Ranking the Environmental Factors of Indoor Air Quality of Metropolitan Independent Coffee Shops by Random Forest Model

Yu-Wen Lin
Fu Jen Catholic University
Chin-Sheng Tang (052340@mail.fju.edu.tw)
Fu Jen Catholic University
Hsi-Chen Liu
Chinese Culture University
Tzu-Ying Lee
Fu Jen Catholic University
Hsiao-Yun Huang
Fu Jen Catholic University
Tzu-An Hsu
Fu Jen Catholic University
Li-Te Chang
Feng Chia University

Article

Keywords: Indoor air quality, independent coffee shop, carbon dioxide, Random Forest, occupant density, meteorological monitoring

Posted Date: June 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1621230/v1

Abstract

Independent coffee shops are the alternative workplaces for people working remotely from traditional offices but are not concerned about their indoor air quality (IAQ). This study aimed to rank the environmental factors in affecting the IAQ by Random Forest models. The indoor environments and human activities of participated independent coffee shops were observed and recorded for 3 consecutive days including weekdays and weekend during the business hours. The multi-sized particulate matter (PM), particle-bound polycyclic aromatic hydrocarbons (p-PAHs), total volatile organic compounds (TVOCs), CO, CO_2, temperature and relative humidity were monitored. Random Forest models ranked the environmental factors. More than 20% of the 15-minute average concentrations of PM_{10}, $PM_{2.5}$, and CO_2 exceeded the World Health Organization guidelines. Occupant density affected TVOCs, p-PAHs and CO_2 concentrations directly. Tobacco smoking dominated PM_{10}, $PM_{2.5}$, TVOCs and p-PAHs concentrations mostly. CO concentration was affected by roasting bean first and tobacco smoking secondly. The non-linear relationships between temperature and these pollutants illustrated the relative low concentrations happened at temperature between 22°C and 24°C. Tobacco smoking, roasting beans and occupant density are the observable activities to alert the IAQ change. Monitoring CO_2 and optimizing the room temperature could also be the surrogate parameters to assure the IAQ.

Introduction

People spend 80-90% of their time in indoor environments, such as homes or workplaces. Therefore, health effects caused by indoor air quality (IAQ) should be addressed. Particulate matter, with aerodynamic diameters £ 2.5 μm ($PM_{2.5}$) and 10 μm (PM_{10}), is the major concerned pollutant in the IAQ. Study showed that most of the indoor $PM_{2.5}$ concentrations were higher than the concentrations of outdoor.[1] In addition, total volatile organic compounds (TVOCs), particle-bound polycyclic aromatic hydrocarbons (p-PAHs) and pollutants from burning solid fuels, such as carbon dioxide (CO_2) and carbon monoxide (CO), are also major indoor air pollutants (IAPs)[2,3]

High concentrations of p-PAHs, such as naphthalene and VOCs were found in coffee shops.[3,4] The Global Workplace Analytics estimates around 4.3 million people work remote at least half the time and as a result the traditional office setting is being replaced by alternative workspaces – the readily-available independent coffee shop is one of choices.[5] More than 70% of independent café consumers surveyed purchase coffee to drink in-store. It was estimated that independent stores served more than 10.5 million cups of coffee each week in UK.[6] So, it is important to understand the IAQ in independent coffee shops as they are served as "workplaces" and the consumers' preferring indoor environment. Besides characteristics of chain coffee shops, independent coffee shops usually roast their own beans on sites. Hence, the levels of carbon monoxide (CO) [7] and VOCs [8] in independent coffee shops shall be addressed.

Some indoor environmental factors affected IAQ. For example, environmental tobacco smoke (ETS) would change the IAQ mentioned in several researches [1,9-11] as well as human activities [1,12-14] and occupant density.[15,16] PM concentration was associated with indoor air flow, humidity, and temperature [17,18]; in addition, it was associated with level of p-PAHs.[19] As a better surrogate of ventilation efficiency and IAQ indicator, CO_2 could be used to represent other pollutants in indoor air other than PM could.[20-25] Therefore, those indoor environmental factors mentioned above might be the indicators to remind staffs of coffee shops to notice IAQ. However, there were few studies assessed which environmental factors would be the useful and simple indicators for IAQ.

Random Forest (RF), a data mining method, has been applied in IAQ studies. Some studies investigated factors associated with selected indoor air pollutants by RF analysis and proofed that RF models had better abilities in prediction than multiple linear regression or other methods.[26-28] Breiman [29] and Horning [30] pointed out that RF can be used to explore relationships among variables without manual settings when data are relatively complicate or the independent and the dependent variables have non-linear relationships. Furthermore, RF has relatively low requirements for the completeness of the data.[29,30] RF seldom used in investigating the importance of factors associated with IAQ. This study described the levels of IAQ and indoor environment in independent coffee shops. The RF models were applied to identify the major factors associated with IAQ and to illustrate the proactive indicators for the IAQ levels in these novel workplaces.

Methods

Recruiting participated coffee shops

We recruited independent coffee shops in the metropolitan area of Taipei, Taiwan and four shops (labeled as A, B, C, and D) participated in this study. The investigations were proceeded from November 2019 to March 2020. At the beginning of each on-site monitoring day, our team members obtained the shops' information including the business hours, the floor plan, type of building, indoor space volume, building materials, smoking area design, ventilation equipment, and window opening situation. The detailed characteristics was listed in Supplement A.

Indoor environments, air pollutants, and meteorological monitoring

Indoor environments of participated shops were collected by activity log and direct-reading instruments. Our team members filled in the activity log on the monitoring day which included the nature of the indoor activity, ventilation status, and the numbers of people with a 15-minute interval. The indoor activities included cooking, roasting beans, cleaning, tobacco smoking, and other behaviors that might change the air quality. The members rechecked the information to assure the correctness after completion. The monitoring was proceeded continuously during the business hours for 3 consecutive days including weekdays and weekends in each coffee shop.

