Atmospheric Pollution Research 6 (2015) 581-588

Atmospheric Pollution Research

www.atmospolres.com

# Evaluation of MARS for the spatial distribution modeling of carbon monoxide in an urban area

**Hamid Taheri Shahraiyni [1,2], Davood Shahsavani [3], Saeed Sargazi [4], Majid Habibi–Nokhandan [5]**

[1] *Institut für Meteorologie, Freie Universität Berlin, Berlin, Germany*
[2] *Faculty of Civil Engineering, Shahrood University, Shahrood, Iran*
[3] *Faculty of Mathematics, Shahrood University, Shahrood, Iran*
[4] *Faculty of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran*
[5] *Climatological Research Institute, Mashad, Iran*

## ABSTRACT

Spatial distribution modeling of CO in Tehran can lead to better air pollution management and control, and it is also suitable for exposure assessment and epidemiological studies. In this study MARS (Multi–variate Adaptive Regression Splines) is compared with typical interpolation techniques for spatial distribution modeling of hourly and daily CO concentrations in Tehran, Iran. The measured CO data in 2008 by 16 monitoring stations were used in this study. The Generalized Cross Validation (GCV) and Cross Validation techniques were utilized for the parameter optimization in the MARS and other techniques, respectively. Then the optimized techniques were compared based on the mean absolute of percentage error (MAPE). Although the Cokriging technique presented less MAPE than the Inverse Distance Weighting, Thin Plate Smooth Splines and Kriging techniques, MARS exhibited the least MAPE. In addition, the MARS modeling procedure is easy. Therefore, MARS has merit to be introduced as an appropriate method for spatial distribution modeling. The number of air pollution monitoring stations is very low (16 stations for 22 zones) and the distribution of stations is not suitable for spatial estimation, hence the level of errors was relatively high (more than 60%). Consequently, hourly and daily mapping of CO provides a limited picture of spatial patterns of CO in Tehran, but it is suitable for estimation of relative CO levels in different zones of Tehran. Hence, the map of mean annual CO concentration was generated by averaging daily CO distributions in 2008. It showed that the most polluted regions in Tehran are the central, eastern and southeastern parts, and mean annual CO concentration in these parts (zones 6, 12, 13, 14 and 15) is between 4.2 and 4.6 ppm.

***Keywords:*** *Carbon Monoxide, Multi–variate Adaptive Regression Splines (MARS), spatial distribution modeling, Tehran, interpolation techniques*

***Corresponding Author:***
*Hamid Taheri Shahraiyni*
☎ : *+49-30-83854366*
🖷 : *+49-30-83854366*
✉ : *hamid.taheri@met.fu-berlin.de*

***Article History:***
*Received: 26 August 2014*
*Revised: 29 December 2014*
*Accepted: 29 December 2014*

**doi:** 10.5094/APR.2015.065

## 1. Introduction

In the last decades, the energy consumption, emission of air pollutants and number of poor air quality days has increased in cities because of increases in population in urban areas (Chan and Yao, 2008). Similar to many other mega cities, Tehran (the capital city of Iran) has serious air pollution problems. The major air pollutants in Tehran are CO, $PM_{10}$, $NO_2$, $SO_2$, HC and $O_3$ (Naddafi et al., 2012). Although only 10% of Iran's population lives in Tehran, about 30% of Iran's vehicles are found in Tehran and 30% of Iran's industrial sites are located around Tehran (Atash, 2007). About 70 and 30% of the air pollution sources in Tehran are non–stationary and stationary sources, respectively (Abbaspour and Soltaninejad, 2004). Of Tehran's many motor vehicles, 40% of those are old vehicles (20 years or older) (Atash, 2007). Low quality fuels and the large number of old and poorly maintained vehicles are responsible for much of the pollutant emissions in Tehran (Abbaspour and Soltaninejad, 2004; Atash, 2007). CO, one of the major air pollutants in Tehran, is mainly formed by primary combustion, particularly in vehicle engines (Givehchi et al., 2013), and there is a direct relationship between CO emission and traffic flow and speed in Tehran (Rashidi and Massoudi, 1980). The sources of CO in Tehran are emissions by transportation (97.1%), household and commercial areas (0.6%), industries (0.3%) and agriculture (0.2%) (Vafa–Arani et al., 2014). The health risks posed by CO are highly related to the concentration and duration of exposure. CO poisoning at low concentrations can cause headaches, dizziness, weakness, nausea, vomiting, confusion, disorientation, visual disturbance and subtle cardiovascular illnesses (Raub et al., 2000). Unconsciousness, coma, convulsions, cardiorespiratory problems, and death may occur after prolonged exposures or acute exposure to high concentrations of CO (Raub et al., 2000). The studies in Tehran showed that increased CO concentration can increase the allergenicity of pollen grains (Majd et al., 2004), daily admission in Tehran's hospitals due to cardiovascular diseases (Hosseinpoor et al., 2005) and the total number of deaths due to respiratory and cardiovascular diseases and cancer (Abbaspour and Soltaninejad, 2004). In addition, it has been found that exposure to increased levels of CO during pregnancy may increase the risk of fetal abnormality (Ziaei et al., 2005).

Modeling of spatial distribution of pollutants can help to estimate pollutant concentration in areas with no air pollution monitoring stations. In addition, it can specify the regions that exceed the air pollution standards. Furthermore, spatial distribution modeling of pollutants can be utilized for the exposure assessment and epidemiological studies (Jerrett et al., 2005; Hoek et al., 2008). Little data is available about the spatial distribution of CO in Tehran, hence there is a need for estimation of CO levels in Tehran's different zones.

