Knowledge E
Engaging minds

Conference Paper

Occupancy profiles modelling based on indoor measurements and clustering analysis: Application in an office building

Dafni Mora[1], Marilena De Simone[2], Gianmarco Fajilla[2], and José R. Fábrega[1]

[1]Universidad Tecnológica de Panamá, Centro de Investigaciones Hidráulicas e Hidrotécnicas (CIHH), Panamà City, Panamá
[2]Department of Mechanical, Energy and Management Engineering, University of Calabria, 87036, Rende, Italy

Corresponding Author:
Dafni Mora
dafni.mora@utp.ac.pa

Received: 15 November 2017
Accepted: 5 January 2018
Published: 4 February 2018

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ESTEC Conference Committee.

Abstract

Sensors were applied in an office building to obtain information regarding user presence and absence intervals. Occupancy was also recorded by manual observation, and indoor parameters such as air temperature, relative humidity, carbon dioxide (CO_2), volatile organic compounds (VOC) were monitored. Occupants' behaviors regarding door/window (open/closed) and electric power were considered. Clustering analysis by manual observation was employed to identify similarities in daily or monthly occupancy and to describe possible occupancy profiles. Similar approach was carried out with each monitored parameter and the results of clustering elaboration were compared with the real occupancy profiles to identify which sensor is more effective to measure office occupancy. Furthermore, data were analyzed to explore relationships between occupancy and the magnitude of indoor environmental changes with the objective to identify daily, weekly, or monthly patterns. Single-linkage, complete-linkage, and average-linkage clustering were applied to each dataset. The cophenetic correlation coefficient was used to verify the quality of the results obtained for each variable, and the complete linkage was selected to define the groups. Comparison between occupancy real data clustering and VOC and open/closed door groups demonstrated not similarities. The electricity consumption and CO_2 data showed some similarities.

Keywords: Occupancy detection, environmental sensor, clustering analysis, Office buildings

1. Introduction

Occupant movements and presence are fundamental to occupant behavior simulation by providing information about whether a room is occupied, the number of occupants,

OPEN ACCESS

How to cite this article: Dafni Mora, Marilena De Simone, Gianmarco Fajilla, and José R. Fábrega, (2018), "Occupancy profiles modelling based on indoor measurements and clustering analysis: Application in an office building" in *6th Engineering, Science and Technology Conference (2017)*, KnE Engineering, pages 711–720. DOI 10.18502/keg.v3i1.1474

or the specific individual in the room. The real occupancy patterns in buildings may differ significantly from each other. Gathering data on occupants-building interaction is a new horizon for achieving energy efficiency in the building sector.

Occupants emit heat, pollutants such as carbon dioxide (CO_2) and odor, and generate sound to the space. These interactions and their effect on the indoor environment can be measured via pertinent environmental sensors (Dong et al. 2010). As the human presence emits heat and pollutants, it is related to the indoor environment (Page et al. 2008). Zhang et al. 2012, concluded significant correlations between the occupancy and the environmental variables with values of 35.70% for CO_2, 32.49% for relative humidity and 11.98% for temperature.

Information regarding the occupancy level can be found also through sensor networks. Dong et al. 2010, investigated the use of ambient sensors for detecting the number of occupants in an office building. The experimental setup was divided into three separate sensor network systems: the wired sensor gas detection network, which measures carbon dioxide (CO_2), carbon monoxide (CO), total volatile organic compounds (TVOC), outside temperature, dew point, small particulates (PM2.5); a wireless ambient-sensing network, which measures lighting, temperature, relative humidity (RH), motion detection and acoustics; and an independent CO_2 sensor network. The authors concluded that there are significant correlations between measured environmental conditions and occupancy status, in particular with humidity, acoustics, and CO_2. However, no significant correlation with temperature data was reported.

Also, Dong and Lam 2011 showed significant correlations between measured environmental conditions and occupancy status, specially acoustics and CO_2 measurements. Also, they found that occupant´s presence has a direct effect on the air quality indexes (temperature, CO_2, and humidity levels) (Ebadat et al. 2013). Overall, the analyzed case studies demonstrated that presence and habits of occupants influence the use of equipment and indoor conditions.

The present work study this correlation by constructing an experimental setup consisting in instrumentation for measuring environmental quantities, electricity consumptions, and occupancy by following the criterion of sensor fusion. Here we will focus on the analysis of occupancy regarding CO_2, electricity consumption, and open/closed door measures.

