The Asian Review of Civil Engineering
ISSN: 2249-6203 (P) Vol.7 No.2, 2018, pp.1-6

DOI: https://doi.org/10.51983/tarce-2018.7.2.2273

Assessment of Air Pollution Modelling using Spatial Interpolation Techniques

M. Goutham Priya[1] and S. Jayalakshmi[2]
[1]Research Scholar, [2]Assistant Professor,
Institute of Remote Sensing, College of Engineering, Chennai, Tamil Nadu, India
E-Mail: gouthampriya.mahendran@gmail.com, j_lak2001@yahoo.co.in

Abstract - **Air pollution has become the growing worldwide threat. Effective assessment of air pollution modelling depends on significant, wide-spread and distinguished instrumental data which are not possible for a developing country like India to have established an extensive network. This shortcoming on the capability to precisely envisage the pollutant concentration at unmonitored stations can be accomplished by spatial interpolation. Spatial Interpolation is a process, in which a parameter can be predicted and enumerated from a limited number of sample data points which is used to model the data for air pollutants like NO_2, SO_2 and RSPM. This study establishes an assessment of interpolation techniques to produce fine-scale air quality data for Tamil Nadu, India. The interpolation techniques which are used to evaluate the air quality data are Kriging, Inverse Distance Weighting, Local Polynomial Interpolation, Global Polynomial Interpolation, Radial Basis Function, Kernel Interpolation, Diffusion Interpolation, Natural Neighbour and Spline. The accuracy assessment was done with the help of statistical error metrics such as Root Mean Square Error, Mean Average Error, Mean Average Percentage Error, Mean Average Relative Error and Index of Agreement. Cross-Validation is executed by considering 20% of the data points as test data and remaining 80% as training data respectively. Spline, Local Polynomial Interpolation and Global Polynomial Interpolation technique showed enhanced resemblance between the observed and estimated values for the pollutants NO_2, SO_2 and RSPM respectively. This study embarks the association between the air pollutant and their distribution through spatial interpolation using Geographical Information System.**
Keywords: **Air Pollution, GIS, Interpolation, Statistical Error Metrics**

I. INTRODUCTION

Air pollution refers to the emission of substances such as particulates, gases and biological molecules into the air that are detrimental to human wellbeing and planet as a whole. Majority of the air pollution is caused from energy use and production. The governing sources of air pollution include but are not limited to automobile exhaust, biomass burning, urbanization, industrialization, construction, and natural disasters such as volcanic eruptions and forest fires. The major pollutants being NO_X, SO_X, Ozone, RSPM and CO. Critics have termed pollution as "Price of Industrialization" and "Disease of Wealth" caused by industrialization and vehicular emission respectively [1]. The major effects of air pollution are respiratory and cardiovascular problems in humans, acid rain, Depletion of ozone layer, deterioration of wildlife, etc. Maintaining the air quality contained by limits has become crucial for the reason that we cannot evade breathing the air that surrounds us since an average adult breathes in about 20000 litres of air per day [2]. The concentration of the pollutant varies with meteorological characteristics such as wind speed and direction at a given space and time. Establishing a network to cover an entire country or state is a concern due to economic status and expansion cost in India [3]. As a result, it is vital to develop an appropriate progression to evaluate and predict the value at unmonitored stations in the area of interest. This can be accomplished by the method of Interpolation, a mathematical field of numerical analysis, which is used to estimate the pollutant at unmonitored location from a set of predefined known original data.

Conventional method of interpolation is repetitive and prolonged, for that reason, spatial interpolation of the GIS toolset would be of huge assistance in obtaining the data in a prompt and accurate manner. Spatial interpolation is one of the dominant tools in modelling the spatial variation of any environmental system. It has been employed in various studies involving parameters such as temperature [4] [5] [6], precipitation [7] [8], air quality [9] [10] [11] and groundwater [12] etc. According to the global urban air pollution ranking released by WHO in May 2016, Chennai is the most polluted city in Tamil Nadu with a global ranking of 314 followed by Trichy and Coimbatore with rankings 370 and 410 respectively, http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/. Hence, Tamil Nadu has become the only state to have three cities in the world ranking. Consequently, there is a need to study the level of pollutants in Tamil Nadu. There are a limited number of monitoring stations which implies a need on spatial interpolation techniques to acquire wide spread data. Thus the objective of this paper is to identify the best interpolation method for NO_2, SO_2 and RSPM through statistical error metrics. This study will help environmental engineers and scientist to rationalize the interpretation over a region of interest for decision making.