There are two general approaches for mapping of air pollutants: dispersion modeling and spatial interpolation (Briggs, 1992). The most serious problem with dispersion models is their severe data demand (Briggs et al., 1997). Different techniques have been developed for the spatial interpolation of air pollutants and they are utilized for the modeling of spatial distribution of air pollutants in urban area. Among these techniques, the IDW (Inverse Distance Weighting), Kriging, Cokriging and TPSS (Thin Plate Smooth Splines) are well–known interpolation techniques and have been widely used for the spatial distribution modeling of air pollutants. Mulholland et al. (1998) used universal Kriging for the spatial interpolation of ozone in Atlanta metropolitan area. TPSS was applied for spatial distribution modeling of $NO_2$ concentrations and its performance was evaluated by the cross validation technique (Ionescu et al., 2000). The results showed that estimated $NO_2$ concentrations by interpolation are reliable. Pikhart et al. (2001) employed geostatistical techniques for the estimation of $SO_2$ levels in Prague, Czech Republic and Poznan, Poland. Ung et al. (2001) utilized the TPSS and Kriging methods for the modeling of pollutant concentration in Strasbourg, France. Jerrett et al. (2001) applied universal Kriging for spatial interpolation of Total Suspended Particulates (TSP) in Hamilton, Canada. Bel (2004) proposed a new technique for evaluation of ozone concentration over Paris, France and compared it with classical Kriging methods. Patil et al. (2003) utilized different interpolation techniques (e.g. IDW and TPSS) for the mapping of $NO_X$ and $SO_X$ spatial variations in the winter in Delhi, India. Finkelstein et al. (2003) mapped the TSP and $SO_2$ in Hamilton using the Kriging. Jerrett et al. (2005) applied cubic splines, Kriging, universal Kriging, and multi–quadratic radial basis function for spatial distribution modeling of $PM_{2.5}$ in Los Angeles, California. They found that universal Kriging and the multi–quadratic radial basis function have better performance than the other methods. Beaulant et al. (2008) employed TPSS for the spatial distribution modeling of Strasbourg, France. Pearce et al. (2009) investigated the potential of the Ordinary Kriging and Kriging with External Drift (KED) techniques to provide high–resolution maps of $PM_{2.5}$ in Cusco, Peru. Statistical evaluations indicated that KED was the most appropriate model. Beelen et al. (2009) applied ordinary Kriging and universal Kriging techniques for the development of EU–wide maps of air pollutants ($NO_2$, $PM_{10}$, $O_3$, $SO_2$, and CO) on a 1×1 km resolution for the global, rural and urban scales, separately. The universal Kriging presented better performance than the ordinary Kriging. Singh et al. (2011) compared the Kriging and Cokriging techniques for spatial interpolation of 8h mean daily maximum ozone concentration and daily mean $PM_{10}$ concentration over the Milan, Italy urban area. The results indicated that the Cokriging technique has better performance than the Kriging technique.

MARS (Multi–variate Adaptive Regression Splines) is a well–known nonparametric regression method that has been utilized for the different spatial and temporal modeling in miscellaneous studies (e.g. Munoz and Felicisimo, 2004; Leathwick et al., 2005; Leathwick et al., 2006; Heikkinen et al., 2007; Storlie et al., 2009), but it has not been used for spatial distribution modeling of air pollutants over urban areas.

MARS is essentially a combination of spline regression, stepwise model fitting and recursive partitioning, and it is able to reveal the underlying nonlinear patterns hidden in complex data sets (Storlie et al., 2009). Number of input variables is not a limitation for MARS and it can be applied for spatial interpolation with different numbers of input variables. In addition, one of its important capabilities is the ability to consider the interaction of different input variables from the product of two or more basis functions (Friedman, 1991a; Friedman, 1991b; Friedman and Rooster, 2005; Hastie et al., 2009). Hence, the MARS construction implies its ability for multivariable interpolation problems. Thus in this paper, MARS is evaluated for the spatial distribution modeling of an air pollutant in Tehran and it is compared with the IDW, TPSS, Kriging and Cokriging techniques. CO is one of the major air pollutants in Tehran (Ziaei et al., 2005) and it is one of the pollutants that requires prevention and control measures to insure adequate human health protection (Raub et al., 2000). Therefore, in this study the spatial distribution of CO in Tehran's different zones is developed by MARS.

## 2. Materials and Methods

### 2.1. Study area

Our study area is the city of Tehran, which is located in the northern part of Iran (between 35.56–35.83°N and 51.20–51.61°E) and its area is 18 909 $km^2$. It is the capital and the most crowded city of Iran and its population is about 11 million. The geographical position of Tehran in Iran is exhibited in Figure 1.


***Figure 1.*** *The position of Tehran in Iran (Source map: US Energy Information Administration), with a view of Tehran and the Alborz mountains in the northern part of Tehran (Photographer: Hansueli Krapf).*

Tehran is bordered by the Alborz Mountains in the northern part (Figure 1) and there are flat plains in the other parts of Tehran. The weather in Tehran is mainly influenced by altitude elements and pollutant emissions. The pollution cannot escape from the city because of a lack of permanent winds. The altitudes in Tehran in the northern, central and southern parts are 1 700, 1 200 and 1 100 meters, respectively. The maximum and minimum temperatures in Tehran are 43.9 and –15 centigrade degrees, respectively. Mean annual rainfall in Tehran is about 316 millimeters. Tehran city and its 22 zones are exhibited in Figure 2. The CO data was retrieved from 16 air pollution monitoring stations shown in Figure 2.

