Identification of Indoor Air Quality (IAQ) Sources in Libraries through Principal Component Analysis (PCA)

A Abu Mansor1*, S Shamsul2, S Abdullah2,5, N C Dom3, N N L Mohd Napi2, A N Ahmed4 and M Ismail1,5*

1 Faculty of Science and Marine Environment, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
2 Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
3 Centre of Environmental Health & Safety, Faculty of Health Sciences, Universiti Teknologi MARA, UiTM Cawangan Selangor, 42300 Puncak Alam, Selangor, Malaysia
4 Institute of Energy Infrastructure (IEI), Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Selangor Darul Ehsan 43000, Malaysia
5 Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

*Corresponding author: marzuki@umt.edu.my

Abstract. Indoor air quality (IAQ) is an important role in affecting visitors and staffs’ health in library. However, good ventilation and suitable types of furniture materials in the indoor environment were important to preserve better indoor air quality for the occupants. In this study, the status of indoor air pollution using in-situ measurement for chemical and physical parameters had been examined. The chemical parameters measured were formaldehyde (CH₂O, ppm), carbon monoxide (CO, ppm), carbon dioxide (CO₂), fine particulate matter (PM₂.₅, mg/m³) and coarse particulate matter (PM₁₀, mg/m³). The physical parameters such as relative humidity (RH, %), temperature (T, °C) and air movement (AM, m/s) were also measured. The dominant sources of IAQ had been determined by using Principal Component Analysis (PCA). Principle Component (PC-2, S1) and PC-1 (S2) (CO, CO₂ and CH₂O) have become the indicator for chemical contaminants and the number of occupants that present in the library. The main contributor of indoor air quality at both areas was thermal comfort, which contributed 56.146% (PC-3, S1) and 60.76% (PC-4, S2). In conclusion, air quality in the libraries was affected by surrounding activities, ventilation performance and physical parameters.

1. Introduction
Indoor Air Quality (IAQ) has a significant impact on human’s health and wellbeing, particularly those with cardiovascular diseases, who are vulnerable and sensitive to air pollutants [1]. Poor IAQ may lead to illness, loss of concentration, drowsiness, and tiredness, as well as adverse health symptoms, such as respiratory problems or headache, lung cancer and acute respiratory infection, heart disease and stroke [2-5]. 9 out of 10 people are estimated exposed to air quality problems and South-East Asia region recorded higher air pollution problems, compared to Eastern Mediterranean region [2]. Nowadays, most people were spending their time over 80% indoors [2-3]. Poor IAQ might become a health risk such as respiratory-related illness to sensitive people. The sources of air pollutants are hard
to be determined as it might be influenced by different geographical locations, types of buildings and climate conditions [2,4]. The occupants might expose to different particles concentration and sizes in indoor and outdoor air pollution [4]. The aim of this study was to investigate the gaseous pollutants footprints and physical parameters by using the Principal Components Analysis (PCA). PCA has two functions in air quality: as data summarisation and reduction of parameters or variables. Data monitored can be quantified as sources of air pollution in a study. Previous studies had successfully utilised the PCA for source apportionment. In a refinery industry, the main source is chemical contaminants, followed by ventilation, and thermal comfort with 35.51%, 21.98%, and 18.48%, respectively [5]. This method can determine the same group of parameters which simultaneously tracing the source of pollution [5-8].

2. Material and Methods

2.1. Site Selection

In this study, two different sites were selected which are Sultanah Nur Zahirah Library (PSNZ) of Universiti Malaysia Terengganu (UMT) (524.498'N, 1035.300' E) (S1) and Public Library Terengganu State (5°18.805'N, 103°7.764'E) (S2) as shown in Figure 1. Sultanah Nur Zahirah Library (S1) was located inside institutional area. PSNZ has two levels of buildings, which contain facilities of furniture, books, and computers. Public Library Terengganu is one of the public libraries located in the city Centre. It has four levels of buildings that contain the same facilities as S1.

![Figure 1. Study areas](image)

2.2. Data Analysis

The monitoring activities were conducted from 0800 hrs until 1700 hrs (working hour) for three days within 15 minutes of the time interval. The physical and chemical parameters such as temperature (T, °C), relative humidity (RH, %), air movement (AM, m/s), carbon dioxide (CO₂, ppm), carbon monoxide (CO, ppm), formaldehyde (CH₂O, ppm) and particulate matter (PM₂.5 and PM₁₀, µg/m³) were measured. The monitoring equipment were placed at least 2m away from windows, doors, or active cooling system, and 1m above the floor. The SPSS software (Statistic Package for Social Science) version 24 was used to proceed with the statistical analysis of the data.