II. MATERIAL AND METHOD

A. Study Area and Data

Tamil Nadu (TN) is situated between Latitude 8° 5' N and 13° 35' N and between Longitudes 76° 15' E and 80° 20'E, is one of the 29 states of India with an area of 1, 30,058 sq km (Fig.1). The Capital City of Tamil Nadu is Chennai. It is bounded by Eastern Ghats on the north, Gulf of Mannar and the Palk Strait on the southeast and the Indian Ocean on the south, the Bay of Bengal in the east, Kerala on the west. TN being a forerunner in industries has the leading number of factories owing to greater pace of air pollution. The state, being more than 50% urbanized, has become mandatory to study the air pollutant concentration on a cautious level.

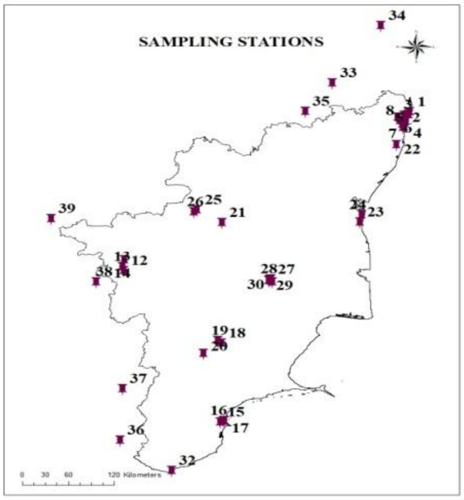


Fig. 1 Study Area with Sampling Locations

The major air pollutants such as NO_2, SO_2, and RSPM have been acquired from central pollution control board (CPCB). The mean annual data of the year 2015 has been used in this study. 31 monitoring stations have been established by both Central and State Government in TN. Since the study is pertained to spatial interpolation, for better accuracy and precision the monitoring stations from neighboring states situated along the border of TN are also included. Therefore 4 and 3 stations from Kerala and Andhra Pradesh respectively have been taken into consideration.

B. Methodology

1. Spatial Interpolation Technique

For air pollution modelling, spatial continuous data play a striking role in augmentation, susceptibility assessment and management. The approach followed in this study to obtain spatial continuous data from air quality monitoring networks which are commonly point data sources is accomplished by interpolation techniques which generate predictions by exploiting the spatio-temporal patterns at unmonitored stations. The interpolation techniques employed in this study are grouped as Deterministic which includes Radial basis Function (RBF), Inverse Distance Weighting (IDW), Local Polynomial Interpolation (LPI), Global Polynomial Interpolation (GPI), Natural neighbour (NN) and Spline (SP), Geostatistical comprises of Kriging (KRIG) and interpolation with barriers encompasses Kernel interpolation (KI), Diffusion Interpolation (DI). The features of the interpolation techniques are detailed in Table I.

TABLE I DESCRIPTION OF INTERPOLATION METHODS

S. No.	Interpolation Method	Features
1	KRIG	Biased linear blend of the known sample values which considers equally the distance and the degree of variation concerning known data points when estimating values in unknown areas
2	IDW	Based on the postulation that the degree of influence of the close proximity sample points should be greater than the effect of more distant points
3	RBF	Exact interpolation technique that creates a surface that captures global trends and picks up local variation and also sensitive to outliers
4	GPI	Typically polynomial, it should be noted that the more multifarious the polynomial, the more complicated it is to attribute substantial significance to it
5	LPI	Fits many polynomials, each within specified overlapping neighbourhoods. These neighbourhoods overlap, and the value used for each prediction is the value of the fitted polynomial at the centre of the neighborhood
6	DI	Similar to kernel interpolation with Gaussian kernel. It predicts on an automatically generated grid unlike other methods which uses triangles with variable sizes
7	KI	Uses the shortest distance between the two points for prediction on the sides of the barrier which are connected by straight lines
8	NN	local, deterministic and abrupt interpolant, where the estimated function is a linearly appropriately weighted average of nearby data points
9	SP	Used to find the minimum curvature surface that passes through a set of irregularly space data points

