Application of multicriteria decision making methods to air quality in the microenvironments of residential houses in Brisbane, Australia

Godwin A. Ayoko*[a], Lidia Morawska [a], Serge Kokot [b] and Dale Gilbert [c]

[a] International Laboratory for Air Quality and Health, School of Physical and Chemical Sciences, Queensland University of Technology, Brisbane, Australia. Email: g.ayoko @ qut. edu. au; Fax: +61 7 3864 1804

[b] School of Physical and Chemical Sciences, Queensland University of Technology, Brisbane, Australia.

[c]Built Environment Research Unit, Queensland Department of Public Works, 80 George Street, Brisbane QLD 4001, Australia.

* Author to whom correspondence should be addressed at the above address:

Abstract

This paper reports the first application of the multicriteria decision making methods, PROMETHEE and GAIA, to indoor and outdoor air quality data. Fourteen residential houses in a suburb of Brisbane, Australia were investigated for 21 air quality-influencing criteria, which included the characteristics of the houses as well as the concentrations of volatile organic compounds, fungi, bacteria, submicrometre and supermicrometre particles in their indoor and outdoor air samples. Ranking information necessary to select one house in preference to all others and to assess the parameters influencing the differentiation of the houses were found with the aid of PROMETHEE and GAIA. There was no correlation between the rank order of each house and the health complaints of its occupants. Patterns in GAIA plots show that indoor air quality in these houses is strongly dependent on the characteristics of the houses (construction material, distance of the house from a major road and the presence of an in-built garage). Marked similarities were observed in the patterns obtained when GAIA and factor analysis were applied to the data. This underscores the potential of PROMETHEE and GAIA to provide information that can assist source apportionment and elucidation of effective remedial measures for indoor air pollution.

Keywords: Residential environments, air quality, multicriteria decision making methods, ranking analysis, pattern recognition

Introduction

Over 80% of people's time is spent in different indoor environments, which include residential houses, workplaces, schools, restaurants, entertainment centres and the interiors of private and commercial vehicles (eg, 1, 2). The quality of air in such environments depends on a multiplicity of variables, which include: the concentration levels and characteristics of airborne pollutants generated indoors or penetrating from outside; thermal and moisture conditions, air movement, noise level and the types of indoor activities and building materials. The relative importance of these variables depends, in turn, on the nature, location and other characteristics of the indoor and outdoor environments, as well as on the preferences and susceptibilities of individuals.

Because of the multivariate nature of the factors influencing indoor air quality, the integrated assessment and comparison of different environments pose significant challenges to indoor air researchers. It is not enough to list or to tabulate the concentration levels of different pollutants or the characteristics of the variables, but a method that would allow for a quantitative assessment and an objective comparison needs to be applied. Thus, a multivariate ranking methodology is required, which is based on some recognised links between pollutants and human health, and capable of providing an overall assessment of the quality of indoor environment in response to a set of variables. Such a method could also be applied in reverse; namely, it could facilitate the identification of the relative significance of individual pollutants or factors, when a database containing human responses to such factors is available from an exposure/ health study.

Although factor analysis/principal component analysis (PCA) (3, 4, 5) and multiple regression/correlation analysis (6) have been successfully applied to airborne organic samples in order to examine the co-variance within a data set, identify patterns, elucidate associations and apportion sources, multi-criteria decision making procedures, in general and the PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluations) and GAIA (Graphical Analysis for Interactive Assistance), in particular, (7, 8, 9, 10, 11) have not been applied to indoor air quality data.

PROMETHEE, an object ranking method, and GAIA, a visual data display method, are examples of methods which assist in the making of decisions for multivariate problems; hence the name multi-criteria decision making methods (MCDM). Many such methods have been discussed and developed in detail for decades in operations research field but have been introduced into chemometrics comparatively recently. A case in point is PROMETHEE and GAIA, which was first discussed by Brans et al in the 1980's in operations research (12), but introduced to chemometrics in a detailed paper by Keller et al (7) in 1991; this paper includes a step – by-step worked example. The attributes of PROMETHEE (and sometimes GAIA) have been compared with other MCDM procedures. For example, Keller et al (7) discussed PROMETHEE, PARETO Optimality as well as ELECTRE, and more recently, SMART (Simple Multi-attribute Rating Technique), ELECTRE III and SMARTER, which is a SMART related centroid approach, were all considered in conjunction with PROMETHEE (13, 14). In general, SMART, which is derived from MAUT (Multiple Attribute Utility Theory), is regarded to be very similar to PROMETHEE but ELECTRE offers additional options for thresholds when defining criteria models. On the other hand, Massart et al (15) suggested that PROMETHEE was a more refined method than ELECTRE in that the former quantifies the degree of preference of an object compared with another for each criterion. In further work, Lerche et al (16) have compared the partial order Hasse Diagram technique (HDT) with several Multi-criteria Analysis (MCA) methods, which included PROMETHEE. They were particularly concerned with studying the relative subjectivity and transparency of the techniques. In their opinion, HDT required minimum external input and it was superior to the MCA methods. However, it was noted that PROMETHEE rated closely behind the HDT and ahead of some of its potential alternatives such as NAIADE (16), ORESTE (16) and the less appropriate method, Analytical Hierarchy Process (AHP)(16).

The application of GAIA, on the other hand, is a particularly useful aid since it provides a data display in the form of a Principal Component Analysis biplot, and in addition, shows a decision making axis, π, which indicates the quality of the decision. Furthermore, PROMETHEE and GAIA (i) preserve as much information as possible (17), (ii) avoid trade-offs (17,18), (iii) permit sensitivity analysis (7, 18), (iv) require no interaction by the user (19), and (v) are readily available in the form of a user friendly software package (17, 18). In the papers noted above, not one except Keller et al. (7) addressed the usefulness of GAIA as a display method for comparing objects, variables, and objects with variables. However, a comprehensive discussion of this method together with an exhaustive listing of the rules for the interpretation of the plots has been reported by Espinasse (20).

In the context of environmental problems, Salminen et al (13) considered PROMETHEE, SMART and ELECTRE III MCDM methods to be particularly suitable. Thus, some examples of the application of PROMETHEE and GAIA in this field include: Martin et al (21), who discussed the development of the Saint Charles River alluvial plain and used the PROMETHEE and GAIA methodology to reach rational decisions based on scientific data and political considerations; Le Teno and Mareschal (17) found that these methods were powerful tools for visualisation and interpretation of Life Cycle Assessment results; and, Ozelkan and Duckstein (22) studied water resource alternatives. More recently, the methods were employed in a number of papers presented at different symposia focussed on environmental issues concerned with air quality (23, 24), and in

2003, a multi-disciplinary investigation combining organo-metallic chemistry with toxicology applied these methods for the screening and ranking of anti-fungal agents (25). Thus, having regard to the fact that PROMETHEE and GAIA are MCDM methods that compare very favourably to some of the alternatives, and have been applied, albeit infrequently, in environmental and analytical fields, they are a reasonable choice for this work.

Consequently, this paper reports the application of the PROMETHEE and GAIA methods to the indoor air quality data obtained for fourteen houses located in a residential suburb of Brisbane, Australia. Studies specifically focused on the measurements and variations in the levels of individual pollutants found in the houses have been described elsewhere (26, 27, 28, 29). The present study reports the multivariate ranking of the houses based initially on 21 air quality-influencing variables. The primary aims of the paper are to: (i) provide ranking information necessary to select one house in preference to all others, on the basis of its air quality (ii) assess the parameters influencing the differentiation of the houses, (iii) attempt to relate the rank order of each house to the health complaints of its occupants (iv) explore the use of GAIA in the identification of the sources of the pollutants and (v) examine the role of building characteristics on quality of air in indoor microenvironments. Additionally, it was hoped that the outcomes of the work would enhance the development of control measures for indoor air pollution in a broader context.

Experimental Methods

Details of the overall study design, sampling site, characteristics of the houses, information obtained from the questionnaire administered to residents, the range of meteorological conditions during sampling, sampling protocol and chemical analyses have been described elsewhere (26, 27, 28, 29).

