

**ENERFICIENCY: User Led Energy Efficiency Management**



**State of the Art**

|  |  |
| --- | --- |
| ***Project Identifier*** | ENERFICIENCY |
| ***Project Title*** | **User Led Energy Efficiency Management** |
|  |  |
| ***Document Version*** | 1 |
| ***Planned Delivery Date*** | 05/03/2014 |
| ***Actual Delivery Date*** | 05/03/2014 |
| ***Document Title*** | ENERFICIENCY State of the Art |
| ***Work Package*** | 1 |
| ***Document Type*** | MS Word |
|  |  |
| ***Abstract*** | This document defines the State of the Art in the computation based building energy efficiency domain for the Enerficiency project. The text originates from the Enerficiency Business and Technical Requirements Document defined within the WP1. |
| ***Keywords*** | NA |

***Versions***

|  |  |  |
| --- | --- | --- |
| ***Version*** | ***Date*** | ***Changes*** |
| Release 1 | 05/03/2014 | Initial |

**TABLE OF CONTENTS**

[State of the Art Definition 1](#_Toc381781667)

[1 State of the Art 4](#_Toc381781668)

[1.1 Context & Big picture 4](#_Toc381781669)

[1.2 Hybrid models or “Grey-Box” approaches 8](#_Toc381781670)

[1.3 Conclusion 10](#_Toc381781671)

[2 Annex 1: Complements on the state of the art 12](#_Toc381781672)

[2.1 Physical models: Building thermal behaviour modelling 12](#_Toc381781673)

[*2.1.1* *Conservation Law* 12](#_Toc381781674)

[*2.1.2* *From Equations to Software* 12](#_Toc381781675)

[*2.1.3* *The CFD Approach* 13](#_Toc381781676)

[*2.1.4* *The Zonal Approach* 14](#_Toc381781677)

[*2.1.5* *The Multizone or Nodal Approach* 15](#_Toc381781678)

[*2.1.6* *Discussion on the Physical Models* 19](#_Toc381781679)

[2.2 Statistical Methods based on Machine Learning 20](#_Toc381781680)

[*2.2.1* *Multiple linear regression or conditional demand analysis (CDA)* 20](#_Toc381781681)

[*2.2.2* *Genetic algorithm (GA)* 21](#_Toc381781682)

[*2.2.3* *Artificial neural network (ANN)* 24](#_Toc381781683)

[*2.2.4* *Support vector machine (SVM)* 27](#_Toc381781684)

[*2.2.5* *Discussion on the Statistical Tools* 28](#_Toc381781685)

[1. Bibliography - State of the Art Section 29](#_Toc381781686)

**INTRODUCTION**

**Description:**

The goal of this document is to describe the state of the art of the Enerficiency project. It is part of the outputs of Task 1.1 High level Business Requirements, State-of-the-art and Gap Analysis. This information has been used as input in order to investigate, develop and validate the technical and high level business requirements of the Enerficiency topics being proposed: communications components, central core (automated energy data management) and computational intelligence. Therefore it investigates the business challenges and opportunities in the building energy efficiency domain as well as the main methodologies and tools used to enable the prediction and forecasting of energy performance. The advantages and disadvantages of the techniques will be studied in detail and one methodology will be focused upon for the purpose of Enerficiency

# State of the Art

This section gives an overview of the state of the art analysis conducted in the context of the Enerficiency Project. This section is divided in three subsections:

* The first sub-section gives the general context and outlines the various classes of approaches studied in the state of the art.
* The second sub-section focuses on the “grey-box” approaches, which are the ones to be deployed in the Enerficiency project
* The third sub-section gives a summary and conclusive elements

***Note:***

* ***This section is only a summary of the full-breadth state of the art that was achieved, which includes also all elements presented in the Annex 1 (section*** 2***) of this document.***
* ***The bibliography with a comprehensive list of scientific references for the State of Art section is given in section 3 of the document.***
* ***The “Enerficiency Modelling Approach” document supports the State of the Art section by giving a short overview of 3 modelling paradigms available for building simulation, the system identification procedure to obtain grey box models and the requirements and potential for using simulation models in the operational phase of a building.***

## Context & Big picture

The building sector in the European Union is considered as the largest consumer of energy, using up to 40% of the final energy consumption [30]. More specifically, residential uses represent about 60% of total energy consumption of the building sector [10][73].

To evaluate the energy performance of both residential and non residential buildings, many parameters are required: thermal characteristics of the building, heating and hot water installations, air conditioning installation, ventilation (natural and artificial), outdoor climatic conditions (position and orientation of the building, solar exposure), passive solar system, indoor climatic conditions and lighting (natural or artificial)[73]. Moreover, Aydinalp et al [4][5][6] have shown a large influence of the socio-economic factors (as the household income or else the number of inhabitants) on the energy consumption.

The building energy end-uses are usually classified in five distinct categories [40] including space heating (electricity, gas, fuel, etc), domestic hot water (immediate DHW, immersion heater, gas water-heater, solar water-heater, etc), appliances (refrigerator, clothes washer, dishwasher, freezer etc), space cooling (chillers, aerators, central ventilation system, fan, central air conditioning system etc) and lighting (halogen, fluorescent, incandescent lights).

Figures 1 and 2 show the breakdowns of energy consumption in EU for residential and non-residential (dealing with the tertiary sector without agriculture) sectors. Despite the huge number of factors influencing the energy performance of a building, we can give an estimation of the energy consumption in European Union rising to about 200 kWh/m2/year. More specifically, the heating energy consumption of a residential and a non residential building amounts, respectively, to about 174 kWh/m2/year [10] and 152 kWh/m2/year [34]. An energy-efficient building amounts to 30-50 kWh/m2/year (for example in Greece), while the heating energy consumption of an energy-consuming building can exceed 450-500 kWh/m2/year (one case in Poland reached 763 kWh/m2/year)[10].

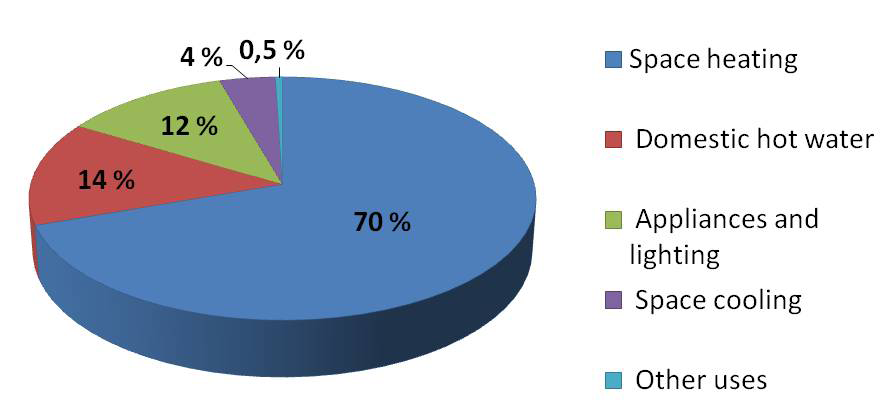


Figure 1: Breakdown of energy uses in buildings in the residential sector in 2001 [10]

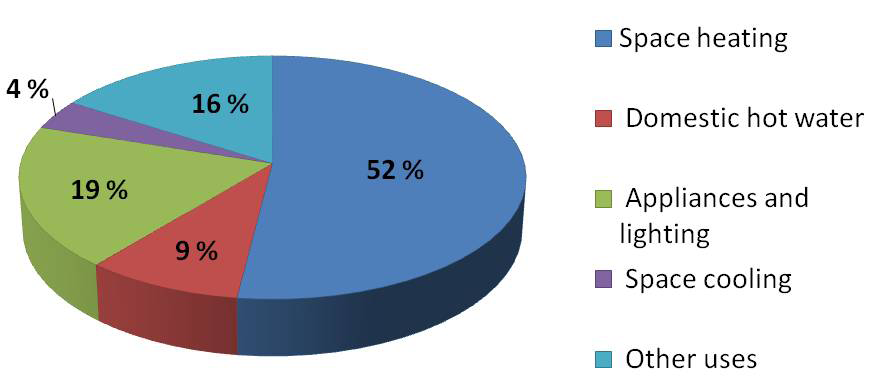


Figure 2: Breakdown of energy uses in buildings in the non-residential sector in 2001 [34]

To increase energy efficiency, the European Union established concrete actions by introducing the EPBD directive (Energy Performance of Building Directive) completely dedicated to the building environmental issue [31]. With the aim of reducing the large differences between the EU members in term of energy savings, this directive suggests to each EU state to target their own objectives. For example, concerning the residential sector, different projects of passive building have emerged in Germany with PassivHaus, in Switzerland with Minergie and with Effinergie in France. A passive house is defined as a building that assures a comfortable indoor climate both in winter and summer without needing the use of conventional heating or cooling system. To be recognized as a passive house, a dwelling must have the following specific characteristics:

* the annual heat demand cannot exceed 15 kWh/m2/year;
* the total energy consumption must be lower than 120 kWh/m2/year (in primary energy).

Several solutions have been proposed to increase the energy efficiency:

* An awareness campaign with the occupants on the environmental issue is necessary to reduce end-use energy consumption [12]. Simple actions could significantly decrease the energy consumption such as changing the space heating behaviour, unplugging the computer or mobile charger and unused devices, configuring the computer to hibernate after a given time of inactivity, avoiding waste of hot water and many other actions [15].
* A second solution is proposed to promote the use of passive solar systems. Many authors have published possible scenarios. Especially, Ihm et al [39] have highlighted the potential of daylighting to increase the energy performance of a building associated with electrical lighting. In the same way, Badescu and Sicre [8][9] evaluated the performance of the solar energy on a passive house in Germany. In order to reduce the energy demand, they coupled a passive solar heating across the large window on the south-facing facade and an active solar collectors creating thermal energy for space heating loads and domestic hot water system, with a ground heat exchanger permitting to regularly supply the house with fresh air during the cold season. They showed that in this case the yearly relative heating demand is about 5.6 kWh/m2/year.
* A third solution is to use control and monitoring systems allowing controlled blackouts during specific moments of a day. Many authors have already demonstrated the efficiency of such systems on the energy performance of buildings [14][70][72]. For example, very recently, Aswani et al. [2] have published a work dealing with a model-predictive control of the HVAC systems able to control the indoor temperature of a room of a computer laboratory in the University of Berkeley. The aim of this study was to reduce the energy consumption of this room. Previously, Mossolly et al. [61] proposed a study comparing several control strategies in order to increase the energy performance in an academic building in Beirut, Lebanon. By determining the optimal control strategy, they recorded energy savings up to 30% during the summer season.
* Coupled with the use of renewable sources, a fourth solution consists of the constructing new dwellings or the refurbishing existing dwellings and bringing energy efficient improvements in agreement with the regulations given above. For example, one way is to favour the exterior insulation when possible or to replace simple glazing windows by double or triple glazing depending on the exposure of the room. The choice of energy efficiency improvements is not obvious with the risk to produce the opposite effects. For example, improving the efficiency of daylighting could decrease the lighting-energy consumption but would also reduce the heat gain of artificial lights, increasing the space-heating energy demand and in the same manner, the building-energy consumption. To avoid these errors, it is necessary to use computing support able to formalize the building behaviour.

