Probabilistic approach for epistemic analysis of lake tapping of koyna dam for future forecasting.

Rajvardhan S. Patil
IITKharagpur, Kharagpur, West Bengal ,721302

Abstract: *The statistical technique like confidence interval (CI) is used to determine the confidence limits of water head over lake tapping site of Koyna Dam (India, Maharashtra) from water level data taken from government of Maharashtra 2000-2010, which helps in future forecasting. It is the statistical technique which gives the Probable Value of the water head in the relative month. By studying these probable values of water heads, monthly variation in the heads seasonal variation can be achieved. Contour maps and variograms are studied to check the variation of Water head from year-2000-2010. It also helps to determine the probable nature by knowing the confidence Interval of heads and it is more useful and helpful for the water management of rain water which is accumulated on ground by the means of large reservoirs. Predication of rainfall and head variation is achieved.*

Keywords: *Statistical Technique, Confidence Interval, surfer10, Contour maps, variograms, Koyna Dam, Water head, Lake Tapping etc.*

I.   Introduction

1.0 Overview:
Seasonal Variation of rainfall is found in INDIA. Due to this variation in rainfall, the water heads in the water bodies or in dams may be erratic. This water is used for various purposes like Domestic purpose, Industrial purpose, and Irrigation purpose and for Power generation. So it is obligatory to go Proper Management in order to avoid the wastage of water and provide the solution to the social problems like water scarcity, drought conditions ETC.

To maintain the water level in the dam and for proper future planning of water management this project is to be implemented. During this project, the detailed study conducted regarding the lake tapping site of **Koyna Dam** and the reservoir of this dam named as Shivajisagar Reservoir. The koyna dam is one of the largest dams in India. It is rubble concrete dam of maximum height of 103m and length 807m. The catchment area of the dam is 891.78sqkm. And the capacity of the dam is 98.78TMC. The water which in the Koyna Dam used by Maharashtra, Karnataka and Andhra Pradesh also, for various purposes like Irrigation, Domestic and for Power generation. The Koyna Project consist of 4 stages of power generation and the total capacity of electricity generation is 1960MW., all the generation units are located underground power houses excavated deep inside the surrounding the mountains of the Western Ghats.

In these 4 stages of power generation, the 4th stage of the generation is done by a unique, innovative and non conventional intake system by piercing the lake from bottom by blasting the rock plug using explosives, the double **Lake Tapping** is done for stage IV and this the first of its kind in Asia on 13 March 1999.

For this project the data of water head available at the Lake Tapping site are taken from IRRIGATION Department Of Government Of Maharashtra. These water heads are the daily water heads from the year 2000 to 2010. By conducting study of heads in each month of 10years, apply the project methodology on the monthly head data by using **Surfer10 Software, Variogram Model** and **CONFIDANCE_Interval Technique** to determine the confidence limits of water head which helps in future forecasting. It is the statistical technique which gives the Probable Value of the water head in the relative month. By studying these probable values of water heads, monthly variation in the heads of water at the site is achieved. And it helps to determine the probable nature by knowing the confidence Interval of heads And it is more useful and helpful for the management of rain water which is accumulated on ground by the means of large reservoirs.

1.1 Lake Tapping:
Under Water Lake tapping is a Norwegian technique developed in that country mainly to tap the inland natural lakes located mountains bellows their normal levels for electricity generation and drinking water supply.

In these 4 stages of power generation, the 4th stage of the generation is done by a unique, innovative and non conventional intake system by piercing the lake from bottom by blasting the rock plug using explosives, The Double **Lake Tapping** Is Done For Stage Iv And This The First Of Its Kind In Asia On 13 March 1999.

Government of Maharashtra water resources department, koyna hydroelectric project- st.4, the fourth phase of the Koyna hydroelectric power station, considered the lifeline of Maharashtra, will be completed by

June 1998. In the fourth phase, four Indian- generators will be set up to increase the power station's supply capacity by 250 megawatts.

Fig.2.1.1 Schematic drawing of lake tapping at stage-4th and stage-4thB in Koyna Dam

1.2 Statistical Techniques:

The Contouring and 3D Surface Mapping for Scientists and Engineers. In Golden SoftwareInc2011. www.GoldenSoftware.com is used powerful contouring, gridding, and surface mapping program for scientists, engineers, educators, or anyone who needs to generate maps quickly and Easily. Surfer is a grid-based mapping program that interpolates irregularly spaced XYZ data into a regularly spaced grid. The Surfer, allowing you to produce the map that best represents your data.

☐ Alaa A. Masoud ,(July 2014) has emphasized on Extensive urban, agricultural and industrial expansions on the western fringe of the Nile Delta of Egypt exerted much load on the water needs and lead to groundwater quality deterioration. Documenting the spatial variation of the groundwater quality and their controlling factors is vital to ensure sustainable water management and safe use. A comprehensive dataset of 451 shallow groundwater samples were collected in 2011 and 2012. On-site field measurements of the total dissolved solids (TDS), electric conductivity (EC), pH, temperature, as well as lab-based ionic composition of the major and trace components were performed. Groundwater types were derived and the suitability for irrigation use was evaluated. Multivariate statistical techniques of factor analysis and K-means clustering were integrated with the geostatistical semi-variogram modeling for evaluating the spatial hydro chemical variations and the driving factors as well as for hydrochemical pattern recognition.

