

COMPUTATIONAL INTELLIGENCE ESTIMATION OF NATURAL BACKGROUND OZONE LEVEL AND ITS DISTRIBUTION FOR AIR QUALITY MODELLING AND EMISSION CONTROL

Herman Wahid^{1*}, Q. P. Ha², and H. Nguyen Duc³

¹ Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Malaysia/ Faculty of Engineering and IT, University of Technology, Sydney, Australia

² Faculty of Engineering and IT, University of Technology, Sydney, Australia

³ Department of Environment, Water and Climate Change NSW, Australia

* Corresponding author (hbinwahi@eng.uts.edu.au)

ABSTRACT: Background ozone, known as the ozone that occurs in the troposphere as a result of biogenic emissions without photochemical influences, has a close relationship with human health risk. The prediction of the background ozone level by an air quality model could cover a wider region, whereas a measurement method can only record at monitoring sites. The problem is that simulation with deterministic models is quite tedious because of the nonlinear nature of some particular chemical reactions involved in the pollutant formulation. In this work, we present a reliable method for determination of the background ozone using the ambient measurement data. Our proposed definition can be used to determine the background level at any part of the globe and in any seasons without relying on data obtained at remote sites. A statistical model approach will be used for the estimation of the background ozone concentration, and a method for extrapolating the site data will be utilised to approximate the spatial distribution on the region. The proposed method will be applied in the Sydney basin to evaluate its effectiveness in background ozone determination. The results show the advantage of the proposed approach as a globally generic and computationally efficient way for the background ozone estimation with a reasonable accuracy.

Keywords: Background Ozone, Air Quality Modelling, Radial Basis Function, Spatial Distribution

1. INTRODUCTION

Natural background ozone level (BOL) can be referred to as the concentration of ozone (O_3) that is formed from only natural emissions (i.e. biogenic or geogenic contaminations) of the precursors in the absence of the local anthropogenic influences. Biogenic emissions (e.g. oxides of nitrogen, NO_x) are generated from the natural sources such as plants, trees, and wildfires, whereas geogenic emissions may come from lightning and soil conditions.

Unfortunately, contributions of the background ozone at most places may include a combination of local natural emissions, photochemical effects as well as long-range transports from other regions. Thus, it is difficult to distinguish between the ozone that is produced by the

anthropogenic and natural effects. One method to determine the BOL is by measuring the ozone concentration at the clean pristine sites of a rural area, considered as a clean environment. This method was widely used in the literature, see e.g., Saitanis [1] and Oltmans et al. [2]. However, Matveev et al. [3] reported that measurements at remote rural sites do not necessarily represent background levels since local pollutants can sometimes be re-circulated.

The US EPA refers the BOL as the ozone concentrations that would occur in United States (US) in the absence of anthropogenic emissions in Continental North America, but including the contributions from the natural sources everywhere in the world [4]. The BOL in the US is estimated by using modelling software called GEOS-

CHEM, with its level being found to vary from 30 to 50 ppb during the typical summer time.

By using a probability distribution analysis, Altshuller and Lefohn [5] related the ozone probability distribution of the data measured at the clean sites with the background ozone level but the relation is not straightforward. By correlations with reactive nitrogen dioxides NO_y (NO_x and its oxidation products), they estimated a daytime BOL of 35 ± 10 ppbv in summer, amounting to almost half of the new air quality standard. Duc and Azzi [6] used the ambient measurements to estimate the non-photochemical BOL by taking the average value of nighttime and early morning hourly ozone concentrations (i.e. from 19:00pm to 8:00am the next morning), when nitrogen oxides (NO_x) are empirically considered as absent, assuming the photochemical processes occur only during daytime. However, the duration time from the definition is too specific and depending on different locations as well as seasons. Furthermore, it is very rare to get measurements with the absence of NO_x values at every hour during that time range. Motivated by this idea, we will refine the definition to be more generic that could be used globally at any period of the year.

In this work, we are addressing two issues. The first is to refine the BOL definition based on [6]. Next, we use a neural network based metamodel to construct the approximate model of BOL and to investigate the long-term trend of BOL in the Sydney basin.

2. METHODOLOGY

2.1 Background ozone level determination

In this work, the determination of BOL is motivated by the definition given in [6]. The first objective is to estimate the start- and end-time of a day period in which the ozone concentration can be used for the determination of BOL. For the case study, the investigation is based on the ambient measurement data collected at various monitoring stations across the Sydney basin in Australia.

