1352–2310(95)00261–8

RECEPTOR MODELING OF VOCs, CO, NO$_x$, AND THC IN TAIPEI

CHANG-CHUAN CHAN and CHIU-KUEI NIEN

National Taiwan University, College of Public Health, Institute of Occupational Medicine and Industrial Hygiene, Rm. 1447, No. 1, Jen-Ai Rd., 1st Sec., Taipei, Taiwan

and

JING-SHIANG HWANG

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan

Abstract—An empirical receptor model based on Markov Chain Monte Carlo simulation was applied to one-year measurements of eight VOCs, CO, NO$_x$, and THC collected in Taipei during 1993. Ambient monitoring data were measured at four monitoring stations in Taipei metropolitan. Among five VOC-based sources (motorcycles, catalyst passenger cars, non-catalyst passenger cars, diesel vehicles, and gasoline vapor), non-catalyst passenger cars had the greatest contributions to eight VOCs (53–61%; 90.0–220.3 μg m^{-3}). Among seven sources based on CO–NO$_x$–THC emissions (catalyst and non-catalyst two-stroke motorcycles, four-stroke motorcycles, catalyst and non-catalyst passenger cars, diesel vehicles, and gasoline vapor), passenger cars had the greatest contributions to NO$_x$ (50–63%; 0.05–0.26 mg m^{-3}), motorcycles had the greatest contributions to CO (70–76%; 0.81–4.97 mg m^{-3}) and gasoline vapor contributed substantially to THC (17–58%; 0.35–0.85 mg m^{-3}). Our empirical receptor models have successfully improved the estimation of source coefficients for VOCs, CO, NO$_x$, and THC by partially solving the collinearity problems among various mobile source profiles. Such an improved methodology is useful for validating source inventory and managing air quality in metropolitan areas.

Key word index: Receptor model, Monte Carlo simulation, VOC, mobile sources, motorcycles.

INTRODUCTION

The receptor modeling method has been applied to apportioning sources of particulate air pollutants successfully in several studies in the past (Miller *et al.*, 1972; Kowalczyk *et al.*, 1982; Dzubay *et al.*, 1988). However, the receptor models based on chemical mass balance (CMB) were recently used to identify sources of non-methane hydrocarbons (NMHC) in Tokyo (Wadden *et al.*, 1986), Newark and Linden (Scheff and Klevs, 1987), Chicago (Scheff and Wadden, 1993), and Detroit (Chung *et al.*, 1994). Until now, no studies of receptor modeling on CO and NO$_x$ were reported. Such a time lag in the application of receptor modeling to gaseous air pollutants can be explained by several inherent constraints in CMB models. First, until recently, extensive field measurements of speciated volatile organic compounds (VOCs), such as benzene, toluene, and xylenes were not carried out. Formerly, all hydrocarbons were measured as a whole without speciation, and termed as total hydrocarbons (THC). Second, valid source profiles based on various gaseous air pollutants have

not been established yet. Third, there seemed to be a higher collinearity among various source profiles based on gaseous species. Last, the reactivity of gaseous air pollutants may need to be taken into account in the models. Even though there is room for improvement for receptor models on gaseous air pollutants, we perceived it as a useful tool in helping to set air pollution control strategies. In Taipei, several studies have shown that mobile-related air pollutants, such as CO and VOC, are associated with the deterioration of air quality (Chan *et al.*, 1993, 1994; Liu *et al.*, 1994). The emission inventory also indicates that mobile sources, including motorcycles, gasoline cars and diesel buses and trucks, may account for about 90% of CO, 60% of NO$_x$, and 90% of THC emissions in Taipei (Taiwan EPA, 1993). These pollutants' emission contributions from different vehicle categories, however, are not well documented because the emission factors and vehicle mileage of motorcycles and cars are unavailable for mobile source modeling in Taipei. In order to have a comprehensive control strategy based on the priorities of emissions, the estimation of emissions from various mobile sources
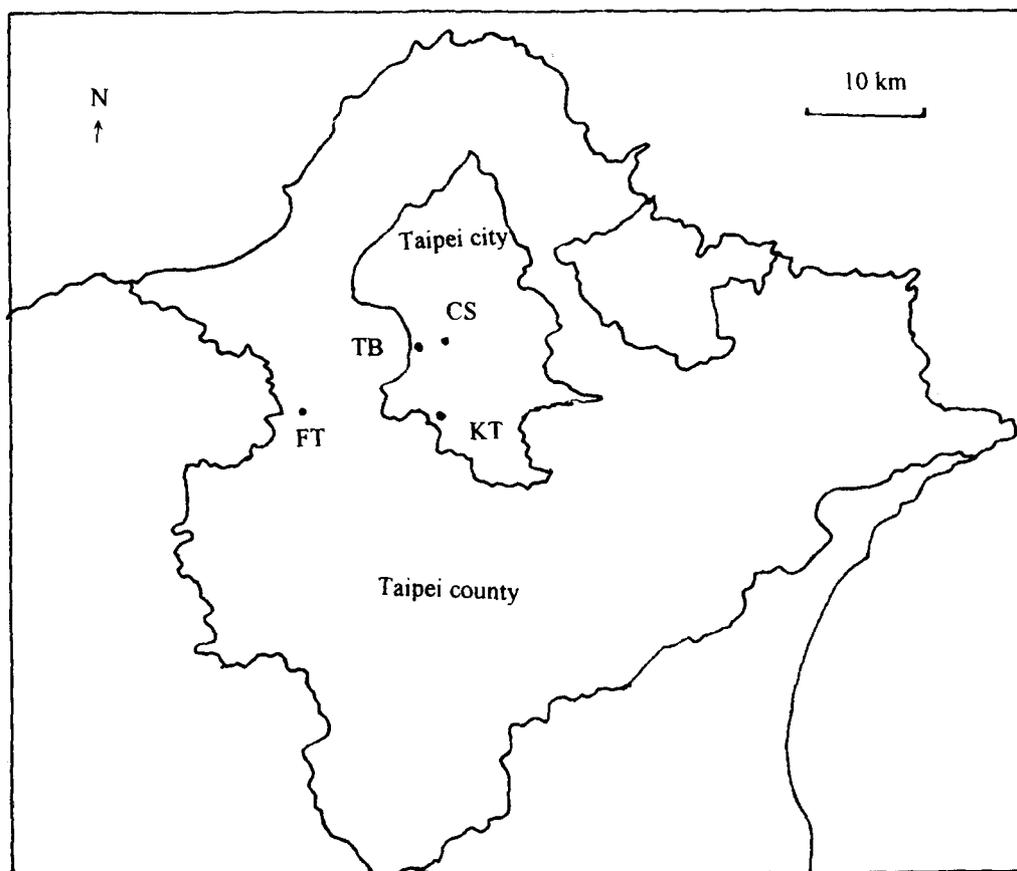
Fig. 1. The locations of monitoring stations in Taipei metropolitan area.

