Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 126 (201709) 250–257

Energy Procedia

www.elsevier.com/locate/procedia

72nd Conference of the Italian Thermal Machines Engineering Association, ATI2017, 6-8 September 2017, Lecce, Italy

On the use of artificial neural networks to model household energy consumptions

F. Martellotta[a,*], U. Ayr[a], P. Stefanizzi[a], A. Sacchetti[b], G. Riganti[b]

[a]*DICAR – Politecnico di Bari, via Orabona 4, Bari, 70125, Italy*
[b]*Tera s.r.l., via C. Balbo 19/D, Conversano, 70014, Italy*

Abstract

Modern houses are more and more frequently characterized by the presence of “smart” metering devices, capable of measuring air temperature, relative humidity, air quality, and in the more sophisticated cases, even electric equipment consumptions. In addition, other relevant parameters such as illuminance may often be determined and they can be used as proxy variables to account for other important aspects (such as solar irradiance) influencing the energy balance of a building. Such information, in combination with weather data which can be retrieved by other sources (or by additional sensors), may conveniently contribute to the creation of a “black box” model in which, given a few input variables it is possible to output a variable which would result from otherwise complex calculations (e.g. an energy balance) requiring many data. The availability of such a “black box” could be helpful under many points of view, such as benchmarking energy consumptions and stimulating virtuous behavior from the occupants. To test whether such approach can be feasible, an EnergyPlus model of a real house was made, trying to accurately reproduce building features, systems set-points, and occupant behaviors. The overall simulated energy consumptions were compared with the real ones resulting from energy bills, thus ensuring a good agreement with reality. The dataset resulting from EnergyPlus was then used to train an artificial neural network (ANN) capable of yielding hourly energy consumptions based on limited input data. Finally, the relative importance of the different input variables was analyzed to understand which might influence prediction accuracy most.

Peer-review under responsibility of the scientific committee of the 72nd Conference of the Italian Thermal Machines Engineering Association

Keywords: Artificial Neural Network (ANN); Energy consumptions prediction; Smart houses; Energy saving

* Corresponding author. Tel.: +39-080-5963631;
E-mail address: francesco.martellotta@poliba.it

Peer-review under responsibility of the scientific committee of the 72nd Conference of the Italian Thermal Machines Engineering Association
10.1016/j.egypro.2017.08.149

1. Introduction

Reduction of energy consumptions is one of the main objectives that need to be addressed in order to limit global warming. Among the different energy uses, the building sector is responsible of a 40% share of the total [1], mostly because of poor envelope characteristics and inefficient HVAC systems. National and international regulations [2] are trying to stimulate a widespread improvement of the abovementioned characteristics, also by means of substantial economic incentives with particular reference to retrofitting existing buildings. However, even when the building and its systems comply with regulations, the actual energy consumption may be hard to predict and, in many cases, the use of conventional approaches may lead to surprising, and contradictory, results [3]. The role of the occupants' behavior appeared in the last years as the key issue to understand in order to get reliable predictions [4], thus suggesting that any effective energy-saving policy needs to involve and engage the final user. In large buildings such problems are solved by means of a building energy management system (BEMS), but in residential buildings it seems unlikely a widespread use of such solutions. On the other hand, the spreading of low cost devices capable of monitoring indoor environmental parameters as well as energy consumptions, combined with the use of more numerous and more powerful portable devices, might significantly contribute to help the occupants to adopt energy efficient lifestyles.

From this point of view, one critical point is understanding how the complex energy flux of a building can be analyzed in order to give the occupants information about the way they use energy. The most obvious answer could be that of using calculation methods complying with the national regulations, but that would imply a collection of data that could be impossible to gather for a non-expert user. Similarly, any dynamic simulation tool (like the well-known EnergyPlus) would similarly require very detailed information about the building, its systems, and the user behavior in order to be accurate. Therefore, a "black-box" approach using a limited number of input variables to predict the energy consumption seems more feasible. The key to the success of such approach is a suitable modeling of the black box behavior. Conventional statistical methods, such as multiple regression analysis, could be used in this case, but growing evidences suggest that artificial neural networks (ANN) and machine learning (ML) techniques may do the job in a more reliable and efficient way [5-12]. The success of ANNs depends on five distinctive features: learning, self-adaptive, fault tolerance, flexibility and real time response. In addition, ANNs can manage complex and ill-defined problems because of their strong nonlinear mapping ability. Neural network models can realize any nonlinear mapping between the input and output, and there is no need to know the mathematical equation describing the load and the influence factors in advance. Thus, it has been popularly applied to predict building energy consumption. Current applications involve investigating the potential of dynamically simulating energy demand of a building using a limited input set [6], dynamically controlling HVAC systems to achieve comfort and energy savings [7,9], predicting electric energy consumptions [8], predicting heat demand based on natural gas consumption [10], standard energy performance of a building using its features as inputs [11,12], and many others. As outlined before, the idea that is investigated in this paper is that of using data potentially available from low-cost monitoring devices, combined with outdoor weather data, to predict optimal energy use for current conditions in a given house and identify any anomalous conditions.

