Learning Individual Thermal Comfort using Robust Locally Weighted Regression with Adaptive Bandwidth

Carlo Manna and **Nic Wilson** and **Kenneth N. Brown**[1]

Abstract.
Ensuring that the thermal comfort conditions in offices are in line with the preferences of the occupants, is one of the main aims of a heating/cooling control system, in order to save energy, increase productivity and reduce sick leave days. The industry standard approach for modelling occupant comfort is Fanger's Predicted Mean Vote (PMV). Although PMV is able to predict user thermal satisfaction with reasonable accuracy, it is a generic model, and requires the measurement of many variables (including air temperature, radiant temperature, humidity, the outdoor environment) some of which are difficult to measure in practice (e.g. activity levels and clothing). As an alternative, we propose Robust Locally Weighted Regression with Adaptive Bandwidth (LRAB) to learn individual occupant preferences based on historical reports. As an initial investigation, we attempt to do this based on just one input parameter, the internal air temperature. Using publicly available datasets, we demonstrate that this technique can be significantly more accurate in predicting individual comfort than PMV, relies on easily obtainable input data, and is fast to compute. It is therefore a promising technique to be used as input to adpative HVAC control systems.

1 INTRODUCTION

One of the primary purposes of heating, ventilating and air conditioning (HVAC) systems is to maintain an internal environment which is comfortable for the occupants. Accurately predicting comfort levels for the occupants can enable one to avoid unnecessary heating or cooling, and thus improve the energy efficiency of the HVAC systems. A number of thermal comfort indices (indicators of human comfort) have been studied for the design of HVAC systems [1,2], the most widely used of which is the Predicted Mean Vote (PMV) index, which was developed by Fanger [1]. This conventional PMV model predicts the mean thermal sensation vote on a standard scale for a large group of people in a given indoor climate. It is a function of two human variables and four environmental variables, i.e. clothing insulation worn by the occupants, human activity, air temperature, air relative humidity, air velocity and mean radiant temperature. The values of the PMV index have a range from -3 to +3, which corresponds to the occupants thermal sensation from cold to hot, with the zero value of PMV meaning neutral.

However, PMV has some drawbacks: *(i)* it requires many environmental data whose retrieval is costly due to the sensors needed, and it requires precise personal dependent data (i.e., clothing and activity level) which are often difficult to obtain in practice; *(ii)* it is a statistical measure which assumes a large number of people experiencing the same conditions, and so may be inaccurate for small groups, or for variable conditions and behaviours within the space, and *(iii)* it requires an expensive iterative evaluation to compute the root of a nonlinear relation.

In this paper, we propose an alternative approach tailored to individual occupants, which relies on historical data on individual responses to internal environment conditions. We apply Robust Locally Weighted Regression [23] with an Adaptive Bandwidth (LRAB), a statistical pattern recognition methods, to learn, automatically, the comfort model of each user based on their history. As a preliminary study, we applied this method with only one input variable (internal air temperature) and compared with PMV, using publicly available datasets [18]. Our experimental results show that LRAB outperforms PMV in predicting individual comfort, and hence it is a promising technique to be used as input to heating/cooling control systems in office environment.

The paper is organised as follows: in the next section, some background on PMV and on alternative techniques are reported. Then, in the section 3, the proposed method is described and in the section 4, the experimental results using a public dataset [18] are shown. Finally, in the section 5, conclusions and future directions are reported.

2 BACKGROUND

The conventional PMV model has been an international standard since the 1980s [3,4]. It has been validated by many studies, both in climate chambers and in buildings [5,6,7]. The standard approach to comfort-based control involves regulating the internal environment variables to ensure a PMV value of zero [8,9,10,11,12].

The PMV model parameters are based on field studies over large populations experiencing the same conditions. For small groups of people within a single room or zone in a building, however, PMV may not be an accurate measure. Kumar and Mahdavi in [17] analysed the discrepancy between predicted mean vote proposed in [1] and observed values based on a meta-analysis of the field studies database made available under ASHRAE RP-884 [18] and finally proposing a framework to adjust the value of thermal comfort indices (a modified PMV). The large field studies on thermal comfort described in [27], have shown that PMV does not give correct predictions for all environments. de Dear and Brager [28] found PMV to be unbiased when used to predict the preferred operative temperature in the air conditioned buildings. PMV did, however, overestimate the subjective warmth sensations of people in warm naturally ventilated buildings. Humphreys and Nicol in [29] showed that PMV was less closely correlated with the comfort votes than were the air temperature or the mean radiant temperature, and that the effects of errors

[1] Cork Constraint Computation Centre, Department of Computer Science, University College Cork, Ireland, e-mail: c.manna@4c.ucc.ie, n.wilson@4c.ucc.ie, k.brown@cs.ucc.ie

in the measurement of PMV were not negligible. Finally the work in [30] also showed that the discrepancy between PMV and the mean comfort vote was related to the mean temperature of the location.