in Taipei is needed. This study, therefore, applies a modified receptor model to apportioning gaseous air pollutants from various mobile sources in metropolitan areas of Taipei.

METHODS

Sampling locations

The ambient concentrations of VOCs, CO, NO_x, and THC are collected at four fixed-site monitoring stations in Taipei metropolitan area. One of them, the Taipei-Bridge station (TB), is located at the intersection of two main traffic routes in downtown Taipei. The TB site is designated as a "traffic monitoring station" by Taiwan EPA. The height of sampling inlet is around 3 m. The other three stations, the Chung-San (CS), the Fu-Ta (FT), and the Ku-Ting (KT), are all located in campus of respective schools (see Fig. 1). They are all designated as "ambient air quality monitoring stations" by Taiwan EPA since their locations are not closely adjacent to any main traffic routes. The heights of the sampling inlets at these three stations are all around 12 m. One particular feature of these stations is that there is a big gasoline service station near the FT station.

Sampling schedule

One-year measurements of air pollutants in 1993 are used for modeling. The sampling frequency of VOCs is one sampling day per week for the first seven months and one samp-

ling day per month for the last three months. On each sampling day, two 12-hour samples (7:00 a.m. to 7:00 p.m. and 7:00 p.m. to 7:00 a.m.) are collected simultaneously at four stations. For NO_x, CO and THC, hourly concentrations measured continuously at four monitoring stations in 1993 are used in receptor modeling calculation.

Sampling and analytical methods

The U.S., EPA TO 1 method is modified to collect and analyze the VOCs. The sampling system consists of a stainless steel tube (178 mm × 5.2 mm ID) with 0.6 g Tenax-GC, and a low-flow sampling pump (Gilian Inc., Model LFS 113D). The sampling rate is adjusted at 50 ml min^{-1} for a duration of 12 h. The VOCs collected in the Tenax-GC are first thermally desorbed and then analyzed by a high resolution gas chromatography-mass spectrometry (GC/MS) technique. Such a measuring method can quantify a total of eight VOCs, including benzene, heptane, heptene, toluene, ethylbenzene, m/p-xylenes, o-xylene, and isopropyl benzene. The details of sampling and analytical conditions and QA/QC procedures have been reported in a previous study (Chan et al., 1993). Ambient concentrations of NO_x, CO, and THC are continuously measured by standard instruments for ambient air monitoring stations, which are a chemiluminescence method for NO_x, a non-disperse infra-red (NDIR) method for CO, and a flame ionization detection (FID) method for THC.

Empirical receptor model

Conventionally, the mathematical receptor models, such as chemical mass balance models (CMB), have a general

form like this:

$$Y_i = X\beta, \quad i = 1, 2, \ldots, n \tag{1}$$

where Y_i denotes the ambient concentration vector of a set of m pollutants, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is a p-vector of total mass concentrations contributed by p individual sources, and X is the $m \times p$ matrix of source profiles. It is well recognized that solutions to this equation usually suffer by the problems of collinearity among source profiles and the negative β.

Several mathematical techniques, such as non-negative least square routines, have also been successfully demonstrated to solve such problems (Wang and Hopke, 1989). In this paper, we propose to apply the Bayesian statistical approaches to solve these problems. In doing so, we treat β as a vector of random variables in our model, rather than solving β directly from equation (1). Therefore, the feature of β can be better described through statistical modeling of the modified CMB equation by the following form:

$$g(Yi) = g(X\beta) + \varepsilon$$

$$\varepsilon \sim N(0, W)$$

$$W = \begin{bmatrix} \delta_1^2 & & 0 \\ & \ddots & \\ 0 & & \delta_m^2 \end{bmatrix} \tag{2}$$

where Y_i, X, and β have the same definitions as in equation (1). To simplify the modeling process, we need first to transform the raw data by a function, g, in order to assure that the vector of errors, ε, is normally distributed. Accordingly, the choice of g is dependent on the characteristics of raw data. Besides that, we must always restrict β to be non-negative in our model because sources cannot produce negative contributions to ambient measurements physically.