A portable IAQ monitor (Smart Indoor Air Quality Sensing Controller Model GiA-K007, NewGreen Tech Co., Taiwan) was employed to monitor temperature and relative humidity (RH) continuously. This IAQ monitor can also measure CO by a built-in electrochemical CO sensor with a detection range of 0 - 500 ppm and CO_2 by a NDIR CO_2 sensor with a detection range of 0 - 10,000 ppm. The portable aerosol analyzer (Model 1.108, Grimm Aerosol Technik GmbH & Co. KG, Ainring, Germany) was used to measure the PMs at a flow rate of 1.2 L/min. The mass concentrations of PM_{10} and $PM_{2.5}$ were selected. Mass concentrations of $PM_{2.5-10}$ (coarse PM) were obtained by subtracting the $PM_{2.5}$ fraction from the concurrent PM_{10} levels. A photoelectric aerosol sensor (PAS2000CE, EcoChem Analytics, League City, TX, USA) was used to measure indoor p-PAHs level with the detection range of 0-4,000 ng/m^3. In addition, we used a ppbRAE 3000 photoionization detector (PID) (model ppbRAE 3000; RAE systems, Inc., USA) with a 10.6 eV lamp with an extended range of 1 – 10,000 ppm to quantify TVOCs. All the monitoring instruments were set to output one value every minute. In addition to the routine calibration and maintenance of the instruments used in this study, the research staff also performed essential calibration for the instrument readings and pump flows before and after each field survey. The sampling spots (i.e., the location of the instruments) in each shop were shown in Figure 1.

Statistical analysis

To ensure the quality of data processing, Microsoft Excel (2019) was used for data management and descriptive analysis. Zero, negative, missing, and unreasonably high and low values were excluded, as were continuous values in a range more than ten folds. The data (pollutants' concentrations and meteorological data) were synchronized with the people counts. The concentrations were expressed as a 15-minute average and the occupant density was the number of indoor people counts dividing by the floor area.

RF model (R, 3.5.1) was built to examine the importance of indoor environmental factors associated with specific indoor air pollutant. The environmental factors include coffee shop, weekday, occupant density, indoor activities, ventilation status, locating on the main traffic street, and meteorological data. As a surrogate of air change rate, CO_2 concentration was also served as a potential determinant of other air pollutants. The 15-minute average concentrations of indoor air pollutants were the dependent variables. The "error rate" code of RF was used to evaluate the classification ability of these models. The "rfcv" code of RF was applied to rank and plot importance of the variables.[31] Then, the partial dependence plots of the independent variables were plotted by partial Plot code of RF. The partial plot illustrated the relationship between a specific independent variable and all dependent variables by controlling other dependent variables. [31,32] The partial dependence plot indicated the percentiles of the independent variable (x-axis) as "rug" for continuous variable. We observed the relationship between the 10^{th} and 90^{th} percentiles of the independent variable and the correspondent dependent variable.

Results

Concentrations of indoor air pollutants

Table 1 showed the concentrations of indoor air pollutants and indoor meteorological parameters (temperature and RH). Although the average concentrations of IAPs were low, the mean of $PM_{2.5}$ concentrations of shop C exceeded the 24-hour average $PM_{2.5}$ of Taiwan Environmental Protection Administration (EPA) IAQ standards 35 μg/m^3 and the World Health Organization (WHO) guidelines 15 μg/m^3. [33,34] The 15-minute averages of IAPs were compared to the Taiwan IAQ standards and WHO guidelines. The $PM_{2.5}$ 15-minuate averages of coffee shop B, C and D had 0.0%, 25.3% and 1.2% exceeded the Taiwan IAQ standard 35 μg/m^3 and 13.6%, 29.5% and 12.1% exceeded the WHO IAQ guidelines 15 μg/m^3 respectively. The portions of 5%, 2.3% and 21.1% of the PM_{10} 15-minute averages of coffee shop A, B and C exceeded 75 μg/m^3 (Taiwan IAQ standard of 24-h average PM_{10}) accordingly. The PM_{10} 15-minute averages of all investigated coffee shops exceeded the WHO guidelines of 24-h average PM_{10} 45 μg/m^3 with the portions of A 5.0%, B 4.5%, C 24.2%, and D 1.2%. The CO_2 15-minute averages of coffee shop A, C, and D showed 5%, 21.1%, and 36.1% exceeded 1,000 ppm (Taiwan IAQ standard of 8-h average CO_2). For 15-minute averages of TVOCs, coffee C and D shop exceeded 0.56 ppm (Taiwan IAQ standard of 1-hour average TVOCs) in the portions of 20% and 1.2% respectively. For comfort parameters, all participated coffee shops controlled the temperature below 26°C, but the 15-minute averages of RH for all shops exceeded 70% with the portions of 100.0%, 34.1%, 23.2% and 55.4% for shop A, B, C, and D accordingly. The RH was set at 70% during occupancy by most Asian countries IAQ standards and guidelines.[35]

Characteristics of environmental factors

Characteristics of indoor environmental factors of four coffee shops were listed in Table 2. The indoor activities were counted every 15 minutes during the on-site surveillance period and summarized as the percent of the total activity counts. The most common indoor activity is cooking with a frequency of 30.0% for Shop A and 21.7% for Shop D. Roasting beans and cleaning are the second frequent indoor activities in these cafes. The major activity was roasting beans for shop B (25.0%). In shop C, 31.6% of the total indoor activities was indoor tobacco smoking. This is the only café allowed indoor tobacco smoking. Indeed, it is prohibited in Taiwan.

The mean occupant density of shop D was 28 occupants/100 m^2 (range: 9 – 66 occupants/100 m^2), higher than the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) recommendation 20 occupants/100 m^2 for coffee stations.[36] During the monitoring periods, 15.9% and 61.4% of the 15-minute interval exceeded the recommended value (20 occupants/100 m^2) for coffee shop B and D respectively.

Table 1. Descriptive statistics of the concentrations* of indoor air pollutants and meteorological data* in the studied coffee shops

	Shop A (N=20)		Shop B (N=44)		Shop C (N=95)		Shop D (N=83)	
	range	mean±SD	range	mean±SD	range	mean±SD	range	mean±SD
PM_{10} (mg/m^3)	2.3-140.9	10.7±30.6	6.6-80.1	23.2±14.1	7.5-161.3	44.2±44.8	4.0-57.2	13.2±8.0
$PM_{2.5\text{-}10}$ (mg/m^3)	0.3-138.3	7.6±30.7	3.9-72.0	13.7±12.3	2.0-18.1	6.5±3.2	1.6-13.0	5.1±2.1
$PM_{2.5}$ (mg/m^3)	1.7-5.4	3.1±1.0	2.7-17.9	9.5±4.9	3.9-152.4	35.8±41.9	1.9-48.7	8.1±6.8
Total VOCs (ppm)	NA	NA	0.1-0.3	0.1±0.0	0.0-1.5	0.3±0.4	0.1-0.8	0.2±0.1
p-PAHs (ng/m^3)	NA	NA	NA	NA	2.6-193.7	20.5±29.3	6.6-24.5	12.4±3.7
CO (ppm)	ND	ND	ND-4.9	0.9±1.4	0.0-4.0	0.5±1.0	ND	ND
CO_2 (ppm)	679.6-1626.8	785.8±201.5	475.0-838.8	590.1±93.4	442.7-1774.9	771.6±309.2	543.3-1607.4	895.4±245.2
Temperature (°C)	21.8-24.6	23.6±0.8	21.5-24.1	23.3±0.6	19.9-25.4	22.7±1.3	20.0-25.8	22.5±1.7
RH (%)	70.3-80.8	75.7±3.3	53.8-80.4	67.5±8.2	53.0-73.6	65.6±5.7	64.5-81.2	71.7±4.6

* Concentrations were 15-minute averages.

