INITIAL DEVELOPMENT OF AN ESTIMATION TOOL FOR VEHICLE-ATTRIBUTED AIR POLLUTION

Rene Val R. TEODORO
University Extension Specialist
National Center for Transp. Studies
University of the Philippines
Diliman, Quezon City
1101 Philippines
Fax: 63-2-9290494

Dr. Olegario G. VILLORIA
Associate Professor
National Center for Transp. Studies
University of the Philippines
Diliman, Quezon City
1101 Philippines
Fax: 63-2-9290494

abstract: This paper seeks to initiate a step to what will be a ladder of empirically based, vehicle-attributed, air pollution estimation tool development. It presents the formulation of an empirical model that estimates the ambient concentration of carbon monoxide in a roadside environment. The present form of the model which is expressed in terms of traffic volume, traffic speed and wind speed at a particular direction applies only to mid-block of a straight road section in a flat terrain. The model is also envisioned to include pollution source and receptor locational parameters and may be extented to cover different road layouts in the near future.

1. INTRODUCTION

Increasing motorization trend, aging vehicle fleet, and worsening traffic congestion are among the most significant factors contributing to the severe degradation of air quality in Metro-Manila. The 1990 Emission Inventory conducted by the Environmental Management Bureau (EMB) showed that motor vehicles contribute about 78% of the total air pollution load in the metropolis. Carbon Monoxide (CO) in particular, the most toxic among the air pollutants, is almost 100% attributed to mobile sources.

The air quality of Metro Manila based on results of recent monitoring activities indicates an exceedance of Suspended Particulate Matter (SPM) concentration by 300% of the National Ambient Air Quality Standards. Nitrogen Dioxide (NO_2) standard on the other hand is occasionally exceeded while Carbon Monoxide (CO) concentration is observed to be in an increasing trend.

Meanwhile, motor vehicle ownership is continually increasing at an alarming rate. Vehicle registration records of the Land Transportation Office (LTO) show that from 1986 to 1995, the total annual vehicle registration records an average increase of 8.66% without appropriate road and infrastructure provisions. This results to a continuous increase in road traffic and, consequently, to the degradation of the air quality by exhaust gases. A study conducted by the Asian Development Bank in 1992 forecasted that assuming there is no implementation of additional controls on vehicular emission, pollution load from vehicles in 1990 will at least double by the year 2005.

In this crisis, it is important that transportation planners and air quality analysts should work more closely than ever in providing mobility while improving air quality at the same time (Wayson, 1992). However, the current status of knowledge demonstrates the inability of the existing information to bridge the gap between local transportation and air quality issues.

Though several traffic forecasting and estimation tools had long been used in conducting traffic studies, none so far have been used locally in estimating the effect of vehicle traffic on air quality.

The circumstances thus bespeaks a need for researches that will provide the transportation community with tools needed to establish the functional relationships which exist between the fields of transportation and environment. Tools that are necessary in the pursuit of sustainable development and the continuation of socially optimal decision-making. In recognition of this need, this study will therefore be an initial step to what will be a ladder of empirically based air pollution investigation and estimation tool development.

2. GENERAL DESCRIPTION

The study is primarily concerned with the development of an empirical model that estimates the ambient concentration of air pollutants particularly carbon monoxide in a road side environment. Specifically, the statistical model is expressed in terms of traffic flow parameters such as traffic volume and traffic speed and simple meteorological parameters such as wind speed and wind direction. The study further aims to verify the accuracy of the model by using traffic flow, meteorological and air pollution data gathered from a different site. The model in its general form is expressed as:

$$CO = f\,(\,Traf\ Speed,\ Traf\ Volume,\ Emission\ Factors,\ W\ Speed,\ W\ Direction)$$

With the use of Horiba mobile air pollution monitoring system, continuous surveys were conducted to measure the hourly concentration of vehicle-attributed pollutants such as oxides of nitrogen (NOx), suspended particulate matter (SPM) and carbon monoxide (CO) as well as wind speed and wind direction. A daily 14-hour classified volume count and spot speed surveys were simultaneously undertaken with the aid of a traffic monitor mounted on a Mitsubishi Pajero. Actual field layout of the Camp Crame survey site is presented in Figure 2.1.

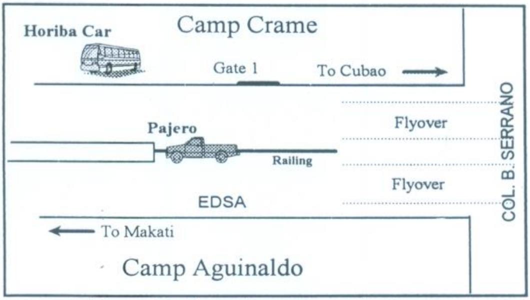


Figure 2.1 Camp Crame survey site layout.

Traffic flow parameters include traffic speed, traffic volume and vehicle composition. Parameters such as vehicle age, vehicle tonnage and engine type, being too difficult to quantify, were not considered both in the conduct of surveys and in modeling. Wind directions with respect to the z-axis were not considered thus simplifying wind into a twodimentional vector. Survey sites were subjected to several locational criteria to simplify topographical considerations and complexity of wind movements brought about by nearby structures.

Atmospheric stability and boundary layer conditions were not considered in actual modeling. The format of the model however was so designed in order to be flexible to further developments such as the inclusion of vertical and lateral dispersion coefficients, both functions of complex meteorological parameters. The chains of chemical reactions occuring in the ambient environment were likewise not included. CO, the main pollutant in focus for this study, is relatively stable as it is difficult to dissolve in water and it does not oxidize without a catalyst. Built-in features of the air pollution monitoring equipment used in the study limits pollutant measurements to a fixed receptor height of 3.5 meters.