2. Methods

In order to investigate the potential offered by the above mentioned devices and understand whether the adoption of a black-box approach in modelling energy consumptions of a given house is feasible or not, a validated EnergyPlus model of a real house was used to generate the dataset required to train the ANN. Validation was carried out using actual energy consumptions retrieved from energy bills related to a four-year period and by carefully matching user behaviors with the relevant schedules used in the simulation tool.

2.1. The case study

The apartment used to create the EnergyPlus model is located in Bari (Italy), in a densely populated neighborhood; it is at the first floor of a 5-story building built in 1960. The overall floor surface is 80 m^2 and the internal height is 2.9 m. Its East and West walls are shared with other apartments and with the stairway that is usually fully ventilated. The dividing walls are made of 20 cm thick Tufa blocks, with plaster on both faces. Internal walls are made of 10 cm thick tufa blocks with plaster on both sides. Floor and ceiling are 30 cm thick and are made of concrete and hollow clay blocks covered with marble-like stone. Heat exchange takes place through the North and South walls, made of two layers of Tufa blocks (15 cm thick) separated by a 5 cm air gap (U=1.39 W/m^{2}K), and through the west wall facing the stairway which is made of Tufa blocks (15 cm thick) and clay blocks (10 cm thick) with no air gap (U=1.52 W/m^{2}K). Windows were completely replaced in 2013 using a 70 mm PVC frame (Uf=1.2 W/m^{2}K) and a glazing system made of two 4 mm panes, divided by a 20 mm air gap (Ug=2.7 W/m^{2}K). Thermally insulated external roller blinds are used as shading devices.

The heating system is made of a condensing boiler (P=4.4/24.5 kW, η_{max} = 106.5%) using methane gas as primary energy vector, and normally operating at a temperature of 55°C. Heated water is then distributed to a set of low-temperature tubular radiators designed to output an overall 3.6 kW power, assuming a temperature difference air and water of 30°C. Heating and ventilation schedules were set according to occupants' typical behavior. Heating is normally turned on from November 15 to March 31, 6:00 to 7:30, and 17:00 to 20:00 with set-point temperature of 20°C, and 20:00 to 23:00 with a set-point temperature of 21.0°C. During weekends, there is no morning heating, and the system starts at 15:00 instead of 17:00.

Ventilation is carried out daily 7:30 to 8:30 (9:00 to 11:00 during weekends) by opening the bedroom window facing North (tilt-in mode, opening area of 0.36 m^2) in combination with the kitchen window (tilt-in mode, opening area of 0.34 m^2) and the bathroom window (tilt-in mode, opening area of 0.08 m^2), both facing South. The latter window remains open until 14:00. In addition, there are two ventilation openings in the kitchen with an overall surface of 0.02 m^2. Additional opening areas due to cracks around windows, roller blinds boxes, and doors was estimated as 0.01 m^2 on each façade.

Internal loads due to the two occupants, lighting, and equipment was determined considering the actual power of the installed devices. Surrounding buildings are located at a distance of 10 m from the North façade and 8 m from the South facade, but while buildings on the North side have the same height of 18 m in excess of the apartment floor, those facing South present a more varied outline which was carefully considered in the modelling (Fig. 2).

