# [ENV12] Pollution pattern of volatile organic compounds (VOCs) in drinking water using chemometrics data analysis methods

**<u>Soh Shiau Chian</u>, Md. Pauzi Abdullah**

Schools of Chemical Sciences and Food Technologies, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan.

## Introduction

The occurrence of volatile organic compounds (VOCs) in drinking water create a major concern over the possible health risk because, even at very low concentrations, many of these materials are toxic, carcinogenic or mutagens (Sacks & Akard, 1994). As considerable quantities of VOCs are manufactured in Malaysia, their use is certainly ubiquitous. VOCs are contained in many manufactured products, including paints, adhesives, gasoline, and plastics (Nikolaou et al., 2002). Therefore, when these products are not stored, used and disposed off with care, VOCs can contaminate our source of drinking water. But apart from anthropogenic emission, VOCs can gain or lost during the water treatment and distribution process. Trihalomethanes (THMs) can form during chlorination at the water treatment plant in the distribution system (Pauzi et al, 2002) and some VOCs concentrations can decrease during water treatment process such as aeration.

This study was done to identify contributing sources with regards to VOCs, and to estimate the pollution pattern in drinking water. For these purposes, we applied principal component analysis (PCA) together with varimax rotation, and parallel factor analysis (PARAFAC) to data sets obtained from one year sampling of drinking water in Peninsular Malaysia.

## Materials and methods

### *Sampling*

Sampling was performed on monthly basis for 12 months (June 2003 to May 2004), at water treatment plants throughout 11 states (Pahang, Selangor, Terengganu, Johor, Perak, Negeri Sembilan, Penang, Kelantan, Kedah, Melaka and Perlis) and Federal Territory (Wilayah Persekutuan) in Peninsular Malaysia. The sampling points for the analysis include treatment plant operation (TPO), service reservoir outlet (SRO) and auxiliary outlet point (AOP) at the water pipelines.

The samples were collected in 100 ml amber glass vials equipped with Teflon-faced septa. Water tap was opened and allowed the system to flush until the water temperature has stabilised. Five drops of 10% natrium thiosulphate ($Na_2S_2O_3$) was added to the sample bottle before sample bottle was filled to overflowing. No air bubbles should pass through the sample as the bottle was filled or be trapped in the sample when the bottle was sealed. The pH was adjusted to $< 2$ by carefully adding two drops of 1:1 acid hydrochloride (HCl) for each 100 ml sample volume. HCl was added to each water sample to prevent biodegradation and dehydrohalogenation. The sample bottles were sealed and kept at the temperature of 4°C.

### *Analytical methodology*

About 40 ml sample was poured into a 50 ml volumetric flask followed by 20 µl of internal calibration standard, fluorobenzene (FB) and surrogate standard, 4-bromofluorobenzene (BFB) at 200 mg/l and then top up to exactly 50 ml. The sample was transferred to a 40 ml glass vial with magnetic stirrer inside and capped with hole cap and Teflon-faced silicone rubber septum.

Solid phase micro-extraction (SPME) device was purchased from Supelco America (Bellefonte, PA) with approximately 7.5 cm of fibre (coated with a 100 µm thick stationary phase of polydimethylsiloxane) attached to a 15 cm of stainless steel needle that is inserted inside the plunger of a Hamilton syringe (Model, 7005) was used. The SPME fibre was previously conditioned (280°C for 30 min in a GC injector). The fibre was immersed into the sample for 5 min at ambient temperature under stirring condition. 5 min absorption time was found to be the optimum time to extract all VOCs in water sample. After sampling time, the SPME fibre was withdrawn into the needle and removed from the sample matrix. The fibre was then immediately inserted into the heated gas chromatography (GC) injection port. A desorption time of 10 min was allowed for the analytes to be desorbed from the fibre and get introduced into the GC column for analyses. Although for non-polar volatile compounds, desorption is virtually complete in a few seconds, desorption should be continued for a few minutes to ascertain that no carryover occurs when a blank is inserted after a sample. A fused-silica capillary column CP SIL 8CB (30 m x 0.25 mm, $d_f = 0.25$ μm; Varian, Inc.) was used on a HP 5890 GC system equipped with mass spectrometry detector (MSD). The oven temperature program started from 35$^o$C holds for 4 min and increased to 160$^o$C at 6$^o$C/min, holding at 160$^o$C for 5 min. The total run time was 30 min. Injector temperature was set at 220$^o$C. The MSD acquisition was performed in scan mode from 35 to 260 a.m.u. and in time scheduled selected ion monitoring mode.