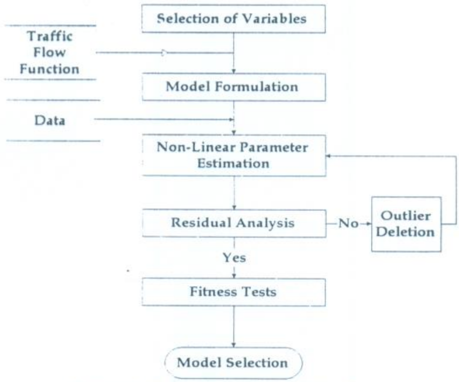


Figure 2.2 Empirical Modelling Process

Empirical modeling primarily utilizes statistical tools like multiple regression and non-linear parameter estimation techniques available in a statistical software. It also utilizes residual analysis and fitness tests to further rectify the model. A schematic diagram of the modeling process is illustrated in Figure 2.2. The effect of wind was considered by classifying the data by wind direction and by performing a separate analysis for each group. Another monitoring activity was conducted for the purpose of testing the applicability and accuracy of the generated models.

3. DATA COLLECTION

In the development of a statistical air pollution model, a basic requirement is the availability of reliable sets of data simultaneously conducted from a particular site. Required data includes hourly measurements of air pollutant concentration, wind speed and direction, traffic volume, fleet composition and average spot speed. In the absence of secondary sets of data, field surveys were conducted. The following sections present the conduct of some of the major data collection and related activities.

3.1 Survey Site Selection

Generally, the ambient sampler should be located outdoor at a place where the public has free access and where the pollutant concentration is highest (De Nevers, 1995). This requirement, together with other obvious considerations such as accessibility, availability of power, enough space for installation of the instruments, security, and distance from other interferring pollutant sources provided proper guidance in site selection. In an attempt to *control* the conditions involved during monitoring and to ease-up the the modeling process with simplifying assumptions, the following criteria in choosing an ideal survey site, as shown in Figure 3.1 was placed atop the above mentioned general practical requirements.

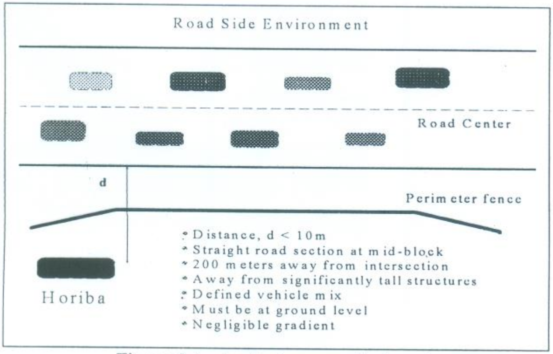


Figure 3.1 An ideal survey site layout.

3.2 Air Pollution Monitoring

The study utilized the state of the art Horiba 350 Series Air Pollution monitoring equipment mounted on a Mitsubishi Rosa. The CO measurement is based on the concept of the absorption of infrared radiation by non-dispersive spectometry. Hourly monitoring was conducted for at least one week per site. Monitoring of oxides of nitrogen utilized the chemiluminescence method while that of the suspended particulate matter utilized the beta-ray absorption method.

3.3 Classified Volume Count Survey

A main variable in road side environment air pollution modeling is traffic volume. Defined as the number of vehicles passing a given point during a specified period of time, traffic volume gives an accurate information on the total traffic from each direction contributing to the pollution level in the area (TTC, 1983). Classified volume count likewise presents details regarding the composition of traffic by vehicle type, and the variation of traffic for a given span of time.

Five general vehicle classes were used in the survey, namely: cars, utility vehicles, buses, trucks and motorcycles. Car includes all *sedan* while utility vehicles include vans, jeepneys, pick-ups and all other vehicles not belonging to the car category and not bigger than the old *Toyota Tamaraw* model. Vehicles bigger than such model, together with trailers are classified as trucks. Buses include conventional buses and mini-buses. Motorcycle inludes both the motorcycles and the tricycles.

Emission rate for every vehicle varies depending on the size of the engine, the type of fuel used, and the weight of the vehicle including its load. The vehicle classification employed in the survey was primarily based on the above-mentioned factors. Engine size and vehicle weight can be accounted for in vehicle size. The composition of the vehicle fleet by fuel type can be derived by assuming all cars and motorcycles to be using gasoline and all buses and trucks to be using diesel. For utility vehicles, a 45.7% gasoline against 54.3% diesel fleet composition which is based on the 1995 vehicle registration by fuel type was safely assumed.

3.4 Spot Speed Survey

Spot speed survey aims to determine the variation of speed at a given location throughout the day (TTC, 1983). As speed affects the rate of emission of vehicles, it likewise supplements the volume count by accurately describing the traffic condition particularly during congestion. Defined as the instantaneous speed of a vehicle at any specified point, spot speed was determined by taking the time it would take for a vehicle to pass through a trap length of a known distance (TTC, 1983). Hourly traffic speed average for the entire direction was determined by taking spot speed samples from the middle lane with the assumption that it approximately represents the average speed for the entire roadway.

The sampling size, *n* that was used in estimating the hourly mean speed was determined by the equation,

$$n = \left(\frac{k\,\sigma}{e} \right)^2 \qquad \text{Eq.(3.1)}$$

where k = level of confidence index
σ = standard deviation
e = allowable error in km/h

Using a 95.0% confidence level and an allowable error of +/- 2 km/h, a fixed number of 180 samples per hour per direction was found to safely meet the required minimum number of samples.

