Spatial analysis of rainfall and lead concentration data using different interpolation methods

Alex Mathew[#1], Surya Durbha [*2]

[#] *Centre of Studies in Resource Engineering, IIT Bombay, Mumbai*

**Centre of Studies in Resource Engineering, IIT Bombay, Mumbai*

[1] pingalex94@gmail.com

[2] surya.durbha@gmail.com

***Abstract* — Spatial interpolation techniques are commonly employed for creating continuous data (raster data) from a distributed set of data points over a geographical region. This paper compares various spatial interpolation techniques like inverse distance weighting, linear trend surface and kriging using spherical, exponential and gaussian model. Each of these techniques are underpinned by different assumptions (e.g. whether the data fits a normal distribution). The implementation is carried out using gstat library in R.**

I. INTRODUCTION

A surface is a continuous field of values that may vary over an infinite number of points. For example, points in an area on the earth's surface may vary in elevation, proximity to a feature, or concentration of a particular chemical. Any of these values may be represented on the z-axis in a three-dimensional x, y, z coordinate system, so they are often called z-values. Because a surface contains an infinite number of points, it is impossible to measure and record the z-value at every point.

A surface model approximates a surface by taking a sample of the values at different points on the surface and then interpolating the values between these points. The interpolation is based on the assumption that spatially distributed objects are spatially correlated; i.e, things that are close together tend to have similar characteristics according to the first law of geography. For example, if it is sunny on one side of the place, you can predict with a high level of confidence that it is sunny 50 m side away from that place. We would be less certain if it was sunny across town and less confident about the state of the weather in the next country. From this analogy, it is easy to see that the values of points close to sampled points are more likely to be similar than those that are farther apart. This is the basis of interpolation.

Here we are comparing three different interpolation techniques - Inverse Distance Weighting (IDW), Trend Surface Interpolation (TSI), and Kriging. IDW Interpolation is an automatic and relatively easy technique, as it requires very few parameters from the operator such as it needs only neighbourhood and exponential parameters. TSI Interpolation is used to fitting a statistical model, a trend surface, through the measured points. Kriging is similar to IDW but the difference is that in this case, weight is based both on the distance between measured and predicted location but also on overall spatial arrangement of points.

II. METHODOLOGY

A. Inverse Distance Weighting (IDW)

IDW interpolation determines cell values using a linearly weighted combination of a set of sample points. The weight is a function of inverse distance. The surface being interpolated should be that of a locationally dependent variable.

Characteristics of interpolated surface can also be controlled by limiting the number of input points used in the calculation of each output cell value. It can be done by two methods either by limiting number of points or by limiting the radius area. The IDW is determined by using following equation:

$$Z_0 = \frac{\sum_1^s Z_i \frac{1}{d_i^k}}{\sum_1^s \frac{1}{d_i^k}}$$

Where, Z_i = Value of known points

Z_0 = Value of unknown points

d_i = Distance from unknown points.

K= Power parameter

S= No of sample points

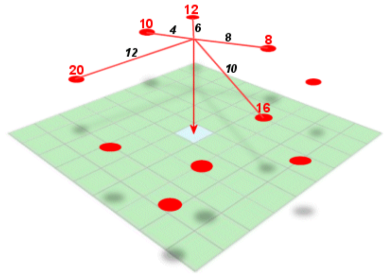

Fig. 1. Inverse Distance Weighted interpolation based on weighted sample point distance

B. Trend Surface

Trend interpolation is just like taking a piece of paper and fitting it in between raised points. Though a flat piece of paper can not capture a landscape containing valley, when you bend a paper then you will get a better fit. Trend is a statistical method that finds the surface that fits the sample points using a least mean squares regression fit. It means surface is created so that for every input points, the sum of the difference between actual value and estimated values is as small as possible. It is an inexact interpolator because output surfaces rarely passes through the input points.