☐ C. Hernández, R. Tutsch ,(2013) have given contrast to classical assembling techniques for high variation processes, the novel approach is presented which is based on the Statistical Feed-Forward Control Model (SFFCM), accounts the dynamic evolution of the variation over time. It is, then, interesting to discover how the presence of response delays affects its effectiveness.

☐ George Makrides , Bastian Zinsser, Markus Schubert,George E. Georghiou, (May 2014) presented a comparison of the annual performance loss rate (PLR) of twelve different grid-connected photovoltaic (PV) technologies based on outdoor field measurements. The annual DC performance loss rates of the installed PV technologies are obtained by using linear regression and classical series decomposition applied on the monthly DC performance ratio (PR) time series over five years (June 2006–June 2011).

☐ A. Naseerutheen,A. Chandrasekaran ,A. Rajalakshmi, R. Ravisankar,(2013) have provided the analysis of archaeological pottery, Energy Dispersive X-ray florescence analysis has been utilized to establish the elemental concentrations up to fourteen chemical elements for each of 14 archaeological pottery samples from Vellore Dist, Tamil Nadu, India. The EDXRF results have been processed using two multivariate statistical cluster and principal component analysis (PCA) methods in order to determine the similarities and correlation between the selected samples based on their elemental composition.

☐ María-Dolores Cubiles-De-La-Vega, Antonio Blanco-Oliver, Rafael Pino-Mejías, Juan Lara-Rubio,(Dec 2013) A wide range of supervised classification algorithms have been successfully applied for credit scoring in non-microfinance environments according to recent literature. However, credit scoring in the microfinance industry is a relatively recent application, and current research is based, to the best of our knowledge, on classical statistical methods.

1.3 Mathematical Modeling:-

ES, Harris S, Soulsbury CD, Richards SA, Stephens PA (2010) (October 25, 2010) Uncertainty in Population Growth Rates: Determining Confidence Intervals from Point Estimates of Parameters Combining this with resampling techniques , we show that confidence intervals for population growth estimates are easy to derive. We used similar techniques to examine the effects of sample size on uncertainty. We show that uncertainty head level can be high, even for relatively well-studied head of water level.

The George Institute for International Health, PO BOX M201, Missenden Road, NSW 2050, Australia.(26 February 2009) Generalized lambda distributions (GLD) can be used to fit a wide range of continuous data. It is useful in estimating confidence intervals for quantiles of continuous data. This technique can also be used to find confidence interval for the mode of a continuous data as well as comparing In Balabdaoui, Rufibach, and Wellner (2009), pointwise asymptotic theory was developed for the nonparametric maximum likelihood estimator of a log-concave density.

In MahdisAzadbakhsh, Hanna Jankowski (24 January 2014) in this theory is used to develop point wise confidence intervals for the true log-concave density. The finite sample size behaviour of these estimated confidence intervals is then studied via a simulation study of the empirical coverage probabilities.

In Devenish Nelson two data sets in terms of quantiles. In Yongsong Qin ShichaoZhang ,Xiaofeng Zhu , Jilian Zhang , Chengqi Zhang was developed Estimating confidence intervals for structural differences between contrast groups with missing data. in the refrance of School of Computer Science and Information Technology, Guangxi Normal University, PR China. We apply our approach to UCI datasets to illustrate its power as a new data mining technique for, such as, distinguishing spam from non-spam emails; and the benign breast cancer from the malign one.

In G. Goffaux, M. Remy, A. VandeWouwer (28 February 2013)In vehicle positioning applications, the confidence level in the position and velocity estimates can be even more significant than accuracy. In this study, a probabilistic interval method is proposed, which combines, through union and intersection operations, the information from a possibly uncertain predictor (the vehicle model) and measurement sensor. The proposed method is compared to Kalman filteringand to guaranteed interval estimation in the context of railway vehicles where security is the key objective.

In Martin J Gardner, Douglas G Altman, the Confidence intervals rather than P values: estimation rather than hypothesis testing BRITISH MEDICAL JOURNAL,15 MARCH 1986 One unfortunate consequence has been a shift in emphasis away from the basic results towards an undue concentration on hypothesis testing. It is more useful to present sample statistics as estimates of results that would be obtained if the total population were studied.The confidence interval thus provides a range of possibilities for the population value, rather than an arbitrary dichotomy based solely onstatistical significance. It conveys more useful information at the expense of precision of the P value. However, the actual P value is helpful in addition to the confidence interval, and preferably both should be presented. If one has to be excluded, however, it should be the P value.Confidence intervals, which also have a link to the outcome of hypothesis tests, should become the standard method for presenting the statistical results of major findings.