The pollutants measured in the overall study are: particles in terms of (i) number concentration and size distribution in the size range 0.015 to 0.697 µm (subsequently called the submicrometre range) and in the size range 0.54 – 19.81 µm (called the supermicrometre range), and (ii) $PM_{2.5}$ mass concentration of particles with aerodynamic diameter smaller than 2.5 µm; fungi and bacteria: number of total colony-forming units per cubic metre air (CFU/m^3); dust: dust mite allergen (Derp1), cat allergen (Feld1) and cockroach allergen (Blag1) (26, 27, 28) and VOCs (this study). Replicated measurements were made for the VOCs in representative houses but because of logistic reasons one measurement was made for each of the other parameters in each of the homes. For the particle number concentrations, the average of the particle number concentrations measured every minute over a 1h sampling period in the absence of indoor sources (ie when the residents were absent and researchers minimised their movements to avoid re-suspension of particles) (26,27) was used for the multivariate analysis. The concentrations of other pollutants as well as the following building characteristics: distance from a major road, wind direction, distance from a park, presence of in-built garage, age of the building and type of building (whether low set or high set) were also used in the PROMETHEE and GAIA procedure. However, the dust mite allergen (Derp1), cat allergen (Feld 1) and cockroach allergen (Blag 1) were not included in the multivariate analysis because these variable were not measured under the “normal ventilation” conditions described by Morawska *et al* (26,27) for all of the 14 houses. Data related to the characteristics of the houses as well as their indoor and outdoor air quality presented as supplementary information were used for the PROMETHEE and GAIA analyses.

Data processing; The data matrix was constructed in the PROMCALC and DECISION LAB software (30), which contain PROMETHEE and GAIA.

The houses were treated as objects in the matrix, and the building characteristics and air pollutants as variables. The data matrix usually consisted of 14 objects and 21 variables. However, whenever it was necessary to focus on the effects of certain objects or variables, appropriate sub-matrices were selected. The algorithms for the two methods have been included as summaries in the Supplementary Information I. Detailed description of the mathematical basis and application tutorials for the PROMETHEE and GAIA procedures are available in the literature (7, 20). In general, PROMETHEE ranks the objects (houses in this work) according to a given set of variables (e.g. concentration of individual pollutants, etc) (30). The method requires that each variable is separately modelled and optimised (ie ranked top-down (maximised) or bottom-up (minimised)). In this study, the concentrations of the pollutants were "minimised" within the framework of the assumption that lower values of these variables indicate better air quality and the recommendation of the European Collaborative Action on Indoor Air Quality (2) that indoor VOC concentrations should be kept As Low As Reasonably Achievable. In contrast, distance from a major road and similar attributes were set to "maximise". The 'V-shaped' preference function with one threshold (details- Supplementary Information I) was used for such criteria. For variables such as the presence of a garage and type of a building (low set or high set), where there are only two possible choices, the "Usual" preference function (details-Supplementary Information I) was used for the "maximised" cases: P = 0 for d ≤ 0, otherwise P =1, and for the "minimised' cases: P = 1 for d < 0, otherwise P = 0, where P expresses the preference of a house over another and d denotes the difference between each pairwise comparison of the variables for two houses. A partial ranking order (PROMETHEE I; φ^{+} and φ^{-}) and a complete ranking order (PROMETHEE II; φ) were obtained for the houses according to a set of rules (Supplementary Information I).

GAIA, on the other hand, is a special type of Principal Component Analysis (PCA) that evaluated and presented PROMETHEE II results as PC1 (principal component 1) versus PC2 (principal component 2) biplots. Thus, in addition to providing rank order for objects, the PROMETHEE procedures also acted as data pre-treatment for GAIA. The results of the GAIA analysis obtained in this work were interpreted according to the guidelines given by Keller et al (7) and Espinasse et al (20). The most important of these are summarised below: (i) the longer a projected vector for a variable, the more variance it contains, (ii) independent variables have orthogonal vectors (ie the covariance is zero), (iii) vectors oriented in the same direction (ie the covariance is >0 and high) are similar (ie they represent equivalent information) while those oriented in opposite directions represent conflicting information (iv) objects projected in the direction of a particular variable are strongly related to that variable (v) dissimilar objects have significantly different PC coordinates while similar objects appear as clusters and (vi) if the decision vector, π, is long, the best objects are those found in its direction and are the farthest from the origin.

Edwards et al (31) recently described the application of Varimax rotation to residential indoor air quality data in Finland. To compare the outcomes of the

PROMETHEE and GAIA with those of factor analysis, the data were subjected to factor analysis using Statistical Package for Social Science (SPSS) for windows version 11.0.

Results and discussion.

Analysis with the PROMETHEE multivariate ranking method. The PROMETHEE partial ranking preference flow chart for the indoor air quality data is displayed in Figure 1, while the PROMETHEE II complete ranking results for the houses for the indoor air data are presented in Table 1.

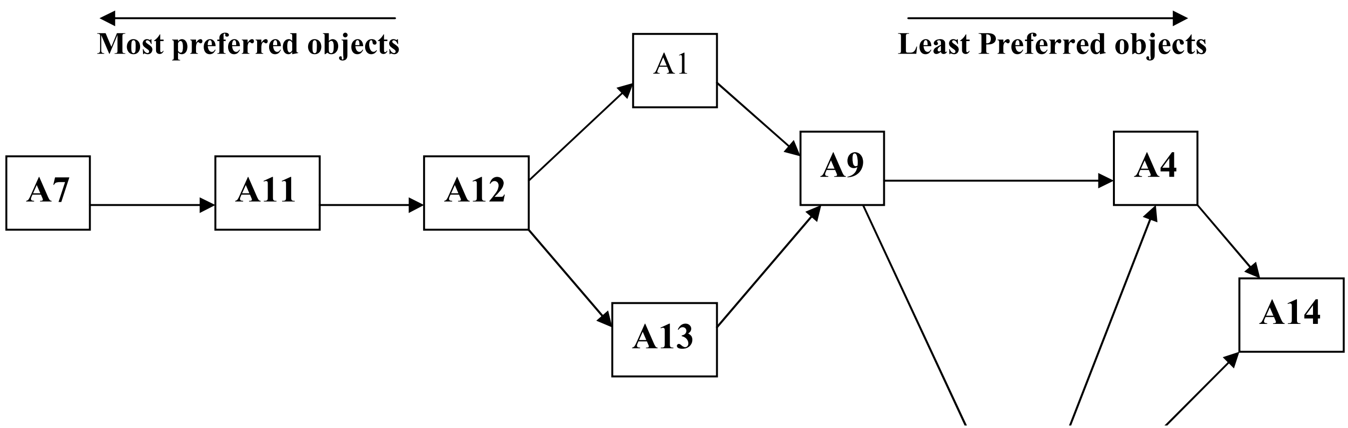
Table 1: PROMETHEE (II) complete ranking results for the houses and the health complaints of the occupants.

House	Net outranking flow* from the indoor data* *	Net outranking flow* from the outdoor data**	Net outranking flow* from the indoor data in which VOCs were the only pollutants examined)**	Health complaints
A7	0.183 (1)	0.090 (4)	0.200 (2)	Cough, wheezing and emphysema
A6	0.138 (2)	0.107 (3)	0.230 (1)	None
A11	0.117 (3)	0.127 (1)	0.153 (4)	Cough and wheezing
A3	0.100 (4)	-0.059 (10)	-0.005 (7)	Emphysema
A12	0.095 (5)	0.125 (2)	0.157 (3)	Hay fever and allergy
A13	0.008 (6)	-0.078 (11)	-0.068 (8)	None
A1	0.004 (7)	0.082 (5)	0.036 (6)	Hay fever and allergy
A8	-0.022 (8)	0.064 (6)	-0.154 (10)	Asthma
A9	-0.033 (9)	-0.015 (8)	0.077 (5)	None