Given n measurements of Y_i in the model, the calculation of equation (2) becomes a problem of computing the posterior distributions of β and δ_1 through δ_m in a Bayesian framework. The posterior distribution is, therefore, directly proportional to the product of the likelihood function, $\ell(y \mid \beta, \delta_1 \ldots \delta_m)$, and a prior density, $p(\beta, \delta_1 \ldots \delta_m)$, which can be expressed by the following equation:

$$\Pi(\beta, \delta_1 \ldots \delta_m \mid y) \propto \ell(y \mid \beta, \delta_1 \ldots \delta_m) \cdot p(\beta, \delta_1 \ldots \delta_m) \tag{3}$$

where

$$\ell(\beta, \delta_1 \ldots \delta_m) \propto |W|^{(n/2)} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} [g(Y_i) - g(X\beta)]' \right.$$

$$\left. \times W^{-1}[g(Y_i) - g(X\beta)] \right\} \cdot I_{(0,\infty)}(\beta).$$

In our model, the prior, $p(\beta, \delta_1 \ldots \delta_m)$, is assumed to be non-informative. Under such an assumption, one simplified form of the non-informative prior, $p(\beta, \delta_1 \ldots \delta_m)$, is chosen to be proportional to $\delta_1^{-1} \times \cdots \times \delta_m^{-1}$.

Equation (3) is expressed as a proportional form only because it is difficult to compute the normalization constant of the higher dimensional posterior distribution needed for an exact form either by traditional analytic or numerical methods in the equation. From a Bayesian approach, we can estimate features of the posterior indirectly by using the Markov Chain Monte Carlo (MCMC) method to draw samples from the product of likelihood function and the prior. Due to its conceptual simplicity in simulation and the availability of calculation algorithms, such as Gibbs Sampler and Metropolis algorithm, the MCMC method has been widely applied to several statistics-related fields, such as statistical physics, image process, and medicine (Smith and Roberts, 1993).

In our application, the MCMC method obtains chains of values as samples to mimic the posterior distributions, which

are generated from the Gibbs Sampler and Metropolis algorithm. To perform the MCMC simulation, a starting value is first assigned to the algorithms. After tens or hundreds of iterations, the chains will start to converge to the posterior. The chains of values are then taken as samples once the chains converge to the posterior. The details of this algorithm are described in the Appendix A.

We used all available measurements from the sampling days in our model fitting, which were about 38–53 d for VOCs and 329–363 d for CO, NO$_x$ and THC at the four monitoring stations. As expected, the raw data of these measurements were not normally distributed. Therefore, we selected log as our g function in equation (3) to transform the raw data into approximate normal distribution. We then assigned 5000 iterations in our simulation in order to generate enough samples for the description of β. However, we found that the simulated chains converged very quickly after 100 iterations. Instead of using all generated samples in our model fitting, we selected only 500 samples systematically from 4900 simulated β values, $(\beta^{101}, \beta^{102}, \ldots, \beta^{5000})$ for further calculation. Accordingly, we can estimate the distribution of source contributions, β, from these 500 samples for individual air pollutants at each monitoring station.

We used three criteria to check the model's internal consistency in our fitting approach. First, we proposed to use the averaging r-square, $\overline{r^2}$, in the following equation to estimate overall goodness-of-fitness of the empirical model.

$$\overline{r^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\hat{Y}_i - \bar{y}_i \cdot \mathbf{1})'(\hat{Y}_i - \bar{y}_i \cdot \mathbf{1})}{(Y_i - \bar{y}_i \cdot \mathbf{1})'(Y_i - \bar{y}_1 \cdot \mathbf{1})}$$

where $\hat{Y}_i = X\bar{\beta}$, $\bar{\beta}$ is the average of these 500 samples of β, $\bar{y}_i$ is the average of m measured concentrations of Y_i, and $\mathbf{1}$ is an $m \times 1$ vector of ones. The model is said to be better fitted if the $\overline{r^2}$ is greater. Second, the proportion of model-calculated mass in the actual measurements, % mass, should be close to 100%. A model is considered to be ill-fitted if the deviation of % mass is too large. Usually, 25% overestimation or underestimation is considered to be unacceptable. Third, the variation in the simulated source coefficients should be within the acceptable range. A smaller variation indicates that a better model is obtained. We used the inter-quartile range (IQR) of source coefficients, which is the range between 75th and 25th percentiles of the samples, as the index of such variation in the model.

Source profile development

As shown in Table 1, 12 source profiles based on VOCs, NO$_x$, CO, and THC are constructed for tailpipe emissions of gasoline cars and motorcycles, diesel vehicles, and gasoline vapors. Five VOC-based source profiles are calculated from normalizing weight concentrations of 8 VOC species in emissions to 100%. We reduced three different motorcycles into one source profile because their VOC-based source profiles were very similar. We constructed another seven source profiles for NO$_x$, CO, and THC by normalizing their weight concentrations in emissions to 100%. For NO$_x$, CO, and THC, the source profiles are significantly different between two-stroke and four-stroke motorcycles. The tailpipe emissions of VOCs, NO$_x$, CO, and THC for gasoline cars and motorcycles are obtained from a study on vehicle emissions tests by chassis dynamometer under standard driving cycles, conducted in Taiwan (Chan et al., 1994). The tailpipe emissions of VOCs for diesel vehicles reported by Scheff et al. and the emissions of NO$_x$, CO, and THC reported by Palmer are used as surrogate emissions for diesel vehicles in Taiwan since the data of diesel emissions are currently unavailable in Taiwan (Scheff et al., 1992; Palmer, 1993). Because the emissions of speciated VOCs from gasoline vapor are unavailable in Taiwan, the source profiles reported by Sigsby et al. are adopted in the current study (Sigsby et al., 1987).

