ISSN 2224-087X. Електроніка та інформаційні технології. 2018. Випуск 10. С. 99–108
Electronics and information technologies. 2018. Issue 10. P. 99–108

UDK 681.5

STATISTICAL ANALYSIS OF THE THERMAL PARAMETERS OF SMART HOMES

O. Sinkevych, L. Monastyrskii, B. Sokolovskyi

Ivan Franko National University of Lviv,
50 Drahomanova st., UA-75005, Lviv, Ukraine

oleh.sinkevych@lnu.edu.ua, lyubomur.monastyrskyy@lnu.edu.ua, bohdan.sokolovskyy@lnu.edu.ua

A statistical analysis of the annual smart home sensor data obtained from REFIT smart homes project is considered. These data, in particular, consist of the high frequency external, internal, radiator surface temperatures and the gas usage values selected from 20 buildings in a period between March 7, 2014 and March 7, 2015. The data were obtained using sensors installed in twenty UK dwellings under REFIT programme dedicated to the research problem of energy savings for the smart homes. Firstly, a sqlite database for the raw data readings was created. Secondly, data preprocessing and resampling of the data were made in order to fill missing values, smooth and decrease the frequency. The correlation analysis of different data readings showed relations which allow one to use the data not only for forecasting problems, but also to formulate inverse problems of the determination of the effective thermophysical coefficients.

Key words: smart home, home automation, data acquisition, data mining.

1. Introduction. The problem of developing the effective energy managing systems (EMS) of smart home is widely studying nowadays. The big part of these studies is dedicated to the problems of analysing the heating behaviour [1], [2] in order to propose the effective algorithm for EMS which takes into account comfortable indoor temperature conditions for the residents of smart home. In the paper [3] authors suggest an artificial neural network which captures human behaviour to learn periodic patterns and adapt it to the minimization of power wastage. In [4] this idea is realized by applying Window Sliding with De-Duplication algorithm. Also, the problem of minimizing power consumption leads researchers to develop effective smart systems for managing electricity demand by efficiently shifting electricity loads of households from peak times to off-peak times [5]. The last one can be integrated with the renewable energy sources like solar panels in combination with batteries [6]. This schema can work without renewable energy sources as well. In addition, the application of the artificial neural network for EMS is studied in [7] where the whole system consists of photovoltaic (PV) local energy generator, an electricity storage system, home grid and automation system.

Last but not least usage of the artificial neural network for the effective EMS is the development of predictive model for calculating ascent time to a target room temperature set by a home resident [8].

Besides these approaches hardware implementations, algorithms for a processing of signal data and optimization, energy disaggregation are still challenging problems for

researches. Promising results in application of recurrent neural network and emergence of different scientific frameworks (tensorflow, keras, etc.) allow researchers to comprehensive studying and to propose new highly effective methods to handle the problems due to smart home and smart grid solutions.

The aim of the present work is the primary statistical analysis of smart home data to study relationships between different classes of sensor data and to estimate data for the application of forecasting and solving the inverse problem which leads to smart home improvements.

2. Sensor data description and pre-processing. The data for the statistical investigation of the smart home thermal parameters were taken from the REFIT (Personalised Retrofit Decision Support Tools for UK Homes Using Smart Home Technology) project. The aim of this project was to create a step-change of a retrofit technology for UK homes and to improve the energy consumption behaviour based on an extensive sensor data collected from 20 householders. We have chosen this dataset for a couple of reasons: 1) we need sufficiently vast and high frequency data to build statistical models and provide quantitative analysis; 2) the fact that climate conditions, for which the data were collected, are comparable enough to Ukrainian climate conditions, allows researchers to use the results for the local smart home retrofit modelling. The data structure is organized in the three parts. The first is the values.csv file with measurements recorded by the sensors placed in the buildings. This file is organized as a list of id's, dates and values for each sensor. The second part is the schema.xsd schema file which describes the structure of the corresponding structure.xml file. The last part is the structure.xml file containing information about sensors and buildings. To reduce PC memory usage caused by reading > 1 GB csv files and minimize time required to work the sqlite database was used (Fig. 1). For each building and for each space the corresponding tables for sensor values were created: building[i] → space[j] → sensor[k]: {sensorType, date, value}.