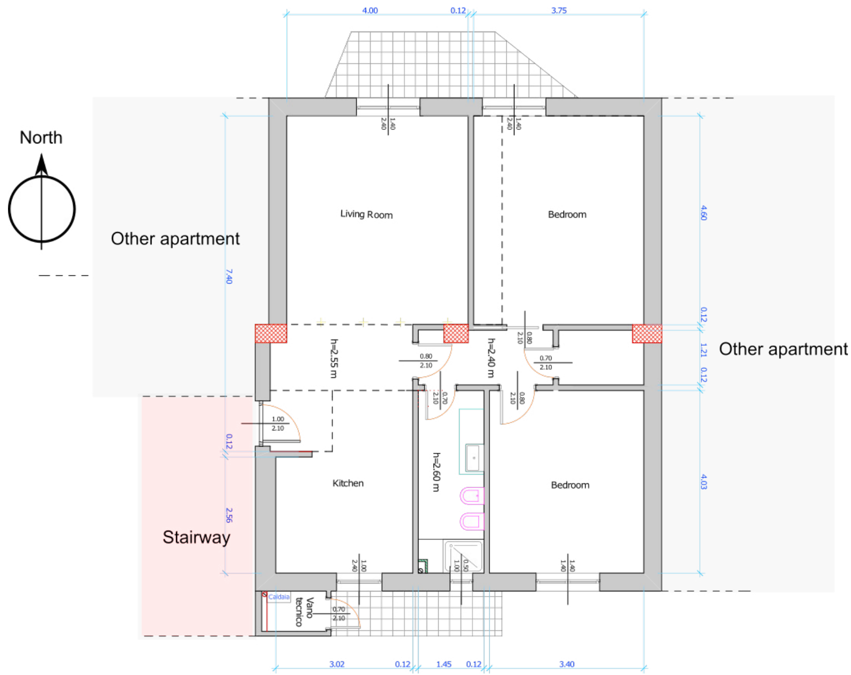

Fig. 1. Plan of the apartment used to create the EnergyPlus model

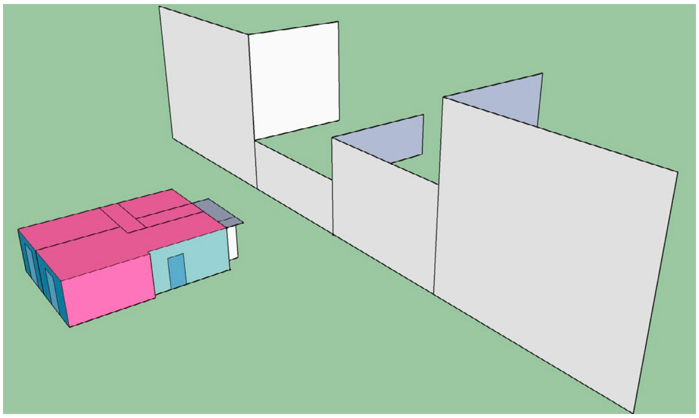

Fig. 2. 3D model of the EnergyPlus model including the surrounding buildings that shade the southern facade

Heating energy consumption was derived from actual energy bills referred to a period of about four years. First the daily consumption was calculated with reference to months without heating and a grand average was calculated excluding (also excluding summer holiday periods). The resulting value was 0.2 Nm^3/day with a standard variation of ±0.045 Nm^3/day. This average value was then used to derive the consumption due to heating during the corresponding months. For each reference period (between subsequent counter readings), the number of days in which heating was expected to be turned on was calculated and the average heating consumptions was finally determined. The result was 1.90 Nm^3/day, with a standard deviation of ±0.33 Nm^3/day due to a variation between a minimum of 1.59 Nm^3/day and a maximum of 2.42 Nm^3/day. Such variations proved to be well related to actual heating degree-days of the corresponding heating period retrieved from historical weather data [13]. So, considering the average daily value and multiplying by the number of days in the conventional heating period, an annual demand of 260±45 Nm^3 of methane gas is finally obtained. Considering that the heating degree-days for the adopted weather file amount to 1441, and that during the four surveyed years the actual degree-days were 1148, 1403, 1218, and 1453, results are in good agreement. For electric consumptions, the 3-year average (always retrieved from bills) returned 1200 kWh per year, including lighting, equipment, and cooling.