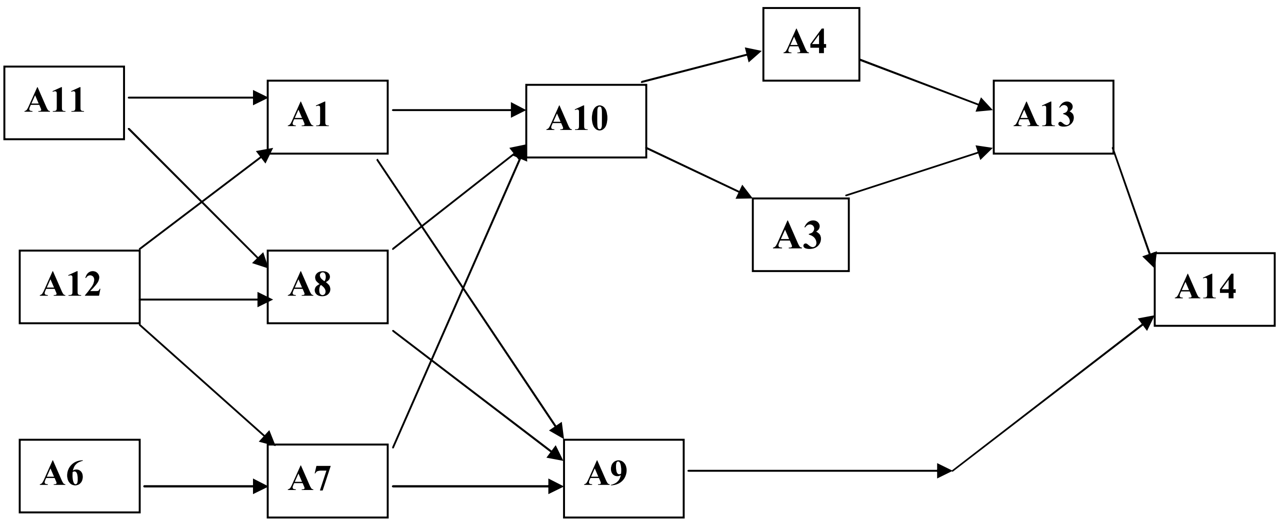
A4	-0.092 (10)	-0.047 (9)	-0.100 (9)	Hay fever, allergy cough, wheezing, emphysema
A10	-0.178 (11)	0.012 (7)	-0.189 (11)	Asthma, hay fever, allergy cough and wheezing
A14	-0.323 (12)	-0.411 (12)	-0.343 (12)	None

*As shown in the Supplementary Information, the net outranking flow (φ) for an object A is such that $\varphi(A) = \varphi^{+}(A) - \varphi^{-}(A)$, where φ^{+} expresses how it outranks other objects and φ^{-} shows how it is outranked by all other objects. The higher the value of φ(A) is, the higher the preference for A. **Figure in parenthesis denotes the rank of the house: most preferred = (1) and least preferred = (12).

(The outdoor data was not used at this stage and houses A2 and A5 were not included in the ranking since the distance of house A2, from the road, and the fungi, bacterial and supermicrometre particle number concentrations of house A5 were not available. The concentration of isopropylbenzene was also not taken into consideration since this pollutant was not detected in many of the houses.) As discussed in detail in the Supporting Information, PROMETHEE I partial ranking (7, 15) highlights one of the following three possible outcomes viz (i) one object is preferred to the other (ii) there is no difference between the two objects or (iii) the objects cannot be compared. As a rule (7,15,30), comparable objects are joined by one or more arrows, incomparable objects are unconnected by arrows and comparable objects to the left of any object are preferred to that object. It can be seen from Figure 1 that the houses are distributed from the most preferred on the left to the least preferred on the right.



(a)



(b)

Figure 1: PROMETHEE I partial preference flow chart showing the rank order of the houses based on (a) indoor air quality influencing variables and (b) outdoor air quality influencing variables. (The houses are ranked from the best performing (on the left) to the least performing (on the right); the numbers in the boxes refer to the codes for individual houses.)

Thus, the best- performing houses (based on the variables) are A6 and A7 and the worst is house A14. However, houses A6 and A7 cannot be compared to each other, which means that the performance of one house on the variables is different, and they are alternative choices.

PROMETHEE II full ranking (Table 1) of the houses based on their indoor air qualities eliminated the incomparability of A7 and A6. Consequently, A7 was identified, as the most preferred house followed by A6, and A14 remains the least preferred house. Although PROMETHEE II appears to be more efficient in ranking the houses, it is less informative than PROMETHEE I. It is also evident that the net outranking flow values, φ, between some of the houses (Table 1) are so close that differentiation between them has little practical significance.

For the outdoor air quality data presented as Supplementary Information III, the PROMETHEE I result shows that A6, A11 and A12 are the most preferred buildings, while the least preferred is house A14 (Figure 1b). Thus, the performance of the houses based on their indoor and outdoor air quality- influencing variables did not produce exactly the same outcome. This was to be expected because some of the pollutants are generated exclusively from indoor or outdoor sources, while some are generated from both indoor and outdoor sources.

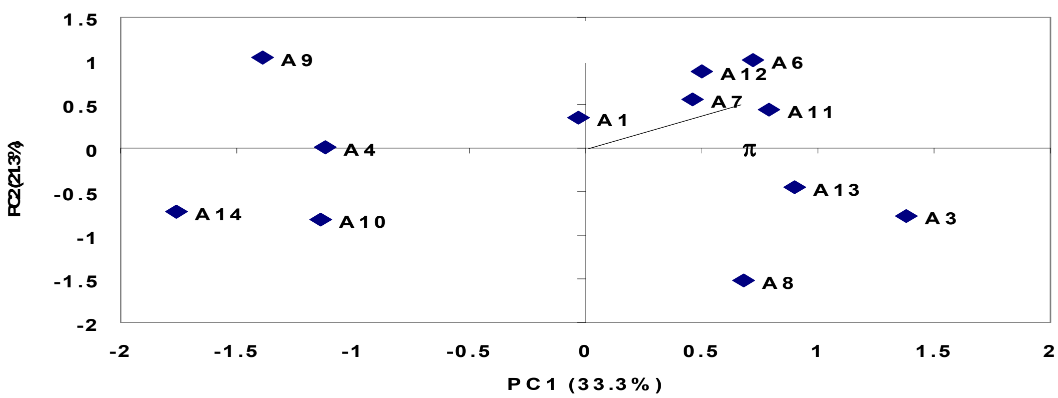
The PROMETHEE II full ranking of the houses based on their outdoor air quality, along with the net outranking flow is also presented in Table 1. This full ranking led to the removal of the incomparability of A6, A11 and A12, and presents the spread of the houses in such a way that the farther apart the net outranking flow of any two houses, the larger the preference of the house with the larger net flow over that with a lower net flow.

The multi-criteria data analysis can be performed in different ways. For example, some of the pollutants or building characteristics may be excluded from the analysis to investigate the effect of such variables on the ranking of the houses. If the submicrometre and supermicrometre particle concentrations were excluded from the indoor air quality data analysis, there was not much difference in the ranking: A6 was one of the best performing houses and A14 is still the worst performing. A similar outcome was obtained when both fungi and submicrometre particle concentrations were excluded to examine the effect of VOC concentrations alone on the indoor air quality. The wind speed, wind direction, temperature and relative humidity observed during the measurements were in the ranges 9-25 km h^{-1}, 110-283^0, 15-23^{0}C and 54-92% respectively. If these variables are included in the indoor air quality data matrix, the ranking does not change appreciably. It therefore appears that the differences in the VOC concentrations among the houses are the dominant factors influencing the ranking of the houses. But this does not imply that other pollutants do not influence indoor air quality. Rather, it suggests that the values of the other pollutants do not vary as widely as those of the individual VOC concentrations.

