Spatio-temporal estimation of air quality parameters using linear genetic programming

Shruti S. Tikhe[1a], K. C. Khare[*2] and S. N. Londhe[3b]

[1]*Department of Civil Engineering, Sinhgad College of Engineering, Pune Maharashtra, 411041, India*
[2]*Department of Civil Engineering Symbiosis Institute of Technology, Pune Maharashtra, 412115, India*
[3]*Department of Civil Engineering, Vishwakarma Institute of Information Technology Pune Maharashtra, 411048, India*

Abstract. Air quality planning and management requires accurate and consistent records of the air quality parameters. Limited number of monitoring stations and inconsistent measurements of the air quality parameters is a very serious problem in many parts of India. It becomes difficult for the authorities to plan proactive measures with such a limited data. Estimation models can be developed using soft computing techniques considering the physics behind pollution dispersion as they can work very well with limited data. They are more realistic and can present the complete picture about the air quality.

In the present case study spatio-temporal models using Linear Genetic Programming (LGP) have been developed for estimation of air quality parameters. The air quality data from four monitoring stations of an Indian city has been used and LGP models have been developed to estimate pollutant concentration of the fifth station.

Three types of models are developed. In the first type, models are developed considering only the pollutant concentrations at the neighboring stations without considering the effect of distance between the stations as well the significance of the prevailing wind direction. Second type of models are distance based models based on the hypothesis that there will be atmospheric interactions between the two stations under consideration and the effect increases with decrease in the distance between the two. In third type the effect of the prevailing wind direction is also considered in choosing the input stations in wind and distance based models. Models are evaluated using Band Error and it was observed that majority of the errors are in +/-1 band.

Keywords: air quality; genetic programming; pune city; spatio temporal modelling

1. Introduction

Air pollution is a complex process which involves pollutant emission from various sources combined with atmospheric dispersion resulting into ground level concentrations. Air pollutants

---

∗Corresponding author, Professor, E-mail: kanchan.khare@sitpune.edu.in
[a]Research Scholar
[b]Professor

circumscribe gaseous and particulate species that may lead to adverse health effects. Air quality models which are based on pollutant concentration data can be used for assessment of potential environment and health effects associated with pollutants. In this regard, availability of accurate and plentiful data and interpretations are an imperative tool which can help the air quality management to take decisions (Asadollahfardi *et al*. 2015). A limited number of air quality monitoring stations limits the strategy of pollution prevention program in the areas at micro-scale level, where air pollution is a serious threat to the community. Most of the epidemiological studies related to air pollution are based on RSPM concentrations at fixed ambient air quality monitoring sites (CPCB 2006). However, the measurement data from these stations do not necessarily represent areas beyond their immediate vicinity, as the concentrations of pollutants in urban areas may vary by orders of magnitude on spatial scales varying from tens to hundreds of meters. Models can provide quantitative estimate of air pollutants in the desired areas, where no stations are available for monitoring of air quality.

Temporal and cause effect modelling is necessary to understand the behavior of the pollutants on a time scale. Advances in mathematical models to describe the formation, emission, transport and disappearance of air pollutants have led to a greater understanding of the dynamics of these pollutants (Tikhe *et al.* 2015). Sometimes a situation may arise when the previous time series of the meteorological parameters or the pollutants are not available due to limited number of monitors, monitoring instrumental failure or extreme environmental conditions. In Environmental studies, spatial modelling is a common practice to distinguish the spatial patterns of the phenomenon by estimating/predicting the values of the un sampled locations based on measurements at sample points. Advantage of the spatial modelling is to know the pollutant concentrations at the locations in above referred problematic events. It helps in policy making by understanding the spatial distribution of the air quality trends.

It was observed that the process of spatial mapping has been tried by researchers for Civil Engineering applications such as estimation of pan evaporation (Chaudhary *et al.* 2014), land use planning (Chen *et al.* 2015), geo hazard analysis (Noack *et al.* 2014), wave height estimation (Londhe 2008) etc. Diem and Comrie (2002) have performed spatial mapping for forecasting of ozone using linear regression methods for Tuscon. Spatial mapping of wave heights using RBF neural networks has been carried out by Kalra *et al.* (2005). Denby *et al.* (2010) used log normal multiple kriging with multiple linear regression to spatially map the O_2 and SO_2 trends in Europe. Kurt and Oktay (2010) proposed geographic model to forecast daily average SO_2, CO & PM_{10} three days in advance for Besiktas district of Istanbul.

