Effects of Weather Conditions and Differential Air Pressure on Indoor Air Quality

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

***Abstract*—**The multivariate methods are widely used in researching of buildings and residential Indoor Air Quality (IAQ). Instead, one of the multivariate methods, namely self-organizing map (SOM) is applied only marginally in the study of correlations of outdoor weather conditions on IAQ. It is also worth noticing, that only few papers detail differential air pressure and its effects on indoor climate. In this study, SOM was applied to resolve the effects of air pressure difference between outdoors and indoors, to the other indoor variables, because there was no linear correlation between measured variables. In addition, the SOM was qualified as a suitable method having a property to summarize the variable's dependencies into easily observable two-dimensional map. Measured data used in this study was collected continuously in a small house in Eastern Finland. The method used could distinguish nine different clusters charactering indoor climate in the study house. The results indicate, that when indoor underpressure increased the other measured values decreased, including indoor relative humidity (RH), and the concentrations of carbon monoxide (CO) and carbon dioxide (CO_2). This probably indicates that the ventilation rate of the study house was adequate during the study period.

***Keywords*—**Differential air pressure, Indoor air quality, Neural networks, Self-organizing map

I. Introduction

INDOOR Air Quality (IAQ) is a widely researched topic, because of manifold effects on health of occupants. Increased interest in energy efficiency is thought to affect negatively on indoor air quality. For instance, in Nature there are many discussions concerning low-energy buildings and their relation to carbon emissions [1], as well as on the use of biological indicators for IAQ [2]. In Science, there are also articles discussing about using and extending smart grids for energy efficiency [3], sustainability [4], and the relationships between healthiness and environment [5].

Neural networks have been used successfully 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]. There are also previous studies on forecasting outdoor air quality parameters using computational methods [11, 12, 13].

This paper describes methodology used in indoor air and outdoor air data analysis, including data processing, pre-processing the raw data, basic idea of self-organizing map (SOM) and clustering. Some results are also presented regarding testing the methodology in a case study.

II. Materials and Methods

A. Data Collection

The collected indoor air quality data consisted of continuous measurements on temperature, relative humidity, CO_2 concentration, CO concentration and air pressure difference between indoors and outdoors in the house in Eastern Finland. Data was collected using a monitoring system developed by the research group of environmental informatics [14]. The measured data from outdoor conditions were collected by weather station installed on the roof of the house located in the same area. Outdoor data consisted of the wind speed, outdoor relative humidity, air pressure and outdoor temperature. Measurements were carried out in 2010 from 29th of August to 1st of November. Measured variables, units and sensor locations are presented in Table I.

TABLE I
MEASURED PARAMETERS AND THE LOCATION OF MEASUREMENTS

Parameter	**Unit**	**Bedroom**	**Living room**	**Bathroom**
Indoor air				
Temperature	°C	X	X	X
Relative humidity	%	X	X	X
Carbon dioxide	ppm	X	X	
Carbon monoxide	ppm		X	
Pressure difference	Pa			
Parameter		**Nearby study house**		
Outdoor air				
Wind speed	m/s	X		
Relative humidity	%	X		
Atmospheric pressure	hPa	X		
Temperature	°C	X		

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)

A. Data Processing and Pre-processing the Raw Data

In the beginning of the data processing (Figure 1), the raw air quality data were pre-processed for the data analysis. This consisted of removing outliers and the variance scaling of the data. It revealed, as visually analysed, that there were outliers in the indoor air temperature. All the data rows with temperature values higher than 40 degrees of Celsius were discharged. Missing data or outliers were 20.8 percent of the data and these rows were removed. The size of the data matrix utilized, was 14837 rows, 14 variables in columns.

The solid 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. Finally, the clusters were analysed concerning effects on indoor air quality.