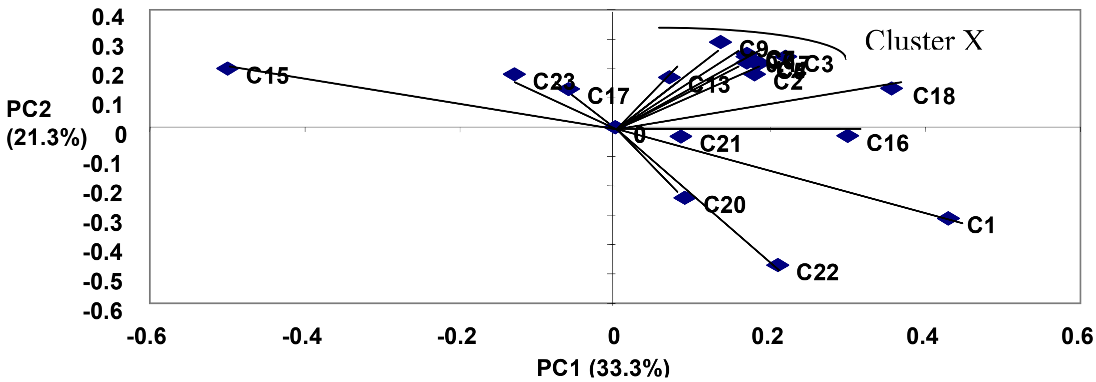
Various studies conducted to explore the association between Total Volatile Organic Compounds (TVOC) concentrations and health effects have produced no consistent outcomes. While some studies (32) suggest that there is a positive association, others (33) found no such associations and some (34) propose that there is a negative association. Since the occupants generally spend more time inside than outside their houses, it is reasonable to assume (32) that indoor air quality will significantly affect their levels of exposure to airborne pollutants more than outside air quality. The net outranking flow values and ranking of indoor environments of the houses obtained from

PROMETHEE II analysis were therefore compared with the information on the health status of the occupants obtained from the questionnaire administered before sampling (Table 1). The comparison reveals that there is no consistent association between the rank order of the houses and the reported health status of the occupants. In particular, no health complaints were recorded in some of the best-(A6), average-(A9 and A1), and worst-(A14) performing houses. Further, while the occupants of some of the worst-performing houses (A10 and A4) had the highest numbers of health complaints, occupants of one of the least preferred houses (A14) had no health complaints. There could be many reasons for the lack of association. One of the most likely is that "self-reporting" by the occupants is not the most objective measure for assessing health status. Secondly, there could be additional differences between the houses, such as susceptibility of the occupants, which were not examined in this analysis. Further, the reported health effects may be associated with other pollutants such as the concentrations of the sub-and supermicrometre particles, fungi and bacteria as well as the various allergens not taken into consideration in the parameters used for the ranking reported in Table 1.

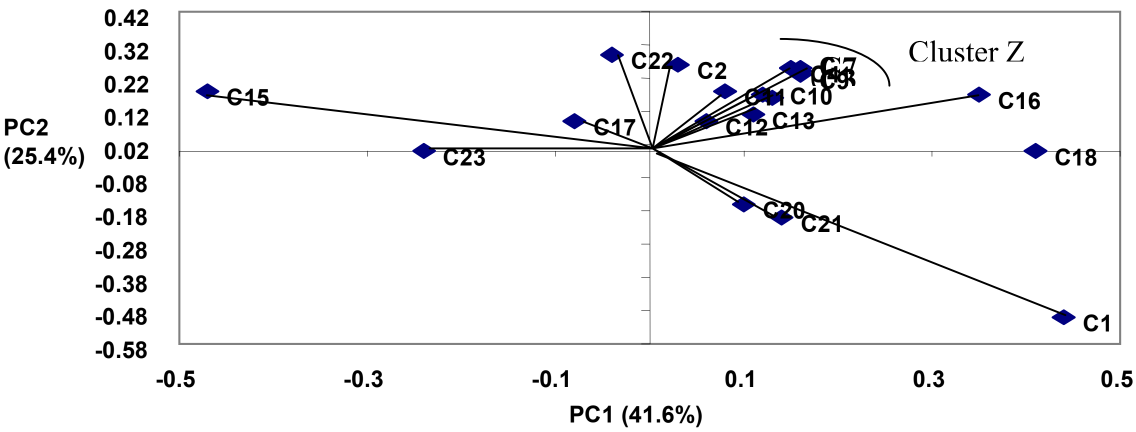
Pattern recognition and significant variables. In order to examine the variables that played the most important role in the ranking of the houses, GAIA analyses of the 14 houses against the 21 variables listed in the Supplementary Information II and III were performed. All variables were given equal weighting and each Principal Component (PC) in the resultant GAIA plots (Figure 2) is associated with a data variance value.



(a)


Cluster X
C15
C23
C17
C9
C13
C2
C3
C18
C21
C16
C20
C1
C22
0
PC2
(21.3%)
PC1 (33.3%)
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

(b)


Cluster Z
C15
C22
C2
C17
C23
C10
C12
C13
C16
C18
C20
C21
C1
PC2
(25.4%)
PC1 (41.6%)
0.42
0.32
0.22
0.12
0.02
-0.08
-0.18
-0.28
-0.38
-0.48
-0.58
-0.5
-0.3
-0.1
0.1
0.3
0.5

(c)

Figure 2: GAIA plot showing (a) scores for the houses based on their indoor air quality, with the decision axis (π) pointing in the direction of the most preferred house, (b) loadings for the indoor air quality influencing variables (clustered vectors (X) consisted of C3, C4, C5, C6, C7, C8, C9, C10, and C13) and (c) loadings for the outdoor air quality influencing variables (clustered vectors (Z) consisted of C3, C4, C5, C6, C7, C8, C9, C10, C11, C12 and C13). (C1 = distance from the park; C2 = Benzene; C3 = Toluene; C4 = Ethylbenzene; C5 = m-Xylene; C6 = p-Xylene; C7 = o-Xylene; C8 = Isopropylbenzene; C9 = n-Propylbenzene; C10 = Trimethylbenzene; C11 = 4-Isopropyltoluene; C12 = Naphthalene; C13 = Hexane; C15 = Distance from a major road; C16 = Garage; C17 = Age; C18 = Type; C20 = Bacteria; C21=Fungi; C22 = Submicrometre particle number concentration; C23 = Supermicrometre particle number concentration. All variables had equal weighting.)

Principal component 1 (PC1) account for the largest data variance, while the subsequent PCs carry variance in decreasing order. The first two PCs account for 54% of the total variance in the original indoor data set and 61% of the total variance in the outdoor data set. It is noteworthy that GAIA provides detail information only on the first two PCs.

Analysing the GAIA plots displayed in Figure 2 (a-c) led to the following conclusions:

Indoor air:

a) Figures 2a and b are the scores and loadings plots for the indoor air quality data respectively. PC1 (in Figure 2a) distinguishes the houses mainly on the basis of their characteristics, with the majority of the high set houses (A6, A7, A11, A12 and A13) having positive scores while most of the low set houses have negative scores (A1, A4, A9, A10 and A14). Since the decision vector, π, points along the PC1 axis, and is relatively long (indicating a high degree of significance), the best performing houses are the high set houses, which lie in the direction of the decision line and are located away from the origin, ie houses A11 and A7. (‘High set’ refers to a house that is elevated above ground on timber or brick stumps, and ‘low set’ indicates a house built directly on a concrete slab; differences in house design and building materials have effects on the air exchange rate.) In Figure 2b, some of the longest positive loadings vectors and therefore, the most significant are due to the distance from the park (C1), type of garage (C16) and type of building (C18) while the longest negative loadings vector is due to the distance from a major road (C15). This confirms that the houses on PC1 of Figure 2a are separated mainly according to the characteristics of the houses. Since the better performing houses (ie A6, A7, A8, A11, A12, and A13) are almost exclusively high set, timber houses without in-built garages, these variables exert considerable influence on indoor air quality and any attempt to reduce indoor air pollution must take this into consideration.

Under the normal ventilation conditions employed for this investigation, air exchange rates in Brisbane ranged between 2 and 5 h^{-1} (35). Therefore the additional ventilation in the high set houses probably contributes to their better indoor air qualities. Interestingly, the fungi and bacteria compositions of these high set (timber) houses were also generally lower than those found in their low set (brick) counterparts (28).