In addition to the relative inaccuracy, the PMV model is a nonlinear relation, and it requires iteratively computing the root of a nonlinear equation, which may take a long computation time. Therefore, a number of authors have proposed alternative methods of calculation to the main one proposed in [1]. Fanger [1] and ISO [4] suggest using tables to determine the PMV values of various combinations between the six thermal variables. Sherman [13] proposed a simplified model to calculate the PMV value without any iteration step, by linearizing the radiation exchange term in Fanger's model. This study indicated that the simplified model could only determine precisely when the occupants are near the comfort zone. Federspiel and Asada [14] proposed a thermal sensation index, which was a modified form of Fanger's model. They assumed that the radiative exchange and the heat transfer coefficient are linear, and they also assumed that the clothing insulation and heat generation rate of human activity are constant. They then derived a thermal sensation index that is an explicit function of the four environmental variables. However, as the authors said, the simplification of Fanger's PMV model results in significant error when the assumptions are not respected. On the other hand, in [15] and [16] different approaches have been proposed in order to compute PMV avoiding the difficult iterative calculation. The former proposes a Genetic Algorithm—Back Propagation neural network to learn user comfort based both on historical data and real-time environmental measurements. The latter proposes a neural network applied to the iterative part of the PMV model that, after a learning phase, based on historical data, avoids the evaluation of such iterative calculation in real-time.

Finally, recent trends in the study of the thermal environment conditions for human occupants are reported in the recently accepted revisions to ASHRAE Standard 55, which includes a new adaptive comfort standard (ACS) [19]. According to de Dear and Brager [20] this adaptive model could be an alternative (or a complementary) theory of thermal perception. The fundamental assumption of this alternative point of view states that factors beyond fundamental physics and physiology play an important role in building occupants expectations and thermal preferences. PMV does take into account the heat balance model with environmental and personal factors, and is able to account for some degrees of behavioral adaptation such as changing one's clothing or adjusting local air velocity. However, it ignores the psychological dimension of adaptation, which may be particularly important in contexts where people's interactions with the environment (i.e. personal thermal control), or diverse thermal experiences, may alter their expectations, and thus, their thermal sensation and satisfaction. In particular, the level of comfort perceived by each individual also depends on their degree of adaptation to the context and to the environmental changes, and therefore the specificity of each individual should be taken into account to learn and predict comfort satisfaction.

For this reason, some authors have proposed techniques based on learning the perception of comfort by individuals. For example, in [21] the author proposes a system able to learn individual thermal preferences using a Nearest Neighbor Classifier, taking into account only four variables (air temperature, humidity, clothing insulation and human activity), acquired by means of wearable sensors. In [22], a Nearest Neighbor Classification-like method was implemented in order to learn individual user preferences based on historical data, using only one variable (air temperature).

In this study we consider such alternative and more practical approaches to predicting thermal comfort through the automatic learning of the comfort model of each user based on his/her historical records. We apply the Robust Locally Weighted Regression [23] technique with an Adaptive Bandwidth (LRAB), one of the family of statistical pattern recognition methods. Non-parametric regression methods, or kernel-based methods, are well established methods in statistical pattern recognition [24]. These methods do not need any specific prior relation among data. Hence, there are no parameter estimates in non-parametric regression. Instead, to forecast, these methods retain the data and search through them for past similar cases. This strength makes non-parametric regression a powerful method due to its flexible adaptation in a wide variety of situations. The Robust Locally Weighted Regression is one of a number of non-parametric regressions. It fits data by local polynomial regression and joins them together. This method was first introduced by Cleveland [23] and further developed for multivariate models [25].

3 THE PROPOSED METHOD

The proposed method is largely inspired by the work in [23]. In the following, we will only describe the proposed LRAB method, while for a more general description of the robust locally weighted regression, the readers should refer to the work in [23].