By analysing the diurnal distribution of the ozone concentration (O_3 , in ppb), nitrogen oxide concentration ($\text{NO}_x = \text{NO} + \text{NO}_2$, in ppb) and hourly temperature (in $^\circ\text{C}$), we found a pattern which can be illustrated in Fig. 1. The

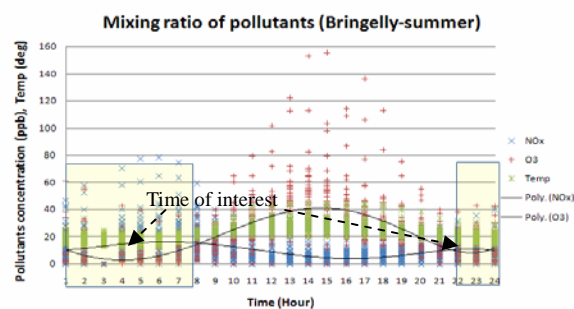


Fig. 1 Diurnal distributions of O_3 and NO_x at Bringelly site during summer 2003.

figure shows the mixing ratio of O_3 and NO_x concentrations against the hourly time at Bringelly station, located in the West of Sydney by using two months' data in summer 2003. Two profiles representing each pollutant are plotted by using a higher-order polynomial line via the least square values for every hour. From the figure, we can see that the ozone starts to form in the morning time and will increase with the temperature. During that time, the NO_x concentration decreases, as contributing to the ozone formulation, and will rise again in the evening. It is suggested the two intersections of O_3 and NO_x concentration profiles in the morning and evening are the limits in the day period to be used for the determination of the non-photochemical BOL. This is meant to exclude the part contributing to the ozone concentration, present during the daytime due to photochemical reactions, which is mainly caused by anthropogenic emissions. These patterns and the corresponding intersection points seem to be different at every site and season. For example, during the winter season, generally the period for the ozone production appears to be shorter during daytime. Thus the time interval to be considered for BOL in winter is longer than in other seasons.

For the generalization purpose, we could repeat the same method by considering data from every site in the region to get the average points, e.g. the BOL definition for the Sydney basin can be determined by considering ozone concentration data from 21:00pm to 08:00am the next morning (i.e. as summer is the period of interest). It shows that the start-time begins later than by the empirical definition [6], i.e. 19:00pm.

The next aim is to use the ozone concentration data, collected at monitoring sites, in the specified time ranges to estimate the BOL to validate the proposed method for BOL determination. In [6], from the ambient measurement data, the stipulation of no nitrogen dioxide present (i.e.) for at least 2 hours consecutively to ensure no ozone reaction with nitrogen oxides is rather unlikely in an urban area. As the zero concentrations of nitric oxide and nitrogen dioxide (i.e. $[\text{NO}]=0$, $[\text{NO}_2]=0$) are difficult to obtain via measurements, it would be hard to determine accurately the non-photochemical BOL.

Fig. 2(a) shows the linear relationships of the mixing ratio of $\text{NO}+\text{NO}_2+\text{O}_3$ with NO_x during nighttime by using the regression analysis. From the idea proposed by Clapp and Jenkin [7], the NO_x -independent and NO_x -dependent contribution appears in the plot. The linear equation is given as follows:

$$[\text{O}_3+\text{NO}_x] = m \cdot [\text{NO}_x] + c \quad \dots(1)$$

where m is the slope value which depends on the NO_x local contribution and c is the intercept value. At the intercept point, $[\text{NO}_x]=0$, we can consider this intercept as the average daily non-photochemical BOL because we only use the nighttime data to develop the regression profile which predominantly excludes the photochemical effect. The variation of hourly BOL is much dependant on the NO_x concentration from local anthropogenic contributions that still exist from the daytime. To simplify the procedure, we can also get the daily BOL value by plotting the oxidant profile against NO_x data and find the intercept value from the regression analysis, as depicted in Fig. 2(b). Next, the hourly BOL can be determined by extrapolating the main curve to the respective hourly data and getting their intercept values, as illustrated therein.