***Data selection***

The variable selection is vital because of the sparseness in the data sets. Only variables that were more than 60% occurrence in the samples were chosen for further analysis. As for the variables with very few values above the method detection limit (MDL), its concentration is set equal to half its MDL (Fharnham 2002). Table 1 shows the abbreviation and MDL used for each compounds.

***Chemometrics data analysis methodology***

PCA and PARAFAC method are used to analyse all data sets statistically. PCA is a method based on bilinear model and assumes a linear model to explain the observed data variance using a reduced number of components (Tabachnick and Fidell, 2001). The varimax rotation was used after PCA as it used the orthogonal factor space and transformations that yield predictions and recognisable factors (Malinowski, 1991).

Where else, PARAFAC is a multi-way method, which the data are arranged in a cube instead of a matrix as in standard multivariate data sets. This is a trilinear model provides unique decompositions and they are the natural extension of bilinear models.

**Results and discussion**

PCA was used to identify the source and pollution types in the drinking water. PCA method is a unique solution and the interpretation of variance is straightforward since scores and loadings are orthogonal. Therefore orthogonal rotation (varimax) was used after the extraction to maximise high correlations and minimize low ones. The goal of varimax rotation is to maximize the variance of factor loadings by making high loadings higher and low ones lower for each factor. From the Negeri Sembilan data sets, only 12 variables have frequency detection more than 60% where 4 of the variables were aromatics and 8 aliphatic VOCs. Figure 1 shows the loadings plot of Negeri Sembilan. It was found that a clear discrimination among the variables selected, which indicate clearly the type of pollution in Negeri Sembilan. Three major types of pollution are found as toluene, dibromochloromethane, chlorobenzene, 1,1,1,2-tetrachloroethane and 1,2-dibromo-3-chloropropane compounds are all belongs to the active components that are used in paint industry. This group is isolated in the negative region of PC 1 and positive region of PC 2. Where else on the positive region of both PC 1 and 2, it was the pollution from agriculture run-off mainly pesticides, insecticides and fumigants. These products used 1,3-dichloropropene, 1,3-dichloropropane, 1,2-dibromoethane, 1,2,3-trichloropropane, 1,3-dichlorobenzene and 1,4-dichlorobenzene as their active ingredients. The third type of pollution is expected to present in all drinking water samples as almost all water treatment plant in Malaysia used chlorination for water disinfection (Sukiman & Pauzi, 1993). Chloroform as the major compound of disinfection by product discriminate from the rest of the variables in the positive region of PC1 and negative PC2. It should be noted that PC1 and PC2 explained 85% of variance in the data sets.

TABLE 1 Abbreviation used for identification of each VOCs compound and its respective MDL.

