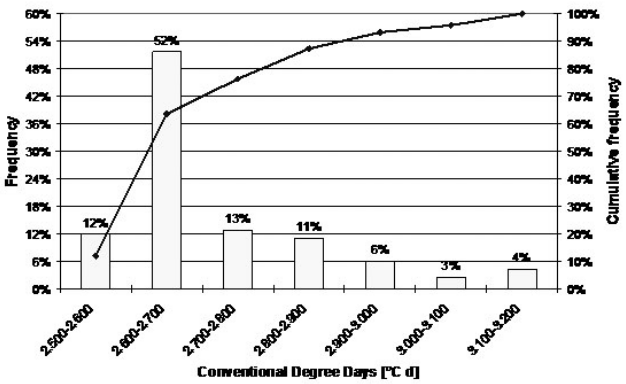


Figure 2-118: Frequency distribution of the building sample on the basis of the Italian conventional heating Degree Days (Italian Government 1993)

As most of the buildings are located in the city of Torino (2617 °C d), more than half of the sample falls within the area with conventional Degree Days between 2600 °C d and 2700 °C d.
Diagram in Figure 2-119 shows the distribution of the sample according to the type of fuel used for the heating system (in the case of methane, a further division into traditional and condensing boilers has been made).

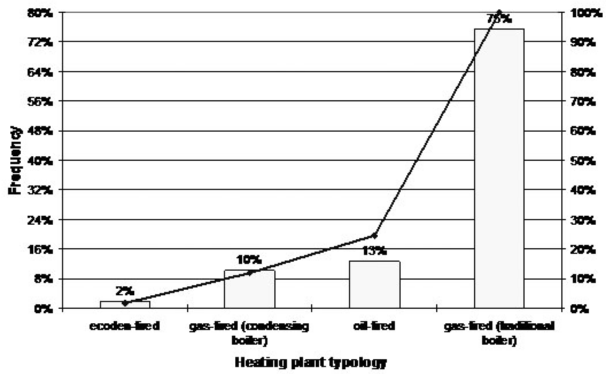


Figure 2-119: Frequency distribution of the building sample according to the type of fuel used for heating system.

The data collected for each building in order to characterize its heating consumptions and, then, to define the heating consumption indicator are:
building location;
geometric data (gross heated volume, floor useful surface);
monthly metered supplied energy for heating for three consecutive heating seasons;
monthly metered primary energy consumption for heating for three successive heating seasons;
conventional heating Degree Days of building site;
measured heating Degree Days in building site for each of the examined seasons;
conventional (standard) heat delivery hours of the heating system;
actual heat delivery hours of the heating system for each examined season;
type of fuel for the heat generator.

As mentioned, data referring to three heating seasons have been considered.
The number of conventional heating delivery hours for the analyzed building stock has been set at 1,098 hours/year, according to the indications of the energy manager on the basis of a standard outdoor climate and a standard use of the building. On the contrary, as regards the actual delivery hours, they were monitored by the energy manager: as an example, Figure 2-120 presents their frequency distribution for the sample with reference to the first analyzed heating season.

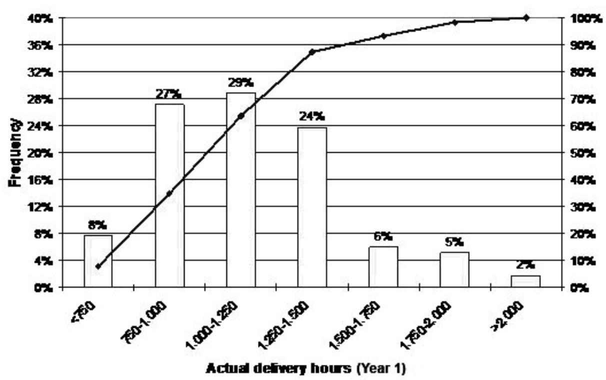


Figure 2-120: Frequency distribution of the building sample on the basis of the actual number of delivery hours for the first analysed heating season.

The monthly efficiency of the heat generator has been calculated for every building through the following equation:

$$\eta = \frac{E_s}{E_p}$$

where:
η = average monthly efficiency of the heat generator;
E_s = Monthly metered supplied energy;
E_p = Monthly metered primary energy.

Also the average seasonal efficiency of the heat generator has been evaluated, using the seasonal metered energies.
In order to verify the consistency of the metered values and as a consequence of the efficiency, the calculated monthly efficiency was checked. When the ratio between monthly supplied energy and monthly primary energy is over 0,90 (1,01 in the case of condensing boilers), the obtained efficiency value is considered as not reliable, due to a probable malfunction in the heat recording system. The pie chart in Figure 2-121 shows the consistency of reliable efficiency data.

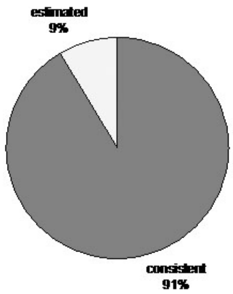


Figure 2-121: Consistency of reliable efficiency data related to the efficiency of the heat generator

Only 9% of the data presents an anomalous efficiency value. For the cases when the efficiency value resulted as unreliable, some standard efficiency values, obtained from an analysis performed on the collected data, have been defined: typically 0,83 for traditional systems, and 0,96 for condensing boilers have been used.

2.6.3 **Method**

In order to carry out a comparison among data obtained from different buildings and/or different seasons, it is necessary to define an energy indicator referred to the heated volume and which neutralizes the effect of any changes among buildings in the heat delivery period, and climate. To this aim, a "conventional" energy performance indicator, already used in the previous study, has been evaluated: such indicator is defined as the ratio between the metered energy supplied by the heat generator (QP) and the gross heated volume (V), referred to the conventional heating Degree Days of the site (DD_c), and to the conventional hours of heat supply (d_c):

$$QP_{s,c} = \frac{QP}{V} \cdot \frac{DD_c}{DD_a} \cdot \frac{d_c}{d_a} \qquad (4)$$

where:

DD_a = actual heating Degree Days of the site;
d_a = actual heat delivery hours during the heating season;
QP = seasonal metered energy supplied by the heat generator.

Such indicator has been calculated for each building, referring to climatic conditions and conventional hours of heat supply during each considered heating season, in order to make possible comparisons at the same conditions, to show trends, or to highlight anomalies.
As an example, Figure 2-122 describes the trend of the indicators calculated each heating season for building n° 1.

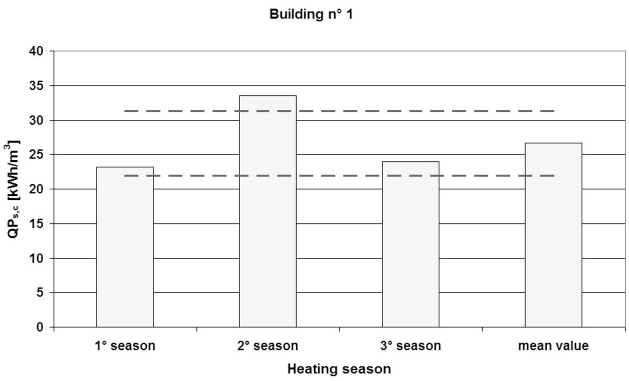


Figure 2-122: Trend of the "conventional" energy performance indicator $QP_{s,c}$ calculated each consecutive heating season for building n° 1

The diagram shows that the indicator presents almost the same value for season 1 and 3. On the contrary, the consumption indicator for the heating season 2 is significantly higher: even if during season 2 the climate was far less rigorous than usual (it was a "warm" winter), in practice such climatic variation did not lead to a proportional decrease in consumption, as highlighted by the high value of conventional consumption. This may be due to poor control strategies on the heating systems, a reduced efficiency of the systems with low-loads, and/or to tendency in slightly increasing the indoor temperature set-points with increasing of outdoor temperature.

2.6.4 **Results and discussion**

Figure 2-123 shows the frequency distribution of the consumption indicator for a representative season of previous research activity, and for a representative season of current research activity (so after the application of energy retrofit actions on the buildings).

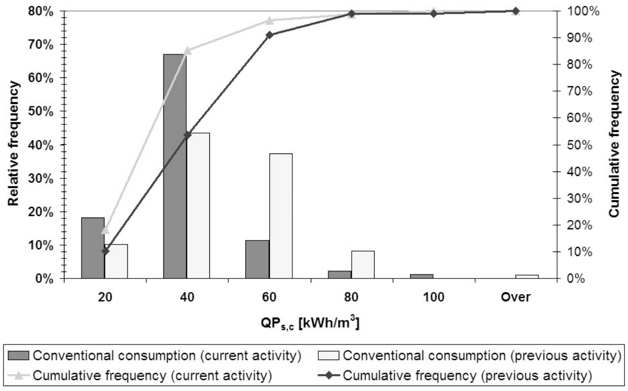


Figure 2-123: Frequency distribution of the "conventional" energy performance indicator $QP_{s,c}$ for a representative season of the previous research activity and of the current research activity

This figure confirms that the retrofit interventions planned after the first investigation campaign have evidently led to an energy efficiency improvement in the whole sample. In fact, while in the first case about 90% of the buildings had a consumption indicator below 80 kWh/m^3, in the second case about 85% of the buildings show a consumption indicator below 40 kWh/m^3: exactly half of the value obtained from the previous study.

Moreover, in the first case the average consumption indicator value was 38 kWh/m^3, while this value is now 29 kWh/m^3. Assuming an average room height of 3 m, the average consumption indicator value referred to the surface unit is 87 kWh/m^2, which is definitely lower than the value of 115 kWh/m^2 obtained from the previous estimation.

The diagram in Figure 2-124 shows the comparison between the actual measured value and the performance indicator value for every building.

The performance indicator has been than used to forecast the consumption of the following heating season. As a consequence, it is necessary to redefine the reference value according to the actual conditions of the analyzed season, and the following equation has been used:

$$QP_{s,c*} = QP_{s,c} \cdot \frac{DD_a}{DD_c} \cdot \frac{d_a}{d_c}$$

where $QP_{s,c}$ is the energy performance indicator for each building, determined as previously described.

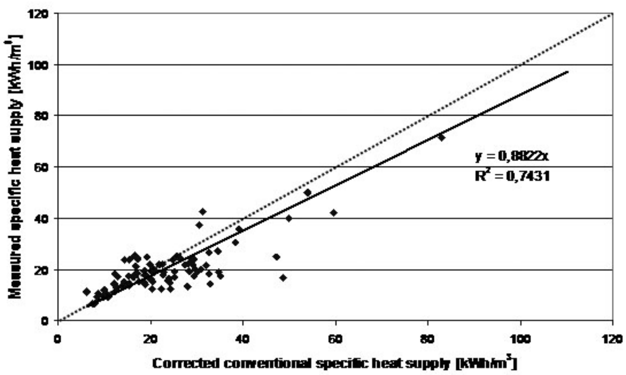


Figure 2-124: Comparison between the actual measured value and foreseen valued from $QP_{s,c}$

The figure confirms the quite good level of accuracy of the energy performance indicator: even if a simple model has been used, the determination coefficient value R^2 is 0,74.
The trend line does not coincide with the bisector of the diagram, which means that the energy performance indicator slightly overestimates the actual consumption value.
One of the main physical reason for such result is that the variables describing the model do not take into account heat gains. Nevertheless, a little overestimation can be considered acceptable when the consumption forecast is used for "energy contract" purposes because it provides a margin of safety of primary importance.

2.6.5 **Conclusions**

The present study, focused on school buildings, aimed at analyzing actual heating energy consumptions and evaluating an energy performance indicator for consumption analysis and forecast of the examined building stock.
The heating energy consumption has been investigated firstly through a comparison between the consumptions obtained from a previous study on the same buildings and the ones related to the current situation. This analysis has demonstrated that the energy re-qualification actions planned and implemented on the buildings according to the results of the previous study, led to an improvement in the energy quality of the analyzed building block.
Then the energy indicator, based on a normalization of the seasonal metered energy consumptions, has been calculated and verified. The indicator resulted reliable, both for the description of the actual consumption in comparison with reference values, and for the future consumption forecast. Moreover, it turned out to be useful to identify reference values for the heating consumption of this school building stock.
Overall, this paper presents a model, indicators and a method for energy consumption analysis and prediction in a public building stock. In this context an accurate energy consumption prediction is very important in order to allocate in advance necessary funds for energy retrofit of schools. Therefore it is important that these models are easy to understand for all those involved in management of analyzed building stock. In school building stock, where heating is the main energy use, consumption analysis involve a not excessively complex system, so prediction accuracy could be higher than in consumption analysis of building stock where, for example, occupant adaptive actions are more free.

2.7 Experience 7: Building energy performance evaluation using daily consumption data in nine individual office buildings in Spain

(Stoyan Danov, Jordi Carbonell, Jordi Cipriano. CIMNE, Spain)

2.7.1 Subject of the work / Abstract

The subject of work is the analysis of the energy performance of nine office buildings in Spain, corresponding to the climatic zone "D" according to the Spanish Building Technical Code. The dimensions of the buildings vary between 340 and 4820 m2.
A method for determining the total heat loss coefficient, the effective heat capacity and the net solar gain of a building is presented. The method uses a linear regressions approach based on daily energy consumption combined with readily available meteorological data.
The effective heat capacity of the building is evaluated by correlating the energy consumption and outdoor temperature changes from the previous day. The net solar gain of the building is assessed by analyzing the data separated into groups by amount of daily solar irradiation. Corrected total heat loss coefficient is determined by explicitly including in the building's energy balance the accumulated heat and the solar gain.
The method has been applied to the analysis of nine public buildings in Spain. An improvement of the estimated heat loss coefficient due to the corrections is observed. The effective heat capacity normalized by the building area is shown to be a useful indicator of the building operation, detecting continuous or intermittent heating.
The estimated parameters in this study can enable specific benchmarking, detecting opportunities for energy savings and evaluating their potential. With the increasing implementation of smart metering technologies, the method is promising for application to the analysis of large building stocks.

2.7.2 Aim of the work

In order to improve the energy efficiency in existing buildings and to design appropriate energy-saving measures, it is important to analyze separately the influence of the building (envelope, energy systems) from the influence of the human behavior factors (heating/ventilating choice, system control).
Ideally, the analysis needs to establish weight of each of the following key factors in the building's total energy consumption: envelope quality; the building use activities; the systems quality and control; occupant behavior factors such as heating and ventilating choice. This detailed knowledge requires expensive monitoring which, in practice, is only possible in a small fraction of the total existing building stock due to cost criteria.

Another approach is the use of energy indicators, usually calculated from billing data sources. Indicators are used for benchmarking to evaluate the savings potential and also the impact of efficiency measures already applied. In this case, the level of detail of the analysis depends on the type of indicators used and on how well they are able to represent the aforementioned key factors for the energy consumption. Furthermore, benchmarking and modeling can be integrated in the same analysis by studying the characteristic building factors with artificial neural network (ANN) techniques [1].
A commonly used indicator for the combined performance of the building and its occupants is the energy use intensity (EUI), or energy consumption per square meter. For characterization of the

building envelope, the total heat loss coefficient (K_{tot}) including both the transmission and ventilation losses of the building, is widely used [2],[3],[4].

The K_{tot} is a quantification of the steady state thermal performance of the building. As the frequency of the data for analysis increases, to the level of using daily data for example, the dynamic performance deserves more attention. One of the ways to model the dynamic effects is the effective heat capacity, defined in [5] as the part of the total heat capacity of a building component that participates in dynamic heat exchange with the environment. In the same reference different analytical models for determining this are presented, all of which require information about the building element properties and dimensions.

The present study aims to analyze building energy performance from the starting point of energy consumption data availability without detailed knowledge about the building, calculating the effective heat capacity and the net solar gain of the building.

2.7.3 **Data used and database characteristics**

The study has been performed on data from 9 public buildings in Spain corresponding to the climatic zone "D" according to the Spanish Building Code. The data has been obtained, from monitoring in the period of April 2009 to March 2010, under a commercial contract with privacy clauses which impede the citation of the exact location. Consequently, the data is used here only as an illustration of the methodology.

