International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:6, No:5, 2012

Effects of Energy Consumption on Indoor Air Quality

M. Raatikainen, J-P. Skön, M. Johansson, K. Leiviskä and M. Kolehmainen

***Abstract*—**Continuous measurements and multivariate methods are applied in researching the effects of energy consumption on indoor air quality (IAQ) in a Finnish one-family house. Measured data used in this study was collected continuously in a house in Kuopio, Eastern Finland, during fourteen months long period. Consumption parameters measured were the consumptions of district heat, electricity and water. Indoor parameters gathered were temperature, relative humidity (RH), the concentrations of carbon dioxide (CO_2) and carbon monoxide (CO) and differential air pressure. In this study, self-organizing map (SOM) and Sammon's mapping were applied to resolve the effects of energy consumption on indoor air quality. Namely, the SOM was qualified as a suitable method having a property to summarize the multivariable dependencies into easily observable two-dimensional map. Accompanying that, the Sammon's mapping method was used to cluster pre-processed data to find similarities of the variables, expressing distances and groups in the data. The methods used were able to distinguish 7 different clusters characterizing indoor air quality and energy efficiency in the study house. The results indicate, that the cost implications in euros of heating and electricity energy vary according to the differential pressure, concentration of carbon dioxide, temperature and season.

***Keywords*—**Indoor air quality, Energy efficiency, Self-organizing map, Sammon's mapping

I. Introduction

INDOOR Air Quality (IAQ) is a relatively widely researched topic, because of its manifold effects on health of occupants. Increased interest in energy efficiency is thought to affect adversely on indoor air quality. For instance, in Science (academic journal of the American Association for the Advancement of Science), there are articles concerning of use and extent of smart grids for energy efficiency [1], sustainability [2], and the relationships between environment and healthiness [3]. In Nature (the world's most cited interdisciplinary scientific journal by the Science Edition of the 2010 Journal Citation Reports) there are many discussions concerning low-energy buildings and their relation to carbon emissions [4], as well as on the use of biological indicators for IAQ [5].

Neural networks have been used widely in the prediction of indoor air quality e.g. feedforward backpropagation [6, 7], recurrent neural networks [8], fuzzy neuro systems [9] and model comparison [10]. Also, studies have been done concerning with forecasting outdoor air quality parameters [11, 12] and air pollution episodes [13] using computational methods.

An international journal, Energy and Buildings, publishes articles with explicit links to energy use in buildings. One of the papers deals with energy performance, energy classification and rating the global environmental quality of school buildings [14]. There are also papers reviewing building energy savings, indoor air quality related standards [15] and enhanced supervision strategies for the reduction of building consumptions [16].

An article in Renewable Energy concerns with assessing the energy performance, based on the monitoring data of the school building, in Greece, in the climate zone with the lowest air temperature during the winter period [17]. Also, using a single compartment of IAQ model, the optimum parameters in terms of IAQ and energy consumption have been determined [18].

This paper describes the methodology used in the analysis of energy consumption data and indoor air quality data, including data processing and combining the data, pre-processing the combined data, basic idea of self-organizing map (SOM) and two clustering algorithms: k-means and Sammon's mapping. Some results are presented regarding with testing the clustering methods in a case study.

II. Materials and Methods

A. Data Collection

The collected consumption and indoor air quality data consisted of continuous measurements on temperature, relative humidity, carbon dioxide concentration, carbon monoxide concentration and differential air pressure between indoors and outdoors in the study house in Eastern Finland, in City of Kuopio. Data was collected using a monitoring system developed by the research group of environmental informatics [19]. Also, the measured data from consumption points of district heat, electricity and water were collected by wired connections installed in the technical room. Measurements were carried out during the time period from 1st of October 2010 to 23th of December 2011 in a house with area of 133.0 m^2 + garage 18.6 m^2 + stockroom 8.3 m^2. The measurements

M. Raatikainen is with the Department of Environmental Science, Research Group of Environmental Informatics, University of Eastern Finland, Kuopio, Finland (e-mail: mika.raatikainen@uef.fi)