2.2. The EnergyPlus model

Considering the characteristics of the apartment, a 3D model was made in SketchUp using the OpenStudio plugin, and subsequently exported to EnergyPlus v. 8.6 in order to perform the dynamic energy analysis. Floor, ceiling, and the side walls shared with other apartments were modelled as adiabatic surfaces. However, although not involved in heat exchange their density, heat capacity and conductivity were provided in order to take into account their contribution to internal mass and heat storage. The wall facing the stairway was modelled as an outdoor wall, but with no sun and wind exposure, so to take into account that the stairway is a fully ventilated indoor space. The remaining surfaces were modelled according to the physical characteristics described above. The effect of surrounding buildings was modelled only on the South façade, where the shading effects were more significant compared to the North façade which, being in front of a long tall building was considered to receive no sun at all. As final results proved to be very sensitive to ventilation and infiltration loads, they were carefully modelled taking into account the actual opening area using the “WindAndStackOpenArea” object, introducing actual opening times using a specific schedule. For ventilation openings and windows cracks no schedule was applied.

In order to determine the heating and cooling energy consumptions in a simple and straightforward way, and also avoid making assumptions on more detailed plant characteristics, an “IdealLoadAirSystem” with no outdoor air was considered. This EnergyPlus object returns both the heating and cooling energy required to meet the temperature set-points that have been provided. To have a more realistic behavior, it is important to set both heating and cooling limits according to the characteristics of the terminals. In this case, the maximum sensitive heating capacity was set to 3.6 kW, in agreement with the characteristics of the radiators. The system was designed according to the actual schedule and actual temperature set points. As the IdealLoadAirSystem returns exactly the thermal energy that must be provided, it is necessary to make some assumptions about the efficiency of the energy conversion.

Among the different output variables that can be returned by the software, hourly values of those more likely to be monitored using low-cost tools were selected for the subsequent analysis: mean indoor air temperature, mean radiant temperature, relative humidity, carbon dioxide (CO_2) concentration, mean illuminance (estimated at the center of the living room area), together with lighting and equipment energy. With reference to outdoor conditions, those more frequently available from weather stations were considered: dry-bulb air temperature, relative humidity, rain depth, and wind speed. Intensity of solar radiation, although playing an important role on energy balance, was not considered directly, as relatively few weather stations can provide this value, but illuminance was considered as a good proxy variable. Finally, hourly values of total heating energy (taken "as is" and expressed in kWh) were also retrieved to be used as target variables in the subsequent analysis.

With reference to the weather conditions, data taken from a large and homogeneous dataset were preferred. Consequently, the IWEC2 (International Weather for Energy Calculations) database developed by ASHRAE within the Research Project RP-1477, "Development of 3012 Typical Year Weather Files for International Locations" [13] was preferred among others. All the analyses were carried out using the climate data for Bari/Palese Macchie.

2.3. The artificial neural network

The ANN was implemented using the neural network toolbox in Matlab [14]. To learn the parameters of the ANN (i.e. the weights between neurons and biases) the network training function was Levenberg-Marquardt backpropagation algorithm [15]. A two-layer feed-forward network with sigmoid hidden neurons and linear output neurons was used. The estimation of the number of neurons in each layer is one of the most difficult task, which is generally carried out using a trial and error procedure. In this case 20 neurons were used as a starting point. Hourly values of total heating energy were used as target values. Hourly values of the ten previously mentioned variables were used as input. In addition, in order to provide the network further elements, hour of the day, and day of the week were also added, together with a running mean of both the indoor and outdoor air temperature in the previous 12 hours (to take into account dynamic behavior of the building). Only values referred to conventional heating days were considered for the analysis resulting in 3288 samples.

To estimate the ANN performance traditional metrics like mean square error (MSE) and regression coefficient R were used. In order to investigate the minimal set of input parameters capable of providing a reasonably accurate estimate of the energy consumptions, several sub-sets of the input variables were also considered.

3. Results

First of all, the results of the validation of the EnergyPlus model are presented. Comparison was made taking into account annual average consumptions. As mentioned above, for natural gas a mean value of 260±45Nm^3 was found, considering that the lower heating value of Methane is 35.16 MJ/Nm^3, and assuming a 99% efficiency for the boiler (as measured during periodic maintenance tests), and 81% for the sub-systems [16], it yields 2.04±0.35 MWh/yr of thermal energy. Compared with the estimated 2.01 MWh/year and considering both the approximations used in the model and the yearly variations, the 1.5% error seems negligible. For electricity consumption, the error is even smaller, being 4.2%. In both cases, the model underestimates actual values.