In the present work, estimation of pollutants at one station is carried out with the help of pollutants recorded at the other station using Linear Genetic Programming.

2. Study area and data

Pune city (Maharashtra State; India) is chosen as the study area. It is one of the highly polluted and fastest developing metropolitan cities in Maharashtra state; India which is considered to be the premier industrial center of the country (www.dnaindia.com). It is located in Western Maharashtra on the Deccan Plateau at the confluence of Mula Mutha rivers at an elevation of about 560 m above mean sea level at Karachi. Pune city is identified by industrial areas of Pimpri Chinchwad and Bhosari at the North West, the major bus station of Swargat at the center and the commercial areas of Karve Road and Nalstop at the south west.
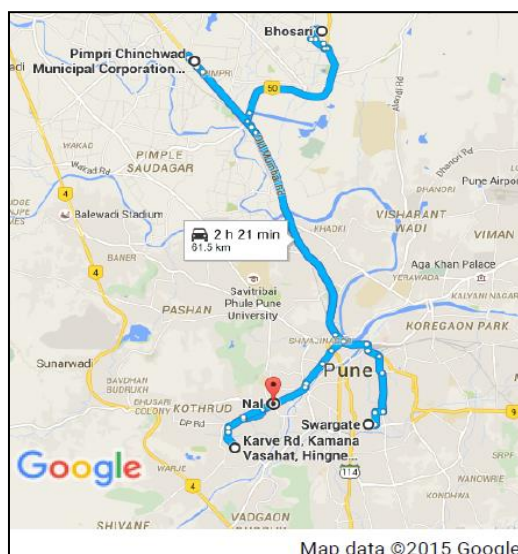
Fig. 1 Location sketch of monitoring stations in pune (Source-www.googlemaps.com)

Rapid industrialization and construction activities have resulted into exponential growth in vehicular population. This has worsened the air quality of Pune which is evident from increased pollutant concentrations for the last the few years (www.mpcb.gov.in).

SO$_2$, NO$_x$ and RSPM are considered as the criteria air pollutants which decide the air quality (www.epa.gov/the Clean Air Act, 1970). According to 2011 census data, population of Pune is over 3 million. Air quality is monitored by Maharashtra Pollution Control Board at five monitoring stations namely Swargate bus depot., Bhosari Industrial area, Nalstop, a busy point on Karve Road and Pimpri Chinchwad Municipal Corporation. Location sketch of the monitoring stations can be seen in Fig. 1.

Karve Road is a continuous monitoring station whereas air quality is monitored for six days a week at PCMC. Air quality at the remaining stations is monitored for two days a week. Chief source of air pollution for Nalstop and Swargate is vehicles. Bhosari area surrounds one of the industrial zones of the State of Maharashtra. Nalstop is very busy commercial area connecting different parts of Pune. One of the major public transport stations is located at Swargate, hence there is a continuous flow of public transport vehicles.

The dataset used for the experiments consists of daily average concentrations of SO$_2$, NOx and RSPM recorded for the period of January 2015 to August 2015 as obtained from Maharashtra Pollution Control Board. Data for all the monitoring stations was preprocessed and missing readings (7%) were filled using linear interpolation method.

3. Modeling strategy

The science (and physics) of meteorology has great bearing on how the air and its constituents move in an urban environment. While the emissions from a region tend to contribute primarily to the local air pollution, depending on the meteorological conditions and the pollutant in discussion, the impact levels can vary. Typically, the emissions are released from (an individual point or an