Normalization was required due to different measurement unit of the input parameters of the model. The range of 0 to 1 produced from normalization process, which can improve the accuracy of the numeric computation. Hence, these normalization steps can interpret all relationships in the data precisely and reduce bias. Equation (1) shows the normalization equation used in this study:

\[
Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}
\]

Where, \(x = (x_1, x_2, ..., x_n)\) and \(Z_i\) is the \(i\)th normalized data.
Principal Component Analysis (PCA) approaches was done by determining interrelationship between several factors and transforming them into a few independent comprehensive variables according to dimension reduction [2, 9-10]. The principle component (PC) shown in the equation below:

$$PC_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n$$ (2)

Where $PC_i$ is $ith$ principle component; $x_1, x_\ldots, x_n$ is the $ith$ independent variables of principle component; $a_{i1}, a_{i2}, \ldots, a_{in}$ is the coefficient of the independent variables in the principles component.

3. Results and Discussion

Principal Component Analysis (PCA) needs Kaiser-Meyer-Olkin of Sampling Adequacy with the value of larger than 0.50 to ensure the adequacy of the data for further analysis. Bartlett Test of Sphericity must be significant at 95 percent level ($p<0.05$) to fulfil the requirement for this analysis. Failures to meet these two requirements caused PCA analysis unable to be proceed. Both data sets for S1 and S2 passed the requirements of PCA, where KMO tests for S1 and S2 were 0.604 and 0.541, respectively, which is more than adequacy of the data (>0.50) and also significant for Bartlett’s Test of Sphericity which as in Table 1. This proved that data set satisfy the requirements for PCA. Extraction values in communalities table was examined to determine the important parameters that contribute at least 50% of the variance contribution in the data set or otherwise, it was removed.

|  | S1  | S2  |
|---|-----|-----|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | 0.590 | 0.517 |
| Approx. Chi-Square | 1413.539 | 465.599 |
| Bartlett's Test of Sphericity | df | 28 | 28 |
| Sig. | .000 | .000 |

After conducting both KMO and Bartlett’s test, all input parameters had an initial proportion of variance 1.0 in Commumalit Table as shown in Table 2. The value of communalities after extraction might be different from initial value but ranged in between 0.0 to 1.0. The communality table shows that extraction values less than 0.5 was remove from PCA. Table 2 shows 6 parameters for S1 which include RH, T, CO$_2$, CO, PM$_{10}$ and PM$_{2.5}$ while 7 parameters that display extraction values more than 0.5 which RH, T, CO$_2$, CO, CH$_2$O, PM$_{10}$ and PM$_{2.5}$.

|  | S1  | S2  |
|---|-----|-----|
| Initial | Extraction |
| Temp  | 1.000 | .525 | .538 |
| RH    | 1.000 | .542 | - |
| AM    | 1.000 | -   | .751 |
| CO$_2$| 1.000 | .625 | .689 |
| CO    | 1.000 | .558 | .554 |
| CH$_2$O| 1.000 | -   | .638 |
| PM$_{10}$| 1.000 | .768 | .674 |
| PM$_{2.5}$| 1.000 | .715 | .693 |
Table 3 shows the eigenvalues that provide a factor that has distinction cleared up by those linear components directly explain eigenvalues in percentage of variance. All elements separated in PCA with eigenvalues more than 1. Table 3 displays the variability factors in both study areas which are 3 factors for 56.146% (S1) and 4 factors with 60.70% (S2). The rotation can affect the factors structure and outcomes of the components that proved before and after rotation. After rotation the factors 1 (21.646%) decrease from initial eigenvalue, factor 2 (21.214%) and factor 3 (13.285%) which shows an increasing percentage of variance for S1. This situation also occurred for S2 which decreasing the percentage of variance after rotation for factor 1 (17.080%) and increasing percentage for factor 2 (16.483%), factor 3 (14.495%) and factor 4 (12.702%).

Table 3. Total Variance Explained for both study areas

| Component | S1  | Extraction Sums of Squared Loadings | Rotation Sums of Squared Loadings | Cumulative % |
|-----------|-----|-------------------------------------|-----------------------------------|--------------|
| 1         | 1.794 | 22.420                              | 22.420                            | 22.420       |
| 2         | 1.638 | 20.469                              | 42.889                            | 42.889       |
| 3         | 1.061 | 13.257                              | 56.146                            | 56.146       |
| 4         | .933  | 11.668                              | 67.814                            | 67.814       |
| 5         | .824  | 10.298                              | 88.112                            | 88.112       |
| 6         | .676  | 8.452                               | 96.564                            | 96.564       |
| 7         | .609  | 7.612                               | 100.000                           | 100.000      |
| 8         | .466  | 5.824                               | 100.000                           | 100.000      |