In this state of the art, we focus on this last point dealing with the choice of the most adapted design for an existing building or computing a model for a new building. Several approaches are usually chosen: some of them based on the thermal knowledge and physical equations of the building and others based on the data collected inside the building:

* ”White Box” approaches (used to model the thermal behaviour of a building). The White Box approach is used for several applications at different scales. For example, the white box scheme allows to evaluate the indoor temperature in a building for different times (year, month, day or hour) and the spatial (the entire building, a room, a cell of a room) scales.
* Statistical or machine learning formulations called ”black box” approaches are mainly used with the aim to deduce a prediction model from a relevant data basis (for example, to forecast energy consumption or heating/cooling load in a given building).
* Finally, some solutions are proposed to couple the white and black box techniques to implement hybrid approaches also called ”grey box” approaches.

More precisely, when it comes to building energy performances characterization and prediction, two main kinds of approaches may roughly be differentiated: white-box approaches, relying on explicit physical models of buildings, and black-box approaches, based on machine learning relying on statistical processing of building energy and the comfort data. While white-box approaches are widespread, both in tools and practices, their limitations are widely known. The most common limitations include the necessity to have exhaustive knowledge of buildings structure characteristics and uses. Furthermore, as the White box models are rigid there is often a disagreement with the real energy data.

On the contrary, black-box approaches are still confined to academic works. The analysis of the available results show that, with respect to black-box methods, the following types of techniques are prominent which include: linear multiple regression, artificial neural network (ANN), genetic algorithms (GA) and support vector machine (SVM). These techniques have been tested and assessed with various application purposes, for e.g. offline consumption data analysis (based on regression techniques), optimization of a cooling systems’ operational planning using GA, or energy demand prediction due to ANN or SVM.

While these approaches can be implemented without a physical model of the building (before deployment) and the ability to dynamically adapt to changes in building characteristics and usage, they usually suffer from a lack of interpretation and require very large amounts of data to be properly configured. Therefore, it is usually beneficial to couple them with physical models in order for them to adapt dynamically, while keeping a sufficient level of interpretability to support energy management strategies. Such approaches are termed hybrid (or grey-box) approaches and have recently yielded promising results. In particular, genetic algorithms have shown to be well adapted to such coupling, with encouraging results in terms of energy conversation measures generation (e.g. building occupancy scenario optimization). The tools deployed in Enerficiency rely on such approaches, with a mix of fully statistical, black-box methods to support energy data analysis and fault detection, and of hybrid, grey-box approaches to support diagnosis and generate appropriate Energy Conservation Measures (ECM). When appropriate, ECM generation is likely to rely also on complementary artificial intelligence techniques (namely expert systems), to enable emulation of energy managers’ decision-making ability. This combination of such advanced techniques is, to our knowledge, a real breakthrough.

In the following section, for the sake of clarity and brevity, we will focus only on the type of approaches that will be deployed in Enerficiency: namely the “grey box approaches”. For a more comprehensive overview of “White-box” and “Black-box” approaches, and for the related bibliography, readers are invited to refer to the annex 1 of this document.

## Hybrid models or “Grey-Box” approaches

The previous section briefly outlined the capacity of both the detailed physical and the statistical methods in the building simulation, (please refer to annex 1 for a comprehensive overview), while also highlighting the limitations of each technique. Specifically, the white box methods assume that all building characteristics, both thermal and geometric, are well-known. This is usually the case for building design but it is more difficult to collect detailed information on existing buildings. However, to establish monitoring strategies, this information is essential. Moreover, these approaches require the ability to describe all physical mechanisms with a high accuracy. Nevertheless, although most of the thermal phenomena are well-known, some of them are based on assumptions and remain difficult to model accurately; for example, natural ventilation is often described by empirical equations.

Black box methods are mainly limited by the large quantity of the data required. Also, it is usually difficult to interpret in physical terms results obtained by such statistical approaches.

Data mining techniques are specific to each building. Each building requires its own model. In contrast, use of general heat transfer equations in white box methods may usually be applied generally. It is possible to overcome the limitations of each method by using them together. The advantages of one method compensates for the drawbacks of the other. For example, by retaining some physical meaning, one improves the interpretability of the model. Moreover, building characteristics can be determined by optimization techniques such as genetic algorithms. Thus, all physical and geometrical input parameters are not required. These hybrid methods, that combine physics and statistics are called ‘grey box” methods.

#### Principles

This approach has been introduced at the beginning of the 1990’s for a specific application which was the automatic control system. For example, Teeter and Chow [82] combined an artificial neural network with a single-zone thermal model to improve the efficiency of the HVAC control by performing the HVAC parameters identification. Other more recent examples are the works of Paris et al. [79][70] who combined the fuzzy logic, a PID controller and a dynamic model describing the thermal behaviour of the building for implementing several heating control schemes. Another application of the hybrid model is the parameter identification. In this approach, the aim is to compute the set of input values corresponding to a given set of outputs. For instance, the objective may be to calculate the optimal thermal properties of the walls (conductivity, capacity, etc) given a target consumption/comfort level. The technique is to combine physical models (used to simulate the thermal behaviour of the building) and statistical technique to retrieve the set of optimal inputs corresponding to the desired outputs.

#### Advantages and limitations of the hybrid methods

The main advantage of the hybrid method is that it allows the consideration of only a limited number of data. Furthermore, the input parameters do not need to be fixed at the initial time of the simulation. Only bounds on physical parameters are required. Thus, a rough description of the building geometry and thermal parameters is sufficient. Furthermore, the hybrid methods allow retention of a physical interpretation. However, some drawbacks pertaining to each type of technique remain in hybrid methods as the free parameters for statistical tool or the computation time requiring both, physical or statistical codes. Another drawback to this hybrid technique is that it uses two distinct scientific domains thus requiring users to train in both scientific methods.

#### Application of the hybrid method

As stated above, the hybrid formulation is applied to the implementation of control systems but also to the prediction and the improvement of energy consumption in building by providing and testing several energy-saving scenarios. Likewise, it can be applied to specific cases as the evaluation of the heating and cooling demand or of the indoor temperature. A reasonable amount of data, collected in a short period of time for the training phase of the statistical tool is adequate. In the literature, some papers focus on the coupling between nodal techniques for the thermal and geometrical representation and genetic algorithms for the parameters identification. We propose to detail some examples using the hybrid method in this specific way.

#### Examples

Lauret et al. [51] implemented a model resolving the state equations in a building with a very simple geometry in the Island of the Reunion to follow the evolution of the indoor dry air temperature. They combined a physical resolution by the finite difference method via CODYRUN (multizone software) with a genetic algorithm. The study is based on experimental data. The authors demonstrated that in previous studies, using the physical model alone was insufficient to predict behaviour that was in agreement with experimental data. [58][60]. In this study, they used a genetic algorithm to isolate the defective node measurement by forcing the value of some temperatures in specific place of the building. The aim is to optimize the value of the indoor dry air temperature.

Znouda et al. [96] studied energy consumption in a Mediterranean building in Tunisia. More specifically, they found the solutions to improve both the energy efficiency and the economic point of view by optimizing architectural parameters. To perform that, they coupled a simplified tool for building thermal evaluation specific to the Mediterranean countries called CHEOPS to a genetic algorithm for the architectural parameters identification. They studied the energetic and economic problem independently. The authors studied a solution adapted both in summer and in winter. They showed that it is difficult to solve this kind of multi-objective issue composed of two independent problems (energetic and economic) because the optimal solutions are different considering either saving energy or saving money.

Wang and Xu [90][91] studied the building thermal transfer in summer in Hong-Kong. The building consists of three different buildings; an office building, another a shopping centre and the last one, a restaurant. The study was based on data collected during a survey conducted in order to deduce the profile of occupancy and use of the lighting and equipment. They used an electrical analogy to predict the heating/cooling load by substituting the building envelope (the roof and the external wall) by two different 3R2C systems and by introducing an internal mass by a 2R2C system. The internal mass corresponds to all others heat storage materials as furniture, carpet, partitions, equipment, etc. Combining the equations resolution with the genetic algorithm for the parameter identification, the authors optimized the values of the resistances and capacitances of the internal mass.

Siddarth et al. [78] have coupled a genetic algorithm and DOE 2.2 in order to establish a data basis allowing them to implement regression functions describing the annual energy consumption. Indeed, they used genetic algorithm to generate several set of parameters. Each set of parameters has been tested in DOE 2.2 which returned the annual energy consumption. Some of these set of parameters are then selected under an annual energy consumption criterion stored in a database, which will be used for the implementation of a regression function. Under this annual energy consumption model, they are, thus, allowed to suggest energy saving strategies.

#### Discussion on the hybrid methods

We have seen in the given examples that the hybrid method is used for optimization applications. Contrary to statistical or physical approaches that we developed above, the aim is not just to predict the thermal behaviour of a specific building but primarily to derive possible strategies to improve energy efficiency. The hybrid technique selects the advantages of each technique and uses them to implement efficient models for monitoring and control applications. For the time being, the physical technique mainly employed is the nodal approach and the most adequate statistical tool used for the parameter identification seems to be the genetic algorithm. However, given that the “grey box” domain is still emerged recently, the approach might evolve in the future.

## Conclusion

This section, together with Annex 1, proposes a review of the main techniques and tools enabling building energy performances prediction. These techniques belong to three categories, each of which is associated to specific scientific paradigms and fields:

The first category includes approaches relying on physical models (”white box” methods). The White box approaches in turn may be divided into three sub-categories, which mainly correspond to a gradual rise of the level of details of building models: the multizone technique which considers the space as a homogeneous volume where all states variables are uniform, the zonal method which divides each room in several cells and the CFD method which describes each zones in several control volumes. Then, methods based on machine learning (or “black box” methods) rely on statistical treatments of building energy and comfort data. Four methods are commonly used- (i) conditional demand analysis, (ii) artificial neural networks, (iii) genetic algorithms and (iv) support vector machine.

The last category uses hybrid approaches which rely on both physical models in order to simulate building thermal behaviour and machine learning techniques in order to optimize input parameters. The outcomes of the study are as follows.