Before giving precise details on the LRAB procedure, we attempt to explain the basic idea of the method. Let (x_i, y_i) denote a response, y_i, to a recorded value x_i, for $i = 1, \dots, n$. In this paper x_i denotes an environmental variable (in our case air temperature) and the response y_i represents the satisfaction degree (integer-valued on a 7-points scale from -3 to $+3$) that the user has given in response to the condition x_i, and then stored in a database. The aim is to assess the response $\hat{y_k}$ (i.e. predict the degree of satisfaction) for a input value x_k. The approach aims to estimate a local mean, fitting the recorded data by means of a local linear regression centered at x_k. This involves, for a fixed entry point x_k, solving a least squares problem, where α_k and β_k are the values that minimize:

$$\sum_{i=1}^{n} \big(y_i - \alpha_k - \beta_k(x_i - x_k)\big)^2 \omega(x_i - x_k; h) \qquad (1)$$

Then α_k is the response $\hat{y_k}$ for the point x_k. The kernel function $\omega(x_i - x_k; h)$, is generally chosen to be a smooth positive function which peaks at 0 and decreases monotonically as $|x_i - x_k|$ increases in size. The smoothing parameter h controls the width of the kernel function and hence the degree of smoothing applied to the data. This procedure computes the initial fitted values. Now, for each (x_i, y_i), a different weight, ψ_i is defined, based on the residual $(\hat{y_i} - y_i)$ (the larger the residual, the smaller the associated weight). Then, the function (1) is computed replacing $\omega(x_i - x_k; h)$ with $\psi_i * \omega(x_i - x_k; h)$. This is an iterative procedure.

3.1 Kernel function and Adaptive bandwidth

In the following we introduce the concepts of kernel function (and adaptive bandwidth) and residuals, then we describe the algorithm in details. There are many criteria to choose the kernel function based on the theoretical model of the function that has to be fitted. For a locally weighted regression, a common choice is a tri-cubic function, which generally can be written as: $\omega(u) = (1 - |u|^3)^3$, for $|u| \leq 1$, and $\omega = 0$ otherwise [23]. Starting from these considerations, we propose a similar kernel function:

$$\omega(x_i - x_k; h) = \left(1 - \left(\frac{|x_i - x_k|}{h}\right)^3\right)^2 \quad (2)$$

for $|x_i - x_k| \leq h$; otherwise $\omega = 0$. The outer exponent is 2 (in place of 3 as in the standard tri-cubic function), because of empirical considerations (preliminary experiments on smaller set of data were carried out to select the shape of the kernel function).

Finally, we need to choose the bandwidth h. The choice here needs to take into account the fact that the density of the recorded data may be variable. In particular, there may be areas in which the data are clustered closely together (which suggests that a narrow bandwidth would be appropriate), while, on other hand, other areas may be characterised by sparse data (in which case a choice of a large bandwidth is better). In view of this, it would be appropriate to have a large smoothing parameter where the data are sparse, and a smaller smoothing parameter where the data are denser (Figure 1). In this situation an adaptive parameter has been introduced. Let the ratio ν/n (where $\nu < n$), describes the proportion of the sample which contributes strictly positive weight to each local regression (for example if the ratio is 0.7, it means that 70% of the recorded data contributes to the regression). Once we have chosen ν/n (that means we have chosen ν, as n is fixed), we select the ν nearest neighbours from the new entry point x_k. Then, the smoothing parameter h is denoted by the distance of the most distant neighbour among the ν neighbours selected. It should be noted that the entire procedure requires the choice of a single parameter setting.

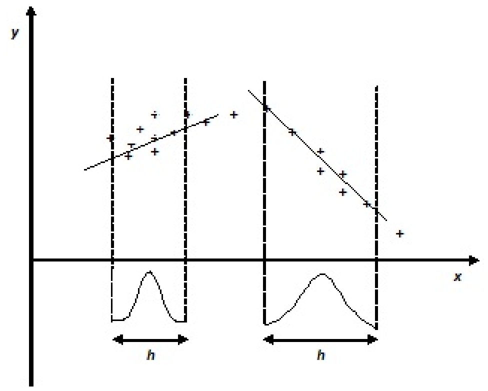


Figure 1. In locally weighted regression, points are weighted by proximity to the current x_k in question using a kernel function. A linear regression is then computed using the weighted points. Here, an adaptive bandwidth h based on the density of the recorded data is proposed.