The original database contains data with 15 minute sampling frequency, with separation of consumption of fuel for heating and total electricity consumption of the building. The energy consumption information has been complemented with readily available meteorological information at 10 minute frequency, taken from nearby meteorological stations. From these data the outdoor temperature and the global solar irradiation on a horizontal surface have been used. In some cases the physical distance from the building to the meteorological station is up to 20 km which can give place to some micro-climatic deviations for the real buildings.

Additionally, information about the typical number of people occupying the building each hour is available from a questionnaire filled by the building operator.

Although the original database contains data with higher frequency, daily integrated values for the energy consumption and the solar irradiation, and average outdoor temperature have been used in the analysis. The reason is that daily data is better correlated with respect to dynamic and solar radiation effects than higher frequency data due to the thermal inertia of the building. The data with higher frequency in this study is used only for more in-depth view of the building use in order to explain the results from the daily data analysis.

2.7.4 **Method**

The global energy balance over a building is given by:

$$Q_{loss} + Q_{dyn} = Q_{hs} + Q_{el} + Q_p + Q_{sol} \quad (1)$$

Where Q_{loss} is the energy loss through the building envelope by transmission and ventilation, Q_{dyn} is the dynamically stored/released heat, Q_{hs} is the energy provided from the heating system. The last three terms on the right are the heat gains respectively from electric use (Q_{el}), metabolic energy from people (Q_p) and solar energy gains (Q_{sol}). Both parts of the equation represent the total heating need of the building for the studied period (Q).

For daily energy balance Q_{dyn} cannot be neglected, as the time constant of buildings is in the magnitude of days. Thus dependence on the previous day's operation and outside temperature exists.

2.7.5 Determining of the total heat loss coefficient

The heat losses of the building are proportional to the inside-outside temperature difference, so we can re-arrange equation (1) in terms of power as:

$$K_{tot}(T_i - T_o) = \dot{Q}_{hs} + \dot{Q}_{el} + \dot{Q}_p + \dot{Q}_{sol} - \dot{Q}_{dyn} \quad (2)$$

Where K_{tot} is the total heat loss coefficient of the building including both transmission and ventilation losses, T_i and T_o are respectively the average internal and external temperatures.

Using daily data for the energy consumption for heating, electricity, occupation and average outside temperature, the K_{tot} can be determined from (2) by energy signature, see Figure 2. This technique does not imply the knowledge of the internal temperature (implicitly supposed constant) and evaluates the heating loss in function only of the outside temperature. As is usual in this kind of modeling, $\dot{Q}_{sol}$ and $\dot{Q}_{dyn}$ are not known, so these terms are omitted and K_{tot} is estimated from the known terms instead of the real heat losses. As a result, the heat loss coefficient is more imprecise and does not reflect the real characteristics of the envelope. Accounting for the dynamic effect and the solar gains will lead to improved linearity of the relation between heat loss and temperature, to obtaining of higher determination coefficient (R^2) for the regression, and thus more precise estimates of K_{tot}.

2.7.6 Evaluation of the dynamic effect

The dynamic contribution to the total energy for heating is due to the thermal inertia of the building and on a daily basis depends on the variation of the average building mass temperature with respect to previous days.

As the average building mass temperature variation cannot be determined with the available information (energy consumption, occupation, meteorological data), the dynamic effect can be evaluated with respect to the variation of the outside temperature.

The dynamic effect on the heating energy demand is related to the effective heat capacity of the building C_{eff} and can be determined from:

$$Q_{dyn} = C_{eff}(T_o^k - T_o^{k-1}) \quad (3)$$

where T_o^k and T_o^{k-1} are the average outside temperature of the actual and previous day respectively.

The effective heat capacity is a quantification of the total heat capacity of the building that participates in the dynamic heat exchange between building and environment. The effective heat capacity can be defined and calculated on basis of admittance [5], [6]. The admittance is the quotient of the heat flux and the temperature oscillation at one surface of a building component. The effective heat capacity of the component is determined as the amplitude of the admittance divided by the oscillation frequency.

Applied to the overall building using daily integrated data, the admittance can be approximated from the oscillation of the heating load in the 24h period to the oscillation of the average outside temperature for the same period.

$$Y = \frac{\dot{Q}^k - \dot{Q}^{k-1}}{T_o^k - T_o^{k-1}} \quad (4)$$

Dividing the admittance from (4) by the oscillation frequency (1/24h) we obtain the effective heat capacity of the building:

$$C_{eff} = \frac{Q^k - Q^{k-1}}{T_o^k - T_o^{k-1}} \quad (5)$$

where Q^{k-1} and Q^k represent the total heating need (including energy from heating system and gains) for two consecutive days, k-1 and k, calculated from (1), without considering the dynamic and solar terms.

Equation (5) is not an exact equation and C_{eff} is determined inversely as the slope of the regression line between the nominator and denominator, see Figure 2-125. It is a negative slope line, ideally passing through the centre of the coordinate system. In order to evaluate C_{eff} considering only the influence from the previous day, from the data set have been excluded the days for which the energy consumption in the previous days is very low or atypical, for example on Mondays or on the first day after the holidays which reflect the dynamic heat effect of more than one day period. These days should be analysed separately.

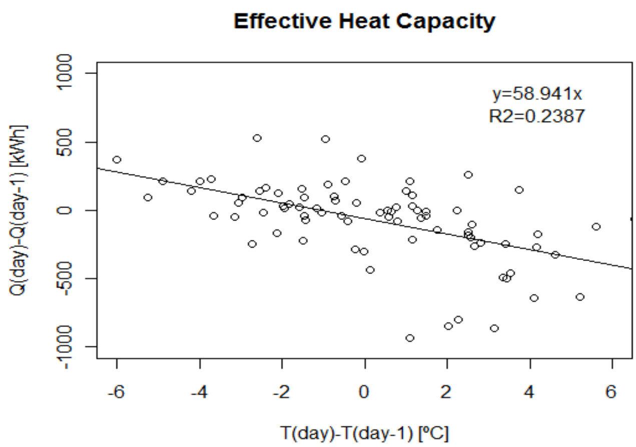


Figure 2-125: Plot for determining of C_{eff} (Building #7)

Analysis of the building solar gain

In the present work, centered in the analysis of the building energy use from daily data energy consumption data and without detailed building description, the focus has been placed on the relation between the global daily solar irradiation on a horizontal surface and the total heating demand of the building.

The solar energy gain contributes to the heating load and is implicitly reflected in the energy requirements for heating. In thermally controlled buildings it can be supposed that higher solar gain reduces the heating necessity of the building. With lower, or nearly zero solar gain, all the necessary heating is covered by the consumed commercially supplied energy. On the other hand, solar gain is related to the available global solar irradiation during the day and can be expected to increase with higher solar irradiation.

From the point of view of analysis of data integrated in a 24h period (including day and night), with the increased amount of daily irradiation the diurnal solar gain will increase. But, this effect will be partly offset by the increased radiation losses during the night due to the lower sky temperature during clear nights, supposing that the coincidence of clear days with clear nights within 24h period is highly probable. These opposite effects will finally give place to a net solar gain or loss, which is the difference between the energy gained during the day and the increased losses during the night. The result for each particular building will depend on its location, design, orientation and surroundings.
In order to evaluate the influence of the solar irradiation on the total heat demand, the whole data set has been divided in four subsets which correspond to different ranges of the daily solar irradiation as shown below.

Set 0: Solar Irradiation level < 800 Wh/m2
Set 1: 800 ≤ Solar Irradiation level < 2000 Wh/m2
Set 2: 2000 ≤ Solar Irradiation level < 3200 Wh/m2
Set 3: Solar Irradiation level ≥ 3200 Wh/m2

The considerations for deciding the division into ranges have been first to assure sufficient data points within each range, and secondly, to provide sufficient step-change in the solar irradiation level as to assure appreciable differences in the output of the analysis.
For each of the subsets, a linear regression between the heat load and the outdoor temperature has been obtained. In order to smooth extraneous effects in the regressions and impose physical meaning by fixing zero energy consumption for the same outdoor temperature for all of the subsets, the regression lines are generated with the constraint of fixing a common point of the regression lines on the temperature axis. As a common point, the cross-point of the regression line obtained with the whole set of data points has been imposed, as shown in Figure 2-126.
The particular arrangement of the regression lines shows the global building energy performance response to the solar irradiation level and represents a sort of building “solar signature”.

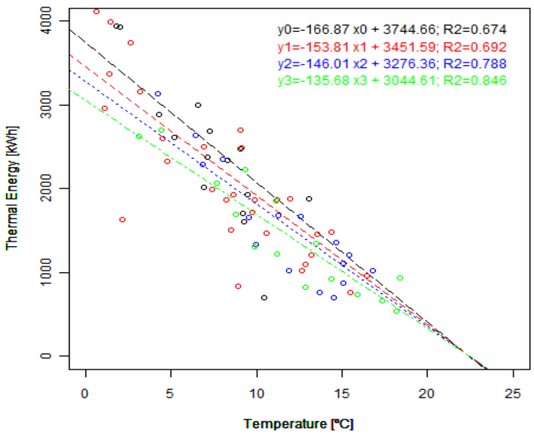


Figure 2-126: Total heat loss coefficient determined by energy signature from sub-sets of data for different levels of solar irradiation with fixed common point.(Building #3)

For the estimation of the solar gain of the building the set of data with minimal solar irradiation (Set0) can be considered as the base set for which the solar gain is zero. All the other sets (Set1 to Set3) are expected to have higher solar gain, resulting in lower demand of energy for heating.

This means that ideally the regression lines of the sets with higher irradiation level will be situated in the graphic lower than those with lower irradiation level, as in Figure 2-126. In the case that the regression line of Set0 is situated below the other sets' lines, this means that with higher irradiation level the building requires more energy for heating, or that the net solar gain is negative.

The daily net solar gain of the building can now be estimated (see Figure 2-126) from the difference between the regression line of Set0 and the line of the corresponding data set (SetX) to which the daily data belongs, calculated for the average outdoor temperature.

$$Q_{sol} = \hat{Q}_{set0} - \hat{Q}_{setX} \quad (6)$$

2.7.7 Results and discussion

Total heat loss coefficient

The total heat loss coefficient for the 9 buildings has been calculated from (2) by linear regression in three ways: i) without considering the dynamic and solar terms (without correction); ii) considering only the dynamic heat effect calculated from eq. (5) (dynamic correction only (K_{tot}^{*})), and iii) considering the dynamic and solar heat effects calculated respectively from eq. (3) and (6) (dynamic + solar correction (K_{tot}^{**})). The results are presented in Table 2-32 calculated per square metre of building area in order to allow comparison between them.

Table 2-32: Values of K_{tot} and R^2 determined from (a) the initially available data without correction; (b) with dynamic correction only; (c) with dynamic and solar correction.

Building	without correction K_{tot} [W/m2K]	R^2 [-]	dynamic correction only K_{tot}* [W/m2K]	R^2* [-]	dynamic + solar correction K_{tot}** [W/m2K]	R^2** [-]	difference (K_{tot}**- K_{tot})/K_{tot} [%]
#1	2,111	0,57	2,202	0,59	2,319	0,73	9,86%
#2	1,586	0,72	1,715	0,77	1,720	0,77	8,45%
#3	0,927	0,48	0,958	0,52	0,891	0,53	-3,86%
#4	0,801	0,38	0,940	0,42	0,880	0,40	9,88%
#5	1,183	0,38	1,209	0,40	1,268	0,47	7,21%
#6	0,758	0,59	0,786	0,65	0,752	0,69	-0,79%
#7	1,231	0,68	1,348	0,75	1,328	0,76	7,91%
#8	1,633	0,68	1,715	0,73	1,685	0,73	3,19%
#9	1,450	0,75	1,558	0,77	1,729	0,80	19,21%

It can be observed that in practically all of the cases the dynamic correction leads to an improvement of the regression from the perspective of the determination coefficient (R^2) values. The addition of the solar correction further improves the regressions' quality, except in the case of building #4, where the R^2 slightly descends in comparison to the K_{tot}^{*} determined with the dynamic correction only.

The regressions obtained with the dynamic and solar corrections show increased linearity of the relationship between the heating need and the outdoor temperature, as expected. The observed corrections of the total heat loss coefficient in the majority of cases lead to an increase of the value within the range of 10%, except for building #9 where an increase of 19% due to the correction is obtained

Effective heat capacity
The quantity of dynamic heat accumulated and released from the building depends on the magnitude of the surrounding temperature variation and the ability of the building to exchange heat with the environment. It is limited to the heat capacity of the materials from which the building is built and is always lower than this, due to temperature cycling over time.

The C_{eff} in this work is obtained by correlating daily integrated values and reflects how the building heat consumption is affected by the changes of the previous day's average external temperature. If the building is intermittently heated during the daily period, it is exposed to larger temperature variations, which are related to the external temperature, and the amount of dynamic heat is expected to be larger. This has been checked using the available higher frequency consumption data and the calculated C_{eff} values for the buildings.

Buildings #1, #4, #7, #8 and #9 present clearly intermittent heating pattern where stopping of the heating system or reducing its operation to minimum during the non-working hours can be observed. Building #2 can also be classified as having intermittent heating although exceptions for some weeks exist where the heating has not been stopped during the night. The control pattern of building #5 presents some irregularities but on the general could be classified as continuous. Figure 2-127 shows the hourly consumption profiles of heating energy of four of the buildings, superposing 6 typical weekly profiles for each. The profiles represent the heating mode of the buildings and are related to the building use. As can be observed from the figure, some of the buildings have well defined control patterns, while others have more irregular operation. The profiles have been evaluated qualitatively, roughly classifying the heating mode as intermittent or continuous.

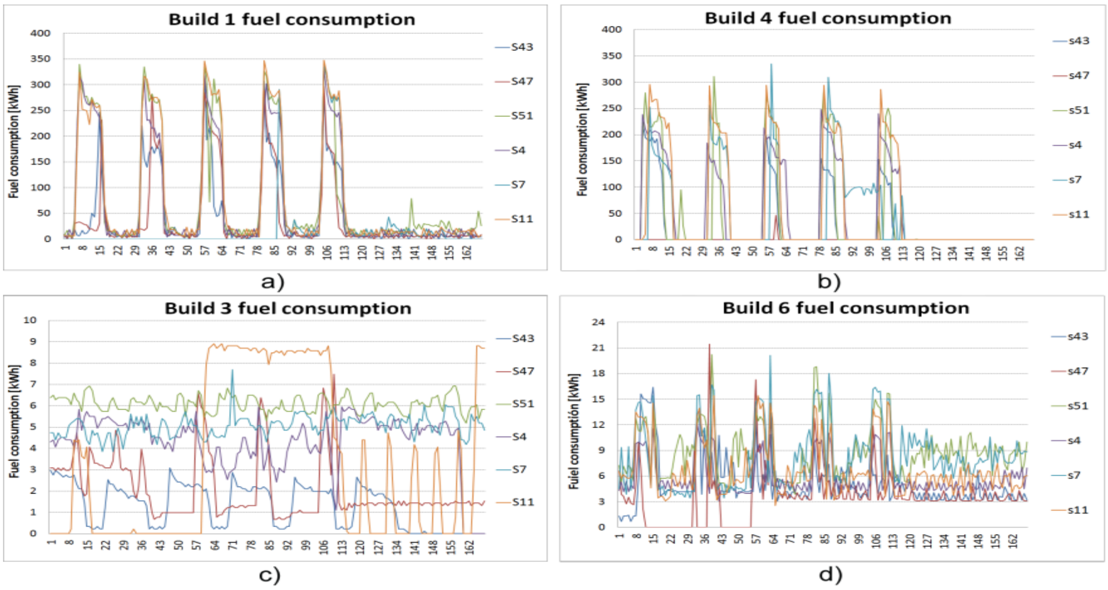


Figure 2-127: Superposed hourly profiles of fuel consumption for heating: six typical weeks for each building are represented. a), b): intermittent operation; c), d) continuous operation

In order to obtain comparable values for the analyzed buildings, the effective heat capacity, determined from equation (4) has been normalized by the net building area. Cross-comparison between the C_{eff} normalized per building area and the detected heating mode is presented in Table 2-33. The obtained results for the C_{eff} for intermittently heated buildings are considerably higher (3 to 5 times higher) than those for continuously heated buildings. This shows clearly that the C_{eff} can be used as indicator for the heating mode of the building: continuous or intermittent.

