Indian J.Sci.Res. 17(2): 58-63, 2018

ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

A STUDY ON COMPARISON OF INTERPOLATION TECHNIQUES FOR AIR POLLUTION MODELLING

[1]Goutham Priya M, [2]Jayalakshmi S, [3]Samundeeswari R
[1,2]IRS, College of Engineering, Chennai
[3]Dept of Civil Engg, University college of Engineering, Panruti

Abstract- Hasty industrialization and automobiles with mounting population are the major causes of air pollution which are habitually termed by critics as "Price of industrialization" or as "the disease of wealth". Air Pollution research relies profoundly on extensive, influential and eminent instrumental data. Spatial Interpolation technique, a context in which a parameter can be derived and quantify spatial variation with existing known data, used to model the data for a choice of air pollutants like NO_2, SO_2 and RSPM. This paper presents a study of comparison between interpolation techniques to produce fine-scale air quality data for the major cities in Tamilnadu, India. The interpolation techniques which were compared and assessed using station-measured air quality data are Inverse Distance Weighting Method (IDW), Local Polynomial Interpolation (LPI), Global Polynomial Interpolation (GPI), Radial Basis Function (RBF), Kernel Interpolation with Barriers (KI), and Diffusion Interpolation with Barriers (DI). Their performances were assessed by the error metrics Root Mean Square Error (RMSE). Cross-Validation is done by taking nearly 20% of the stations as test data and the remaining 80% as training data. The study shows that IDW, RBF and GPI are the best interpolators for the air pollutants NO_2, SO_2 and RSPM respectively which showed low RMSE.

Key words: Air Pollution, Interpolation, GIS, RMSE

I. Introduction

Air Pollution in urban and rural environment, is a crucial factor for the solemn health hazards and life implications on humans. Clean air is required to decrease the mortality rate caused by air pollution. The main causes for air pollution can be either natural such as volcanoes, forest fires, dust, pollen grains, inorganic decays, fog, mists etc or anthropogenic such as automobiles, industries, shipping factories, domestic activities etc. Therefore, in general, the actions of humans have contributed immensely to air pollution and with the increasing population, the air pollution have also popped up in recent years in all the major metropolitan cities like Delhi, Mumbai, Kolkata and Chennai in India. Air Pollution monitoring network is required to know the baseline status of various air pollutant parameters like NO_2, SO_2, RSPM, CO, O_3 etc. But establishing monitoring stations to cover the entire country or state is not possible for a country like India because of its economic status due to its expansion survey cost. Moreover, the air pollutant measured at a particular location need not be the concentration emitted exactly at the same location. It may have also moved in from a different location due to wind velocity and direction. Therefore with limited stations established for monitoring, there is need for a method to obtain the concentration of the pollutant throughout the study area. Interpolation comes in picture, which is a mathematical method of predicting or estimating data of un-sampled location from a set of predefined known original data. Conventional method of interpolation is monotonous and protracted; therefore, interpolation in GIS would be of enormous assist in obtaining the data in a speedy and precise manner. Interpolation is one of the triumphant tools in modelling the spatial variation of the environmental system.

In Tamil Nadu, air Pollution is widespread in urban areas due to the driving factors like industrialization, raging increase in vehicular population and haphazard burning of garbage. Consequently, there is a need to study the major air pollutants in major cities in Tamil Nadu. Interpolation is a significant tool and has been used in various studies like temperature [1] [2] [3], precipitation [4] [5], air quality [6] [7] [8], ground water [9] etc. The objectives of this paper are (a) To study and compare the various interpolation techniques such as Radial basis Function (RBF), Inverse Distance Weighting (IDW), Local Polynomial Interpolation (LPI), Global Polynomial Interpolation (GPI), Kernel interpolation with Barriers (KI), Diffusion Interpolation with Barriers (DI) to spatially interpolate NO_2, SO_2, RSPM for the sampling locations in Tamil Nadu. (b) To evaluate the best optimal interpolation model for NO_2, SO_2, RSPM by accuracy assessment using Statistical Accuracy Metrics.