The first category of methods - those relying on physical models - are mostly applicable to contexts in which building design data are available, and especially in the scope of the design of a new building. These methods rely on detailed descriptions of buildings, notably entailing geometry, material properties, and energy systems features. While this information can be considered to be easily extractable from design data in the case of a new building, this is less than obvious for existing buildings (e.g. in the scope of a refurbishment). This is true for the most basic of these methods - the nodal one - but all the more true when we consider more advanced ones (zonal, CFD). When it comes to comparing the methods among each other, the conclusion is quite straightforward that it is better to use more detailed models (CFD) for the sake of reliability and precision of the simulation result, but models are more tedious to build and computation times are higher. Zonal methods can be considered as good trade-offs, but still, most simulation tools used today in “live” projects are based on nodal approaches. Nevertheless, a possible trend is a gradual shift to CFD methods with computers becoming more powerful.

The second category of methods, which are based on machine learning techniques, are extremely useful in other situations, i.e. those in which one has real energy and comfort data from the building but has little or no information about the design. However, the reliability of these techniques is highly dependent on the quality and quantity of available data, as were the physical approaches dependent on the complexity of the underlying model. It is quite difficult to perform a qualitative and comparative assessment of the various techniques devised in this field, since - again - their performance will depend on the training data used as input. Compared to physical approaches, machine learning ones require less information about the building and may appear as easier to deploy. However, physical approaches are better where interpretation of physical phenomena is desired.

At last, hybrid approaches appear as a very promising field for the near future [20]. They can be considered as a good trade-off between physical and machine-learning based methods, and relax their drawbacks by combining them. Hybrid methods may be appreciated in situations where a physical building model is available, but is incomplete or does not cover enough details, and must be adapted and / or completed. When dealing with existing buildings, where it is usually difficult to rebuild detailed physical model, such approaches could be of great help.

# Annex 1: Complements on the state of the art

This section gives a more comprehensive overview of the study performed on “White-box” and “black-box” approaches, as a complement to section .

## Physical models: Building thermal behaviour modelling

Physical models are employed to model the thermal behaviour in different varieties of buildings with their own specific needs: dwelling, office, hospital, school, firms etc. Some of them include models of space heating ,natural ventilation , air conditioning system , passive solar , photovoltaic panel , hydrothermal effects , financial issue , occupants behaviour , climate environment etc. The physical techniques are based on the resolution of the equations describing the physical behaviour of the heat transfer.

### *Conservation Law*

We give here a simplified formulation of the physical heat transfer issue. The first step is to list incoming and out-coming fluxes and to write the energy conservation equation. presents a scheme of the different fluxes occurring in a heat transfer system.

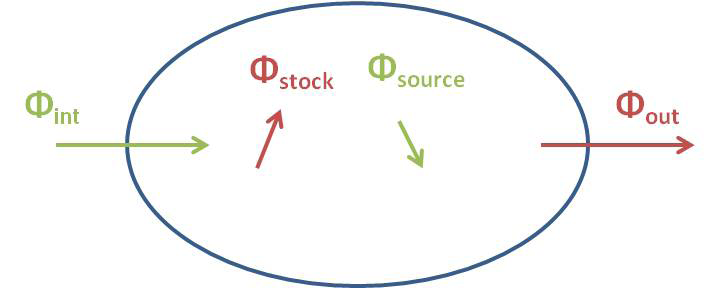


Figure 3: Scheme of the different flux occurring in a heat transfer system

According to the first law of thermodynamics, the energy conservation equation is written as in the following equation:



Φint is the heat flux entering the system, Φsource the heat flux of a potential heat source, Φout the heat flux leaving the system and Φstock the heat flux stored.

### *From Equations to Software*

To solve such physical problems, a large number of numerical software is available. Many authors proposed benchmarks to compare these software to which the reader is invited to refer for further information on comparisons between these environments. Theoretically, each building software is able to include each of mechanisms described in the previous section. They give the choice to users to select the mechanisms and the associated equations occurring in the system. But, in reality, Woloszyn and Crawley showed that many software are badly adapted to take into account the consequences of the humidity and that, generally, the effects of the latent heat are neglected. Three main thermal building models are currently used: the multizone, zonal and CFD (Computing Fluid Dynamic) methods. Each of them has its own application and actually the choice of the physical method depends essentially on the problem considered. In the following, we will detail and give some examples for each of these methods.

### *The CFD Approach*

#### Principles

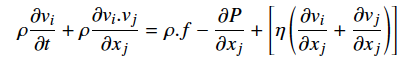
The most complete approach in the physical building simulation is the CFD (Computational Fluid Dynamics) simulation. This is a microscopic approach of the ventilation modelling allowing to detail the flow field. It is based on the decomposition of each zone of the building in a large number of control volumes with a global mesh uniform or not. So, the CFD technique is recognized as a three dimensional approach. is a schematic representation of a problem solved with the CFD method.

Software using the CFD simulation are essentially based on the resolution of the Navier-Stokes equation. Assuming the airflow as an incompressible flow, the Navier-stokes equations are given by the equations that follow.

The mass conservation equation is enounced as:



where vi is the air velocity in the direction i. The momentum conservation equation is written as:



where **ρ** is the density of the fluid, **η** the dynamic viscosity and P the pressure. f is all the forces applied to the fluid.



Figure 4: Schematic representation of the streamlines in a problem solved with the CFD method

There is a huge number of CFD software such as FLUENT, COMSOL Multiphysics, MIT-CFD, PHOENICS-CFD etc. Usually, their application fields are very large and do not have building simulations as unique application but every systems considering detailed flow description.

#### Advantages and application field of CFD methods

The CFD method is mainly employed for its ability to produce a detailed description of the different flows inside buildings (airflow, pollutant flow, etc). Consequently, the CFD is very well-adapted to the study of the particle transport as pollutant particles. Moreover, we saw that the volume is divided into several discrete control volume with a mesh that can be uniform or not. Thus, it allows the study of very complex geometries of the building by minimizing locally the mesh of some specific parts.

#### Limitations

The main disadvantage of the CFD approach resides in its huge computation time. This is due to the fact that a complete detailed 3D-description of a building requests a very fine mesh that is a huge number of control volumes. Consequently, the smaller the mesh, the larger the computation time. However, given that the air velocity in at least 75% of the building is less than 0.5 m/s, it is not always necessary to apply the CFD technique in the entire building but just to specific constituents of the building as HVAC (Heating, Ventilation and Air Conditioning) equipment or appliances. Thus, it allows considerable reduction of computation time. That is why the CFD is frequently coupled with one of the methods described hereafter (nodal or zonal). Tan et al. compared the full CFD simulation and the coupling between the CFD and multizone method for modelling the natural ventilation across large openings or atrium. They showed that the full CFD simulation would take more than ten hours, whereas the coupled method needs less than one hour. Moreover, the CFD method is quite limited by the complexity of the model implementation. Indeed, it is not easy to use without previous knowledge on fluid dynamics and software. Furthermore, the CFD is also largely limited when it comes to model of the turbulence. In some cases, it is not necessary to describe flows in the building so finely and a way to overcome the difficulties enforced by the CFD is to model the building behaviour in a simpler manner by giving a less detailed description of the interested zone. The zonal approach is a way to obtain this simpler modelling while maintaining the complexity in a 2D map.

### *The Zonal Approach*

#### Principles

The zonal method is an intermediate one between the CFD and the multizone techniques. This approach is a fast way to detail the indoor environment and to estimate a thermal comfort zone. Practically, it consists in dividing each building zone into several cells. One cell corresponds to a small part of a room. So, the zonal method can be assumed to a two dimensional approach. represents a scheme in case of zonal methods.

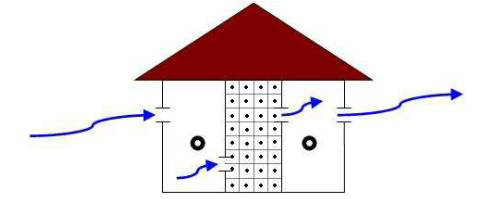


Figure 5: Schematic representation of a problem solved with the zonal method [85]

#### Advantages and application field of the zonal approach

The zonal formulation can treat a large volume space and the coupling between the system and its environment. The physical equations are solved for each cell of the zonal system. Consequently, it allows the determination of the local variables in a 2D-map. Thus, it permits the evaluation of the spatial distribution of different fields such as, temperature, pressure, concentration or air velocity remaining with a quite reasonable computation time. Wurtz et al. showed that the zonal simulation is a suitable method for an accurate estimation of the temperature field in a room and of the indoor thermal comfort. It allows also the visualization of the system airflow. One building modelling software using the zonal technique is mainly employed to represent indoor airflows: SimSPARK. Equations are solved by the object-oriented environment called SPARK. Moreover, some researchers implemented their own zonal software as POMA .

#### Limitations

As mentioned above, the zonal approach is a minimization of the complexity of the CFD method. Thereby, it is obvious that some studies normally well implemented via the CFD are not anymore feasible via the zonal method. Notably, some limitations reside in the following aspects:

* this technique requires previous knowledge of the flow profiles
* it is not able to provide accurate results on the detailed description of the flow field.
* the study of the pollutant transport remains limited.

Although, it has been hugely enhanced with the zonal approach, it also remains a problem with the computation time and the complexity of the equation implementation. To overcome these difficulties, another simplification is possible considering no more a 2D or 3D description of the building behaviour but a simple 1D visualization of the phenomenon occurring in the system.

### *The Multizone or Nodal Approach*

#### Principles

This last approach, which is probably the simplest one is called the multizone technique (also called nodal method). It considers the following assumption: each building zone is a homogeneous volume characterised by uniform state variables. Thus, one zone is approximated to a node that is described by a unique temperature, pressure, concentration etc. Generally, a node represents a room or the exterior of the building but it can be more specific like a load. The physical equations are solved for each node of the system. So, the nodal method can be considered to a one-dimensional approach. is a scheme of the nodal modelling. TRNSYS, EnergyPlus, IDA-ICE, ESP-r, Clim2000, BSim and BUILDOPT-VIE are softwares using the nodal approach employed for building simulations.

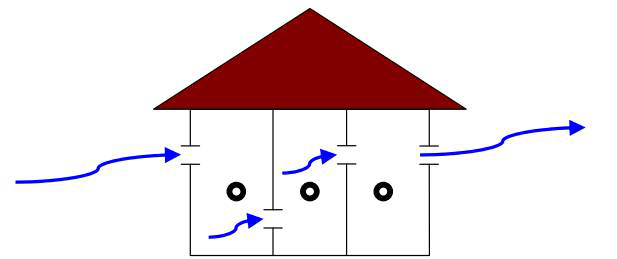


Figure 6: Schematic representation of a problem solved with the multizone method [85]

#### Advantages and application field of the nodal approach

The huge advantage of this technique resides in its ability to describe the behaviour of a multiple zone building on a large time scale with a small computation time. It is a particularly well-adapted tool for the estimation of the energy consumption and the time evolution of the space-averaged temperature into a room. Moreover, it can be used to predict the building air exchange rates and the airflow distribution between different rooms of a building. Some other applications as the ventilation efficiency or the pollutant transport for entire buildings can also be studied by this formulation .