In the current research external temperature measurements, indoor temperature measurements for each space[j] of the building[i], radiator surface temperatures for each space, where the radiator is mounted, and boilers gas consumption were selected from the REFIT database (using sql queries and Python 3.5). These types of measurements were chosen in order to provide statistical analysis related to heating processes of the smart homes. It should be mentioned that gas was used not only for the direct heating process, but also for other appliances, e.g. showers, dishwashers, etc. The gas usage is assumed to be taken into account only for the direct heating of houses, hence energy disaggregation problem to separate heating/non heating resources will be considered in the future work.

To convert gas measurements $g_{i|}|_{m^3}$ obtained in cubic meters to energy usage $g_i|_{kWh}$ (kWh), where $i = 1, N$, N – total number of readings, the next formula was used [1]:

$$g_i|_{kWh} = (g_i|_{m^3} \cdot c \cdot cv) c_f, \quad (1)$$

where $c = 1,02264$ is the industry standard conversion factor, $cv = 39,3$ is the calorific value and $c_f = 3,6$ is the conversion factor.

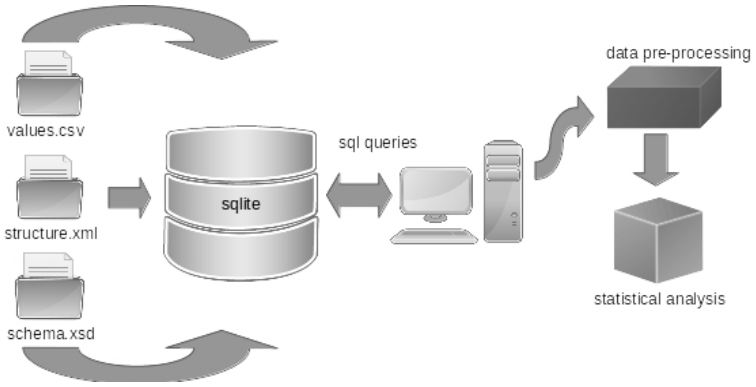


Fig. 1. Schema of data reading and operation

The reasons why the data interval between March 7, 2014 and March 7, 2015 was chosen were monthly measurement gaps up to March 7, 2014 and the interest to study not fragment but annual data. The last ones guarantee acquiring seasonal patterns in temperature and gas usage which allows using them for the determination of effectiveness of heating process and forecasting.

Raw data $(d_1, x_1), (d_2, x_2), ..., (d_i, x_i), ..., (d_N, x_N)$, where d_i is the date time, x_i is the corresponding value obtained by a sensor, were preprocessed in the three steps: 1. filling small gaps and missing values; 2. smoothing data; 3. resampling by selected time intervals to get clean data for the statistical analysis. It is assumed that the data step is constant for each sensor reading.

1. Filling small gaps and missing values. In the case of missing value x_i and known x_{i-1}, x_{i+1} the simple averaging formula can be used

$$x_i = (x_{i-1} + x_{i+1}) / 2 . \quad (2)$$

If neighboring sequences of length n correspond to the same pattern, this formula can be expanded to

$$x_i = \frac{1}{2} \sum_{j=i-n}^{i+n} x_j , \; j \neq 1 . \quad (3)$$

In case of larger gaps in the data, the well-known state-of-art interpolation methods [9] should be considered, e.g.:

$$x_i = (1-\alpha) x_a + \alpha x_b , \; x_i \in [x_a, x_b] , \quad (4)$$

where a and b are the endpoints of missed interval (gap), $\alpha \in [0,1]$ is the interpolation factor.

For this study the linear interpolation method was chosen, because the lengths of gaps in selected interval weren't too large. For larger gaps the more precise methods are recommended.