Table 2-33: Summary results.

Building	**Net area [m2]**	**Av.daily EUI [Wh/m2.day]**	**K_{tot}** [W/m2K]**	**Solar fraction [%]**	**C_{eff} [Wh/m2K]**	**Heating mode [-]**
#1	2470	659,13	2,319	-3,79%	22,78	intermittent
#2	1969	598,18	1,720	1,85%	26,28	intermittent
#3	340	348,29	0,891	8,41%	4,85	continuous
#4	3845	474,64	0,880	-6,31%	24,26	intermittent
#5	500	389,74	1,268	18,69%	6,61	continuous
#6	516	622,50	0,752	3,91%	6,76	continuous
#7	2773	468,65	1,328	-2,09%	20,58	intermittent
#8	1770	666,37	1,685	-4,78%	18,20	intermittent
#9	4817	416,07	1,729	6,79%	28,91	intermittent

Analysis by level of solar irradiation

The results from the analysis by level of solar irradiation are presented in Table 2-34. Here we can see the corrected total heat loss coefficient, determined by introducing the daily solar gain in the buildings' energy balances, and the solar gain of the buildings for the analysed period (September – April), calculated using the method described in point 3.3.

For the present method, the precision in determining of the base set with nearly zero solar gain (Set0) is crucial as it is the reference for determining of the solar gain of the building and can influence strongly the assessment. Analysis with more data is necessary in order to establish the acceptance criteria. A minimum number of data points in the subsets, or fixing minimal value of the determination coefficient of the regression could be part of the criteria.

Table 2-34: Ktot determined from subsets of data divided by level of solar irradiation and calculated solar gain of the building for the analysed period.

	Ktot by solar irradiation level, [W/m2K]				
Building	$K_{tot\ set0}$	$K_{tot\ set1}$	$K_{tot\ set2}$	$K_{tot\ set3}$	**Solar gain**
#1	2,31	2,60	2,48	2,11	-3,79%
#2	1,65	1,63	1,53	1,63	1,85%
#3	1,08	1,04	0,99	0,81	8,41%
#4	0,93	1,03	0,98	0,95	-6,31%
#5	1,65	1,58	1,22	0,81	18,69%
#6	0,41	0,41	0,39	0,36	3,91%
#7	1,35	1,40	1,49	1,24	-2,09%
#8	1,69	1,77	1,77	1,88	-4,78%
#9	1,60	1,52	1,44	1,33	6,79%

The results of the analysis in this study show that by using daily integrated data for obtaining the total heat transfer coefficient K_{tot} its value, in general, decreases with the increase of the solar irradiation; and that the heating demand is lower when the solar irradiation is higher, which is logical because the solar gain contributes to the heating.

2.7.8 **Conclusions**

The present work suggests a simple linear regression-based method for determining the building's total heat loss coefficient, effective heat capacity and solar gain by using daily energy consumption data. Consequently, introducing the calculated dynamic and solar gain terms explicitly into the overall energy balance of the building enables an improved total heat loss coefficient to be determined. It has been observed that the dynamic heat correction leads to an improvement of the regression between energy consumption and outdoor temperature in all of the cases studied. The addition of the solar gain correction further improves the regressions, except for one of the buildings where the regression's determination coefficient slightly decreases.

The analysis of the detailed hourly consumption profile of the buildings showed that the thus obtained effective heat capacity, normalised by the building area, is closely related to the building's operational pattern and is a clear indicator for intermittent or continuous heating. Intermittently heated buildings present effective heat capacity values from 3 to 5 times higher than those heated continuously.

The three parameters - the corrected total heat loss coefficient, the effective heat capacity and the solar gain - can be used as performance indicators for specific benchmarking in order to detect underlying building operational patterns with available only daily data. In order to establish clearer criteria for interpreting of the results, additional studies and analysis of larger number of buildings is necessary.

With the increasing application of smart metering in building energy management systems and for billing by the utility companies, daily or even hourly consumption data is available, giving the possibility to evaluate these energy performance indicators in a large stock of buildings without any additional measurement cost. The analysis can be used for preliminary evaluation of the energy saving potential and development of energy saving strategies by service companies, or for development of additional services for utility companies' customers.

2.7.9 **References**

[1] X. Cipriano, J. Carbonell, J. Cipriano, Monitoring and modelling energy efficiency of municipal public buildings: case study in Catalonia region, International Journal of Sustainable Energy Vol. 28, Nos. 1–3, March–September 2009, 3–18.

[2] C. Ghiaus, Experimental estimation of building energy performance by robust regression, Energy and Buildings 38 (2006) 582-587.

[3] J.-U. Sjögren, S. Andersson, T. Olofsson, An approach to evaluate the energy performance of buildings based on incomplete monthly data, Energy and Buildings 39 (2007) 945-953.

[4] J.-U. Sjögren, S. Andersson, T. Olofsson, Sensitivity of the total heat loss coefficient determined by the energy signature approach to different time periods and gained energy, Energy and Buildings 41 (2009) 801-808.

[5] Jan Akander, The ORC Method – Effective modelling of thermal performance of multi-layer building components, PhD thesis, Bulletin no 180. KTH – Building Technology, Stockholm, 2000.

[6] G. Johannesson, Active heat capacity - Models and Parameters for Thermal Performance of Buildings. Doctoral Dissertation, Report TVBH-1003, TTH, Lund, Sweden, 1981.

2.8 Experience 8: Experimental estimation of building energy performance by robust regression (Synthetic contribution)

(Christian Ghiaus. INSA, Lyon)

2.8.1 Subject of the work

The subject of work is to estimate the HVAC energy consumption from hourly or daily data by using the concept of free-running temperature.

2.8.2 Aim of the work

Estimation of energy performance indexes, such as the heating curve or the energy signature, requires robust regression of the heating losses on the outdoor temperature. The solution proposed in this paper is to use the range between the 1st and the 3rd quartile of the quartile – quartile (q-q) plot to check if the heating losses and the outdoor temperature have the same distribution and, if yes, to perform the regression in this range of the q-q plot. The result is a model that conserves its prediction performance for data sets of the outdoor temperature different of those used for parameter identification. The robust model gives the overall heat transfer coefficient and the base temperature, and it may be used to estimate the energy consumption for data sets of the outdoor temperature coming from different locations or time intervals.

2.8.3 Database characteristics

The database contains hourly data for one school:

- Consumed fuel energy (hourly)
- Consumed electrical energy (hourly)
- Number of occupants in the building (hourly)
- External temperature (hourly)
- Global solar irradiation (hourly)

2.8.4 Method/Methods applied for the data analysis

The energy signature of the building is related to the overall heat loss coefficient of the building (or U-value) which is the mean thermal transmittance through building envelope to the external environment by conduction and by ventilation. Linear regression finds out a relationship between two variables by fitting a linear model to observed data. The regression model of heating load as a function of outdoor temperature and the frequency of occurrences of outdoor temperature may be used to estimate the energy consumption. The assumptions made for linear regression are that the outdoor temperature has a normal distribution and that the heating load is a random variable of mean.

The above conditions are not satisfied in real situations: the building is not air-conditioned at a constant temperature for the whole range of the outdoor temperature. Consequently, the outdoor temperatures which correspond to the heating period does not have a normal distribution.

A robust regression based on quantile – quantile plot is proposed to mitigate this problem (quantiles indicate the number of elements of a random variable that are in a given range).

2.8.5 Main results

Many modern buildings are equipped with Building Energy Management Systems (BEMS) that control the indoor temperature and record the energy consumption and the outdoor temperature. These data may be used to assess the energy performance of the building, such as the heating load as a function of the outdoor temperature. This relation may be used to evaluate the overall heat transfer coefficient of the building represented by the slope of the heating load.

2.8.6 Related publications

C. Ghiaus (2006) Experimental estimation of building energy performance by robust regression, Energy and Buildings, 38, pp. 582-5987

2.9 Experience 9: Equivalence between the load curve and the free-running temperature in energy estimating methods (Synthetic contribution)

(Christian Ghiaus. INSA, Lyon)

Synthetic contribution is missing

2.9.1 Related publications

C. Ghiaus (2006). Equivalence between the load curve and the free-running temperature in energy estimating methods. Energy and Buildings 38 (2006) 429–435

2.10 Experience 10: Mining Hidden Patterns from Real Measured Data to Improve Building Energy Performance (Synthetic contribution)

(Zhun (Jerry) Yu, Fariborz Haghighat)

2.10.1 Subject of the work

In order to mine hidden useful knowledge about building energy performance improvement from measured building-related data, we proposed a rational data analysis process and a systematic data mining framework within the building engineering domain. For demonstration purposes, a number of efficient data analysis methodologies were developed based on the framework to account for the interactions between building energy consumption and its influencing factors. These methodologies were applied to 80 Japanese residential (both single-family and multi-family houses) which are located in six different districts of Japan.

2.10.2 Aim of the work

- To develop a new data analysis methodology of establishing reliable building energy demand models, which are interpretable and can be easily used by common users without a priori knowledge in advanced mathematics and statistics.
- To develop new data analysis methodologies for studying building occupant behavior, such as quantitatively identifying the effect of occupant behavior on building energy consumption, and identifying occupant behavior that needs to be modified.

2.10.3 Database characteristics

Number of Buildings: 80 residential buildings distributed in six different districts in Japan
House type: both single-family and multi-family
Period: Dec. 2002 to Nov. 2004
Measured data: energy data + indoor environmental data (temperature+ relative humidity)
End users: three-level classification
On-line database: available (energy indicator: secondary energy)

2.10.4 Method/Methods applied for the data analysis

- Based on the decision tree method, a new methodology for establishing building energy demand predictive models was developed.
- Based on a basic data mining technique (cluster analysis), a new methodology for examining the influences of occupant behavior on building energy consumption was developed. Moreover, to deal with data inconsistencies, min-max normalization was performed as a data preprocessing step before clustering. Grey relational grades, a measure of relevancy between two factors, were used as weighted coefficients of different attributes in cluster analysis.
- Based on three basic data mining techniques: cluster analysis, classification analysis, and association rules mining, a methodology for identifying and improving occupant behavior in existing residential buildings was developed. End-use loads were divided into two levels (i.e. main and sub-category), and they were used to deduce corresponding two-level user activities (i.e. general and specific occupant behavior) indirectly. Cluster analysis and classification analysis were combined to analyze the main end-use loads, so as to identify energy-inefficient

general occupant behavior. Then, association rules were mined to examine end-use loads at both levels, so as to identify energy-inefficient specific occupant behavior.

2.10.5 Main results

- A new methodology for establishing building energy demand predictive models was developed. The developed model estimates the building energy performance indexes in a rapid and easy way. This method's advantage lies in the ability to generate accurate predictive models with interpretable flowchart-like tree structures that enable users to quickly extract useful information. To demonstrate its applicability, the method was applied to estimate residential building energy performance indexes by modeling building energy use intensity (EUI) levels (either high or low). The results demonstrate that the use of the decision tree method can classify and predict building energy demand levels accurately (93% for training data and 92% for test data), identify and rank significant factors of building EUI automatically.
- A new methodology for examining the influences of occupant behavior on building energy consumption was developed. To demonstrate the applicability of the proposed method, the method was applied to a set of residential buildings' measurement data. The results show that the method facilitates the evaluation of building energy-saving potential by improving the behavior of building occupants, and provides multifaceted insights into building energy end-use patterns associated with the occupant behavior. The results obtained could help prioritize efforts at modification of occupant behavior in order to reduce building energy consumption, and help improve modeling of occupant behavior in numerical simulation.
- A methodology for identifying and improving occupant behavior in existing residential buildings was developed. In order to demonstrate its applicability, this methodology was applied to a group of residential buildings in Japan, and one building with the most comprehensive household appliances was selected as the case building. The results show that, for the case building, the method was able to identify the behavior which needs to be modified, and provide occupants with feasible recommendations so that they can make required decisions. Also, a reference building can be identified for the case building to evaluate its energy-saving potential due to occupant behavior modification. The results obtained could help building occupants to modify their behavior, thereby significantly reducing building energy consumption. Moreover, given that the proposed method is partly based on the comparison with similar buildings, it could motivate building occupants to modify their behavior.

2.10.6 Related publications

[1] Z. Yu, F. Haghighat, B.C.M. Fung, H. Yoshino. A decision tree method for building energy demand modeling. Energy and Buildings. 42(10) (2010) pp. 1637-1646.

[2] Z. Yu, B.C.M. Fung, F. Haghighat, H. Yoshino, Edward Morofsky. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and Buildings. 43(6) (2011) pp. 1409-1417.

[3] Z. Yu, F. Haghighat, B.C.M. Fung, H. Yoshino, Edward Morofsky. A Methodology for identifying and improving occupant behavior in residential buildings. To appear in Energy

3. Statistical analysis of national building stock

3.1 Introduction

Statistical analysis of national building energy consumption is aimed to define a general overview of the energy end use due to the construction sector, at national level. Actually, the knowledge of national building energy use has remained under-investigated, due to a lack of information regarding the overall characteristic. With the aim of building strong national databases, national agencies and institutions (CBECS in the U.S., MOHURD in China, TABULA in Europe) have gathered real energy use data and physical characteristics on the national building stock. Specifically, China has collected data of government office buildings and large-scaled commercial building [1], U.S. has built a national sample database of commercial buildings [2], whereas the European countries have collected data characterizing national building residential stock [3].
Subject of the task is to collect and subsequently to elaborate data characterizing national building stock in order to offer a realistic interpretation of typical building energy consumption. Different approaches have been tested and used for the databases statistical analysis.

Wei, Xiao and Jiang [1] adopted two statistical research methods: boxplot and key statistical parameter of energy use data and frequency distribution analysis. Both these two approaches have been presented as effective and suitable for future analysis and international comparison. Database characteristics have been gathered based on regional government website releases and included Gross Floor Area (GFA) and annual electricity consumptions (excluding district heating) of 4600 offices buildings. Cluster analysis showed that the average national stock electric consumption is 107 kWhe/m^{2}a for business office buildings and 67.6 kWhe/m^{2}a for government office buildings.

Hong and Wang [2] analyzed utility bills (monthly energy use for electricity and natural gas) of the CBECS U.S. sample survey and broke them down into energy end use for commercial building national stock. Statistical regressions and engineering modeling approaches were used to estimate national end use based on consumption data. Average energy consumption for the commercial buildings in the U.S. - emerged from monthly regression models of 1518 gathered buildings - is 292.6 kWh/m^2, whereof the single largest part (35.3%) is due to space heating.

The European project TABULA (Typology Approach for Building Stock Energy Assessment) presented by Talà [4] and Becchio, Corgnati, Ballarini and Corrado [3] aimed to create a homogeneous database for European Residential Building Typologies. The research tested three statistical methods with the final goal to estimate the energy consumption of residential building stocks and therefore, to predict the potential energy efficiency measures impact of benchmark models at national level (singular evaluation for each European country participating in the project). These methodologies shoot for the enhancement of the potential impact of energy saving measures and carbon dioxide reduction, by means of the selection of the more adequate energy retrofitting strategies and interventions in existing buildings [3] [4]. Model calculations aimed to estimate the energy saving potential of national residential building stocks (Energy Balanced Method) were developed by four countries (Denmark, Germany, Italy and Czech Republic) representative of main European climatic regions, by using the national EPBD asset rating method [3]. Moreover, the same modeling method (EBM) can be possibly extended for the energy performance assessment of the whole national building stock.

For each country, two levels of building retrofit were considered: (a) standard refurbishment, applying standard national measures; (b) advanced refurbishment, applying the best national technologies available [4]. Specifically, Italian database contained records for more than 66.000 houses rated across Piedmont region and gathered information on physical characteristics and calculated energy requirements of single houses. On the base of three independent variables elaborated by means of statistical analysis (location, age, form of the building), a total of 84 building types (archetypes) representative of the Italian residential building stock were generated [4].

All these kind of approaches, which use statistical analysis of national building stock sample, are very effective. As a matter of fact, average predictions of energy consumption at national level are made available. Public existing building energy use has remained for a long time at a micro-perspective [2] due to a lack of shared definitions and outdated information [3]. Nonetheless the development and the statistical analysis of strong national energy-use datasets, could be one element towards a more robust estimation of the overall energy consumption of the national building stocks.