### 2.2. Air pollution data

The air pollution monitoring stations in Tehran displays different pollutant concentrations in the air by bar charts on an electronic panel and they save the concentration of pollutants as hourly averaged data. All of the stations measure the level of air pollutants such as CO, $NO_2$, $SO_2$, $NO_X$, $O_3$ and $PM_{10}$. The hourly CO data of 16 air pollution monitoring stations was provided by the Tehran Air Quality Control Corporation (AQCC). The latitude, longitude and altitude of the employed stations are presented in Table 1. The accuracy of CO measurements is ±10%. The air pollution data from 2008 was utilized in this study. A database of hourly CO concentrations measured at different stations in 2008 was created. Then the average of the hourly data was considered as the mean daily CO concentration and the database of daily CO concentration in 2008 was created.

### 2.3. Elevation data

The elevation data in Tehran is utilized as the auxiliary variable for the Cokriging technique, or it can be used as an input variable for the MARS model. The implementation of MARS with three input variables (latitude, longitude and elevation), also known as MARS3, requires elevation data as an input variable.

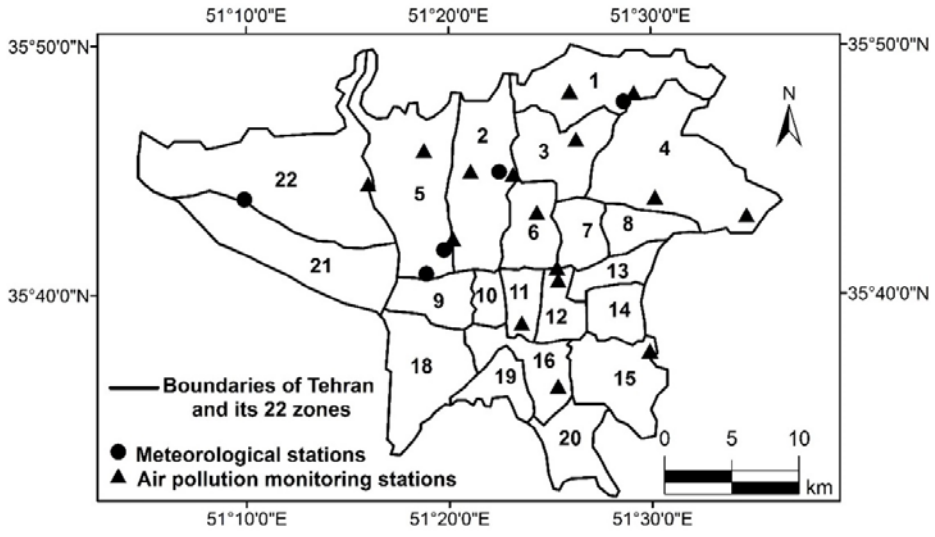


***Figure 2.*** *Tehran's geographical situation and its 22 zones with the positions of meteorological and air pollution monitoring stations.*

***Table 1.*** *Location of the air pollution monitoring stations*

| Altitude (m) | Longitude ( °, ', '') | Latitude ( °, ', '') | Longitude (UTM) | Latitude (UTM) |
|---|---|---|---|---|
| 1 260 | 51, 24, 28.475 | 35, 43, 16.388 | 536893 | 3953105 |
| 1140 | 51, 25, 23.563 | 35, 40, 38.060 | 538298 | 3948233 |
| 1 575 | 51, 29, 6.114 | 35, 48, 0.21 | 543825 | 3961881 |
| 1 390 | 51, 23, 10.705 | 35, 44, 48.242 | 534928 | 3955927 |
| 1 067 | 51, 25, 32.08 | 35, 36, 16.474 | 538547 | 3940175 |
| 1 299 | 51, 16, 4.39 | 35, 44, 23.597 | 524223 | 3955132 |
| 1 480 | 51, 18, 51.611 | 35, 45, 46. 415 | 528415 | 3957696 |
| 1 302 | 51, 30, 22.073 | 35, 43, 51.69 | 545771 | 3954234 |
| 1 178 | 51, 29, 56.480 | 35, 37, 48.097 | 545185 | 3943029 |
| 1 299 | 51, 25, 34.003 | 35, 43, 48.490 | 538565 | 3954101 |
| 1 211 | 51, 20, 13.997 | 35, 42, 16.121 | 530506 | 3951224 |
| 1 140 | 51, 23, 41.144 | 35, 38, 48.826 | 535737 | 3944857 |
| 1 422 | 51, 26, 21.895 | 35, 46, 21.068 | 539717 | 3958807 |
| 1 431 | 51, 34, 57.879 | 35, 43, 0.588 | 552708 | 3952968 |
| 1 596 | 51, 26, 3.355 | 35, 48, 21.942 | 539235 | 3962529 |
| 1 152 | 51, 25, 24.498 | 35, 41, 4.574 | 538318 | 3949050 |

### 2.4. Interpolation techniques

Interpolation techniques that are considered in this research are the MARS, IDW, TPSS, Kriging and Cokriging techniques. The IDW, TPSS, Kriging and Cokriging techniques are well known and they are described in Webster and Oliver (2001).