| Abbreviation | Compound | MDL(μg/l) | Half the MDL (μg/l) |
|---|---|---|---|
| c1 | 1,1-Dichoroethylene | 0.924 | 0.462 |
| c2 | Methylene chloride | 1.122 | 0.561 |
| c4 | 1,1-Dichloroethane | 0.116 | 0.058 |
| c5 | 1,2-Dichloroethylene | 0.734 | 0.367 |
| c6 | 2,2-Dichloropropane | 0.224 | 0.112 |
| c7 | Chloroform | 0.222 | 0.111 |
| c8 | Bromochloromethane | 0.236 | 0.118 |
| c9 | 1,1,1-Trichloroethane | 0.164 | 0.082 |
| c10 | 1,1-Dichloropropene | 0.108 | 0.054 |
| c11 | 1,2-Dichloroethane | 0.036 | 0.018 |
| c12 | Benzene | 0.038 | 0.019 |
| c13 | Carbon tetrachloride | 0.024 | 0.012 |
| c14 | 1,2-Dichloroprapane | 0.214 | 0.107 |
| c15 | Trichloroethylene | 0.114 | 0.057 |
| c16 | Dibromomethane | 0.218 | 0.109 |
| c17 | Dichlorobromomethane | 0.206 | 0.103 |
| c19 | 1,3-Dichloropropene | 0.036 | 0.018 |
| c20 | Toluene | 0.038 | 0.019 |
| c21 | 1,1,2-Trichloroethane | 0.096 | 0.048 |
| c22 | 1,3-Dichloropropane | 0.184 | 0.092 |
| c23 | Dibromochloromethane | 0.162 | 0.081 |
| c24 | 1,2-Dibromoethane | 0.096 | 0.048 |
| c25 | Tetrachloroethylene | 0.184 | 0.092 |
| c26 | Chlorobenzene | 0.168 | 0.084 |
| c27 | 1,1,1,2-Tetrachloroethane | 0.166 | 0.083 |
| c28 | Ethylbenzene | 0.038 | 0.019 |
| c30 | Xylenes | 0.058 | 0.029 |
| c31 | Bromoform | 0.116 | 0.058 |
| c32 | Styrene | 0.09 | 0.045 |
| c33 | 1,1,2,2-Tetrachloroethane | 0.012 | 0.006 |
| c34 | 1,2,3-Trichloropropane | 0.126 | 0.063 |
| c35 | Isopropylbenzene | 0.006 | 0.003 |
| c36 | Butylbenzene | 0.088 | 0.044 |
| c40 | 1,3,5-Trimethylbenzene | 0.07 | 0.035 |
| c42 | 1,2,4-Trimethylbenzene | 0.064 | 0.032 |
| c43 | 1,3-Dichlorobenzene | 0.144 | 0.072 |
| c44 | *sec*-Bromobenzene | 0.136 | 0.068 |
| c45 | 1,4-Dichlorobenzene | 0.036 | 0.018 |
| c46 | *p*-Isopropylbenzene | 0.086 | 0.043 |
| c47 | 1,2-Dichlorobenzene | 0.026 | 0.013 |
| c49 | 1,2-Dibromo-3-chloropropane | 0.166 | 0.083 |
| c50 | 1,2,4-Trichlorobenzene | 0.084 | 0.042 |
| c52 | 1,2,3-Trichlorobenzene | 0.062 | 0.031 |

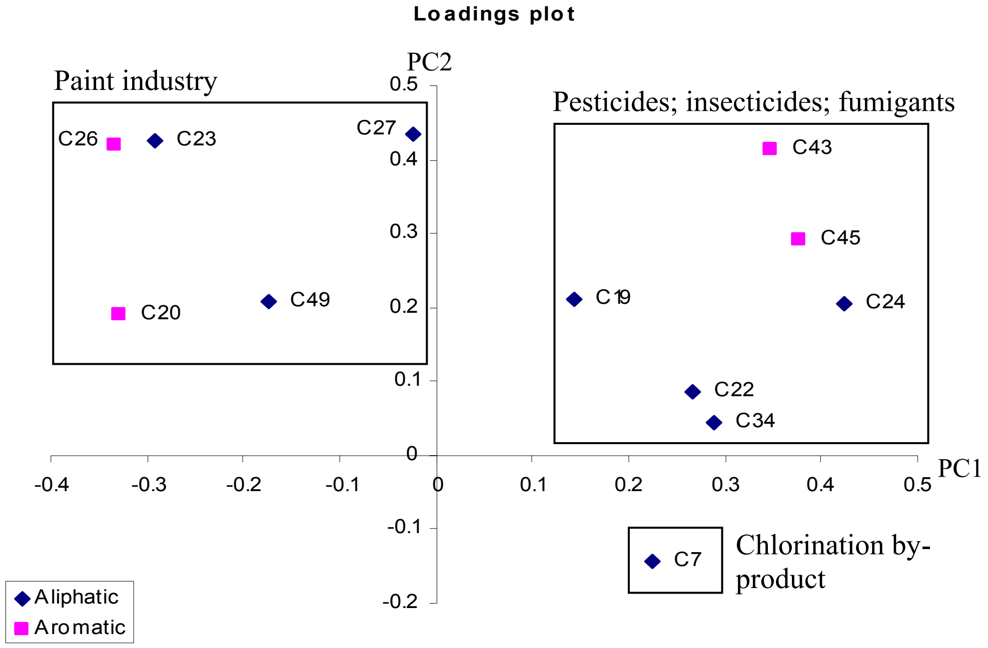

FIGURE 1 Loadings plot of Negeri Sembilan, which shows the type of pollutions found in drinking water.