Hence, to summarize this section, we can form a generic function for the new BOL definition which probably can be used globally at various geographical locations and in different seasons, as given in the following equation:

$$\text{BOL}_{NP}(\text{O}_3, \text{NO}_x, t') = \text{intercept}(\text{O}_3(t') : \text{NO}_x(t')), \quad \dots(2)$$

where BOL_{NP} is the non-photochemical BOL, intercept is the function of the intercept of the regression line, and t' is the period of interest determined by:

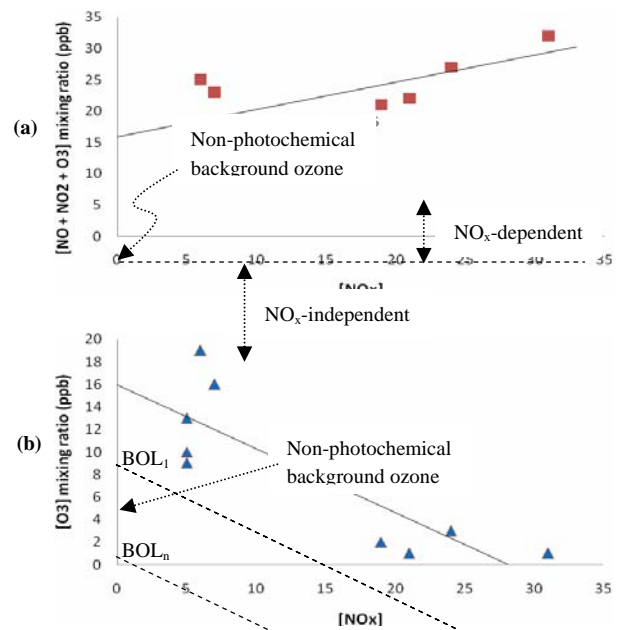


Fig. 2 Non-photochemical BOL determination: (a) daily average shown by intercept of the linear regression line, (b) hourly average shown by intercept of the dashed lines.

$$t' = \text{interval}(\text{polynomial}(\text{NO}_x'), \text{polynomial}(\text{O}_3')), \quad \dots(3)$$

where interval is the function to find the two intersection points of the higher order polynomial lines, NO_x' is the mixing ratio of the NO_x component, and O_3' is the mixing ratio of the O_3 component for the season in consideration.

2.2 Radial basis function neural network (RBFNN)

The RBFNN structure is basically similar to a feed-forward artificial neural network (ANN) architecture which consists of an input layer, a hidden layer and an output layer. The neurons in the hidden layer work are the processing elements to perform a non-linear transformation of the input data to approximate the output data. The main difference with other ANNs is that this structure uses a radial function such as Gaussian as the activation function, and the number of the hidden neurons can be adjusted appropriately to meet the best convergence subject to a prescribed goal. Another advantage is that the networks can be designed in a considerably small amount of time as compared to others due to the process of getting the weights associated with a large number of hidden nodes by using linear regression rather than the iteration process such as in the steepest descent. Therefore, many authors

suggested that RBFNN offers good performance in terms of accuracy, efficiency and simplicity (e.g. in [8-9]). The RBFNN structure available in the Matlab neural network toolbox [10] is used in the current analysis.

2.3 The estimation of BOL using RBFNN model

The model development covers for two aims, namely, to predict the future concentration of the background ozone using data collected from monitoring stations for the model training, and to extrapolate the prediction results from each site to estimate the spatial distribution for a wider area.

To construct the model in the first task, we will use the ambient measurement data for pollutants recorded at the monitoring sites. A proper selection of the model's input variables is essential to obtain the best approximation function for modelling. For the determination of BOL, several combinations of the input variables have been investigated in [11] and [12], which showed that time information, pollutants' precursor data, ozone data and meteorological data have a major influence on the model performance. However, each site was separately treated with a different model for the network training and validation.

In this work, we will construct a generic model that covers several sites in the region. To achieve such model, we have to come up with a proper training technique that can be mixed between the sites. For the training dataset, we use the following variables as the inputs: the x and y locations as the site's identification, the NO concentration, the NO₂ concentration, the O₃ concentration, the ambient temperature (TEMP), the wind direction (WDR), and the wind speed (WSP). The considered input datasets are the hourly data at the specified time period of interest for the BOL determination which has been described in Section 2.1. Furthermore, the BOL defined in (2) is used as the target output of the network. All input-output variables are normalized using their minimum and maximum values so that they are in the same range of the radial basis function used. After the learning process by RBFNN is finished and the accuracy obtained with some test sets is satisfactory, the network can be used for the prediction with other available data for the following years.