In contrast to the long loadings vector for the submicrometre particle number concentration, the vectors for the concentrations of 4-isopropyltoluene, naphthalene and fungi as well as age of the building are short. This suggests that the concentrations of the submicrometre particles strongly influence the ranking of the houses (7) while the other variables exert weak influences on the rank order. The observed weak influence of the age of the building on the rank order is consistent with a Swedish study, in which the indoor Total Volatile Organic Compounds (TVOC) levels of office buildings aged between 1.5 and 10 years showed no apparent effect of the age of the buildings (36). Although new and newly renovated buildings are subject to elevated levels of VOCs, sometimes up to 20 times the recommended maximum limit, the effect diminishes within 12 weeks (37,38,39). In agreement with this well-established effect of renovation on indoor air quality, Figure 2a shows that house A8, which was renovated at the time of the study, is atypical with a high negative score on PC2 in contrast to the positive PC2 scores observed for most of the other high set houses. Since the loadings vector for “age of the building” is short, this effect would be expected to disappear after a few weeks as reported in the literature (37, 38,39).

b) A few broad groups of variables are apparent from the GAIA results in Figure 2b. Group A consists of the bacterial concentration (C20) and submicrometre particle number concentrations (C22), whose loadings vectors are oriented in the same direction and are also fairly close to the vectors for the park level (C1), suggesting that the park is a possible source of indoor bacteria and submicrometre particles**.** Group B has two members: Park (C1) and fungi (C21) in agreement with the finding that the fungi originate from the park (28). Group C has garage (C16) and type of building (C18), in agreement with the observation that low set houses almost always had in-built garages and vice versa. The vectors for concentrations of benzene (C2), toluene (C3), ethylbenzene (C4), m-xylene (C5), p-xylene (C6), o-xylene (C7), and trimethybenzene (C10) are oriented in the same direction and are grouped together as group D. In a study conducted in Finland (31), because of the strong dependence of the concentrations of these compounds on wind direction, it was concluded that they originated from sources far away from the study sites. By contrast, in the present study, the exclusion or inclusion of wind direction in the multivariate analysis had no noticeable effect on the ranking of the houses and on the outcomes of the GAIA plots. Therefore, it is reasonable to assume that the compounds have predominantly indoor sources such as particleboards, floor and wall coverings as well as cleaning and personal care products (40). Group E consists of n-propylbenzene (C9), and hexane (C13). The two compounds are known to have traffic origin (31) but their vector loadings are almost orthogonal to that for distance from the road (C15), suggesting (7, 20) that these compounds are independent of the distance from the road (7, 20) and may not have predominantly traffic origin in this study. The loadings vectors for 4-isopropyltoluene (C11) and naphthalene (C12) are oriented in the same direction (not apparent in Figure 2b) and are designated as group F. However, the loadings vector for each of these variables is so short that the amount of variance they account for is negligible. Since these vectors are in the opposite direction to those for benzene, toluene and trimethylbenezene, which are thought, in this study, to have predominantly indoor sources, it is reasonable to suggest that the naphthalene and 4-isopropyltoluene, in these houses, originate from sources different from those of benzene, toluene and trimethylbenezene. Group G contains distance (C15), age (C17) and supermicrometre particle number concentration (C23), which have rather weak correlations with one another. While re-suspension of supermicrometre particles may occur more readily from older than newer houses, the main source of such particles is probably vehicle emission. Morawska *et al* (27) showed that the concentrations of large particles indoors tend to closely follow the concentrations outdoors. Therefore, it is reasonable to expect a correlation between supermicrometre particles number concentration and the distance of a house from a major road.

c) An additional feature of the GAIA plot (Figure 2b) is that some of the variables are oriented at approximately 180^0 to each other. For example, the loadings vectors for naphthalene (C12) and 4-isopropyltoluene (C11) on one hand, and benzene (C2), tolulene (C3), ethylbenzene (C4), trimethylbenzene (C10) and o-, m-, p-xylene (C5, C6, C7), on the other, are oriented in opposite directions on PC1 (Figure 2b). Similarly, the vectors for the age of the building (C17) and bacterial concentration (C20) are oriented at approximately 180^o to each other. Such vectors have been described as "conflicting criteria" by Keller et al (7) and would be expected to have opposing effects on the ranking of the houses. In addition, where the vectors represent the concentrations of pollutants, their conflicting orientations may reflect differences in the sources of the pollutants.

d) The vector for 'Park' (C1) (Figure 2b) is almost orthogonal to the vectors for benzene (C2), toluene (C3), ethylbenzene (C4), trimethylbenzene (C10) and o-, m-, p-xylene (C5, C6, C7). This means that these VOCs are independent (7) of the distance from the park and may or may not originate from the park. Similarly, the vector for 'supermicrometre particle number concentration' (C23) is independent of those for benzene (C2), toluene (C3), ethylbenzene (C4), trimethylbenzene (C10) and o-, m-, p-xylene (C5, C6, C7) as they are unlikely to have identical origins. The vector for 'Distance' (C16) is also independent of those for 'Hexane' (C13), suggesting that the hexane found in the houses do not arise predominantly from vehicular exhaust emission.

Outdoor air:

e) For outdoor air (Figure 2c), the loadings vectors for the fungal

concentration (C21), bacterial concentration (C20 and park (C1) strongly correlates (7) since their vectors are oriented in the same direction. This confirms the strong correlation obtained by means of correlation analysis (28) for the bacterial and fungal outdoor concentrations and suggests that the two are from the same source, possibly the park. Figure 2c also shows that the vectors for garage (C16) and type of house (C18) are oriented in approximately the same direction, in keeping with the fact that high set houses almost always had no in-built garages while the converse is true for low set houses. Most of the projected vectors for the VOCs are oriented in the same direction, possibly because they have the same outdoor source, and therefore, affect the ranking of the houses in the same way. Unlike the situation in the indoor environment, naphthalene correlates strongly with other VOCs like toluene, trimethylbenzene, xylene and ethylbenzene in the outdoor environment. Thus naphthalene and the other VOCs do have similar outdoor but different indoor origins. The vector for the submicrometre particle number concentration correlates only weakly with variables such as toluene

(C2), age (C17), and distance (C15). Although vehicle emissions, which are likely to be the main sources of VOCs in outdoor environment is also a probable source of submicrometre particles, it is conceivable that the bulk of the submicrometre particles found in these houses arise from other combustion processes and long range transport. Variables such as supermicrometre particle concentrations, distance from the road and age of the building also correlate but the variance accounted for by the distance from the road and supermicrometre particle concentration variables are much higher than that contained by the age of the building. Since traffic emissions are well-established sources of particles (26, 27), the correlation between distance and particle number concentration is expected.

The original indoor and outdoor matrices contained 21 data variables. Since several of these showed close correlation, it may be advantageous to replace such group(s) with one representative variable. Thus, the data was examined with different numbers of variables, which had the highest loadings as well as those, which are representatives of the pollutants that showed collinearity. As shown in the Table 2, the amount of data variance accounted for increased when fewer variables were employed but the full outranking flow remained practically unchanged.

Table 2: The effect of the variables examined on the PROMETHEE and GAIA analysis

Microenvironment	Variables/ (number of variables)	% Variance accounted for	Best 3 objects	Worst 3 objects
Indoor	All **(21)**	55	A6, A11, A7	A4, A10, A14
Indoor	C1, C2, C3, C5, C7, C9, C12, C13, C15, C16, C18,	56	A6, A7, A11	A4, A9, A14

	C20, C21, C22, C23 **(15)**			
Indoor	C1, C2, C3, C5, C9, C15, C16, C18, C20, C22, C23 **(10)**	60	A6, A7, A11	A4, A10, A14
Indoor	C1, C3, C5, C15, C18, C23 **(6)**	73	A6, A7, A12	A4, A10, A14
Outdoor	All (**21**)	66.5	A12, A8, A11	A4, A9, A14
Outdoor	C1, C2, C3, C5, C7, C9,C12, C13, C15, C16, C18, C20, C21, C22, C23 **(15)**	67	A8, A12, A7	A9, A4, A14
Outdoor	C1, C2, C3, C5, C9, C15, C16, C18, C20, C22, C23 **(10)**	67	A8, A7, A12	A4, A9, A14
Outdoor	C1, C3, C5, C15, C18, C23 **(6)**	77	A8, A12, A7	A9, A4, A14

Hence, the top three performing houses and the worst ones are generally the same when different variables were employed. Since the worst performing houses are usually associated with high VOC concentrations, any control measures should target these pollutants.

Factor analysis. When the indoor air quality data was subjected to factor analysis, seven factors accounting for approximately 91% of the total variance were retained on the basis of the well-established eigenvalue-greater-than-one rule. Factors 1, 2, 3, 4, 5, 6 and 7 explained 30%, 17 %, 14%, 10%, 8%, 7% and 5% of the variance respectively. The factor loadings along with the communality of each of the variables for the rotated (Varimax) and unrotated PCs are presented as Supplementary Information IV. The varimax rotation was carried out to maximise the variance in order to facilitate the interpretation of the results. However, as observed by Scrimshaw et al (41) the unrotated and rotated PCs gave broadly similar results. Therefore the former was compared with the GAIA results. Comrey and Lee (42) suggested that factor loadings of the order 0.55 (30 % overlapping variance) are considered good and those of the order of 0.45 (20% overlapping variance) are fair. Interpretation of the factors have therefore been limited to those with loadings greater than 0.50.