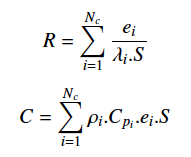
#### Limitations of the nodal approach

Due to the simplification enclosed in the multizone approach, it has obviously some limitations for treating some specific cases, better supported by the very complete CFD method.

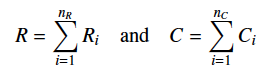
* The study of the thermal comfort and the air quality inside a zone is quite difficult.
* The impact of loads on their close environment is not addressed (for example, a radiator with a plume).
* Despite the fact that it is a well-adapted method to study a multiple zone building, it is quite difficult to apply the nodal form to a room with a large volume.
* Although it is a good way to visualize the distribution of pollutant between building zones, it does not allow the consideration of the local effects of a heat or pollutant source. Reduction of the computation time and the technical complexity is always a concern, our team is particularly interested in a variant of the nodal method based on the analogy with electrical circuits.

#### The electrical analogy: a variant of the nodal method

Another way to formalise and simplify a nodal problem is to use an analogy with an electrical circuit to rewrite the conductive transfer across a wall. This method introduced by Rumaniovski is very useful since it simplifies drastically the physical problem through a linearization of the equations and thus, reduces the computation time. The principle of the electrical analogy is to associate a thermal resistance R and a thermal capacity C to a wall by introducing the following relations:



**λ**i is the thermal conductivity, **ρ**i the mass density, Cpi the thermal capacity, Nc is the number of distinct layers into the wall, S the surface of the body and ei is the spatial interval between two different layers. Moreover, let us introduce the components of the RC model: the resistances Ri and the capacities Ci. They verify the following conditions:



nR being the number of resistances and nC the number of capacities both in the RC model. The more the number of components is important, the more the model is accurate. However, it is obvious that a higher system leads to a significant increase of the computation time. Fig (7) presents three examples of descriptions of a wall with different levels of granularity.

The analogy gives the following equivalence with the Ohm’s law:



The temperature **θ** is equivalent to voltage U, the heat flux **Φ**L to current I and the thermal resistance e/**λ**.S to electrical resistance R. We propose in the following some examples of these methods.

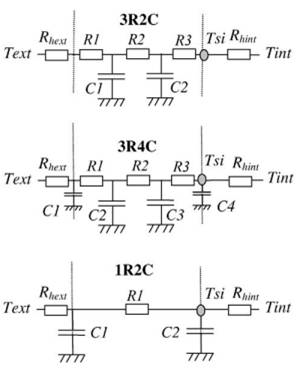


Figure 7: Examples of different electrical circuit of a wall with three levels of granularity [33]

#### Examples

Kalogirou used a multizone software TRNSYS (Transient Simulation Program) to determine the energy consumption in a building in Nicosia, Cyprus. More specifically, the aim was to see how the energy demand behaves with a hybrid photovoltaic-thermal solar system (coupling of a normal PV panel and a heat exchanger) rather than a standard photovoltaic panel. The authors chose the nodal method because on the one hand, their studied system was constituted of several interconnected zones and on the other hand, they were interested in a specific macroscopic variable (energy consumption) and not in the distribution field.

For the same reasons, Ibanez et al. used the TRNSYS software to study the efficiency of the phase change materials (PCM) in Lleida, Spain. To perform that, they considered a uniform indoor temperature in the room and determined its time evolution. By using the TRNSYS software, they evaluated the influence of the PCM on different part of the envelope of the room (wall, ceiling and floor).

In the same way, Zhai et al. [95] studied the effects of the ventilation in summer on simulated data of indoor temperature with the multizone software EnergyPlus . To achieve that, they compared experimental and simulated measures of indoor temperature in three distinct building offices: a single-story building with automatically controlled air ventilation in Belgium, a three-story building with manually controlled air ventilation in Denmark and another three-story building with an automatically controlled air ventilation in United Kingdom.

Likewise, Cron et al. used the electrical analogy to estimate the performance of hybrid ventilation. The system is composed of fan assisted natural ventilation incorporating a control demand strategy based on indoor air temperature and CO2 concentration.

Musy et al. studied the indoor thermal comfort in a room through a zonal software called SPARK. Particularly, the aim was to determine the vertical profile of temperature and the pollutant concentration repartition inside the room. The authors justified the choice of a zonal approach by the necessity to reduce the computation time compared with a CFD and the inability of the nodal method to provide detailed temperature and flow distribution and, in the same way, to predict the thermal comfort.

Other authors chose this strategy for the same reasons:

Tittelein et al. focused their works on the passive house and the methods to reduce the energy consumption of a building located in the region of Chambery, France. They compared the effects of a counter-flow ventilation and a single-flow ventilation on the energy efficiency.

Likewise, Stephan et al. were interested in inverse methods for improving the performance of the natural ventilation in a room. By coupling SimSPARK with an optimization calculation, they deduced the optimal size of the openings needed to maximize the performance of the natural ventilation.

Moreover, Haghighat et al. implemented a software using the zonal approach called POMA (Pressurized zOnal Model with Air-diffuser). This software allows the prediction of the airflow pattern and the temperature distribution in a room which is naturally or mechanically ventilated.

Furthermore, Jiru and Haghighat computed the airflow and the temperature distribution in a ventilated double skin facade, using the zonal method. Specifically, they compared the time evolution of the temperature in three positions inside the facade. Parametric studies have been accomplished in order to test the influence of the cavity height, the flow rate and the presence of venetian blinds on the inlet-outlet temperature difference.

Zhai et al. coupled EnergyPlus and MIT-CFD programs to predict the cooling or heating demand both in an office and in an auto racing complex. The authors used EnergyPlus to determine the cooling or heating demand and MIT-CFD to find the airflow and temperature distribution in the zone volume. The necessity of using the CFD here is due to the fact that it deals with very large volumes (office and auto racing complex) where the convective mechanisms are really complex. We mentioned above that the nodal approach assumes that the convection depends on the constant parameter h. So, it does not allow the treatment of large zones with a high accuracy. Thus, the use of the CFD is necessary. At each time step, EnergyPlus passed the information to the CFD program that used them as boundary conditions. Then, the CFD program deduced the distribution of the air temperature in the thermal boundary layer and the convective heat transfer coefficients into the office. Finally, these outputs are injected in EnergyPlus as inputs to improve the accuracy of the heating load prediction.

Other authors chose the same strategy. For example, Wang and Wong used ESPr (a software using the nodal method) and FLUENT (a flow software using the finite element method) to simulate the natural ventilation in residential buildings. The ESP-r simulation contained the geometric information, the construction thermal properties and the airflow network for the whole building. The place studied is a double zone building. To reduce the computation time, the authors chose to apply the CFD simulation only in one zone and to pilot the system by imposing pressure as opening boundary conditions. The ESP-r simulation results provided boundary conditions to the CFD simulation.

Moreover, Srebric et al. coupled a multizone tool called CONTAM with a CFD tool called PHOENICS-CFD to evaluate the contaminant distribution in a building. First, they determined the airflow rates and the contaminant transport between zones. Then, they applied the CFD simulation only in the contaminant sources to deduce the airflow profile and the concentration distributions. These results are injected as fluxes in a new CONTAM simulation excluding the CFD domains. Finally, they evaluated the contaminant distribution. The authors showed that the coupled method is efficient in the zones very near the contaminant sources. However, in the other zones, the multizone approach remains the more appropriate method.

### *Discussion on the Physical Models*

The previous sections have described several physical methods employed in the building modelling. We saw through the principles of each technique and the previous examples that each physical method has its own field of application. The most complete and detailed approach is the CFD. It allows the very fine description of each mechanism occurring in the building system. Especially, it is particularly adapted for modelling the convective phenomenon taking place in a large zone volume. However, it is difficult to simulate all phenomena by using CFD because of the huge computation time. This is the reason why it is usually coupled with nodal software as EnergyPlus or TRNSYS. The nodal approach is really well adapted to treat global resolution as the determination of uniform field. Contrary to the CFD, phenomena are described less finely. The aim is to simplify as far as possible the resolution system by linearizing the major part of the equations (when it is physically possible). Thus, the technical complexity is significantly reduced and the computation time too.

The zonal method is an intermediate technique between nodal and CFD approaches. It is less accurate than the CFD but retains more information compared to the nodal technique. Enerficiency focuses on building control and monitoring. Such applications are part of those which it is really important to have simple tools with computation time as short as possible. Thus, the technical simplicity and the reduced computation time of the nodal approach and more particularly the electrical analogy make them the most relevant for our scope. Generally, an important drawback of the physical formulation is the fact that it requires a detailed description of the physical behaviour. So, it implies extensive knowledge on the physical system, especially on the mechanisms occurring inside and on the building geometry. Unfortunately, it is far from being always the case. In contrast, the statistical tools have the great faculty to product a model only from measures.

## Statistical Methods based on Machine Learning

The particularity of statistical models compared with physical methods is the fact that they do not require any physical information. No heat transfer equations, no thermal or geometrical parameters are preliminary needed. Indeed, statistical models are based on the implementation of a function deduced only from samples of training data describing the behaviour of a specific system. Thus, these methods are well adapted when the physical features of the considered building are not known. Several statistical tools are able to build prediction model using learning methods. The great power of these techniques is the fact that they do not need to have much knowledge about the building geometry or the detailed physical phenomena to deduce an accurate prediction model. In contrast, they are totally based on measures and in such cases where it is difficult to collect data, it can become a real issue. We propose in the following part of this state of the art to describe the statistical techniques mainly employed in the field of the building energy forecasting: the linear multiple regression, the artificial neural network, the genetic algorithm and the support vector machine. These techniques belong to the branch of the artificial intelligence.

### *Multiple linear regression or conditional demand analysis (CDA)*

The conditional demand analysis (CDA) is a linear multivariate regression technique applied to the building forecasting. The linear regression was introduced by Galton in 1886. In 1980, Parti and Parti were the first to propose a new method using the linear regression for the prediction of energy consumption in buildings : the conditional demand analysis . The idea was to deduce the energy demand from the sum of several end-use consumption added to a noise term. In this way, they could infer the monthly and yearly residential end-use consumption from household invoices in San Diego.