The paper is structured as follows: presentation of the experimental setup, description of the measured environmental parameters, and application of clustering analysis.

The data were analyzed to explore relationships between the occupancy and the magnitude of indoor environmental changes with the aim to identify what sensor is more suitable to describe the occupancy in an office.

2. Methodology

2.1. Office and Occupancy Description

Data were collected in an office building at the University of Calabria (Italy). The office room has an area of 19 m^2 and a height of 2.50 m. The room presents a single wall facing outside Westwards and a two-wing window of 68x76 cm. The room is equipped with desktop computers and printers and the heating and cooling system is autonomous. The level of analysis is individual: one office generally used by one person. The information about how the office is utilized was collected by interviewing.

2.2. Sensors characteristics and positioning

The sensors' location was carefully selected to ensure that the instruments are triggered when occupants are in the office. The indoor environment sensors were placed on internal walls at the height of roughly 1.8 above the floor (Andersen 2009), avoiding direct sunlight. One CO_2 sensor is installed near the desk at nose level (when seated, 1.1 m) above the ground (Dong et al. 2010). All occupancy sensors were factory calibrated, and control systems were commissioned before data collection. For management and inventorying the WuTility Version 4.30 tool was used (Wiesemann & Theis 2015). Table 1 gives the characteristics of the sensors used for the experimental setup.

2.3. Occupancy Manual Observation

Data collected for weekdays, holidays and weekends from May 13 through September 30, 2016 were used. The range of data was limited to working days (from Monday to Friday) from 8:00 to 21:00. In addition to the sensor measurements, people recorded manually their presence each minute. The sequences in turn are constructed by 780 characters, one for each 1-minute time step, with a value that corresponds to binary data (1 is Occupied and 0 is unoccupied). Furthermore, the analyzed days were divided

DOI 10.18502/keg.v3i1.1474

into time slots: morning (08:00-13:00), lunch time (13:00-15:00), afternoon (15:00-21:00) and all day (08:00-21:00). The days were identified with the first letter of the month, then the day of the week and the number of days (e.g. JF24 is Friday, June 24).

Table 1: Specification of sensors employed in this study

Sensor	Variable	Measuring error	Measuring range and resolution
Wieseman & Theis 57018 CO_2 sensor	Carbon dioxide [ppm]	Measuring range: 0..2000ppm CO_2 ± 30ppm,± 5%	Analog
Wieseman & Theis 57645 AC Device	Electricity power	Range 0..50A AC, 30-6000Hz (all waveforms)	-
ABUS FU7350W Abus rectangular, NC,0.2 A Reed Switch	Window/door position (open/closed) Air conditioner (on/off)	Contact sensor	-
Wieseman & Theis 57618 Web-Graph Air Quality	Volatile organic compounds [ppm]	Measuring range: 450..2000ppm VOC as CO_2 equivalent	
	Air temperature [° C]	typ. @ 25 ° C ± 0.3 ° C max. @ 0..50° C ± 1.2° C	1/10° C,
	Relative humidity [%]	typ. @ 25° C ± 3% rH max. @ 0..50° C ± 7% rH (0-100% rH)	1/10% rH

2.4. Clustering Analysis

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). A distance measure is a metric (or quasi-metric) on the feature space used to quantify the similarity of patterns. In the single link method, the distance between two clusters is the minimum of the distances between all pairs of patterns drawn from the two clusters, while in the complete-link algorithm, the distance between two clusters is the maximum of all pairwise distances between patterns in the two clusters (Jain et al. 1999).

The cophenetic correlation coefficient is used to compare the results of clustering the same data set using different distance calculation methods or clustering algorithms (Aldenderfer and Blashfield 1984).

$$C_{coph} = \frac{\sum_{i<j} (d(i,j) - d)(t(i,j) - t)}{\sqrt{\left[\sum_{i<j} (d(i,j) - d)^2\right]\left[\sum_{i<j} (t(i,j) - t)^2\right]}}$$

The authors applied single-linkage, complete-linkage and average-linkage clustering on the dataset for three different hours combinations of the edit distance. The cophenetic correlation coefficient was used to verify the quality of the results obtained for each variable, and the complete-linkage was selected to define the groups.

The statistical software R (R Development Core Team 2014) was used for the statistical analysis.