Abbreviations: PM_{10}, particulate matter with an aerodynamic diameter less than 10 μm; $PM_{2.5\text{-}10}$, particulate matter with an aerodynamic diameter between 2.5 and 10 μm; $PM_{2.5}$, particulate matter with an aerodynamic diameter less than 2.5 μm; Total VOCs, total volatile organic compounds; p-PAHs, particulate polycyclic aromatic hydrocarbons; CO, carbon monoxide (ppm); CO_2, carbon dioxide (ppm);RH, relative humidity (%); NA, not available due to instrumental malfunction; ND, not detected.

Table 2. Summary of the indoor environmental characteristics in the investigated coffee shops

	Shop A (N=20)		Shop B (N=44)		Shop C (N=95)		Shop D (N=83)	
	n	%	n	%	n	%	n	%
Indoor activities†								
Cooking	6	30.0%	1	2.3%	10	10.5%	18	21.7%
Roasting beans	0	0.0%	11	25.0%	2	2.1%	9	10.8%
Cleaning	0	0.0%	3	6.8%	0	0.0%	7	8.4%
Smoking	0	0.0%	0	0.0%	30	31.6%	0	0.0%
Others	0	0.0%	3	6.8%	0	0.0%	1	1.2%
Ventilation status†								
AC on/Window or door open	0	0.0%	44	100.0%	0	0.0%	0	0.0%
AC on/Window or door closed	20	100.0%	0	0.0%	86	90.5%	83	100.0%
AC off/Window or door closed	0	0.0%	0	0.0%	9	9.5%	0	0.0%
Locating on the main traffic street	Yes		No		No		No	
Occupant density (person/100m^2)	8-13‡	11±1§	7-33‡	17±6§	3-20‡	10±3§	9-66‡	28±3§

† Recorded by 15-minute interval. ‡ range. §mean±SD.

Abbreviations: AC, air conditioner

Ranking the environmental factors by RF models

The occupant density, CO_2, temperature, indoor activities, and RH were identified as the top 5 indicators by the variable importance plots resulting from the RF models of each indoor air pollutant. For PM_{10} and $PM_{2.5}$, occupant density and human activities were the top two indicators (Figure 2(A), (F)). Excluding the outliers of the occupant density, the concentrations of PM_{10} and $PM_{2.5}$ slightly increased as occupant density increased that was found from the partial dependence plot in Figure 2(B) and (G). The highest partial average level of PM_{10} and $PM_{2.5}$ occurred during the indoor activity “tobacco smoking” and the difference from other activities were 14 μg/m^3 and 16 μg/m^3, respectively (Figure 2(C), (H)). The CO_2 concentration was the third important indicator for indoor concentration of PM_{10} and $PM_{2.5}$. When the concentration of CO_2 increased from 900 ppm to about 1300 ppm, the concentrations of PM_{10} and $PM_{2.5}$ were proportional to the concentration of CO_2, and the increase concentration of PM_{10} and $PM_{2.5}$ were 6.5 μg/m^3 and 7.5 μg/m^3, respectively (Figure 2(D), (I)). Temperature was the fourth important indicator for PM_{10} and $PM_{2.5}$. The correlations between PM (PM_{10} and $PM_{2.5}$) and temperature are nonlinear. The lowest concentrations happened at 22°C. Then, the PM concentrations remained stable at 24 μg/m^3 for PM_{10} and 18 μg/m^3 for $PM_{2.5}$ as the temperature maintaining at 23°C - 26°C. T (Figure 2(E), (J)). The R^2 of RF model of PM_{10} and $PM_{2.5}$ were 0.71 and 0.80, respectively.

The R^2 of RF model for $PM_{2.5\text{-}10}$ was 0.21 and the top four important indicators were RH, temperature, CO_2 and occupant density for $PM_{2.5\text{-}10}$ (Figure 3(A)). Excluding the outlying RH (>80%), the concentrations of $PM_{2.5\text{-}10}$ slightly decreased as RH increased from around 58% to 78% (Figure 3(B)). The relationship between temperature and $PM_{2.5\text{-}10}$ was shown in “W” shape. The bottom was at about 23.2°C and 24°C. When the temperature was lower than 23.2°C, the relationship between temperature and $PM_{2.5-10}$ was complicated. On the other hand, when the temperature was higher than 24°C, the concentrations of $PM_{2.5-10}$ was positively proportional to temperature (Figure 3(C)). When the concentration of CO_2 was higher than about 1050 ppm, the concentration of $PM_{2.5-10}$ was proportional to the CO_2 concentration (Figure 3(D)). Excluding the outlying occupant density, the concentrations of $PM_{2.5-10}$ slightly increased as occupant density increased from about 11 to 33 person/100m^2 (Figure 3(E)).

The top four indicators shown in RF model for CO were CO_2, indoor activity, occupant density and temperature (Figure 3(F)) with the R^2 of 0.46. When indoor concentration of CO_2 was around 680 to 1450 ppm, the concentration of CO was proportional to the concentration of CO_2 (Figure 3(G)). The highest partial average level of CO was occurred with activity "roasting beans" at about 0.7 ppm, and the second highest partial average level of CO was occurred with indoor activity "tobacco smoking" at about 0.5 ppm (Figure 3(H)). The CO concentration had a negative relationship with occupant density when the density was lower than 0.3 person/m^2. When density was higher than 0.3 person/m^2, the concentration of CO was not affected by the density (Figure 3(I)). The relationship between the concentration of CO and temperature was shown in "U" shape. Basically, CO was relatively low when temperature was between 21.2 to 24.5°C (Figure 3(J)).