II.   Statastical Modeling

2.0 Confidence interval:

In statistics, a confidence interval (CI) is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate. A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter (see precision). A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter. It is an observed interval (i.e. it is calculated from the observations), in principle different from sample to sample, that frequently includes the parameter of interest if the experiment is repeated. How frequently the observed interval contains the parameter is determined by the confidence level or confidence coefficient. More specifically, the meaning of the term "confidence level" is that, if confidence intervals are constructed across many separate data analyses of repeated (and possibly different) experiments, the proportion of such intervals that contain the true value of the parameter will match the confidence level; this is guaranteed by the reasoning underlying the construction of confidence intervals. Whereas two-sided confidence limits form a confidence interval, their one-sided counterparts are referred to as lower or upper confidence bounds.

Confidence intervals consist of a range of values (interval) that act as good estimates of the unknown population parameter. However, in infrequent cases, none of these values may cover the value of the parameter. The level of confidence of the confidence interval would indicate the probability that the confidence range captures this true population parameter given a distribution of samples. It does not describe any single sample. This value is represented by a percentage, so when we say, "we are 99% confident that the true value of the parameter is in our confidence interval", we express that 99% of the observed confidence intervals will hold the true value of the parameter. After a sample is taken, the population parameter is either in the interval made or not; it is not a matter of chance. The desired level of confidence is set by the researcher (not determined by data). If a corresponding hypothesis test is performed, the confidence level is the complement of respective level of significance, i.e. a 95% confidence interval reflects a significance level of 0.05. The confidence interval contains the parameter values that, when tested, should not be rejected with the same sample. Greater levels of

variance yield larger confidence intervals, and hence less precise estimates of the parameter. Confidence intervals of difference parameters not containing 0 imply that there is a statistically significant difference between the populations.

In applied practice, confidence intervals are typically stated at the 95% confidence level. However, when presented graphically, confidence intervals can be shown at several confidence levels, for example 50%, 95% and 5%.

A confidence interval is always qualified by a particular confidence level, usually expressed as a percentage ; thus one speak of a "95% confidence interval". The end points of the confidence interval are referred to as confidence limits. For a given estimation procedure in a given situation, the higher the value of confidence level, the wider the confidence interval will be.

The calculation of a confidence interval generally requires assumptions about the nature of the estimation process – it is primarily a parametric method – for example, it may depend on assumptions that the distribution of errors of estimation is normal.

2.0.1 Confidence Limits

Confidence limits are the lower and upper boundaries / values of a confidenc interval that is, the values which define the range of a confidence interval. The upper and lower bounds of a 95% confidence interval are the 95% confidence limits. These limits may be taken for other confidence levels, for example, 90%, 95% 5%.

2.0.2 Why confidence interval...?

- A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.
- If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.
- The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter (see precision). A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.
- Confidence intervals are more informative than the simple results of hypothesis tests (where we decide "reject H0" or "don't reject H0") since they provide a range of plausible values for the unknown parameter.

2.0.3 Formula for confidence interval:

$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

Where, x = mean of sample, σ = Standard deviation
n = nos. of samples, z = Margin of error , $\bar{x}$ = point estimate
Values of Zc; For 95% CI= 1.9599
For 50% CI = 0.6744
For 5% CI = 0.1427

2.1 Variogram:

The variogram characterizes the spatial continuity or roughness of a data set**.** Ordinary one dimensional statistics for two data sets may be nearly identical, but the spatial continuity may be quite different**.** Variogram analysis consists of the experimental variogram calculated from the data and the variogram model fitted to the data**.** The experimental variogram is calculated by averaging one-half the difference squared of the z-values over all pairs of observations with the specified separation distance and direction**.** It is plotted as a two-dimensional graph**.**

Details about the mathematical formulas used to calculate the experimental variogram. The variogram model is chosen from a set of mathematical functions that describe spatial relationships**.** The appropriate model is chosen by matching the shape of the curve of the experimental variogram to the shape of the curve of the mathematical function.

2.1.1 Variogram Depiction:

1) The variogram characterizes the spatial continuity or roughness of a data set**.**
2) Ordinary one dimensional statistics for two data sets may be nearly identical, but the spatial continuity may be quite different**.**
3)Refer to the Surfer User's Guide and the topic Variogram Model Graphics in the Surfer Help for graphs illustrating the curve shapes for each function.

4)To account for geometric anisotropy (variable spatial continuity in different directions), separate experimental and model variograms can be calculated for different directions in the data. Set.

$$\gamma(\Delta x, \Delta y) = \frac{1}{2N(\Delta x, \Delta y)} \sum_{i,j \epsilon s(\Delta x, \Delta y)} (Zi - Zj)^2$$

Where N(h)=Set of pairs of observations i&j

2.1.2Parameters:
o Nugget :-the height of the jump of the semivariogram at the discontinuity at the origin.
o Sill:-limit of the variogram tending to infinity lag distances.
o Range:-the distance in which the difference of the variogram from the sill becomes negligible.

2.1.3 Varigram models:
o Exponential Variogram model
o Spherical Variogram model
o Gaussian Variogram model
Gaussian Variogram model is applied……………………..!