3. Results and Discussion

3.1. Measured data and Correlation with Occupancy

Figure 1 shows hourly data of $CO_{2,}$ VOC, air temperature, relative humidity, window state, electric power and door opening for two typical summer days. Occupancy is illustrated by dashed line. Correlation analysis was realized in order to know the relationship between the occupancy and the measured parameters. Figure 2 shows the correlation plot and the corresponding Pearson's (r) correlation coefficients, positive correlations are indicated in blue and negative correlations in red color. The p-values are less than 2.2e-16 for all correlations. Correlations are significant for CO_2 (r=0.62), electric power (r=0.64), and door status (r=0.50), in contrast with the other parameters that demonstrate not significant correlations.

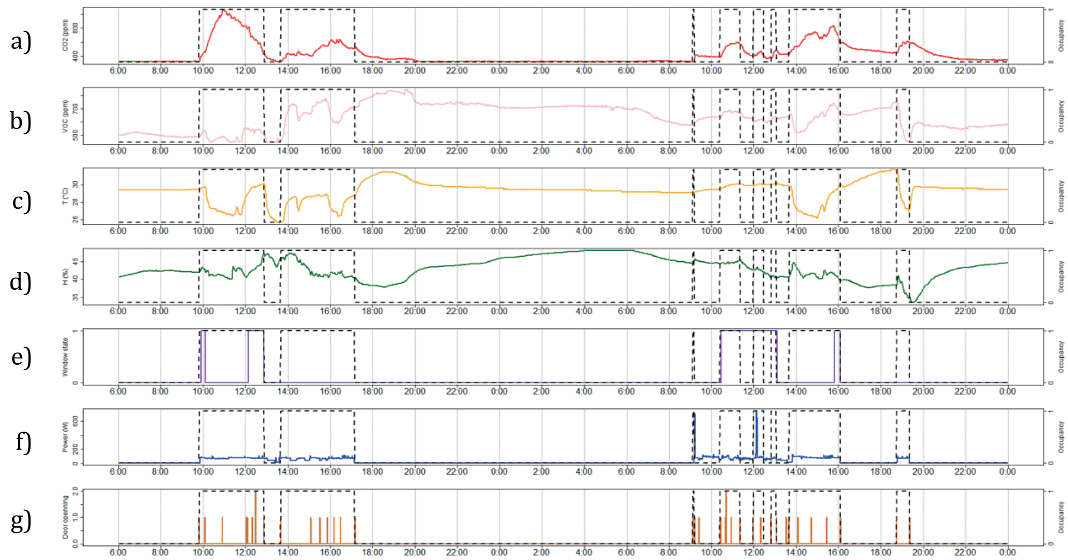

Figure 1: Measurements and occupancy profiles for two continuous typical summer days. a) CO2 (ppm), b) VOC (ppm), c) T (° C), d) Relative humidity (%), e) window state, f) electric power (watts) and g) door opening

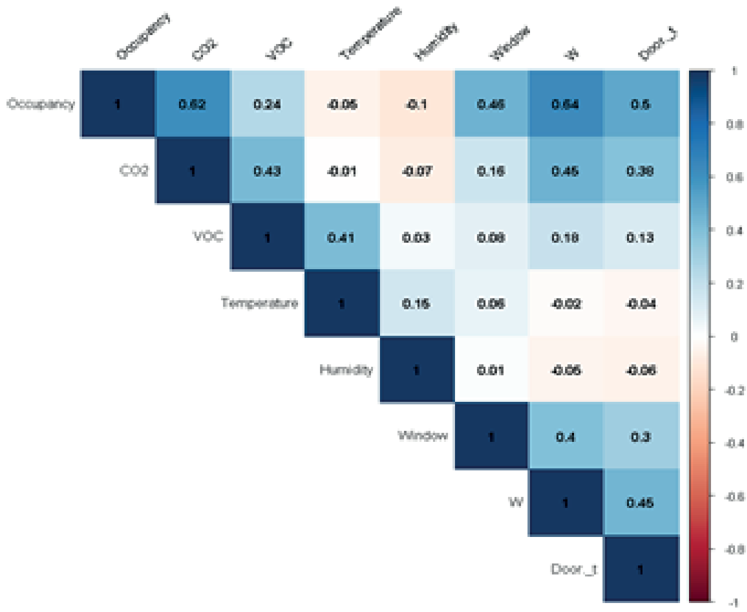

Figure 2: Correlation plot. Pearson's correlation coefficients, by color and values for 48 hr of continuing measurements corresponding to Figure 1 data

3.2. Clustering Analysis application

By considering the results of the correlation **b**etween variables, three parameters were identified and used in the cluster analysis to identify similarities in the days and possible occupancy profiles and determine the minimum quantity of sensors necessary to define the occupancy patterns. Three clusters were identified for all day schedule and two clusters for the other time slots. For reasons of brevity, this paper only presents the dendrogram results for occupancy data by considering the all day schedule.