The spatial distribution model is important as it can be used to estimate the background ozone concentration at locations other than the monitoring sites. Extrapolation of the concentration information at a site to used to estimate the distribution around them. In this work, the extrapolation method by Sandwell [13] is adopted, in which part of the featuring code is compatible with Matlab software (i.e. 'v4' interpolation). For extrapolation, the inputs are x and y location of monitoring stations, and the yearly average BOL, which is predicted from RBFNN model. The output results will be the extrapolated grid data at every 2km×2km distance in the region surrounding the monitoring sites in consideration.

3. APPLICATION IN THE SYDNEY BASIN

3.1 Validity of the proposed method

To verify the new definition of non-photochemical BOL and evaluate the proposed method for BOL determination, we test the consistency with several monitoring sites and compare our method with several approaches. First, we try to determine the BOL based on daily mean ozone concentration data collected during summer 2003 at a semi-pristine site in Bargo located in the greater south-west of Sydney which may be considered as a clean site. Fig. 3 shows the difference of the BOL obtained from the proposed method and the mean 24-hour ozone concentration (i.e. including the daytime ozone) measured for about one month. As can be seen, the levels are about similar in most of the evaluated days. Some differences exist at the event days, which are expected due to the site is not completely a clean site. We assume that some high ozone concentration during the daytime is also affected by the photochemical ozone which is transported to this area instead of its natural processes. The average background ozone concentration given by our method is 18.4 ppb compared to 20.7 ppb as obtained from by measurement data at the site.

Next, we summarise the variations of the BOL at various sites in Sydney as listed in Table 2. The results point out that the East and South-West of Sydney in general exhibit a higher BOL than other areas, in which Oakdale and Randwick sites are the highest on average. In average, the

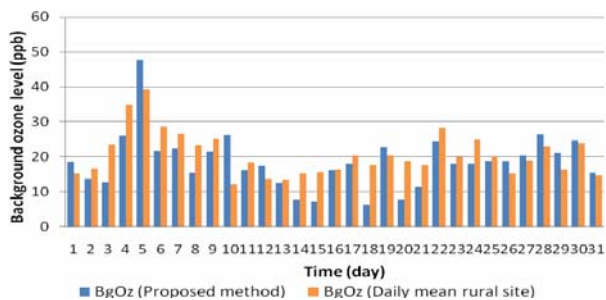


Fig. 3 Comparison of background ozone level between the proposed method and mean concentration at semi-pristine site.

Table 2 Variations of background ozone level at Sydney region in summer 2003 based on proposed definition.

Region	Site	Min	Max	Mean	Duc & Azzi [6]
East	Earlwood	1.8	31.4	13.5	20.0
	Randwick	8.7	48.5	20.5	22.5
	Rozelle	2.0	45.8	16.0	19.0
North-west	Richmond	4.9	37.0	17.0	19.0
	Vineyard	6.7	39.1	16.7	19.0
West	St Marys	1.7	39.0	12.9	18.5
South-west	Bringelly	2.3	51.2	16.1	20.0
	Oakdale	11.0	42.0	21.1	-
Average		4.9	41.8	16.7	19.7

level is 3 ppb higher than the proposed definition due to the different time range used, and they only considered data with the absence of NO rather than NO_x. Table 2 also shows the BOLs for 2003 as reported in [6]. Generally, the BOL in the Sydney basin ranges from 5 to 42 ppb with some incidences occurring to the maximum of 51 ppb. This result is consistent with the analysis of Fiore et al. [14], which found that the natural BOLs were typically between 10 to 25 ppb with some incidences between 40 to 50 ppb in western US by using GEOS-CHEM modelling software.

Furthermore, we compare the hourly BOL corresponding to other air pollutant concentrations. Fig. 4 shows an example of hourly data for BOL, ozone concentration and NO_x concentration. We found that the concentration level is similar when the NO_x value is equal to zero, which prompts to the absence of photochemical reactions. Thus, the ozone concentration at those points can be considered as BOL, which is in line with the proposed definition.