J-P. Skön is with the Department of Environmental Science, Research Group of Environmental Informatics, University of Eastern Finland, Kuopio, Finland (e-mail: jukka-pekka.skon@uef.fi)

M. Johansson is with the Department of Environmental Science, Research Group of Environmental Informatics, University of Eastern Finland, Kuopio, Finland (e-mail: markus.johansson@uef.fi)

K. Leiviskä is with the Control Engineering Laboratory, University of Oulu, Oulu, Finland (e-mail: kauko.leiviska@oulu.fi)

M. Kolehmainen is with the Department of Environmental Science, Research Group of Environmental Informatics, University of Eastern Finland, Kuopio, Finland (e-mail: mikko.kolehmainen@uef.fi)

were logged from five rooms (Figure 1); living room (OH), bedroom (MH), kitchen (KT), bathroom (PH) and technical room (TR). Measured variables, units and sensor locations are presented in Table I.

TABLE I
MEASURED PARAMETERS AND THE LOCATION OF MEASUREMENTS

Parameter	Unit	Living room (OH)	Bedroom (MH)	Kitchen (KT)	Bathroom (PH)
Indoor air					
Temperature	°C	X	X	X	X
Relative humidity	%	X	X	X	X
Carbon dioxide	ppm	X	X	X	X
Carbon monoxide	ppm	X			
Parameter		**Technical room**	**(TR)**		
Consumption					
District heat	kWh	X			
Electricity	kWh	X			
Water	m^3	X			
Differential pressure	Pa	X			

B. Data Processing and Combining the data

Concerning the result of measurements there were two different data; consumption data and IAQ-data. The consumption data consisted of 541200 rows and 4 variables; time stamp, and consumptions of heat, electricity and water. All the data rows with water consumption higher than 300 had been removed (35 rows). After removal, the hourly consumption values were calculated and converted into cents per hour using the prices in euros of heat (MWh) and electricity (kWh) informed by local energy supplier. After these operations there were 8968 rows in the consumption data. In the IAQ-data there were 3785743 rows and 14 variables in columns. In the original data, IAQ samples had been taken every 5 seconds. There was no need to remove any of the rows and the hourly mean values for indoor air parameters were calculated. The consumption data and IAQ-data were combined into a matrix having 8445 rows and 19 columns (variables) explained in more details in table II.

The data matrix was modelled using self-organizing map (SOM). The reference vectors of SOM could be used as a basis for further analysis more easily than original measurement vectors due to the reduced number of data. The reference vectors were classified to clusters by k-means clustering algorithm, and for reference to express distances and groups in the data by Sammon's mapping. Finally, the clusters were analysed concerning indoor parameters' effects on energy efficiency.

Fig. 1 Layout of the study house.

C. Self-Organizing Map

The Self-Organizing Map (SOM) is a neural network algorithm developed by a Finnish academician Teuvo Kohonen in early 1980ies. The common purpose of SOM-method is to perform data analysis by mapping n-dimensional input vectors to the neurons, and visualizing results in a two-dimensional lattice [20]. In the two-dimensional lattice, the input vectors with common features effect on the same or neighbouring neurons, preserving the topological order of the original data. The SOM learning process is unsupervised: there is no need for a priori classifications for the input vectors. A large variety of SOM-based applications have been developed during the last three decades. The common application fields of SOM are, for example, in machine vision, signal processing, exploratory data analysis and in pattern recognition [20].

The training of SOM produces, as a result, a topological arrangement of output neurons. Each of these neurons has a special reference vector describing its hits, or input vectors. Each neuron in SOM is defined by the reference vector, which has the same dimensionality as the input vectors, and by its location. The reference vector can be defined as follows (Eq. 1):

$$r_m = (r_{m1}, r_{m2}, \dots, r_{mn}), \quad (m = 1, \dots, M), \tag{1}$$

where n is the number of variables, and M refers to the number of neurons in the map.

Firstly, in the beginning of the training, SOM must be initialized. In linear initialization SOM is initialized along the map dimensions, according to the greatest eigenvectors of the training data. In random initialization, the map is initialized by using arbitrary values for the reference vector. The use of linear initialization results in an ordered initial state for reference vectors instead of arbitrary values generated by random initialization [20].