Table 1. Twelve source profiles normalized to eight VOCs and CO–NO$_x$–THC (wt%)

Compounds	Motorcycles			Car_NC	Car_C	Diesel	Gasoline vapor
Benzene	20.3			18.4	21.2	22.1	4.9
Heptene	2.9			2	2.8	0	0
Heptane	9.5			9.8	13.7	5.1	3.8
Toluene	34.4			32.4	36.7	29.1	62.8
Ethylbenzene	10.1			8.9	6.7	4.9	3.9
m/p-Xylene	12.1			20.3	13.2	6.9	15.6
o-Xylene	9.3			7.7	5.2	17.7	6.4
Isopropyl benzene	1.4			0.5	0.5	14.3	2.7
	2S_NC	2S_C	4S_NC	Car_NC	Car_C	Diesel	Gasoline vapor
CO	58	66	82	79	69	26	0
NO$_x$	0	0	4	12	18	14	0
THC	41	34	14	9	13	59	100

2S_NC: Two-stroke motorcycle without catalytic converter. 2S_C: Two stroke motorcycle with catalytic converter.

4S_NC: Four-stroke motorcycle without catalytic converter. Car_NC: passenger car without catalytic converter.

Car_C: passenger car with catalytic converter.

RESULTS AND DISCUSSION

Ambient measurements of air pollutants

The average ambient concentrations of 8 VOCs, CO, NO$_x$, and THC over the sampling period in four monitoring stations are summarized in Table 2. Mean concentrations are 67.1 μg m^{-3} for benzene, 129.6 μg m^{-3} for toluene, 6.52 mg m^{-3} for CO, 0.43 mg m^{-3} for NO$_x$, and 2.59 mg m^{-3} for THC at the TB station. The mean concentrations are about 17.3–23.2 μg m^{-3} for benzene, 58.7–85.2 μg m^{-3} for toluene, 1.15–1.78 mg m^{-3} for CO, 0.08–0.14 mg m^{-3} for NO$_x$, and 1.33–1.47 mg m^{-3} for THC at three ambient sampling stations (FT, CS, and KT). Apparently, the concentrations of all air pollutants are the highest at the traffic monitoring station (TB). The ambient measurements, however, show no significant differences among three ambient air monitoring stations. The ratios of ambient concentrations between the traffic station and the ambient station are about 1.5–3.9 for 8 VOCs, 3.7–5.7 for CO, 3.1–5.4 for NO$_x$, and 1.8–2.0 for THC. Obviously, mobile sources seem to contribute a large amount of concentrations of these pollutants measured at the traffic station.

Source coefficients for 8 VOCs

The five emission sources considered in modeling 8 VOCs are motorcycles, catalyst passenger cars and non-catalyst passenger cars, diesel vehicles, and gasoline vapor. The mean and IQR of the estimated source coefficients, and the simulation model's %mass and $\bar{r}^2$ are shown in Table 3. Overall, the model's internal consistency is reasonably good because $\bar{r}^2$ values, 0.70–0.75, are moderately high and %mass, 73–89%, are close to 1 for four monitoring stations. The variation of source coefficients tends to vary with the absolute values of estimated source coefficients. As

source coefficients become smaller, their relative variations become larger. For example, the IQR/mean ratios are always less than 1 for the non-catalyst passenger cars, which is always the greatest contribution source at all monitoring stations. In contrast, for the second largest contribution source, the motorcycles, and other minor sources, their IQR/mean ratios are either very close to 1 or greater than 1.

Non-catalyst passenger cars are the greatest contribution sources to 8 VOCs among five mobile sources. They account for 53–61% mass concentrations of 8 VOCs measured at four monitoring stations, which are equal to 92.0–220.3 μg m^{-3}. In contrast, catalyst passenger cars have relatively small contributions to 8 VOCs. The difference in source contributions between these two car types is generally in agreement with the difference in vehicle numbers and emission factors between non-catalyst and catalyst cars in Taipei. It is estimated that less than 20% passenger cars are currently equipped with catalysts in Taipei because catalytic converters are only required for the 1991 model of domestically made and imported passenger cars. Moreover, the emission factors of eight VOCs for the two car types are quite different, differing by about a factor of two. Motorcycles are the second highest contribution sources for eight VOCs at the TB station, which contribute 30.8 μg m^{-3} to total VOC mass concentrations. The high contributions from motorcycles at TB station are believed to be due to a relatively higher motorcycle traffic volume near the station where specific traffic lanes are designated for motorcycles. The contributions of gasoline vapor to eight VOCs are relatively higher at three ambient monitoring stations than the traffic station. This indicates that the importance of fugitive emissions increases as the contributions of tail-pipe emissions decrease. It is also understandable that diesel vehicles have the least contributions to eight VOCs in all

Table 2. Yearly averaging concentrations of air pollutants at four receptor locations in Taipei in 1993