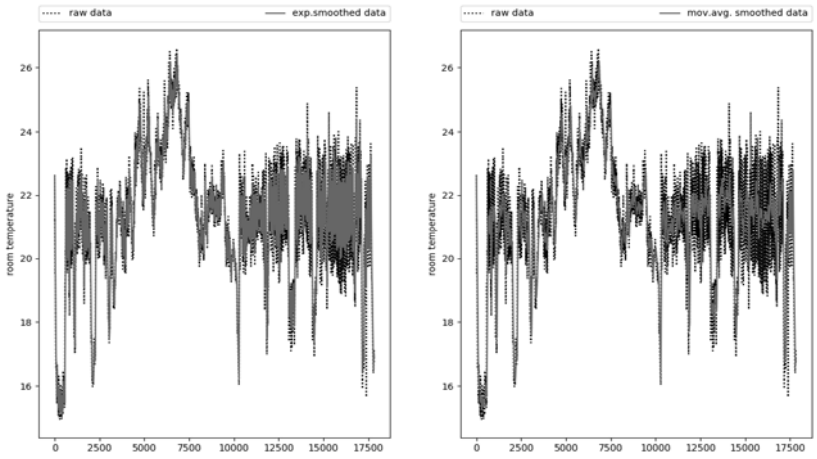


Fig. 2. Raw data and exponential/moving average smoothed data

2. Smoothing raw data. To smooth 'spikes' due to high frequency of the data and possible errors which were accumulated during gathering the sensor data, such methods like moving average [10] or exponential smoothing [11] are considered to be the optimal choice because of simplicity and algorithm's speed. In Fig. 2 the application of these approaches is shown for the room temperature data.

For an exponential smoothing method defined as

$$e_1^s = x_1,\ e_{i+1}^s = \alpha x_{i+1} + (1-\alpha)e_i^s,\ \alpha \in [0,1], \tag{5}$$

where parameter where parameter α was set to be equal to 0,1.

For instance, for a moving average method

$$m_i^a = \frac{1}{w}\sum_{j=1}^{w-1} x_{i-j}, \tag{6}$$

where the parameter w is the length of data points for calculation of the moving average value, was chosen equal to 48 (number of points obtained by sensors during the half of day).

A selection of proper method for the smoothing needs advanced investigation based on type of the data, frequency, detection of seasonal component, etc. Hence, for the sake of simplicity a simple exponential smoothing was used to remove a minor amount of 'spikes' which might have significant impact on resulting data.

3. Resampling smooth data. A daily time step resampling by the averaging smoothed data readings was done with the use of the next formula

$$x_d = \frac{1}{dp}\sum_{j=1}^{dp} x_j,\ \ x_j \in [x_{k_0}, x_{k_{dp}}], \tag{7}$$

where x_d is the daily averaging value, dp is the number of points during the corresponding day, x_j is the input values during the corresponding day, $[x_{k_0}, x_{k_{dp}}]$ is the interval of data

points for the corresponding day. For more precise investigation, maximum/minimum data values can be considered and resampling interval can be reduced to the smaller parts of the day.

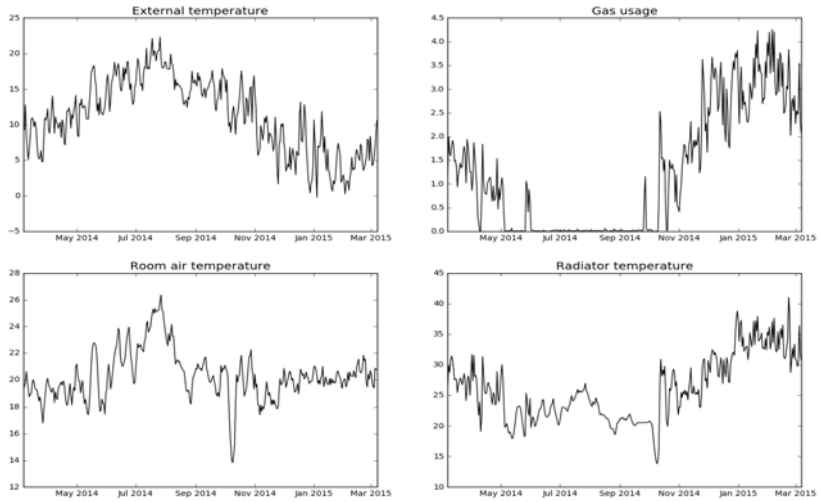


Fig. 3 The averaged resampled data

The calculated via these algorithms resampled data were stored in the database in accordance to the following structure: building → {external temperature {date, value}, gas usage [kWh] {date, value}, internal temperature {date, space {values $\overline{1,k}$ }}, radiator surface temperatures {date, value}}.