The Best Matching Unit (BMU) is the neuron located at the smallest Euclidean distance from the input vector (Eq. 2):

$$\beta(x_i, R) = \arg\ \min_j \|x_i - r_j\| \tag{2}$$

where β is the index of the BMU, x_i is the input vector, and R includes the reference vectors of SOM.

The BMU and the group of its neighbouring neurons can be trained using the following update rule: [20] (Eq. 3):

$$r_m(k+1) = r_m(k) + h_{\beta m}(k)[x_i - r_m(k)] \tag{3}$$

where k is a number of iteration rounds and m implies the index of the neuron updated. The reference vectors gradually become weighted averages of the original data samples assimilated to each of them. The neighbourhood function is usually assumed to be Gaussian [20] (Eq. 4):

$$h_{\beta m}(k) = \alpha(k)\exp\left(-\frac{\|v_\beta - v_m\|^2}{2\sigma^2(k)}\right), \tag{4}$$

where v_β and v_m are the location vectors of the corresponding nodes, α refers to the factor of the learning rate, and σ(k) defines the width of the kernel.

In summary, the training of SOM proceeds as follows: 1) finding the BMU for one input vector according to the minimum Euclidean distance, 2) moving the reference vector (using the update rule) of the BMU towards this input vector, 3) moving the reference vectors (using the update rule) of neighbouring neurons towards this input vector, 4) repeating steps 1-3 for the next input vector until all input vectors have been used, 5) repeating steps 1-4 until the algorithm converges, 6) finding the final BMU for each input vector according to the Euclidean distance.

D. K-means clustering

The k-means clustering is a well-known non-hierarchical cluster algorithm [21]. The basic version of k-means begins by randomly picking k cluster centres and assigning each point to the cluster whose mean is closest in the sense of Euclidean distance. The next steps include computing the mean vectors of the points assigned to each cluster and using these as new cluster centres in an iterative approach. The clusters are determined by minimizing the sum of squared errors:

$$J_K = \sum_{k=1}^{K} \sum_{i \in C_k} (x_i - m_k)^2, \tag{5}$$

where x_i is a vector representing the ith data point and m_k is the centroid of the data points in C_k.

In the case of specific application, the number of clusters may not be known a priori. In the k-means algorithm the number of clusters has to be predefined. It is common that the algorithm is applied with the different number of clusters, and the best solution is selected using a validity index [22] or using enlightened asset expertise.

E. Sammon's Mapping

Sammon's mapping [23] is a non-linear mapping algorithm. The aim of the algorithm is to represent the points of a p-dimensional dataset onto a subspace of two dimensions preserving, however, the inter-pattern distances as far as possible. The goal is to optimize an error function E by following the steepest descent process.

$$E = \frac{1}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij}} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{(d_{ij} - d'_{ij})^2}{d_{ij}} \tag{6}$$

where n is the number of data points, d_{ij} is the (Euclidean) distance between two points x_i and x_j in the original space, and d'_{ij} is the (Euclidean) distance between the corresponding points x'_i and x'_j in the lower dimensional space.

F. Analysing combined consumption and IAQ-Data

The combined consumption and IAQ data were coded into inputs for the self-organizing map. All the input values were normalized by variance scaling and permutated before training the map. After training, a SOM having 100 neurons in 10 x 10 hexagonal grid was constructed. Linear initialization and batch training algorithm were used in training. The Gaussian function was used as the neighbourhood function. The map was taught with 10 iterations and the initial neighbourhood had the value of 6. The SOM Toolbox version 2.0 (Aalto University, Laboratory of Computer and Information Science) was used in the data analysis under a Matlab® software platform (Mathworks, Natick, MA, USA).

The k-means algorithm was used to cluster the trained map, or more precisely, to cluster the reference vectors. The Davies-Bouldin index (DBI) was used to evaluate the goodness of the clustering for each cluster. After the clustering, the desired reference vector elements of clustered neurons were visualized in a two-dimensional lattice to reveal the possible interactions between variables.