Pollutants	Unit	CS			KT			FT			TB		
		Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
Benzene	μg m⁻³	17.42	14.63	39	23.18	26.57	44	17.32	12.49	54	67.07	35.74	47
Heptene	μg m⁻³	6.26	10.40	38	5.37	5.62	44	6.43	8.96	54	19.43	19.49	47
Heptane	μg m⁻³	10.76	12.42	39	12.68	21.30	42	13.68	16.13	54	29.21	21.10	46
Toluene	μg m⁻³	58.89	32.34	39	58.65	35.10	44	85.24	51.09	54	129.64	65.45	47
Ethyl benzene	μg m⁻³	16.14	12.04	39	13.55	13.88	44	15.82	9.59	54	39.62	18.28	47
m/p-Xylene	μg m⁻³	29.82	23.75	39	28.12	20.43	44	32.90	22.67	53	88.73	71.07	47
o-Xylene	μg m⁻³	14.89	10.67	38	14.55	12.92	44	14.31	9.04	53	43.10	27.62	47
Isopropyl benzene	μg m⁻³	1.07	1.01	38	1.39	2.30	44	1.36	1.41	53	3.13	3.94	47
CO	mg m⁻³	1.78	0.50	363	1.77	0.54	329	1.15	0.35	347	6.52	1.28	353
NO$_x$	mg m⁻³	0.14	0.04	365	0.12	0.03	358	0.08	0.03	339	0.43	0.12	360
THC	mg m⁻³	1.39	0.20	362	1.33	0.29	352	1.47	0.47	336	2.59	0.35	362

Table 3. CMB source coefficients (μg m⁻³), %mass, and averaging r^2 for modeling eight VOCs at four receptor sites

Source	Site											
	CS			KT			FT			TB		
	Mean	IQR	%mass	Mean	IQR	%mass	Mean	IQR	%mass	Mean	IQR	%mass
Motorcycle	15.3	11.0	9.9	19.5	13.4	12.4	12.0	10.9	6.4	129.1	12.8	30.8
Catalyst passenger car	5.2	5.5	3.4	5.2	5.4	3.3	5.5	5.6	2.9	17.2	26.1	4.1
Non-catalyst passenger car	94.6	15.4	60.9	92.0	15.6	58.5	103.8	16.8	55.5	220.3	33.3	52.5
Diesel	0.8	0.5	0.5	0.8	0.8	0.5	0.6	0.6	0.3	1.5	1.7	0.4
Gasoline vapor	13.7	9.4	8.9	7.5	7.6	4.8	14.7%	8.8	7.9	6.9	8.4	1.6
Sum of source contributions	129.3		83.3	125.0		79.4	136.5		73.0	375.0		89.3
Averaging r^2			0.70			0.74			0.72			0.75

Note: IQR (inter-quartile range): distance between 25th and 75th percentile.

stations because the diesel engine's hydrocarbons emissions consist of higher carbon numbers than the eight VOCs in this study.

Source coefficients for CO, NO_x, *and THC*

The source coefficients of CO, NO_x, and THC for seven different mobile sources are shown in Table 4. Apparently, the overall performance of modeling CO, NO_x, and THC is even better that of VOCs. The $\overline{r^2}$ values, 0.77–0.98, are high and the %mass is close to 1. The %masses are about 97–100% for CO, 100–102% for NO_x, and 93–101% for THC. The results of IQR/mean show that variation in modeling CO, NO_x, and THC is comparable to modeling VOCs.

Most of the CO concentrations measured at four monitoring stations are contributed by three motorcycle types and the non-catalyst cars. In total, they account for about 70% of CO's total mass concentrations measured at four monitoring stations. The passenger cars account for another 20% of CO concentrations, while the diesel emissions explain the other 1–5%. The differences in CO contributions among these three vehicle categories simply reflect the differences in their CO emission factors. On average, motorcycles' CO emission factors are about three times higher than passenger cars' (Chan et al., 1994). Comparing source contributions among motorcycles alone, we found that 2-stroke motorcycles' CO contributions are the greatest. At three ambient monitoring stations, two-stroke motorcycles' contributions are about 2.5–3 times that of four-stroke motorcycles. The differences in CO contributions between two-stroke and four-stroke motorcycles are generally in agreement with their differences in CO emission rates. As reported by Chan et al., a two-stroke motorcycle's CO emission rates are about twice that of a stroke motorcycle (Chan et al., 1995). However, we found equal CO contributions from two-stroke and four-stroke motorcycles at the TB station, whose contributions are 39.4 and 39.9%, respectively. We believe this may simply reflect differences in traffic volumes between two-stroke and four-stroke motorcycles near the TB station. The mean source contributions of CO in terms of mass concentrations are 0.59–2.37 $mg\,m^{-3}$ from two-stroke motorcycles, 0.22–2.60 $mg\,m^{-3}$ from four-stroke motorcycles and 0.26–1.46 $mg\,m^{-3}$ from passenger cars.

For NO_x, passenger cars are the most important contributing sources, which account for 50–62% of NO_x concentrations measured at four stations. The second important sources for NO_x are diesel vehicles at three ambient monitoring stations (25–42%) and four-stroke motorcycles at the TB station (30%). In contrast, two-stroke motorcycles have no significant contributions to NO_x. Again, such comparisons can be explained by differences in emission factors among vehicles and variation in traffic volumes among monitoring stations. It is known that two-stroke motorcycles are low NO_x emitters and diesel cars are

high emitters (Chan et al., 1995). The mean source contributions of NO_x in terms of mass concentrations are 0.05–0.27 $mg\,m^{-3}$ from passenger cars and 0.02–0.5 $mg\,m^{-3}$ from diesel cars.