3.2.1. Occupancy

Hierarchical clustering found different occupancy profiles for each time slot. The cophenetic correlation coefficients and the distance used for each method are listed in Table 2. The method that presents the highest value of the cophenetic correlation is the complete, and the results of this approach are presented for each parameter and compared with clustering of real occupancy data. For all combinations, single-linkage clustering performs the worst and complete-linkage clustering performs the best (Table 2).

In addition to the cophenetic correlation, a visual inspection of each dendrogram was done, and the results confirm that the complete method provides well-defined clusters for each time slot.

Table 2: Cophenetic correlation coefficients for all methods and time slots considered

Distance measure	Clustering method	Morning	Lunch	Afternoon	All day
Euclidean	Single	0.75	0.77	0.78	0.50
	Complete	0.80	0.71	0.81	0.70
Squared Euclidean	Average	0.78	0.79	0.80	0.66

In the dendrogram for all day slot, three clusters were identified and defined as three occupancy levels: low level with a mean daily occupancy of 3.8 hours, medium level with mean occupancy of 5.9 hours and high level with 6.4 of mean daily occupancy hours. For the first cluster, 36% of cases are Monday, in the second cluster 27% are Tuesday, and in the last cluster, 50% of cases are Friday. Figure 3 presents the dendrogram obtained for each method. In particular, Figure 3b shows the three clusters identified by different colors for each occupancy level.

3.2.2. Electrical usage

Daily measurements of electric power were clustered in the same way as occupancy data. Visual inspection of the dendrograms showed the similarities reported in Table 3. For the all day period, similarities were found between the medium occupancy profile and two electric power clusters (42% and 88% respectively) and with high occupancy and another cluster (67%).

For lunch time, non-similarities were found. A possible explanation is that when the user leaves the office for lunch, equipment is not turned off. For morning and afternoon period, clusters were in accordance in more than a half of cases.

Table 3: Comparison between occupancy and electric power clusters for different time slots

Time Slots		Occupancy Clusters		
		1- Low	2- Medium	3- High
All day	Electric power clusters	-	42% 88%	67%
Morning		63%	-	56%
Lunch time		Non similarities were found		
Afternoon		95%	-	77%