In Fig. 3 the averaged re-sampled data (external temperature, arbitrary selected room air and surface radiator temperatures, gas usage) for building #4 are shown.

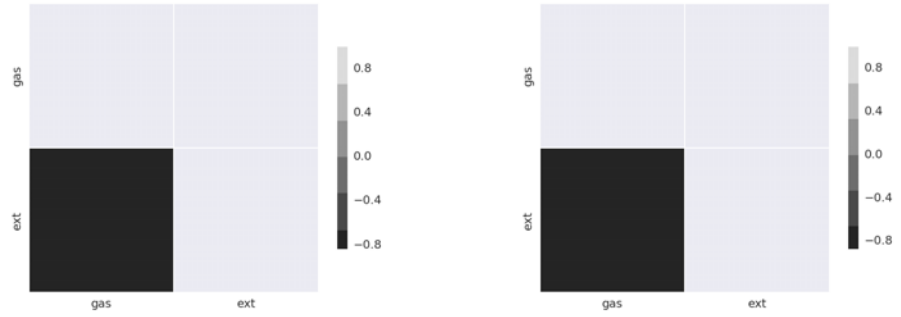
3. Determination of statistical parameters of the preprocessed data. To investigate the relations between the obtained external temperatures, room air temperatures, surface radiator temperatures and gas usage, the calculation of Pearson's correlation coefficients $p_{(X,Y)}$ was performed via the following formula

$$p_{(X,Y)} = \frac{\sum_{i=1}^{N}(x_i - x_{mean})(y_i - y_{mean})}{\sqrt{\sum_{i=1}^{N}(x_i - x_{mean})^2 \sum_{i=1}^{N}(y_i - y_{mean})^2}}, \quad (8)$$

where X and Y are the averaged resampled data sequences, $x_i \in X$, $y_i \in Y$, x_{mean} and y_{mean} are the corresponding mean values of X and Y.

In Fig. 4-7 the correlation values are shown for different combinations of the data: a) the external temperatures and the gas usage; b) the gas usage and surface radiator temperatures for each space where a radiator is mounted; c) the external temperatures and surface radiator

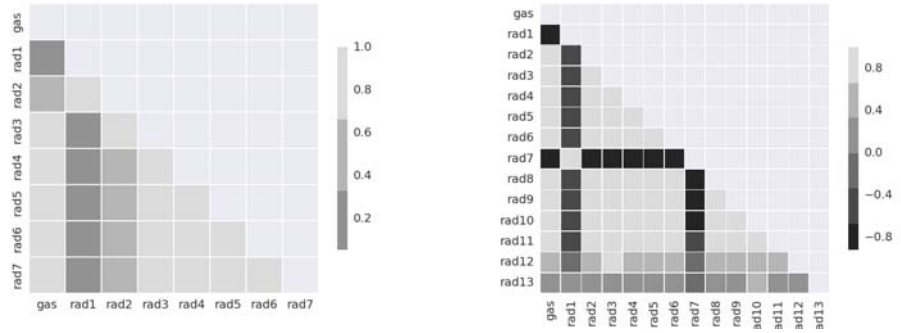
temperatures, d) ΔT_k and surface radiator temperatures, where ΔT_k is defined as $\Delta T_k = T_{int_k} - T_{ext}$ (T_{int_k} is the internal space temperatures for space *k*, where the corresponding radiator is mounted, T_{ext} is the external temperature). These correlations were calculated for two arbitrary chosen buildings: #1 and #4. The magnitude of correlations in these figures is reflected in grayscale: the greater correlation is, the less the intensity.