Occupant density, CO_2, temperature, and indoor activities were the top four important indicators for TVOCs and p-PAHs (Figure 4(A), (F)), and the R^2 of RF model were 0.77 and 0.55, respectively. The concentrations of TVOCs and p-PAHs were positively proportional to occupant density between 0.15 and 0.33 person/m^2 (Figure 4(B), (G)). When the concentration of CO_2 was around 450 to 1,200 ppm, the concentration of TVOCs was proportional to the concentration of CO_2 and the increase of TVOCs was about 0.2ppm (Figure 4(C)). When the concentration of CO_2 was around 500 to 1,150 ppm, the concentration of p-PAHs was proportional to the concentration of CO_2 and the increase of p-PAHs was about 13.1 ng/m^3 (Figure 4(H)). Excluding outlying temperature, the concentrations of TVOCs slightly increased as temperature increased from about 22 to 24.5°C (Figure 4(D)). When the temperature was greater than 23.5°C, the concentrations of p-PAHs slightly increased as the temperature increased (Figure 4(I)). The highest partial average level of TVOCs and p-PAHs was occurred with indoor activity "tobacco smoking" and the difference from other activities was 0.07 ppm and 6.5 ng/m^3, respectively (Figure 4(E) and (J)).

The top four indicators that affect the concentrations of CO_2 were occupant density, temperature, RH, and indoor activity as shown in Figure 5(A) and the R^2 of RF model was 0.53. Excluding outlying occupant density, the concentrations of CO_2 significantly increased from about 723 to 885 ppm when the occupant density increased from about 0.08 to 0.43 person/m^2 as showed in Figure 5(B). The relationship between temperature and CO_2 was complicated and was basically shown in "U" shape. The bottom was at about 23.5°C, when the temperature increased from 23.5°C to about 25.8°C, the concentrations of CO_2 increased from about 740 to 848 ppm (Figure 5(C)). Basically, CO_2 was relatively low when RH was lower than 62 %, but when the RH increased from about 62% to 80%, the concentrations of CO_2 increased from about 740 ppm to 850 ppm (Figure 5(D)). The highest partial average level of CO_2 was occurred with indoor activity "tobacco smoking" and the difference from other activities was about 100 ppm (Figure 5(E)).

Discussion

Compliance with regulations

In Taiwan, on average, each person drank 104 cups of coffee in 2020 according to annual coffee bean import statistics from the International coffee organization (ICO). Taiwan's coffee shop density is the highest globally, and coffee chains

Starbucks and Louisa have both exceeded 500 stores.[37,38] People might have more chances to stay at coffee shops in urban area. This environment is not only a living and leisure space, but also a workplace. However, the Taiwan IAQ regulation does not regulate these shops up till now. Customers won't carry any IAQ instruments usually. The aim of this study was to identify the observable factors that can be the significant indicators of IAQ concurrently. The time of spending in a coffee shop is ranged from 4-6 hours for working or studying to 15 minutes for grabbing a cup of coffee for to-go. So, we decided to inspect the 15-minute average concentrations to reflect the exposure of group with short staying periods and also an assurance for all groups. These time-weight averages were applied to check the compliance with the IAQ standards/guidelines. Among all investigated café, 11.2%, 18.2%, and 21.1% of the 15-minute averages of PM_{10}, $PM_{2.5}$, and CO_2 accordingly did not meet the WHO guidelines and 8.3% of TVOCs exceeded Taiwan IAQ standard. However, these comparisons could only be a reference, as IAQ standard regulated the time-weighted concentration of 1 hour, 8 hours or 24 hours. Our results found the exceedances of short term (15-minutes) concentration remind the long-term time weighted average might underestimate the exposure of customers and employee at certain periods. The IAQ of café should be addressed as a workplace and/or a public environment to compliance with the regulations and to assure the healthy environments of people in these indoor spaces.

Cooking and bakery are the main sources of indoors' PAHs. In coffee shop C and D, the p-PAHs were quantified in the range of 2.6 and 193.7 ng/m^3 resulting from preparing the light meals. Abdullahi et al.[39] reviewed the cooking emission studies and found the PAHs concentrations were variated with cooking styles, ranged from 0.2 - 1,590 ng/m^3. Ielpo et al.[40] reported the mean concentration was 7.4 ng/m^3 (range: 5 - 10 ng/m^3) from bakery. All these were similar with our findings in p-PAHs. The difference might due to different cooking patterns and gaseous PAHs.

Significance of ranking

This study is the first one to rank the importance of indoor environmental factors and examine the relationships between factors and air pollutants in independent coffee shops as novel workplace. Results of this study showed that occupant density, indoor activities, CO_2 concentration, and temperature can remind the concentration changes of PM_{10}, $PM_{2.5}$, TVOCs, p-PAHs, and CO. Limited researches have evaluated the determinants of IAP in coffee shops. In this study, occupant density was found being the most important determinant of the indoor concentrations of PM_{10}, $PM_{2.5}$, TVOCs, p-PAHs, and CO_2. On the other hand, the occupant density was the third and fourth important determinant of the concentrations of CO and $PM_{2.5\text{-}10}$. We found slightly positive dose-response relationship between occupant density and air pollutants, particularly in the increasing of CO_2 concentration. This was also found in other peer studies.[23,41] Previous study showed that the lower occupant density, the lower the PM concentration in preschools classrooms [16], but this study didn't find dose-response effects between occupant density and concentrations of PM_{10}, $PM_{2.5}$, or $PM_{2.5\text{-}10}$. The higher occupant density, the higher the concentrations of specific VOC.[15] We found the occupant density was the most important indicator for TVOCs. Coffee shops serve different functions for metropolitan people, such as social gatherings, studying, working and business meetings. Most of the shop design was the open-kitchen style, no significant segregation between dining and cooking areas. According the ANSI/ASHRAE standard 62.1, these investigated independent cafés were fitted in the occupancy category, cafeteria/fast-food dining and kitchen (cooking) both. While, the default occupant density with recommended ventilations of these two categories are different [35]. So more empirical researches are recommended to assure the appropriate occupant density of these shops to compliance with ASHRAE recommended air class.

Indoor activities were the second important determinant of the concentrations of PM_{10}, $PM_{2.5}$, and CO. For TVOCs, p-PAHs, and CO_2, the indoor activities were the fourth important determinant. Besides, this study further pointed out the major human activity associated IAQ was tobacco smoking. Previous studies showed that concentrations of indoor air pollutants, such as VOCs,[1,9] $PM_{2.5}$,[1,11] p-PAHs,[1] and CO,[42] were associated with tobacco use. Only shop C allowed indoor smoking among the 4 investigated shops. The partial dependence effect of smoking in increasing the pollutant concentrations were stronger than other indoor activities by RF modeling with other environmental factors being controlled. Thus, it is important to ban the indoor smoking to reduce IAPs in coffee shops. Moreover, previous study pointed out the relationship between roasting coffee beans and concentration of CO,[7] this study further showed that roasting beans was a more important human activity associated with indoor CO concentration than tobacco smoking. As due to a limited number of studied coffee shops, this result need further confirmed.