Combined speed, *C_Speed* for both directions was calculated by simply taking the average which is expressed as,

$$C\_Speed = \frac{Speed_1 * Vol_1 + Speed_2 * Vol_2}{Vol_1 + Vol_2} \qquad \text{Eq.(3.2)}$$

where $Speed_i$ = average speed for direction i
Vol_i = total traffic volume for direction i

3.5 Meteorological Monitoring

Meteorological monitoring involves measurements of basic weather parameters such as wind speed and wind direction, the minimum required parameters for an air pollution modeling. Measurements were conducted using an anemometer and anemoscope raised to an elevation of 9.0 meters to have it cleared of any windward obstacle. Hourly measurements of wind speed were expressed in m/s with wind speed of 0.4 m/s or less considered as calm. Most prevalent hourly wind directions were established using the 16 compass points.

The power-law function of height commonly used to estimate the mean wind speed at desired elevation given a set of measurements taken from a different altitude was adopted as expressed in the equation,

$$U = U_0 \left(H/H_0\right)^a \qquad \text{Eq.(3.3)}$$

where, U : Assumed wind speed (m/s) at height H (m)
U_0 : Wind speed (m/s) at standard height H_0 (m)
a : Exponential index

The value for the parameter **a** tends to become bigger as surface roughess increases as shown in Table 3.1. For this study, a value of 1/5 was used representing a convertion index for a sub-urban setting for the relatively flat Camp Crame-Camp Aguinaldo areas.

Table 3.1 Wind speed conversion index for different topograghy.

Land Use Condition	Exponential Index (a)
Urban	1/3
Sub-Urban	1/5
Plain Lands without Obstacle	1/7

Wind speed readings were not right away converted to coincide with the 3.5 meter receptor height. Since conversion is done by simply multiplying all wind measurements by a factor $(H / H_0)^a$ which is just equal to a constant 0.827876 for H=3.5 and H_0=9.0, it was expected not to affect the parameter estimate when conducting the modeling. Instead, the coefficient *b* that *absorbs* the correction factor was later corrected.

3.6 Background Air Pollution Estimation

The daytime background concentration of carbon monoxide on the study area was estimated to be within the range of 0.7 ppm to 1.4 ppm. The range was based on the results of the NCTS survey (Fig.3.2)and the graphical estimation technique (Figure 3.3) that was employed using the EDSA data. This value will be later used as bases for a relatively acceptable modeling-generated intercept of the NCTS survey and the graphical estimation technique that was employed using the EDSA data. This value will be later used as bases for a relatively acceptable modeling-generated intercept.

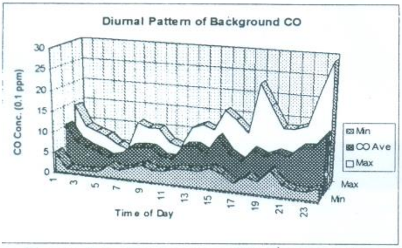


Figure 3.2 Background CO Concentration

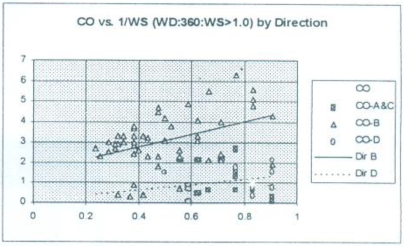


Figure 3.3 Graphical Background CO

4. EMPIRICAL MODEL BUILDING

Air pollution modeling is based on the several generally accepted assumptions. Among them is that pollutant concentration is directly proportional to traffic volume while inversely proportional to wind velocity. Traffic speed, being in general inversely proportional to traffic volume, is likewise assumed to be inversely proportional to the pollutant concentration. Results of previous studies that vehicles traveling in speed lower than 60 km/h tend to emit a more polluted exhaust (Hamilton, 1991) further established the inverse relationship .

Wind direction being a significant factor in the fluctuation of pollutant was accounted for by classifying the data by wind directions. An initial graphical analysis that was conducted showed that relationship between wind speed and CO concentration improved when data are grouped by wind direction. Modeling later focused on the data set where wind blows crossing the road towards the direction of the receptor (Dir B) since it was observed that it gave the most consistent and most significant relationship between the parameters being tested.

Practical modeling considerations included the accuracy, simplicity, applicability and flexibility of the model. Accuracy refered to the model's predictive performance while simplicity was dependent on the amount and availability of the required input. Applicability, on the other hand refered to the vastness of the model's application while flexibility refered to its ability to cater and adopt further modeling developments.

4.1 Model Formulation

Specific format estabishing the relationship between CO and wind speed, was determined by performing a graphical analysis. The analysis was done by generating a scatterplot and drawing a trendline over various sets of data classifications. Data are grouped by wind direction, time of day, and their combinations. A representative format is that having a trendline of higher R^2 and lower intercept.

With EDSA coinciding the NNW-SSE axis, the four groupings by wind direction labeled as Directions A to D are shown in Figure 4.1. In cases where wind direction of a data point coincided with the boundary, the data is accounted for in both adjacent groups. *Directions*

A and *C* were further combined by symmetry, finally reducing the originally 16-point directional classification into three. Various groupings by time of day and combinations thereof were also analyzed.