3.2 Computing the residuals and weights update

In this section we introduce the update mechanism for the weighted function (2), based on the residuals $(\hat{y}_i - y_i)$ for $i = 1, \ldots, n$, as mentioned at start of Section 2. Define the bisquare function:

$$\Gamma(\xi) = (1 - \xi^2)^2 \quad (3)$$

for $|\xi| < 1$; otherwise $\Gamma = 0$.

Then, for a fixed new entry point x_k, let:

$$\rho_i = (\hat{y}_i - y_i) \quad (4)$$

be the residuals for $i = 1, \ldots, n$, between the original points y_i and the estimated points $\hat{y}_i$ (i.e. by means of α_k and β_k), and let m be the median of the $|\rho_i|$. As described in [23], we now choose robustness weights by:

$$\psi_i = \Gamma(\rho_i/6m) \quad (5)$$

At each step of the proposed procedure, the equation (5) is used to update the weight of the function (2) based on the residual ρ_i. In this way the value of the kernel function (2) at each recorded points x_i, is decreased (increased) where the residual value in x_i (i.e. ψ_i) is too high (too low), so as to improve the regression for the next step.

3.3 The algorithm

The proposed method can be described by the following sequence of operations:

LRAB
1: **Initialize**: set parameters ν
2: **For each** entry point x_k:
2.1: **minimise** (1)
2.2: **while** *iterations* $<$ *max iterations* **do**:
2.2.1: **for each** i compute (5)
2.2.2: **minimise** (1) replacing ω with $\psi_i * \omega$
2.3: **end while**
3: **end**

The algorithm is initialized by setting only one parameter (step 1). Then, for each new entry point x_k (step 2), it first computes an initial fitting (step 2.1), then it strengthens the initial regression by the steps 2.2.1 to 2.2.2, performing the sub-procedure described in the previous section, iteratively. If we have K new entry points x_k in total, the steps from 2.1 to 2.3 are repeated K times (one time for each new entry point).

4 EXPERIMENTS

This section describes the experimental results obtained from a comparison between the proposed method and the PMV. Although PMV is not based on a learning approach, in this paper, we compare our method with PMV since the latter is the international standard used to predict comfort in current building design and operation [32,33].

In particular, LRAB has been compared with the PMV index on real data from ASHRAE RP-884 database [18]. This collection contains 52 studies with more than 20,000 user comfort votes from different climate zones. However, some of these field studies contain only a few votes for each user. Thus they are not well suited for testing the proposed algorithm. This is because our approach seeks to learn the user preferences based on their votes, and it requires sufficiently many data records. For this reason, only the users with more than 5 votes have been used to compute the proposed LRAB. After removing the studies and records as described above we were left with 5 climate zones, 226 users and 7551 records (Table 1).

As a starting point, we consider only one environmental variable (i.e. inside air temperature) to evaluate the proposed LRAB. LRAB has been implemented in Matlab™, using the *trust-region method* to minimize the problem in (1), with a termination tolerance of 10^{-6}. The experiments have been performed through *leave-one-out validation*, for each user (i.e., using a single observation from the original

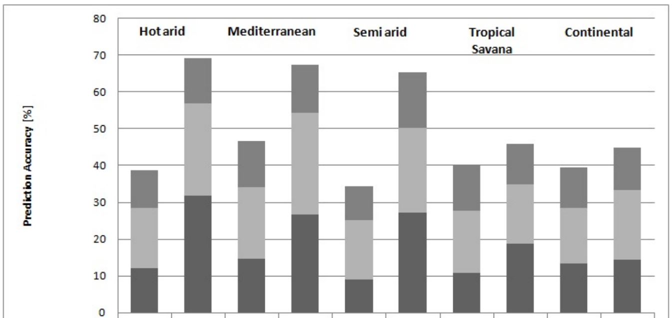
Climate zone	users	records
Hot arid	59	2594
Mediterranean	51	1899
Semi arid	21	2185
Tropical Savana	54	476
Continental	41	397

Table 1. Number of users and records divided by climate zones, and used in the experiments

sample as the validation data, and the remaining observations as the training data).