MARS is a novel nonparametric modeling technique that was developed by Jerome Friedman (Friedman, 1991a) and its algorithm is explained briefly here. Suppose that we are going to model a system where $X$=($X_1$, $X_2$,…, $X_m$,.., $X_p$) is the vector of $p$ input variables and $Y$ is the output function of the system. Imagine, there are $n$ samples (observations) of the system so that the $k^{th}$ sample ($k$=1,…, $n$) is denoted by $x_k$=($x_{k1}$, $x_{k2}$,…, $x_{km}$,…, $x_{kp}$) and $y_k$. The principles of the MARS technique are based on the piecewise linear basis functions (BF) of the following forms:

$$BF1(x) = |x-t|_+ = \max(0, x-t) = \begin{cases} x-t & if\ x > t \\ 0 & if\ x \le t \end{cases}$$
$$BF2(x) = |t-x|_+ = \max(0, t-x) = \begin{cases} t-x & if\ x < t \\ 0 & if\ x \ge t \end{cases} \quad (1)$$

where, $t$ is called the *Knot* and $|.|_+$ means the positive part. These functions are also known as reflected pairs, mirror–image functions, or hockey stick functions and can be defined for each input variable $X_m$ at its observed values $x_{km}$ , $k$=1, 2,…, $n$ as:

$$BF1 = \max(0, X_m - x_{km})$$
$$BF2 = \max(0, x_{km} - X_m) \quad (2)$$

Consider the set $C$ involving all $2n\times p$ possible basis functions, i.e.,

$$C_i = \{ (X_i - t)_+ , (t - X_i)_+ \};\ t \in \{x_{1i}, x_{2i}, \dots, x_{ji}, \dots, x_{ni}\},\ i = 1, \dots, p \quad (3)$$

The MARS modeling technique is performed in a forward–backward stepwise fashion. In the forward step, a model consists of only the intercept being considered ($\hat{y} = \beta_0$), then MARS adds a reflected pair functions from $C$ to the model that minimizes the sum of squares of the residuals. That is, $n\times p$ models of the form $\hat{y} = \beta_{0j} + \beta_{1j}(X_i - x_{ji})_+ + \beta_{2j}(x_{ji} - X_i)_+ , j = 1, \dots, n$ are fitted and the one with the minimum $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is selected as the appropriate one. Suppose that the current model is denoted by model no. 1. In the next stage, each of the remaining ($n$–1)×$p$ reflected pairs in $C$ are candidates to add to the model no.1. In each stage, the coefficients are re–estimated by the least square method. In this way, the model no. 2 is formed by selecting paired functions that provide the minimum *SSE*. The process of adding the new terms can be repeated until the variation of *SSE* is not significant or the number of basis functions in the model is equal to the user defined maximum number of basis functions. This upper limit should be large enough to ensure that the correct model can be captured. At the end of the forward step, the resulting model has many redundant knots and clearly has poor performance to predict the test data. In other words, this is an over–fitted model. Hence the basis functions that contribute least to the overall fit must be removed from the model in the backward pruning step.

The backward step is started by the final model at the end of the forward step. Suppose that this model is denoted by $f_M$ including $M$ basis functions ($M$/2 pair basis functions) and one bias term. In each step of the backward procedure, one basis function is removed from the current model to provide models with $M$–1, $M$–2,…1, and 0 basis functions. If the resulting models are named by $f_{M-1}$, $f_{M-2}$,…, $f_1$, $f_0$, respectively, then each of these can be a candidate as the final MARS model. The criterion to choose the best model is generalized cross validation (GCV) (Hastie et al., 2009). This measure, for the $l^{th}$($l$=0,1,…,$M$–1) model, $f_l$, is defined as Equation (4):

$$GCV_l = \frac{SSE_l}{(1 - (vm_l/n))^2} \quad (4)$$

where, $m_l$ and $SSE_l$ are the number of basis functions and the sum of squares of residuals of the $l^{th}$ model in the deletion step, respectively. The user–defined $v$ is the smoothing parameter that is normally selected between 2 and 4. The model with the lowest GCV value is the final MARS model.

The only remaining question is that of which basis function should be removed in each deletion step. Suppose that $b_1$, $b_2$,…, $b_l$ are the basis functions in model $f_l$. The reduced model $f_{l-1}$ is formed by removing a basis function from the set $b_1$, $b_2$,…, $b_l$ whose removal will result in the smallest increase in $SSE_l$.

The above explanations are about the MARS technique without consideration of interaction among input variables. But MARS can also be implemented with consideration of interaction among input variables if interactive basis functions are defined in the set of possible basis functions ($C$). An interactive basis function can be defined by multiplication of the basis functions of different variables. For example, $I(X_1, X_2)$=max(0, $X_1$–$x_{k1}$)×max(0, $X_2$–$x_{s2}$) is an interactive basis function which defines the interaction between the two input variables $X_1$ and $X_2$ in a localized region where $X_1$>$x_{k1}$ and $X_2$>$x_{s2}$ ($x_{k1}$ and $x_{s2}$ are the $k^{th}$ and $s^{th}$ values of input variables $X_1$ and $X_2$, respectively). This means that different sub–regions of predictors might have different interaction patterns that can be expressed explicitly in MARS.

### 2.5. Algorithm of study

Multiple steps of the research algorithm are shown in Figure 3. In the first step, the hourly carbon monoxide concentration data for the year 2008 were collected from the air pollution stations. Then the new database was generated by converting the hourly data to daily data.