II. Material and Method

A. Study Area and Data

Tamil Nadu 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 of 1, 30,058 sq km area (Fig.1). Chennai is its capital. It is bordered by the Bay of Bengal in the east, Kerela on the west, Eastern Ghats on the north, Gulf of Mannar and the Palk Strait on the southeast,

[1]**Corresponding Author**

and the Indian Ocean on the south. On top of requisites of Industrialization, in recent years, Tamil Nadu has turn out to be a forerunner of industries. Consequently, it has become mandatory to study the base line concentration of air pollutants on a vigilant level. For the study, the main source data has been acquired from Central Pollution Control Board (CPCB) for the pollutants NO_2, SO_2, and RSPM. There are 31 monitoring stations established by State and Central Government in the study area. In view of the fact that this study focuses on interpolation, to obtain more accuracy, sampling points from the neighbouring states are also taken into consideration. Four and three sampling stations are taken from Kerala and Andhra Pradesh respectively. The annual mean of the year 2015 of the pollutants NO_2, SO_2 and RSPM were used for the study.

B. Methodology

1) Spatial Interpolation Techniques:

Spatial continuous data take on a noteworthy role in enhancement, vulnerability evaluation, and administration in environmental organization. Air Pollution data collected at the sampling stations is characteristically point sources. However, for decision making, spatial continuous data are required for environmental managers and scientists over the region of interest for justified interpretations.

Figure.2 represents the methodology in a comprehensive approach. To facilitate effortless utilization and analysis, the first step in this study was to accumulate the necessary data and amalgamate them into a spatial database using GIS. An exploratory data analysis is done to detect abnormalities in the data and also allows us to study the spatial distribution and the variation among them within the study area. The second step is to identify the interpolation methods for the study. The interpolation methods used for this research are Inverse Distance Weighting (IDW), Radial Basis Function (RBF), Local Polynomial Interpolation (LPI), and Global Polynomial Interpolation (GPI) which falls under deterministic approach of interpolation provides no estimation of errors with predicted values and Kernel Interpolation with Barriers (KI) and Diffusion Interpolation with Barriers (DI) come under interpolation with barriers category.

Inverse Distance Weighting (IDW) is based on a hypothesis that the quantity of influence of in close proximity sample points ought to be larger than the effect of more remote points. Equation 1 gives the general formula for IDW.

$$Z(s_o) = \square \lambda_i Z(s_i) \qquad \text{Eq. 1}$$

Where,

$Z(s_o)$ – value being predicted for the target location s_o

λ_i – Weights assigned to each measured point

$Z(s_i)$ – Observed value at location s_i

Radial basis Function (RBF) is suitable estimators for irregularly distributed data. The equation is generalised in Eq. 2. It is based on an assumption that the estimated value can be approximated to several degree of meticulousness by summing up of standard mathematically distinct values

Figure 1: Study Area with Sampling Locations

$$. \; y(x) = \square \, w_i \varphi (\| x - c_i \|) \qquad \text{Eq. 2}$$

Where,

y – Estimation function

φ – Radially symmetrical Function

c_i– Centres (data points)

w_i– Coefficients

Global polynomial interpolation (GPI) is a deterministic approach of interpolation that fits a plane, typically polynomial, through the measured data points. It is usually used to fit a surface where the region varies slowly over the area of interest.Local Polynomial Interpolation (LPI) method uses a particular sample of recognized points from the entire dataset and a polynomial equation to assess unidentified values.