As with the field study [22, 31], the algorithms are evaluated considering the difference ΔV between the computed votes by both LRAB (evaluated) and PMV (reported in the database) and the actual vote (reported in the database) on a three-level accuracy scale [22, 31] as reported below:

- *Precise*: $\Delta V < 0.2$
- *Correct*: $0.2 \leq \Delta V < 0.5$
- *Approximation*: $0.5 \leq \Delta V < 0.7$



	PMV	LRAB	PMV	LRAB	PMV	LRAB	PMV	LRAB	PMV	LRAB
Approximation	10.1	12.41	12.37	13.06	9.34	15.01	12.61	10.92	11.08	11.59
Correct	16.46	25.21	19.69	27.8	16.11	23.16	17.02	16.18	15.11	18.89
Precise	12.1	31.73	14.53	26.64	9.02	27.09	10.71	18.7	13.35	14.36

Figure 2. Average accuracy of predicting user comfort in 5 different climate zones.

Figure 2 illustrates how accurately the LRAB predicts the actual comfort vote of each user compared with PMV. In all 5 climate zones, LRAB predicts the actual vote better than PMV especially in the accuracy level $\Delta V < 0.2$ and has up to 200% of the number of occupants for whom a precise value is predicted. However, having a close look to the Figure 2, we notice that in the first three climate zones (i.e. hot arid, mediterranean and semi-arid), LRAB achieves more than 25% accuracy ($\Delta V < 0.2$), while in the last two climate zones (i.e. tropical savana and continental) it achieves only 16% and 18% of the same accuracy level respectively. We think that this discrepancy is because there is a lack of data for the last two climate zones compared the first three. In fact, in the former we have 476 records on 54 users (tropical savana) and 397 records on 41 users (continental), as shown in Table 1. Hence, in these case studies we have 8.8 and 9.6 records per user on average, respectively. Conversely, in the first three climate zones, we have 44, 37 and 104 records per user on average, respectively (table 1). As the proposed LRAB essentially *learns from the data*, it requires a sufficient *amount of data* to give the best results.

5 CONCLUSIONS AND FURTHER STEPS

In the present paper, we have applied robust locally weighted regression with an adaptive bandwidth to predict individual thermal comfort. The approach has been characterized and compared with the standard PMV approach. The experiments were carried out using publicly available datasets: they have shown that our LRAB outperforms the traditional PMV approach in predicting thermal comfort. Since LRAB can be computed quickly, and requires only a single setting parameter that is easily obtained, then if individual comfort responses are available, this method is feasible for use as a comfort measure in real time control.

The next step will be:*(i)* the comparison of our LRAB to other (nonparametric) regression mechanisms (e.g., CART, neural networks, k-NN) and *(ii)* the extension of the method to accept multiple environment variables (for example humidity, external air temperature etc.) in order to improve the above results. This mainly means the choice of a different kernel function to the one used here, in order to avoid a bias problem on the boundaries of the predictor space, a kind of problem that may be arise especially in the multidimensional case [26].

This work is part of the Strategic Research Cluster project ITOBO (supported by the Science Foundation Ireland), for which we are acquiring occupant comfort reports and fine grained sensor data, and constructing validated physical models of the building and its HVAC systems. The intention is then to use the comfort reports and sensor data as input to our LRAB method, and then to use the output of LRAB as the input to intelligent control systems which optimise the internal comfort for the specific individual occupants.

6 ACKNOWLEDGEMENTS

This work is supported by Intel Labs Europe and IRCSET through the Enterprise Partnership Scheme, and also in part by SFI Research Cluster ITOBO through grant No. 07.SRC.I1170.

REFERENCES

[1] Fanger, P. O. Thermal comfort: analysis and applications in environmental engineering. New York: McGraw-Hill; 1972.
[2] Gagge, A.P., Fobelets, A.P., Berglund, L.G. A standard predictive index of human response to the thermal environment. ASHRAE Trans 1986;92(2B):70931.
[3] ASHRAE-55. Thermal environmental conditions for human occupancy. American Society of Heating, Refrigerating and Air-conditioning Engineers Inc.; 1992.
[4] International Standard ISO 7730, Moderate thermal environments-estimation of the PMV and PPD indices and specification of the conditions for thermal comfort, Geneva; 1987.
[5] Busch, J.F., A tale of two populations: Thermal comfort in air conditioned and naturally ventilated offices in Thailand. Energy Buildings 1992;18:23549.
[6] de Dear, R. J., Leow, K.G., Ameen, A., Thermal comfort in the humid tropics. Part I. Climate chamber experiments on temperature preferences in Singapore. ASHRAE Trans 1991;97(1):8749.
[7] de Dear, R.J., Auliciems, A., Validation of the predicted mean vote model of thermal comfort in six Australian field studies. ASHRAE Trans 1985;91(2):45268
[8] Funakoshi, S., Matsuo, D., PMV-based train air-conditioning control system. ASHRAE Trans 1995;101(1):42330.
[9] MacArthur, J.W., Humidity and predicted-mean-vote-base (PMV-based) comfort control. ASHRAE Trans . 1986;92(1B):517.
[10] Scheatzle, D.G., Arch, D., The development of PMV-based control for a residence in a hot and arid climate. ASHRAE Trans 1991;97(2):100219.
[11] Wai, L.T., Albert, T.P.S., Implementation of comfort-based air-handling unit control algorithms. ASHRAE Trans . 2000;106(1):2944.