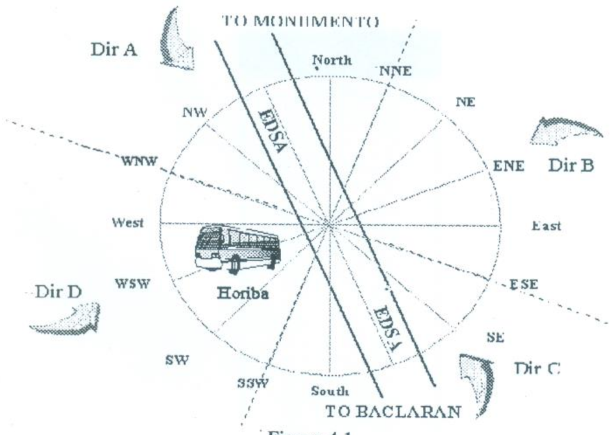


Figure 4.1

The negative exponential and the linear-inverse exhibited a better fit on CO and wind speed relationship. With a and b as the estimated parameters, the former and the latter take the forms:

$$CO = a * exp(-b * WS) \qquad \text{Eq. (4.1)}$$

and, $$CO = a + b / WS \qquad \text{Eq. (4.2)}$$

Further, scrutiny resulted to a slight preference to the latter since it does not return a zero CO estimate as wind speed approaches the value of infinity. Such behavior is more

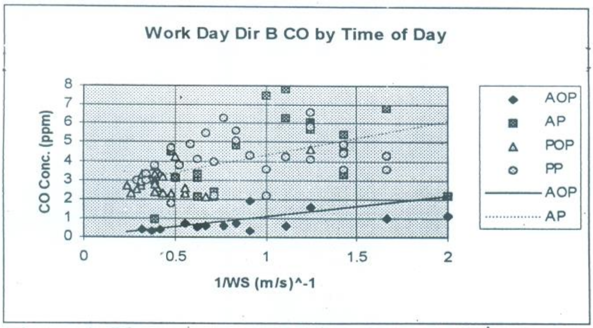


Figure 4.2 CO concentration vs. 1/WS for Direction B by time of day.

realistic, for regardless of the extent of dispersion over an area, there is still an initial concentration representing the background pollution. This is in addition to the higher R^2

values and the lower intercepts the latter exhibited. The second term of the resulting equation is very similar to the simplified Gaussian dispersion equation used by Colwill, as well as the concentration equation adopted by Manins. Figure 4.2 below shows the linearized curves using the factor (1/WS) for *Dir B* and different time of day groupings.

4.2 Empirical Model Development

Trendline analysis identified wind direction B (Dir B), as the group which exhibits statistically significant CO-wind speed relationship. This set of data was therefore used in model building. Starting from Equation 4.2, intercept a' was made to represent the background pollution and the second term as the general dispersion equation used by Manins. This generated a modeling equation in the form,

$$CO = a + k * Q / WS \qquad \text{Eq. 4.3}$$

where emission rate, Q, is a function of traffic flow parameters while coefficient k is a dimensionless parameter that is a factor of averaging time, location of the source and receptor, and the turbulence in the atmosphere (Manins, 1991). This transforms Equation 4.3 into:

$$CO = a + b * f(TP)_i / WS \qquad \text{Eq. 4.4}$$

where $f(TP)_i$ can be any traffic flow function and coefficient b a dimensionless parameter empirically derived to approximate the term (k*Q/f(TP)), thus, preserving the equality. Note that by doing so, coefficient b accounted for the factors concerning atmospheric stability.

Several traffic flow functions were evaluated in the model, each of which are depicting a unique blend of simplicity and accuracy. Among the functions are Equations 4.5-a, 4.5-b and 4.5-c. as shown below.

$$F(TP)_1 = \Sigma (Veh_i * E.F._i) * Speed^c \qquad \text{Eq. (4.5-a)}$$
$$f(TP)_2 = Total\ Vol * (\Sigma\ Speed / n)^c \qquad \text{Eq. (4.5-b)}$$
$$f(TP)_3 = Total\ Vol / (\Sigma\ Speed / n) = C\_VSR \qquad \text{Eq. (4.5-c)}$$

where,
Veh_i = no. of vehicle of type i
E.F.i = emission factor of vehicle i
$Speed^c$ = road section average traffic speed raised to a constant c.
Total Vol = total volume

Equation 4.5-a considers vehicle classification by adopting emision rate factors generated by the 1992 ADB Study. Equations 4.5 b & c on the other hand assumes that gasoline and diesel-engined vehicles emit the same amount of carbon monoxide, a simplifying assumption used by Colwill and Hickman on a similar study in 1982 justifying that the lower CO concentration of diesel exhaust is being offset by the larger volume of exhaust produced by large diesel engine. The last function, the simplest, only combines volume and speed into a single traffic flow parameter, *combined volume-speed ratio*.

4.3 Modeling Results

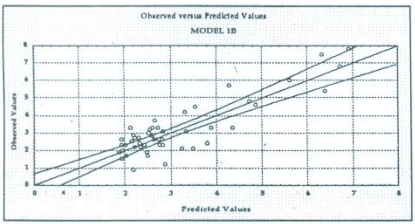
The Non-linear Estimation procedure of STATISTICA, a statistical package by StatSoft, Inc. was used in generating the empirical models. An initial run using all sets of data was conducted followed by a residual analysis. The generated model then was further rectified based on the results of the analysis. Table 4.1 summarizes the statistics of the three modeling equations containing the traffic flow function $f(TP)_i$ plus a fourth equation (Model 1-b) which is basically $f(TP)_i$ with exponent c=-1.

Table 4.1 Summary of the modeling estimates and the fitness tests results.