References (contributions to the Annex 53)

[1] Experience 1: Qingpeng Wei, He Xiao, Yi Jiang, National Database of Office Building Energy Use in China

[2] Experience 2: Tianzhen Hong, Liping Wang. The U.S, Commercial Buildings Energy Consumption Survey (CBECS)

[3] Experience 3: Cristina Becchio, Stefano P. Corgnati, Ilaria Ballarini and Vinecenzo Corrado, Energy saving potentialities by retrofitting the European residential sector

[4] Experience 4: Novella Talà. National/Regional investigation level, Single & Multifamily houses in Italy

3.2 Experience 1: National Database of Office Building Energy Use in China

(Qingpeng WEI, He XIAO, Yi JIANG. School of Architecture, Tsinghua University, Beijing, China)

3.2.1 Introduction

There is a very limited scope survey of energy use in public buildings [1], instead of small scaled investigations, case studies or scenario analysis in China. Thus, the knowledge of public building energy use always remains at a micro-perspective, lacking the understanding of overall characteristics on a regional or national level. In order to build a strong energy data collection system and improve certain problems, the Ministry of Housing and Urban-Rural Development (MOHURD) has set up a long-term strategic plan [2] to gather and collect real energy use data of governmental office buildings and large-scaled commercial buildings in dozens of demonstration provinces, cities and municipalities from 2007. The first round data opened to the public was analyzed in this study. This research offers a realistic interpretation for office building energy use in China, studies its statistical distribution characteristics.

The initial 24 demonstrated provinces, autonomous regions, municipalities under direct control of the Central Government and cities with separate budgets from the central finance have distributed their surveyed data for governmental offices and large-scaled commercial building since 2007. The surveyed data includes Gross Floor Area (GFA) and annual electricity consumption (total value including lighting, office appliances, lift, escalator, HVAC system, circulating pumps, and electrical heating (not including heating energy use which consumes natural gas, steam, etc., including heating energy use of electrical heater, electrical boiler, etc.), and other electrical devices energy use, excluding district heating) of each building (EUI excl. DH), which have been released on regional government website by the end of 2007 and gathered. Those data is distinguished by region or building primary activity, used to calculate the statistical characteristic value (such as median or quartile value), to analyze the dual sector feature by cluster analysis and to calculate Gini-coefficient by each city what follows in this research.

3.2.2 Database characteristics

- Building activity: governmental or business office building
- Energy data period: January 2007 to December 2007
- Contents: building name, building GFA, annual electricity consumption
- Interval: annual
- Total sample number: 4600 office building in 13 cities or provinces (detailed info is shown in Table 2-35)
- Online database: the local department of construction website releases the data online, but the information is written in Chinese
- (For example, Tianjin-
http://www.tjcac.gov.cn/jzjn/detail.asp?articleid=9303&classid=1&parentid=0

Table 2-35: Summary of key data information and sampling size

City/Province code	Urban population (million)	Annual per capita GDP ($USD/capita)	HDD18	CDD26	City scale(1)	Initial sampling size		
						Total	Business	Governmental
A	14.9	10871	2616	103	SL	136	52	84
B	12.4	12188	1648	203	SL	861	601	260
C	6.0	9655	2709	116	SL	892	228	664
D	2.5	6352	289	365	EL	611	241	370
E	5.4	5478	1549	286	SL	467	75	392
F	6.1	5381	1276	60	SL	315	129	186
G	4.3	6904	1684	151	Province	76	N.A(2)	N.A
H	2.7	7759	1481	213	EL	172	N.A	N.A
I	3.5	3293	2135	167	EL	115	54	61
J	2.4	7415	2743	61	EL	226	N.A	N.A
K	1.2	2456	495	291	L	72	N.A	N.A
L	2.6	5187	605	274	EL	597	57	540
M	3.4	2971	105	469	Province	133	N.A	N.A

Note: (1) SL: super large, the urban population of the city is larger than 4 million. EL: extra large, the urban population of the city is larger than 2 million. L: large, the urban population is larger than 1 million. (2) N.A: Not available. Province G, M and City H, J and K, there is only the total number of office building instead of the breakdown of business and governmental ones separately. (3) electric consumption includes lighting, office appliances, lift, escalator, HVAC system, circulating pumps, and electrical heating (not including heating energy use which consumes natural gas, steam, etc., including heating energy use of electrical heater, electrical boiler, etc.), and other electrical devices energy use.

3.2.3 **Methodology**

Frequency distribution

To divide the energy use data into groups and observe the frequency distribution features of EUI excl. DH in 13 selected cities or provinces, the interval was determined as an empirical equation given by H.A.Sturges:

$K = 1 + \frac{\lg n}{\lg 2}$, where n=sampling size of each city or province.

At the same time, the polynomial fitting method is applied to approach the frequency distribution of electricity consumption data, as shown on Figure 2-130. The blank column illustrates the sampling size within each EUI range, while the blue solid line represents the curve fitting using a polynomial with 4th order.

Cluster analysis

Cluster analysis (CA) is a multivariate statistical technique which can group the observations into classes or clusters so that the greater the homogeneity within a group and more distinctions between groups can be easily seen. This method can help us to get a better understanding of the dependencies existing among a set of inter-correlated variables.

The traditional Agglomerative Hierarchical Clustering Method, which starts with the points as individual clusters and, at each step, merging the closest pair of clusters, has been proposed in the study to determine the reasonable classification of office building EUI in each city or province in China.

Samples were defined as (xi, yi) in which xi and yi refer to EUI and GFA of each building sample respectively. The gravity of each cluster was calculated according to equations below, where i representing each building and N representing total number of building samples. It could be considered as a “typical building” representing characteristics of each cluster.

$$EUI_{A,B} = \frac{\sum_{i=1}^{N}(EUI_i \times A_i)}{\sum_{i=1}^{N} A_i} \qquad \text{Equation (1)}$$

$$GFA_{A,B} = \frac{\sum_{i=1}^{N}(EUI_i \times A_i)}{\sum_{i=1}^{N} EUI_i} \qquad \text{Equation (2)}$$

3.2.4 Results and discussion

Boxplot

By creating box plots of annual Electricity Use Intensity excluding District Heating (EUI excl. DH), the maximum and minimum intensity of each city or province is compared, as well as the 25th and 75th percentile limits, as presented in Figure 2-128 and 2-129.

Two significant features are summarized:

The median EUI excl. DH of City-A and City-B is obviously higher than other cities or provinces. While City-I is general lower than that of the others. Take a business office building for instance, the median is 107.0 kWhe/(m^2.a) in City-A and 89.8 kWhe/(m^2.a) in City-B, which is obvious higher than 33.4 kWhe/(m^2.a) in City-I.

The EUI excl. DH of first-tier cities is general higher than the one of second-tier cities; governmental office buildings are lower than business office buildings.

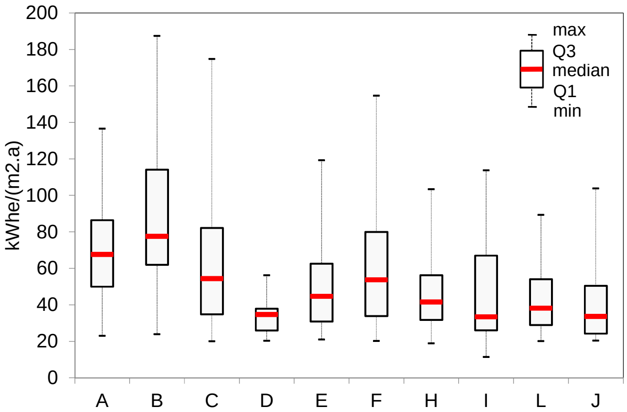


Figure 2-128: Box plot of annual electricity use intensity (excluding district heating) of governmental office buildings in China

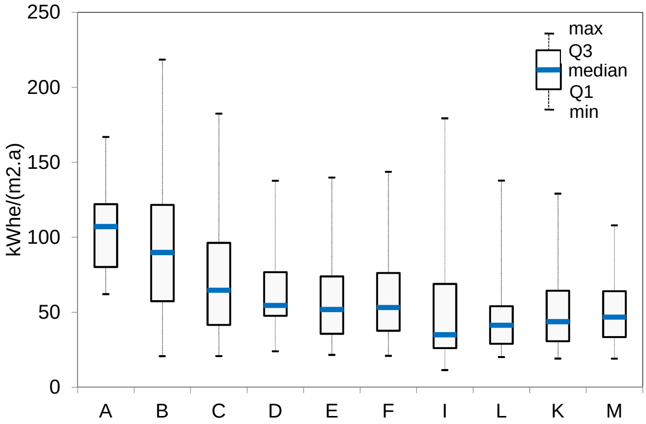
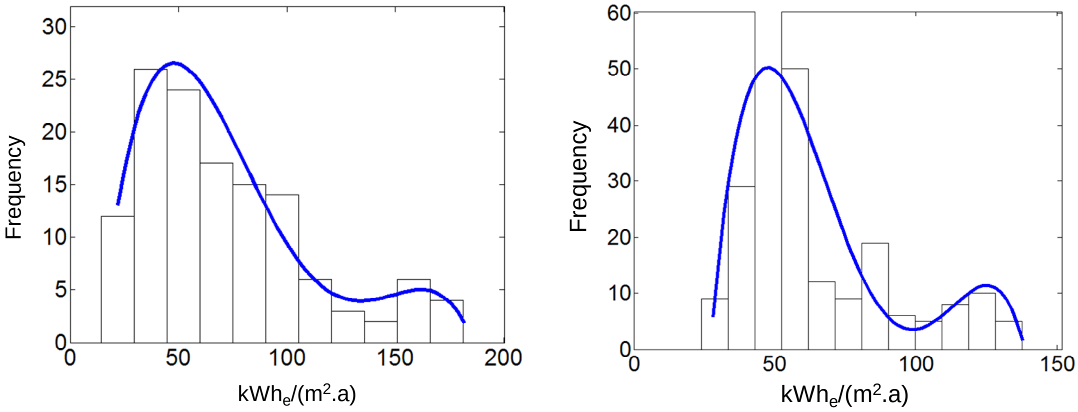


Figure 2-129: Box plot of annual electricity use intensity (excluding district heating) of business office buildings in China

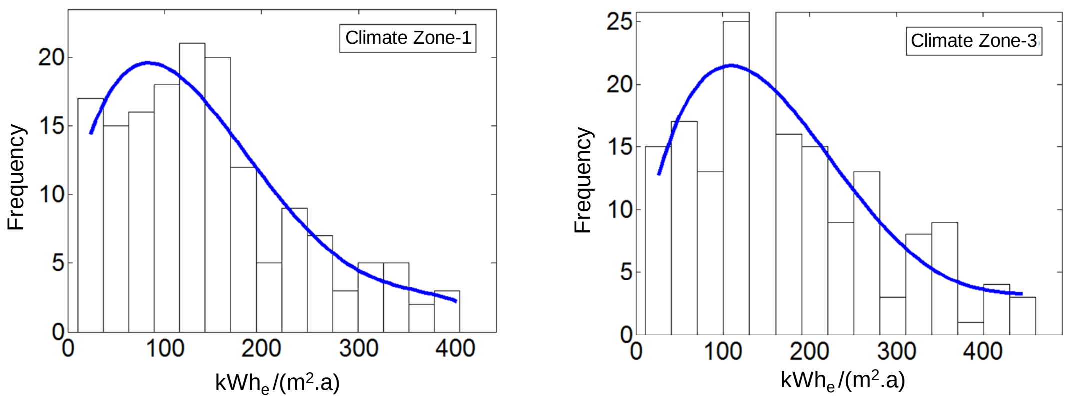
Frequency distribution



(a) City-A: 52 business office buildings (b) City-D: 225 business office buildings

Figure 2-130: Frequency distribution and polynomial fitting plot of electricity use intensity (excl. district heating) of business office building in typical cities or province

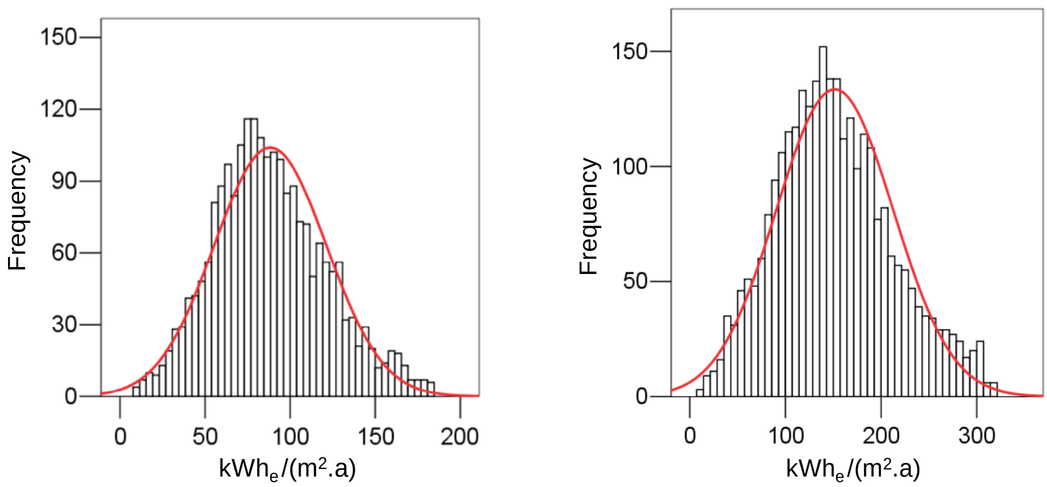
Figure 2-130 illustrates a unique distribution feature of business office buildings' EUI excl. DH in China, which the majority of are centralized over a lower energy range, while the minority of are distributed at a higher energy range. The polynomial curve of electricity consumption data appears to have double peaks, the phenomenon observed based on the large sample survey is defined as "Duel Sector Model", which exists extensively among the 13 cities selected in China. The following Figure 2-131 presents the frequency distribution of office buildings in two typical Climate Zones (CZ) in the United States. The sufficient data contains 6,837 office building samples nationwide, which was selected from the Commercial Buildings Energy Consumption Survey (CBECS) database [3]. The Skewness ad Kurtosis test has been applied in the study, and results shows that the EUI of office buildings in the United States appears the "Right-skewed Single Peak Distribution" feature, which is widely different from China.



(a) U.S. Climate Zone-1 (b) U.S. Climate Zone-3

Figure 2-131: Frequency distribution and polynomial fitting plot of electricity use intensity of office building of climate zone 1 and 3 in the United States

Besides China and the U.S., the statistical distribution feature of office building EUI in Japan also captures our concerns. The Japanese Association for Sustainable Development has developed the Database of Energy Consumption of Commercial Building (DECC) [4], which consists of basic information and monthly energy consumption data for 41,000 commercial buildings, including 2,339 governmental office building samples and 2,951 business office building samples all over the country. After the check of statistical data of DECC, a total of 5104 valid samples are analyzed. The median EUI of governmental office building is 85.4 kWhe/(m^2.a) (only the electricity consumption, not including heating and cooling use, such as heat for domestic hot water or steam), 146.2 of business office building fall below the median, and the average by type is very close to the median. The result of the Kolmogorov- Smiromov (K-S) normality test indicates a normal distribution for EUI of office buildings in Japan, as seen in Figure 2-132.



(a) Governmental office building (b) Business office building

Figure 2-132: Normal distribution for electricity use intensity of office building in Japan

The EUI of office buildings in the U.S. and Japan has the “Single Sector Distribution” feature instead of the “Duel Sector Distribution” characteristics as seen in China. However, slight differences exist between the U.S. and Japan. The former shows Right- Skewness, but the latter one shows a normal curve. The order of EUI from smallest to largest is China, Japan and U.S. What needs to be emphasized is that the EUI in office buildings in China does not include district heating but does include some packaged electrical heating equipment instead.