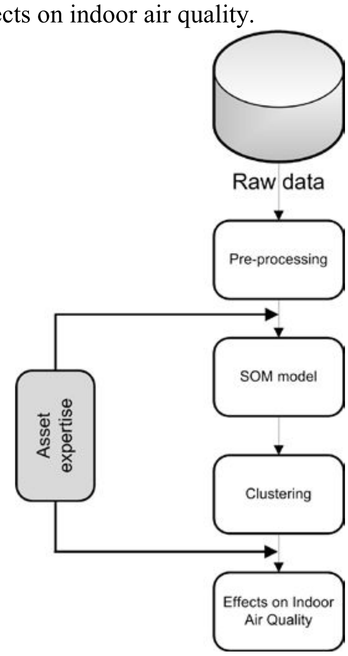

Fig. 1 A diagram of the data processing.

B. 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 [15]. 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 three decades. The common application fields of SOM are, for example in machine vision, signal processing, exploratory data analysis and in pattern recognition [15].

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. On one hand each SOM-neuron is defined by the reference vector, which has the same dimensionality as the input vectors, and secondly by its location. The reference vector can be defined as follows (Eq. 1):

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

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, the 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 [15].

The Best Matching Unit (BMU) is the neuron being 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 the SOM.

The BMU and the group of its neighbouring neurons can be trained using the following update rule: [15] (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 [15] (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 learning rate, and σ(k) defines the width of the kernel.

In summary, the training of the 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 that input vector, 3) moving the reference vectors (using the update rule) of neighbouring neurons towards that 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

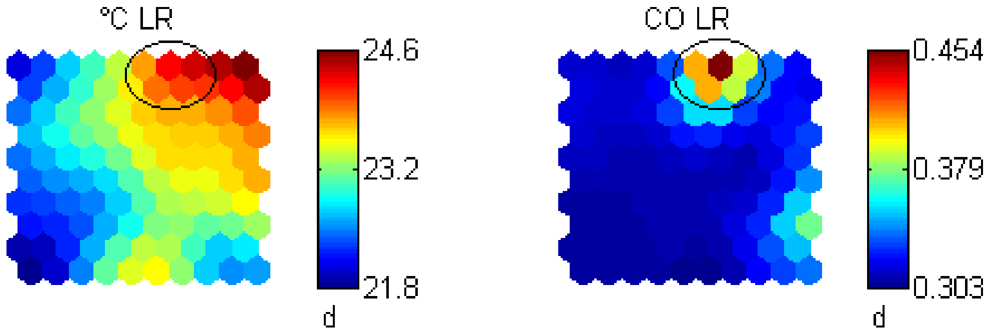

Fig. 2 Dependencies of measured temperature and carbon monoxide. LR means Living Room.

according to the Euclidean distance.

C. K-means clustering

The k-means clustering is a well-known non-hierarchical cluster algorithm [16]. The basic version of k-means begins by randomly picking k cluster centres, 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, using these as new clusters in an iterative approach. The clusters are determined by minimizing the sum of squared errors:

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

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 [17] or using enlightened asset expertise.

D. Analysing IAQ data

The IAQ data were coded into inputs for the self-organizing map. All the input values were normalized by variance scaling 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 the map 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, precisely, to cluster the reference vectors. The Davies-Bouldin index (DBI) was used to evaluate the goodness of the clustering for each cluster number. After the clustering, the 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, temperature and carbon monoxide. As an example, Figure 2 shows how the increase of temperature and carbon monoxide concentration cluster in the same area in SOM, indicating the correlation of these variables. The dependence is due to the fact that there has been fire in the living room fireplace at the moment of time considered. The concentration of carbon monoxide did not rise to the level of health hazard during this time.

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 9. Clusters are named with the letters from A to I, and they are visualized on the surface of SOM in Figure 4. Each cluster corresponds to a particular indoor air state in the house.

When comparing the clusters in Figure 4 to the component levels of SOM in Figure 3 and analysing Figures 5 and 6, the characterization for the clusters described below could be found:

A. This cluster describes the situation, when the house is unoccupied. In this case, all measured indoor air values are low, and there is an underpressure in the house. The outdoor temperature is low.

B. This cluster describes the situation when the fireplace is heated, or it just has been heated. There is an underpressure in the house, and the temperature in the living room is higher than the average value. The carbon monoxide concentration measured in the living room, is slightly elevated, and the relative humidity is low.