area) source, which after entering the atmosphere, depending on the meteorological conditions like wind speed, wind direction, pressure, temperature, moisture content, etc., interact with other pollutants, and depending on the local canopy, either deposit on a surface (as dry or wet) or linger in the air in the form of pollution, which we breathe (Guttikunda Sarath 2011). Considering the significant influence of oxides of nitrogen and sulphur on human health and environment, it is important to monitor their concentration and examine the mechanisms involved in their production, transport and decomposition in atmospheric condensed phase and surface water (Zuo *et al.* 2006). It was observed that high levels of nitrite and nitrate in dew water droplets may constitute an important source of hydroxyl radicals in the sunny early mornings. The oxidation reactions of SO_2 in atmospheric water droplets change during day and night time (Zuo and Zhan 2005). Dissolved organic compounds and transition metals such as iron ions are involved in the photochemical formation of hydrogen peroxide and other photo-oxidants in atmospheric waters (Zuo and Hoigne 1993). The rate of Fe(III)-catalyzed oxidation of S(IV) species is found to increase with increasing light intensity. The effects of sunlight on the Fe(III)-catalyzed oxidation of S(IV) should be taken into account when predicting the daytime rates of sulfuric acid formation in atmospheric water droplets (Zuo *et al.* 2005). Thus the real air quality depends on meteorological conditions, traffic characteristics, along with interactions amongst the pollutants and the atmospheric agents and the sinks of the pollutant with varying rates at day and night times. The meteorological studies for Pune city has revealed that air masses transported to Pune are from northwest to southeast direction (Yadav *et al.* 2015).

The major air pollutants such as SO_2, NO_x and RSPM are found to vary seasonally as well as diurnally. Highest concentrations of pollutants are observed in winter as compared to summer and lowest values of the pollutants are observed in rainy season (MPCB Report 2014). The pollutant concentrations also vary during peak and off peak hours (MPCB Report 2014). Highest pollution loads of RSPM and NOx is observed in central part of Pune with major commercial activities and high population and road densities and the major source of SO_2 emissions comes from industries located in the North West of the city.

In the present study, each monitoring station has a specific cause of air pollution hence the continuous air quality data for each season is essential but the air quality at all the monitoring stations except Karve Road is not being monitored continuously which increases the possibility of missing of the pollution episodes. In such a situation, the pollutant concentrations at the un-sampled locations can be estimated using spatio temporal modelling.
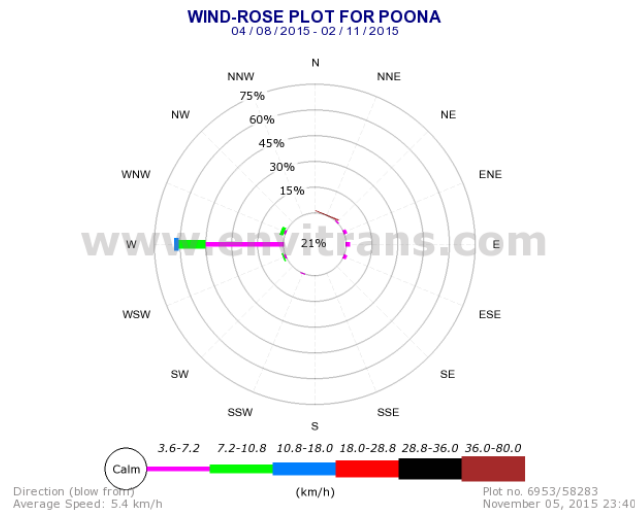
Models have been developed for estimation of criteria air pollutants with certain assumptions such as

(a) The drift and diffusion of the pollution depends on the wind direction.

(b) Wind direction is uniform and it is Westerly.

(c) The emissions are equally mixed in an urban environment under the mixing layer, for the same emissions, a lower mixing height means higher ambient concentrations.

(d) There will be atmospheric interactions between the areas covered by two stations.

(e) The previous values of the pollutant concentrations take in to account the combined effect of atmospheric interactions, sinks of pollutants and diurnal variations.

Initially the models are developed considering only the pollutant concentrations at the neighboring stations. These models do not consider the effect of distance between the stations as well as the significance of the prevailing wind direction. Distance based models are based on the hypothesis that there will be atmospheric interactions between the two stations under consideration and the effect increases with decrease in the distance between the two (Kurt and Oktay 2010). The

Table 1 Aerial distance between monitoring stations

Station	Aerial Distance Between Stations (km)				
	Swargate	Nal stop	Karve Road	PCMC	Bhosari
Swargate	0	4.36	6.48	17.36	15.29
Nal stop	4.36	0	2.93	16.31	19.05
Karve Road	6.48	2.93	0	15	22.22
PCMC	17.36	16.31	15	0	9.56
Bhosari	15.29	19.05	22.22	9.56	0

Fig. 2 Wind rose diagram for Pune (Source-http://envitrans.com/windrose.php)

effect of the prevailing wind direction is also considered in choosing the input stations in wind and distance based models.