For THC, gasoline vapor and two-stroke motorcycles are two of the most important contributing sources, which account for 17–58% and 25–53% of THC concentrations measured at four monitoring stations, respectively. Gasoline vapor becomes an important contributing source to THC because several light-chain hydrocarbons evaporate from the vehicular gasoline tanks easily in the subtropical weather of Taiwan. Two-stroke motorcycles are another important contributing source because of their high THC emission rates. On average, two-stroke motorcycles' THC emission rates are about 3–6 times higher than that of other vehicles (Chan et al., 1995). The source contributions of THC in terms of mass concentrations are 0.38–0.85 $mg\,m^{-3}$ from gasoline vapor and 0.17–1.39 $mg\,m^{-3}$ from two-stroke motorcycles.

Model comparisons and limitations

The results of apportioning CO, NO_x, or THC measurements to various motor vehicles have not been reported in the previous CMB receptor modeling studies. The lack of reliable source profiles, which are based on CO, NO_x, and THC emissions, is the main difficulty in modeling these pollutants. Although there are some CMB receptor modeling studies on VOCs, mobile sources are not further classified to various vehicle types in these studies. The high collinearity of source profiles among mobile sources is one major problem for traditional CMB receptor models in apportioning mobile sources. For VOC-based source profiles in this study, there is a high collinearity among two-stroke motorcycles, four-stroke motorcycles and passenger cars because their condition indexes (CI) are greater than 30 in the regression collinearity diagnosis. Accordingly, the CMB7 model can only partition VOCs into two categories, the gasoline vapor and the tailpipe emissions of all vehicles. Apparently, the calculation methodology applied in our empirical receptor model is more robust toward collinearity problems because it has successfully attributed contributions of VOCs, CO, NO_x and THC to 5–7 different mobile sources. However, the collinearity problems are still not fully solved by our approach because source coefficients' variations are still relatively large for minor contributing sources in our models.

Traditional CMB models also often encounter the problem of overestimating or underestimating total source contributions. This is not found to be a serious problem for our empirical receptor models. For CO, NO_x and THC, their ambient concentrations at four monitoring stations are very close to the total emitted %mass by seven sources in Taipei metropolis. The ambient concentrations of VOCs at most monitoring stations except the FT station are also well explained by five sources considered in Taipei. Such modeling

Table 4. CMB source coefficients (μg m^{-3}), %mass, and averaging r^2 for modeling CO, NO$_x$, THC at four receptor sites

							Site						
	CS			KT			FT			TB			
Source	Mean	IQR	%mass	Mean	IQR	%mass	Mean	IQR	%mass	Mean	IQR	%mass	
CO													
2S_NC	0.26	0.29	14.6	0.51	0.51	28.8	0.31	0.25	27.0	0.86	0.90	13.2	
2S_C	0.65	0.34	36.5	0.58	0.52	32.8	0.28	0.26	24.3	1.51	1.63	23.2	
4S_NC	0.36	0.38	20.2	0.22	0.24	12.4	0.22	0.25	19.1	2.60	2.64	39.9	
Car_NC	0.27	0.27	15.2	0.23	0.25	13.0	0.19	0.00	16.5	1.08	0.15	16.6	
Car_C	0.12	0.17	6.7	0.07	0.14	4.0	0.07	0.09	6.1	0.38	0.08	5.8	
Diesel	0.09	0.07	5.1	0.10	0.07	5.6	0.04	0.06	3.5	0.07	0.04	1.1	
Gasoline vapor	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	
Sum of seven source contributions	1.75		98.3	1.71		96.6	1.11		96.5	6.50		99.7	
Modeling error	0.29			0.33			0.32			0.21			
NO$_x$													
2S_NC	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	
2S_C	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	
4S_NC	0.02	0.02	14.3	0.01	0.01	8.3	0.01	0.01	12.5	0.13	0.13	30.2	
Car_NC	0.04	0.04	28.6	0.04	0.04	33.3	0.03	0.00	37.5	0.17	0.02	39.5	
Car_C	0.03	0.04	21.4	0.02	0.04	16.7	0.02	0.02	25.0	0.10	0.02	23.3	
Diesel	0.05	0.04	35.7	0.05	0.04	41.7	0.02	0.03	25.0	0.04	0.02	9.3	
Gasoline vapor	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00	0.0	
Sum of seven source contributions	1.14		100.0	1.12		100.0	0.08		100.0	0.44		102.3	
Modeling error	0.29			0.29			0.39			0.29			
THC													
2S_NC	0.18	0.21	12.9	0.36	0.37	27.1	0.22	0.18	15.0	0.61	0.64	23.6	
2S_C	0.33	0.17	23.7	0.30	0.27	22.6	0.15	0.13	10.2	0.78	0.84	30.1	
4S_NC	0.06	0.06	4.3	0.04	0.04	3.0	0.04	0.04	2.7	0.43	0.44	16.6	
Car_NC	0.03	0.03	2.2	0.03	0.03	2.3	0.02	0.00	1.4	0.12	0.02	4.6	
Car_C	0.02	0.03	1.4	0.01	0.03	0.8	0.01	0.02	0.7	0.07	0.01	2.7	
Diesel	0.19	0.15	13.7	0.23	0.16	17.3	0.08	0.12	5.4	0.17	0.09	6.6	
Gasoline vapor	0.58	0.25	41.7	0.35	0.29	26.3	0.85	0.89	57.8	0.43	0.15	16.6	
Sum of seven source contributions	1.39		100.0	1.32		99.2	1.37		93.2	2.61		100.8	
Modeling error	0.15			0.27			0.46			0.14			
Averaging r^2	0.77			0.96			0.80			0.98			