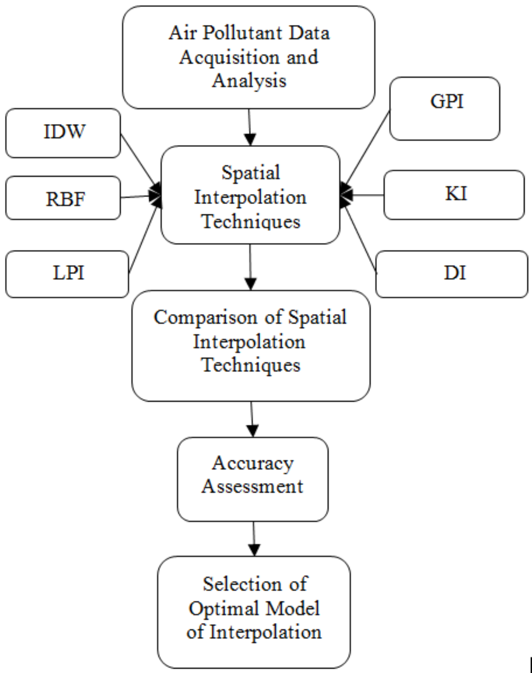

Figure 2: Methodology -Flowchart

Diffusion interpolation with Barriers (DI) depicts how particles diffuse inside a particular barrier. It refers to the gradual flow of particles. Kernel Interpolation with Barriers (KI) is a variation of first order LPI in which the volatility in the calculation is prohibited by means of regression coefficients similar to ridge regression.

For the choice of optimal interpolation method for each of the air pollutant (NO_2, SO_2, RSPM), many tests has been conducted with the interpolation techniques using GIS. 20% of the point data were taken as testing samples and remaining 80 % are considered as training samples. For each of the test conducted, 20% of known sampling stations were eliminated from the spatial database and the interpolation was conducted with the remaining 80% training data sets using GIS. Then, the interpolated values of the five sampling stations are identified and compared with their known original input dataset values by three iterations. As a result a total of three maps are obtained for each pollutant which is then overaid with equal weightage to acquire the interpolation map.

C. Accuracy Assessment:

To decide the optimal model, this study uses error metrics for cross-validation which is a prevailing tool. There are different types of accuracy metrics available such as Mean Absolute Error (MAE), Geometric Mean Absolute Error (GMAE), and Mean square Error etc. The statistical accuracy method used in this study is Root Mean Square Error (RMSE).

RMSE (Eq.3) is more powerful and useful when large errors are particularly undesirable, since it is been squared even before it is averaged. It gives relatively high weight to normal errors. It also increases with the variance of the frequency distribution of error magnitudes.

$$RMSE = \sqrt{} \, [1/n \, \Box \, | O_i - E_i|^2] \quad \text{Eq.3}$$

Where,

O_i – Observed Value

E_i – Estimated Value

This method allows us to validate and assess the accuracy of each interpolation method for the study area. It consisting in predicting the values of a sampling point based on the remaining points. Later the difference between the measured and observed values can be calculated and the RMSE is applied, thus using statistical accuracy indicators. Indeed, all the six interpolation methods are run in the same manner for each of the pollutant for all the stations in Tamil Nadu. To maintain more accuracy a few sampling stations were identified in the neighbouring states (Kerala and Andhra Pradesh) along the boundary of Tamil Nadu.

III.Results and Discussions

Figure .3 graphically represents the comparison between the sampled values and the interpolation values of all the six interpolation methods for the air pollutant NO_2. It is very apparent from the graph that the IDW follows more or less identical inclination as the sampled values, therefore having the lowest RMSE among all the other interpolators. The comparison of RMSE for each of the individual interpolation method for the pollutant NO_2 is represented in Fig.4.

It is palpable from the above graph that IDW is the top interpolator for NO_2 which has the least RMSE value of 3.385 followed by KI with value 3.683. RBF and LPI have a medium range RMSE of 4.135 and 4.197 respectively. While GPI has the highest range implicating that employing the method for interpolation will rise to spurious results.

The order of the interpolation methods according to the value of RMSE is IDW < KI < RBF <LPI < DI < GPI.

The evaluation flanked by the sampled values and the interpolated values of SO_2 is given in the structure of graphical illustration in Fig.5. All the six interpolators are plotted by the side of sampled values. It is evident that RBF and GPI follow a very close tendency in the direction of the original values followed by IDW and LPI with trivial variations. Obviously KI and DI did not follow any inclination towards the sampled values.

Figure .6 shows the RMSE values calculated for all the spatial interpolation methods for the pollutant SO_2. Evidently RBF has the lowest RMSE value of 2.876 followed by GPI with a value of 2.959. IDW and LPI almost have similar values 3.028 and 3.041 respectively with a minute difference of 0.013. As a result they both occupy more or less same position after RBF and GPI. Also, KI and DI have similar values of 3.240 and 3.272 with a variation of 0.032; consequently they occupy almost similar position after IDW and LPI. The orders of RMSE of the interpolators are in the following order: RBF < GPI < IDW < LPI < KI < DI.