Correlation of the pollutants at one station with the other was calculated and the input stations are arranged in the descending order of their correlation with the output and following models are developed.

(a) Plain models

In plain models pollutant concentrations are estimated on the basis of pollutant concentrations measured at four stations, which are geographically connected.

Plain models can be written as,

$$SO_2 \text{ (Swargate)} = f \text{ (}SO_2\text{(Bhosari, Nalstop, Karve Road, PCMC))} \tag{1}$$

$$NO_x\text{(Swargate)} = f \text{ (}NO_x \text{ (Bhosari, Nalstop, PCMC, Karve Road))} \tag{2}$$

$$RSPM\text{(Swargate)} = f \text{ (RSPM (Bhosari, Nalstop, Karve Road, PCMC))} \tag{3}$$

Similarly the models are developed for other stations and evaluated using LGP (Discipulus). These models consider only the pollutants without consideration of the effect of distance or wind direction.

Table 2 Input stations chosen for wind and distance based models

Station under consideration	Input Stations
Swargate	Nalstop, Karve Road, PCMC, Bhosari
Nalstop	PCMC, Bhosari
Karve Road	PCMC, Bhosari

(b) Distance based models

Distance based models take into account the distance between the two stations under consideration in addition to the pollutant concentration. Five stations were chosen for the study which lie within the radius of about 25 km. It was assumed that there will be atmospheric interactions between the areas covered by two stations. These complex interactions play a significant role in deciding the air quality at both the stations. It was also assumed that as the distance between the stations decreases, the resemblance between the pollutant concentrations at both the stations should be higher (Kurt and Oktay 2010). The model is based on the studies which have demonstrated that the effects of air pollutant levels of the neighboring station are inversely proportional to the distance between the two (Kurt and Oktay 2010).

In this model the distance values are used to compute a weighted average of air pollutant from the air pollutant values of the neighboring stations. A new weighted average air pollutant variable is created using the corresponding variables and the aerial distances between nearby stations. Aerial distances are presented in Table 1.

The new variable can be written as,

$$\text{New variable for } SO_2 \text{ Swargate} = (0.1046*(SO_2 \text{ Nalstop}) + (0.1075*(SO_2 \text{ Karve Road})) + (0.3440*(SO_2 \text{ PCMC}) + (0.4516* (SO_2 \text{Bhosari})) \tag{4}$$

The coefficient represents the weighted average distance between the two stations under consideration. In the similar fashion the new variables were created for all the pollutants and for all the stations and the models are developed. The distance based GP model would have this additional input which will take care of the assumed atmospheric interactions between the two stations.

(c) Wind and distance based models

Many atmospheric factors influence the way the air pollutants are dispersed including wind direction, wind speed, type of terrain and heating effects. A wind rose is a graphic tool used by meteorologists to give a succinct view of how wind speed and direction are typically distributed at a particular location. Using a polar coordinate system of gridding, the frequency of winds over a time period is plotted by wind direction, with color bands showing wind speed ranges. The direction of the longest spoke shows the wind direction with the greatest frequency. Fig. 2 presents the wind rose diagram for Pune city. The prominent direction of the wind for Pune is Westerly.

In these types of models, additional input is obtained on the basis of the wind direction. The stations which lie along the wind direction are only chosen as input.

PCMC and Bhosari are along the north east of the Nal stop and Karve Road so while calculating new variable for Nalstop and Karve Road only those two stations are considered. Similarly Swargate is on the south west of all four stations hence all the stations are considered while estimating pollutants at Swargate. Table 2 lists the input stations chosen for estimation of the three pollutants.