TABLE II
STATISTICAL PROPERTIES OF THE VARIABLES

Variable	Unit	Minimum	Maximum	Mean	Standard	Kurtosis	Skewness
Indoor							
Temperature (BR)	°C	20,6	25,0	22,5	0,7	3,2	0,2
Temperature (LR)	°C	21,2	26,0	23,1	0,9	2,9	0,4
Temperature (BA)	°C	20,9	28,4	22,6	0,9	4,9	0,9
Relative humidity (BR)	%	20,0	49,1	32,3	6,2	2,5	0,4
Relative humidity (LR)	%	18,0	47,4	31,0	6,3	2,3	0,3
Relative humidity (BA)	%	20,2	90,1	35,6	8,9	8,1	1,5
Carbon dioxide (BR)	ppm	359,8	997,6	549,2	114,0	5,2	1,6
Carbon dioxide (LR)	ppm	432,1	1345,5	656,4	145,5	4,7	1,3
Carbon dioxide (BA)	ppm	0,2	4,5	0,3	0,1	423,3	18,8
Outdoor							
Differential air	Pa	-9,3	41,2	-1,0	2,6	31,3	4,0
Wind speed	m/s	0,1	11,6	1,9	1,4	5,8	1,7
Relative humidity	%	46,0	98,2	82,4	11,8	2,9	-0,8
Atmospheric pressure	hPa	989,5	1035,2	1015,4	11,0	2,0	-0,2
Temperature	°C	-1,4	19,5	8,4	4,2	2,6	-0,2