III. RESULTS

The statistical properties of the original data are presented in Table II. The aptitude of SOM as an analysis tool has been illustrated by examining the dependence of two variables, differential pressure and cost of electricity. As an example, Figure 2 shows how the under pressure and the high cost of electricity cluster in the same area in SOM, indicating the correlation of these variables.

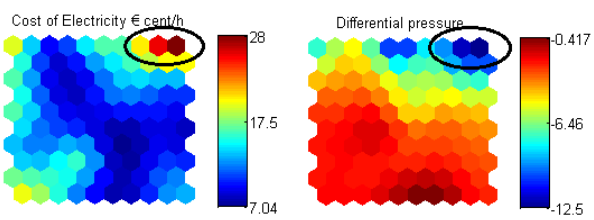

Fig. 2 Dependence of differential pressure and the cost of electricity.

Vectors of SOM were clustered by using k-means clustering algorithm and evaluated by using the Davies-Bouldin index. As a result, the number of clusters was defined as 7. Clusters are named with the numbers from 1 to 7, and they are visualized on the surface of SOM in Figure 4. Sammon's mapping in Figure 5 supports the clustering made by k-means illustrating the distances of vectors. Each cluster corresponds to a particular state of indoor air and costs of electricity and heat in the house. When comparing the clusters in Figure 4 to the component levels of SOM in Figure 3 in more detail, clusters can be characterized as follows:

1) This cluster describes the situation when the under pressure is very low. Relative humidity values are elevated and carbon dioxide values are quite low. Room temperatures are a little bit elevated and costs of electricity and heating are low. Probably it is summertime and the study house is unoccupied. Ventilation rate is very low or it is turned off or windows and doors are open.
2) This cluster describes the situation when the under pressure is quite a low. Relative humidity values are elevated and carbon dioxide values are elevated (600 – 700 ppm), too. Room temperatures are at the normal level (22 °C) and costs of electricity and heating consumptions are average. The study house is occupied. The season is supposedly late summer.
3) This cluster describes the situation when there is a small under pressure (c. 2 Pa) in the study house. Relative humidity values are quite average (15%) and carbon dioxide values are at the normal level (550 ppm). Room temperatures are lower than normal (19 °C). Cost of electricity consumption is low and cost of heat consumption is average. The study house is occupied and the ventilation rate is low. Probably the season is autumn.
4) This cluster describes the situation when under pressure is at high level. Relative humidity values are quite low (7-9%) and carbon dioxide values are the lowest (440 – 510 ppm). Room temperatures are elevated (24 °C). Cost of electricity consumption is at 17 € cent/h and cost of heating is 31 € cent/h. Definitely, the study house is occupied and ventilation rate is high. Because of low humidity and elevated temperature the season is probably spring.
5) This cluster describes the situation when under pressure is at level 6 Pa. Relative humidity values are quite low (7-9%) and carbon monoxide and carbon dioxide values are elevated in the living room. That means that fireplace is heated, or it just has been heated. Room temperatures are high (26 °C). Cost of heating is the highest. House is occupied and it is cold winter season.
6) This cluster describes the situation when differential pressure is close to zero. Relative humidity values are at highest level (25%) and carbon monoxide level is elevated. Carbon dioxide values are also elevated (600 ppm). Room temperatures are slightly elevated (23 °C). Cost of electricity consumption is higher than mean (20 € cent/h) and cost of heating is close to mean (22 € cent/h). Probably ventilation has been switched off during the fireplace has been heated. Because of high humidity it is supposedly rainy autumn season. House is occupied.
7) This cluster describes the heavily occupied house situation, when under pressure is the highest (13 Pa). Relative humidity values are less than average (13%) and the room temperatures are slightly low (20 °C). Carbon dioxide values are quite low (450 – 500 ppm). Cost of electricity is the highest (28 € cent/h) and cost of heating is average (23 € cent/h). Ventilation rate has been high and there has been lots of electricity consumption. Probably it is late winter.