2.1.4The mathematical definition of the variogram:

$$\gamma(\Delta x, \Delta y) = \frac{1}{2}\varepsilon[\{Z(x + \Delta x. y + \Delta y) - Z(\quad)x. y\}^2]$$

where , Z(x, y) is the value of the variable of interest at location (x, y), and € [ ] is the statistical expectation operator. Note that the variogram, γ( ), is a function of the separation between points (Δx, Δy), and not a function of the specific location (x, y). This mathematical definition is a useful abstraction, but not easy to apply to observed values.

Consider a set of n observed data: {(x1, y1, z1),(x2, y2, z2), … (xn, yn, zn)}, where (xi,yi) is the location of observation i, and zi is the associated observed value. There are n (n - 1)/2 unique pairs of observations. For each of these pairs we can calculate the associated separation vector: (Δxi,j, Δyi,j) = (xi-xj, yi-yj)

When we want to infer the variogram for a particular separation vector, (Δx, Δy), we will use all of the data pairs whose separation vector is approximately equal to this separation of interest: (ΔXi,j, ΔYi,j) » (Δx, Δy)

Let S(Δx, Δy) be the set of all such pairs:
S(Δx, Δy) = { (i,j) | (Δxi,j, Δyi,j) » (Δx, Δy) }

Furthermore, let N(Δx, Δy) equal the number of pairs in S(Δx, Δy). To infer the variogram from observed data we will then use the formula for the experimental variogram.

$$\gamma(\Delta x, \Delta y) = \frac{1}{2N(\Delta x, \Delta y)} \sum_{i,j \epsilon s(\Delta x, \Delta y)} (Zi - Zj)^2$$

That is, the experimental variogram for a particular separation vector of interest is calculated by averaging one-half the difference squared of the z-values over all pairs of observations separated by approximately that vector.

2.2 Software used:

Surfer is a grid based contour program. Gridding is the process of using original data points(observations) in an XYZ data file to generate calculated data points on a regularly spaced grid (agrid [GRD] file). Interpolation schemes estimate the value of the surface at locations where no original data exists, based on the known data values (observations).

2.2.1 Using Map Projections in Surfer 10

Map projection support is a very powerful new feature in Surfer 10. It allows you to combine data (data, grid and base map files) from different coordinate systems, and convert that data to a new coordinate system for the map to display. It saves you time and effort since you don't have to do this conversion separately in a different program before using the data in Surfer. For example, you could create a contour map from a grid file in State Plane coordinates, add a post map layer using data in lat/long coordinates, and display the map in any coordinate system you wish

**When Do I Use Coordinate Systems?**

You do not always need to set the coordinate system for the map layers or the map. You only need to specify the coordinate systems for the map layers and map if: A) You load data in one coordinate system (i.e. State Plane 1983 – South Carolina) and want to display the map in another coordinate system (i.e. Lat/Long). B)

You want to create map layers from data that are in multiple coordinate systems, (i.e. one SHP file in State Plane, another in UTM, or another in Lat/Long) and want them to all overlay correctly in the same map. C)You want to create a map, and export it to a file (such as a SHP or GeoTIFF) and retain the coordinate system information with the exported file.

Objectives:
- Determination of Head of water available at surface.
- Determination of ground water table.
- To prepare variogram of available data.
- To prepare contour map.
- To prepare 3d map of strata.
- For determination Surface modeling, contouring, landscape visualization.
  …….And by using this software, we get **variogram data, contour map & grid data** of available data

III. Methodology Adopted
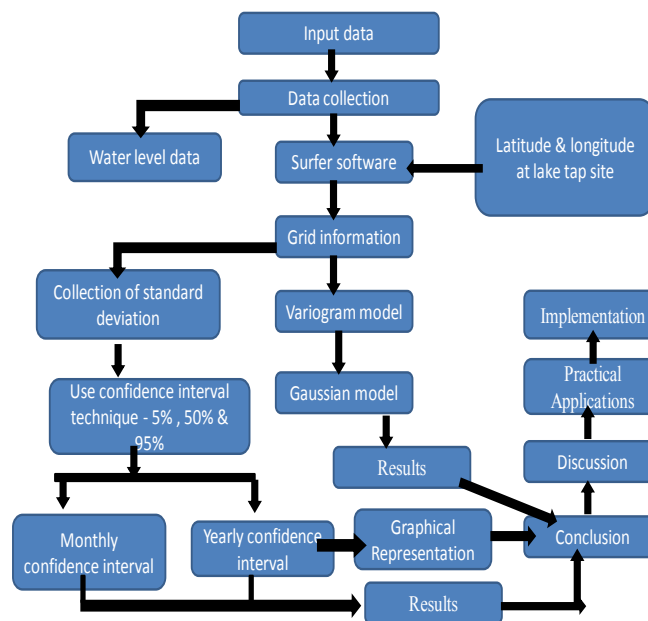
IV. Results

4.0 Gridding Report :-
Thu Jan 30 21:37:44 2014
Elapsed time for gridding: 0.02 seconds
Data Source
Source Data File Name: I:\2000\Jan\lat & long.xlsx (sheet 'Sheet1')
X Column: A
Y Column: B
Z Column: C
Univariate Statistics

_____

	X	Y	Z	
_____

Count:		31	31	31
1%%-tile:		17.446138	73.727682	32.443
5%%-tile:		17.446169	73.727682	32.519
10%%-tile:		17.446179	73.727725	32.672
25%%-tile:		17.446507	73.728036	33.205
50%%-tile:		17.447018	73.72869	34.12