CO_2 is a global indicator of IAQ and a rough indicator of the effectiveness of ventilation. High CO_2 level implies the possibilities of indoor IAPs accumulations,[20,23-25] our results showed that the concentrations of PM, TVOCs, p-PAHs, CO were proportional to the concentration of CO_2. This confirmed the findings of other studies.[20,43] Besides, we found that the CO_2 were the top three important predictors of the concentrations of multi-size PM, TVOCs, p-PAHs, and CO. The partial dependence plots indicated that the concentrations of IAPs increased proportional to the CO_2 concentration once it exceeded 1,000 ppm (Taiwan IAQ standard of 8-h average CO_2). So, we recommend the low-cost CO_2 monitor shall be installed in coffee shops to monitor CO_2 and alert the IAQ. CO_2 was affected by the occupants. However, RF modeling can overcome the collinearity of CO_2 and occupant to ranking the importance of these two determinants.

Temperature and RH are the important factors of thermal comfort. Previous studies reported that temperature and RH were related to the IAPs levels positively.[44-47] We found that the temperature was the top four important predictors of the concentrations of multi-size PM, TVOCs, p-PAHs, CO, and CO_2. Relationships between indoor pollutants concentrations and temperature were complicated, and the partial dependence plots of RF models in our study show that the IAP concentrations were consistently increased as the temperature increased within a certain range. The air conditioners were turned on during the business hours of these investigated shops. So, the temperature was kept constant with small variation (range 19.9 - 25.8°C). The results were complied with the IAQ standards/guidelines of major Asian countries. Therefore, if temperature was selected to be an indicator to alert the levels of IAQ, future research should include the indoor temperature with big variation and be caution of the non-linear relationship with the IAP concentrations.

RH is less important than temperature to be an indicator of IAQ, and our partial dependence plots results showed the complicated non-linear relationships between IAPs and RH. We observed the RH was the most important predictor of the $PM_{2.5-10}$ concentrations. $PM_{2.5-10}$ decreased slightly as the RH increasing. Oliveira et al. found the concentrations of CO_2 and $PM_{2.5}$ were affected by the RH inversely in the kindergartens.[43] Some studies reported the positive correlations between RH and IAPs (e.g.CO_2, HCHO).[20,23] The relationships were inconsistent among different researches. As we found, the relationship between RH and IAPs was complicated and nonlinear. RH is not an appropriate indicator for IAPs. Still, RH is relevant on the IAQ study, because it affects perceived IAQ comfort, synergistic effects may occur with air pollutants as well.[48] High RH provides the optimal condition for bacteria, fungi and viruses proliferation.[49] Indoor RH is not easy to control in Taiwan's subtropical climate. Our monitoring data of coffee shops resulted in 42.6% of the RH over 70%. Currently, the IAQ standards or guidelines of Asian and European countries for temperature and humidity criteria are different.[35] The humidity (RH) of coffee shops shall be maintained within a comfort range according to the climate conditions.

The factors, weekday, ventilation status, shop's pattern, and locating on the main street were less important in predicting the IAPs as they were not listed by the rfcv module in RF model analysis. However, it did not mean that they had no effects of IAQ. The possible reasons were due to the small variations of the four investigated shops. For example, only three situations of the ventilation status were observed (table 2), so the significance of these factors can't be identified by the statistics analysis.

Limitations and strength

This study was limited by the scale variations of the café, but we proved that RF was able to deal the limited numbers of data and build models to examine relatively complex and non-linear relationship between IAPs and determinant variables. Meanwhile, the RF model calculation considered all environmental factors simultaneously and provided insight in potential causal relationship between air pollutants and environmental factors, particularly the temperature and RH. The total picture and interrelationships between different enironmental parameters were illustrated. On the other hand, our RF model identified the most important determinat of CO_2 concentraiotn was occupant density (Figure 5(A)) and the most important indoor activity in affecting PM_{10} and $PM_{2.5}$ levels was occurred with "smoking"(Figure 2(C), (H)). A Rome's study reported the tobacco smoking increased the indoor $PM_{2.5}$ concentrations by two to three times of the non-smoking sites.[1] This was consistent with the present study and proved the reliablility of the RF analysis results.

Recommendations

Although the strength of this study is that it used RF models to examine indoor environmental indicators for reminding levels of IAQ in novel independent coffee shops, there still were some limitations. First, some studies showed that ventilation could influence IAQ, but this study didn't measure outdoor air change rate in these participated shops. However, previous studies showed indoor ventilation was non-significantly associated with concentration of CO and PM.[50,51] Moreover, previous findings showed natural ventilation, such as window opening, and outdoor air pollutants would influence IAQ.[17,50] Therefore, it is suggested that future researchers could consider outdoor air pollution when investigating IAQ if there are natural ventilation in coffee shops. In addition, this study did not include chain coffee shops. In Taiwan, most of the chain coffee shops are located at the commercial buildings with central air conditioning systems which are different from the independent café with individual air-conditioners of this study. More coffee shops with different air-conditioning designs will be included in our future studies to validate and extend the applicability of the results of this study. Second, only four independent coffee shops participated in this study. Hence the findings cannot be inferred to other types of coffee shops. The importance ranking of determinants that affect IAPs may be changed due to the large variations of environmental factors if various types of coffee shops are included in future study. Last, our sampling time did not include the summer season, future study should evaluate potential seasonal variations and their influence.

Conclusions

The application of Random Forest models in assessing and ranking the environmental factors that affect the IAPs of independent coffee shops was demonstrated. Meanwhile, the RF was able to illustrate the complicated non-linear relationship between IAPs and determinant variables. Customers and staffs in the independent coffee shops can be reminded the change of indoor concentrations of PM, CO, CO_2, TVOCs, and p-PAHs by observing the occupant density and human activities, such as tobacco smoking and roasting beans. Monitoring CO_2 and maintaining the room temperature at appropriate range could also be the surrogate parameters to assure the acceptable IAQ.

Declarations

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article and its supplementary information files.