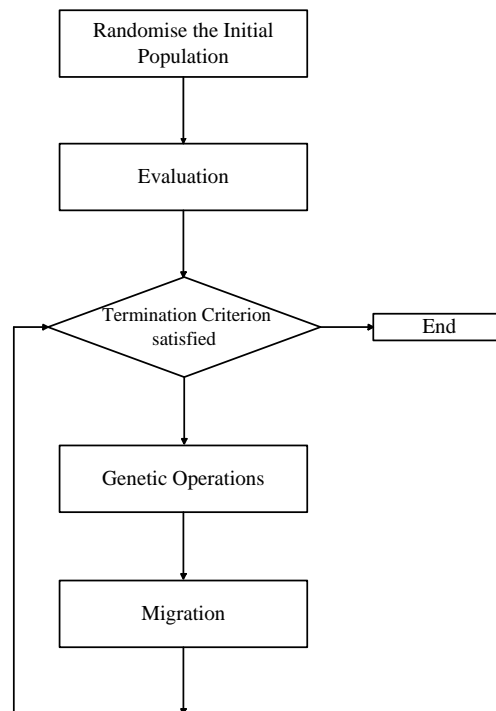All the models are developed using GP.

```
┌─────────────────────┐
│ Randomise the Initial│
│     Population       │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│     Evaluation       │
└─────────────────────┘
           │
           ▼
      ╱─────────╲
     ╱ Termination╲        ┌──────────┐
    ╱  Criterion   ╲──────▶│   End    │
    ╲   satisfied  ╱       └──────────┘
     ╲─────────────╱
           │
           ▼
┌─────────────────────┐
│  Genetic Operations  │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│      Migration       │
└─────────────────────┘
```

Fig. 3 Typical GP Flowchart (Source Pires *et al.* 2011)

4. Genetic programming

Genetic Programming (GP) is an evolutionary based data driven modeling approach. It is similar to Genetic Algorithms (GA) which follows the Darwinian principle of the survival of the fittest and obtains the solution to the problem through the process of crossover, mutation and reproduction. Solution of GP is a computer program or an equation as against a set of numbers in GA (Koza 1992). Hence GP can be conveniently used as a Regression tool. The process of GP is illustrated in Fig. 3.

Explanation of various concepts related to Genetic Programming can be referred from Koza 1992. GP starts with population of randomly generated computer program initially constructed from the data set on which genetic operations are performed using function and terminal set. Function set comprises of the operator to be used such as addition, subtraction, logarithm, square root etc. Terminal set consists of values such as inputs, constants, temporary variables on which function set operates. From the population so generated, four programs are randomly selected and a tournament is conducted. Performance of each program is measured by GP and the two programs are selected on the basis of performance. GP algorithm copies the two winner programs and transforms these copies into two new programs via crossover and mutation as per fitness i.e., winners now have the 'children'. These two new child programs are then inserted into the population of programs, replacing the two loser programs from the tournament. Crossover is inspired by the exchange of genetic material occurring in biological process of reproduction. The creation of off-springs continues (in an iterative manner) till a specified number of off-springs in a generation is produced and further till another specified number of generations is created. The

resulting offspring at the end of all this process (an equation or a computer program) is the solution of the problem. The GP thus transforms one population of individuals into another one in an iterative manner by following the natural genetic operations like reproduction, mutation and cross-over. The generated equation or program can be directly applied to unseen data to obtain required predictions.

The tree based GP corresponds to the expressions (syntax trees) from a 'functional programming language'. In this type, functions are located at the inner nodes; while leaves of the tree hold input values and constants. A population of random trees representing the programs is initially constructed and genetic operations are performed on these trees to generate individuals with the help of two distinct sets; the terminal set T and the function set F.

The second variant of GP is Linear genetic Programming (LGP) which uses a specific linear representation of computer programs. The name 'linear' refers to the structure of the (imperative) program representation only and does not stand for functional genetic programs that are restricted to a linear list of nodes only. On the contrary, it usually represents highly nonlinear solutions. Each individual (Program) in LGP is represented by a variable-length sequence of simple C language instructions, which operate on the registers or constants from predefined sets. The function set of the system can be composed of arithmetic operations (+, - , X, /), conditional branches, and function calls (f {x, xn, sqrt, ex ,sin, cos, tan, log, ln }). Each function implicitly includes an assignment to available which facilitates use of multiple program outputs in LGP. LGP utilizes two point string cross-over. A segment of random position and random length of an instruction is selected from each parents and exchanged. If one of the resulting children exceeds the maximum length, this cross-over is abandoned and restarted by exchanging equalized segments. An operand or operator of an instruction is changed by mutation into another symbol over the same set. The readers are referred to Londhe and Dixit 2012 for further details.