Rotated component matrix appeared in Table 4 and displays the stacking of every factor onto each factor, where esteem under 0.4 were deleting from yield [11-12]. Principal Component (PC) coefficient shows the commitment of each factors to a PC either positive or negative contribution. The loading factors less than 0.3 are considered as weak, 0.4-0.49 as moderate and strong contribution is more than 0.5 [13]. PC-1 (S1) and PC-3 (S2) shows contribution of PM_{10} and PM_{2.5} up to 21.262% and 48.055%, respectively which being triggered by walking activities in the library besides contribution of the dust from the bookshelves proved by previous studies [13-15]. PC-2 (42.861%, S1) which shows contribution of RH, CO, CO\textsubscript{2} and CH\textsubscript{2}O and PC-1 (17.080%, S2), display contribution of furniture and occupant in the library itself due to previous study proved that CO\textsubscript{2}, CO and CH\textsubscript{2}O depend on RH understandable considering that a ventilation system often used to remove indoor pollutants [5,14]. PC-3 (48.055%, S1) and PC-4 (60.76%) were highest contribution which include air movement inside the building itself. Library ventilation and air-conditioning require properly ventilated areas, which affect the air movement in the library. Proper air-conditioning and ventilation are an important aspect of both reading rooms, where many visitors read, study or use computers, and library halls, where paper slowly acquires its aroma; a regular exchange of air needed [16]. Different sources percentage contribution of air pollutants and physical parameters in both libraries caused by different arrangement in the library and activities in the library itself.
Table 4. Rotated Component Matrix

| Component | S1 | S2 |
|-----------|----|----|
| Temp      | .763 | .721 |
| RH        | .584 | -|-593 |
| AM        | .655 | .632 | .974 |
| CO₂       | .733 | .729 |
| CO        | .515 | .654 |
| CH₂O      | .847 | .729 |
| PM₁₀      | .789 | .727 |
| PM₂.₅     |     |    |

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

4. Conclusion
The potential sources of both libraries were quantified by applying PCA. The findings of this study show that the amount of CO₂ concentrations was influenced by the presence of occupants, and other pollutants such as formaldehyde which caused by the furniture, detergents and permanent-press fabrics. Both pollutants with presence of relative humidity known as the main contributor with the range of 42.861% (PC-2, S1) and 17.080 % (PC-1, S2). The main factor for both sites was air movement, which contribute 56.146% (PC-3, S1) and 60.760% (PC-4, S2) that caused by the insufficient air movement, especially in the occupied zone, and affects most of the indoor air quality.

5. References
[1] Becerra J A, Lizana J, Gil M, Barrios-Padura A, Blondeau P and Chacartegui R 2020 Identification of potential indoor air pollutants in schools, *J. Clean. Prod.* 242 118420
[2] World Health Organization (WHO) (2019). World health statistics 2019: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization, Retrieved from https://apps.who.int/iris/bitstream/handle/10665/324835/9789241565707-eng.pdf?sequence=9&isAllowed=y
[3] Ghorani-Azam A, Riahi-Zanjani B and Balali-Mood M 2016 Effects of air pollution on human health and practical measures for prevention in Iran, *J. Res. Med. Sci.* 21(65)
[4] Radaideh J A 2017 Effect of Meteorological Variables on Air Pollutants Variation in Arid Climates, *J. Anal Toxicol* 07(04) 2161-0525
[5] Abdullah S, Ismail M and Fong S Y 2017 Multiple Linear Regression (MLR) models for long term PM₁₀ concentration forecasting during different monsoon seasons. *J Sustain Sci Mang.* 12(1) 60–69
[6] Peixoto M S, de Oliveira Galvão M F and Batistuzzo de Medeiros S R 2017 Cell death pathways of particulate matter toxicity, *Chemosphere* 188 32–48.
[7] Aung W Y, Noguchi M, Pan-Nu Yi E E, Thant Z, Uchiyama S, Win-Shwe T T and Mar O 2019 Preliminary assessment of outdoor and indoor air quality in Yangon city, Myanmar, *Atmos. Pollut. Res.* 10(3) 722–730
[8] Gui F, Ren S, Zhao Y, Zhou J, Xie Z, Xu C and Zhu F 2019 Activity-based allocation and optimization for carbon footprint and cost in product lifecycle. *J. Clean. Prod.* 236 117627
[12] He C, Morawska L, Hitchins J and Gilbert D 2004 Contribution from indoor sources to particle number and mass concentrations in residential houses. Atmos. Environ. 38(21), 3405–3415
[13] Abdullah S, Ismail M, Fong S Y and Ahmed N A 2016 Evaluation for Long Term PM$_{10}$ Concentration Forecasting using Multi Linear Regression (MLR) and Principal Component Regression (PCR) Models 9(2) 101–110
[14] Ul-Saufie A Z, Yahaya A S, Ramli N A, Rosaida N and Hamid H A 2013 Estimating fine particulate concentration using a combined approach of linear regression and artificial neural network, Atmos. Environ 219 117050
[15] Chatousidou S E and Lazaridis M 2019 Assessment of the impact of particulate dry deposition on soiling of museum/libraries, J. Cult. Heri. 39 221-228
[16] Jaiswal A, Samuel C and Kadabgaon V M 2018 Statistical trend analysis and forecast modeling of air pollutants, Global Journal of Environmental Science and Management 4(4) 427–438

Acknowledgments
This study is funded by the Fundamental Research Grant Scheme (FRGS) by the Malaysian Ministry of Education (FRGS/1/2019/TK10/UMT/02/4) VOT59606. We thanks to the Talent and Innovation Center, University Malaysia Terengganu for the funding. We also would like to thanks to the Air Quality and Laboratory, University Malaysia Terengganu for the provision of instrumentations for the data collection.