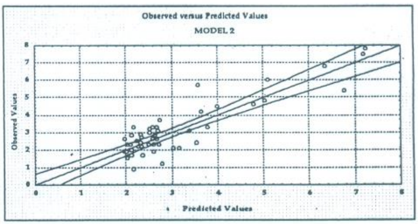
MODELING & FITNESS TESTS RESULTS				
Parameters / Indexes	**Model 1**	**Model 1B**	**Model 2**	**Model 3**
A	1.60572	1.38811	1.60126	1.37553
Std. Error	0.13738	0.15927	0.13761	0.16050
t(53)	11.68859	8.71568	11.63592	8.57006
p-level	0.00000	0.00000	0.00000	0.00000
B	0.00266	0.00030	0.10689	0.01156
Std. Error	0.00019	0.00002	0.00757	0.00088
t(53)	14.12293	13.14376	14.12248	13.10047
p-level	0.00000	0.00000	0.00000	0.00000
C	-1.64916		-1.66399	
R	0.88886	0.87478	0.88885	0.87410
Fin. Loss	24.94719	27.89751	24.94846	28.03861

The evaluation of model fit involved the examination of the Observed vs. Predicted Values; the Normality Plot of the Residuals and the Plot of the Fitted Functions. Also included were the assessment of statistical indexes such as correlation coefficients, standard error of estimates and percentage of explained variance.

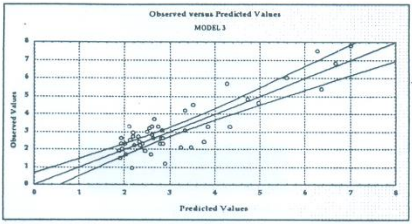
Modeling and fitness tests results as shown in Table 4.1, generally depict set of well fitted models. The The *t-statistics* for instance as indicated by the *p-levels* denoted highly significant parameter estimates. Very low *standard errors* particularly for the coefficients *b* proved the exactness of the estimates to yield the least combined residuals. The *correlation coefficients, R*, with values ranging from 0.874 to 0.889 for a total of 55 samples surpassed



(a) Observed vs. Predicted Plot of Model 1B



(b) Observed vs. Predicted Plot of Model 2



(c) Observed vs. Predicted Plot of Model 3

Figure 4.3 Observed vs predicted scatterplots with a 99.0% Confidence band width.

the critical values of the Spearman's Rank Correlation Coefficients of 0.478 for $\alpha = 0.005$ for even a smaller sample. The *final loss* values representing the summation of the loss function, which in this case equal to $(OBS\text{-}PRED)^2$, posted relatively low values indicating a good set of estimates. The *percentage of explained variance* were simply the equivalent of the coefficient of determination R^2. Figures 4.3 presents some of the Observed vs. Predicted plots of the generated models.

4.4 Endorsement of the Best Model

All four models primarily met the basic statistical guidelines regarding parameter estimates and goodness of fit. An examination of the the models *Observed vs. Predicted (O-P)* plots showed a similar number of points found inside and outside the confidence level band width. Though in general, a very slight difference in the O-P was observed in favor of Model 2 attributed to an extra parameter *c*, the slight edge is leveled off by the presence of few data points that are way off O-P proportionality line. In terms of accuracy, the differences particularly on the indexes and test values among models were found to be so minimal to merit major consideration in the selection of the best model.

Model 3 is the simplest followed by Model 2 with only *combined volume-speed ratio* (*C_VSR*) for the former, and traffic speed and volume for the latter as key modeling inputs. Both though, except for the exponent *c* and some minor calculation requirements in Model 2 are essentially similar with C_VSR being a function of both traffic speed and volume. Model 1 on the other hand is the most complex followed by Model 1B. Both will require intricate calculations and breakdown of traffic volume by vehicle type which is hard to secure as data from traffic existing detectors does not contain information on vehicle composition. The complex models, however, offers a wider use as it is not limited to mere forecasting and estimation. Among its potential applications included the estimation of pollution load contribution by vehicle type as well as the environmental evaluation of vehicle-type based traffic management scheme like truck ban and several car restraint policies.

All four models were almost equally adaptable. A main development projected to be considered in future studies is the inclusion of the seasonal variation of atmospheric stability. For the mean time, the combined characteristics concerning atmospheric stability and other unconsidered factors were represented by the empirical coefficient *b*. The inclusion of lateral and vertical diffusion coefficients, σ_x and σ_y, using the same data set for instance will simply replace the coefficient b with $d / (\sigma_x * \sigma_y)$ where d is roughly equal to $(\sigma_x * \sigma_y) * b$ thus preserving the model.

A parameter which differentiates the models over the other is the modeling intercept. Theoretically, models with lower intercepts are better since intercept accounts for the the relative error in the value contributed by the other parameters. Such models likewise has a wider range of guess values as it is capable of generating low estimates due the lower initial cut off. This can be observed in the O-P plot wherein Model 2, having a higher intercept, rarely generated a predicted value lower than the 2.0 ppm. The intercept must likewise fall within the established background pollution range.

Based on the above discussions, the study endorses Model 3 as the best model to be further developed to improve its application in the area of on-site air pollution estimation and forecasting. A three-dimentional plot of Model 3 is presented in Figure 4.4. It's simplicity, relative accuracy and rational numerical features made it the best model relating a traffic flow function and an air pollutant concentration over the others.