The performance of the ANN, as described in Fig. 3 is actually impressive. MSE referred to the testing set (15% of the overall sample) when using 70% to train the network is 0.00242 kWh and regression coefficient is 0.999.

Table 1. Summary of the validation of the EnergyPlus model

An example of a column heading	Actual consumptions	Estimated consumptions	Difference
	[MWh/yr]	[MWh/yr]	[%]
Thermal energy (heating)	2.04	2.01	+1.5%
Electricity (lighting, equipment, air cond.)	1.20	1.15	–4.2%

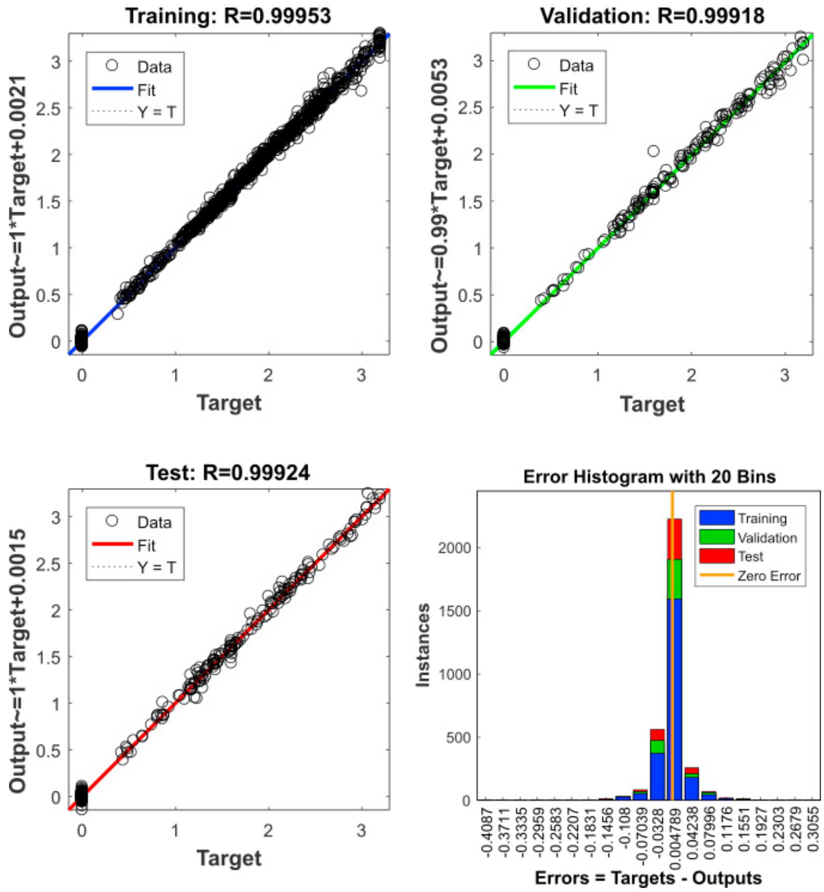

Fig. 3. Regressions resulting from each sub-set of data: training, validation, and test, and histogram of the distribution of absolute errors

The absolute errors obtained by comparing target and output values of hourly heating energy were generally within a ±0.1 kWh range (on a hourly base), which is a very good result considering that target values spanned between 0 and 3.2 kWh. A more detailed analysis of the relative errors showed (Fig. 4) that, excluding the cases in which the target value was zero, errors were within ±5% in 92% of the cases. The largest errors were of about 25% and appeared only in a few cases. Detailed analysis of such cases showed that they mostly correspond to the starting time of the heating system.

Reducing the number of samples used to train the network to 30% (984) yielded a MSE=0.004 kWh and R=0.9978. A further test was made by using data pertaining to one single month (January) to train the ANN, and then test it on the subsequent months. Again, results were good, with MSE=0.0064 kWh, and R=0.9963, thus suggesting that even a short-term training may provide acceptable results when all the input variables are included.

At this point it was important to understand whether input variables equally contributed to the accuracy of the network. In order to investigate this aspect, a sensitivity analysis was carried out. Three different samples of input variables were considered, including an hour in which system is off, one in which it is starting, and one in which system is approaching stationary conditions. Each input variable was changed by ±20% and the resulting output variation was analyzed. Results showed that outdoor relative humidity, rain depth, and wind speed had limited influence on the results. Among the indoor parameters, those showing limited influence on the output were relative humidity and illuminance. Hour and day number also proved of limited usefulness. Conversely, mean radiant temperature, indoor air temperature and its running mean proved to have the highest influence.