#### Principles

The principle of the linear multivariate regression is to predict Y as a linear combination of the input variables (X1, X2,…, Xp) plus an error term **ε**i:



n is the number of sample data, p the number of variables and **α**0 a bias. For example, if the predicted output is the internal temperature, there can be as inputs the external temperature, the humidity, the solar radiation and the lighting equipment.

#### Advantages and limitations of the CDA

The CDA technique can be used both for prediction or forecasting and for data mining. This method has a main advantage which is the simplicity of use by beginners since no parameter has to be tuned. Indeed, no specific expertise of the method is required to manage such type of prediction method. However, the multiple linear regression presents a major limitation due to its inability to treat non linear problems. It leads to a lack of flexibility in forecasting but also a real difficulty to manage the multicollinearity inside the prediction results (that is the correlation between several variables). A possible solution to overcome these difficulties is to use a preliminary feature selection formulation.

#### Application field in CDA

In the building sector, the CDA is mainly used for forecasting energy consumption or comparing the evolution of energy demand between two different periods. But the constraint is mostly present on data. Indeed, a large amount of data is required for a proper prediction and moreover the non collinearity between variables is necessary. In the following, we present some examples using the CDA for a building application.

#### Examples

Lafrance and Perron studied the evolution of the residential electricity demand at the regional level of Quebec in Canada. More specifically, they used the CDA as a signal processing tool and compared three years of data: 1979, 1984 and 1989.

Tiedermann analysed the annual end-use consumption and the energy savings in the region of British Columbia in Canada. They studied also the energy consumption month by month and found two sudden increases: the first peak corresponds to November, December, January and February and is probably due to the use of the electric space heating and heating water. The second peak concerns the months of June, July and August and is related to the use of the air conditioning (central or portable).

Aydinalp-Koksal and Ugursal used the CDA to model the residential end-use energy consumption in Canada at the national level. They kept their interest on several end-uses : appliances, lighting, space cooling, space heating and domestic hot water. Different energy sources have been studied: electricity, natural gas and oil. Each end-uses for each kind of energy were described by a linear regression.

More recently, Aranda et al. has implemented a multiple regression model which allows to predict the energy consumption in the banking sector in Spain and to suggest energy saving strategies to increase the energy efficiency. The authors chose a model able to combine the simplicity of the evaluation method and the accuracy in the result without needing a huge amount of input data. Nevertheless, due to the non flexibility of the method, it is very difficult to apply the regression techniques for the data analysis. The following method is able to predict both linear and non linear problem. It is called genetic algorithm.

### *Genetic algorithm (GA)*

The genetic algorithm (GA) is a stochastic optimization technique deduced from an analogy with the evolution theory of Darwin. This artificial intelligence method has been introduced in 1975 by Holland its use as an optimization tool for the building simulation started in the 1990’s.

#### Principles

The principle of the genetic algorithm is based on the faculty of a given species to adapt itself to a natural environment and to survive extreme conditions. The genetic information is given by the gene sequences contained in the chromosome of an individual. In the GA process, all input variables are contained into one chromosome. This information can be coded in different way : binary, character string and tree. We will describe now the different step of the GA.

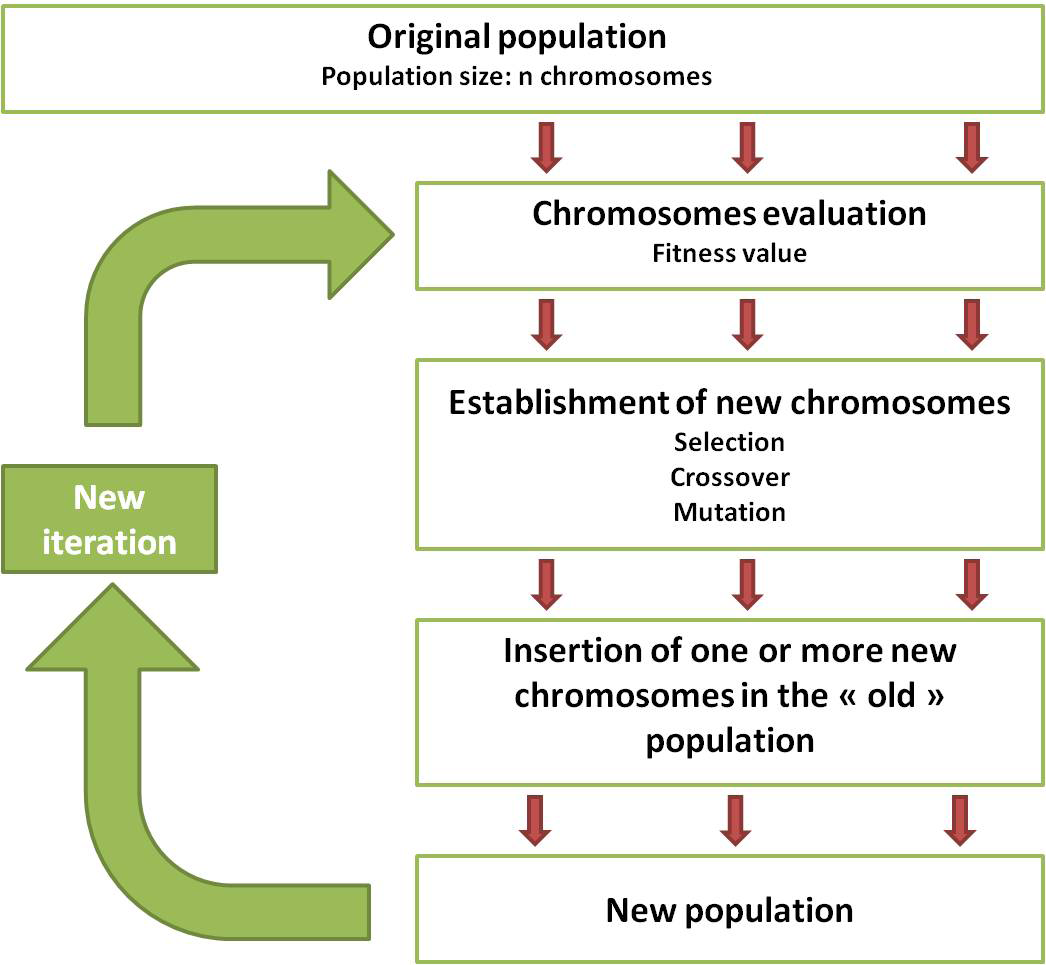


Figure 8: Scheme of the general operation in the genetic algorithm

(1) Production of the original population. A population contains a given number n of chromosomes or individual (which is the population size). Usually, this number is ranged between 50 and 1000.

(2) Evaluation of each chromosome based on the fitness value. The fitness value provides information on the adaptation performance of an individual. Its evaluation technique depends on the studied problem.

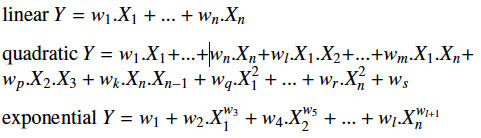
(3) Selection, crossover and mutation. The selection is responsible for selecting (at least) two chromosomes. Depending on the problem, it can be done randomly or based on the fitness value. When the selection is done, a crossover can be applied to the chromosomes. It deals with the exchange of a part of the information between the parents chromosomes. From two ”parents”, we can generate two ”children” randomly or not. The crossover is carried out with a probability pc ranged between 0.2 and 0.8. Finally, a last operation called mutation can occur with a very small probability pm usually ranged between 0.01 and 0.1. The mutation consists in the substitution of a part of a chromosome by another.

(4) Insertion of the new chromosomes in the population. At the end of the above processes (selection, crossover and mutation), we need to select new chromosomes to be added in the old population for creating the new one. For example, we can evaluate the fitness value of one (or more) new chromosome(s) and compared it with one (or the same number than the selected new chromosomes) old chromosome(s). If the quality of the new chromosome is better than the old one, it is injected in the population instead of the old chromosome. Thus, a new population is generated.

(5) Process reiteration. Once we have reach this step, the process restarts with the second step on the new population. The results of all steps is called a generation. Thus, the process continues until the user specified generation number NG is completed. Once again, no systematic law allows to calculate this specific number.

We can sum up the parameters needed in the genetic algorithms: the population size n, the number of generation NG, the crossover probability pc and the mutation probability pm. Moreover, we need also to know the way to evaluate the fitness and the selection process. shows a scheme of the general operations in the genetic algorithm.

In building simulation, GA is used to find a prediction model. The goal is to find a simple equation able to fit the problem. The form of the equation imposed by the user can have the following forms:



Y is the output (for example the energy demand), Xi are the input variables (for example the outdoor temperature, the humidity, the solar radiation and the exposure) and wi are the weighting of each input variables. The GA is used to optimize the weighting wi of each variables.

#### Advantages and limitations of the GA

An important advantage of genetic algorithm is the fact it deals with a powerful optimization method able to resolve every problems provided the convexity of the describing function. Another essential advantage of the genetic algorithm is its ability to give several final solutions to a complex problem with a large number of input parameters. It allows the user to choose with his own judgement the most probable one. Obviously, this is also a drawback by the fact that the user can never be sure to have chosen the best solution, especially as the GA will not necessary generate the optimal solution. Another disadvantage of the GA is the large computation time. Some authors try to reduce this computation time by coupling the genetic algorithm with other statistical methods. Especially, Magnier and Haghigat associated an artificial neural network to a genetic algorithm for estimating energy consumption and thermal comfort in a building. Another difficulty of the GA is the adjustment of the algorithm. Indeed, no rules are able to determine the number of individuals in the population, the number of generation or crossover and mutation probability. So, the only way to adjust the model is to test different combinations. Another important limitation of the GAs is their capacity to generate local optimum leading to study the system locally instead of globally. Finally, the performance of the GA is really limited when the individuals present a similar evaluation value. In this case, the genetic algorithm can no longer evolve. Moreover, in this specific case, an important drawback is the fact that it is absolutely essential to postulate the form of the describing function.

#### Application field of the GA

In building simulation, the genetic algorithm is mainly used for the determination of simple prediction models of the energy consumption and for the optimization of the equipment /load demand. The data basis can be both simulated or real and can contain instantaneous samples on several time scale (hourly, monthly or yearly) or samples averaged in time and/or space. As the ANN and the CDA, a large amount of data is required.

We propose now to give some examples of works using the genetic algorithm.

#### Examples

Ooka and Komamura were interested in the energy efficiency in building during a day. With two genetic algorithms, they provided the optimized combination of equipment capacity and optimized operational planning for cooling system during a period of 24 h with an electric turbo refrigerator and a heat pump and water heating system with two distinct heat pumps for hot water. For the equipment capacity, the authors used an algorithm with a population size of 10 individuals (2 sub-populations with a size of 5 individuals), a number of generation of 30, a crossover probability of 1, a mutation probability of 0.01 and a migration probability of 0.5. For the operation planning, the GA presented a population size of 24 individuals (3 sub-populations with a size of 8 individuals), a number of generation of 750, a crossover probability of 1, a mutation probability of 0.01 and a migration probability of 0.5. This work was applied to an hospital of Tokyo in Japan on a period of 24 hours.