ACKNOWLEDGEMENT

This work was supported by Wang Jhan-Yang Charitable Trust Fund (Grant No. WJY2020-AP-01)

AUTHOR CONTRIBUTIONS

Yu-Wen Lin: Conceptualization (equal); funding acquisition (lead); writing – original draft (equal); review and editing (lead). His-Chen Liu: Writing – original draft (equal); review and editing (supporting); Tzu-Ying Lee: Formal analysis (supporting); data curation (equal); writing – review and editing (supporting); Hsiao-Yun Huang: Formal analysis (equal); writing – review and editing (supporting); Tzu-An Hsu: data curation (equal); methodology (supporting); writing – review and editing (supporting); Li-Te Chang: writing – review and editing (supporting); Chin-Tseng Tang: Conceptualization (equal); methodology (lead); form al analysis (equal); writing – original draft (equal); writing – review and editing (supporting)

References

1. Romagnoli, P. *et al.* Indoor air quality at life and work environments in Rome, Italy. Environ. Sci. Pollut. Res. Int. **23**, 3503–3516; http://doi.org/10.1007/s11356-015-5558-4 (2016).
2. Sarigiannis, D. A. Indoor Air Quality Indicator in *Environmental Indicators* (eds. Armon, R. H. & Hänninen, O.) 827–841 (Dordrecht, 2014).
3. WHO. WHO guidelines for indoor air quality: selected pollutants. World Health Organisation https://www.who.int/publications/i/item/9789289002134 (2010).
4. Bruno, P., Caselli, M., de Gennaro, G., Iacobellis, S., & Tutino, M. Monitoring of volatile organic compounds in non-residential indoor environments. Indoor Air **18**, 250–256; http://doi.org/10.1111/j.1600-0668.2008.00528.x (2008).
5. Awair. Does Working From A Coffee Shop Help or Hurt our Health? *Awair* https://blog.getawair.com/does-working-from-a-coffee-shop-help-or-hurt-our-health (2018).
6. Global Coffee Report. Allegra report details rise of independent cafés in the UK. Global Coffee Report https://www.gcrmag.com/allegra-report-detai (2019).
7. LeBouf, R. F., & Aldridge, M. Carbon monoxide emission rates from roasted whole bean and ground coffee. J. Air Waste Manag. Assoc. **69**, 89–96; https://doi.org/10.1080/10962247.2018.1515125 (2019).
8. Kabir, E. & Kim, K. H. An investigation on hazardous and odorous pollutant emission during cooking activities. J. Hazard Mater **188**, 443–454; https://doi.org/10.1016/j.jhazmat.2011.01.113 (2011).
9. Cancelada, L. *et al.* Heated Tobacco Products: Volatile Emissions and Their Predicted Impact on Indoor Air Quality. Environ. Sci. Technol. **53**, 7866–7876; https://doi.org/10.1021/acs.est.9b02544 (2019).
10. Jia, C., & Batterman, S. A critical review of naphthalene sources and exposures relevant to indoor and outdoor air. Int. J. Environ. Res. Public Health **7**, 2903–2939; https://doi.org/10.3390/ijerph7072903 (2010).
11. Kim, H., Lee, K., An, J., & Won, S. Determination of secondhand smoke leakage from the smoking room of an Internet café. J. Air Waste Manag. Assoc. **67**, 1061–1065; https://doi.org/10.1080/10962247.2017.1338205 (2017).
12. Adhikari, S., Mahapatra, P. S., Pokheral, C. P., & Puppala, S. P. Cookstove smoke impact on ambient air quality and probable consequences for human health in rural locations of Southern Nepal. Int. J. Environ. Res. Public Health **17**, 550; https://doi.org/10.3390/ijerph17020550 (2020).

13. Mitova, M. I. *et al.* Human chemical signature: Investigation on the influence of human presence and selected activities on concentrations of airborne constituents. Environ. Pollut. **257**, 113518; https://doi.org/10.1016/j.envpol.2019.113518 (2020).
14. Singh, V., Sokhi, R. S., & Kukkonen, J. An approach to predict population exposure to ambient air $PM_{2.5}$ concentrations and its dependence on population activity for the megacity London. Environ. Pollut. **257**, 113623; https://doi.org/10.1016/j.envpol.2019.113518 (2020).
15. Cheng, M. *et al.* Factors controlling volatile organic compounds in dwellings in Melbourne, Australia. Indoor Air **26**, 219–230; https://doi.org/10.1111/ina.12201 (2016).
16. Yu, K. P., Lee, Y. C., Chen, Y. C., Gong, J. Y., & Tsai, M. H. Evaluation of PM_1, $PM_{2.5}$, and PM_{10} exposure and the resultant health risk of preschool children and their caregivers. J. Environ. Sci. Health A. Tox Hazard Subst. Environ. Eng. **54**, 961–971; https://doi.org/10.1080/10934529.2019.1623598 (2019).
17. Tong, X. *et al*. Prediction model for air particulate matter levels in the households of elderly individuals in Hong Kong. Sci. Total. Environ. **717**, 135323; https://doi.org/10.1016/j.scitotenv.2019.135323 (2020).
18. Kim, J. J., Hann, T., & Lee, S. J. Effect of flow and humidity on indoor deposition of particulate matter. Environ. Pollut. **255**, 113263; https://doi.org/10.1016/j.envpol.2019.113263 (2019).
19. Sharma, D., & Jain, S. Carcinogenic risk from exposure to $PM_{2.5}$ bound polycyclic aromatic hydrocarbons in rural settings. Ecotoxicol. Environ. Saf. **190**, 110135; https://doi.org/10.1016/j.ecoenv.2019.110135 (2020).
20. Hwang, S. H., Roh, J., & Park, W. M. Evaluation of PM_{10}, CO_2, airborne bacteria, TVOCs, and formaldehyde in facilities for susceptible populations in South Korea. Environ. Pollut. **242**, 700–708; https://doi.org/10.1016/j.envpol.2018.07.013 (2018).
21. Jaber, A. R., Dejan, D., & Marcella, U. The Effect of Indoor Temperature and CO_2 Levels on Cognitive Performance of Adult Females in a University Building in Saudi Arabia. Energy Procedia **122**, 451–456; https://doi.org/10.1016/j.egypro.2017.07.378 (2017).
22. Branco, P. T. B. S., Alvim-Ferraz, M. C. M., Martins, F. G., & Sousa, S. I. V. Children's exposure to indoor air in urban nurseries-part I: CO_2 and comfort assessment. Environ. Res. **140**, 1–9; https://doi.org/10.1016/j.envres.2015.03.007 (2015).
23. St-Jean, M. *et al.* Indoor air quality in Montreal area day-care centres, Canada. Environ. Res. **118**, 1–7; https://doi.org/10.1016/j.envres.2012.07.001 (2012).
24. Griffiths, M., & Eftekhari, M. Control of CO_2 in a naturally ventilated classroom. Energy and Buildings **40**, 556–560; https://doi.org/10.1016/j.enbuild.2007.04.013 (2008).
25. Jones, A. P. Indoor air quality and health. Atmos. Environ. **33**, 4535–4564; https://doi.org/10.1016/S1352-2310(99)00272-1 (1999).
26. Ryan, P. H., Brokamp, C., Fan, Z. H., & Rao, M. B. Analysis of Personal and Home Characteristics Associated with the Elemental Composition of $PM_{2.5}$ in Indoor, Outdoor, and Personal Air in the RIOPA Study. Res. Rep. Health Eff. Inst. **185**, 3–40; http://www.ncbi.nlm.nih.gov/pubmed/26934775 (2015).
27. Wei, W., Sivanantham, S., Malingre, L., Ramalho, O., & Mandin, C. Predicting the rate constants of semivolatile organic compounds with hydroxyl radicals and ozone in indoor air. Environ. Pollut. **266**, 115050; https://doi.org/10.1016/j.envpol.2020.115050 (2020).
28. Yuchi, W. *et al.* Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. Environ. Pollut. **245**, 746–753; https://doi.org/10.1016/j.envpol.2018.11.034 (2019).
29. Breiman L. Random Forests. Machine Learning **45**, 5–32; https://doi.org/10.1023/A:1010933404324 (2001).