Figure 2 illustrates the comprehensive approach of the methodology. The required air pollutant data (NO_2, SO_2, and RSPM) are acquired and a thorough study is conceded to avoid anomalies. Air pollutant database is developed for the study area using GIS. To determine the best interpolation technique, a number of iterations have been carried out by considering 80% of the known sample point data as training data and the remaining 20% as test data. The iterations are conducted by rotating the training and test data until all the sampling stations are covered. The estimated values are compared with the observed values for accuracy assessment. The same procedure is followed until all the monitoring stations from the study are covered for each of the pollutant NO_2, SO_2 and RSPM.

Pollutant concentration maps are generated using overlay analysis associated with fuzzy logic. The advantage of fuzzy over weighted overlay is that it is used to address situations when the boundaries between classes are not clear. Unlike weighted overlay it defines the possibility of a phenomenon belonging to a class. To accommodate fuzzy overlay, initially, the input raster has to be fuzzified. The process of transforming the original values to a scale of 0 to 1 is called fuzzification process. Therefore fuzzy membership is allotted to the raster of each iteration process of the best interpolation technique. Finally the overlay is executed with gamma function for normalized results.

2. Accuracy Assessment

The accuracy of the interpolation techniques is assessed by using statistical error metrics, which is considered to be the best prevailing tool for cross-validation. Statistical error metrics can be classified as dimensional and non-dimensional. In this study the error metrics considered are Root Mean Square Error (RMSE) as defined in equation (1) and Mean Average Error (MAE) as defined in equation (2) under the category of scale dependent metrics, Mean Average Percentage Error (MAPE) as defined in equation (3) grouped under percentage error metrics, Mean Average Relative Error (MARE) as defined in equation (4) in the class of Relative error metrics and finally Index of Agreement (IOA) as defined in equation (5) from non-dimensional statistics.

$$\text{RMSE} = \sqrt{[\frac{1}{n}\ \sum_{i=1}^{n} |\ O_i - E_i\ |^2]} \quad (1)$$

$$\text{MAE} = \frac{1}{n}\ \sum_{i=1}^{n} |\ O_i - E_i\ | \quad (2)$$

$$\text{MAPE} = \{\frac{1}{n}\sum_{i=1}^{n} \frac{O_i - E_i}{O_i}\} * 100 \quad (3)$$

$$\text{MARE} = \frac{1}{n}\sum_{i=1}^{n} \frac{O_i - E_i}{O_i} \quad (4)$$

$$\text{IOA} = 1 - [\frac{\sum_{i=1}^{n}(O_i - E_i)^2}{\sum_{i=1}^{n}(|O_i - \mu| + |E_i - \mu|)^2}] \quad (5)$$

Where, O_i is the observed value, E_i is the estimated value, n is the number of values and μ is the mean of the observed values.

The accuracy of the estimated values is determined by employing the statistical error metrics. The values of RMSE, MAE should be minimal for the best method as it measures the average magnitude of the error in a set of predictions without considering its direction. Whereas, MAPE is a percentage error metric which, calculates the relative overall fit in percentage. Output of MAPE indicates the percentage of error in the estimated values. Therefore the value of MAPE should be minimal for the best method.