TABLE II
STATISTICAL PROPERTIES OF THE COMBINED DATA VARIABLES

Variable	**Unit**	**Minimum**	**Maximum**	**Mean**	**Median**	**Standard deviation**
Cost of electricity	€ cent/h	3,7	93,1	12,4	9,3	10,3
Cost of heating	€ cent/h	4,7	65,7	23,1	18,8	9,4
Differential pressure	Pa	-31,9	6,1	-4,0	-3,2	4,1
Temperature, Bedroom	°C	17,1	26,8	22,0	21,7	2,3
Relative humidity, Bedroom	%	2,7	32,4	15,1	15,4	5,4
Carbon dioxide, Bedroom	ppm	466,0	1469,6	571,6	559,7	63,2
Temperature, Living room	°C	17,2	27,8	22,9	22,8	2,4
Relative humidity, Living room	%	4,4	35,2	16,5	16,8	5,0
Carbon monoxide, Living room	ppm	0,4	2,3	0,5	0,4	0,1
Carbon dioxide, Living room	ppm	395,6	1224,2	498,0	484,9	68,5
Temperature, Bathroom	°C	21,2	29,0	23,6	23,2	1,6
Relative humidity, Bathroom	%	3,5	32,1	15,2	15,7	4,6
Temperature, Kitchen	°C	17,7	27,6	22,9	22,6	2,2
Relative humidity, Kitchen	%	3,9	33,2	15,6	16,0	5,3
Carbon dioxide, Kitchen	ppm	418,2	1428,3	511,4	504,2	60,1

International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:6, No:5, 2012

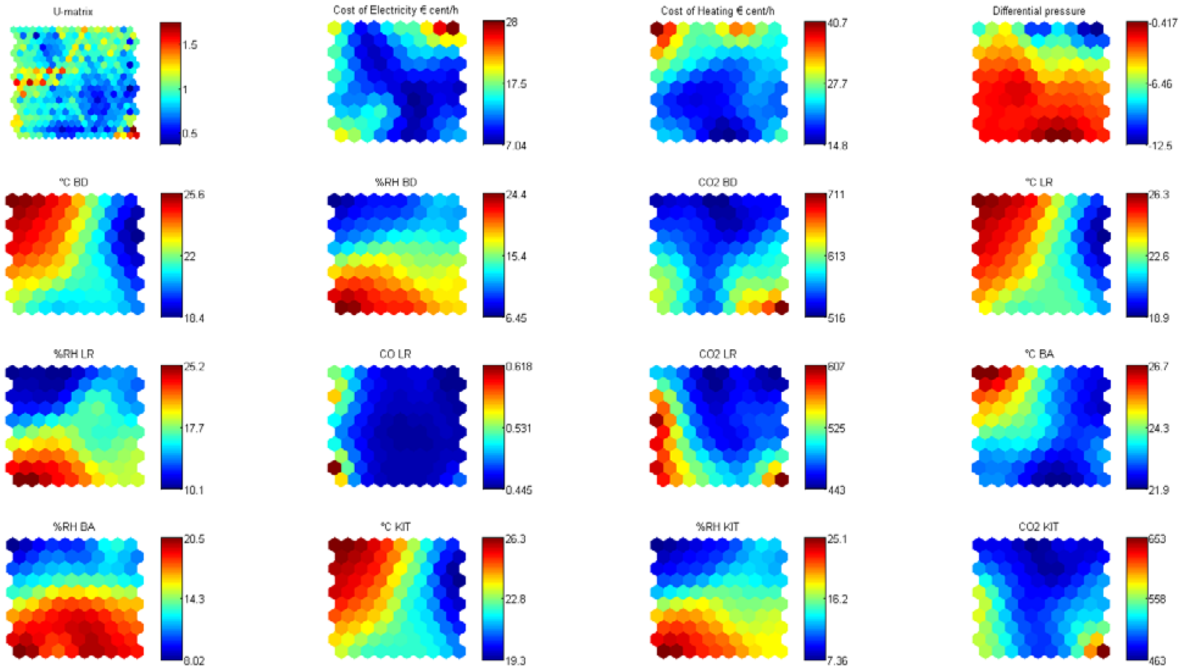

Fig. 3 Component levels of SOM, describing different variables. LR means living room, BD means bed room, BA means bathroom and KIT means kitchen.