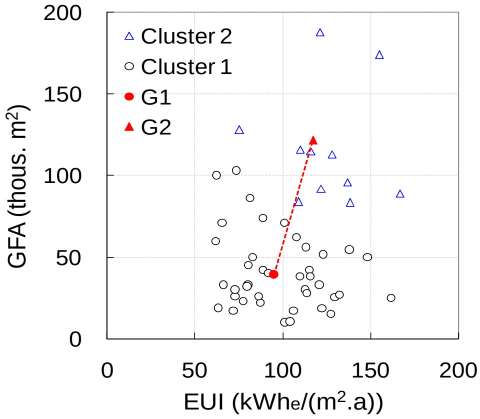
Cluster analysis

Typical results of the cluster analysis in two typical cities are shown in Figure 2-133. Several conclusions can be summarized, including:

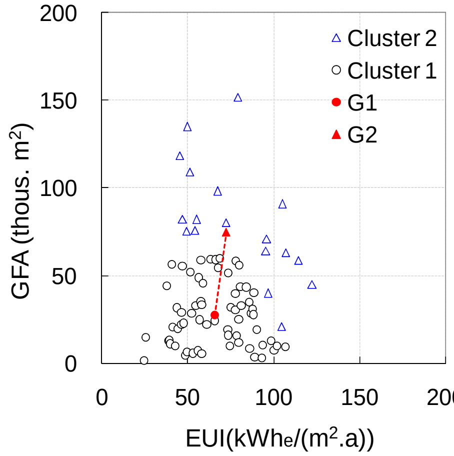
Two clusters in each city or province were identified based on the analysis, which confirms the similar phenomenon on frequency distribution analysis. The distance of two gravities illustrates the difference of two clusters, and the gravity itself represents the energy intensity of the cluster.

The gravity of first-tier cities is usually larger than that of second-tier cities. The gravity of double clusters in each city is shown in Table 2-36. For instance, the two cluster gravities of business office buildings in city-A is (39, 95.1) and (122, 117.2), while the two gravities of city-F are (14, 61.8) and (58, 65.3), the first figure represents GFA, and the latter one refers to EUI.

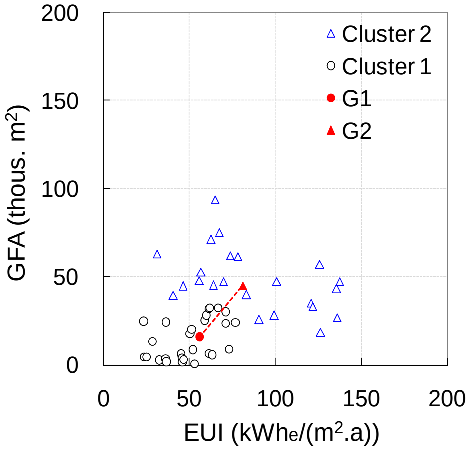
The gravity of government office buildings is usually smaller than that of business office buildings, as shown in Table 2-36, which means that government office buildings are smaller and less energy intense consumers than business office buildings.



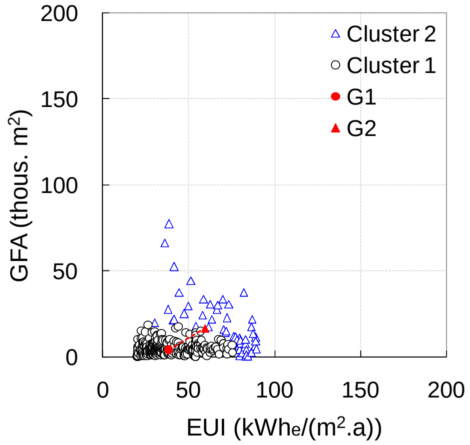
(a-1) City A-84 business office building



(a-2) City A-54 government office building



(b-1) City L-53 business office building



(b-2) City L-498 government office building

Figure 2-133: Cluster Analysis of EUI excl. DH of business and governmental office building in typical cities

Table 2-36: Gravity of clusters in each city or province

City / Indicator			A		B		C		D		E	
			BO	GO	BO	GO	BO	GO	BO	GO	BO	GO
Cluster 1	GFA	thous. m^2	39	27	25	3	10	4	37	3	3	5
	EUI excl. DH	kWhe/(m^2.a)	95.1	65.8	93.6	80.4	68.9	65.3	62.5	33.5	59.6	48.2
Cluster 2	GFA	thous. m^2	122	75	97	31	51	58	118	36	33	25
	EUI excl. DH	kWhe/(m^2.a)	117.2	72.4	149.8	87.5	70.0	61.6	71.4	37.8	59.0	59.3

City / Indicator			I		L		F		G	K	J	M
			BO	GO	BO	GO	BO	GO	O	O	GO	O
Cluster 1	GFA	thous. m^2	20	3	16	4	14	4	7	4	3	7
	EUI excl. DH	kWhe/(m^2.a)	50.9	38.7	56.1	38.5	61.8	52.8	84.0	48.7	35.5	39.0
Cluster 2	GFA	thous. m^2	51	5	45	17	58	14	28	26	13	13
	EUI excl. DH	kWhe/(m^2.a)	78.8	49.3	80.9	59.5	65.3	68.6	121.7	71.0	58.3	69.8

3.2.5 **Conclusions**

This is the first chance to collect such a large amount of office building energy use survey data in several cities in China. Based on those samples, the EUI excl. DH was analyzed in typical cities in China. Take city-A for example, the most well-developed city in China, the range of business office building EUI was from 62.1 to 166.9 kWhe/(m^2.a), with an average of 107.0 kWhe/(m2.a); and from 23.0 to 136.6 kWhe/(m2.a), with an average of 67.6 kWhe/(m^2.a) for government office buildings.

A frequency distribution analysis, as well as polynomial fitting method was conducted. It was found that the EUI of office buildings in China have a unique “Dual Sector” characteristic, which means a large proportion of buildings distributed at the range with smaller EUI while there always existing a small proportion of buildings with a higher EUI level. This feature was definitely different from the U.S and Japan. By analyzing national CBECS survey data, the EUI frequency distribution in the U.S. was found to be right-skewed single-peak. While the EUI distribution in Japan, based on DECC investigation data, fitted normal distribution.
Furthermore, cluster analysis of “GFA” together with “EUI” for office buildings in China was considered. Two clusters in each city or province in China were identified based on the traditional Agglomerative Hierarchical Clustering Method. It was found that the gravity of first-tier cities is usually higher than second-tier cities and business offices are higher than government offices. The slope and distance between two gravities was different from each city or province, reflecting the interregional disparity in China.

Due to the national database of office buildings in China only provides the GFA and annual electricity consumption, it is very difficult to study the occupant behavior's impact through statistical research. However, the statistical research method used in this paper is effective and suitable for application. Two key methods can be adopted in the future analysis or international comparison: Boxplot and key statistical parameter of energy use data. Due to most of the national energy distribution is not normal distribution, thus, average are not suitable for comparison. Five key statistical parameter, including median, 3/4 quartile, 1/4 quartile, maximum and minimum are more objective to depict the distribution feature of the data. Boxplots like Figure 2-128 and Figure 2-129 is clear and comprehensive to illustrate those five parameters.
Frequency distribution analysis. Huge differences are appeared after comparing the frequency distribution of China, Japan and U.S. For understanding the regional or national statistical feature, frequency distribution analysis can be used. Several statistic software has the function of Kolmogorov-Smiromov (K-S) normality test, such as SPSS 13.0, R, etc.

3.2.6 **References**

[1] H. Zhou, N.P. Li, etc. A survey on energy consumption and operation management status of public buildings in Changsha, China. Proceedings – 6th International Symposium on Heating, Ventilation and Air Conditioning, ISHVAC 2009, 58-64.

[2] Ministry of Housing and Urban-Rural Development, Principal of the Implementation on Strengthening Energy-saving Management of Governmental Office and Large-scaled Commercial Building, MOHURD, Beijing, 2007.

[3] Energy Information Administration, Commercial Building Energy Consumption Survey (CBECS), http://www.eia.gov/emeu/cbecs/cbecs2003/.

[4] The Japanese Association for Sustainable Development, Energy Consumption of Commercial Building (DECC), http://www.jsbc.or.jp/decc_download/outline.html#map.

3.3 Experience 2: The U.S. Commercial Buildings Energy Consumption Survey

(Tianzhen Hong, Liping Wang. Lawrence Berkeley National Laboratory, USA)

3.3.1 Introduction

Although no one building type dominates the commercial buildings sector in the US, office buildings are the most common and account for more than 800,000 buildings or 17 percent of total commercial buildings. Offices comprised more than 12 billion square feet of floor space, 17 percent of total commercial floor space, the most of any building type. The aim of the work is to describe the data analysis method employed by the CBECS 2003 that took utility bills (monthly energy use) and broke them down into energy end uses for office buildings.

3.3.2 Database characteristics

CBECS is a national sample survey, developed by the U.S. Energy Information Administration, that collects information on the stock of U.S. commercial buildings, their energy-related building characteristics, and their energy consumption and expenditures. Commercial buildings include all buildings in which at least half of the floor space is used for a purpose that is not residential, industrial, or agricultural, so they include building types that might not traditionally be considered "commercial," such as schools, correctional institutions, and buildings used for religious worship. The building survey covers many topics – building size and use; ownership and occupancy; energy sources, uses, and equipment; energy consumption and cost. The CBECS was first conducted in 1979; the eighth, and most recent survey, was conducted in 2003. The 2012 CBECS interviews will be conducted between April 2013 and September 2013. CBECS is currently conducted on a quadrennial basis. The sample size is historically in the range of 5000 to 7000 buildings across the country which were statistically sampled and then weighted to represent the entire stock of commercial buildings in the U.S. For CBECS 2012, the overall building sample size is increased to 8400 buildings. There are 878 office buildings surveyed in the 2003 CBECS.

3.3.3 Method

The energy end-use consumption tables for 2003 CBECS provide estimates of the amount of electricity, natural gas, fuel oil, and district heat used for ten end uses: space heating, cooling, ventilation, water heating, lighting, cooking, refrigeration, personal computers, office equipment (including servers), and other uses. There are four basic steps in the end-use estimation process:

- Regressions of monthly consumption on degree-days to establish reference temperatures for the engineering models,
- Engineering modeling by end use,
- Cross-sectional regressions to calibrate the engineering estimates and account for additional energy uses, and
- Reconciliation of the end-use estimates to the CBECS total building energy consumption.

Monthly Regression Model

Monthly consumption data were available for 1,518 CBECS buildings for electricity and 1,021 buildings for natural gas in the 2003 building samples. These data allow us to analyze the dependence of consumption on temperature. The degree-day terms were calculated for the periods defined by the actual meter-read dates each month for each case. The results of the analysis determine appropriate degree-day bases for modeling heating and cooling energy use.

For each of the monthly data cases, we fit a model of the form

$$Y_m = \alpha_m + \beta_h h_m(\tau_h) + \beta_c c_m(\tau_c) + \varepsilon_m$$

Where

- Y_m = consumption per day for month m
- τ_h = heating degree-day basis that maximizes the regression's R^2
- τ_c = cooling degree-day basis that maximizes the regression's R^2
- $h_m(\tau_h)$ = heating degree-days per day base τ_h for month m
- $c_m(\tau_c)$ = cooling degree-days per day base τ_c for month m
- ε_m = residual error
- α_m, β_h, β_c = regression coefficients

The regression iteratively searches for the degree-day bases τ_h and τ_c that give the best R^2, and uses the corresponding coefficients α_m, β_h, β_c, from the regression with these bases. These are the maximum likelihood estimates.

Engineering Models for End Use Estimates

The engineering end-use models in the end-use estimation procedure were from ASHRAE, Illuminating Engineering Society of North America (IESNA), and other standard engineering handbooks. Parameter values came from these handbooks and from large-scale field studies of commercial buildings.

a) Space Heating and Cooling

The heating and cooling models estimate the energy consumption of heating systems (primary and secondary) and cooling systems for all energy sources. The models account for building heat loss (or gain) as a function of the building's weighted average conductance and heating (or cooling) degree days. The model accounts for ventilation heat loss (or gain) as a function of the volume of external air brought into a building each day, the temperature difference between the outside air and the inside air, and the heat capacity of air. Starting with CBECS information on the equipment type and estimated percentage of heated or cooled floor space, the model relies on average estimates for equipment efficiency, and on calculations for conduction and ventilation losses (or gains).

To estimate heating and cooling consumption, the engineering sub-models make use of degree-days. Generally, the form of degree-day calculations is as follows:

$$\text{Heating} = \text{Heat Loss Coefficient} * \text{HDD} / \text{Efficiency}_{\text{Heating}}$$

Where

- HDD is a term for heating degree-days

- Heat Loss Coefficient is a term that encompasses ventilation and conductive losses for a building
- $\text{Efficiency}_{\text{Heating}}$ is heating efficiency

For cooling, cooling degree-days are used, along with cooling equipment efficiencies. Also, because outdoor temperature is higher than the set indoor temperature, infiltration and conductive 'losses' result in net gains that increase load.

By incorporating combining the HDD and CDD terms with the heat loss coefficient and separating ventilation losses from conduction. The resulting form is:

$\text{Heating} = (\text{Loss}_{\text{Bldng}} + \text{Loss}_{\text{Vent}}) / \text{Efficiency}_{\text{Heating}}$

Where

- $\text{Loss}_{\text{Bldng}}$ is heat loss from the building due to conductance, including HDD
- $\text{Loss}_{\text{Vent}}$ is heat loss from the ventilation system's intake of external air, including HDD

$\text{Cooling} = (\text{Gain}_{\text{Bldng}} + \text{Gain}_{\text{Vent}}) / \text{Efficiency}_{\text{Cooling}}$

Where

- $\text{Gain}_{\text{Bldng}}$ is heat gain from the building due to conductance, including CDD
- $\text{Gain}_{\text{Vent}}$ is heat gain from the ventilation system's intake of external air, including CDD.

65°F (18.33°C) is a commonly referenced degree-day base. However, buildings may vary in their internal gains. Therefore, rather than using the 65°F (18.33°C) base, the engineering model uses modified degree-day bases, as informed by the monthly regression models.

b) Ventilation

The engineering model for ventilation estimates supply and return fan energy use. The model accounts for differences in static pressure by system type and by building floor space. Typical meteorological year data helped develop estimates of variable air-volume energy factors by climate zone.

The ventilation engineering submodel estimates supply and return fan energy use. At its most basic, the equation for ventilation energy use is as follows:

$$\text{Ventilation} = \frac{1{,}000 * \text{CFMV} * \text{VentHrs} * 365 * \text{WG}}{8{,}520 * \text{VentEff}}$$

Where

- CFMV = total ventilation air volume (ft^3/minute),
- VentHrs = ventilation system operating hours,
- WG = static pressure (inch of water gauge, WG),
- 8,520 = conversion factor (ft^3-in/min-kW),
- VentEff = ventilation efficiency.

The submodel uses this form to develop estimates for supply and return fan energy, for a

VentilationTotal = VentSupply + VentReturn (6)

Where

- VentSupply = supply fan energy, and
- VentReturn = return fan energy.

To estimate total ventilation air volume, the model relies on the external air ventilation volumes discussed in the heating and cooling submodels. It inflates these values for some central and packaged HVAC systems by assuming that the outdoor air volume is 25 percent of the total air flow rate, except for labs where it is 100 percent. For heat pump systems, the model assumes a central ventilation system that circulates fresh air. The model also accounts for differences in static pressure by system type, and by building floor space.

c) Lighting

The lighting model estimates electricity consumption from internal and external lighting for all building types. The model calculates energy use as a factor of average lamp power per floors pace and average annual operating hours. The interior lighting portion relies on information from CBECS on percentage floor space lit by each lamp type, and building operating hours. The model assumes average lamp system efficacy (lumens per watt) for each lamp type, and recommended average illuminance levels by building type.

Lighting = LightingInterior + LightingExterior (7)

The submodel calculates energy use as a factor of average lamp power per floorspace and average annual operating hours. The interior lighting portion relies on information from CBECS on percentage floorspace lit by each lamp type, and building operating hours. External assumptions include average lamp system efficacy (lumens per watt) for each lamp type, and recommended average illuminance levels by building type. The exterior lighting portion assumes a fixed average power density per lamp type, by exterior lighting application: exit signs, exterior architecture, parking, exterior signs, and exterior landscaping. Average annual operating hours by building type are also assumed.