75%%-tile:	17.447418	73.729216	34.882	
90%%-tile:	17.447796	73.729827	35.186	
95%%-tile:	17.447807	73.729827	35.186	
99%%-tile:	17.447827	73.729838	35.263	
Minimum:	17.446138	73.727682	32.443	
Maximum:	17.447858	73.729838	35.263	
Mean:	17.4470029032	73.7287199355	34.0237096774	
Median:	17.447018	73.72869	34.12	
Geometric Mean:	17.4470028933	73.728719932	34.0111470655	
Harmonic Mean:	17.4470028833	73.7287199285	33.9985339953	
Root Mean Square:	17.4470029132	73.728719939	34.0362159704	
Trim Mean (10%%):	17.4469738214	73.7286771429	33.9916428571	
Interquartile Mean:	17.4469786875	73.72864725	34.033875	
Midrange:	17.446998	73.72876	33.853	
Winsorized Mean:	17.4470015161	73.7287223548	34.033516129	
TriMean:	17.44699025	73.728658	34.08175	
Variance:	3.58926156989e-007	5.31130129031e-007		0.879549946237
Standard Deviation:	0.000599104462501	0.00072878675141		0.937843241825
Interquartile Range:	0.000910999999999	0.00117999999999		1.677
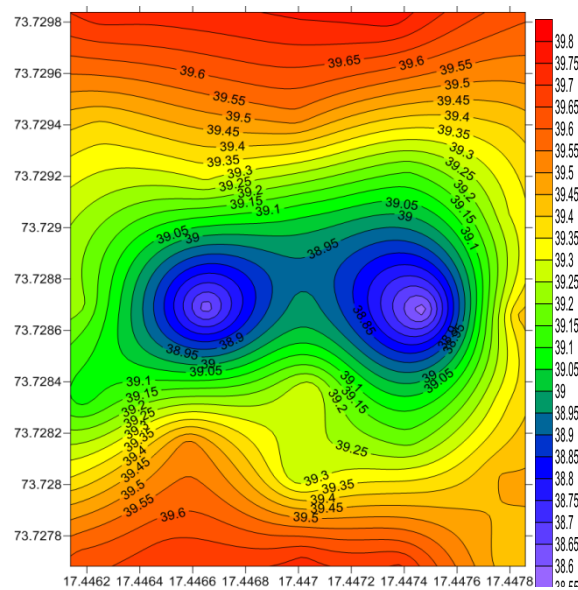
4.1 Contour maps :-

fig. 4.1.1 Contour map

Contour map represents variation in water head data. Contour maps are two-dimensional representations of three dimensional data. Contours define lines of equal Z values across the map extents. The shape of the surface is shown by the contour lines. Contour maps can display the contour lines and colours or patterns between the contour lines.

4.2 variogram
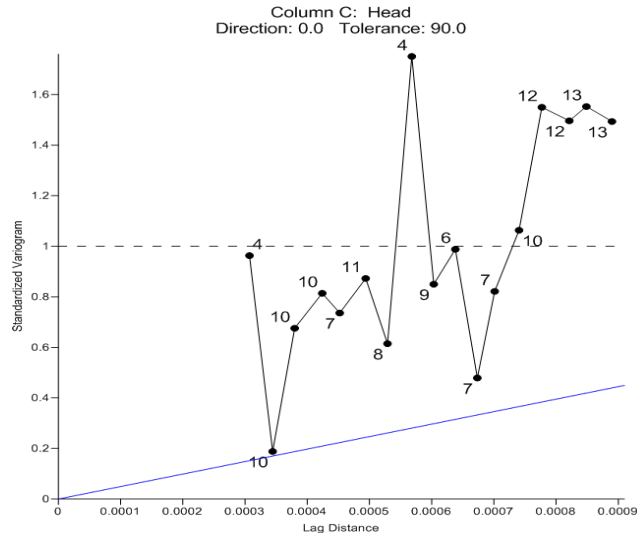
Fig. 4.2.1 variogram

   The variogram characterizes the spatial continuity or roughness of a data set. Ordinary one dimensional statistics for two data sets may be nearly identical, but the spatial continuity may be quite different. Refer to the Surfer User's Guide and the topic Variogram Model Graphics in the Surfer Help for graphs illustrating the curve shapes for each function. To account for geometric anisotropy (variable spatial continuity in different directions), separate experimental and model variograms can be calculated for different directions in the data Set.