To the best of our knowledge Genetic Programming has not been used for spatio temporal estimation of air pollutants.

5. Results and discussion

Air quality forecasting spatio temporal models are developed for five monitoring stations of Pune with an objective of continuous monitoring of the pollutants so as not to miss the episodes and also take the immediate action during the episodes. The above mentioned three types of models are developed for three pollutants observed at five stations. All forty five models are evaluated using band error. In air pollution modelling it is better to estimate the band error as the authority is interested to know the upper limits of the pollutants or the number of days for which the air quality standards have been violated. Error is expressed as band error to assess the performance of the air quality models in the most objective way. It represents the difference between the observed and the estimated intervals in which the observed and the estimated values fall (Kurt and Oktay 2010, Feng *et al.* 2015). The range of pollution value is normally divided into five equal intervals (U.S. EPA 2009). The maximum concentration recorded during study period for SO_2, NO_x and RSPM is 68, 198 and 428 μg/cum respectively. Band ranges are decided as under on the basis of maximum values of the pollutants (refer Table 2).

SO_2 (1-14), (15-28), (29-42), (43-56), (57-70)

NOx (1-40), (41-80), (81-120), (121-160), (161-200)

RSPM (1-90), (91-180), (181-270), (271-360), (361-450)

Table 3 Maximum concentration of pollutants

Station	SO₂ (μg/cum)	NOx (μg/cum)	RSPM (μg/cum)
Swargate	66	167	391
Nalstop	64	130	306
Karve Road	54	198	428
Bhosari	66	167	391
PCMC	68	158	386

Table 4 Estimation error

Station	Error band	SO₂			NOx			RSPM		
		Plain model	Distance based model	Wind + distance based model	Plain model	Distance based model	Wind + distance based model	Plain model	Distance based model	Wind + distance based model
Swargate	+/- 1	29 (26%)	21 (19%)	21 (19%)	28 (25%)	32 (29%)	32 (29%)	21 (19%)	22 (20%)	22 (20%)
	+/- 2	4 (4%)	1 (0.9%)	1 (0.9%)	0	2 (2%)	2 (2%)	0	1 (0.9%)	1 (0.9%)
	+/- 3	0	0	0	0	0	0	0	0	0
Nalstop	+/- 1	25 (23%)	25 (22%)	33 (30%)	37 (33%)	41 (37%)	37 (34%)	30 (27%)	33 (29%)	26 (24%)
	+/- 2	2 (2%)	2 (1.8%)	3 (3%)	0	0	1 (0.9%)	0	0	1 (0.9%)
	+/- 3	0	0	0	0	0	0	1 (0.9%)	0	1 (0.9%)
Karve Rd	+/- 1	27 (24%)	39 (35%)	30 (27%)	51 (46%)	46 (41%)	48 (43%)	50 (45%)	46 (41%)	44 (40%)
	+/- 2	0	3 (3%)	1 (0.9%)	12 (11%)	0	10 (9%)	0	0	0
	+/- 3	0	0	0	0	0	0	2 (2%)	0	0
Bhosari	+/- 1	36 (36%)	29 (29%)	---	23 (21%)	30 (27%)	---	29 (26%)	27 (24%)	---
	+/- 2	0	2 (2%)	---	2 (2%)	1 (0.9%)	---	1 (0.9%)	1 (0.9%)	---
	+/- 3	0	0	---	0	0	---	0	0	---
PCMC	+/- 1	30 (27%)	81 (73%)	---	84 (75%)	75 (68%)	---	46 (41%)	35 (32%)	---
	+/- 2	1 (0.9%)	0	---	0	0	---	0	0	----
	+/- 3	0	0	---	0	0	---	0	0	---

First number in the results represent number of days with error out of 111 output readings and the number in bracket indicate the % error.