Working days in Iran are Saturday to Wednesday. Friday is the only official weekend holiday, but some companies, state offices and organizations don't work or work fewer hours than regular working days on Thursdays. Therefore, there are different traffic intensities on Fridays, Thursday and Saturdays–Wednesdays, and consequently different air pollution patterns and intensities. In the second step, a Friday, a Thursday and a Monday were selected and the hourly CO concentration data of the 16 monitoring stations for the selected three days (72 hours) in 2008 was transferred to GIS environment of ArcGIS software. Then, the different interpolation techniques (IDW, TPSS, Kriging and Cokriging) were implemented on these hourly maps and the parameters of the above techniques for each hour were calibrated separately. Then a Friday, a Thursday and a Monday were selected in each month in 2008 (36 days), the daily CO concentration data was calculated using the hourly CO measurements from the 16 monitoring stations and a similar interpolation procedure was implemented on the 36 different daily datasets.

Similarly, the datasets were imported to MATLAB for implementation of MARS2 (MARS with two input variables (longitude and latitude)) and MARS3 (MARS with three input variables (longitude, latitude and elevation)) and generated the spatial distribution models of CO in hourly and daily scales (Step 3).

In the fourth step, the optimum values of the parameters of the typical interpolation techniques were determined using the cross–validation (CV) technique (leaving out the observation points in each map one at a time).

Similarly, the optimum models for MARS2 and MARS3 were determined using the generalized cross validation (GCV) technique (step 5).

Mean Absolute Percentage Error ($MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|$; where, $n$ is number of observations. $y_i$ and $\hat{y}_i$ are the measured and estimated CO concentrations, respectively) was utilized as evaluation criteria for all of the studied techniques and the best method for spatial distribution modeling of hourly and daily CO concentration was determined (step 6). In the last step (step 7), the hourly, daily and annual spatial distribution maps of carbon monoxide concentration in Tehran were generated using the best method and the pollution condition in different parts of Tehran was discussed.

## 3. Results and Discussion

The input variables for the IDW, TPSS and Kriging methods are latitude and longitude data and the output is CO concentration. Cokriging not only utilizes longitude and latitude as input variables, but also the elevation data as auxiliary input variables. The optimum conditions for different interpolation techniques were determined by the cross–validation technique in the ArcGIS environment and have been presented in Sargazi et al. (2011). The cross validations showed that the optimum weight exponent for hourly and daily interpolation by the IDW technique are 2.1 and 1.1, respectively. These weight exponents are in the range of weight exponents, presented by de Mesnard (2013), for the dispersion of pollution plumes. In addition, the optimum smoothing parameter for the TPSS technique is equal to $1.1\text{x}10^{-11}$ and the best semi–variogram model for the Kriging and Cokriging is spherical model.

The MPAE of each optimized method for 36 daily and 72 hourly datasets was calculated. Tables 2 and 3 show the MAPE values of the IDW, Kriging, Cokriging, MARS2 and MARS3 techniques for spatial modeling of hourly and daily CO concentration in Tehran. MAPE values for typical interpolation techniques ranged from 79.7 (Cokriging technique) to 110.5 (TPSS technique) and from 75.9 (Cokriging technique) to 109 (TPSS technique) for hourly and daily interpolations, respectively.

The Cokriging technique is the best typical interpolation technique for the spatial modeling of CO in Tehran because its MPAE for both hourly and daily modeling is lower than the other studied interpolation techniques (Tables 2 and 3). It has better results than the IDW, TPSS and Kriging techniques, because it uses not only the latitude and longitude data as input variables but also implicitly uses the elevation data.

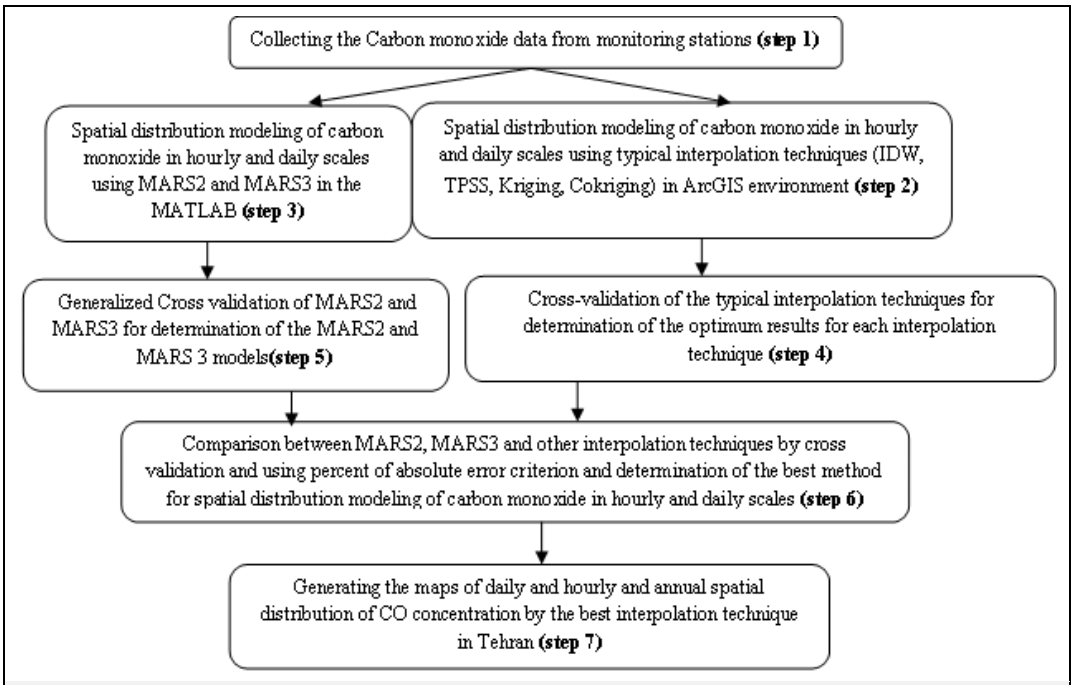


***Figure 3.*** *The main steps of research algorithm.*

***Table 2.*** *Mean absolute percentage error (MAPE) for hourly spatial distribution modeling of CO concentrations in Tehran for different interpolation techniques*