Sadeghi et al. used the GAs to implement optimized prediction models of the annual electricity consumption per inhabitant in residential sector in Iran. Three forms of simple equations are tested: linear, quadratic and exponential. Their variables are the annual gross domestic product, the annual real price of electricity and the annual real price of natural gas. The population size is 60 individuals, the number generation raises to 400, the probability crossover is equal to 0.5 and the probability mutation 0.02. The fitness was evaluated by the reverse of the sum squared error. Thus, the criterion was to maximize the fitness value. The selection process was the roulette-wheel method.

In the same manner, previous works of Ozturk et al. studied the annual electricity consumption estimation in Turkey evaluated in the industrial sector and in the total electricity demand. The authors implemented two prediction models of the annual electricity consumption for both industrial and total Turkish demand, allowing the prediction of the annual electrical demand from 2002 to 2025. Nevertheless, it remains the problem due to the postulate of the kind of function. The following technique overcomes that by running as a black-box system. It is called the artificial neural network.

### *Artificial neural network (ANN)*

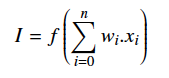
The artificial neural network (ANN) is a non-linear statistical technique principally used for the prediction. This artificial intelligence method was inspired by the central nervous system with their neurons, dendrites, axons and synapses. It has been introduced in its mathematical form by McCulloch and Pitts in 1943. They published with Lettvin and Maturana the first works on the neural network in 1959 .

#### Principles

The basic ANN containing just two layers (input and output neurons) is described as the following steps:

(1) Choice of the inputs xi considering the output(s). For example, in the building field, we could have the energy consumption in the building as output with the following inputs: the outdoor temperature, the heating or cooling demand, the humidity and the solar radiation. Initially, each input is associated with a weight wi randomly chosen. The inputs are the neurons of the first layer.

(2) Application of the activation function f on the aggregation function. Most of the time, the aggregation function is a linear combination as:



n is the number of input neurons and the product for i = 0 is the bias. The activation function is responsible for converting the weighted input into an output activation (it is the analogy with a synapse). It returns a number between 0 and 1, allowing to maintain the convergence. It can be formalised by several non linear functions as the sigmoid, Heaviside step or hyperbolic function. More the number is close to 1, more the information is right. In contrast, when the activation function returns a number close to 0, the information is wrong. Fig (9) shows a scheme of one neuron layer where we can see the application of the activation function.

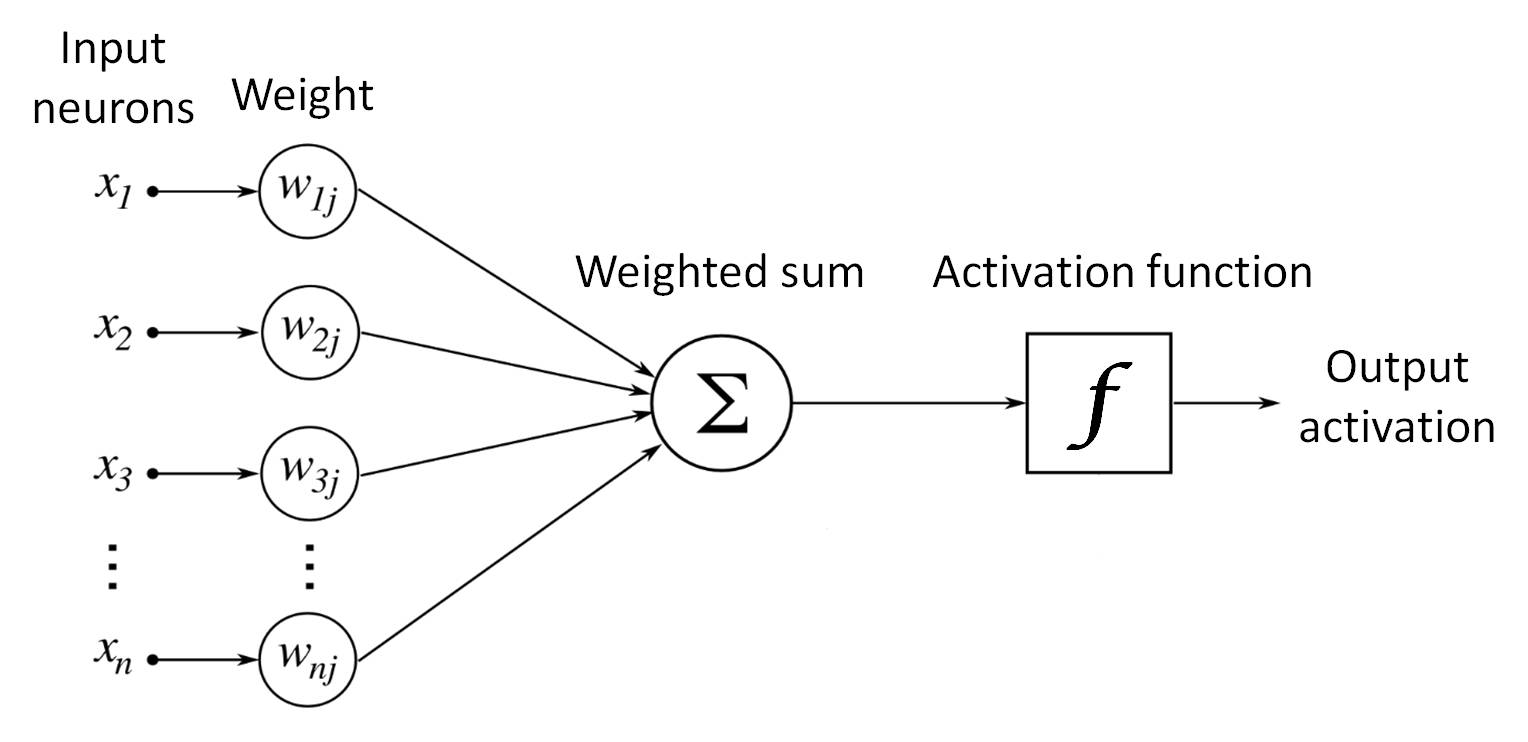


Figure 9: Scheme of one neuron layer with the application of the activation function

(3) Error calculation and application of the learning algorithm. The output is produced from the other steps. The global error corresponds to the sum of the training error calculated considering each data of the learning basis. To minimize the global error, a learning algorithm depending on a learning value is used to adjust the weight of each input neurons. The process is reiterated from step 2 to step 3 until reaching the error criterion. This description is available for a mono-layer ANN. Some details on multi-layer ANN are given in Annexe.

#### Advantages and limitations of the ANN

As we mentioned above, an advantage of the ANN is that it does not need to detect the potential collinearity. Moreover, given their training faculty, another advantage of the ANN is its ability to deduce from data the relationship between different variables without any assumptions or any postulate of a model. Furthermore, it overcomes the discretization problem and is able to manage the data unreliability. Finally, the ANN suggests a large variability of the predicted variable form (yes/no, binary 0 or 1, continuous value etc). However, the ANNs are significantly limited by the fact that this implies to have a relevant data basis. Indeed, it is really important to train an ANN with an exhaustive learning basis with representative and complete samples (for example, samples in different seasons or in different moments of the day or during week-end/holidays etc and samples with each the same amount of information). Another disadvantage of the ANN is its large number of undetermined parameters (with no rules to determine them).

#### Application field of the ANN

In the building simulation, the artificial neural network are usually used for the prediction of the energy consumption or the forecasting of energy use as the cooling or heating demand without knowing the geometry or the thermal properties of the building. Different kind of data basis can be considered depending on the time scale as the hour, the month or the year and the nature of the data (real or simulated and instantaneous or time/space-averaged data). One main condition is absolutely essential for applying the artificial neural network technique: the completeness of the learning data. Kalogirou has published many works on the building applications using the ANN. Particularly, in 2000, he presented a bibliographic review summing up the applications of the ANN in the field of energy-engineering systems . To illustrate that, we propose now to give some examples using the ANN for building applications.

#### Examples

Kalogirou and Bojic published a paper dealing with the prediction of the energy consumption of a passive solar holiday home in Cyprus during a day in summer and in winter. The inputs are the season, characteristics of the insulation, the masonry thickness, characteristics of the heat transfer coefficient and time of the day. The output is the energy consumption in kWh with a time-step of 10 minutes. The authors used a recurrent neural network containing four layers with 23 neurons on the hidden layers.

Aydinalp et al. studied the Canadian annual electricity consumption in residential sector of appliances, lighting and cooling in a first paper (ALC) [4], and of space heating (SH) and domestic hot water (DHW) in a second paper . In the first one, many inputs were used as appliances, weather, lighting, total heated area, socio-economic factors, etc. This information were propagated along a feedforward network containing one input layer with 55 neurons, three hidden layers each of them including 9 neurons, and one output layer with one neuron representing the average of the annual electricity consumption due to the ALC.

Neto and Fiorelli compared both an ANN model and a building software EnergyPlus for the forecasting of the energy demand in an administration building in Sao Paulo, Brazil. Two ANN model were tested: the first is a feed-forward neural network containing three layers : one input layer with 5 neurons (external temperature, humidity, two solar radiation parameters and daytype), one hidden layer with 21 neurons and one output layer with 1 neuron (daily total consumption). The second is a simpler ANN with only the external and internal temperature as inputs. The results for both simple ANN and complex ANN appeared to be very near, indicating that the humidity and the solar radiation were certainly less significant than the external temperature for the forecasting of energy demand in this specific building study.

Recently, Kwok and Lee [48] studied the influence of the occupancy on the cooling load in Hong-Kong, China. They compared three different neural networks called probabilistic entropy-based neural network (PENN) to predict the total building cooling load : a first ANN containing 6 neurons on the external layer each of them characterizing a weather parameter, a second ANN with one more neuron (so 7 external neurons in total) for the hourly total occupancy area and a third ANN with another one more neuron (so 8 external neurons in total) corresponding to the occupancy rate (modification induced by the human presence). They found the best fitting between real data and the prediction for the last model (with 8 external neurons). It shows the huge influence of the occupancy on the building cooling load.

Moreover, Escriva-Escriva et al. predicted the energy consumption based on building end-uses in University of Valencia, Spain. They used an ANN with multi-layer perceptron architecture consisting of three layers. The input layer contains four neurons (maximum temperature, minimum temperature, average temperature over just one day period and the average temperature of the previous day), the hidden layer contains 3 neurons and the output layer consists in one neuron characterizing the energy consumption. However, the ANN is hugely limited by its lack of interpretability and the fact that it requires a large amount of learning data and mainly a relevant and completeness data basis (that is no missing data in the data basis and same amount of information for each variable). The following technique overcomes these difficulties given that it supports heterogeneous data basis and introduces a describing function. This method is called the support vector machine.