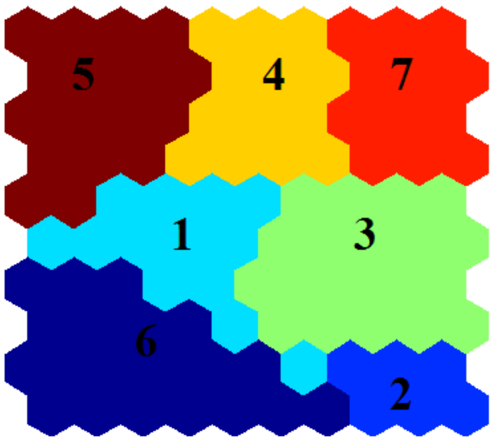

Fig. 4 Clusters of SOM.

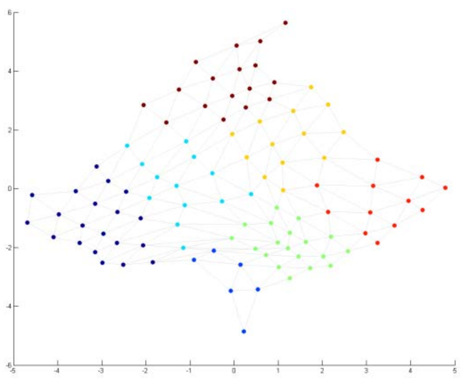

Fig. 5 Sammon's mapping of the results of SOM analysis.

IV. DISCUSSION

As mentioned before, this paper presents preliminary results of data analysis on energy consumption and indoor air quality. In particular, SOM analysis showed that there were seven different phenomena characterizing indoor air quality in the case study house.

Based on measurements, it can be concluded, that the growth of pressure difference has an impact on electricity consumption. In addition, compared to IAQ conditions of bed room and kitchen, CO_2 concentrations seem to be similar. It is significant, that some of the seasons can be identified from Figures 3 and 4. For example, cluster 1 represents the summer season and cluster 5 represents the winter season. The clusters seem to describe well normal acceptable indoor air quality. Whole, reviewed indoor air quality was good concerning the variables presented in Table I.

Figure 3 shows that there is a positive correlation between the cost of heating and room temperatures. In addition, cost of electricity consumption has a negative correlation to differential pressure and carbon dioxide concentration. When house is occupied, the electricity energy is needed partly of human usage but, also for ventilation needs.

V. CONCLUSION

SOM-based method can reveal dependencies between data variables relatively fast and easily. Also SOM result is good when the clustering behaviour of the data is unknown before the data analysis. The results presented in this paper show that the applied SOM-based neural network is an efficient way to analyse indoor air quality data and Sammon's mapping can support and express further features from clusters. Nowadays, buildings are more energy efficient and airtight. This can cause problems with indoor air quality. In this study, the ventilation rate of the study house was sufficient concerning indoor parameters, when the study house was occupied. Finally, we will extend our research work to include several apartment and school buildings, and we are also extending the continuous measurement period. In addition, we will research the application possibilities of neural network modeling further in the field of energy efficiency and healthy housing.

In this study, preliminary results are interesting. Particularly, further research on variable distribution of clusters and seasonal variation of indoor air quality and its effect on energy consumption, is needed.

ACKNOWLEDGMENT

This research was done as a part of the Finnish AsKo-project (*Asuinrakennusten korjaus- ja täydennysrakentamisen vaikutukset asumisen energiatehokkuuteen ja sisäilman laatuun*; The effects of renovation and complimentary construction on energy efficiency and indoor air quality). For financial support, the authors would like to thank the Ministry of the Environment.

REFERENCES

[1] N. Gershenfeld, S. Samouhos, and B. Nordman: "Intelligent Infrastructure for energy efficiency," Science, vol. 372, pp.1086-1088, Feb. 2010.