3.2 Background ozone distribution

For the BOL model development, we utilise the hourly data of the pollutant concentrations and meteorological data which have been carefully recorded at 10 monitoring stations in the Sydney region. The data cover six years

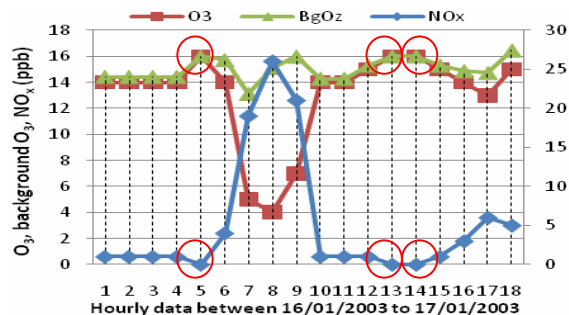


Fig. 4 Hourly data at Randwick station showing BOL equal to ozone concentration when NO_x=0. (The right axis represents NO_x scale in ppb).

from 2003 to 2008 in which summer is chosen as the season of interest. For the Sydney region, we use hourly data from 21:00pm to the 08:00am the next morning for the determination of BOL.

We prepared two groups of data, which were for the training and testing purposes. For the training dataset, the input and target data were taken partly from the summer 2004 from every site, whereas the remaining data from 2003 to 2008 can be used as the testing datasets. By using RBFNN with an initial setting of 1.0 for the spread and 0.01 for the MSE goal, the model is constructed with 5 hidden neurons.

Figure 5 shows an example for the reliability of the model to predict the BOL with a reasonably good result. The six-year prediction results for the other sites are depicted in Fig. 6, which correspond to the yearly BOL profiles. The results point out that there is an upward tendency at each site from 2003 to 2005 which is in line with the results reported in [6]. It is shown also that the patterns start to decrease after 2005 until 2008, and reach a level below 15 ppb. However, the BOL remains high at Randwick site, which may require a further investigation.

An example of the spatial distribution of the BOL is

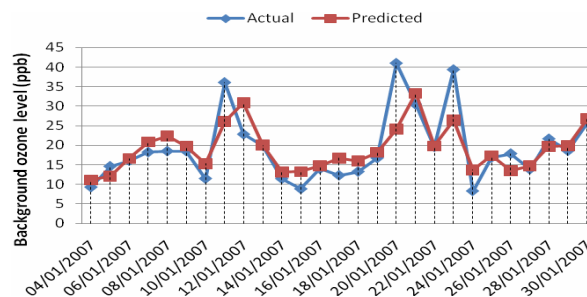


Fig. 5 Comparison of daily BOL in summer between RBFNN predictions and actual values (at Bringelly site).

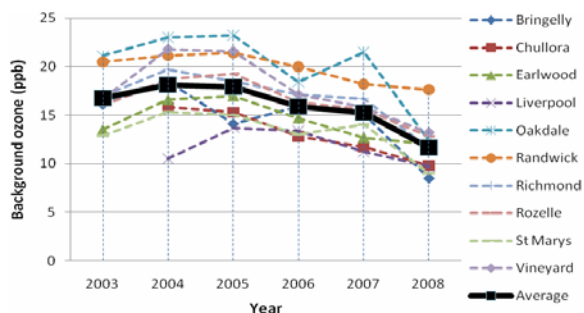


Fig. 6 BOL profiles in the Sydney basin.

shown in the contour image of Fig. 7, where the extrapolation data cover the frame of 250 to 350 km easting and 6210 to 6310 km northing, using the Australian map grid (AMG) coordinates. The grid data outputs are extracted from the spatial distribution and several site results are listed in Table 3. The results from other sites indicate the similar BOL growth until 2005 and a downward tendency after that year. A comparison at Macarthur site between predicted and recorded BOL also reveals the same remark on the BOL distribution over the period. The results indicate that the region around Liverpool has the lowest BOL, which suggests that this area is not affected by the transportation of high ozone concentration from the urban area in the south-east to the semi-rural area in the west of Sydney.

4. CONCLUSION

A new method for determination of background ozone level has been presented in this paper. The validity and reliability of the proposed definition has been evaluated in several ways. Notably, the approach is generic in determining BOL according to the period of interest for non-photochemical activities. Thus, it can be applied at any location and in any season. Using this approach, we have developed a regional computational intelligence model for estimation of the long-term BOL using neural networks. Coordinate location of the sites, nitrogen oxides and meteorological data as the input, and non-photochemical background ozone as the target output. Using data collected at Sydney's monitoring sites for several years, the model is capable of predicting the long-term BOL with fair accuracy. Once the BOL at a site has been predicted, its spatial distribution can be constructed to estimate the BOL for the whole region. For accuracy improvement, a higher

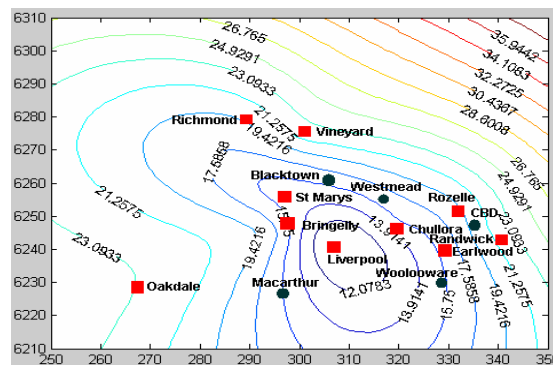


Fig. 7 Spatial distribution of Sydney BOL in 2004 (squares mark sites used in the trained model while circles mark the predicted locations).