ext. temp. ~ gas usage
-0.84

ext. temp. ~ gas usage
-0.87

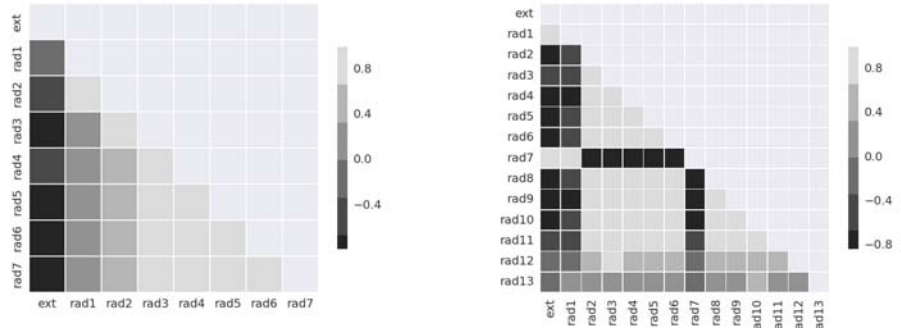
Fig. 4. Correlations between the external temperature and gas usage for buildings #1, #4



gas usage ~ rad.1, rad.2, …, rad.7
0.28 0.69 0.92 0.73 0.98 0.97 0.97

gas usage ~ rad.1, rad.2, …, rad.13
-0.81 0.93 0.78 0.94 0.95 0.93 -0.91
0.88 0.89 0.90 0.76 0.44 0.17

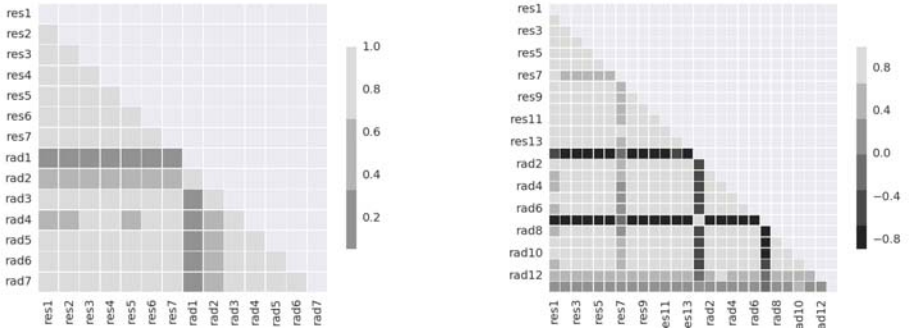
Fig. 5. Correlations between the gas usage and surface radiator temperature for buildings #1, #4



ext. temp. ~ rad.1, rad.2, …, rad.7
-0.17 -0.4 -0.7 -0.56 -0.78 -0.75

ext. temp. ~ rad.1, rad.2, …, rad.13
0.89 0.75 -0.52 -0.73 -0.73 -0.71 0.87
-0.67 -0.75 -0.7 -0.57 -0.19 -0.03

Fig. 6. Correlations between the external temperature and surface radiator temperatures for buildings #1, #4



$\Delta T_1, \Delta T_2, ..., \Delta T_7$ ~ rad.1, rad.2, …, rad.7
0.23 0.54 0.8 0.71 0.89 0.88 0.89

$\Delta T_1, \Delta T_2, ..., \Delta T_{13}$ ~ rad.1, rad.2, …, rad.13
-0.53 0.86 0.73 0.85 0.67 0.83 -0.26
0.82 0.87 0.85 0.74 0.48 0.23

Fig. 7. Correlations between ΔT_k and surface radiator temperatures for buildings #1, #4

4. Results and discussion. As it was expected there are high negative correlations: -0.84 and -0.87 between the gas usage and external temperatures calculated for building #1 and #4 (see Fig. 4). Moreover it was discovered that the distinction between two correlation values calculated for two buildings obviously related, in particular, to the difference in: 1) thermophysical parameters such the coefficients of thermal conductivity and heat exchange; 2) spaces (rooms) ventilation; 3) space volumes.

Also, it was expected that correlations between the gas usage and the surface radiator temperatures (see Fig. 5) would be positively equivalent to the correlations between the gas usage and external temperatures but some negative correlation values were found for building #4. The conducted analysis of data-out allows one to identify that these anomalies were caused by the absence of heating in respective spaces (Fig. 8). In these spaces the temperature distributions are determined not only by the gas usage, but also by the external temperature and

temperature distributions in the neighbouring spaces. These results in combination with the disaggregated gas usage can be as an indication of the absence of heating and can be used to identify non-heated spaces.