Taking in to account the correction due to the wind measurement altitude as discussed in Section 3.5, a correction factor was introduced to the coefficient *b* yielding the model's final form. The corrected model, Model 3 in its final format is expressed as follows:

$$CO = 1.37553 + 0.013963 * C\_Volume\ Speed\ Ratio / Wind\ Speed \qquad \text{(Eq. 4.6)}$$

where, CO = carbon monoxide ambient concentration in ppm
C_Volume Speed Ratio = the ratio of ombined traffic volume in veh/h and average traffic speed in km/h
Wind Speed = wind speed measurement at 3.5 m altitude in m/s

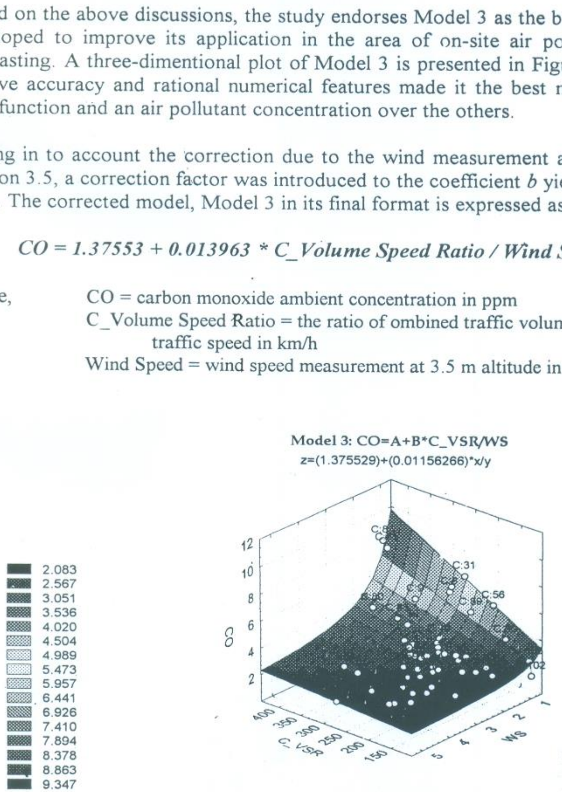


Figure 4.4 Three-dimensional surface plot of Model 3.

5. MODEL PERFORMANCE TEST

Model performance test aimed to evaluate the predictive performance of the model and diagnose the conditions associated with the inaccuracies in the model's prediction (TRB, 1981). To test the performance of the model, another monitoring activity was conducted to gather the same set of parameters from a different site. The model was utilized to estimate the carbon monoxide levels given a new set of wind and traffic flow measurements. The generated estimates were then compared to the actual measurements through the conduct of fitness tests and residual analysis.

The monitoring activity was conducted along Commonwealth Avenue inside the Asian Institute of Tourism (AIT) compound. Using the same equipment, the mobile air pollution monitor was set up just 6- meters from the side of the road. The survey gathered traffic, air pollution and meteorological parameters using the same data gathering procedures.

The orientation of Commonwealth Ave. results to the following groupings namely: (1) Wind Dir A: wind is blowing to Quezon City Memorial Circle (QMC) parallel to the road (NNE to ENE quadrant); and (2) Wind Dir B: wind is blowing towards the receptor (E to S quadrant); (3) Wind Dir C: wind is blowing to Fairview parallel to road (SSW to W quadrant); and (4) Wind Dir D: wind blowing away from the receptor (WNW to N quadrant). Due to the prevalence of the northeast monsoon, only three data points fall outside the Wind Directions A and B.

5.1 Residual Analysis

Using the calculated CO as the *expected* and the gathered data as the *observed* values, the analysis of the residuals exhibited promising results. Out of the 60 data points, only four residuals were found to be lying beyond the +/-2 ppm residual range. Further, almost 80% of the entire data were within the +/-1 ppm and about 50% were within the +/-0.5 ppm expanse. The frequency distribution in Figure 5.1 further shows that most of the residual falls at the left of the bell curve, this initially indicated that with most data points, the model generated a conservative estimate by predicting a slightly higher concentration

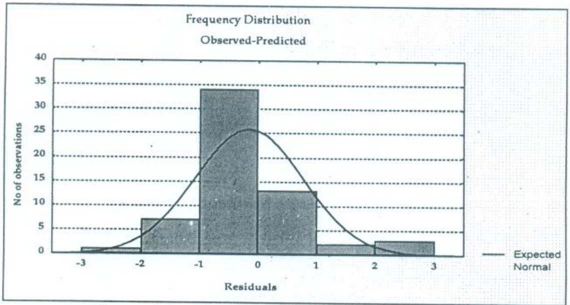


Figure 5.1 Frequency distribution of residuals.

values than that of the observed. It was found out later from the *observed-expected* scatterplot that the distribution of samples and an estimation bias cause the uneven distribution of the residual.

5.2 Fitting Linear Function

The scatterplot as shown in Figure 5.2 was intended to supplement the model performance analysis by residual distribution. Assuming a perfect fit, the data points in the scatterplot will theoretically coincide with the *observed=expected* line. The scattering of the plots with respect to the line was evaluated by employing a linear regression with zero intercept. A good fit should be indicated by a high correlation coefficient together with a parameter estimate (slope) that is very close to 1.0. The result of the regression is summarized in Table 5.1.

Table 5.1 Linear Regression with zero intercept.

STAT. REGRESS.	Mult. R = 0.68273 R-Square = 0.46612		Adj. R-Square = 0.44917 Fin. Loss = 51.48881	
N=60	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
EXPECTED	0.96512	0.05066	19.05118	0.00000

The estimated slope of 0.96512 showed that the fitted function almost coincided with the observed vs. expected line. The correlation coefficient of 0.68273 relatively exhibits a good linear relationship considering 60 samples. The R-square of 0.46612 means that

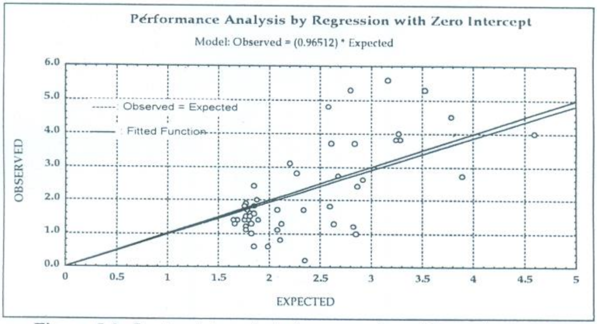


Figure 5.2 Scatterplot analysis by regression with zero intercept.

approximately 46.6% of the variation in the *observed* values is accounted for by a linear relationship with the *expected*. The p-value of almost equal to zero and the t-statistics of 19.05118 for 58 degrees of freedom indicates that the correlation coefficient is significantly greater than 0 even at $t_{0.995}$ (0.005 significance level).