As for PARAFAC analysis, from the Core Consistency Diagnostic (CORCONDIA), it was found that the number of significant factors is two, as shown in Table 2. The CORCONDIA ideally should be 100% therefore a value of between 80 to 100% means that the model is valid, while a value below 40% means that the model is not valid. The CORCONDIA will mostly decrease with the increased in the number of components but decrease very sharply where the correct number of components is exceeded. Hence, the appropriate number of components is the model with the highest number of components and a valid CORCONDIA. Since PARAFAC models are not nested, one may not fit models by augmenting an existing model by adding an extra component. For that reason, the PARAFAC analysis was repeated by using only two factors. Figure 2 shows the plot of variables for two factors. From this plot, a very clear discrimination was found. There are two groups showing three major species of pollution i.e. pesticides together with insecticides, paint industry and disinfection by-products.

Apart from disinfection by-products found in drinking water, both analysis method PCA with varimax rotation and PARAFAC results agreed that three major type of pollutions present in Negeri Sembilan drinking water, i.e. paint industry, agricultural run-off and disinfection by-products. The same approach is conducted to the rest of the data sets for other states in Peninsular Malaysia. Table 3 summarised the overall analysis results.

TABLE 2 The Core Consistency Diagnostic (CORCONDIA)

| No. of Factors | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2 | **100** | **99.98** | **81.44** | **59.40** | **98.34** | **99.36** | **60.50** | **99.11** | **25.42** | **98.90** |
| 3 | 14.42 | 48.90 | 56.09 | 15.53 | 49.51 | 33.80 | 62.18 | 62.36 | 85.81 | 75.13 |

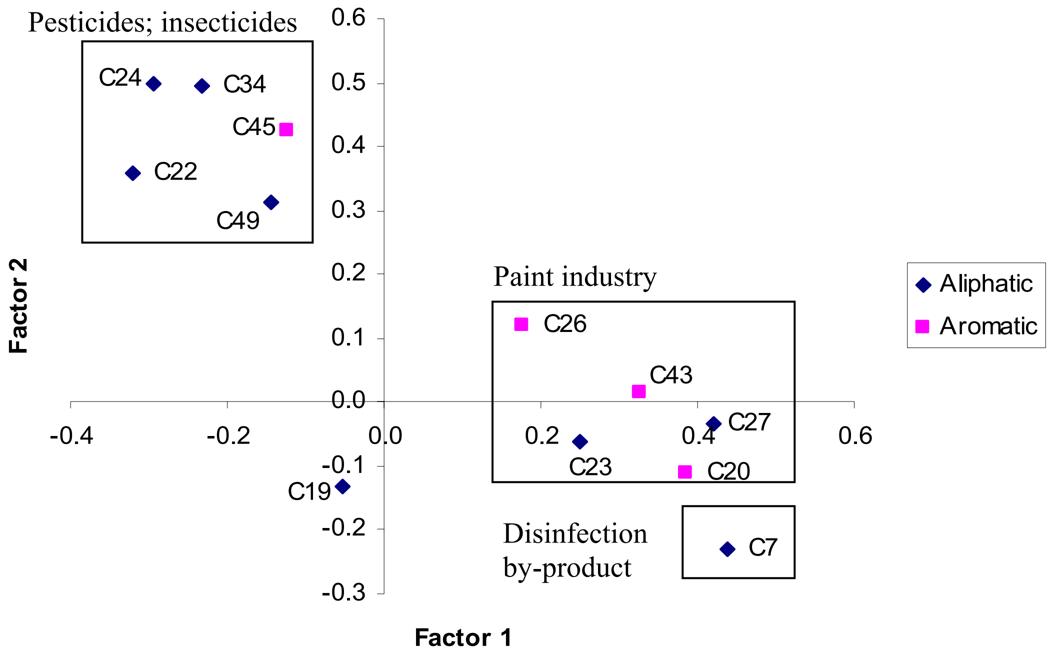

FIGURE 2 PARAFAC analyses for Negeri Sembilan dataset that shows three major types of pollution.