### *Support vector machine (SVM)*

The support vector machine (SVM) has been introduced in 1995 by Vapnik and Cortes This artificial intelligence technique is usually used to solve classification and regression problems. We will focus our interest only on regression.

#### Principles

The principle of the SVM for regression is to find the optimal generalization of the model, in order to promote sparsity.

Let us consider a given training data [(x1, y1),…, (xn, yn)], xi being in the input space and yi in the output space. In a non linear problem, the basic idea is to overcome the non-linearity by transforming the non linear relation between x and y in a linear map. The way to do that is to send the non linear problem in a high dimensional space called the feature space. As all regression techniques, the aim is to determine the function f (x) that fits best the behaviour of the problem. These non linear functions are called kernel function. The particularity of the SVM is the fact that it authorizes an error or an uncertainty **ε** around the regression function.

#### Advantages and limitations of the SVM

The main difficulty in the SVM is to select the best kernel function corresponding to a dot product in the feature space and the parameters of this kernel function. The main advantage of the SVM is the fact that the optimization problem is based on the structural risk minimization principle (SRM). It deals with the minimization of an upper bound of the generalization error consisting of the sum of the training error. This principle is usually confronted to the empirical risk minimization (ERM) which only minimizes the training error. Another advantage is the fewer free parameters of optimization. Indeed, using the SVM technique required the adjustment of the regularization constant C and the margin **ε**. In contrast, the ANN method requires to know the topology of the inter-connections between neurons, the aggregation function, the number of hidden layers, the number of neurons on each hidden layers, the activation function, the learning algorithm (with the error calculation) and the learning value. In the same way, to implement a GA, we need to adjust the population size, the number of generation, the crossover probability and the mutation probability.

#### Application field of the SVM for regression

In the building field, the SVM is mainly used for the forecasting of energy consumption or temperature. The system can be trained from different kind of data with various time scale (year, month, hour) and various nature (instantaneous or space/time averaged). There is usually no restriction on the data basis except the fact that vector data are required. And a huge advantage is the fact that it supports a heterogeneous data basis that a data basis where all variables do not have the same amount of information or where we can find missing data. We will now on the use of SVM in building prediction, with some examples.

#### Examples

The use of support vector machine in the forecasting of energy consumption in buildings is quite recent. In 2005, Dong et al. were the first to use SVM for the prediction of the building energy consumption. The aim was to predict the monthly energy consumption in four offices in Singapore. The input variables are the mean outdoor dry-bulb temperature, the relative humidity and the global solar radiation. The kernel function used is the radial basis function kernel.

Li et al. used the SVM in regression for the prediction of hourly cooling demand in Guangzhou, China. The aim was to predict the cooling demand hour by hour during summer in an office building. The input parameters are the outdoor dry-bulb temperature, the relative humidity and the global solar radiation. The SVM used as the kernel function a radial basis function.

Moreover, Kavaklioglu used the support vector regression method to predict the electricity consumption in Turkey until 2026. The kernel function is the radial basis function. The input variables are socio-economic parameters as population, Gross National Product, imports and exports.

Likewise, Paniagua-Tineo et al. employed support vector regression method to model and predict the daily air outdoor temperature in several European countries. The model depends on many prediction variables as the maximum and minimum temperature, the precipitation, the relative humidity, the air pressure, the global radiation, the specific synoptic situation of the day and the so called monthly cycle. The kernel function is a Gaussian function.

Furthermore, Che et al. proposed to develop an adaptive fuzzy rule based prediction system combining the SVM in regression and a fuzzy inference method with the aim to forecast the electrical load in New South Wales. The authors used the radial basis function as kernel function.

Chen et al. estimated the monthly mean daily solar radiation in Chongqing, China via the support vector machine method. More particularly, the aim was to improve the state of data collected in the station. The authors chose to test three different kernel function: linear, polynomial and radial basis function. Also, they proposed to experiment seven combinations of input variables only based on the maximum temperature and the minimum temperature. Finally, they implemented 21 different SVM systems.

### *Discussion on the Statistical Tools*

Contrary to the physical techniques where each can be associated to a specific application, such classification is not possible with statistical tools. However, it is possible to classify them by complexity. Indeed, linear multiple regression is probably the easier statistical method. It is able to give good prediction and does not need real expertise to be implemented. However, it is greatly limited by the fact that it assumes a linear description of phenomena. The genetic algorithm is a less limited because it is able to treat both linear and non linear problems. But it suggests that the function describing the system behaviour is well-known. However, it is rarely the case. The artificial neural network overcomes this problem given that it does not need to give specific description. Nevertheless, it runs as a black-box system which makes the interpretability very difficult. Moreover, an important drawback of the ANN is the fact that it requires a large amount and a completeness of learning data. In contrast, the support vector machine has the huge advantage to do not need an exhaustive data basis. And due to the known kernel function, the problem remains interpretable. However, contrary to the artificial neural network, it requires the assumption of the form of the kernel function. Therefore, it can be considered each of these statistical techniques has its own advantages and drawbacks and that the choice of the method depends mainly on the intended usage and expectations.