The volatile organic compounds loaded on three main factors: 1, 4 and 6. As observed in the result obtained with GAIA, toluene, ethylbenzene, o-xylene, p-xylene, m-xylene and 1, 3, 5-trimethylbenzene correlated on Factor 1, confirming that they have common sources in these indoor environments. Other similarities between the results obtained by GAIA and unrotated component matrix include: (i) correlation between bacterial concentration and submicrometre particle number concentration, (ii) increase in hexane and submicrometre particle concentrations as the distance from the road decreased (iii) increase in fungal concentration as the distance from the park decreased. and (iv) the analogy between the seven groups of variables identified from the GAIA results and the seven factors in the unrotated component matrix. However, the factor analysis did not give any ranking information about the quality of the air in the houses. In addition, there are two minor differences in the results obtained by the unrotated component matrix and those obtained by GAIA for the indoor environments. In the GAIA results, the bacteria and fungal concentrations are weakly correlated, but in the unrotated components matrix, they are not correlated. Secondly, the age of the house correlated with the supermicrometre particle number concentration in the GAIA analyses but the unrotated component matrix shows that the age of the house correlated negatively with the type of garage. There are no immediate explanations for these differences.

For the outdoor air data, 86% of the variance was explained by four unrotated factors with factors 1, 2, 3 and 4 accounting for 53%, 19.5%, 7.5% and 6% of the variance respectively. Patterns recognized from this unrotated PC matrix also reinforce those observed using GAIA procedure. Again, unlike PROMETHEE, factor analysis did not provide ranking information on the quality of air in these microenvironments.

Conclusions

This study demonstrated the ability of the Multi-criteria decision making methods (MCDM), PROMETHEE and GAIA, to provide partial pre-order and net ranking information necessary to select one house in preference to all others, on the basis of its air quality. Such ranking analysis has not previously been reported in indoor air literature. The study has also shown that patterns in GAIA plots cannot only assist the identification of the plausible sources of airborne pollutants in various microenvironments, but also provide information on the significant variables that are essential for the discrimination of objects, and those that are important to be monitored in their own right or as representatives of their class. Since the better performing houses are almost exclusively high set timber houses without in-built garages but with relatively low submicrometre particle and toluene concentrations, these variables exert considerable influence on indoor air quality and any effort to reduce indoor air pollution must take them into consideration.

Attempts to relate PROMETHEE ranking information with the health complaints in the buildings produced limit success possibly because psychosocial and other non-quantifiable response modifiers play significant roles in building related health complaints.

Overall, PROMETHEE and GAIA (i) preserve as much information as possible (17), (ii) avoid trade-offs (17,18), (iii) permit sensitivity analysis (7, 18), (iv) require no interaction by the user (19), and (v) are readily available in the form of a user friendly software package (17, 18). Therefore, they offer a hitherto unexplored potential to assist ranking analysis of air quality, pattern recognition and source apportionment of airborne pollutants as well as elucidation of effective remedial measures for indoor air pollution and evaluation of exposure-response relationship.

Acknowledgements

The authors acknowledge the financial support of the Built Environment Research Unit (BERU) of the Queensland Department of Public Works and Australian Research Council, through SPIRT Grant No C69804416. Thanks are also due to Elizabeth Christensen and Jane Hitchins for assistance with the study and to Akin Ayoko for drawing the Figures.

Supporting Information Available

Supplementary Information containing the algorithm for PROMETHEE and GAIA, the primary data used in the multicriteria decision making procedures and the factor analysis results is provided.

Literature cited

(1) Lee, S-C.; Guo, H.; Li, W-H.; Chan, L-Y. *Atmos. Environ*. **2002**, 36, 1929 and references therein.

(2) ECA (European Collaborative Action on Indoor Air Quality and its Impart on Man), Total Volatile Organic Compounds (TVOC) in Indoor Air Quality Investigations. Report No 19, EUR 17675EN, **1997**.

(3) Veltkamp, P.R.; Hansen, K. J.; Barkley, R. M.; Sievers, R. E. *J. Geophys. Res,* **1996,** 101, 19495.

(4) Li, H.; Banner, C. D.; Mason, G. G.; Westerholm, R. N.; Rafter, J..J. *Atmos. Environ.* **1996**, 30, 3537

(5) Lee, E.; Chan, C. K.; Paatero, P. *Atmos. Environ*, **1999**, 33, 3201.

(6) Mitra, S.; Ray, B. *Atmos. Environ.* **1995**, 29, 3345.

(7) Keller, H.R.; Massart, D. L.; Brans, J. P. *Chemom. Intell. Lab. Syst.* **1991**, 11, 175.

(8) Brans, J. P.; Mareschal, B.; PROMETHEE-GAIA Visual Interactive Modelling for Multicriteria Location Problems, Internal Report University of Brussels, STOOTW/244, 1989.

(9) Kokot, S.; King, G.; Keller, H. R.; Massart, D. L. *Anal. Chim. Acta* **1992,** 268, 81.

(10) Kokot, S.; King, G.; Keller, H. R.; Massart, D. L. *Anal. Chim. Acta*, **1992**, 259, 267.

(11) Kokot, S.; Phuong, T.D. *Analyst*, **1999**, 124**,** 561.

(12) Brans, J. P.; Mareschal, B.; Vincke, P. H.; PROMETHEE: a new family of outranking methods in MCDM. In Brans, J. P. (Ed.), Operational Research 84.North-Holland, Amsterdam **1984**, pp 477-490

(13) Salminen, P.; Hokkanen, J.; Lahdelma, R. *Eur J. Oper. Res*., **1998***,* 104, 485.

(14) Olson, D. L. *Eur. J. Oper. Res.*, **2001**, 130, 576.

(15) Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J.; In Data Handling in Science and Technology 20A, Handbook of Chemometrics and Qualimetrics Part A, Elsevier, Amsterdam, Chapter 26, **1997**, pp771-804.

(16) Lerche, D.; Brüggemann, R.; Sørensen, P.; Carlsen, L.; Nielson ,O. J. *J. Chem. Inf. Comput. Sci.,* **2002**, 42, 1086.

(17) Le Teno , J.F.; Mareschal, B. *Eur J. Oper. Res.*, **1998**, 109, 522.

(18) Macharis, C.; Springael, J.; De Brucker, K.; Verbeke, A. *Eur J. Oper. Res*., **2004**, 153, 307.

(19) Mahmoud, M.R.; Garcia, L.A. *Environ. Model. Software*, **2000,** 15, 471.

(20) Espinasse, B,; Picolet G.; Chauraqui, E. Eur. J. Oper. Res. **1997**, 103, 389.

(21) Martin, J.M.; Fajardo, W.; Blanco, A.; Requena, I. *Int. J. Intell. Syst.* **2003**, 18, 711.

(22) Ozelkan, E.C.; Duckstein, L. *J. Environ. Manage.* **1996**, 48, 69.

(23) Ayoko, G. A.; Uhde, E.; Moraswka, L.; Kokot, S.; Gilbert, D. Presented at Healthy Building 2003, 7th International Conference, National University of Singapore, Singapore, 7-11December, 2003.

(24) Ayoko, G. A.; Morawska, L.; Ristovski, Z.D.; Jayaratne, R.; Lim, M.; Kokot, S.; Abstracts of the Sixth International Aerosol Conference, Taipei, Taiwan, September 8-13, **2002**, 1, 597.

(25) Ayoko G, A.; Bonire, J. J.; Abdulkadri, S.S.; Olurinola, P. F.; Ehinmidu, J. O.; Kokot, S.; Yiasel, S. *Appl. Organometal. Chem*. **2003,** 17, 749.

(26) Morawska, L.; He,C.; Hitchins, J.; Mengersen, K.; Gilbert, D. *Atoms. Environ*. **2003**, 37, 4195.

(27) Morawska, L.; He, C.; Hitchins, J.; Gilbert, D.; Parappukkaran, S. *Atmos. Environ.* **2001**, 35 , 3463.

(28) Hargreaves, M.; Parappukkaran, S.; Morawska, L.; Hitchins, J.; He, C.; Gilbert, D. *Sci. Tot. Environ.* **2003,** 312, 89.