Further, the scattering of data points with respect to the *observed=expected* line was found to be unevenly distributed. Taking the observed value as reference, CO measurements greater than 3.0 ppm tended to fell below the *observed=expected* line while those greater than 3.0 were plotted above the line. The observation was examined by fitting a linear function (with intercept) that would best correlate the *observed* and the *expected* values, and then comparing the fitted function to that of the *observed=expected* line. The result of linear regression is presented in Table 5.2.

Table 5.2 Observed Vs. Expected Fitted function.

STAT. REGRESS.	Mult. R = 0.71563 R-Square = 0.51212		Adj. R-Square = 0.50371 Fin. Loss = 47.05158	
N=60	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.96673	0.41336	-2.33875	0.02282
EXPECTED	1.35481	0.17363	7.80275	0.00000

The negative intercept (-0.96673) indicated that the model was inclined to initially overestimate the low-level CO concentration (Fig.5.3). Ideally, an intercept indicating a good fit should be close to zero. The calculated coefficient of 1.35481 (greater than 1.0) indicated that as CO level increases, the observed value gradually outpaces the corresponding increase in the expected value, thus, neutralizing the overestimating effect of the intercept until such level that the model underestimates the observed CO concentration. As a result, good estimates were made between observed CO values of 1.0 ppm and 4.5 ppm. With most AIT measurements falling within this range, a positive result of the previous test was generated. The result can be partly attributed to the measurement ranges of the two data sets. Note that EDSA data, which was used in the model calibration ranges from 0.9 ppm to 7.8 ppm, encompassed the AIT range of 2.1 ppm to 5.6 ppm.

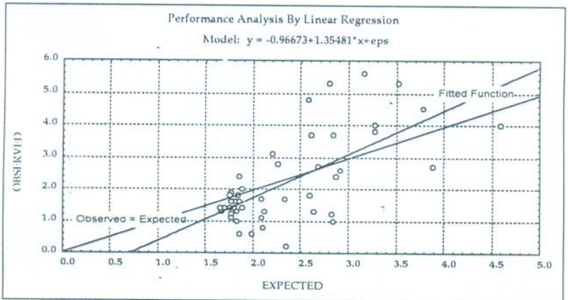


Figure 5.3 Observed vs. expected scatterplot analysis by linear regression.

5.3 Diagnostic Analysis

The uneven distribution or the *bias* in the scattering of the observed vs. predicted plot was attributed to certain conditions that are different from each site. This particularly pertained to the factors accounted for by the estimated parameters intercept *a*, and coefficient *b*. The negative intercept in Table 5.2 for instance signified that a lower intercept was more suitable in describing the pollution level at AIT. Likewise, the coefficient of 1.35481 indicated that a higher slope would be more appropriate. Since intercept *a* accounted for the background pollution in the area, and that EDSA was relatively closer to other major roads than that of AIT, then it is just but logical and consistent to conclude that a higher background pollution which exists at EDSA, caused the over estimation of low level CO concentration at AIT.

On the other hand, parameter b mainly accounted for factors represented by the dimensionless parameter K in Eq.(4.3) such as averaging time, location of the source and receptor, and turbulence in the atmosphere. Considering the similarities in averaging time, and atmospheric conditions leaves the location of the source and receptor as the potential factor causing the difference. Source and receptor location is basically a distance parameter-distance between the roadside and the receptor as well as the width of the road section, with road center being considered as the source point.

With Horiba at EDSA positioned farther by four (4) meters from the roadside compared to that of AIT, in addition to being a 10-lane two-direction stretch against the 6-lane two-direction Commonwealth Avenue, EDSA differs in distance between the receptor and the road center by at least 11.0 m. The difference in distance is inclined to be even more significant as the wind blowing from the road towards the receptor decreases. Note that assuming equal traffic flow measurements, the decrease in wind speed is characterized by an increase in pollution level. This supposition stands consistent with the identified trend that the difference between *observed* and *expected* was increasing as the CO level increases.

5.4 Performance Test Summary of Results

To sum up, the encouraging result indicated by a good percentage of insignificant residuals was proved to be limited to a certain range of CO values. CO levels within the 1.0 ppm to 4.5 ppm were observed to have tolerable residuals while those beyond the range had the tendency to generate larger errors of estimate. Aside from the obvious technical similarities on the conduct of both surveys, the fit within the range could be attributed to the following similarities:

(a) topography of the site
(b) meteorological similarities
(c) CO concentration range

Though EDSA and Commonwealth Ave. were basically different in terms of road classification, size and traffic characteristics, the selection of both sites were based on a single yardstick. Specifically on topography, both sites were free of significant structures that obstructs the path of the wind that carried the polluted air from the road towards the location of the receptor. Both surveys were meteorologically similar being conducted during fair days of a Northeast monsoon season. Wind measurements also indicated that both are within the same wind velocity and are not affected by the land breeze and sea breeze circulation.

With EDSA having a bigger traffic and a volume-speed ratio ranging from 89 to 560 vehicles per kilometer of road section, it was expected that it will generate a wider range of CO measurement than that of Commonwealth Ave. However, AIT with volume-speed ratio ranging only from 50 to around 170 vehicles per km of road section was expected not to be fully explained by the model as volume-speed ratios lower than EDSA's 89 veh/km were not represented in the calibration. This partially explained the uneven distribution of larger residuals.