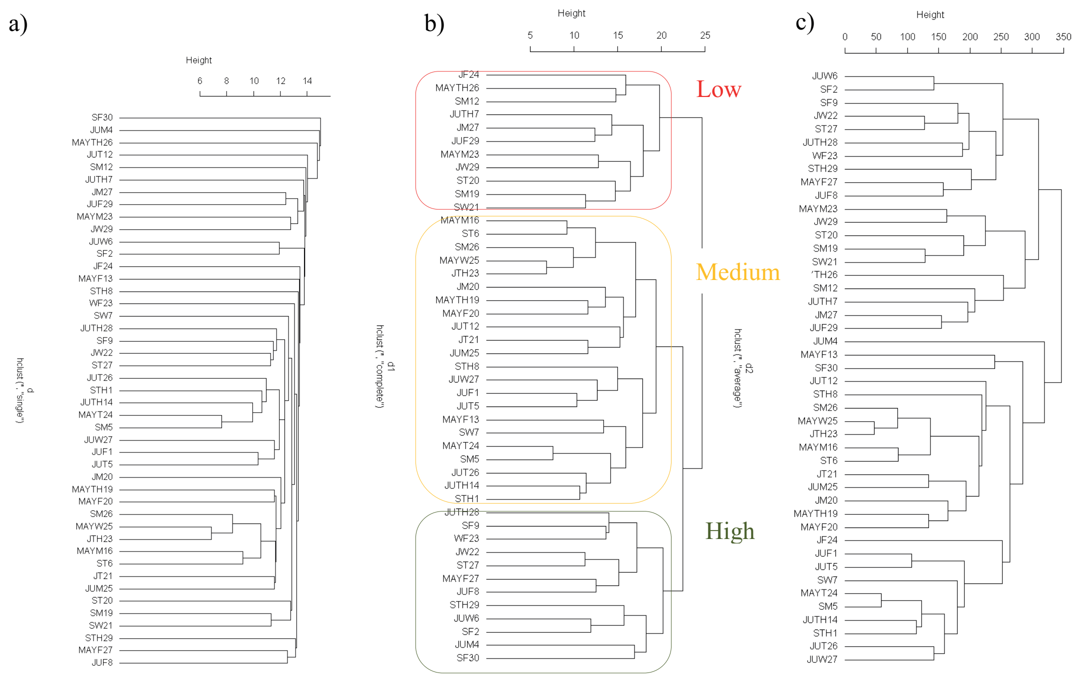

Figure 3: Dendrograms from hierarchical cluster analysis with single (a), complete (b) and average (c) linkage (All day slot)

3.2.3. CO_2 values

The occupancy and CO_2 clusters are compared in Table 4. In the diverse time slots, correspondences were found: for all day the occupancy was registered as medium and high, while in the other intervals the presence can be identified in case of low and high level.

TABLE 4: Comparison between occupancy and CO_2 clusters for different time slots

Time Slots		Occupancy Clusters		
		1- Low	2- Medium	3- High
All day	CO_2 clusters	-	62%	55%
Morning		61%	-	100%
Lunch		82%		85%
Afternoon		60%	-	70%

DOI 10.18502/keg.v3i1.1474

3.2.4. Door Status

Magnetic reed switch to detect door open/close events are used for binary occupancy detection. Table 5 summarizes the similarities between groups. For all day the occupancy was registered with low and medium level almost in the same percent and for other intervals more than 50% of the presence can be identified as low or high.

Table 5: Comparison between occupancy and door status clusters for different time slots

Time Slots		Occupancy Clusters		
		1- Low	2- Medium	3- High
All day	Door status clusters	59%	56%	-
Morning		63%	-	-
Lunch		-	-	59% 79%
Afternoon		67% 52%	-	-

4. Conclusions

In this article, the authors conducted a preliminary analysis of occupancy in an office by using human in-the-loop with manual observations method and sensor network data. Data processing allowed to identify typical occupancy profiles for an office university building and, at the same time, which sensor is more suitable to describe occupancy characteristics.

The comparison between occupancy real data clustering and VOC and open/closed window groups did not demonstrate similarities. Otherwise, electricity consumption, carbon dioxide, and door status showed some similarities, confirming that occupancy profiles could be estimated with a single sensor installation. On the other hand, the quality of the results is affected by the limitations of lack of fine-grained granularity related to the characterization of the spatial resolution of occupancy information which can be obtained.

References

[1] Aldenderfer MS, Blashfield RK (1984) Cluster Analysis. In: Cluster Analysis (Quantitative applications in the social sciences. SAGE Publications, Inc., pp 11–28

[2] Andersen RV (2009) Occupant behaviour with regard to control of the indoor environment. Technical University of Denmark

[3] Dong B, Andrews B, Lam KP, Höynck M, Zhang R, Chiou YS, Benitez D (2010) An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network. Energy and Buildings 42:1038–1046. doi: 10.1016/j.enbuild.2010.01.016

[4] Dong B, Lam KP (2011) Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network. Journal of Building Performance Simulation 4:359–369. doi: 10.1080/19401493.2011.577810

[5] Ebadat A, Bottegal G, Varagnolo D, Wahlberg B, Johansson KH (2013) Estimation of building occupancy levels through environmental signals deconvolution. In: Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13. pp 1–8

[6] Jain AK, Murty MN, P.J. Flynn (1999) Data clustering: a review. ACM Computing Surveys 31:264–323. doi: 10.1145/331499.331504

[7] Page J, Robinson D, Morel N, Scartezzini J-L (2008) A generalised stochastic model for the simulation of occupant presence. Energy and Buildings 40:83–98. doi: 10.1016/j.enbuild.2007.01.018

[8] R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing.

[9] Wiesemann & Theis (2015) WuTility - management and inventorying tool.

[10] Zhang R, Lam KP, Chiou Y-S, Dong B (2012) Information-theoretic environment features selection for occupancy detection in open office spaces. Building Simulation 5:179–188. doi: 10.1007/s12273-012-0075-6

Authorization and Disclaimer

Authors authorize ESTEC to publish the paper in the conference proceedings. Neither ESTEC nor the editors are responsible either for the content or for the implications of what is expressed in the paper.