(29) Ayoko, G. A.; Hitchins, J.; Morawska, L.; Gilbert, D.; Christensen, E. *to be submitted for publication.*

(30) Brans, J. P. PROMCALC (Version 3:1), Centre for Statistics and Operations Research, Free University of Brussels, Brussels, 1991; Decision Lab 2000 Executive edition, 1999, Getting started guide, Visual Decision Inc., Montreal, Canada.

(31) Edwards, R. D.; Jurvelin, J.; Koistinen, K.; Saarela, K.; Jantunen, M. *Atmos Environ.* **2001**, 35, 4829.

(32) Hodgson, M. J.; Frohliger, J.; Permar, E.; Tidwell, C.; Traven, N. D.; Olenchock, S. A.; Karpf, M. *J. Occup. Med*. **1991,** 33, 527.

(33) Nagda, N. L.; Koontz, M. D.; Albercht, R. J.; Effect of ventilation rate in a health building. In Proc. IAQ 91 "*Healthy Building*" ISBN 0-910110-84-0. Atlanta, GA 30329: ASHRAE, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 1991, 101.

(34) Sundell, J.; Anderson, B.; Anderson, K.; Lindvall, T. *Indoor Air*, **1993**, 3, 82.

(35) Morawska, L.; Jamriska, M. Time evolution of environmental tobacco smoke and other combustion aerosols. Proceedings of the 12th International Conference of CASANZ, Perth, **1994,** 603.

(36) Ekerg, L. E. *Atmos. Environ.* **1994,** 28, 3571.

(37) Brown, S.K. Assessment and control of VOCs and house dust mites in Australian buildings *in Proceedings of the 7th International Conference on Indoor Air Quality and Climate, Nagoya, Japan, 21-26 July 1996* (Indoor Air'96), **1996**, 2, 97.

(38) Brown, S.K. Air toxics in a new Victorian dwelling over an eight -month period *in Proceedings of the 15th International Clean Air and Environment Conference, Sydney, Australia, 26-30 November* **2000***,* 1*, 458*.

(39) Wolkoff, P.; Clausen, P.A.; Nielsen, P.A.; Mϕlhave, L. *Indoor Air*, **1991**, 1, 478.

(40) Bluyssen, P. M.; Fernandes, E. D. O.; Groes, L.; Clausen, G.; Fanger, P.O.; Valbjϕrn, O.; Bernhard, C.A.; Roulet, C. A. *Indoor Air*, **1996**, 6, 221.

(41) Scrimshaw, M. D.; Lester, J. N. *Environ. Sci. Technol*., **2001**, 35, 2676.

(42) Comrey, A. L.; Lee, H. B. A first course in factor analysis (2nd ED.) Hillsdale, NJ: Erlbaum, **1992**.

Cover Sheet for Supplementary Information

Authors: Godwin A. Ayoko*[a], Lidia Morawska , Serge Kokot and Dale Gilbert

Title: An application of multicriteria decision making methods to air quality in the microenvironments of residential houses in Brisbane, Australia

Number of pages: 8

Number of table: 5

Supplementary Information I: Algorithm for PROMETHEE and GAIA

PROMETHEE is a non-parametric method, which ranks objects (houses, in this paper) on the basis of a range of variables. For each variable, the decision maker must indicate: (i) a preferred ranking sense i.e. top-down (maximised) or bottom-up (minimised), (ii) a weighting – set to 1 by default but can be altered, if decision making experiments require analysis of alternative scenarios (iii) a preference function, *P (A, B),* which defines how one object is to be chosen relative to another. The stepwise procedure (7) is presented below:

Step 1: Conversion of the raw data matrix to a difference matrix.

For each variable, the column entries, y, of the raw data matrix are subtracted from each other in all possible ways to generate a difference, d, matrix.

Step 2: Application of the preference function, *P (A ,B)*

For each variable, one of the six preference functions available in the PROMCALC (30) or Decision Lab 2000 (30) software is applied to decide how much an outcome *A* is preferred to *B*. The six preference functions are described in the table below.

Step 3: Computation of an overall or global preference index, π

The following equation provides an overall or global index, π, for the comparison of the preference of object *A* over *B*

$$\pi\,(A, B) = \sum_{j=1}^{k} w_j \times P_j(A,B) \qquad (1)$$

w_j = weightings

Step 4: Computation of outranking flows

The positive (φ^+) and negative outranking flows (φ^-) are calculated as shown below.

$$\varphi^{+}(A) = \sum_{x=A} \pi(A,x) \quad (2)$$

$$\varphi^{-}(A) = \sum_{z \in A} \pi(x,A) \quad (3)$$

φ^{+} indicates how an object outranks all others while φ^{-} shows how all others outrank each object. The higher the φ^{+} and the lower the φ^{-}, the higher is the preference for an object.

Step 5: Comparison of outranking flows.

A partial ranking or partial pre-order of the objects is obtained by pairwise comparisons (of *A* and *B*) of all experimental results using the rules below.

1. *A* outranks *B* if:

$$\varphi^{+}(A) > \varphi^{+}(B) \text{ and } \varphi^{-}(A) < \varphi^{-}(B) \quad (4)$$

or

$$\varphi^{+}(A) > \varphi^{+}(B) \text{ and } \varphi^{-}(A) = \varphi^{-}(B) \quad (5)$$

or

$$\varphi^{+}(A) = \varphi^{+}(B) \text{ and } \varphi^{-}(A) < \varphi^{-}(B) \quad (6)$$

2. *A* is indifferent to *B* if:

$$\varphi^{+}(A) = \varphi^{+}(B) \text{ and } \varphi^{-}(A) = \varphi^{-}(B) \quad (7)$$

3. *A* cannot be compared with *B*: in all other cases where B does not outrank A on the basis of rules similar to those outlined in 1 above.

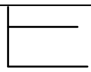
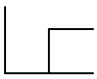
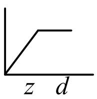
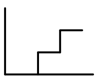
Step 6 Computation of net outranking flow

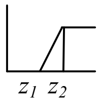
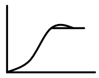
$$\varphi(A) = \varphi^{+}(A) - \varphi^{-}(A) \quad (8)$$

This relationship eliminates the incomparability rule 3 in Step 5 and removes the partial pre-order. Although the net outranking flow, φ, is intuitively more convenient, it is less informative.

GAIA, a data display method, complements the PROMETHEE ranking and provides guidance about the principal variables that contribute to the rank order of the objects. GAIA is also crucial for experimenting with different variable weightings; in this context a special sensitivity decision vector, π, is plotted. A GAIA plot is a PC1 versus PC2 biplot obtained from a matrix that has been formed from a decomposition of the PROMETHEE net outranking flows (7). The interpretation of the GAIA plot is essentially the same as for a PCA biplot. In addition, Espinasse et al. (20) provide a comprehensive list of rules for the interpretation of GAIA plots.

The preference functions are illustrated below (shapes represent only the 'maximise' part of each function; z = threshold value of d when P = 1)

Preference Function	Shape	Mathematical Justification
Usual (No threshold)		$\begin{matrix} y(z)=0 \\ y(z)=1 \end{matrix} \begin{cases} z<0 \\ z\geq 0 \end{cases}$
U-shape(q threshold)		$\begin{matrix} y(z)=0 \\ y(z)=1 \end{matrix} \begin{cases} x<1 \\ x\geq 1 \end{cases}$
V-Shape (p threshold)	 Slope m= 1/z	$\begin{matrix} y(z)=mz \\ y(z)=1 \end{matrix} \begin{cases} z<d \\ z\geq d \end{cases}$
Level (q and p thresholds)		$\left.\begin{matrix} y(z)=0 \\ y(z)=\frac{1}{2} \\ Y(z)=1 \end{matrix}\right\} \begin{matrix} z<d \\ z=0 \\ z>d \end{matrix}$

Linear (q and p thresholds)		$$\left.\begin{array}{c} y(z)=0 \\ y(z)=mz+c \\ y(z)=1 \end{array}\right\} \begin{array}{c} z, z_1 \\ z_1 > z < z_2 \\ z > z_2 \end{array}$$
Gaussian (s threshold)		$$y(z)=\frac{e^z}{1+e^z}$$