The results showed agreed and coherent with the degree of development in each state. It shows that the number of significant compounds detected increased with the extent of infrastructure growth in the states. It should be noted that, the effect of urbanisation and the resultant increase in population and activities associated with urban life normally change the quality of runoff within a catchment and it's receiving waters (Cunningham & Saigo, 1999). The problem of water pollution is now becoming more serious with reports indicating a downward trend of about 1% per annum in river water quality (Malaysia 2001). If this trend persists, then the costs for the river rehabilitation and water treatment facilities for water supply will require higher investments. As Malaysia is blessed with a unique situation of having enough resources annually, but this should not be taken for granted. In order to have a security of water supply with good quality, the consumers on a whole should also aware of water conservation.

TABLE 3 Summary of number of significant compounds and anthropogenic emissions

| State | No. significant compounds | Anthropogenic emissions (exclude disinfection by-products) |
|---|---|---|
| Kedah | 8 | Pesticides |
| Perak | 4 | Pesticides, insecticides, paint industry |
| Penang | 11 | Manufacture of plastics, pharmaceutical industries, adhesive solvent |
| Negeri Sembilan | 12 | Pesticides, paint industry |
| Selangor | 6 | Manufacture solvents, pharmaceutical products, pesticides, cleaning agents for textiles |
| Melaka | 9 | Manufactures for oils, fats, resins and chlorinated rubber, pesticides |
| Johor | 12 | Manufacture of plastics, resins, productions of dyes, insecticides, solvents for paints |
| Pahang | 11 | Pesticides, solvents for chlorinated rubber, paint industry |
| Kelantan | 6 | Pesticides, dry cleaning agent |
| Terengganu | 11 | As scavenger for lead in gasoline, manufacture paints, rubber, pesticides |

## Acknowledgements

The authors gratefully acknowledged Universiti Kebangsaan Malaysia for the laboratory facilities. The financial support form Malaysian Government under R&D grant, IRPA program 08-02-02-0009EA179, Department of Chemistry and Ministry of Health is gratefully acknowledged. Ministry of Sciences, Technology and Innovation is also acknowledged for financial support to SC Soh.

## References

Cunningham, W.P. and Saigo, B.W. (1999). *Environmental science: A global concern.* Boston: McGraw Hill International Book Company.

Fharnham, I.M., Singh, A.K., Stetzenbach, K.J. and Lohannesson, K.H. (2002). Treatment of non-detects in multivariate analysis of groundwater geochemistry data. *Chemom. Intell. Lab Syst.* 60: 265-281.

Malaysia. (2001). *Malaysia environmental quality report 2001.* KL, Malaysia: Ministry of Sciences, Technology and the Environment.

Malinowski, E.R. (1991). *Factor analysis in chemistry*. New York, USA: Wiley-Interscience Publication.

Nikolaou, A.D., Golfinopoulos, S.K., Kostopoulou, M.N., Kolokythas, G.A. and Lekkas, T.D. (2002). Determination of volatile organic compounds in surface and treated waste water in Greece. *Wat. Res,* 36: 2883-2890.

Pauzi, A., Yew, C.H, Mohamad Salleh, R. and Rahmah, A. (2002). Validation of solid phase microextraction technique for analysis of Trihalomethanes (THMs) in Malaysian drinking water. *Malaysian Journal of Chemistry*. 4: 27-34.

Sacks, R. and Akard, M. (1994). High-speed GC analysis of VOCs: Tuneable selectivity and column selection. *Environ Sci Technol.* 28: 428A-433A.

Sukiman, S. and Pauzi, A. (1993). *Chemical quality of Malaysian drinking water sources. Drinking water Quality: Microbiological and Public Health Aspects.* Bangi, Selangor: Universiti Kebangsaan Malaysia.

Tabachnick, B.G. and Fidell, L.S. (2001). *Using multivariate statistics.* Needham Heights, USA: Allyn & Bacon, Inc.