Table 3 Trends of the BOL at other sites than in the model which extracted from the spatial prediction.

Sites	2003	2004	2005	2006	2007	2008
Blacktown	13.0	16.1	16.8	13.3	13.0	10.2
Macarthur	17.8	16.8	16.6	16.9	15.2	10.9
Sydney CBD	16.5	18.9	19.4	16.9	15.7	13.7
Westmead	13.0	16.7	17.0	13.4	12.8	10.3
Woolooware	16.6	15.2	17.6	16.7	13.3	15.3
Macarthur (actual)	-	-	15.7	15.9	15.0	-

number of monitoring stations may be required to yield a better interpolation within the sites as well as better extrapolation for areas surrounding the sites.

REFERENCES

- [1] C.J. Saitanis, "Background ozone monitoring and phytodetection in the greater rural area of Corinth-Greece," *Chemosphere*, Vol. 51(9), 2003, pp. 913-923.
- [2] S.J. Oltmans, A.S. Lefohn, J.M. Harris, and D.S. Shadwick, "Background ozone levels of air entering the west coast of the US and assessment of long-term changes," *Atmospheric Environment*, Vol. 42(24), 2008, pp. 6020-6038.
- [3] V. Matveev, M. Luria, D. Alper-Siman Tov, and M. Peleg, "Long range transportation of air pollutants from Europe towards Israel," *Israeli Journal of Earth Sciences*, Vol. 51, 2002, pp. 17-28.
- [4] U.S. Environmental Protection Agency, "Air quality criteria for ozone and related photochemical oxidants (final)", Research Triangle Park, NC: Office of Research and Development, 2006 EPA/600/R-05/004af.
- [5] A.P. Altshuller and A.S. Lefohn, "Background ozone in the planetary boundary layer over the United States," *Journal Air Waste Management Assoc.*, Vol. 46, 1996, pp. 134-141.

- [6] H. Duc and M. Azzi, "Analysis of background ozone in the Sydney basin," *Proc. 18th IMACS & MODSIM Congress, Cairns Australia*, 2009, pp. 2307-2313.
- [7] L.J. Clapp and M.E. Jenkin, "Analysis of the relationship between ambient level of O₃, NO₂ and NO as a function of NO_x in the UK," *Atmos. Env.*, Vol. 35, 2001, 6391-6405.
- [8] R.G. Regis and C.A. Shoemaker, "Constrained global optimisation of expensive black box functions using radial basis functions," *J. Global Optimisation*, Vol. 31(1), 2005, pp. 153-171.
- [9] G. Poshal and P. Ganesan, "An analysis of formability of aluminium preforms using neural network," *J. Material Processing Technology*, Vol. 205, 2008, pp. 272-282.
- [10] M.H Beale, M.T. Hagan, and H.B. Demuth, "Radial basis networks," *Neural Networks Toolbox 7: User's guide*, 2010, pp. 6-2 – 6-14.
- [11] H. Wahid, Q.P. Ha, and H. Nguyen-Duc, "Adaptive neural network metamodel for short-tem prediction of background ozone," *Proc. IEEE-RIVF 2010*, 2010, pp. 250-253.
- [12] H. Wahid, Q.P. Ha, and H. Duc, "A metamodel for background ozone level using radial basis function neural networks," *Proc. ICARCV*, Singapore, 2010, pp. 959-963.
- [13] D.T. Sandwell, "Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data," *Geophysical Research Letters*, Vol. 14(2), 1987, pp. 139-142.
- [14] A. Fiore, D.J. Jacob, H. Liu, R.M. Yantosca, T.D. Fairlie, and Q. Li, "Variability in surface ozone background over the United States: implications for air quality policy," *J. Geophysical Research*, Vol. 108, 2003, pp. 19-1 – 19-12.