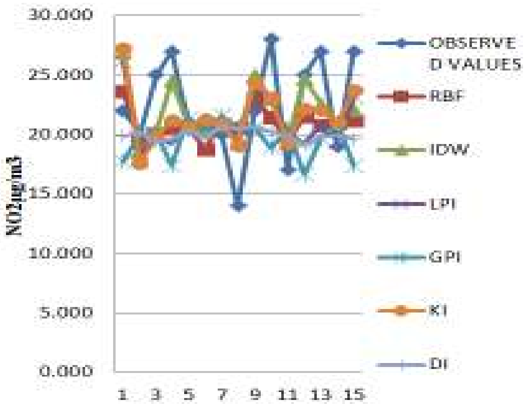

Figure 3: Comparison of observed and estimated values of NO_2 by the Interpolation Methods

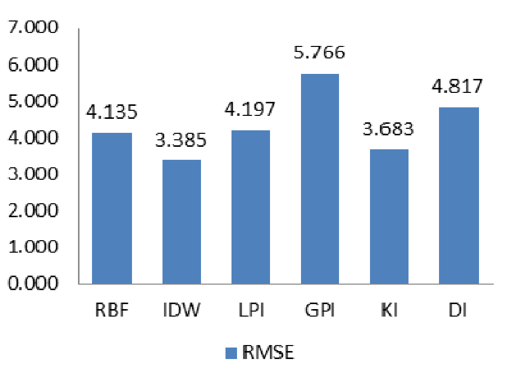

Figure 4: RMSE for each interpolation Method for NO_2

Correspondingly, Fig.7 depicts the association between the sampled values and the interpolated values for the pollutant RSPM for all the six interpolators. It is unambiguous that the interpolator GPI is the most accurate following the movement of the original known values whilst others does trail with varying values which visibly indicates that GPI is the superlative method for interpolating RSPM in the study area.

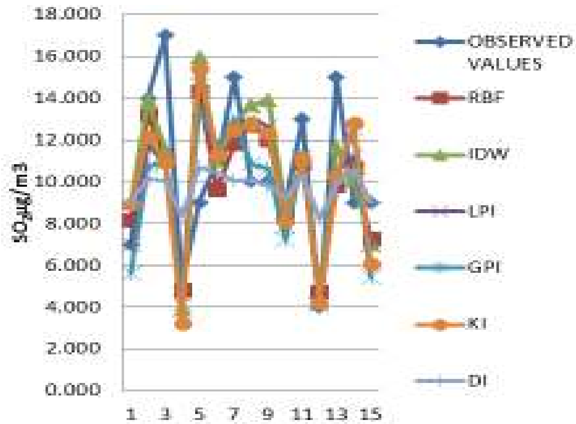

Figure 5: Comparison of observed and estimated values of SO_2 by the Interpolation Methods

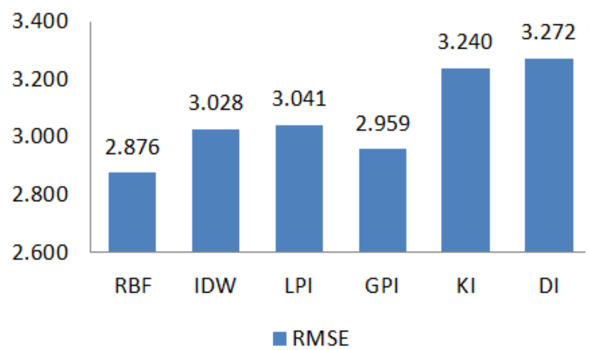

Figure 6: RMSE for each interpolation Method for SO_2

Also, Fig.8 represents the assessment of the RMSE values calculated for the six interpolators for the pollutant RSPM. Unmistakably, GPI has the lowest value RMSE of 21.528 thereby the best method for RSPM. DI has the least RMSE value of 22.958 next to GPI. The methods RBF and IDW have values of 23.210 and 23.788 respectively. LPI and KI have the largest RMSE value when compared with other methods thereby not suitable for the interpolation of RSPM. The order of the RMSE is in the following order: GPI < DI < RBF < IDW < LPI <KI.