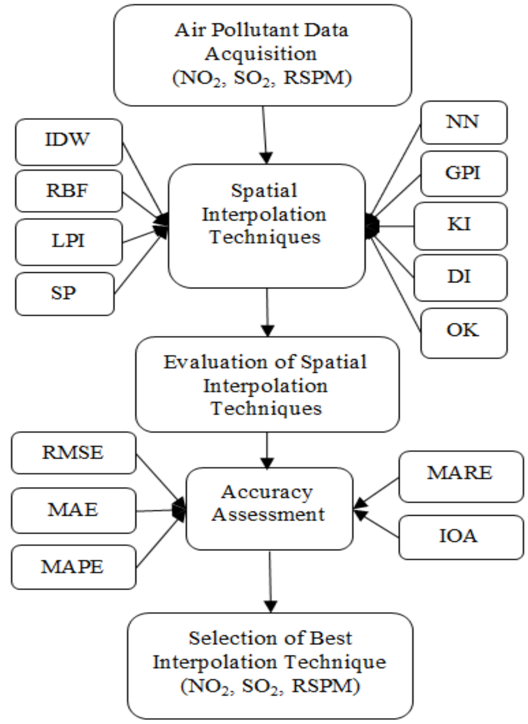


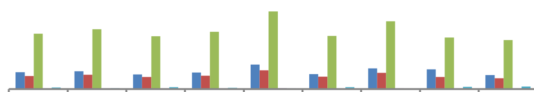
Fig. 2 Methodology - Flowchart

Also, MARE gives the relative error of the average between the observed and estimated values which has to be least. IOA is a non-dimensional statistical parameter which used to determine the relationship between the actual and interpolated data. It is a dimensionless parameter and the value ranges from 0 to 1 indicating better agreement.

III. RESULTS AND DISCUSSION

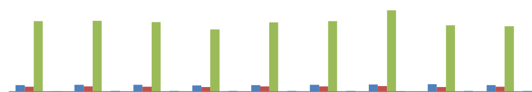
Figure 3 shows the graph depicting the comparison between the statistical error metrics for all the interpolation methods considered for NO_2. It is apparent that SP has an RMSE of 3.238, MAE of 2.529, and MARE 0.114, all having the desirable least values whereas IOA yields highest value of 0.563 which is closer to 1 when compared to the other techniques. It is perceptible that SP is followed by IDW as it carries an RMSE of 3.380, MAE of 2.801, MARE of 0.122 and IOA of 0.415. But when considering the percentage of accuracy arrived from MAPE, approximately all the interpolation techniques are in close proximity to one another with SP and IDW contributing 88.587% and 87.762% respectively.

GPI has the lowest percentage of 81.972%. Therefore on the grounds of considering all the error metrics Spline has been identified as the best interpolator and IDW has been identified as the next best interpolator for NO_2 for the study area respectively. GPI holds the highest value of all the statistical metrics representing as the least method to be adopted for the pollutant.



	KRIG	RBF	IDW	LPI	GPI	KI	DI	NN	SP
RMSE	3.875	4.108	3.380	3.824	5.651	3.431	4.781	4.594	3.238
MAE	3.031	3.281	2.801	3.129	4.373	2.864	3.792	2.770	2.529
MAPE	12.874	13.884	12.238	13.325	18.028	12.352	15.763	11.949	11.413
MARE	0.129	0.139	0.122	0.133	0.180	0.124	0.158	0.119	0.114
IOA	0.321	0.202	0.415	0.296	0.148	0.395	0.043	0.480	0.563

Fig. 3 Comparison of the statistical metrics of the interpolators considered for NO_2



	KRIG	RBF	IDW	LPI	GPI	KI	DI	NN	SP
RMSE	3.172	3.388	3.440	3.086	3.282	3.310	3.587	3.646	3.164
MAE	2.485	2.627	2.506	2.330	2.688	2.655	2.782	2.275	2.457
MAPE	32.979	33.205	32.574	29.166	32.449	33.035	38.051	31.021	30.655
MARE	0.330	0.332	0.326	0.292	0.324	0.330	0.381	0.310	0.307
IOA	0.468	0.491	0.550	0.580	0.530	0.529	0.244	0.580	0.600

Fig. 4 Comparison of the statistical metrics of the interpolators considered for SO_2

In case of the pollutant SO_2 the nearest association of the observed values is observed in LPI than the other interpolators. The values of RMSE, MAE, and MARE are 3.086, 2.330, and 0.292 respectively which are least when compared to the other methods. Whereas in view of MAPE, it can be clearly justified that LPI has the highest percentage of accuracy of 70.834% whilst other methods has a lower percentage value. But in the event of IOA Spline has the highest value of 0.600.