There are two types of Trend interpolation:

a) Linear trend   b) Logistic trend

Here, we deal with 1st order polynomial and second order polynomial linear trend surface methods:

$$Z = a_0 + a_1 x + a_2 y \qquad (1^{st} \text{ order polynomial})$$

$$Z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \qquad (2^{nd} \text{ order polynomial})$$

where,
Z = value at any point
a_i = coefficients estimated in the regression model

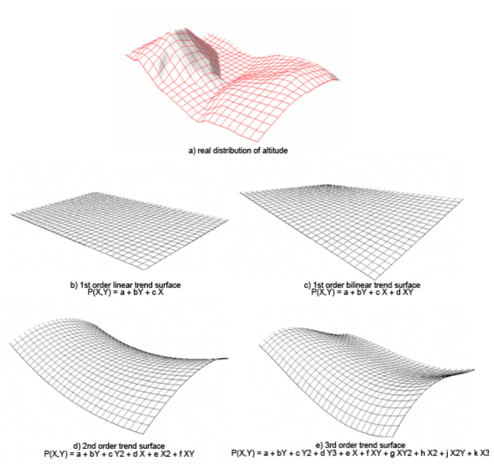

Fig. 2.Some common trend surfaces compared to a real surface

C. Kriging method

Kriging is a stochastic model which assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The Kriging tool fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location. Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface. Kriging is most appropriate when you know there is a spatially correlated distance or directional bias in the data. It is often used in soil science and geology.

Kriging is similar to IDW in that it weights the surrounding measured values to derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

where,

$Z(s_i)$= the measured value at the i th location

λ_i = an unknown weight for the measured value at the i th location

s_0 = the prediction location

N = the number of measured values

In IDW, the weight λi fully depends on the distance to predicted location. However, in kriging method, the weights are based not only on distance between measured point and predicted location but also on overall spatial arrangement of points. Kriging involves two steps:

1. *Plotting the semi-variogram* : It is a plot of semi-variance v/s the distance. For all pairs of locations separated by distance h, the semi-variance formula is given by:

   Semi-variance (h) = 0.5 * average ((value$_i$ – value$_j$)

2. *Fitting a model to the empirical semi-variogram* :

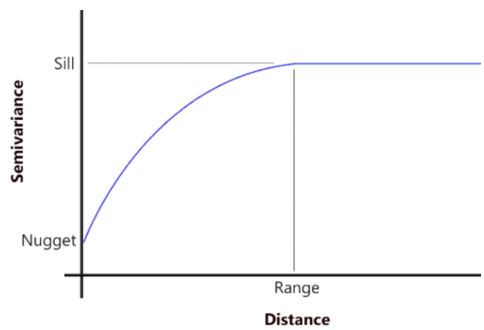

Fig. 3.Illustration of Range, Sill and Nugget in a semi-variogram

To fit a model to the empirical semi-variogram, select a function that serves as your model-

i. *Spherical model:* Shows a progressive fall in spatial autocorrelation (equivalently, an increase of semi-variance) until some distance, beyond which autocorrelation is zero.

$$y(h) = \begin{cases} 0 & , |h| = 0 \\ c_0 + c_1\left\{\frac{3|h|}{2a} - \frac{1|h|^3}{2a}\right\} = c_0 + c_1\left[sph\left(|h|\right)\right], & 0 < |h| \le a \\ c0 + c1 & , |h| > a \end{cases}$$

where,

c_0 = Nugget
c_0+c_1= Sill
a= Range
h = Distance

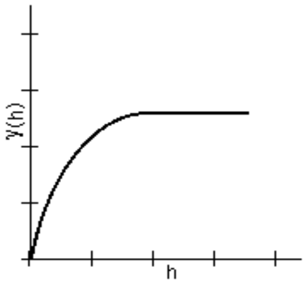

Fig. 4. Spherical semi-variance model illustration

b. Exponential model: Applied when spatial autocorrelation decreases exponentially with increasing distance.

$$y(h) = \begin{cases} 0 & , |h| = 0 \\ c_0 + c_1\left\{1 - \exp\{-\frac{|h|}{a}\}\right\} = c_0 + c_1\left[exp\left(|h|\right)\right], & |h| \ne 0 \end{cases}$$

where,

c_0 = Nugget
c_0+c_1= Sill
a= Range
h = Distance

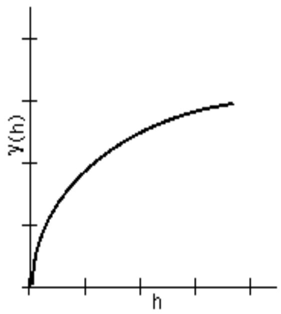

Fig. 5. Exponential semi-variance model illustration

III. RESULTS

Following results were obtained from various interpolation techniques in gstat library in R for Rain and lead concentration datasets:

A. *Inverse Distance Weight (IDW)*

Results of IDW interpolation method using different inverse exponent powers on rain and lead concentration data is shown below (Fig 6 to Fig 13). It can be observed that for higher exponent powers much finer variations can be obtained.