4.3 Monthly Results:
4.3.1 Contour Map – Year 2000 (Monthly)
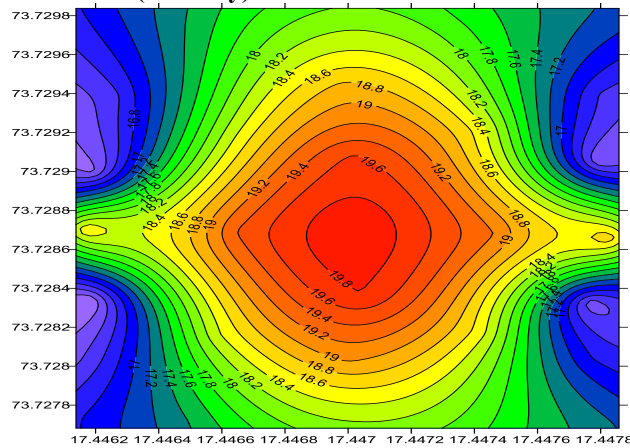
Fig. 4.3.1.1 contour map; May-2000
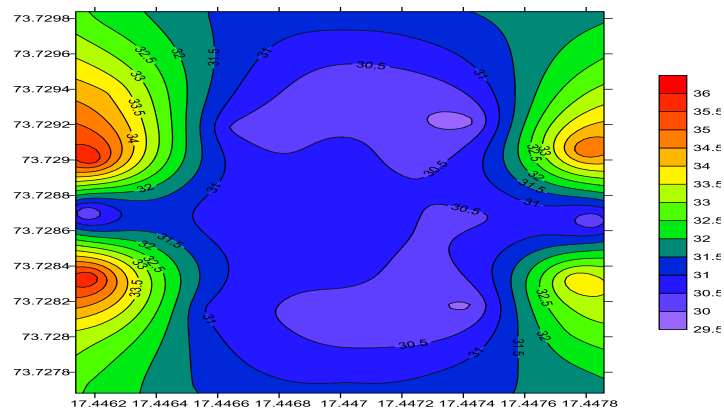
Fig. 4.3.1.2 contour map; Aug-2000

4.3.2 Variograms – Year 2000:

Column C: Head
Direction: 0.0   Tolerance: 90.0
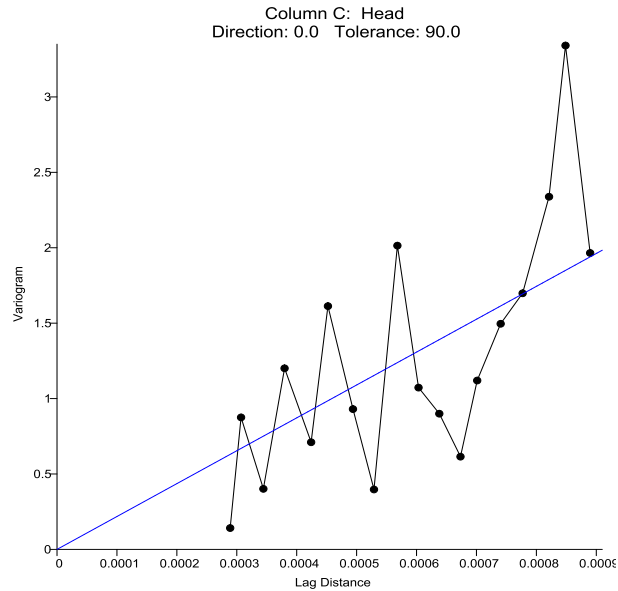
Fig. 4.3.2.1 Variogram; May-2000

Column C: Head
Direction: 0.0   Tolerance: 90.0
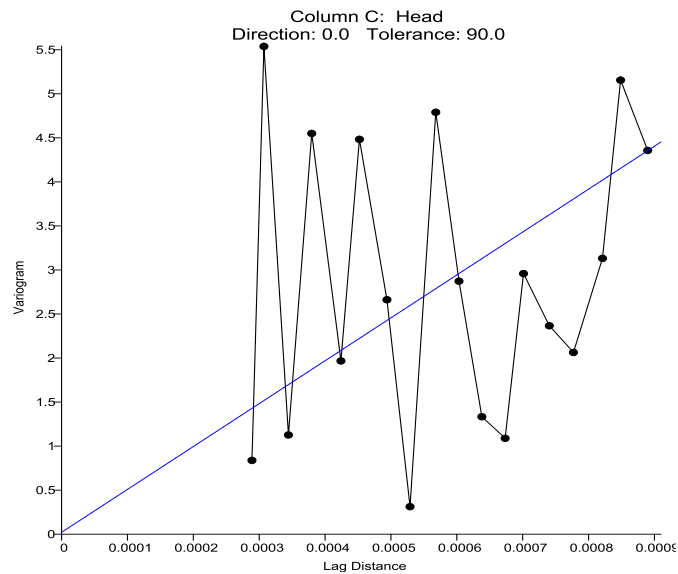
Fig. 4.3.2.2 Variogram; Aug-2000