Further, the analysis using fitted function with intercept identified the differences in background pollution, road width and distance from receptor to the side of the road as the cause of the the biased estimates. These very parameters are needed to be considered in improving the estimation range and accuracy of the model.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Empirical Modeling Results

A simple statistical model designed to estimate the ambient concentration of an air pollutant in a roadside environment was developed. The model was able to particularly estimate carbon monoxide level given basic traffic flow and meteorological parameters. The model was generated using first-hand. hourly day-time data points with wind blowing from the road towards the direction of the receptor. Considerations on coming up with an appropriate format included the adoption of commonly accepted assumptions, some general similarities with other estimation models and the utilization of basic statistical techniques.

At present, the model can handle CO level estimation given traffic speed and volume, or simply the *combined volume speed ratio*, and wind speed blowing from the road towards the point being evaluated. An on site model performance test was conducted yielding positive results and identifying potential areas of improving the model. The generated model is as shown below:

$$CO = 1.37553 + 0.013963 * Comb.Vol\_Spd\ Ratio / Wind\ Speed$$

6.2 Model Performance Analysis

The result of the model performance test was encouraging revealing a good correspondence between the expected and the observed values though only for a certain range of CO concentration. Factors contributing to the generation of a good set of estimates were attributed to similarities in general meteorology, the sites' topography and the coinciding CO concentration range of the two study sites. The model however seemed to overestimate CO concentration outside the lower bound of the range and underestimate CO level beyond the upper bound of the range. The biased miscalculations were attributed to the difference in background air pollution, road width and distance between the receptor and the side of the road. Appropriate considerations on the identified parameters together with wind direction are expected to significantly improve the estimating capability of the model.

In general, the results of the test hinted at the possibility of the models applicability to several other sites of different road type and traffic composition. Data points of significant residual values on the other hand echoed a universal truth: that model estimates should not be regarded as numerically accurate description of the actual air quality or prediction of a projected scenarios (Hickman, 1982). Rather, model estimates simply gives a general description of the most likely situation given a particular condition.

6.3 Recommendations for Further Development

The genetrated model however was far from perfect. Data used were practically limited to a particular relative wind direction, flat terrain, straight road sections, a road side location and a dry-Northeast monsoon season. Numerous simplifying assumptions were likewise adopted in the course of model formulation thus further limiting the applicability of the model to specific conditions. Researches dealing with the effect of a variation in the identified parameters to CO concentration is highly recommended to further the model's applicability. For instance, the inclusion of wind direction as a continuous variable expressed as a function of an angle with respect to the position of the road will make the model applicable to all wind directions.

Model calibration of other pollutants, perhaps, SPM, Hydrocarbons and NO_2, is viewed to be likewise necessary. Though establishing the empirical relationships between the concentration of CO and other pollutants could be possible, an estimation of the concentration of another pollutant based on its relationship to CO can be very unreliable.

The formulation of a simulation program is likewise one of the most immediate steps identified beyond the coverage the study. Simulation program can be very useful in further evaluating and developing the existing models. Being replicable, it will be very useful in conducting evaluation runs with minimal conduct of field surveys. Simulation is likewise a potential start towards a comprehensive software development.

REFERENCES

Asian Development Bank (1992): "Vehicle Emission Control Planning Project for Metro-Manila" Manila, Philippines.

De Nevers, Noel (1995): Air Pollution Control Engineering. McGraw-Hill Inc., Chemical Engineering Series, Singapore.

Environmental Management Bureau (1990):. *The Philippine Environment in the Eighties*, Environmental Management Bureau, Department of Environment and Natural Resources, Republic of the Philippines. pp. 1-32

Hamilton, R. S. and Harrison R. M. (1991): Highway Pollution. Studies in Environmental Science 44; Elsevier Science Publishers B.V., Amsterdam. pp. 8-10

Hickman, A J and D M Colwill (1982): *The Estimation of Air Pollution Concentrations From Road Traffic*, Department of the Environment Department of Transport, TRRL Report LR 1052, Transport and Road Research Laboratory, Crowthorne, Berkshire.

Manins, Peter (1991): Model for Air Pollution Planning. WHO Consultant, Environmental Management Bureau, Department of Environment and Natural Resources, Philippines.

National Center for Transportation Studies (1995): NCTS Environmental Text Series. NCTS, University of the Philippines System, Diliman, Quezon City.

StatSoft, Inc. (1994): STATISTICA for Windows (Volume I: General Conventions and Statistics I and Volume III: Statistics II). StatSoft, Inc. Tulsa, Oklahoma, USA. pp. 3005-3053

Teodoro, Rene Val R. (1996): Empirical Analysis on the Relationship Between Air Pollution and Traffic Flow Parameters (Masters Thesis). College of Engineering. University of the Philippines

Transport Research Board (1981): *Methodology for Evaluating Highway Air Pollution Models*, National Cooperative Highway Research Program Report 245. TRB, National Research Council, Washington, D.C. pp.4-5

Transport Training Center (1983): Course Notes on Transportation and Traffic Technology. TTC, University of the Philippines System, Diliman, Quezon City.

Waterfield, V H and A J Hickman (1982): *Estimating Air Pollution From Road Traffic: A Graphical Screening Method*, TRRL Report SR 752. Department of the Environment Department of Transport, Transport and Road Research Laboratory, Crowrthorne, Berkshire.

Wayson, Roger L. (1992): *Transportation Planning and Air Quality*.Conference Proceedings. Urban Transportation Division, American Society of Civil Engineers (ASCE), New York.p.iii