| Interpolation Technique | TPSS | IDW | Kriging | Cokriging | MARS2 | MARS3 |
|---|---|---|---|---|---|---|
| MPAE | 110.5 | 85.3 | 81.6 | 79.7 | 73.4 | 69 |

***Table 3.*** *Mean absolute percentage error (MAPE) for daily spatial distribution modeling of CO concentratiosn in Tehran for different interpolation techniques*

| Interpolation Technique | TPSS | IDW | Kriging | Cokriging | MARS2 | MARS3 |
|---|---|---|---|---|---|---|
| MAPE | 109 | 95.2 | 86.2 | 75.9 | 65.9 | 64.4 |

MPAEs of MARS2 for hourly and daily modeling are 73.4 and 65.9, respectively. Although the MARS2 utilized only the two input variables, it presented better results than the Cokriging technique, which used the three input variables. The better results of MARS2 over the Cokriging technique in the daily and hourly scales implies the capability of MARS for spatial distribution modeling. It was mentioned previously that MARS is able to consider the interaction among the input variables. Thus, an example is presented to show how MARS considers the interaction among input variables. For spatial modeling of hourly CO concentration in Tehran by MARS2 on September 11th, 2008 (04:00 am), two input variables—Longitude ($x_1$) and Latitude ($x_2$)—are employed. The final MARS2 model is as in Equation (5):

$$CO = 1.3 + 227.2 \times BF1 + 2350.7 \times BF2 - 272 \times BF3 \quad (5)$$

where, CO is the hourly CO concentration (ppm) on September 11th, 2008 (04:00 am), $BF1$=max(0,35.43–$x_2$), $BF2$=max(0,35.43–$x_2$) ×max(0, $x_1$–51.25) and $BF3$=max(0,35.42–$x_2$).

It can be noted that $BF1$ and $BF3$ show the one–variable influences on the CO concentration and $BF2$ exhibits the effects of interaction of two input variables on the CO concentration. In $BF2$, when $x_1$>51.25 and $x_2$<35.4, the interactive effect appears explicitly. In addition, the intensity of $BF2$ effects on CO concentration is about 10.3 (2 350.7/227.2) and 8.6 (2 350.7/272) times more than $BF1$ and $BF3$, respectively. For instance, the basis functions and the CO concentration in a location in Tehran with $x_1$=51.2 and $x_2$=35.42 is estimated below:

$BF1$=0.009, $BF2$=0.00045, $BF3$=0 and CO=1.3+227.2×0.009+ 2 350.7×0.00045–272×0=4.4 ppm.

MARS3 provided better results than the MARS2. It is easy to judge that MARS3 is the best method for the retrieval of spatial distribution of CO in hourly and daily time scales. Spatial modeling using MARS is very easy and straightforward. MARS spatial distribution is performed by definition of the model input parameters without any pre–processing of data. MARS can perform multivariate spatial modeling and the number of input variables is not a limitation for MARS.

The performance of interpolation techniques is related to characteristics of the spatial variation of the pollutants and the characteristics of monitoring stations (e.g. sampling density and distribution) (Briggs et al., 1997). The number of air pollution monitoring stations is very low (16 stations for 22 zones) and the distribution of stations is not suitable for spatial estimation; hence the level of errors, presented in Tables 2 and 3, is relatively high. Consequently, mapping of CO provides a limited picture of spatial patterns of CO in Tehran, but it is suitable for mapping of relative CO levels in different zones in Tehran. Therefore, the daily CO concentration maps in 2008 were generated by the MARS3 method. Then the daily maps for 2008 were averaged and the map of average annual CO concentration in Tehran was generated (Figure 4). This figure exhibits that the southern, southeastern and eastern parts of Tehran are more polluted than the other parts. The most polluted part of Tehran is the southeastern part. The annual CO concentration in each zone was averaged and then ranked. The ranked zones with their average CO concentrations have been presented in Table 4. The results show that the 22nd, 21st and 5th, zones (western parts of Tehran) have a lower CO concentrations than the other parts of Tehran and the 13th, 14th and 15th zones (southeastern part of Tehran) are the most polluted regions in Tehran. Figure 4 implies that the central, eastern and southeastern parts of Tehran (zones 6, 12, 13, 14 and 15) need more effective air pollution mitigation programs and strategies. The suitable mitigation strategies for Tehran are improvement of new vehicles and motorcycles built in Iran (more energy efficient and less polluting), discarding old vehicles, improvement of fuel quality and adopting appropriate alternatives, inspection and maintenance of vehicles, expanding the public transportation system as well as its delivery, a traffic management system and training programs, and promotion of public awareness and education about air pollution conditions and its health effects (Atash, 2007). In addition, it is proposed that the most vulnerable people (infants and young children, the elderly, and people with respiratory conditions and heart diseases) avoid heavy traffic into these polluted zones.