LightingInterior = OpHrs * SqFt * LPT (8)

Where

- OpHrs = annual operating hours (hrs),
- SqFt = building floor space (ft^2),
- LPT = $\sum_{\substack{LampType,i \\ Buildingtype,b}} IL_b * LampP_i \div LampLPW_i$,

Where

- IL_b = recommended lighting illumance levels by building type (lumens),
- $LampP_i$ = percentage of floorspace lit by a lamp type (%),

- $LampLPW_i$ = average system efficacy, accounting for fixture efficiency and lumen degradation over time (lumens per watt).
- LightingExterior = $\sum_{\substack{\text{ExteriorLightingType, i} \\ \text{Building type, b}}} OpHrs_b * Pwr_i$

Where

- $OpHrs_b$ = annual operating hours, by building type (hrs),
- Pwr_i = weighted average wattage per square foot by exterior lighting category (W).

CBECS information on percent lit by each lamp type does not sum to 100% because of overlap in lighting types for given applications. However, to estimate lighting consumption, the submodel renormalizes these percentages to develop average shares of lighting types. To estimate average system efficacies, the model chooses typical lamp (and ballast) systems for each lighting category. In addition, it considers the presence of specular reflectors, which it estimates as improving light output by twenty percent. The presence of electronic ballasts indicated which type of fluorescent lamp was present, T12 or T8.

d) Office Equipment

The office model estimates electricity consumption from office equipment for all building types. The model divides office equipment electricity consumption into four components. One division separates office electricity use into computer equipment versus other office electric loads. Computer equipment includes PCs, monitors and printers. The non-computer-based equipment includes copiers, faxes, cash registers, and servers. The other division separates office electricity used during building on-hours from electricity used during building off-hours.

Office = OfficePCOn + OfficePCOff + OfficeNonPCOn + OfficeNonPCOff;

Where,

- OfficePCOn = energy use of PC’s, printers and monitors during building on-hours
- OfficePCOff = energy use of PC’s, printers and monitors during building off-hours
- OfficeNonPCOn = energy use of servers, faxes, copiers and cash-registers during building off-hours
- OfficeNonPCOff = energy use of servers, faxes, copiers and cash-registers during building off-hours.

e) Water Heating

The water heating model uses system efficiency to convert water heating load to total energy consumed, where the load is the amount of energy needed to heat a given amount of water to a given temperature. Additional energy is used in systems which distribute hot water throughout the building or systems with storage tanks. To account for the variation in energy use by system type, the model uses indicators about equipment type and whether the water is supplied by instant-heating types to determine whether storage and distribution are used.

Water Heating = Load / WHEff = $[(T_{in}\text{-}T_{out}) * GPD * C_w * C_p * Days]$ / WHEff, Where,

- T_{in} = inlet water temperature (°F)
- T_{out} = temperature of delivered water (°F)
- GPD = gallons of water used per day (gallon/day)
- Days = days per year (day)
- C_w = the specific heat of water (btu/lb°F)
- C_p = the density of water (lb/gallon)
- WHEff = system efficiency (%)

f) Refrigeration

The refrigeration submodel calculates electricity consumption from commercial refrigeration. The submodel relies predominantly on end-use intensity estimates, by building type, from CEUS. However, it also incorporates CBECS information on the number of refrigerators.

Refrigeration = x_ElRf * m_MonUse/12 * 1000 * RFUnits * x_RfElPerUnit

Where

- x_ELRf = 0, 1 depending on whether refrigeration is indicated with RFGEQP8,
- m_MonUse/12 = fraction of the year the building is open (%), m_MonUse is based on MonUse8 where present, or defaults to 12 where missing
- 1000 = conversion factor to convert kilowatt-hours to watt-hours
- RFUnit = total number of refrigerator units in building =
- m_RFGCLN + m_RFGOPN + m_RFGRSN + m_RFGVNN + m_RFGWIN,
- and x_RfElPerUnit = x_RfElEUI / x_RFDensityBldng,

Where

- SqFtRf = SqFt8 * x_ElRf,
- RFDensity = $\dfrac{\sum_{\text{CEUS\_BType}} \text{RFUnits}}{\sum_{\text{CEUS\_BType}} \text{SqFtRf}}$.

g) Other

The electricity and district heat models rely on engineering estimates. Since many types of equipment use electricity, CBECS does not explicitly ask if electricity is used for unspecified "other" uses of electricity. Therefore, the engineering model estimates "other" electricity use by applying the CEUS (2005 California Commercial End-Use Survey) intensities for miscellaneous, process equipment, motors, and air compressors to the CBECS floorspace. These estimates were then adjusted for the number of months of building operation per year. Since district heat is primarily used for heating, water heating, cooling, and cooking end-uses, which were explicitly modeled, and given the relatively small number of cases and lack of information, the district model does not calculate "other" consumption. For fuel oil and natural gas, the model for other energy use is based on regression estimates.

Cross-sectional Regressions

Cross-sectional regression models were used to calibrate the natural gas and fuel oil engineering estimates. The cross-sectional regression models for natural gas were fit with consumption per square foot as the dependent variable and the independent variables were defined on a corresponding scale. Besides the engineering estimates, independent variables included dummy variables for the presence of a laundry, cleaners, or central plant, and to indicate natural gas use for manufacturing or electricity generation. Additional dummy variables were defined to indicate whether natural gas or some other fuel was used as a secondary heating source in the building.

The cross-sectional fuel oil regression models were fit with CBECS consumption per square foot as the dependent variable. The independent variables, defined on a corresponding scale, included the engineering estimates and dummy variables for the presence of a central plant or the use of fuel oil for manufacturing or electricity generation. Additional dummy variables were defined to indicate whether fuel oil or some other energy source was used as a secondary heating source in the building.

Final Reconciliation

For electricity, reconciliation with the total consumption took two steps. First, the monthly model results were used to provide approximate estimates of annual heating and cooling use. For each case with monthly regression estimates, the ratio of that heating or cooling estimate to the corresponding preliminary engineering estimate was calculated. The median ratios were then reviewed by building size, activity, and age, as well as by climate zone. Since the results showed a definite variation by climate zone, the median ratios were used to adjust the engineering estimates for electric primary heat and electric cooling for all cases. Second, the adjusted engineering estimates were prorated to match the CBECS estimate of total building electricity consumption.

For natural gas and fuel oil, the adjusted engineering estimates were prorated to match the total building consumption. For district heat, the engineering estimates were prorated to match the total building consumption.

3.3.4 **Results**

Figure 2-134 shows the geographical areas as defined by the U.S. Bureau of Census including the four Census Regions and nine Census Divisions. Office building samples distributions by floor area, vintage and location are shown in Figures 2-135 to 2-137 respectively. Figure 2-138 lists the site energy use intensities (EUI, in kWh/m²) for office buildings in nine census divisions. The average site EUI is 292.6 kWh/m² for the office buildings in the 2003 CBECS, with HVAC (Space Heating + Cooling + Ventilation) consuming 50.5% followed by lighting 24.9%. The single largest end use is space heating which consumes 35.3% of total site energy.

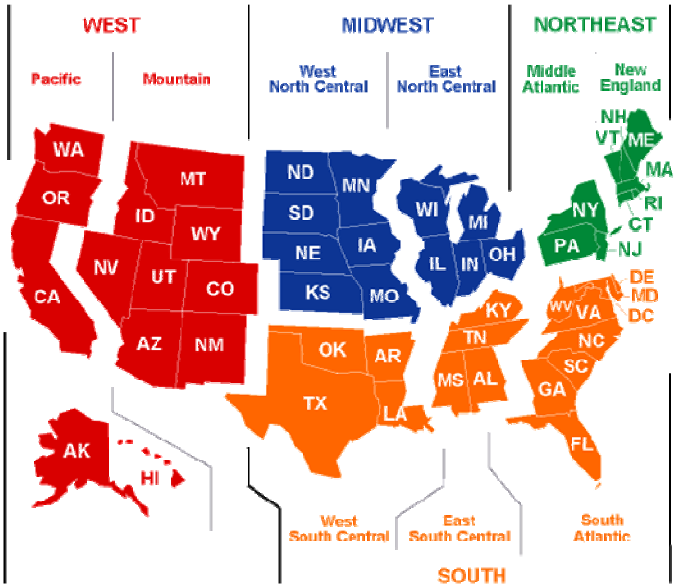


Figure 2-134: CBECS Census Regions

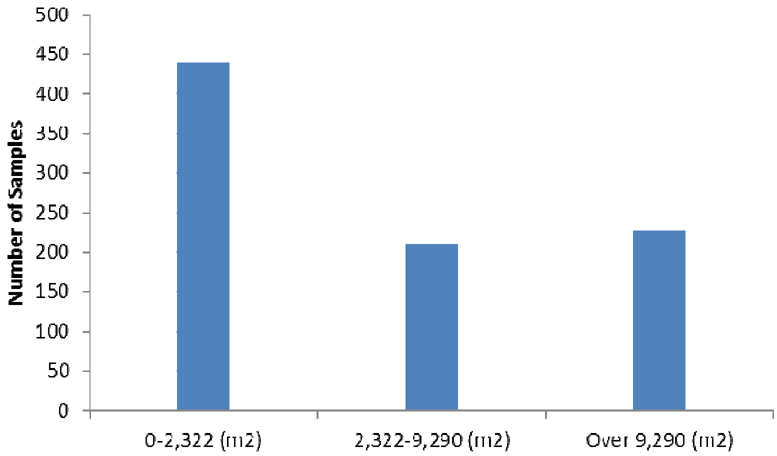


Figure 2-135: Distribution of office building samples by floor area

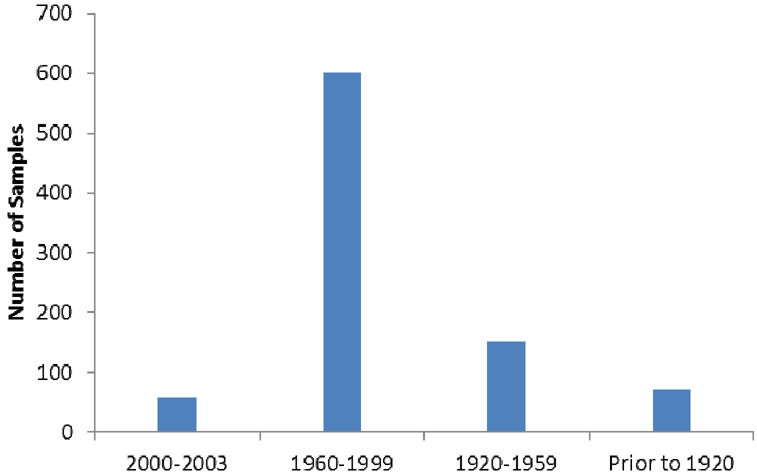


Figure 2-136: Distribution of office building samples by vintage

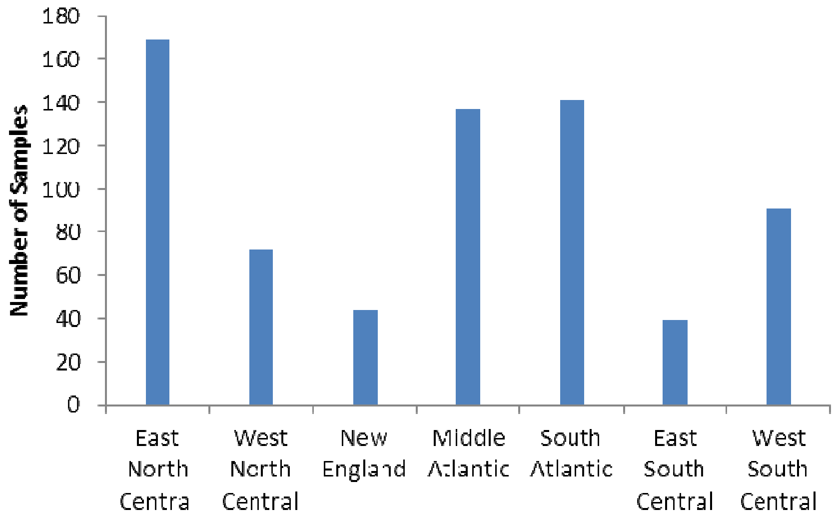


Figure 2-137: Distribution of office building samples by location

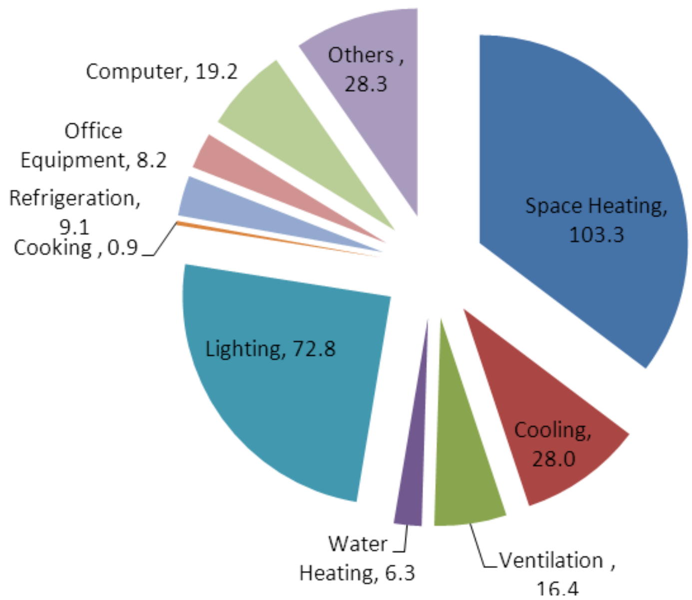


Figure 2-138: Average site energy use intensities (kWh/m^2) for office buildings in nine census divisions

3.3.5 Conclusions

This document describes the data analysis method used by CBECS 2003 for end use estimates for office buildings in the U.S. Statistical regressions and engineering modeling approaches were used to calculate end uses based on monthly consumption data for electricity and natural gas, and collected building system characteristics in the survey. The top four major end uses in office buildings are space heating, lighting, space cooling, and plug loads (office equipment + computers).

3.3.6 **References**

[1] Energy Information Administration, prepared by KEMA, Energy End-Use Consumption Estimates for Commercial Buildings, Sub-Task 3.2, Documentation of 2003 CBECS EUC Estimation, Final Report, 2008

[2] Energy Information Administration, http://www.eia.gov/emeu/cbecs/tech_end_use.html

[3] California Commercial End-Use Survey, http://www.energy.ca.gov/ceus/

3.4 Experience 3: Energy saving potentialities by retrofitting the European residential sector

(C. Becchio, S. P. Corgnati, I. Ballarini and V. Corrado, Politecnico di Torino, Italy)

Energy saving

The national *building typologies* can be used as data sources for forecasting and evaluating the energy saving potential and the carbon dioxide emission reductions for each European country. Thereby the main objective of the IEE TABULA project has been to create a harmonized structure of the European building typologies and to identify representative building types. This purpose has come from the need to assess the energy consumption of the national building stock and consequently to predict the impact of different energy efficiency measures in order to select effective retrofit strategies on the existing buildings. Two levels of building retrofit have been considered: a *standard refurbishment*, applying measures which are commonly used in the country; an *advanced refurbishment*, applying measures which reflect the best available technologies. The evaluation of each reference building type has been performed in each country by using the national EPBD asset rating method and by showing the energy performance before and after the refurbishment.

Additional statistical information about the frequency of constructions and of heating systems types has made possible the use of the reference building types as models for the assessment of the energy performance of the whole national building stock.

The present paper reports the first outcomes of the application of the above described methodology to the national residential building stocks of four countries representative of the North, Middle, South and East European Countries. It summarizes the results presented in the TABULA report "*Application of Building Typologies for Modelling the Energy Balance of the Residential Building Stock*".