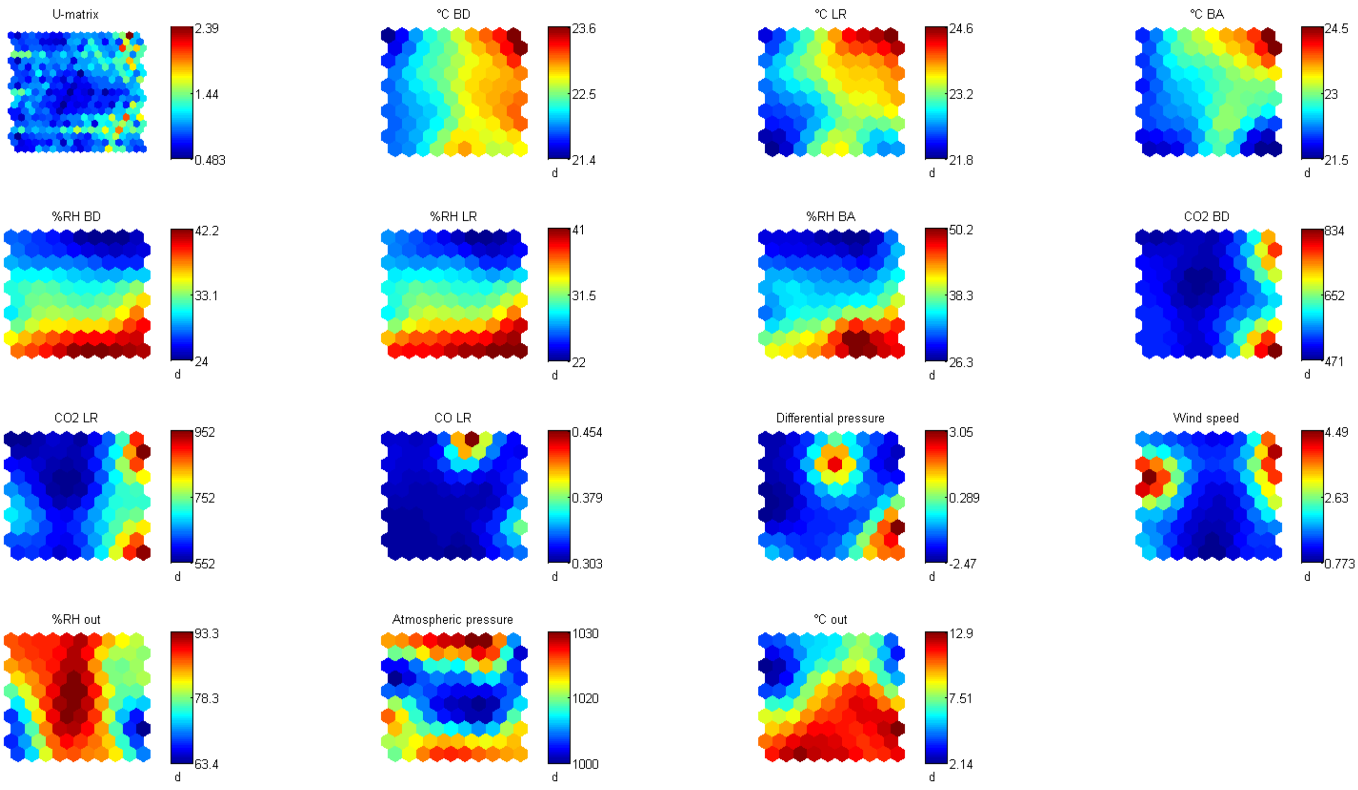

Fig. 3 Component levels of SOM, describing different variables. LR means Living Room, BD means Bed Room and BA means Bathroom.

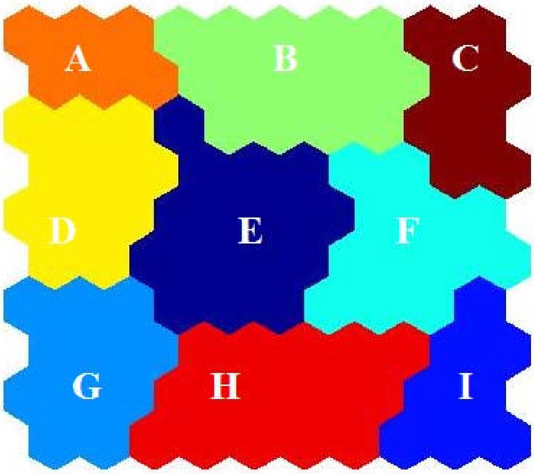

Fig. 4 Clusters of SOM.

C. This cluster describes the fact that the house is occupied when the outdoor pressure and temperature are low. In this case, there is an underpressure in the house. All rooms have a low relative humidity, room temperature is higher than the average and carbon dioxide values are slightly elevated.

D. Again, this cluster describes the situation when the house is unoccupied. In this situation, all the measured indoor air values are low. The atmospheric pressure and outdoor temperature are low and wind speed is high.

E. In this cluster, all the measured indoor values are on the average level. Outdoor measurements noticed that humidity is higher and atmospheric pressure is lower than the average. Apparently it is raining outside.

F. Also, this cluster describes the house occupied situation, when there is an underpressure outside. Outdoor relative humidity and temperature are lower than average. The temperatures in all rooms are higher than average and carbon dioxide values are slightly elevated.

G. In this cluster indoor temperature is also low, although the indoor humidity is elevated. Carbon dioxide and carbon monoxide levels are average. Atmospheric pressure is slightly elevated and ambient humidity is low.

H. This cluster differs from the cluster G, so that all the room temperatures are slightly higher than average, further atmospheric pressure and outdoor relative humidity are also high.

I. This cluster describes the situation, that there is a excess pressure in the house at the same time, when relative humidity is high. Also, carbon dioxide and carbon monoxide levels are elevated, thus the house is populated. It should be noted that the living room and bathroom temperatures are low, although the bedroom temperature is slightly elevated. There is a high-pressure outside, it is warm and outdoor humidity is low. Indoor air pressure is higher than average.

In Figures 5 and 6 boxplot graphs describing variables of each cluster are presented. The blue colour in the rectangles describes 25% and 75% fractiles, and the red line inside the rectangle denotes the median. The transversal black lines show the extreme values of a variable.

Figures 5 and 6 support the characterization drawn based on SOM component planes in Figure 3.