The interpolation map of NO_2 by IDW method for the annual mean of the year 2015 is given in Fig.9 (a). The interpolation maps of SO2 and RSPM by RBF and GPI method of interpolation is given in Fig.9 (b) and 9(c) respectively.

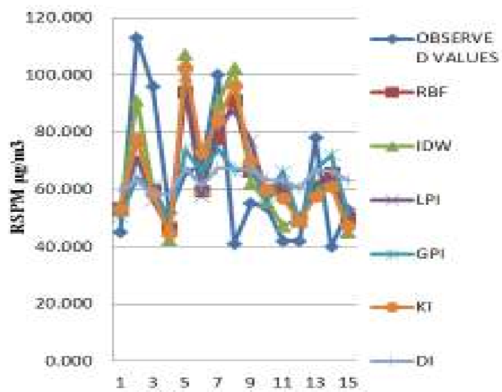

Figure 7: Comparison of observed and estimated values of RSPM by the Interpolation Methods

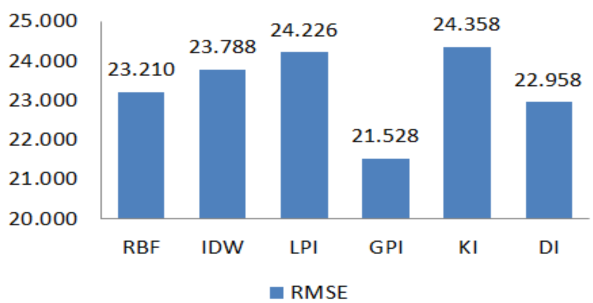

Figure 8: RMSE for each interpolation Method for RSPM

The interpolation map of NO_2 by IDW method for the annual mean of the year 2015 is given in Fig.9 (a). The interpolation maps of SO2 and RSPM by RBF and GPI method of interpolation is given in Fig.9 (b) and 9(c) respectively.

The causes of NO_2 emissions are predominantly heavy vehicular traffic in major cities. Fig.9 (a) depicts that the concentration of NO_2 is more concentrated in the interior parts of Tamil Nadu. This is due to the fact that the mode of transportation is primarily by road through public transportation or own vehicles and lack of local railway transport when compared with major urban cities like Chennai. Moreover, the increase the vehicles lead to traffic congestion and release many toxic pollutants. The quantum of vehicular pollutants emitted is highest in Chennai followed by Coimbatore, Salem, Madurai, Trichy and Thirunelveli. Also the pollution is more concentrated in the major cities like Chennai, Trichy and thoothukudi due to the increase in heavy vehicles population owing to the increase in industries. Concentration is low in various parts such as Nilgiri, Kanyakumari and parts of Vellore.

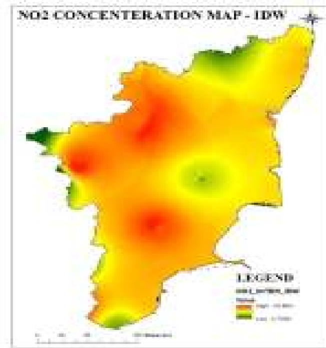

Figure 9 (a): Spatial Interpolation of NO_2 by Inverse Distance Weighting Method (IDW)

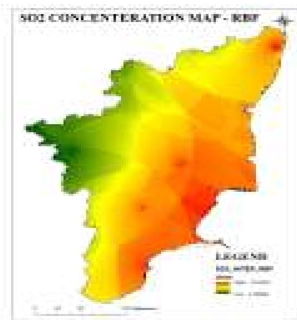

Figure 9 (b): Spatial Interpolation of SO_2 by Radial Basis Function (RBF)

The Central Pollution Control Board (CPCB) has acknowledged seventeen categories of industries as drastically polluting which includes exceedingly polluting industries such as integrated iron and steel, thermal power plants, copper/zinc/aluminium smelters, cement, oil refineries, petrochemicals, pesticides and fertiliser units. Tamil Nadu has a significant occurrence in cement, sugar and fertilizer industries which leads to the increase emission of SO_2.