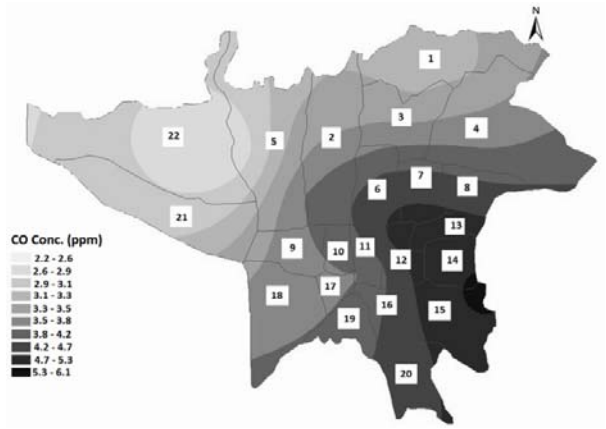


***Figure 4.*** *The mean annual CO concentrations in 22 zones of Tehran in 2008, generated by MARS3 method.*

***Table 4.*** *Ranking of average annual CO concentrations in 22 zones of Tehran (The position of zones have been presented in Figure 2)*

| Zone | CO Conc. (ppm) | Zone | CO Conc. (ppm) |
|---|---|---|---|
| 22 | 2.61 | 19 | 3.65 |
| 21 | 2.85 | 11 | 3.79 |
| 5 | 2.98 | 16 | 3.86 |
| 1 | 3.1 | 8 | 3.94 |
| 3 | 3.36 | 20 | 4.1 |
| 2 | 3.4 | 7 | 4.12 |
| 9 | 3.39 | 6 | 4.19 |
| 18 | 3.43 | 12 | 4.26 |
| 17 | 3.43 | 13 | 4.44 |
| 4 | 3.44 | 14 | 4.57 |
| 10 | 3.63 | 15 | 4.58 |

## 4. Conclusions

The results showed that the Cokriging technique has better performance than the other studied typical interpolation techniques. The MARS2 and MARS3 presented better results than the Cokriging technique in hourly and daily scales. This shows that MARS has capability for spatial modeling. In addition, MARS3 is the best method for spatial modeling of CO in Tehran and its MAPEs for hourly and daily CO estimation are 69 and 64.4, respectively. The spatial modeling using MARS is easy and straightforward. According to the results and capabilities of MARS, it can be introduced as an appropriate method for spatial distribution of air pollutants. The map of average annual CO concentration in Tehran showed that the central, eastern and southeastern zones of Tehran are the most polluted regions and mean annual CO concentration in these zones is between 4.2 and 4.6 ppm. These parts of Tehran (zones 6, 12, 13, 14 and 15) need appropriate mitigation programs and strategies.

## Acknowledgments

The authors thank Ms. Chelsea Wright for her valuable proof-reading of this paper.

## References

Abbaspour, M., Soltaninejad, A., 2004. Design of an environmental assessment model on the effect of vehicle emission in greater Tehran on air pollution with economic sensitivity. *International Journal of Environmental Science & Technology* 1, 27–38.

Atash, F., 2007. The deterioration of urban environments in developing countries: Mitigating the air pollution crisis in Tehran, Iran. *Cities* 24, 399–409.

Beaulant, A.L., Perron, G., Kleinpeter, J., Weber, C., Ranchin, T., Wald, L., 2008. Adding virtual measuring stations to a network for urban air pollution mapping. *Environment International* 34, 599–605.

Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K., Briggs, D.J., 2009. Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the Total Environment* 407, 1852–1867.

Bel, L., 2004. Non parametric variogram estimator. Application to air pollution data, in *geoENV IV—Geostatistics for Environmental Applications*, edited by Sanchez–Vila, X., Carrera, J., Gomez–Hernandez, J.J., Springer, Netherlands, pp. 29–40.

Briggs, D.J., 1992. Mapping environmental exposure, in *Geographical and Environmental Epidemiology: Methods for Small–Area Studies*, edited by Elliott, P., Cuzick, J., English, D., Stern, R., Oxford University Press, Oxford, pp. 158–176.

Briggs, D.J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., van Reeuwijk, H., Smallbone, K., van der Veen, A., 1997. Mapping urban air pollution using GIS: A regression–based approach. *International Journal of Geographical Information Science* 11, 699–718.

Chan, C.K., Yao, X., 2008. Air pollution in mega cities in China. *Atmospheric Environment* 42, 1–42.

de Mesnard, L., 2013. Pollution models and inverse distance weighting: Some critical remarks. *Computers & Geosciences* 52, 459–469.

Finkelstein, M.M., Jerrett, M., DeLuca, P., Finkelstein, N., Verma, D.K., Chapman, K., Sears, M.R., 2003. Relation between income, air pollution and mortality: A cohort study. *Canadian Medical Association Journal* 169, 397–402.

Friedman, J.H., Rooster, C.B., 1995. An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research* 4930, 197–217.

Friedman J.H., 1991a. Multivariate adaptive regression splines. *The Annals of Statistics* 19, 1–141.

Friedman J.H., 1991b. Estimating Functions of Mixed or Ordinal and Categorical Variables using Adaptive Splines, Technical Report, LCS108, Department of Statistics, Stanford University.