Therefore, the best technique is identified based on how many statistical error components are acceptable for the given pollutant. Since LPI has four out of five error metrics (RMSE, MAE, MAPE and MARE) satisfactory, it can be considered as the best method for the interpolation of SO_2 among the other methods. To rationalize, fig 5 shows the comparison between the statistical metrics for all the interpolation methods considered for SO_2. DI has the highest and lowest values of the dimensional and non-dimensional statistics respectively confirming that it is the least method having accuracy of 61.949%.

The concentration map of NO_2 is shown in figure 5 which distinctly manifest that Chennai, Kancheepuram, Villupuram, Cuddalore, part of Vellore, Trichy, Coimbatore and Salem are severely affected. The main source of NO_2 are burning fuels from automobiles, power plants and the equipment or the machineries engaged in factories for the production which operates on fuel. It is very indisputable that the towering vehicle population such as cars, two-wheelers, trucks and buses when compared to the other districts in spite of high-quality railway transportation owing to high population.

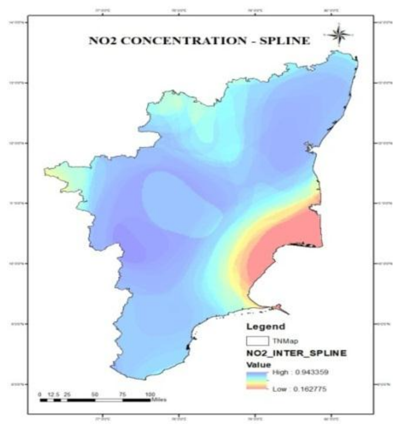


Fig. 5 NO_2 Concentration Map

Also, these areas are known for soaring number of power plants and factories. The high concentration in the interior parts is due to the fact that the primary mode of transportation is roadways. Also, due to the inconsistency of the railway transport there is a high dependency on public transport or own vehicles, adds to the cause. The areas around the districts Nagapatinam, Thiruvarur, Thanjavur, Pudukottai, Sivagangai and Ramanathapuram are the ones that have low concentration of NO_2. These districts are least urbanized. Moreover, the main sources of economy are sea-borne trading, fishing, agriculture, tourism, and service oriented industry. Figure 6 depicts the SO_2 Concentration Map for the study area. It is noticeable that the concentration fades out from the coastal areas to the interior parts. The concentration is high along the coastal areas like Chennai, Kancheepuram, Villupuram, Cuddalore, Nagapatinam, Pudukottai, Ramanathapuram, Pudukottai and Tuticorin and becomes low at the areas which include Nilgiris, parts of Coimbatore, Dharmapuri and Vellore.

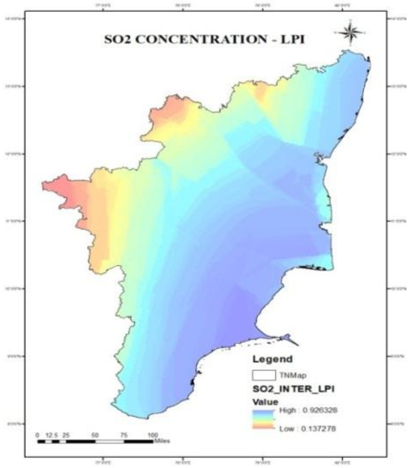
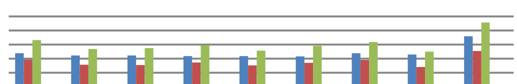


Fig. 6 SO_2 Concentration Map

SO_2 being an invisible gas are sourced mainly from anthropogenic activities like industrial activity that involves coal, oil or gas that contains sulphur. Diesel-driven vehicles can also be considered as a source of SO_2. A significant increase in sugar, cement and fertilizer industries in the recent years have contributed to the increased emission of SO_2.