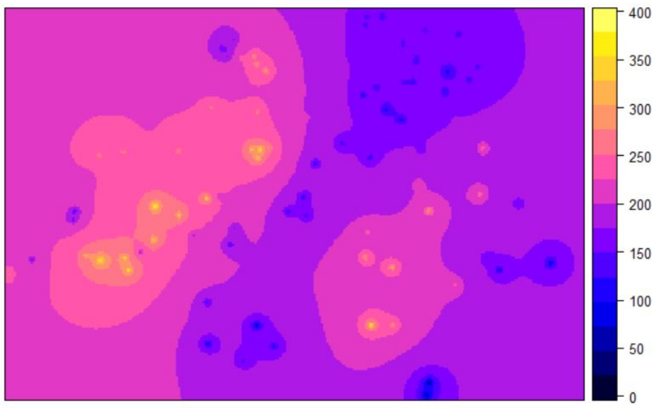

Fig. 6. *Rain : IDW, power = 1*

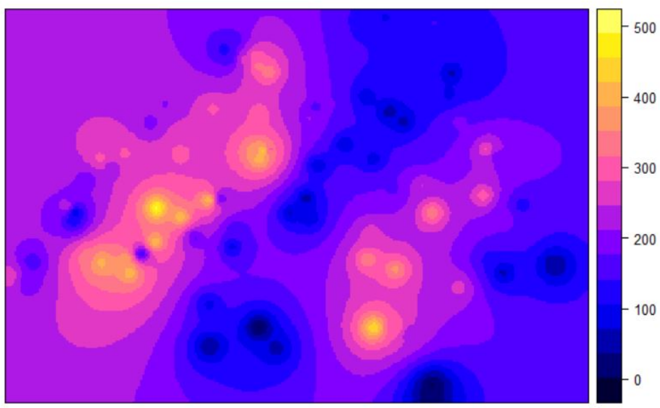

Fig. 7. *Rain : IDW, power = 2*

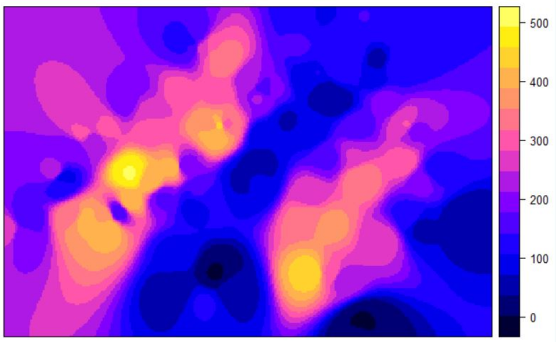

Fig. 8. *Rain : IDW, power = 5*

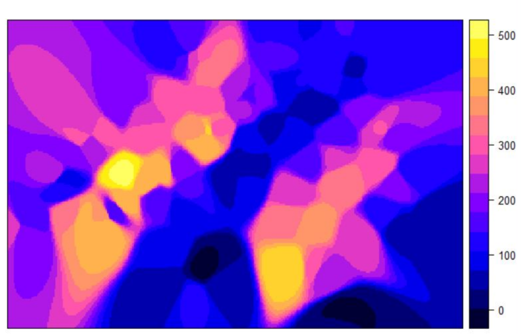

Fig. 9. *Rain : IDW, power = 10*

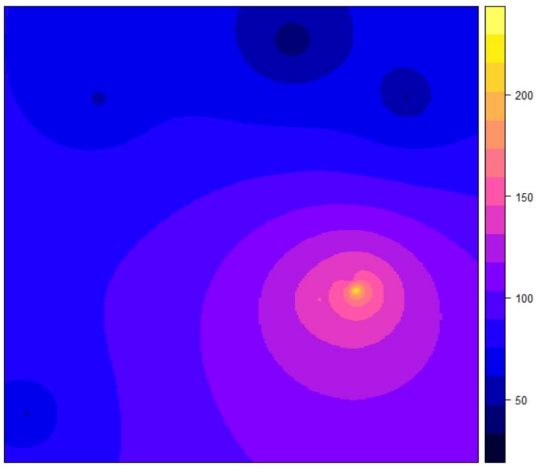

Fig. 10. *Pb concentration : IDW, power = 1*

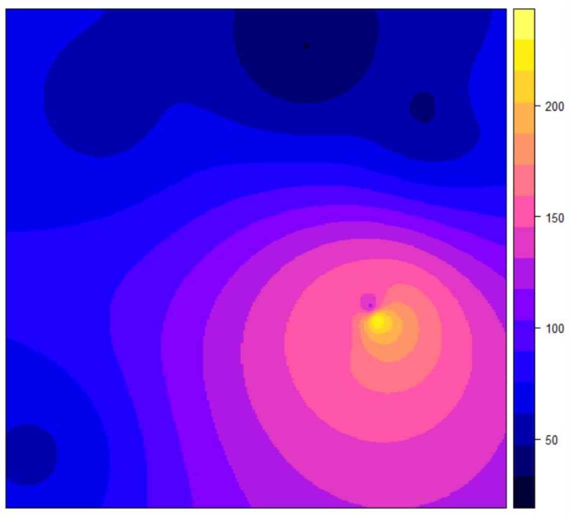

Fig. 11. *Pb concentration : IDW, power = 2*

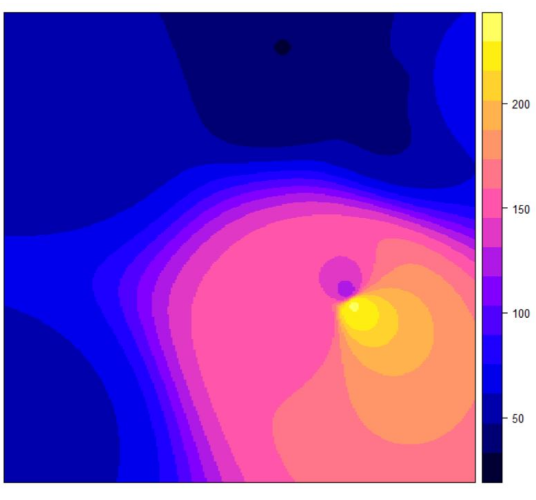

Fig. 12. *Pb concentration : IDW, power = 5*