Comparing each observed value with its predicted, the band in which it lies is identified and the error is estimated and presented in Table 4.

It has been observed from the results that performance of the GP as a virtual monitor is satisfactory for all the stations except PCMC. When the measured concentrations at PCMC were closely observed it was found that there were about 21% of missing readings which affect the consistency of the series as in spatial estimation sufficient number of observations are required in minimizing prediction error.

The majority of the errors (about 85%) for all the models and for all the stations are in +/-1 band and relatively less error is reported in +/-2 (13%) and +/-3 (2%) band.

It can be seen that high concentrations of SO_2 is recorded for PCMC whereas high values of NOx and RSPM are recorded for Karve Road. PCMC area has primarily automobile industries and the industrial area of Bhosari also surrounds PCMC which results into high SO_2 concentrations. Pune city's one of the fastest growing suburbs is Karve Road. High vehicular population is observed at Karve Road which results into highest concentrations of NOx and RSPM. For all five monitoring locations of Pune city, LGP could reasonably map the variations in the concentrations of the pollutants.

As compared to plain models the results of the distance based models have been improved for Swargate (SO_2), Karve Road (NOx, RSPM), Bhosari (SO_2, RSPM) and PCMC (NOx, RSPM). Results are further improved for Nalstop (NOx and RSPM) and Karve Road (RSPM) when the input stations are chosen on the basis of prevailing wind direction.

The possible reason lies in the physics behind pollutant dispersion. Dispersion of the pollutants is controlled by geographical location of the station and local meteorological conditions such as mixing layer height, wind speed, wind direction, temperature, pressure, moisture content etc. In the present work five stations are lying along a relatively flat terrain, hence the meteorological conditions regulate the dispersion of the pollutants. The fundamental parameter in the movement of the contaminants is the wind, its speed and direction, which in turn are interlinked with the vertical and horizontal temperature gradients. The greater the wind speed, the greater is the turbulence and the more rapid and complete is the dispersion of the contaminants in the air. All the stations of the study area lie within the radius of about 25 km with respect to other stations. Hence general trend of the wind direction as given by wind rose diagram has been considered in order to determine the choice of the input station. The general trend of the wind direction for Pune is Westerly which implies that the stations lying along the South East of the other stations will have the effect of the pollutants from the station at the North West due to dispersion. Amongst the five stations, Swargate lies along the South East West of the other four stations and hence will have the maximum effect of dispersion. From the results of the wind and distance based model, it can be seen that the spatio-temporal estimation at Swargate station has lesser error as compared to other stations.

It can be seen from all the models that LGP results are satisfactory and it could trace the non linearity in the variation of the air quality. It can be said that LGP performs well as it can self-modify, through the genetic loop, a population of function trees in order to finally generate an "optimal and physically interpretable" model. Hence, in Spatio temporal modelling the choice of the input stations can be made considering the prevailing wind direction. However more trials are necessary with seasonal data to test the capability of LGP for spatial mapping.

6. Conclusions

The present work is carried out to study the spatio temporal estimation of air quality parameters using LGP. The data consists of eight months measured concentrations of air pollutants from five

monitoring stations. Spatio temporal models are developed considering the pollutants recorded at the neighbouring stations, atmospheric interactions between satellite cities and the prevailing wind direction.

From all the models, it can be seen that LGP could map the variations in the observed air quality. In general models achieved acceptable statistical measures in terms of band error. In most of the cases wind and distance based models are proved to be superior to other models as they consider the physics behind pollution dispersion. In this manner, the study of movements of pollutants over urban areas can help us understand their impact on pollution planning. Furthermore in depth study is necessary with respect to variation in wind direction, mixing height and pollution dispersion in spatio temporal modelling.

References

Asadollahfardi, G., Zamanian, M., Mirmohammadi, M., Asadi, M. and Tameh, F.I. (2015), "Air pollution study using factor analysis and univariate box-jenkins modelling for the northwest of Tehran", *Adv. Environ. Res.*, **4**(4), 233-246.

Central Pollution Control Board (CPCB), New Delhi, India (2006), *Report on Air Quality Trends and Action Plan for Control of Air Pollution from Seventeen Cities*.