It is apparently evident in figure 7 that the Natural Neighbour (NN) has the estimated values in close proximity with the observed values having an MAE of 14.097, MARE of 0.251 and IOA of 0.557. It is also palpable from MAPE that 74.909% of the values are accurate. But in Figure 8 depicting the concentration map of RSPM by NN interpolator which distinctly depicts that the certain areas under study are not covered by the interpolation as it creates a Delaunay Triangulation of the input data points and selects the closest nodes that form a convex hull around the interpolation point, and then weights their values by proportionate area. This method is most appropriate where sample data points are evenly distributed throughout the study area. Hence NN cannot be identified as the best interpolator. Therefore Global Polynomial Interpolation (GPI) replaces NN as it has the next nearest value to the observed values having RMSE of 21.822, MAE of 15.405, MARE of 0.257. The percentage of accuracy of GPI is 74.269 and it is closer to that value of NN. As a result GPI is identified as the best method of interpolation for RSPM in the study area.



	KRIG	RBF	IDW	LPI	GPI	KI	DI	NN	SP
RMSE	23.84	22.38	22.38	21.88	21.82	21.71	23.90	22.90	35.82
MAE	19.36	15.65	15.42	17.33	15.40	17.12	19.08	14.09	25.41
MAPE	33.22	26.81	27.53	29.48	25.73	29.03	31.86	25.09	45.56
MARE	0.332	0.268	0.275	0.295	0.257	0.290	0.319	0.251	0.456
IOA	0.183	0.468	0.543	0.422	0.388	0.455	0.100	0.557	0.458

Fig. 7 Comparison of the statistical metrics of the interpolators considered for RSPM

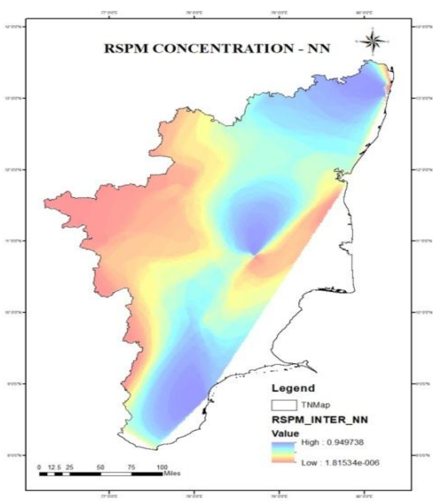


Fig. 8 RSPM Concentration Map by NN

According to WHO particulate matter affects the living more than any other pollutant. The major components are sulphate, nitrates, ammonia, sodium chloride, black carbon and mineral dust. It is visible that the concentration is high along the coastal areas and fades out towards the interior parts of TN due to escalating speed in growth of new industries and expansion of new constructions along the coastal zones. Also the re-suspension of dust plays a major role in contributing to the increase of particulates.

The particulates can also pop in the study area by long-range transportation of the pollutants due to meteorological parameters like wind direction and wind velocity. Therefore, the areas under high concentration need not specifically indicate that the particulates are generated from the same area. Also there are thermal power plants and coal industries situated amid the high concentration areas.