results are empirically plausible because there are relatively fewer industrial or fugitive emission sources compared to mobile sources in Taipei. The VOC measurements at the FT station are significantly underestimated because our empirical model can only explain 73% of total mass concentrations. This implies that additional contributions to VOCs may be from emissions other than mobile sources near the FT station. However, a more accurate and comprehensive emission inventory of these pollutants in Taipei is needed in order to have a precise validation of the modeling results.

It has also been reported that variation in the hydrocarbon's photochemical reactivity is a problem in CMB receptor modeling for VOCs (Aronian et al., 1989) and PAHs (Li and Kaments, 1993). The measurements of pollutants need to be modified by individual pollutant's reactivity in order to meet mass conservation requirement in CMB models. The reactivity problem is believed not to be a major problem in our model for the following three reasons. First, the reactivities of eight VOCs are about the same except heptene. The experimental results of these VOCs' reactivity to OH radicals in the laboratory show that their reactivities are in the same order according to the reports of Atkinson (1986). Second, the distances between the ambient measurements and the emission sources at station are relatively short. Therefore, variation of these air pollutants' half-lives may not have major effects on the modeling results. Third, the effect of meteorological variation on source patterns can also be neglected because one-year averaging measurements are used in our models. However, a more accurate and precise estimation of source contributions can still be expected by including the reactivity of VOCs and the meteorological variation in our empirical models.

The representativeness of source profiles is deemed to be the major limitation of our empirical model. The real-world emission factors of mobile sources can be very different from the emission factors obtained from the dynamometer because the driving cycles and vehicular ages in the standard testing procedures may not be representative. Therefore, the mean contributions estimated by our empirical receptor model may overestimate the contributions from "clean vehicles" but underestimate the contributions from "dirty vehicles".

CONCLUSION

An empirical receptor model based on the Markov Chain Monte Carlo simulation has been successfully applied to apportioning the contributions of the ambient measurements of eight VOCs, CO, NO_x, and THC to six various mobile sources and one fugitive source in Taipei metropolitan. Non-catalyst passenger cars are found to be a major source of VOCs and NO_x in Taipei. The other important source of NO_x is

the diesel cars. The motorcycles, particularly the two-stroke motorcycles, are found to be a major source of CO and THC in Taipei. The gasoline vapor is found to be another major source of THC in Taipei. The source contributions estimated by our empirical receptor models are generally in agreement with emission factors and number of various motor vehicles in Taipei metropolitan. Our empirical receptor models are also more robust to the issues of VOC's reactivity and source profile's collinearity than traditional CMB models. This study successfully demonstrates that empirical receptor models can be applied to apportioning organic and inorganic air pollutants to various mobile sources in the metropolitan areas where their air pollution problem are mainly traffic related. The proposed receptor models are believed to be very useful for validating inventory of various emission sources or designing effective management programs of air quality controls.

REFERENCES

Aronian P. F., Scheff P. A. and Wadden R. A. (1989) Wintertime source-reconciliation of ambient organics. *Atmospheric Environment* **27**, 911–920.

Atkinson R. (1986) Kinetics and mechanisms of gas phase reactions of the hydroxyl radical with organic compounds under atmospheric conditions. *Chem. Rev.* **86**, 69.

Chan C. C., Lin S. H. and Her G. R. (1993) Student's exposure to volatile organic compounds while commuting by motorcycle and bus in Taipei city. *J. Air Waste Man. Ass.* **43**, 1231–1238.

Chan C. C., Lin S. H. and Her G. R. (1994) Office worker's exposure to volatile organic compounds while commuting and working in Taipei city. *Atmospheric Environment* **28A**, 2351–2359.

Chan C. C., Nien C. K., Tsai C. Y. and Her G. R. (1995) Comparison of tail-pipe emissions from motorcycles and passenger cars. *J. Air Waste Man. Ass.* **45**, 116–124.

Chung J., Wadden R. A. and Scheff P. A. (1994) VOC receptor modeling applied to prediction of ambient ozone in Detroit. In *Proc. 87th Annual Meeting of the Air and Waste Management Association*, paper no. 94-TA30.02.

Dzubay T. G., Stevens R. K., Gorden G. E., Olmez I., Scheffield A. E. and Courtney W. J. (1988) A composite receptor method applied to Philadelphia aerosol. *Envir. Sci. Technol.* **22**, 46–52.

Kowalczyk G. S., Gordon G. E. and Rheingrover S. W. (1982) Identification of atmospheric particulate sources in Washington, D.C., using chemical element balances. *Envir. Sci. Technol.* **16**, 79–90.

Li C. K. and Kaments R. M. (1993) The use of polycyclic aromatic hydrocarbons as source signatures in receptor modeling. *Atmospheric Environment* **27A**, 523–532.