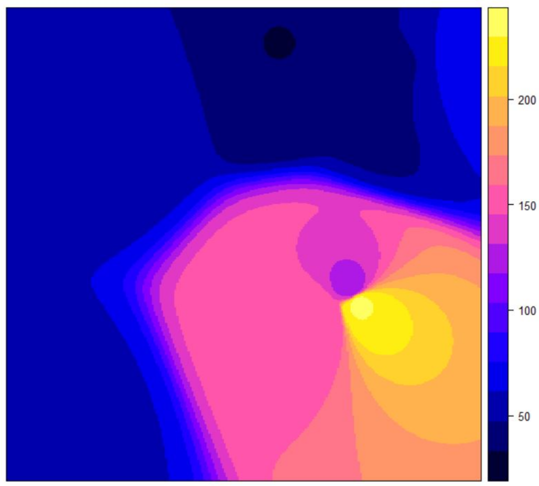

Fig. 13. *Pb concentration : IDW, power = 10*

B. *Trend Surface Interpolation (TSI)*

We employ first order and second order polynomial functions to interpolate on the rain and lead concentration data. Fig 14 and Fig 15 shows the output for rain and lead data respectively using first order. Fig 16 and Fig 17 shows it using second order polynomial. It is clear that the surfaces are smoother when a higher order polynomial is used. On comparing to IDW, in trend surface output the plot is varying in a uniform manner since the function is applied globally.