Chaudhary, N., Londhe, S. and Khare, K. (2014), "Spatial mapping of pan evaporation using linear genetic programming", *J. Hydrol. Sci. Technol.*, **4**(3), 234-244.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X. and Mills, J. (2015), "Global land cover mapping at 30 m resolution: A POK based operational approach", *ISPRS J. Photogram. Rem. Sens.*, **103**, 7-27.

Daily News and Analysis (2016), http://www.dnaindia.com/, India.

Denby, B., Sundvor, I., Cassiani, M., De Smet, P., De Leeuw, F. and Horalek, J. (2010), "Spatial mapping of ozone and SO_2 trends in Europe", *Sci. Tot. Environ.*, **408**(20), 4795-4806.

Diem, J. and Comrie, A. (2002), "Predictive mapping of air pollution involving sparse spatial observations", *Environ. Pollut.*, **119**, 99-117.

Environmental Protection Authority (2016), http://www.epa.wa.gov.au/the clean air act 1970/, U.S.A.

Feng, X., Zhu, Q.L.Y., Hou, J., Jin, L. and Wang, J. (2015), "Artificial neural networks forecasting of $PM_{2.5}$ pollution using air ass trajectory based geographic model and wavelet transformation", *Atmosph. Environ.*, **107**, 118-128.

Guttikunda, S. and Gurjar, B. (2010), "Role of meteorology in seasonality of air pollution in megacity Delhi, India", *Environ. Monitor. Assess.*, **185**(5), 3199-3211.

Kalra, R., Deo, M.C., Kumar, R. and Agarwal, V.K. (2005), "RBF network for spatial mapping of wave heights", *Mar. Struct.*, **18**(3), 289-300.

Koza, J.R. (1992), Genetic Programming, MIT Press.

Kurt, A. and Oktay, A.B. (2010), "Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks", *Exp. Syst. Appl.*, **37**(12), 7986-7992.

Londhe, S.N. (2008), "Soft computing approach for real-time estimation of missing wave heights", *Ocean Eng.*, **35**(11-12), 1080-1089.

Londhe, S.N. and Dixit, P.R. (2012), *Genetic Programming-New Approaches and Successful Applications*, Tech Publications, 199-224.

Maharashtra Pollution Control Board (2016), http://mpcb.gov.in/, Maharashtra State, India.

MPCB (2013-2014), *Air Quality Report*.

Noack, S., Knoblosh, A., Etzold, S.H., Barth, A. and Kallmeier, E. (2014), S*patial Predictive Mapping Using Artificial Neural Networks*, The International Achieves of the Photogrammetry , Remote Sensing and Spatial Information Sciences, XL-2, 79-86.

Pires, J.C.M. and Martins, F.G. (2011), "Prediction of tropospheric ozone concentrations: Application of a methodology based on the Darwin's theory of evolution", *Exp. Syst. Appl.*, **38**(3), 1903-1908.

Tikhe, S.S., Khare, K.C. and Londhe, S.N. (2015), "Multicity seasonal air quality index forecasting using soft computing techniques", *Adv. Environ. Res.*, **4**(2), 83-104.

Yadav, S., Pravin, O. and Satsangi, G. (2015), "The effect of climate and meteorological changes on particulate matter in pune", *Environ. Monitor. Assess.*, **187**, 402.

Yuegang, Z. and Jian, Z. (2005), "Effects of oxalate on Fe-catalyzed photooxidation of dissolved sulfur dioxide in atmospheric water", *Atmosph. Environ.*, **39**, 27-37.

Yuegang, Z. and Jürg, H. (1993), "Evidence for photochemical formation of H_2O_2 and oxidation of SO_2 in authentic fog water", *Sci.*, **260**, 71-73.

Yuegang, Z., Chengjun, W. and Thuan, V. (2006), "Simultaneous determination of nitrite and nitrate in dew, rain snow and lake water samples by ion-pair high performance liquid chromatography", *Talanta*, **70**, 281-285.

Yuegang, Z., Jian, Z. and Taixing, W. (2005), "Effects of monochromatic UV-visible light and sunlight on Fe(III)-catalyzed oxidation of dissolved sulphur dioxide", *J. Atmosph. Chem.*, **50**, 195-210.

WL