3.4.1 Introduction

TABULA (Typology Approach for Building Stock Energy Assessment) [1] was a project within the European program "Intelligent Energy Europe" (IEE) with the participation of thirteen European countries (Germany, Greece, Slovenia, Italy, France, Ireland, Belgium, Poland, Austria, Bulgaria, Sweden, Czech Republic and Denmark). The project objective has been to create a harmonized structure of the European building typologies [2]. Each participant developed a building typology classification that allowed to divide national existing buildings in categories: for each category, a building type was identified as representative of a defined climatic region, period of construction, building size, etc. In many European countries, the classification of building types is a concept already used at national and/or regional level. However, both at national and at European level, a number of problems rise up due to lack of shared definitions, to unknown or not updated data about existing buildings, to the difficulties in defining a common concept of building typology. In practice, it is impossible to compare the types of buildings among European countries without uniform and shared definitions. As a consequence, TABULA firstly aimed to create a harmonized structure to classify building types in Europe: the project focused on residential buildings, but a possible extension to other uses is also possible.

Building typologies developed during the TABULA project can be exploited as a basis for analysing the national housing sector. In fact, a crucial goal of the project has been to estimate the energy consumption of residential building stocks at national level and, consequently, to predict the *potential impact of energy efficiency measures* (addressed to building envelope and space heating and DHW systems) in order to select effective strategies for upgrading existing buildings. In particular, during the TABULA project six of the European partners (Belgium, Czech Republic, Denmark, Germany, Greece, Italy) carried out model calculations aimed to image the energy consumption and estimate the energy saving potentials of their national residential building stocks (Energy Balance Method).

Specifically, as shown in Figure 2-139, starting from global statistics at national and regional level and from the corresponding available residential building samples divided in classes, some reference building types have been selected in order to obtain a relevant characterization of the analyzed buildings. They have been chosen as representative of a large portion of the national residential building stock. Different modelling approaches were chosen by the partners depending on the available statistical data. Some defined a set of synthetic buildings reflecting building stock averages; others applied a set of generic example buildings from the national TABULA typologies.

The methodology provided by the European standards supporting the Energy Performance of Buildings Directive (EPBD, 2002/91/EC) has been applied for the evaluation of the energy demand of the selected building types and to assess the energy saving potential due to energy retrofit actions. In fact, for each reference building type two refurbishment measures have been considered: a *standard refurbishment* through the application of measures commonly applied within the country; an *advanced refurbishment* through the introduction of measures that reflect the use of the best available technologies. Finally additional information about the number and the frequency of each specific building type has made possible the application of statistical models in order to estimate the overall energy performance, energy saving potentialities, carbon dioxide emissions reductions of the building stock at national/regional level.

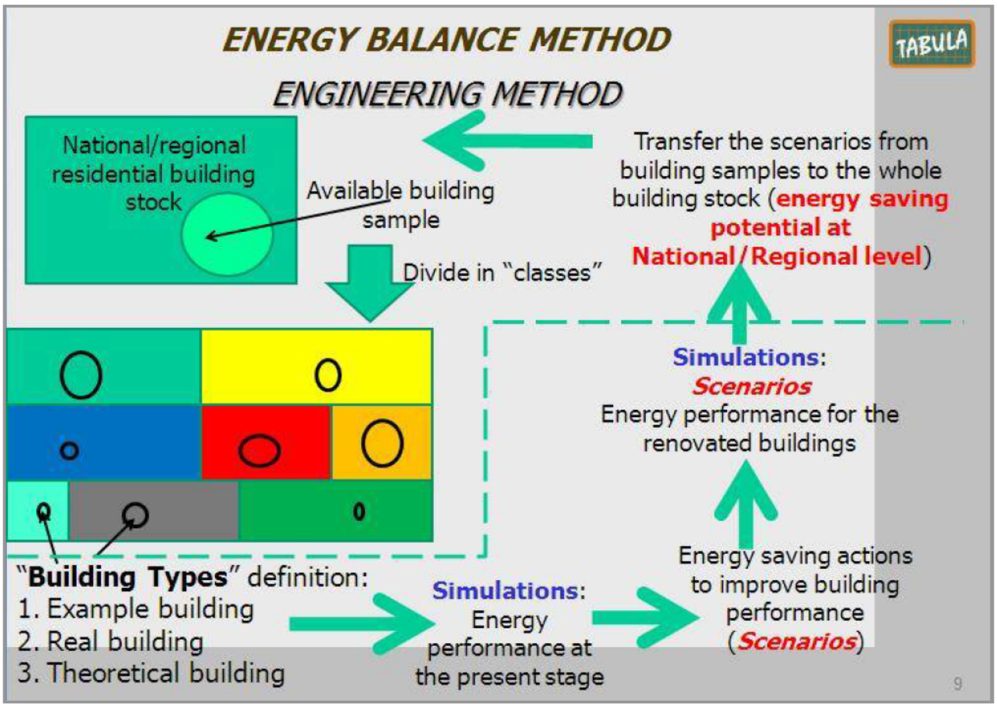


Figure 2-139: Procedure for Energy Balance Method used in the TABULA project to predict the potential impact of energy efficiency measures on national housing sector.

This contribute shows the first outcomes of the application of the above described Energy Balance Method at the national residential building stock of four countries:

- Denmark, as a representative of the North European countries;
- Germany, as a representative of the Middle European countries;
- Italy, as a representative of the South European countries;
- Czech Republic, as a representative of the East European countries.

The data presented in this paper have been extrapolated from the TABULA report "Application of Building Typologies for Modelling the Energy Balance of the Residential Building Stock" [3] and from the "National Scientific Report" on the TABULA project of the four analysed countries [4-7].

3.4.2 **Denmark**

The energy balance of the Danish residential buildings was calculated using synthetical average buildings. These were split within nine different construction periods and three building types (single family houses SFH, terraced houses TH, block of flats AB).
In order to estimate energy saving potentials the national Energy Balance method was used.
Refurbishment measures were applied only to the envelope and consisted in two different levels of thermal insulation: the standard refurbishment is associated with a high thickness of insulating material (300 mm for the ceiling, more than 100 mm for the wall), while the advanced refurbishment is associated with a higher thickness of insulating material (400 mm for the ceiling, more than 200 mm for the wall). Consequently, the energy saving potential was calculated only in term of net energy demand for heating and DHW. The results of the analysis are presented in term of energy saving and CO_2 emission reduction in Table 2-37.

Table 2-37: Annual energy saving potentialities (in terms of net energy demand for space heating and DHW) and CO_2 emissions reductions by standard and advanced refurbishment for Danish residential building stock.

Reference building type	Original State		Standard Refurbishment			Advanced Refurbishment		
	$Q_{H,W,p}$	t_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}
	[10^3GWh]	[10^6t]	[10^3GWh]	[-]	[10^6t]	[10^3GWh]	[-]	[10^6t]
SFH and TH	31.5	---	14.6	-46%	---	15.6	-50%	---
AB	12.1	---	5.3	-44%	---	5.9	-49%	---
	43.6	---	19.9	-46%	3.1	21.5	-49%	3.4

3.4.3 **Germany**

The analysis of the German building stock was conducted on a set of six synthetical average buildings. Two building size classes (single family houses with one or two dwellings and multifamily houses with three or more dwellings) and three construction periods according to different levels of energy saving national regulations were considered (Table 2-38).

Table 2-38: Classification of the German building stock.

Reference building type		Construction period	Number of buildings
Single Family House	SFH I	until 1978	9'610'000
Single Family House	SFH II	1979 – 1994	2'710'000
Single Family House	SFH III	1995 – 2009	2'670'000
Multi-Family House	MFH I	until 1978	2'340'000
Multi-Family House	MFH II	1979 – 1994	440'000
Multi-Family House	MFH III	1995 – 2009	270'000
			18'040'000

The energy balance model was developed on basis of the available statistical input data. The energy demand for space heating of the considered six building types was calculated according to a seasonal energy balance approach. In this way an estimation of energy saving potentials in the German building stock for heating and hot water supply was carried out.

The refurbishment measures consisted in the application of insulation material on walls, floors and roofs and in the replacement of windows. The standard refurbishment is characterized by U-values of 0.24 W/(m^{2}K) for walls, roofs and upper floor ceilings, U-values of 0.3 W/(m^{2}K) for ground floors and cellar ceilings and U-values of 1.3 W/(m^{2}K) for windows. The advanced refurbishment is characterized by U-values of 0.16 W/(m^{2}K) for walls, U-values of 0.14 W/(m^{2}K) for roofs and upper floor ceilings, U-values of 0.20 W/(m^{2}K) for ground floors and cellar ceilings and U-values of 0.80 W/(m^{2}K) for windows. With reference to the retrofit of the space heating and DHW systems, at the standard level it was considered to replace the heat generator, while at the advanced level the measures consisted in the improvement of efficiency of the distribution and generation subsystem, in the application of an heat recovery ventilation system and in the installation of a solar thermal plant.
Energy saving potential obtained by retrofitting the German residential building stock is reported in Table 2-39.

Table 2-39: Annual energy saving potentialities (in terms of primary energy for space heating and DHW) and CO_2 emissions reductions by standard and advanced refurbishment for German residential building stock.

Original State		Standard Refurbishment			Advanced Refurbishment		
$Q_{H,W,p}$	t_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}
[10^{3}GWh]	[10^{6}t]	[10^{3}GWh]	[-]	[10^{6}t]	[10^{3}GWh]	[-]	[10^{6}t]
661	136	304	-46%	63	512	-77%	100

3.4.4 **Italy**

In Italy, six reference building types were created to represent the housing stock for the purpose of Energy Balance analysis, as shown in Table 2-40.

Table 2-40: Classification of the Italian building stock.

Reference building type	Construction period	Number of buildings

Single Family House	SFH.01	until 1900	1'046'278
Single Family House	SFH.03	1921 – 1945	559'336
Multi-Family House	MFH.04	1946 – 1960	707'563
Apartment Block	AB.05	1961 – 1975	869'056
Apartment Block	AB.06	1976 – 1990	1'214'773
Apartment Block	AB.07	1991 – 2005	358'765
			4'755'771

These reference buildings were chosen according to statistical analysis: they are representative of a suitable significant portion of the entire national building stock considering both the construction age and the building size (i.e. number of apartments, floor area) and they belong to the "Middle Climatic Zone" (from 2100 to 3000 heating degree days), which is the most representative of the Italian climate (about 4250 municipalities on a total number of 8100). Specifically, the first two reference buildings (single family houses) are "Theoretical Buildings", chosen on the basis of statistical data (Piedmont Regional Database of Building Energy Performance Certificates). The other reference buildings (multi-family house and three apartment blocks) are "Example Buildings", i.e. real buildings defined typical according to the experience.

The official national calculation method (Technical Specification UNI/TS 11300 - National Annex to CEN Standards) for energy certificates was applied for the evaluation of the energy demand of the selected reference buildings and to assess the energy saving potential due to energy retrofit actions according to two different scenarios (standard and advanced refurbishment). In regard to the envelope, the refurbishment measures consisted in the application of insulation material on walls, floors and roofs and in the replacement of windows. The considered U-values correspond to the requirements established by the new regulations on energy performance of buildings in Piedmont Region (D.G.R. n. 46-11968), that belongs to the "Middle Climatic Zone". The U-values applied for the standard refurbishment are the U-value limits set by the Piedmont Region regulation (0.33 W/(m^{2}K) for walls, 0.30 W/(m^{2}K) for roofs, ceilings and floors, and 2 W/(m^{2}K) for windows), while the U-values applied for the advanced refurbishment are the optional U-value targets set by the Piedmont Regional regulation (0.25 W/(m^{2}K) for walls, 0.23 W/(m^{2}K) for roofs, ceilings and floors, and 1.7 W/(m^{2}K) for windows).
With reference to the refurbishment of the space heating and DHW systems, some measures were considered in order to improve the efficiency of emission, distribution and generator subsystems and to exploit renewable energies with the installation of a thermal solar plant (advanced refurbishment).
Energy saving potentialities obtained applying the mentioned retrofit measures at the Italian residential building stock are reported in Table 2-41.

Table 2-41: Annual energy saving potentialities (in terms of primary energy for space heating and DHW) and CO_2 emissions reductions by standard and advanced refurbishment for Italian residential building stock.

Reference building type	**Original State**		**Standard Refurbishment**			**Advanced Refurbishment**		
	$Q_{H,W,p}$	t_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}
	[10^3GWh]	[10^6t]	[10^3GWh]	[-]	[10^6t]	[10^3GWh]	[-]	[10^6t]

SFH.01	50.6	10.3	38.8	-77%	7.9	42.8	-85%	8.7
SFH.03	22.1	4.5	17.8	-81%	3.6	19.4	-88%	3.9
MFH.04	127.2	25.8	98.2	-77%	19.9	105.5	-83%	21.4
AB.05	419.5	85.2	301.2	-72%	61.2	349.9	-83%	71
AB.06	364.3	74	204.4	-56%	41.5	255.4	-70%	51.9
AB.07	76.6	15.6	32	-42%	6.5	42.3	-65%	8.6
	1060.5	215.3	692.5	-65%	140.6	815.4	-77%	165.5

3.4.5 **Czech Republic**

Six reference building types were created to represent the Czech Republic housing stock for the purpose of Energy Balance analysis. This set of buildings was categorized by size and age as shown in Table 2-42.

Table 2-42: Classification of the Czech Republic building stock.

Reference building type		Construction period	Number of buildings
Single Family House	SFH.1	until 1979	1'649'756
Single Family House	SFH.2	1980 – 2001	424'172
Single Family House	SFH.3	2002 – 2010	139'293
Multi-Family House and Apartment Block	APT.1	until 1979	1'277'705
Multi-Family House and Apartment Block	APT.2	1980 – 2001	574'438
Multi-Family House and Apartment Block	APT.3	2002 – 2010	165'648
			4'159'902

The buildings are theoretical buildings based on the analysis of available statistical data and on the knowledge of historical standard requirements for the U-values of the building envelope and the usual efficiency of the heating and DHW systems.

The energy balance model was created on basis of the statistical data. The delivered energy and the energy demand for space heating of the considered six groups of buildings was calculated using national calculation method.

In this case the refurbishment measures were fixed on the basis of recent studies. In fact, it was estimated by experts that by achieving U-values prescribed by the latest version of the Czech standard CSN 730540 following amount of energy can be saved:

- 20% of energy in average can be saved by applying ETICS (External Thermal Insulation Composite Systems) to the exterior walls;
- 10% of energy in average can be saved by roof insulation;
- 25% of energy in average can be saved by windows replacement;
- heating control systems would bring savings ranging approximately between 5 and 15%;
- the losses can be reduced up to 50% by insulating properly the pipes.

The above mentioned percentages were considered in the calculation energy balance model and distributed over the categories of buildings. The results are shown in Table 2-43.

Table 2-43: Annual energy saving potentialities (in terms of primary energy for space heating and DHW) and CO_2 emissions reductions by standard and advanced refurbishment for Czech Republic residential building stock.

Reference building type	**Original State**		**Refurbishment**		
	$Q_{H,W,p}$	t_{CO2}	$\Delta Q_{H,W,p}$	Δ% savings	Δt_{CO2}
	[10^3GWh]	[10^6t]	[10^3GWh]	[-]	[10^6t]
SFH.1	11.9	5.5	7.7	-65%	3.6
SFH.2	12.7	5.9	4.8	-38%	2.2
SFH.3	5.5	2.6	1.1	-20%	0.5
APT.1	6.1	2.9	3.2	-52%	1.5
APT.2	15.2	6.5	5.3	-35%	2.3
APT.3	5.4	2.6	1	-19%	0.5
	56.8	26	23.1	-41%	10.6

3.4.6 **Conclusion**

The analysis shows that building typologies can be a helpful tool for modelling the energy consumption of national building stocks and for carrying out scenario analyses beyond the TABULA project. The consideration of a set of representative buildings, which reflect the current state of the building national stock, makes it possible to have a detailed view on various packages of refurbishment measures for the complete buildings stock or for its sub-categories. The effects of different insulation measures at the respective construction elements as well as different system supply measures including renewable energies can be considered in detail with fast analysis.

As general rule, when two different level of retrofit were considered it is noted that the standard refurbishment is associated with high relative percentage of energy saving (Figure 2-140): the energy saving due to a standard refurbishment is bigger than the saving variation between a standard refurbishment and an advanced refurbishment. In fact, national building stock is often characterized by low energy performance and even the application of basic energy renovations may provide significant increases in energy performance and consequent reduction of CO_2 emission (the case of Italy is exemplificative of this trend). Thereby from an economic point of view it is more convenient to apply standard refurbishment measures at the national building stock than advanced ones that are the most expensive.
It was highlighted that, even with standard refurbishments, energy saving over 45% can be achieved. As a consequence of this big saving potential, suitable policies to address energy retrofit actions of existing buildings are crucial.
Finally, the quality of future model calculations will depend very much on the availability of statistical data. For reliable scenario analyses, information about the current state of the building stock and about the current trends is needed. The availability and regular update of the relevant statistical data will be an important basis for the development of energy strategies in the building sector.