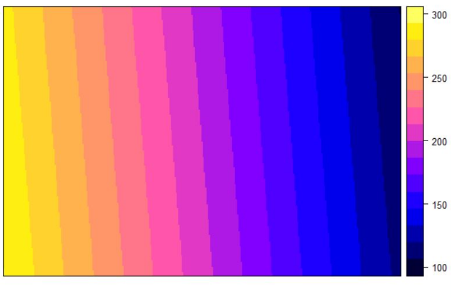

Fig. 14. *Trend surface for rain data (1st order polynomial)*

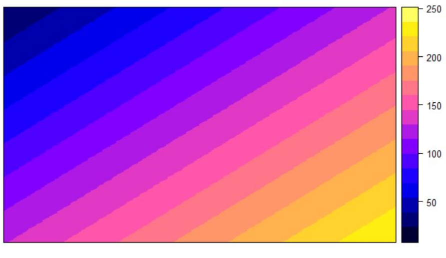

Fig 15. *Trend surface for lead data (1st order polynomial)*

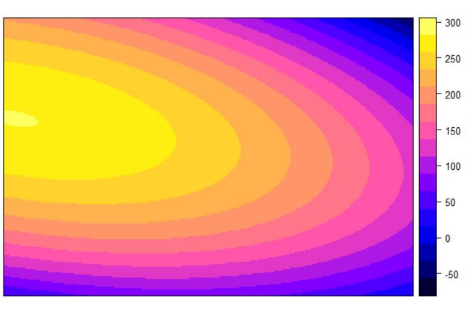

Fig 16. *Trend surface for rain data (2nd order polynomial)*

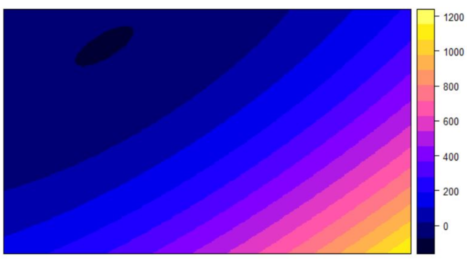

Fig 17. Trend surface for lead data (2nd order polynomial)

C. *Krigging*

Semivariograms for rain and lead data are created by giving respective sill, nugget and range values for different models like exponential, spherical and gaussian are plotted (Fig 18 to Fig 22). Corresponding kriging surfaces are plotted for each case (Fig 23 to Fig 27).