Givehchi, R., Arhami, M., Tajrishy, M., 2013. Contribution of the Middle Eastern dust source areas to $PM_{10}$ levels in urban receptors: Case study of Tehran, Iran. *Atmospheric Environment* 75, 287–295.

Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd Edition, Springer Verlag, New York.

Heikkinen, R.K., Luoto, M., Kuussaari, M., Toivonen, T., 2007. Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landscape and Urban Planning* 79, 347–357.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land–use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* 42, 7561–7578.

Hosseinpoor, A.R., Forouzanfar, M.H., Yunesian, M., Asghari, F., Naieni, K.H., Farhood, D., 2005. Air pollution and hospitalization due to angina pectoris in Tehran, Iran: A time–series study. *Environmental Research* 99, 126–131.

Ionescu, A., Candau, Y., Mayer, E., Colda, I., 2000. Analytical determination and classification of pollutant concentration fields using air pollution monitoring network data – Methodology and application in the Paris area, during episodes with peak nitrogen dioxide levels. *Environmental Modelling & Software* 15, 565–573.

Jerrett, M., Burnett, R.T., Ma, R.J., Pope, C.A., Krewski, D., Newbold, K.B., Thurston, G., Shi, Y.L., Finkelstein, N., Calle, E.E., Thun, M.J., 2005. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16, 727–736.

Jerrett, M., Burnett, R.T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., Brook, J.R., 2001. A GIS – Environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environment and Planning A* 33, 955–973.

Leathwick, J.R., Elith, J., Hastie, T., 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* 199, 188–196.

Leathwick, J.R., Rowe D., Richardson, J., Elith, J., Hastie, T., 2005. Using multivariate adaptive regression splines to predict the distribution of New Zealand's freshwater diadromous fish. *Freshwater Biology* 50, 2034–2052.

Majd, A., Chehregani, A., Moin, M., Gholami, M., Kohno, S., Nabe, T., Shariatzade, M.A., 2004. The effects of air pollution on structures, proteins and allergenicity of pollen grains. *Aerobiologia* 20, 111–118.

Mulholland, J.A., Butler, A.J., Wilkinson, J.G., Russell, A.G., Tolbert, P.E., 1998. Temporal and spatial distributions of ozone in Atlanta: Regulatory and epidemiologic implications. *Journal of the Air & Waste Management Association* 48, 418–426.

Munoz, J., Felicisimo, A.M., 2004. Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* 15, 285–292.

Naddafi, K., Hassanvand, M.S., Yunesian, M., Momeniha, F., Nabizadeh, R., Faridi, S., Gholampour, A., 2012. Health impact assessment of air pollution in megacity of Tehran, Iran. *Iranian Journal of Environmental Health Science & Engineering* 9, art. no. 28.

Patil, U.D.A.Y., Raavan, S., Kaushal, A., 2003. GIS based air pollution surface modeling. *GIS Development* 7, 45–47.

Pearce, J.L., Rathbun, S.L., Aguilar–Villalobos, M., Naeher, L.P., 2009. Characterizing the spatiotemporal variability of $PM_{2.5}$ in Cusco, Peru using kriging with external drift. *Atmospheric Environment* 43, 2060–2069.

Pikhart, H., Bobak, M., Gorynski, P., Wojtyniak, B., Danova, J., Celko, M.A., Kriz, B., Briggs, D., Elliott, P., 2001. Outdoor sulphur dioxide and respiratory symptoms in Czech and Polish school children: A small–area study (SAVIAH). *International Archives of Occupational and Environmental Health* 74, 574–578.

Rashidi, M., Massoudi, M.S., 1980. A study of the relationship of street level carbon monoxide concentrations to traffic parameters *Atmospheric Environment* 14, 27–32.

Raub, J.A., Mathieu–Nolf, M., Hampson, N.B., Thom, S.R., 2000. Carbon monoxide poisoning – A public health perspective. *Toxicology* 145, 1–14.

Sargazi, S., Taheri Shahraiyni, H., Habibi–Nokhandan, M., Sanaeifar, M., 2011. Application of GIS for the modeling of spatial distribution of air pollutants in Tehran. *Proceedings of SPIE 8181, Earth Resources and Environmental Remote Sensing/GIS Applications II*, 81810I, October, 2011, Prague, Czech Republic.

Singh, V., Carnevale, C., Finzi, G., Pisoni, E., Volta, M., 2011. A cokriging based approach to reconstruct air pollution maps, processing measurement station concentrations and deterministic model simulations. *Environmental Modelling & Software* 26, 778–786.

Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C.J., 2009. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety* 94, 1735–1763.

Ung, A., Weber, C., Perron, G., Hirsch, J., Kleinpeter, J., Wald, L., Ranchin, T., 2001. Air pollution mapping over a city – virtual stations and morphological indicators. *Proceedings of 10th International Symposium of Transport and Air Pollution*, September 2001, Boulder, Colorado, United States.

Vafa–Arani, H., Jahani, S., Dashti, H., Heydari, J., Moazen, S., 2014. A system dynamics modeling for urban air pollution: A case study of Tehran, Iran. *Transportation Research Part D–Transport and Environment* 31, 21–36.

Webster, R., Oliver, M., 2001. *Geostatistics for Environmental Scientists*, John Wiley & Sons, Chichester.

Ziaei, S., Nouri, K., Kazemnejad, A., 2005. Effects of carbon monoxide air pollution in pregnancy on neonatal nucleated red blood cells. *Paediatric and Perinatal Epidemiology* 19, 27–30.