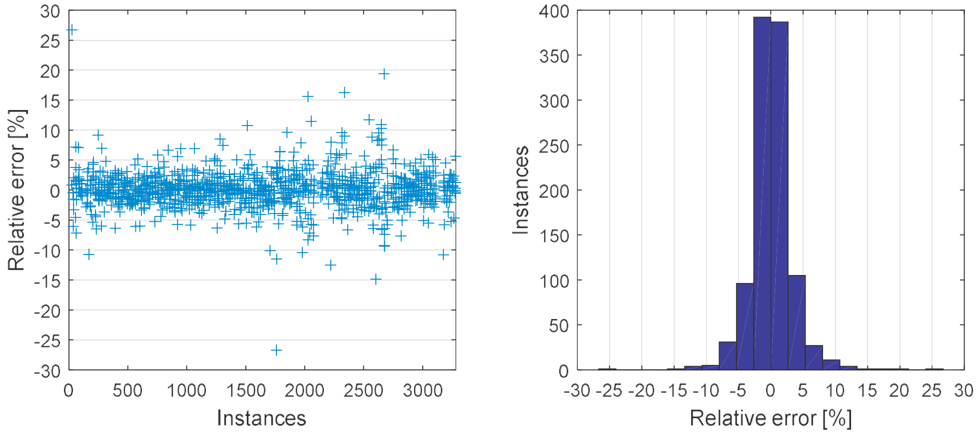

Fig. 4. Analysis of relative errors between ANN output and the corresponding EnergyPlus results considering the whole sample of data (heating season). Cases having target values equal to zero were not included in the plot.

Taking into account previous results, the predictive potential of a reduced set of input values where the above mentioned variables were removed was investigated. Results confirmed that the ANN obtained using 70% of the whole sample for training, and 15% for testing, had a MSE=0.015 kWh and R=0.999, with a slightly worsened performance. Analysis of relative errors also confirmed that a very limited number of cases exceeded a 10% error. A comparison of target (EP) and predicted (ANN) values of hourly heating energy, showed (Fig. 5) that differences were usually negligible and that ANN accurately predicted on/off cycles, also including daily variations (e.g. different schedule between working days and weekend).

A final test was performed removing also mean radiant temperature, which is difficult to measure without specialized equipment, and joining all electricity consumptions in a single variable, which could be more easily measured. Under these conditions, using 70% of the whole sample to train the ANN, and 15% for testing, MSE remained nearly the same as before, while R "dropped" to 0.992. However, relative errors got bigger, with maxima approaching 60% variation in a few cases, and only 42% of the test data within the ±5% limit and 70.8% within the ±10%. Finally, removing even electricity consumptions and CO_2 concentration raised MSE to 0.082, while R dropped to 0.955, suggesting that having variables that account for internal heat gains and for air exchange may considerably improve the accuracy of the system.

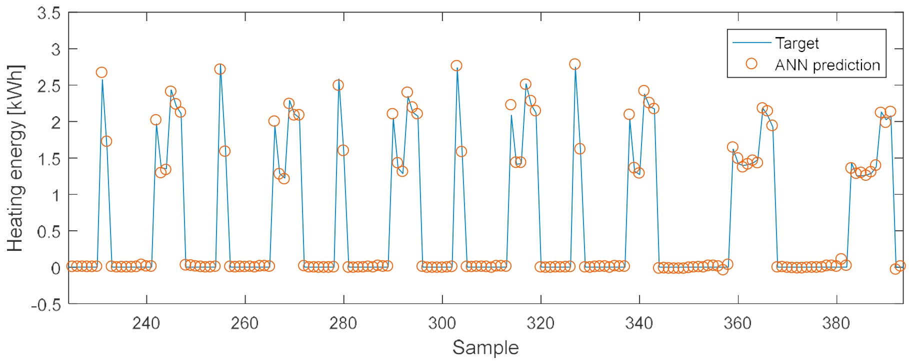

Fig. 5. Comparison between EnergyPlus and ANN energy consumption data referred to a typical week. ANN was trained using the reduced dataset (8 input variables).