IV. Discussion

Based on measurements, it can be concluded, that the growth of pressure difference has an impact on relative humidity of indoor air, as well as on concentrations of indoor carbon dioxide and carbon monoxide. In particular, when the ambient pressure increased higher than the indoor air pressure, all measured values were found to be low. Based on SOM analysis, it showed that there were nine different phenomena characterizing indoor air quality in the studied house. There were three occasions, when the studied house was unoccupied. In these phenomena, carbon dioxide concentration in the house was low. There were six clusters, where the studied house was occupied. In these situations indoor temperature was normal or warmer than usual. Also, carbon dioxide levels were elevated, but CO_2 consentration levels remained at normal level. The indoor air quality levels are considered to be satisfactory, when CO_2 concentration is less than 1500 ppm. The clusters seem to be descriptive considering normal acceptable indoor air quality. Whole, reviewed indoor air quality was good concerned the variables presented in Table I.

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 shows that the applied SOM-based neural network method is an efficient way to analyse indoor air quality data. 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 during the study period concerning indoor parameters. Developing new indoor air quality monitoring systems and new methods for analysing the data are very important. Finally, we will extend our research work to include several apartment buildings, and we are also extending the measurement period. In addition, we research the application possibilities of neural network modeling further in the field of energy efficiency and healthy housing.

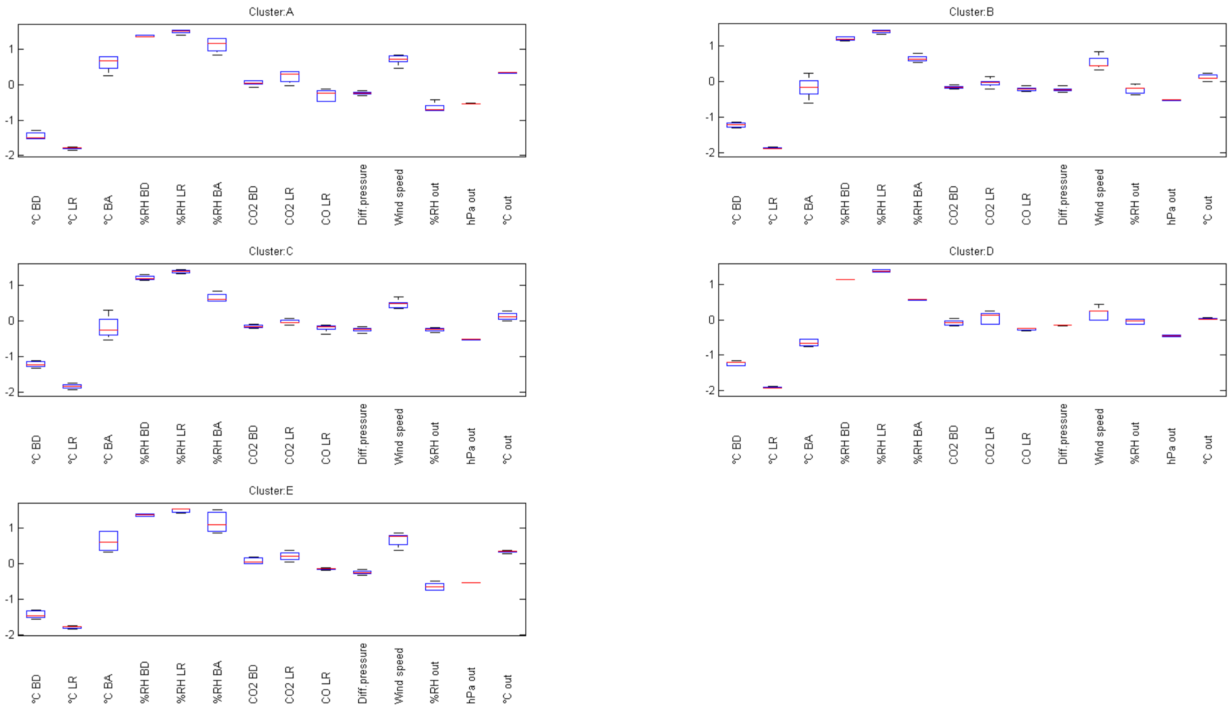

Fig. 5 Variable distributions of clusters from A to E are described as a boxplot graphs.