## Bibliography - State of the Art Section

1. Aranda, G. Ferreira, M. Mainar-Toledo, S. Scarpellini, E.L. Sastresa, Multiple regression models to predict the annual energy consumption in the spanish banking sector, Energy and Buildings 49 (2012) 380–387.
2. A. Aswani, N. Master, J. Taneja, D. Culler, C. Tomlin, Reducing transientand steady state electricity consumption in hvac using learning based model-predictive control, Proceedings of the IEEE 100 (2012) No.1.
3. J. Axley, Multizone Airflow Modelling in Buildings: History and Theory, HVAC and Research 13 (2007) No. 6.
4. M. Aydinalp, V.I. Ugursal, A.S. Fung, Modeling of the appliance, lighting,and space- cooling energy consumptions in the residential sector using neural networks, Applied Energy 71 (2002) 87–110.
5. M. Aydinalp, V.I. Ugursal, A.S. Fung, Modelling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks, Applied Energy 79 (2004) 159–178.
6. M. Aydinalp-Koksal, V.I. Ugursal, Comparison of neural network, conditional demand analysis, and engineering approaches for modelling end-use energy consumption in the residential sector, Applied Energy 85 (2008) 271–296.
7. V. Badescu, Simple and accurate model for the ground heat exchanger of a passive house, Renewable Energy 32 (2007) 845–855.
8. V. Badescu, B. Sicre, Renewable energy for passive house heating II. Model, Energy and Buildings 35 (2003) 1085–1096.
9. V. Badescu, B. Sicre, Renewable energy for passive house heating part I. Building description, Energy and Buildings 35 (2003) 1077–1084.
10. C.A. Balaras, A.G. Gaglia, E. Georgopoulou, S. Mirasgedis, Y. Sarfidis, D.P. Lalas, European residential buildings and empirical assessment of the hellenic building stock, energy consumption, emissions and potential energy savings, Building and Environment 42 (2007) 1298–1314.
11. I.L. Balemans, Review of correlations for convective heat transfer from flat surfaces in air, Chaleur et Climats 590 (1987) 450.
12. S. Barlow, D. Fiala, Occupant comfort in UK offices - How adaptive comfort theories might influence future low energy office refurbishment strategies, Energy and Buildings 39 (2007) 837–846.
13. J.L. Bertez, The PASSIVE stake. Strategic overview on a global, structured and sustainable way for ”efficient building”, Zenergie (2009).
14. B. Bohm, P. Danig, Monitoring the energy consumption in a district heated apartment building in Copenhagen, with specic interest in the thermodynamic performance, Energy and Buildings 36 (2004) 229–236.
15. E. Carroll, E. Hatton, M. Brown, Residential energy use behavior change pilot, CMFS Project Code B21383 Franklin Energy (2009).
16. J. Che, J. Wang, G. Wang, An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting, Energy 37 (2012) 657–664.
17. J.L. Chen, H.B. Liu, W. Wu, D.T. Xie, Estimation of monthly solar radiation from measured temperatures using support vector machines e a case study, Renewable Energy 36 (2011) 413–420.
18. Q. Chen, Ventilation performance prediction for buildings: A method overview and recent applications, Building and Environment 44 (2009) 848–858.
19. T. Chow, J. Hand, P. Strachan, Building-integrated photovoltaic and thermal applications in a subtropical hotel building, Applied Thermal Engineering 23 (2003) 2035–2049.
20. J. Clarke, J. Cockroft, S. Conner, J. Hand, N. Kelly, R. Moore, T. OBrien, P. Strachan, Simulation-assisted control in building energy management systems, Energy and Buildings 34 (2002) 933–940.
21. J. Clarke, C. Johnstone, N. Kelly, P. Strachan, The simulation od photovoltaic-integrated building facades, Building Simulation (1997)
22. C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (1995) 273–297.
23. D.B. Crawley, J.W. Hand, M. Kummert, B.T. Grith, Contrasting the capabilities of building energy performance simulation programs, Building and Environment 43 (2008) 661–673.
24. F. Cron, C. Inard, R. Belarbi, Numerical analysis of hybrid ventilation performance depending on climate characteristics, International Journal of Ventilation 1 (2003) HybVent–Hybrid Ventilation Special Edition.
25. D. Datta, S.A. Tassou, D. Marriott, Application of neural networks for the prediction of the energy consumption in a supermarket, Proceedings of the Clima 2000 Conference Brussels (1997) Belgium.
26. M. Davies, Optimal designs for star circuits for radiant exchange in a room, Building and Environment 18 (1983) 135–150.
27. C. Dobrzynski, Adaptation de maillage anisotrope 3D et application à l aéro-thermique des bâtiments, Ph.D. thesis, Universit Pierre et Marie Curie-Paris 6, 2005.
28. B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy and Buildings 37 (2005) 545–553.
29. EnergyPlus, www.eere.energy.gov/buildings/energyplus, U.S. Department of Energy (DOE).
30. A. Entrop, H. Brouwers, A. Reinders, Evaluation of energy performance indicators and financial aspects of energy saving techniques in residential real estate, Energy and Buildings 42 (2010) 618–629.
31. EPBD, On the energy performance of buildings, Official Journal of the European Union Directive 2010/31/EU of the European Parliament and of the Council (2010).
32. G. Escriva-Escriva, C. Alvarez-Bel, C. Roldan-Blay, M. Alcazar-Ortega, New articial neural network prediction method for electrical consumption forecasting based on building end-uses, Energy and Buildings 43 (2011) 3112–3119.
33. G. Fraisse, C. Viardot, O. Lafabrie, G. Achard, Development of a simplified and accurate building model based on electrical analogy, Energy and Buildings 34 (2002) 1017–1031.
34. A.G. Gaglia, C.A. Balaras, S. Mirasgedis, E. Georgopoulou, Y. Sarafidis, D.P. Lalas, Empirical assessment of the Hellenic non-residential building stock, energy consumption, emissions and potential energy savings, Energy Conversion and Management 48 (2007) 1160–1175.
35. F. Gugliermetti, G. Passerini, F. Bisegna, Climate models for the assessment office buildings energy performance, Building and Environment 39 (2004) 39–50.
36. F. Haghighat, Y. Li, A.C. Megri, Development and validation of a zonal model POMA, Building and Environment 36 (2001) 1039–1047.
37. J. Holland, Adaptation in natural and artificial systems, Ann Arbor, 1975.
38. M. Ibanez, A. Lazaro, B. Zalba, L.F. Cabeza, An approach to the simulation of PCMs in building applications using TRNSYS, Applied Thermal Engineering 25 (2005) 1796–1807.
39. P. Ihm, A. Nemri, M. Krarti, Estimation of lighting energy savings from daylighting, Building and environment 44 (2009) 509–514.
40. M. Jaggs, J. Palmer, Energy performance indoor environmental quality retrofit a european diagnosis and decision making method for building refurbishment, Energy and Buildings 31 (2000) 97–101.
41. T.E. Jiru, F. Haghighat, Modelling ventilated double skin facadea zonal approach, Energy and Buildings 40 (2008) 1567–1576.
42. S.A. Kalogirou, Applications of artificial neural-networks for energy systems, Applied Energy 67 (2000) 17–35.
43. S.A. Kalogirou, Use of TRNSYS for modelling and simulation of a hybrid pvthermal solar system for cyprus, Renewable Energy 23 (2001) 247–260.
44. S.A. Kalogirou, M. Bojic, Artificial neural networks for the prediction of the energy consumption of a passive solar building, Energy 25 (2000) 479–491.
45. S.A. Kalogirou, C.C. Neocleous, C.N. Schizas, Building heating load estimation using artificial neural networks, Proceedings of Clima 2000 Conference (1997).
46. K. Kavaklioglu, Modelling and prediction of Turkey’s electricity consumption using support vector regression, Applied Energy 88 (2011) 368–375.
47. K. Klein, A transient simulation and program., Madison (WI): Solar Energy Laboratory (1996).
48. S.S. Kwok, E.W. Lee, A study of the importance of occupancy to building cooling load in prediction by intelligent approach, Energy Conversion and Management 52 (2011) 2555–2564.
49. G. Lafrance, D. Perron, Evolution of residential electricity demand by end-use in quebec 1979-1989: A conditional demand analysis, Energy Studies Review 6 (1994) Art.4.
50. P. Lagonotte, Y. Bertin, J.B. Saulnier, Analyse de la qualité de modèles nodaux réduits l’aide de la méthode des quadriples, International Journal of Thermal Sciences 38 (1999) 51–65.
51. P. Lauret, H. Boyer, C. Riviere, A. Bastide, A genetic algorithm applied to the validation of building thermal models, Energy and Buildings 37 (2005) 858–866.
52. J. Lettvin, H.R. Maturana, W.S. McCulloch, W.H. Pitts, What the frog’s eye tells the frog’s brain, Proc. Inst. Radio Engr. 47 (1959) 1940–1951.
53. Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, Applied Energy 86 (2009) 2249–2256.
54. Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks, Energy Conversion and Management 50 (2009) 90–96.
55. Y. Li, A. Delsante, J. Symons, Prediction of natural ventilation in buildings with large openings, Building and Environment 35 (2000) 191– 206.
56. Z. Liao, A. Dexter, A simplied physical model for estimating the average air temperature in multi-zone heating systems, Building and Environment 39 (2004) 1013–1022.
57. L. Magnier, F. Haghighat, Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and artificial neural network, Building and Environment 45 (2010) 739–746.
58. T.A. Mara, F. Garde, H. Boyer, M. Mamode, Empirical validation of the thermal model of a passive solar cell test, Energy and Buildings 33 (2001) 589–599.
59. T.P. McDowell, J.W. Thornton, S. Emmerich, G. Walton, Integration of airflow and energy simulation using CONTAM and TRNSYS, ASHRAE Transactions 109 (2003) KC–03–10–2.
60. F.W. Miranville, H. Boyer, T. Mara, F. Garde, On the thermal behavior of roof-mounted radiant barriers under tropical and humid climatic conditions: modelling and empirical validation, Energy and Buildings 35 (2003) 997–1008.
61. M. Mossolly, K. Ghali, N. Ghaddar, Optimal control strategy for a multizone air conditioning system using a genetic algorithm, Energy 34 (2009) 58–66.
62. M. Musy, F.Winkelmann, E.Wurtz, A. Sergent, Automatically generated zonal models for building air ow simulation: principles and applications, Building and Environment 37 (2002) 873–881.
63. A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy and Buildings 40 (2008) 2169–2176.
64. R. Ooka, K. Komamura, Optimal design method for building energy systems using genetic algorithms, Building and Environment 44 (2009) 1538–1544.
65. A. Oppenheim, Radiation analysis by the network method, ASME Transactions 78 (1956) 725–735.
66. M. Ordenes, R. Lamberts, S. Guths, Estimation of thermophysical properties using natural signal analysis with heat and moisture transfer model, Energy and Buildings 41 (2009) 1360–1367.
67. H.K. Ozturk, H. Ceylan, O.E. Canyurt, A. Hepbasli, Electricity estimation using genetic algorithm approach: a case study of turkey, Energy 30 (2005) 1003–1012.
68. A. Paniagua-Tineo, S. Salcedo-Sanz, C. Casanova-Mateo, E. Ortiz- Garcia, M. Cony, , E. Hernandez-Martn, Prediction of daily maximum temperature using a support vector regression algorithm, Renewable Energy 36 (2011) 3054–3060.
69. B. Paris, J. Eynard, S. Grieu, M. Polit, Hybrid pid-fuzzy control scheme for managing energy resources in buildings, Applied Soft Computing 8 (2011) 5068–5080.
70. B. Paris, J. Eynard, S. Grieu, T. Talbert, M. Polit, Heating control schemes for energy management in buildings, Energy and Buildings 42 (2010) 1908–1917.
71. M. Parti, C. Parti, The total and appliance-specific conditional demand for electricty in the household sector, The Bell Journal of Economics 11 (1980) 309–321.
72. S. Petersen, S. Svendsen, Method for simulating predictive control of building systems operation in the early stages of building design, Applied Energy 88 (2011) 4597–4606.
73. B. Poel, G. van Cruchten, C.A. Balaras, Energy performance assessment of existing dwellings, Energy and Buildings 39 (2007) 393–407.
74. Proposal-EPBD, Low energy building in Europe: current state of play, definitions and best practice (2009).
75. M. Qin, G. Walton, R. Belarbi, F. Allard, Simulation of whole building coupled hygrothermal airflow transfer in different climates, Energy Conversion and Management 52 (2011) 1470–1478.
76. P. Rumaniovski, J. Brau, J.J. Roux, An adapted model for simulation of the interaction between a wall and the building heating system, Proceedings of the thermal performance of the exterior envelopes of buildings IV conference Orlando, USA (1989) 224–233.
77. H. Sadeghi, M. Zolfaghari, Mohamad.Heydarizade, Estimation of electricity demand in residential sector using genetic algorithm approach, International Journal of Industrial Engineering and Production Research 22 (2011) 43–50.
78. V. Siddharth, P. Ramakrishna, T. Geetha, A. Sivasubramaniam, Automatic generation of energy conservation measures in buildings using genetic algorithms, Energy and Buildings 43 (2011) 2718–2726.
79. J. Srebric, J. Yuan, A. Novoselac, On-site experimental validation of a coupled multizone and cfd model for building contaminant transport simulations, ASHRAE Collections NY-08-033 (2008).
80. L. Stephan, A. Bastide, E. Wurtz, Optimizing opening dimensions for naturally ventilated buildings, Applied Energy 88 (2011) 2791–2801.
81. G. Tan, L.R. Glicksman, Application of integrating multi-zone model with CFD simulation to natural ventilation prediction, Energy and Buildings 37 (2005) 1049–1057.
82. J. Teeter, M.Y. Chow, Application of functional link neural network to HVAC thermal dynamic system identification, IEEE Transactions on Industrial Electronics 45 (1998) NO. 1.
83. K. Tiedermann, Using conditional demand analysis to estimate residential energy use and energy savings, European Council for an Energy Efficient Economy Summer Study 6 (2007) 1279–1283.
84. P. Tittelein, E. Wurtz, G. Achard, Simspark platform evolution for low-energy building simulation, International Scientific Journal for Alternative Energy and Ecology 62 (2008).
85. M. Trocmé, Aide au choix de conception de bâtiments économes en énergie, Ph.D. thesis, Ecole Nationale Supérieure des Mines de Paris, 2009.
86. V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.
87. M. Wall, Distribution of solar radiation in glazed spaces and adjacent buildings. a comparison of simulation programs, Energy and Buildings 26 (1997) 129–135.
88. L. Wang, N.H. Wong, Coupled simulations for naturally ventilated residential buildings, Automation in construction 17 (2008) 386–398.
89. S.Wang, Dynamic simulation evaluation of building vav air-conditioning system and of emcs on-line control strategies, Building Simulation 34 (1999) 681–705.
90. S. Wang, X. Xu, Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm, Energy Conversion and Management 47 (2006) 1927–1941.
91. S. Wang, X. Xu, Simplified building model for transient thermal performance estimation using ga-based parameter identification, International Journal of Thermal Sciences 45 (2006) 419–432.
92. M. Woloszyn, C. Rode, Tools for performance simulation of heat, air and moisture conditions of whole buildings, Building Simulation 1 (2008) 5– 24.
93. E. Wurtz, L. Mora, C. Inard, An equation-based simulation environment to investigate fast building simulation, Building and Environment 41 (2006) 1571–1583.
94. Z. Zhai, Q. Chen, P. Haves, J.H. Klems, On approaches to couple energy simulation and computational fluid dynamics programs, Building and Environment 37 (2002) 857–864.
95. Z.J. Zhai, M.H. Johnson, M. Krarti, Assessment of natural and hybrid ventilation models in whole-building energy simulations, Energy and Buildings 43 (2011) 2251–2261.
96. E. Znouda, N. Ghrab-Morcos, A. Hadj-Alouane, Optimization of Mediterranean building design using genetic algorithms, Energy and Buildings 39 (2007) 148–153.