Liu J. J., Chan C. C. and Jeng F. T. (1994) Predicting personal exposure levels to carbon monoxide (CO) in Taipei, based on actual COD measurements in microenvironments and a Monte Carlo Simulation Method. *Atmospheric Environment* **28**, 2361–2368.

Merz W. K. and Biswas P. (1992) Receptor modeling of VOCs accounting for chemical reactivity. *Proc. 85th Annual Meeting of the Air and Waste Management Association*, paper no. 92-104.01.

Miller M. S., Friedlander S. K. and Hidy G. M. (1972) A chemical element balance for the Pasadena aerosol. *J. Colloid Interface Sci.* **39**, 165–176.

Palmer F. H. (1993) Environmental challenges facing road transport fuel and vehicles outside the U.S.A. SAE Paper No. 932682.

Scheff P. A. and Klevs M. (1987) Source-receptor analysis of volatile hydrocarbons. *J. Envir. Engng* **113**, 994–1005.

Scheff P. A. and Wadden R. A. (1993) Receptor modeling of volatile organic compounds. 1. Emission inventory and validation. *Envir. Sci. Technol.* **27**, 617–625.

Scheff P. A., Wadden R. A., Keil C. B., Jean G. and Jeng J. (1992) Composition of volatile compound emissions from spark ignition and diesel vehicles, coke ovens, wastewater treatment plants and wood combustion. *Proc. 85th Annual Meeting of the Air and Waste Management Association*, paper no. 92-66.02.

Sigsby J. E., Tejada S., Ray W., Lang J. M. and Duncand J. W. (1987) Volatile organic compound emissions from 46 in-used passenger cars. *Envir. Sci. Technol.* **21**, 466–475.

Smith F. M. and Roberts G. O. (1993) Bayesian computation via the Gibbs Sampler and related Markov Chain Monte Carlo Methods. *J. R. Statist. Soc. B* **55**, 3–23.

Taiwan EPA (1993) Taiwan emission data system, version 2.0 (TEDS-2.0).

Wadden R. A., Uno I. and Wakamatsu S. (1986) Source discrimination of short-term hydrocarbon samples measured aloft. *Envir. Sci. Technol.* **20**, 473–483.

Wang D. and Hopke P. K. (1989) The use of constrained least squares to solve the chemical mass balance problem. *Atmospheric Environment* **23**, 2143–2150.

APPENDIX A: MARKOV CHAIN MONTE CARLO METHODS (MCMC)

We follow the notation of the MCMC methods used by Smith and Roberts (1993) to describe Gibbs sampler and Metropolis algorithms in this paper. The objective of using the MCMC in this study is to draw p-vector samples $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ from the distribution $\Lambda(\theta)$, which is proportional to the product of a likelihood function and a prior. To serve such a purpose, we need to construct Markov chains from two well-established simulation procedures, the Gibbs sampler and the Metropolis algorithms. The systematic form of the Gibbs sampler algorithm is used in the following two steps. The first step is to input one set of starting values $\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_p^0)$ arbitrarily to initiate the algorithm. In the second step, successively random drawings can be obtained from the full conditional distributions described in the following:

$$\theta_1^1 \text{ from } \Lambda(\theta_1 \,|\, \theta_2^0, \ldots, \theta_p^0)$$
$$\theta_2^1 \text{ from } \Lambda(\theta_2 \,|\, \theta_1^1, \theta_3^0, \ldots, \theta_p^0)$$
$$\vdots$$
$$\theta_p^1 \text{ from } \Lambda(\theta_p \,|\, \theta_1^1, \theta_2^1, \ldots, \theta_{p-1}^1).$$

By completing the above cycle once, we can get a transition from $\theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_p^0)$ to $\theta^1 = (\theta_1^1, \theta_2^1, \ldots, \theta_p^1)$. Therefore, the iterations of this cycle, in turn, will produce a sequence, $\theta^0, \theta^1, \theta^2, \ldots, \theta^t, \ldots$, which is a realization of a Markov chain. The key feature of this algorithm is that we only sample from one-dimensional distributions, which are performed by the Metropolis algorithm described in the following paragraph.

In generating a value for the next realized state from the current state realization θ_j^t by the Metropolis algorithm, we can randomly draw θ_j^* from the neighbor of θ_j^t and calculate the ratio of

$$\frac{\Lambda(\theta_j^* \,|\, \theta_1^{t+1}, \ldots, \theta_{j-1}^{t+1}, \theta_{j+1}^t, \ldots, \theta_p^t)}{\Lambda(\theta_j^t \,|\, \theta_1^{t+1}, \ldots, \theta_{j-1}^{t+1}, \theta_{j+1}^t, \ldots, \theta_p^t)}.$$

We then compare the ratio with a random number, which is drawn uniformly between 0 and 1. We will accept $\theta_j^{t+1} = \theta_j^*$ if the ratio is greater than this random number. Otherwise, we will reject the value θ_j^*, and set $\theta_j^{t+1} = \theta_j^t$.

Therefore, the Markov chain of $\theta^0, \theta^1, \theta^2, \ldots, \theta^t, \ldots$ can be successfully constructed by using the Metropolis algorithms within the Gibbs sampler as explained in the previous discussions. When t is large enough, $\theta^t = (\theta_1^t, \theta_2^t, \ldots, \theta_p^t)$ will approximately reach the equilibrium distribution, $\Lambda(\theta)$. In other words, the simulation will start to converge after certain iterations. The chains of values after convergence, accordingly, become available samples to mimic $\Lambda(\theta)$.