4. Conclusions

The application of ANN to predict hourly energy consumptions for heating with reference to a house simulated using EnergyPlus proved very effective. Apart from ANN performance metrics, which, due to the large number of data, always offer encouraging results, more interesting results were obtained by analyzing the relative errors. With reference to the ANN using all the initial 15 input variables, such relative errors are generally very low, rarely exceeding a ±5% tolerance. The largest variations (in the range between 15% and 30%) appear when the heating system turns on, likely because of the dependence of the transient conditions on a larger number of factors. However, this behavior appeared only on limited occasions. A sensitivity analysis demonstrated that many of the input variables initially selected offer limited contribution to the ANN, so a simplified version, using only 8 input variables was tested. Results differed very little from those obtained with the initial set, confirming that using such reduced set may be a good choice. Any further change in the input variables had major effects on the prediction accuracy, with relative errors getting larger. Finally, removing mean radiant temperature and joining electricity consumptions in a single variable yielded relative errors approaching ±60%, with only 42% of the sample within the ±5% range and 70% within the ±10% range. This means that ANN may still provide an acceptable prediction of hourly energy consumptions for heating, but in case this prediction is used to benchmark actual consumptions a larger "alert" threshold should be adopted. From this point of view, further studies are required in order to clarify the mutual interaction of the different input variables, the ANN being clearly non-linear, as demonstrated by sensitivity analysis. In addition, the same set of data could be used to train other ANN to monitor other aspects, such as windows opening, temperature setpoint, etc., thus assisting the occupants in an energetically responsible management of the house.

References

[1] European Environment Agency (2012) "Annual European Union greenhouse gas-inventory 1990-2010 and inventory report".

[2] EPBD - Energy Performance of Buildings Directive 2010/31/CE of the European Parliament and of the Council on the Energy Performance of Buildings,18 June 2010.

[3] Majcen, D., Itard, L.C.M., Visscher, H., (2013) "Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications" *Energy Policy*, 54 (2013): 125-136.

[4] Andersen, R.K. , Fabi, V., Corgnati, S.P., (2016) "Predicted and actual indoor environmental quality: Verification of occupants' behaviour models in residential buildings" *Energy and Buildings,* 127(2016): 105-115.

[5] Deb, C., Eang, L.S., Yang, J., Santamouris, M. (2016) "Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks" *Energy and Buildings*, 121(2016): 284-297.

[6] Neto, A.H. and Fiorelli, F.A.S. (2008) "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption" *Energy and Buildings*, 40.12 (2008): 2169-2176.

[7] Moon, J.W. and Kim, J.-J. (2010) "ANN-based thermal control models for residential buildings" *Building and Environment*, 45.7 (2010): 1612-1625.

[8] Pombeiro, H., Santos, R., Carreira, P., Silva, C., Sousa, J.M.C., (2017) "Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks" *Energy and Buildings 146*(2017): 141-151.

[9] Ferreira, P.M., Ruano, A.E., Silva, S., Conceicao, E.Z.E. (2012) "Neural networks based predictive control for thermal comfort and energy savings in public buildings", *Energy and Buildings* 55(2012): 238-251.

[10]Spoladore, A., Borelli, D., Devia, F., Mora, F., Schenone, C, (2016) "Model for forecasting residential heat demand based on natural gas consumption and energy performance indicators", *Applied Energy* 182(2016): 488-499.

[11]Buratti, C., Barbanera, M. Palladino, D. (2014) "An original tool for checking energy performance and certification of buildings by means of Artificial Neural Networks" *Applied Energy* 120 (2014): 125-132.

[12]Buratti, C., Orestano, F.C., Palladino, D. (2016) "Comparison of the Energy Performance of Existing Buildings by Means of Dynamic Simulations and Artificial Neural Networks" *Energy Procedia* 101 (2016): 176-183.

[13]ASHRAE (2012) "International Weather for Energy Calculations (IWEC Weather Files) Version 2.0".

[14]MATLAB R.2016b, Neural Network Toolbox.

[15]Marquardt, D., (1963) "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM Journal on Applied Mathematics* 11.2 (1963): 431–441.

[16]Decreto interministeriale 26 giugno 2015 - Applicazione delle metodologie di calcolo delle prestazioni energetiche e definizione delle prescrizioni e dei requisiti minimi degli edifici