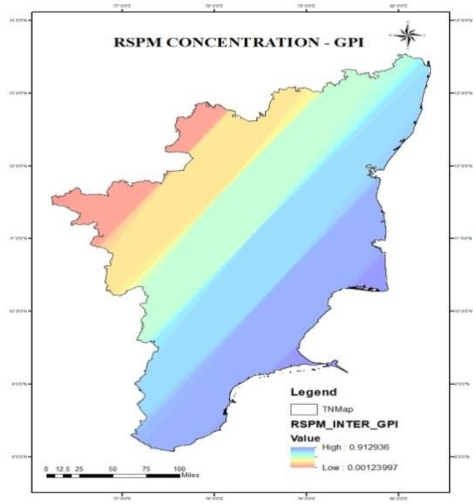


Fig. 9 RSPM Concentration Map by GPI

IV. CONCLUSION

Spatial Interpolation of air pollution data contributes useful maps to understand the spatial distribution in space. Such techniques can be used to estimate the pollutant concentration from unmonitored stations for sustainable management. Such techniques also requires continuous, evenly and wide-spread data to obtain consistent results. In this study, various interpolation techniques (KRIG, RBF, IDW, LPI, GPI, KI, DI, NN and SP) were considered to estimate the values of the air pollutants NO_2, SO_2, RSPM in Tamil Nadu. Accuracy Assessment was done using dimensional (RMSE, MAE, MAPE, MARE) and non-dimensional (IOA) statistical error metrics. Spline, Local Polynomial Interpolation and Global Polynomial Interpolation are considered to be the best method of interpolation for the pollutants NO_2, SO_2 and RSPM respectively. Also in the year 2015 there was incessant rain during monsoons which dissipated most of the pollutants. Since the annual average of the year is taken for the study without seasonal variations, there are certain regions in the study area which showed low concentration which are supposed to be high. The findings also indicate that there is no global method to identify the best technique for a study area [13]. It also depends on the spatial configuration of monitoring networks. To recognize the same, the different interpolation techniques can be compared and assessed using statistical error metrics for the study area considered. The results of this research may also be combined with other parameters like altitude, seasonal pattern of wind velocity and wind direction to study the dispersion pattern and to identify the critical sampling points from the network.

REFERENCES

[1] M. N. Rao and H. V. N. Rao, “Air Pollution”, McGraw Hill Education Private Limited, New Delhi, India, pp. 1-2, 2016.

[2] Bhawna Dubey, “Application of air pollution models and remote sensing in Air Quality Management”, *Indian Journal of Applied Research*, Vol. 4, No. 5, pp. 266-268, 2014.

[3] Kanakiya r s, Singh S K, and Shah U, “GIS application for spatial and temporal analysis of the Air pollutants in Urban area”, *International Journal of Advanced Remote Sensing and GIS*, Vol.4, No. 1, ISSN 2320 – 0243, pp 1120-1129, 2015.

[4] Willmott.C.J and Matsuura K, “Smart Interpolation of annually averaged air temperature in the United States” *Journal of Applied Meteorology*, Vol. 34, pp. 2577 – 2586, 1995.

[5] Daniel Kurtzman and Ronen Kadmon, “Mapping of temperature variables in Israel: a comparison of different interpolation methods”, *Climate Research*, Vol. 13, pp. 33–43, 1999.

[6] Mustafa GÜLER and Tekin KARA, “Comparison of Different Interpolation Techniques for Modelling Temperatures in Middle Black Sea Region”, *Journal of Agricultural Faculty of Gaziosmanpasa University*, Vol.31, Issue.2, ISSN:1300-2910, pp 61 – 71, 2014.

[7] Xian Luo and Youpeng Xu, Yi Shi, “Comparison of Interpolation Methods for Spatial Precipitation under Diverse Orographic Effects”, *IEEE*, 2011.

[8] Alan Mair and Ali Fares, “Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island”, *Journal of Hydrologic Engineering*, Vol. 16, pp 371 – 383, 2011.

[9] Phatarapon Vorapracha, Pongtep Phonprasert, Suparada Khanaruksombat and nuchanaporn Pijarn, “A Comparison of Spatial Interpolation Methods for predicting concentrations of Particle Pollution (PM_{10})”, *International Journal of Chemical, Environmental & Biological Sciences*, Vol. 3, No. 4, pp. 302 – 306, ISSN: 2320 - 4087, 2015.

[10] D Rojas Avallaneda, “Spatial interpolation Techniques for estimating the levels of pollutant concentrations in the atmosphere”, *Revista Mexicana De Fisica*, Vol. 53, No. 6, pp. 447 – 454, 2007.

[11] Shareef M, Husain T and Alharbi B, “Optimization of Air Quality Monitoring Network Using GIS Based Interpolation Techniques”, *Journal of Environmental Protection*, Vol.7, pp. 895 – 911, 2016.

[12] Yue Sun, Shaozhong Kang, Fusheng Li and Lu Zhang, “Comparison of interpolation methods for depth to ground water and its temporal and spatial variations in the Minqin oasis of northwest china”, *Environment Modelling and Software*, Vol. 24, No. 10, pp. 1163 – 1170, 2009.

[13] Lamiaa Khazaz, Hassane Jarar Oulidi, Saida El Moutaki and Abdessamad Ghafiri, “Comparing and Evaluating Probabilistic and deterministic Spatial Interpolation Methods for Ground Water Level of Haouz in Morocco”, *Journal of Geographic Information System*, Vol. 7, pp. 631-642, 2015.