4.3.3 Contour Map – Year 2009:
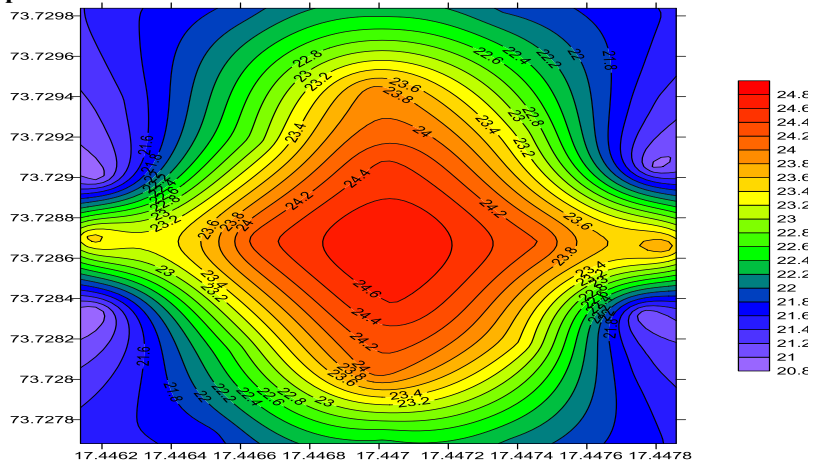
Fig. 4.3.3.1 contour map; may-2009
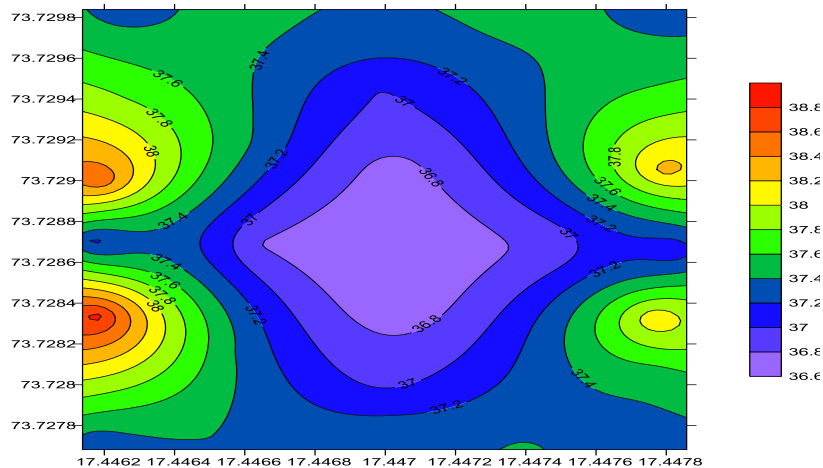
Fig. 4.3.3.2 contour map; Aug-2009

4.3.4 Variograms – Year 2009
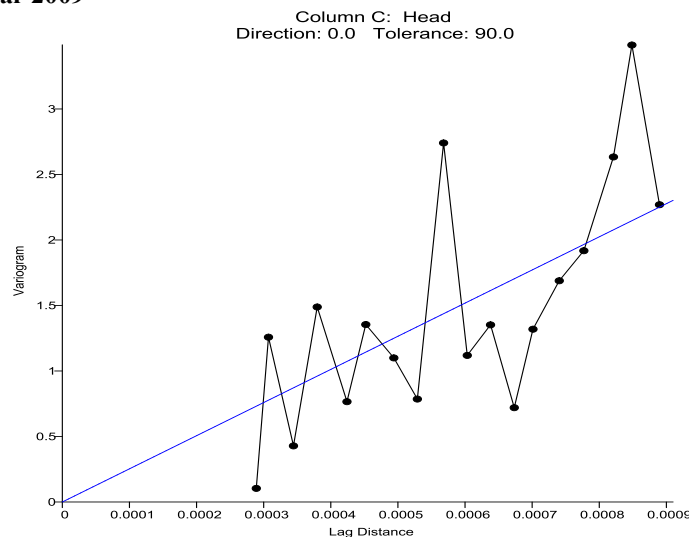
Fig. 4.3.4.1 Variogram; May-2009
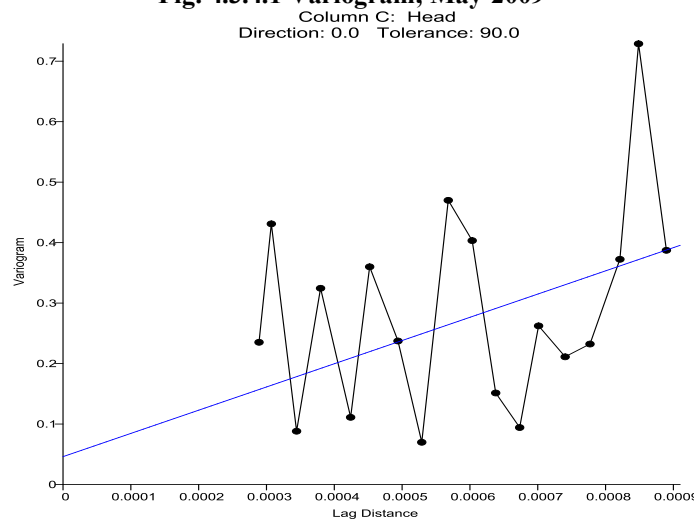

Fig. 4.3.4.2 Variogram; Aug-2009

From above graphs we can predict that variation of water head over site can be seen in such a way that in may -2000,2009, head at the centre is maximum and at the outside it is minimum. While on other hand aug-2000,2009 head at the centre is minimum and at the outside it is maximum. From such results we conclude that silting may takes place as well as variation of head over site as per seasonal variation. From data and graphical

results of contour maps and variograms we can use it for future forecasting and proper water management ,storage application for domestic use etc**.**