30. Horning, N. Random Forests: An algorithm for image classification and generation of continuous fields data sets. Geoinformatics International https://gisws.media.osaka-cu.ac.jp/gisideas10/viewabstract.php?id=342 (2010).
31. Liaw, A., & Wiener, M. Classification and regression by Random Forest. R news **2**, 18–22; https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf (2002).
32. Hastie, T., Tibshirani, R., & Friedman, J. Random Forest in *The Elements of Statistical Learning* (eds. Hastie, T., Tibshirani, R., & Friedman, J.) 587–604 (New York, 2013).
33. Taiwan Air Quality Monitoring Network. Air Quality Standards. *Environmental Protection Administration* https://airtw.epa.gov.tw/ENG/Information/Standard/Rules.aspx (2012).
34. World Health Organization. Ambient (outdoor) air pollution. *World Health Organization* https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (2021).
35. Abdul–Wahab, S. A., En, S. C. F., Elkamel, A., Ahmadi, L., & Yetilmezsoy, K. A review of standards and guidelines set by international bodies for the parameters of indoor air quality. Atmospheric Pollution Research **6**, 751–767; https://doi.org/10.5094/APR.2015.084 (2015).
36. ASHARE. Interpretations for standard 62.1–2019. *ASHARE* https://www.ashrae.org/technical-resources/standards-and-guidelines/standards-interpretations/interpretations-for-standard-62-1-2019 (2020)
37. International coffee organization. Taiwan: coffee consumption in 2020 reached 2.4 billion cups, or 104 cups per capita. *International coffee organization* https://www.comunicaffe.com/taiwan-coffee-consumption-reach-2-4-billion-cups-of-coffee-or-104-cups-per-capita/ (2021).
38. International coffee organization. World coffee consumption. *International coffee organization* https://www.ico.org/prices/new-consumption-table.pdf (2021).
39. Abdullahi, K. L., Delgado-Saborit, J. M., & Harrison, R. M. Emissions and indoor concentrations of particulate matter and its specific chemical components from cooking: a review. Atmos. Environ. **71**, 260–294; https://doi.org/10.1016/j.atmosenv.2013.01.061. (2013).
40. Ielpo, P. *et al*. Polycyclic aromatic hydrocarbons in a bakery indoor air: trends, dynamics, and dispersion. Environ. Sci. Pollut. Res. **25**, 28760–28771; https://doi.org/10.1007/s11356-018-1513-5 (2018).
41. Branco, P. T. B. S., Alvim-Ferraz, M. C. M., Martins, F. G., & Sousa, S. I. V. Quantifying indoor air quality determinants in urban and rural nursery and primary schools. Environ. Res. **176**, 108534; https://doi.org/10.1016/j.envres.2019.108534 (2019).
42. Bahcebasi, T., Kandis, H., Baltaci, D., & Kara, I. H. Factors affecting exhaled carbon monoxide levels in coffeehouses in the Western Black Sea region of Turkey. Toxicol. Ind. Health **27**, 195–204; https://doi.org/10.1177/0748233710383888 (2011).
43. Oliveira, M., Slezakova, K., Delerue-Matos, C., Perira, M. C., & Morais, S. Indoor air quality in preschools (3- to 5-year-old children) in the Northeast of Portugal during spring-summer season: pollutants and comfort parameters. J. Toxicol. Environ. Health A **80**, 740–755; https://doi.org/10.1080/15287394.2017.1286932 (2017).
44. Krupińska, B., Van Grieken, R., & De Wael, K. Air quality monitoring in a museum for preventive conservation: Results of a three-year study in the Plantin-Moretus Museum in Antwerp, Belgium. Microchemical Journal **110**, 350–360; https://doi.org/10.1016/j.microc.2013.05.006 (2013).
45. Mirmohammadi, S. Indoor air quality assessment with emphasis on flour dust: A crosssectional study of a random sample from Iranian bakeries workers. Iranica Journal of Energy and Environment **4**, 150–154; https://doi.org/10.5829/idosi.ijee.2013.04.02.12 (2013).
46. Fonseca, J., Slezakova, K., Morais, S., & Pereira, M. C. Assessment of ultrafine particles in Portuguese preschools: levels and exposure doses. Indoor Air **24**, 618–628; https://doi.org/10.1111/ina.12114 (2014).

47. Canha, N. *et al.* Assessment of ventilation and indoor air pollutants in nursery and elementary schools in France. Indoor Air **26**, 350–365; https://doi.org/10.1111/ina.12222 (2016).
48. Wolkoff, P., & Nielsen, G. D. Non-cancer effects of formaldehyde and relevance for setting an indoor air guideline. Environ. Int. **36**, 788–799; https://doi.org/10.1016/j.envint.2010.05.012 (2010).
49. Alves, C. *et al.* Air quality in a school with dampness and mould problems. Air Qual. Atmos. Health **9**, 107–115; https://doi.org/10.1007/s11869-015-0319-6 (2015).
50. Rostami, R. *et al.* The effects of ventilation and building characteristics on indoor air quality in waterpipe cafés. J. Expo. Sci. Environ. Epidemiol. **30**, 805–813; https://doi.org/10.1038/s41370-020-0240-4 (2020).
51. Seidenberg, A. B., Orlan, E. N., Travers, M. J., & Sutfin, E. L. Air quality and presence of air ventilation systems inside waterpipe cafés in North Carolina. Tob. Control **28**, 356–358; http://dx.doi.org/10.1136/tobaccocontrol-2018-054361 (2019).