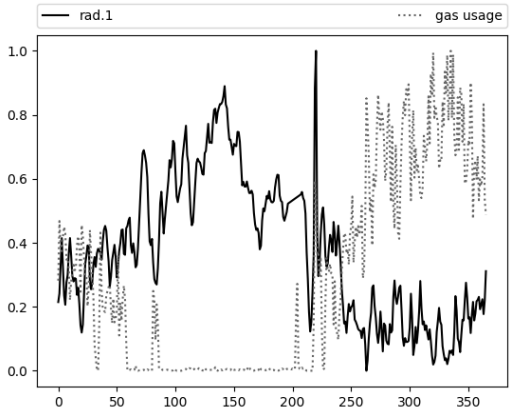


Fig. 8. Normalized gas usage and surface radiator temperature (radiator #1) for building #4

During the correlation analysis of the external temperatures and the surface radiator temperatures (see Fig. 6) the negative correlation values were confirmed except for the cases of absence of the compulsory heating. If for building #1 the correlation values were predictable enough, then for building #4 the anomalies of the same nature as discovered in previous paragraph were observed.

Except for the correlations between the direct temperatures, the correlation analysis between difference temperatures ΔT_k and the surface radiator temperatures was conducted (see Fig. 7). It provided more precise interpretation of statistical processing of the results. The cases of anomaly negative correlation values reconfirmed preceding results.

The carried out correlation analysis can be helpful for prediction problems of energy demands for smart homes in the case of availability of a meteorological forecast. Based on the high correlation values the accuracy of such forecasting will be high enough too.

Statistically studied data can be used for formulation and solving the inverse problem of estimation of the effective thermophysical parameters (coefficients of thermal conductivity and heat exchange) of the smart homes which are sometimes difficult to obtain via direct calculations.

5. Conclusions. In the current work open access smart home sensor data downloaded from REFIT smart homes project were used to conduct the statistical analysis. During the research the data were stored in developed sqlite database in order to provide quick and convenient data readings. Several techniques to clean, smooth and resample available data were used in the preprocessing stage. The correlation analysis by the Pearson method was done to study the relations between the gas usage, external temperatures, surface radiator temperatures and difference temperatures. This analysis discovered different types of relations depending on weather conditions, compulsory heating processes and thermophysical

parameters. Based on the obtained results a possibility to use this or similar data for solving forecasting and inverse problems was proposed.

REFERENCES

1. *Kane T.* Heating behaviour in English homes: An assessment of indirect calculation methods / T. Kane, S. Firth, T. Hassan, V. Dimitrou // Energy and Buildings. – 2017. – № 148. – P. 89–105.
2. *Zehnder M.* Energy Optimization in Smart Homes Using Customer Preference and Dynamic Pricing [Електронний ресурс] / M. Zehnder, H. Wache, H. Witschel // Smart Cities Conference (ISC2), 2015 IEEE First International. – 2015. – Режим доступу: https://web.fhnw.ch/personenseiten/holger.wache/Papers/zehnder_etal-15.pdf.
3. *Teich T.* Design of a Prototype Neural Network for Smart Homes and Energy Efficiency / T.Teich, F. Roessler, D. Kretz, S. Franke. // Procedia Engineering. – 2014. – №69. – P. 603 – 608.
4. *Schweizer D.* Using consumer behavior data to reduce energy consumption in smart homes / D. Schweizer, M. Zehnder, H. Wache, H. Witschel. // Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on. – 2015. – P. 1123–1130.
5. *Hu R.* A Mathematical Programming Formulation for Optimal Load Shifting of Electricity Demand for the Smart Grid [Електронний ресурс] / R. Hu, R. Skorupski, R. Entriken, Y. Ye // IEEE TRANSACTIONS ON SMART GRID, 10(10). – 2015. – Режим доступу: https://web.stanford.edu/~yyye/OLSforIEEEv5.pdf.
6. *Saha A.* A Home Energy Management Algorithm in a Smart House Integrated with Renewable Energy [Електронний ресурс] / A. Saha, M. Kuzlu, W. Khamphanchai // Conference: IEEE PES ISGT-Europe Conference. – 2014. – Режим доступу: https://www.researchgate.net/publication/271520194_A_Home_Energy_Management_Algorithm_in_a_Smart_House_Integrated_with_Renewable_Energy.
7. *Matallanas E.* Neural network controller for Active Demand-Side Management with PV energy in the residential sector / E. Matallanas, M. Castillo-Cagigal, A. Gutiérrez. // Applied Energy. – 2012. – №91 (1). – P. 90–97.
8. *Moon J.* Performance of a Predictive Model for Calculating Ascent Time to a Target Temperature [Електронний ресурс] / J. Moon, M. Chung, H.Song, S. Lee // Energies 2014, 9(12). – 2016. – Режим доступу: http://www.mdpi.com/1996-1073/9/12/1090/xml.
9. *Kress R.* Numerical Analysis / Raimer Kress., 1998. – 326 с. – (Springer Science & Business Media).
10. *Droke C.* Mastering Moving Averages / Clif Droke., 2014. – 140 с. – (Perfect Paperback). – (ISBN# 978-0-9792572-6-3).
11. *Hyndman R.* Forecasting with Exponential Smoothing: The State Space Approach / R. Hyndman, A. Koehler, J. Ord, R. Snyder., 2008. – 362 с. – (Springer Series in Statistics). – (ISBN-13: 978-3540719168).