[2] R. J. Jackson, "Environment Meets Health, Again," Science, 315(5817), pp.1337, Mar. 2007.

[3] J. P. Holdren, "Energy and Sustainability," Science, 315(5813), pp. 737, Feb. 2007.

[4] R. Armstrong and N. Spiller, "Synthetic biology: Living quarters," Nature, 467, pp. 916-918, Oct. 2010.

[5] D. Butler, "Architects of a Low-energy Future," Nature, 452, pp. 520-523, Apr. 2008.

[6] S. C. Sofuoglu, "Application of artificial neural networks to predict prevalence of building-related symptoms in office buildings," Building and Environment, vol. 43, pp. 1121-1126, 2007.

[7] H. Xie, F. Ma and Q. G. Bai, "Prediction of indoor air quality using artificial neural networks," Fifth International Conference on Natural Computation (ICNC '09), vol. 2, pp. 414-418, 2009.

[8] M. H. Kim, Y. S. Kim, J. J. Lim, J. T. Kim, S. W. Sung and C. K. Yoo, "Data-driven prediction model of indoor air quality in an underground space," Korean Journal of Chemical Engineering, vol. 27, pp. 1675-1680, 2010.

[9] T. E. Alhanafy, F. Zaghlool and A. S. El Din Moustafa, "Neuro fuzzy modeling scheme for the prediction of air pollution," Journal of American Science, vol. 6, pp. 605-616, 2010.

[10] T. Lu and M. Viljanen, "Prediction of indoor temperature and relative humidity using neural network models: model comparison," Neural Computing & Applications, vol.18, pp. 345-357, 2009

[11] M. Kolehmainen, H. Martikainen, T. Hiltunen, and J. Ruuskanen, "Forecasting air quality parameters using hybrid neural network modelling," Environmental Monitoring and Assessment, vol. 65, pp. 277-286, 2000.

[12] M. Kolehmainen, H. Martikainen and J. Ruuskanen, "Neural networks and periodic components used in air quality forecasting," Atmospheric Environment, vol. 35, pp. 815-825, 2001.

[13] H. Niska, T. Hiltunen, M. Kolehmainen and J. Ruuskanen, "Hybrid models for forecasting air pollution episodes," International Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA'03), University Technical Institute of Roanne, France April 23-25, 2003.

[14] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M.N. Assimakopoulos, R. Mitoula, S. Zerefos, "Using intelligent clustering techniques to classify the energy performance of school buildings," Energy and Buildings, vol. 39, pp.45-51, 2007.

[15] Z. Wang, Z. Bai, H. Ya, J. Zhang, T. Zhu, "Regulatory standards related to building energy conservation and indoor-air-quality during rapid urbanization in China, "Energy and Buildings, vol. 36, pp.1299-1308, 2004.

[16] J. Yao, N. Zhu, "Enhanced supervision strategies for effective reduction of building energy consumption – A case study of Ningbo, "Energy and Buildings, vol. 43, pp.2197-2202, 2011.

[17] A. Simoudi, P. Kostarela, "Energy monitoring and conservation potential in school buildings in the C´ climate zone of Greece, "Renewable Energy, vol. 34, pp.289-296, 2009.

[18] A. Elkilani, W. Bouhamra, "Estimation of optimum requirements for indoor air quality and energy consumption in some residences in Kuwait," Environment International, vol. 27, pp.443-447, 2001.

[19] J-P. Skön, O. Kauhanen and M. Kolehmainen,"Energy Consumption and Air Quality Monitoring System, "Proceedings of the 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 163-167, Adelaide, Australia Dec. 6-9, 2011.

[20] T. Kohonen, "Self-organizing maps," 3rd ed., Springer-Verlag, Berlin Heidelberg, 2001.

[21] J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297, 1967.

[22] D. Davies and D. A. Bouldin, "Cluster separation measure," IEEE Trans Pattern Anal Mach Intell, vol. 2, pp. 224–7, 1979.

[23] J. W. Sammon, "A nonlinear mapping for data structure analysis," IEEE Transactions on Computers 18: 401–409, 1969.