[12] Yang, K.H., Su, C.H., An approach to building energy savings using the PMV index. Building Environment 1997;32(1):2530.
[13] Sherman, M., A simplified model of thermal comfort. Energy Buildings 1985;8:3750.
[14] Federspiel, C.C., Asada, H., User-adaptable comfort control for HVAC systems. Trans ASME 1994;116:47486.
[15] Bingxin, M., Jiong, S., Yanchao, W.,Experimental design and the GA-BP prediction of human thermal comfort index, Seventh International Conference on Natural Computation,vol.2, pp. 771 - 775, 2011
[16] Atthajariyakul, S.,Leephakpreeda, T., Neural computing thermal comfort index for HVAC systems, Energy Conversion and Management, Volume 46, Issues 1516, September 2005, Pages 2553-2565, 10
[17] Kumar, S.,Mahdavi, A., Integrating thermal comfort field data analysis in a case-based building simulation environment, Building and Environment, Volume 36, Issue 6, pp 711-720, 2001
[18] de Dear, R.J., A global database of thermal comfort experiments, ASHRAE Trans., 104(1):1141-1152, 1998.
[19] ASHRAE. ASHRAE standard 55: thermal environmental conditions for human occupancy. American Society of Heating, Refrigerating and Air-conditioning Engineers; 2004.
[20] de Dear, R.J., Brager, G.S., Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55, Energy and Buildings, Volume 34, Issue 6, Pages 549-561, 2002
[21] Feldmeier, M.C., Personalized Building Comfort Control, PhD Thesis, Massachusetts Institute of Technology, Sept. 2009
[22] Schumann, A., Wilson, N., Burillo, M., Learning User Preferences to Maximise Occupant Comfort in Office Buildings, IEA/AIE (1) 2010: 681:690.
[23] Cleveland, W.S., Robust Locally Regression and Smoothing Scatterplots, J. of the American Statistical Association,Vol.74, No. 368, 829-836, 1979.
[24] Hastie , T., Tibshirani, R., Friedman, J., The elements of Statistical Learning, Springer, 2001
[25] Cleveland, W.S., 1988. Locally weighted regression: an approach by local fitting. Journal of the American Statistical Association, 88, 596-610.
[26] Hastie , T. and Loader, C., Local regression: automatic kernel carpentry, Statist. Science 8, 120-143, 1993.
[27] Humphreys, M. A., Nicol, J. F., The validity of ISO-PMV for predicting comfort votes in every-day thermal environments, 2002, Energy and Buildings, 34(6), 667-684
[28] de Dear, R.J., Brager, G.S., Developing an adaptive model of thermal comfort and preference, Field studies of thermal comfort and adaptation, ASHRAE Technical Data Bulletin 14 (1) (1998) 27-49; ASHRAE Transactions 104 (1) (1998).
[29] Humphreys, M.A., Nicol, J.F., Effects of measurement and formulation error on thermal comfort indices in the ASHRAE database of field studies, ASHRAE Transactions 106 (2) (2000) 493-502.
[30] Humphreys, M.A., in: Proceedings of the World Renewable Energy Congress on the Recent Progress in the Adaptive Approach to Thermal Comfort, Brighton, UK, 2000, 438-443.
[31] IS0/DIS 7730:2003, Ergonomics of the thermal environment Analytical determination and interpretation of thermal comfort suing calculation of the PMV and PPD indices and local thermal comfort, 2003.
[32] ASHRAE-55. Thermal environmental conditions for human occupancy. American Society of Heating, Refrig- erating and Air-conditioning Engineers Inc.; 1992.
[33] International Standard ISO 7730, Moderate thermal environments-estimation of the PMV and PPD indices and specification of the conditions for thermal comfort, Geneva; 1987.