СТАТИСТИЧНИЙ АНАЛІЗ ТЕПЛОВИХ ПАРАМЕТРІВ РОЗУМНИХ БУДИНКІВ

О. Сінькевич , Л. Монастирський, Б. Соколовський

Львівський національний університет імені Івана Франка,
вул. Драгоманова, 50, 79005, Львів, Україна

oleh.sinkevych@lnu.edu.ua , lyubomur.monastyrskyy@lnu.edu.ua ,
bohdan.sokolovskyy@lnu.edu.ua

У роботі проведений кореляційний аналіз відкритих для використання даних "розумних будинків", які були отримані в рамках дослідницького проекту REFIT. Метою проекту було здійснення дослідження проблематики енергозбереження в "розумних будинків" для побудови ефективних алгоритмів мінімізації енерговитрат в умовах індустрії 4.0 (Industry 4.0). Серед усіх зібраних даних, зокрема, були виокремленні набори даних, що містять розподіли зовнішньої та внутрішньої температур, температури на поверхні обігрівальних елементів та розподіл споживання газу у 20 будинках за період від 7 березня 2014 року до 7 березня 2015 року.

Для швидкого і зручного доступу до отриманих даних було створено sqlite базу даних, інтегровану з Python API. З використанням розроблених алгоритмів здійснено обробку сирих даних, яка включала заповнення пропущених значень у часових рядах методом інтерполяції, згладжування нехарактерних "піків" та пониження частоти даних. Оброблені таким чином дані були розміщені в базі даних для подальшого їх використання у дослідженні.

Для оцінки залежностей між наборами даних (споживання газу та зовнішніми температурами/температурами на поверхні обігрівальних елементів, температурами на поверхні обігрівальних елементів та зовнішніми температурами, різницею між внутрішньою і зовнішньою температурами та температурами на поверхні обігрівальних елементів) проведений кореляційний аналіз та обчислені кореляційні коефіцієнти. Значення цих коефіцієнтів підтвердили гіпотезу про високу кореляцію між наборами даних та дали кількісний критерій для визначення неопалювальних приміщень будинків. Також, на основі результатів кореляційного аналізу була встановлена можливість використання даних такого типу не лише для задач прогнозування витрат, а й для формулювання обернених варіаційних задач визначення ефективних теплофізичних коефіцієнтів "розумних будинків", наприклад, ефективного теплообміну та теплоємності.

Ключові слова: розумний будинок, автоматизація будинку, система збору даних, інтелектуальний аналіз даних.

Стаття: надійшла до редакції 23.05.2018,
доопрацьована 28.05.2018,
прийнята до друку 29.05.2018