Figures

Figure 1

Floor plans and measured spots of four coffee shops

Figure 2

Variable importance rank and partial dependence plot of PM_{10} and $PM_{2.5}$ from Random Forest models.

(PM_{10}: particulate matter with an aerodynamic diameter £ 10 μm (μg/m^3); $PM_{2.5}$: particulate matter with an aerodynamic diameter £ 2.5 μm (μg/m^3) ACTIVE: Indoor activities; CO_2: carbon dioxide (ppm); DENSITY: Occupant density (person/m^2); RH: relative humidity (%); SHOP: coffee shop A, B, C, D; STREET: locating on the main traffic street; TEMP: temperature(°C); VENT: ventilation status)

Figure 3

Variable importance rank and partial dependence plot of $PM_{2.5\text{-}10}$ and carbon monoxide (CO) from Random Forest models.

($PM_{2.5\text{-}10}$, particulate matter with an aerodynamic diameter between 2.5 and < 10 μm (μg/m^3); ACTIVE, Indoor activities; CO_2, carbon dioxide(ppm); DENSITY, Occupant density (person/m^2); RH, relative humidity (%); SHOP, coffee shop A, B, C, D; STREET, locating on the main traffic street; TEMP, temperature(°C); VENT, ventilation status)

Figure 4

Variable importance rank and partial dependence plot of total volatile organic compounds (TVOCs) and particle-bound polycyclic aromatic hydrocarbons (p-PAHs) from Random Forest models.

(ACTIVE, Indoor activities; CO_2, carbon dioxide (ppm); DENSITY, Occupant density (person/m^2); RH, relative humidity (%); SHOP, coffee shop A, B, C, D; STREET, locating on the main traffic street; TEMP, temperature(°C); VENT, ventilation status)

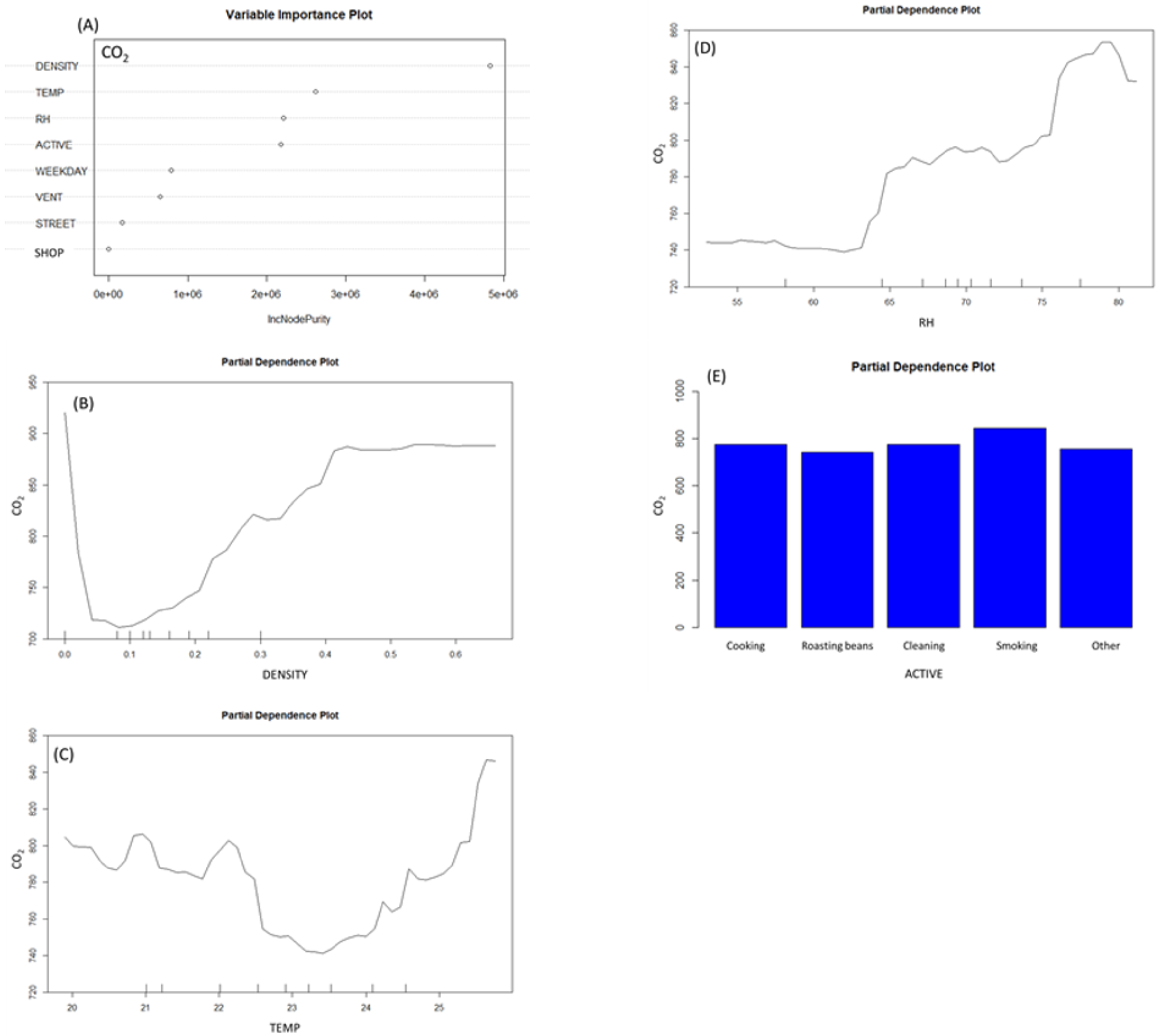

Figure 5

Variable importance rank and partial dependence plot of carbon dioxide (CO_2) from Random Forest models.

(ACTIVE, Indoor activities; CO_2, carbon dioxide(ppm); DENSITY, Occupant density (person/m^2); RH, relative humidity (%); SHOP, coffee shop A, B, C, D; STREET, locating on the main traffic street; TEMP, temperature(°C); VENT, ventilation status)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Rawdatacoding.csv
- Rawdataset.csv