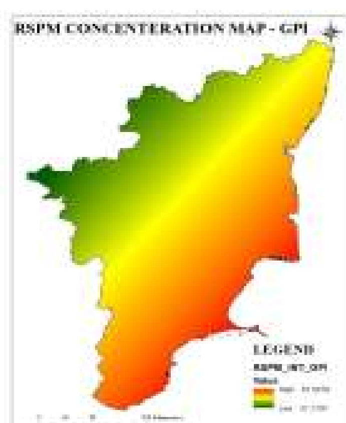

Figure 9 (c): Spatial Interpolation of RSPM by Global Polynomial Interpolation (GPI)

Figure.9 (b) and Figure.9 (c) conveys that the SO_2 and RSPM concentration is more in the places in close proximity to coastal areas which is inhabited with industries and construction activities and it fades out towards the interior parts of Tamilnadu. The spatial distribution of the pollutant is also governed by the factors like wind direction and wind velocity which tend to move the pollutant in any direction from its source. Also the year 2015 had predominant rains in most of the parts of Tamil Nadu which had resulted in dissipation of the pollutants which results in low concentration of the pollutant which were expected to be high. This study has revealed that pollution is intense in industrial sectors and that the input to pollution is habitually inconsistent to its contribution to the industrial output.

IV. Conclusion

To measure the concentration of the air pollutant, visiting each and every location in the area of study is typically complicated or expensive. As an alternative, distributed sample point locations can be selected and a predicted value can be assigned to all other locations. In this study the various interpolation techniques (IDW, RBF, GPI, LPI, GI, and KI), which can be employed for air pollution point estimation has been discussed. Also the best optimal interpolation model has been recognized for the air pollutants NO_2, SO_2, RSPM from iterations by assessing the accuracy amid the observed and the estimated values. This was accomplished by the statistical error metrics RMSE. Cross-validation procedure by means of RMSE played a significant responsibility in assessing the accuracy of each and every interpolation model. It was found that for the study area, for NO_2, IDW has been acknowledged as the most advantageous interpolation model. Furthermore, for SO_2 and RSPM, RBF and GPI has been recognized as the best interpolation method for the study area respectively. The use of identifying the optimal spatial interpolation provides maps to understand the spatial distribution and variation of the air pollutant. This study also reveals that there is no universal rule in identifying the best interpolation method. It is better if, for every study area, comparing methods and assessing the accuracy through statistical error metrics is the best way [10]. More accurate results can be obtained if the seasonal data had been included; the difference in spatial distribution of the pollutant concentration had been more useful in seasonal predictability. Other parameters like wind velocity and wind direction can also be included for the study which will be useful to justify the interpretations over the pollutant maps.

References

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

[2] Daniel Kurtzman and Ronen Kadmon, 1999, ‘Mapping of temperature variables in Israel: a comparison of different interpolation methods’, CLIMATE RESEARCH, Vol. 13, pp 33–43.

[3] Mustafa GÜLER and Tekin KARA, 2014, ‘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.

[4] Xian Luo and YoupengXu, Yi Shi, 2011,’ Comparison of Interpolation Methods for Spatial Precipitation under Diverse Orographic Effects’, IEEE.

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

[6] PhataraponVorapracha, PongtepPhonprasert, SuparadaKhanaruksombat and nuchanapornPijarn, 2015, ‘A Comparison of Spatial Interpolation Methods for predicting concentrations of Particle Pollution (PM10)’, International Journal of Chemical, Environmental & Biological Sciences, Vol.3, Issue.4, ISSN: 2320 - 4087, pp 302 - 306.

[7] D Rojas Avallaneda, 2007, ‘Spatial interpolation Techniques for estimating the levels of pollutant concentrations in the atmosphere’, Revista Mexicana De Fisica, Vol.53, Issue.6, pp 447 – 454.

[8] Shareef M, Husain T and Alharbi B, 2016, ‘Optimization of Air Quality Monitoring Network Using GIS Based Interpolation Techniques’, Journal of Environmental Protection, Vol.7, pp 895 – 911.

[9] Yue Sun, Shaozhong Kang, Fusheng Li and Lu Zhang, 2009, ‘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, Issue.10, pp 1163 – 1170.

[10] LamiaaKhazaz, HassaneJararOulidi, Saida El Moutaki, Abdessamad Ghafiri,2015, ‘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.