House	Volatile Organic Compounds (μg/m^3)												Other pollutants in various units*				House characteristics				
	Ben	**Tol**	**EBen**	**m-xyl**	**p-xyl**	**o-xyl**	**IPBen**	**n-Pben**	**1,3,5-TMben**	**4-IPTol**	**Nap**	**Hex**	**Bact**	**Fungi**	**SMPS**	**APS**	**Dist (m)**	**Gar**	**Age (years)**	**Park (m)**	**Type**
A1	0.44	7.30	0.82	2.19	0.88	0.44	0.03	0.11	0.20	0.11	0.28	28.32	610	1250	4300	1470	400	1	3	60	1
A2	0.53	5.65	0.87	*2.08*	0.86	0.45	0.03	0.08	0.18	0.67	32.52	36.51	250	560	3500	1060	ND	2	31	270	2
A3	0.78	3.73	1.27	3.13	1.29	0.69	0.04	0.11	0.19	0.05	0.22	34.11	160	510	1050	5200	100	2	15	400	1
A4	3.81	21.39	2.36	6.47	2.76	1.12	0.05	0.17	0.34	0.08	0.19	120.84	440	400	7800	1550	500	1	27	60	1
A5	0.45	5.48	1.32	4.01	1.61	0.72	0.04	0.19	0.36	0.13	0.72	37.84	ND	ND	700	ND	10	2	39	150	2
A6	0.42	2.55	0.59	1.41	0.57	0.29	0.02	0.05	0.10	0.04	0.33	20.79	850	750	10900	1320	400	2	40	120	2
A7	0.84	5.76	0.93	2.07	0.89	0.41	0.00	0.08	0.18	0.22	0.91	34.61	460	6250	8000	1430	500	2	23	200	2
A8	0.66	21.93	4.10	10.76	4.66	2.78	0.00	1.21	1.77	0.12	0.38	68.05	150	510	2000	4720	400	2	29	400	2
A9	0.53	11.24	2.02	4.95	1.14	1.17	0.26	0.26	0.60	0.08	0.27	55.60	475	1325	12900	2810	600	1	100	10	1
A10	1.47	28.19	4.55	10.74	4.24	2.43	0.14	0.49	1.17	0.10	0.25	175.70	230	1675	3700	1110	500	2	2	20	1
A11	0.38	5.17	0.85	2.34	0.95	0.45	0.02	0.09	0.19	0.08	0.81	46.65	340	910	7700	4150	400	2	59	90	2
A12	0.45	11.44	1.35	2.71	0.45	0.53	0.03	0.12	0.23	0.23	0.35	45.10	710	500	18700	1110	450	2	4	220	2
A13	1.22	9.46	1.37	3.30	1.37	0.65	0.00	0.13	0.23	0.12	0.21	189.66	80	910	8400	1800	200	1	29	430	2
A14	1.19	23.69	7.77	20.76	8.54	5.10	0.00	1.09	2.58	0.09	0.54	47.58	750	590	16100	1990	500	1	29	140	1

Supplementary information II: Variation of indoor air quality influencing parameters from one house to another ;*Taken from Morawska *et al* 2001; ND = Not determined; Ben = benzene; Tol = toluene; EBen = Ethylbenzene; xyl = xylene; IPBen = Isopropybenzene; n-Pben = n-Propylbenzene; 1,3,5 TMben = Trimethybenzene; 4-IPTol = 4-Isopropyltoluene; Nap = Naphthalene; Hex = hexane; Bact = Bacteria; SMPS = Submicrometre particle number concentration; APS = Supermicrometre particle number concentration; Dist =Distance; . Gar =Garage

House	Volatile Organic Compounds (μg/m^3)												Other pollutants*			
	Ben	Tol	Eben	m-xyl	p-xyl	o-xyl	IPBen	n-Pben	1,3,5-TMben	4-IPTol	Nap	Hexane	Bact	Fungi	SMPS	APS
A1	0.13	1.66	0.29	0.68	0.28	0.13	0.01	0.03	0.04	0.01	0.03	6.11	250	1660	2945	750
A2	0.63	18.28	3.16	6.89	1.60	1.81	0.11	0.40	0.94	0.07	0.94	101.86	560	725	5150	1280
A3	0.48	4.64	0.71	1.80	0.73	0.37	0.02	0.08	0.11	0.02	0.09	11.48	300	650	11670	4640
A4	0.89	8.24	1.01	2.67	1.14	0.51	0.02	0.08	0.00	0.02	0.08	34.27	200	575	8040	1740
A5	0.54	1.73	0.68	1.74	0.71	0.31	0.01	0.05	0.08	0.02	0.06	0.00	210	460	815	ND
A6	0.41	3.45	0.67	1.79	0.72	0.34	0.00	0.06	0.10	0.02	0.12	11.41	475	1640	8720	1808
A7	0.77	4.56	0.61	1.34	0.30	0.31	0.02	0.05	0.09	0.02	0.08	15.27	188	690	7770	1430
A8	0.58	3.04	0.49	1.42	0.60	0.27	0.01	0.06	0.15	0.02	0.07	25.12	75	650	2170	5240
A9	0.43	4.80	0.77	2.10	0.85	0.47	0.03	0.09	0.21	0.02	0.10	20.93	1625	3125	1420	1390
A10	0.41	4.72	0.95	1.89	0.76	0.47	0.02	0.08	0.19	0.02	0.12	47.50	250	2190	3010	980
A11	0.21	2.03	0.31	0.77	0.32	0.15	0.00	0.03	0.04	0.01	0.07	2.64	310	925	8560	4660
A12	0.52	1.49	0.16	0.38	0.16	0.07	0.00	0.01	0.01	0.00	0.01	22.04	390	850	7600	3980
A13	1.26	5.86	0.96	2.33	0.95	0.45	0.02	0.06	0.11	0.03	0.10	19.67	310	860	9545	1765
A14	0.98	20.81	3.19	8.17	3.37	1.89	0.00	0.29	0.72	0.05	0.39	62.18	113	810	13880	2140

Supplementary Information III: Variation of outdoor air quality influencing parameters from one house to another. Taken from Morawska *et al* 2001

SUPPLEMENTARY INFORMATION IV
Varimax rotated component matrix for the indoor air quality influencing variables investigated in the houses (Rotation converged to 11 iterations.)

Variable	Communality	Factor						
		1	2	3	4	5	6	7
Benzene	0.885						**-0.909**	
Toluene	0.943	**0.743**						
Ethylbenzene	0.992	**0.973**						
m-xylene	0.992	**0.918**						
p-xylene	0.984	**0.978**						
o-xylene	0.998	**0.992**						
n-propylbenzene	0.838							
Trimethylbenzene	0.992	**0.978**						
Isopropyltoluene	0.727					**0.813**		
Naphthalene	0.871					**0.799**		
Hexane	0.950		**0.898**					
Distance	0.914			**0.758**				
Garage	0.857						**0.540**	**-0.525**
Age	0.879							**0.882**
Type	0.722						**0.513**	
Bacteria	0.995		**-0.566**		**0.744**			
Fungi	0.935		**0.583**	**0.649**				
Submicrometre Particle number	0.797				**0.817**			
Supermicrometre Particle number	0.959				**-0.822**			
Park	0.968			**-0.925**				

Unrotated component matrix for the indoor air quality influencing variables investigated in the houses

Variable	Communality	Factor						
		1	2	3	4	5	6	7
Benzene	0.671					**-0.534**		
Toluene	0.923	**0.884**						
Ethylbenzene	0.985	**0.969**						
m-xylene	0.989	**0.833**						
p-xylene	0.978	**0.952**						
o-xylene	0.988	**0.947**						
n-propylbenzene	0.928				**0.546**		**0.547**	
Trimethylbenzene	0.988	**0.906**						
Isopropyltoluene	0.781				**0.681**			
Naphthalene	0.650							
Hexane	0.970		**-0.612**					
Distance	0.880		**0.711**					
Garage	0.812						**-0.565**	
Age	0.906						**0.732**	
Type	0.859							
Bacteria	0.703		**0.890**					
Fungi	0.977			**-0.681**		**0.640**		
Submicrometre Particle number	0.843		**0.734**					
Supermicrometre Particle number	0.913			**0.546**	**-0.542**			
Park	0.937		**-0.631**	**0.644**				