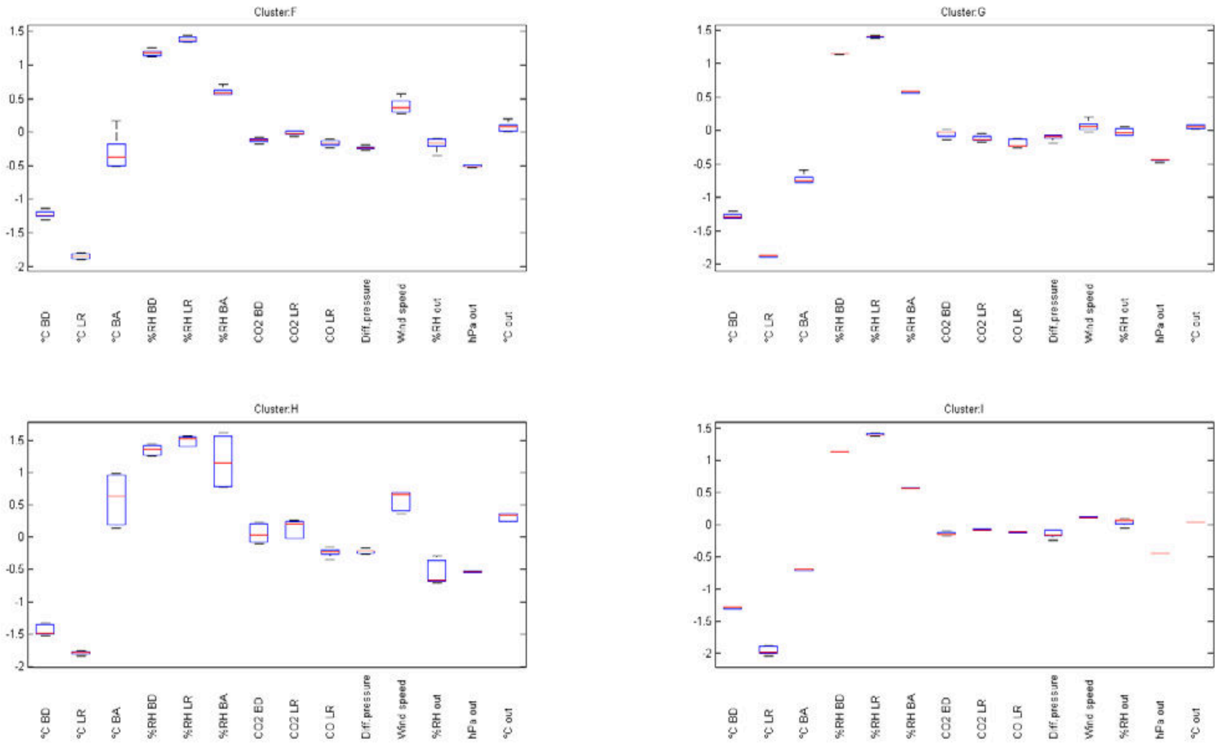

Fig. 6 Variable distributions of clusters from F to I are described as a boxplot graphs.

ACKNOWLEDGMENT

This research was done as a part of the Finnish TERTU-project (*Asumisen terveellisyys ja turvallisuus*; Housing Healthy and Safety) and 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, Regional Council of Pohjois-Savo, City of Kuopio and for cooperation YIT Kiinteistötekniikka Oy, Saint-Gobain Rakennustuotteet Oy, Rakennusliike Parha Oy, FinnEnergia Oy and Osuuskunta Suomen Asuntomessut.

REFERENCES

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

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

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

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

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

[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] 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.

[15] S. Haykin, "Neural Networks–A Comprehensive Foundation," 2nd ed., New Jersey: Prentice-Hall Inc., 1999.

[16] 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.

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