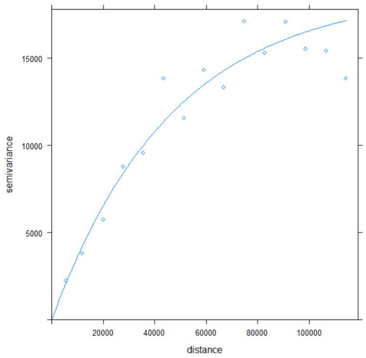

Fig 18. *Semivariogram - rain data (Exponential model)*

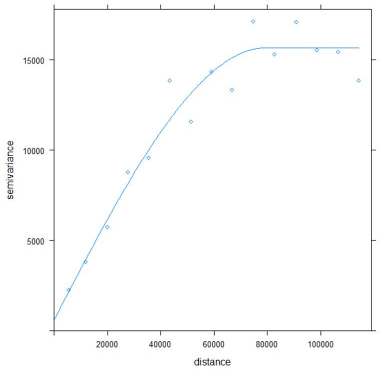

Fig 19. *Semivariogram - rain data (Spherical model)*

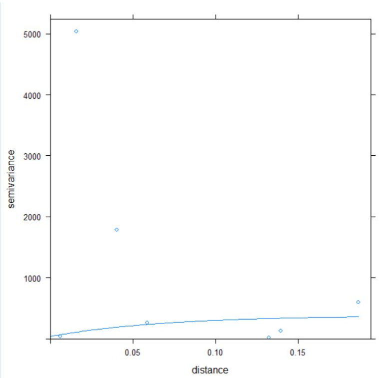

Fig 20. *Semivariogram - lead data (Exponential model)*

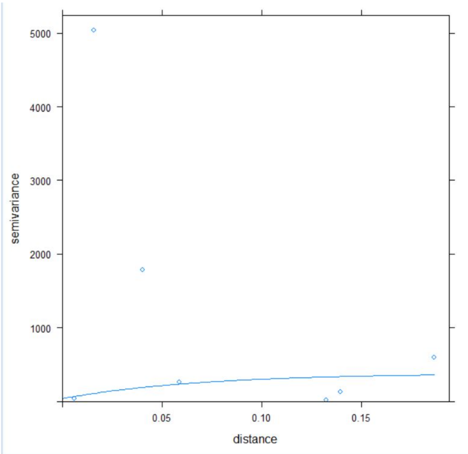

Fig 21. *Semivariogram - lead data (Spherical model)*

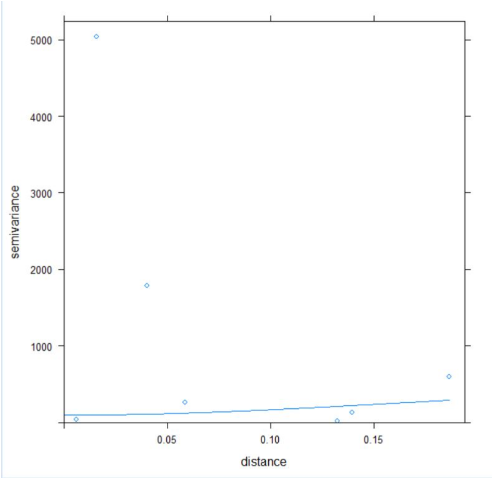

Fig 22. *Semivariogram - lead data (Gaussian model)*

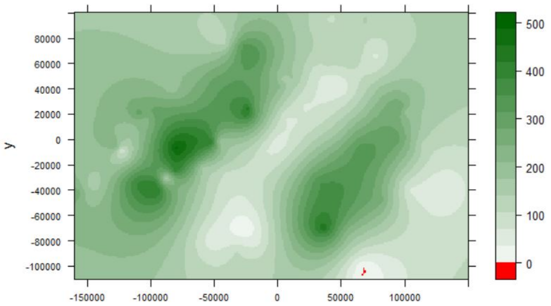

Fig 23. *Kriging surface - rain data (Exponential model)*

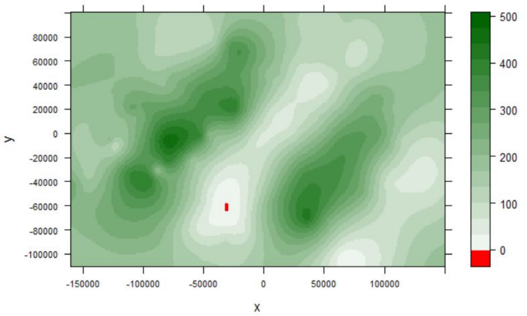

Fig 24. *Kriging surface - rain data (Spherical model)*

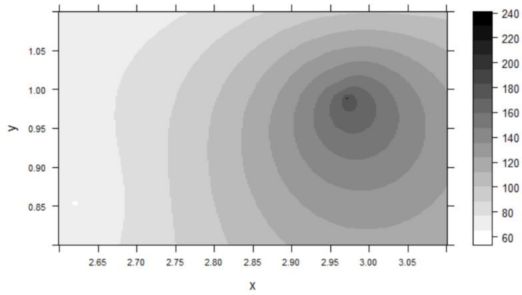

Fig 25. *Kriging surface - lead data (Exponential model)*

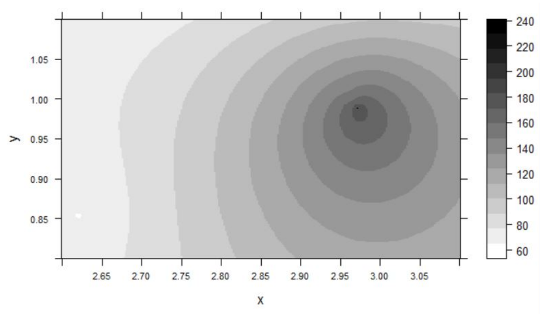

Fig 26. *Kriging surface - lead data (Spherical model)*

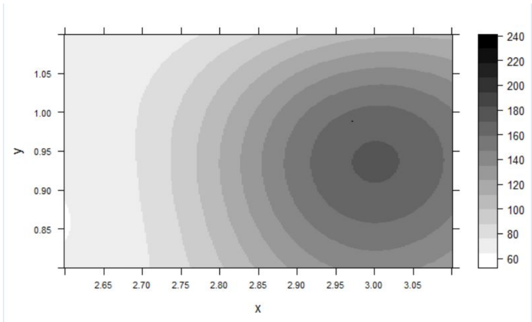

Fig 27. *Kriging surface - lead data (Gaussian model)*

IV. CONCLUSION

The available rain and lead data samples were interpolated using three different methods — the inverse distance weighted method, trend surface method and kriging method. Trend surface is giving a surface which shows the general behaviour of the whole region under consideration. IDW and Kriging are able to provide variations more on local basis, and kriging method is able to provide smoother result since IDW is just based on the distance from the sample point.

The output greatly depends on the accuracy and spread of the sample data. In our example the rain data contains a lot of sample points with a clear pattern whereas the lead concentration data has very a smaller number of sample points which appeared more like random scattering. The difference in accuracy of the output due to the nature of sample data can be observed by comparing the output surfaces. The accuracy can be assessed via cross-validation.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to Prof. Surya S Durbha, CSRE Dept., IIT Bombay for his support and guidance. I would also like to thank my batchmates and seniors who gave important suggestions to improve my work.

REFERENCES

1. http://planet.botany.uwc.ac.za/nisl/GIS/spatial/
2. https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/kriging-in-geostatistical-analyst.html
3. https://www.rdocumentation.org/packages/gstat/versions/2.0-3