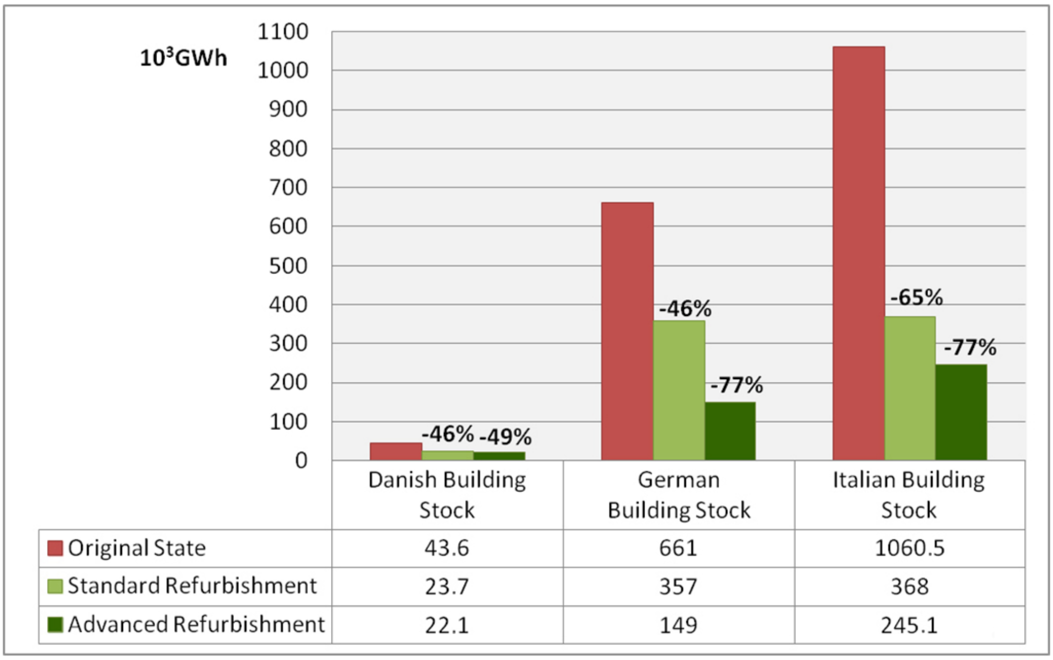


Figure 2-140: Comparison between annual energy saving potential by applying a standard refurbishment and an advanced one to the Danish, German and Italian building stock.

3.4.7 References

[1] TABULA Project, www.building-typology.eu

[2] TABULA Project Team, Use of Building Typologies for Energy Performance Assessment of National Building Stocks. Existent Experiences in European Countries and Common Approach, Synthesis Report SR1, June 2010

[3] TABULA Project Team, Application of Building Typologies for Modelling the Energy Balance of the Residential Building Stock, Thematic Report N° 2, February 2012

[4] Corrado V., Ballarini I., Corgnati S., *National scientific report on the TABULA activities in Italy*, Politecnico di Torino, Torino, May 2012

[5] Loga T., Diefenbach N., Stein B., Born R., *TABULA - Scientific Report. Germany Further Development of the National Residential Building Typology*, Institut Wohnen und Umwelt GmbH, Darmstadt, October 2012

[6] Wittchen K. B., Kragh J., *Danish building typologies. Participation in the TABULA project*, Danish Building Research Institute, Aalborg University, Hørsholm, 2012

[7] VV. AA., National Scientific Report. Czech Republic, STÚ-K, a.s., Prague, May 2012

3.5 Experience 4: National/Regional investigation level, Single & Multifamily houses in Italy

(N. Talà, TEBE Research Group, Department of Energy, Politecnico di Torino, Italy)

3.5.1 Introduction

The IEE-Project TABULA is aimed to create a harmonized structure for European Building Typologies, focusing on residential buildings: the topic of the research is how to collect, elaborate and analyze data characterizing national building stock in order to define "typical" buildings able to express a Building Typology. Different strategies with different level of information details can be adopted for "typical" building definition.
In this work, the different approaches for defining the "building typologies" are tested. In particular, three methods are explained to show the developed benchmark models: the first method identifies building types based on several assumptions deduced by an expert without statistical data; the second method processes empirical data to pick out real buildings that are representative of the stock; finally, the third method provides a building that is the most probable of a group of buildings.
Then, these approaches are applied to some Italian case studies: example building characteristics, statistical analysis on residential building database, Italian building typologies structures.

Type of building	Residential buildings
Dimension	80÷400 m^2
Location	Piedmont, Italy
Thermal characteristics	Variable
Type of observed spaces	Whole building
Year of construction	Seven building age classes (1900÷2005)
No. of floors	Variable (2÷10)
Windows, orientation	Either N, E, S, W
Window opening	Variable
Shading devices	Variable
Sources of heat gains	-
Activity, sex and age of occupants	Variable
Origin of occupants	Variable

3.5.2 Aim of the work

The project objective is to create a harmonized structure on the building types in Europe. Each participant develops a "building types" classification at national level: each identified "building type" is representative of a defined period, size, etc.
Another important outcome of the project is the development of an interactive web tool where the "building types" classification can be used with different objectives in the building energy sector: advice for energy retrofitting, energy performance assessment of building stocks, comparison of energy performance among buildings and building stocks. In particular the web tool contains a data structure of "building-types", characterized by dimensions, shape factors, thermo-physical properties

(e.g. thermal transmittance of building components), efficiency of heating systems and other energy indicators.
A crucial goal of the project is to estimate the energy consumption of residential building stocks at national level and, consequently, to predict the potential impact of energy efficiency measures in order to select effective strategies for upgrading existing buildings. To this aim, it is fundamental the application of a methodology for the definition of “building types”, which allows the classification of existing buildings in categories (“buildings-types”) to be analyzed and investigated.

3.5.3 **Database characteristics**

Number of buildings	7104	
Period of measurement	Not applicable	-
Duration (days)	-	-
Number of observed spaces	-	-
Number of observed spaces with window sensors	-	-
	Items	Interval
IF1. Climate	Heating Degree Days	-
IF2. Building envelope	U-value, Window to wall ratio	-
IF3. Building service & Systems	Type of space heating system: space heating – centralized/decentralized	-
IF4. Operation & Maintenance	-	-
IF5. Indoor environmental quality	-	-
IF6. Occupants’ activities and behavior	-	-
IF7. Social and economical aspects	-	-

The database contains records for more than 66.000 houses rated across Piedmont. The 66.000 house records represent the result of the information collected by EP certification schemes.
The database contains information on physical characteristics and calculated energy requirements of each house. Each submission includes more than 40 information fields.
The data includes:

- location;
- construction period;
- form;
- heated gross volume;
- net floor area;
- window average thermal transmittance;
- calculated energy demands and indicators.

The purpose of the EPCs database is also to gather the individual energy analyses data. Once an energy advisor successfully completes the energy assessment of a house, the resulting energy analysis data is collected and stored into the database.
In order to validate the quality of data and to simplify the analysis the amount of data is restricted to only 7104 certificate schemes. In particular, apartment blocks, multi-family houses, terraced houses

and single-family houses have been considered. Such data are conveniently illustrated by means of the pie charts in Figure 2-141 and Figure 2-142.

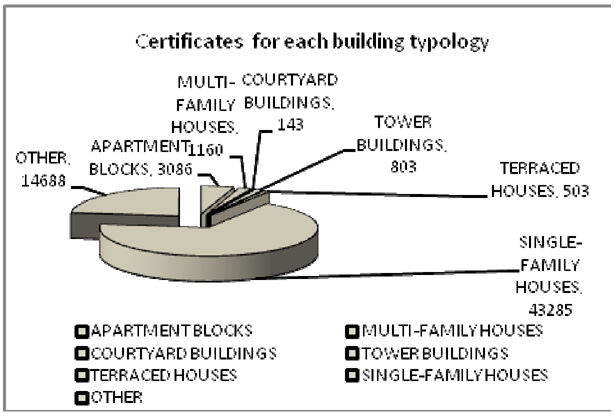


Figure 2-141: Split of Energy Performance Certificates for each building typology (66063 certificates).

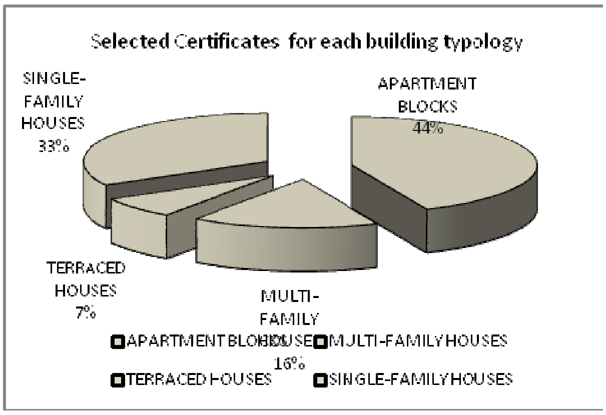


Figure 2-142: Split of the selected Energy Performance Certificates for each building typology (7104 certificates).

3.5.4 Methods applied for the data analysis

In order to define a typical house useful for describing the thermal and geometric characteristics of a group of houses, the first step consists of identifying independent variables influencing the multitude of parameters that are specific to the building.

The TABULA project has fixed three independent variables which are: location, age and form. In the specific Italian case, the three-dimensional space that generate appropriate reference building includes 3 climatic zones, 7 ages and 4 forms of Italian housing (single-family homes, multi-family homes, terraced house, apartment block), the combinatorial process produces 84 building typologies.

First Approach

According to the first approach the definition of the representative building (Example Building) is based on construction period, size and region by an expert using rule-of-thumb to compensate the lack of statistical information.

Second Approach

The second method identifies the typical building (Real Building) employing a statistical analysis. Collected data are statistically elaborated to pick out the real building with geometrical and thermo-physical characteristics similar to the average of the building sample.

The Piedmont Regional Database of Energy Performance Certificates has been used to define the building typologies within the categories single family homes and terraced houses.

The following steps outline the method adopted.

Based on the available data, representative parameters of geometric and thermal features have been selected. These parameters are: volume, net floor area, envelope area to volume ratio, number of levels, number of dwellings, opaque envelope average thermal transmittance, window average thermal transmittance.

For each parameter the number of data were sufficient (at least 10 observations) to calculate statistical functions such as mean, median, 25th percentile, 75th percentile.
Interquartile ranges (IQRs) are evaluated for all the parameters. This step allows to identify, for each parameter, the 50% of the buildings close to the median value. The intersection of all IQRs permits to select the single real building whose parameters are the closest to the median values.

If this procedure gives more than one or no real building, IQRs can be tuned by means of suitable criteria in order to pick out only one real building.
The available data of the real building identified by such procedure are not sufficient to perform energy analyses. Additional parameters have to be specified according to experience or statistical analysis.

Third Approach
The third method identifies the typical building (Theoretical Building) as an archetype, that is ‘‘a statistical composite of the features found within a category of buildings in the stock’’ (ECBCS,2004).
The main steps for developing an archetype can be summarized as follows:

Identification of primary independent variables (Xi) for describing the parameters (Pj) of a specific house (Bk) in the stock. For example the following parameters (Pj) that characterize the building can be considered: building external shape, internal layout, window to wall ratio, thermal insulation…
On the other hand, among the independent variables (Xi) it is possible to consider: floor area, construction year, location, main heating source…
Determination of the trend of each parameter based on independent variables by means of several engineering hypotheses, analysis or rule of thumb. This permits to Figure out which are the most significant independent variables xi for each parameter.
Determination of the analytical relation between the jth parameter and its significant independent variables using statistical analysis (e.g. regression techniques).
An example showing the application of this approach for the window average thermal transmittance (WTT).

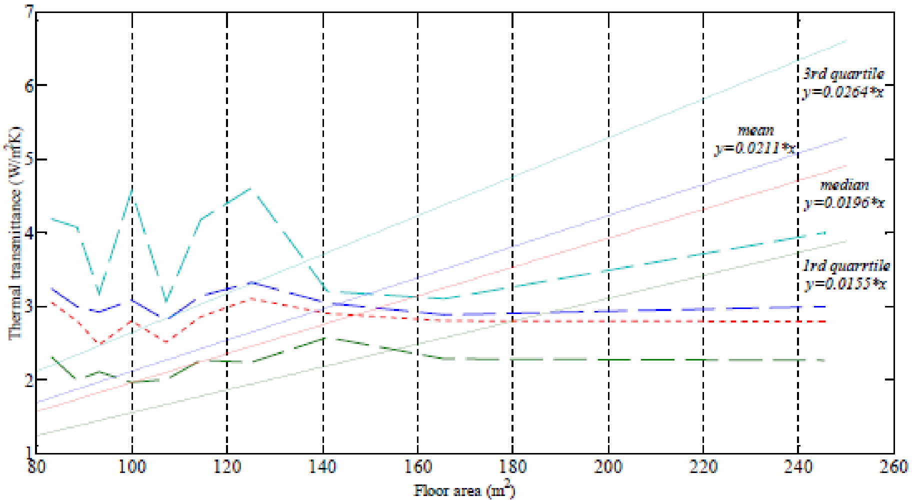


Figure 2-143: Thermal transmittance analysis.

As result of step 2, this specific parameter presents a significant relation with the floor area. A sampling of 339 buildings was selected from the Piedmont Regional Database of Energy Performance Certificates. The WTT values associated with the 1st quartile, mean, median and 3rd quartile are evaluated for each of the 10 groups defined by the deciles of the population in terms of floor area. In this particular case, each group contains approximately 34 buildings. Figure 2-143 presents the four dispersion values for each central value of these 10 groups. Regression curves and their analytical expression (STEP3) for the 1st quartile, mean, median and 3rd quartile are reported.

3.5.5 **Main results and discussion**

The prediction of energy consumption of residential building stock and the measure of energy savings due to efficient strategies at national level are very topical research items. To this aim, it is important to define a methodology for the generation of "building types", representative of the categories to be investigated. It is fundamental to establish the rules to develop the building typology in order to compare the most suitable configurations and scenarios for the implementation of efficiency measures. The three methods introduced for defining the "building types" permits to face, in practice, the problem of the definition of "building typologies" when different level of information details are available.

3.5.6 References

[1] Anil Parekh, M.A Sc., P.. Development of archetypes of building characteristics libraries for simplified energy use evaluation of houses, Proceedings of the 9th International IBPSA Conference,August 2005.

[2] ECBCS, 2004. Stock aggregation,in Annex31-Energy related environmental impact of buildings, IEAECBCS (International Energy Agency-Energy Conservation in Buildings and Community Systems).

[3] Ente Nazionale Italiano di Unificazione. UNI/TS 11300-1, Energy performance of buildings – Part 1:Evaluation of energy need for space heating and cooling, May 2008.

[4] Intelligent Energy Europe (IEE). Typology Approach for Building Stock Energy Assessment (TABULA), in Description of the Action, Annex I, SI2.528393, April 2009.

[5] Simon Sansregret, Jocelyn Millette. Development of a functionality generating simulations of commercial and institutional buildings having representative characteristics of a real estate stock in Québec (Canada), Proceedings of the 11th International IBPSA Conference, July 2009.

Published by Institute for Building Environment and Energy Conservation, Zenkyoren building Kojimachikan, 3-5-1, Kojimachi Chiyoda-ku, Tokyo 102-0083 Japan

ISBN: 978-4-9907425-6-0

Participating countries in EBC:
Australia, Austria, Belgium, Canada, P.R. China, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Republic of Korea, the Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States of America.

Additional copies of this report may be obtained from:

EBC Bookshop
C/o AECOM Ltd
Colmore Plaza
Colmore Circus Queensway
Birmingham B4 6AT
United Kingdom
Web: www.iea-ebc.org
Email: essu@iea-ebc.org