4.4.1 Yearly Representation of Confidence interval for the year 2000 and 2009:-
Maximum values are taken from that maximum values mathematical modeling is done to take out the results :-

Graphical representation of 5%,50%,95% confidence interval for the year 2000:
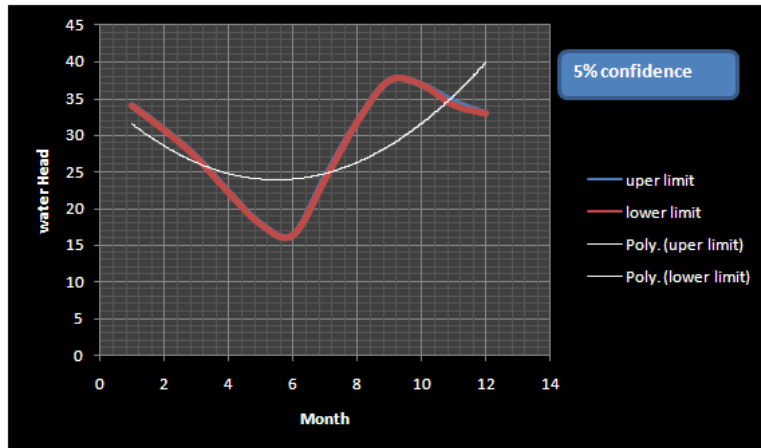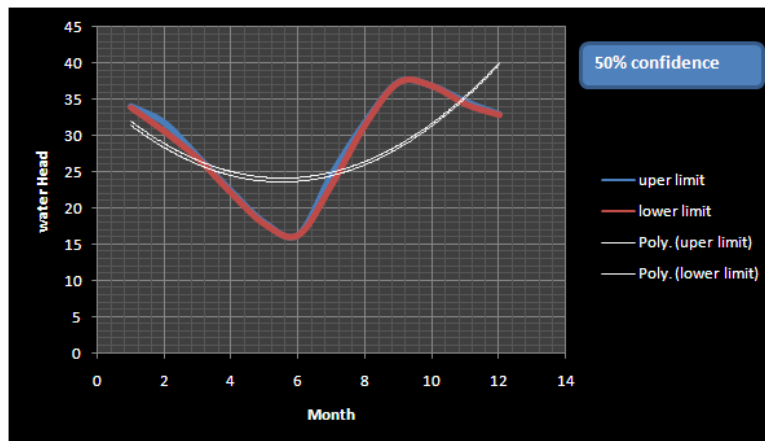

Fig. 4.4.1   5% confidence (Year 2000)

Fig. 4.4.1.2 50% confidence (Year 2000)
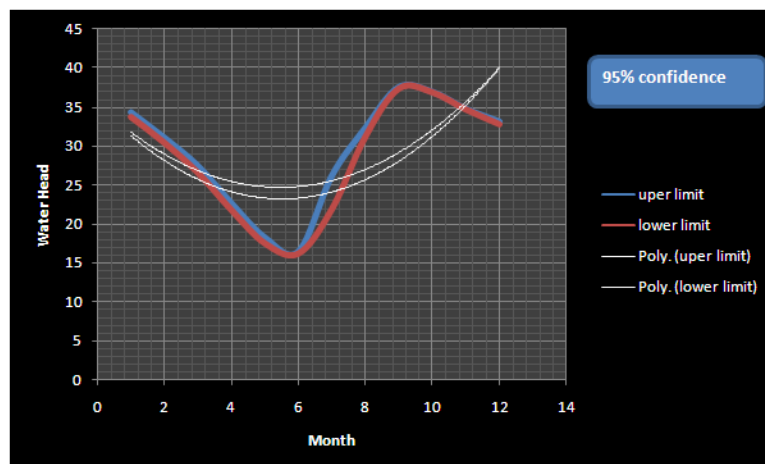
Fig. 4.4.3 95% confidence (Year 2000)

4.4.2 Graphical representation of 5%,50%,95% confidence interval for the year 2009:
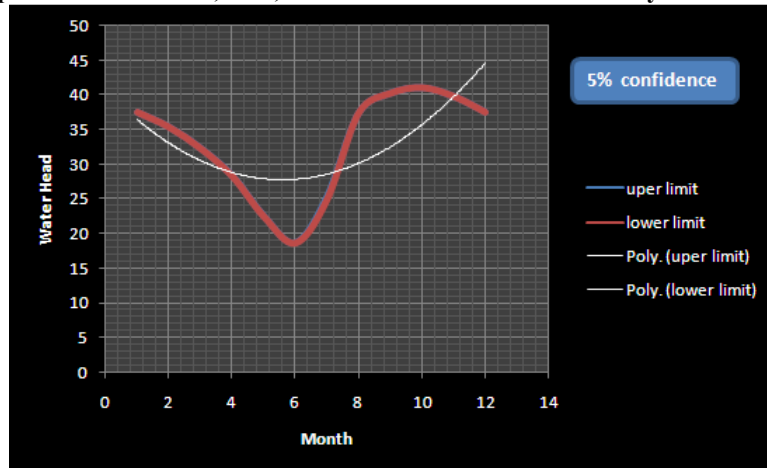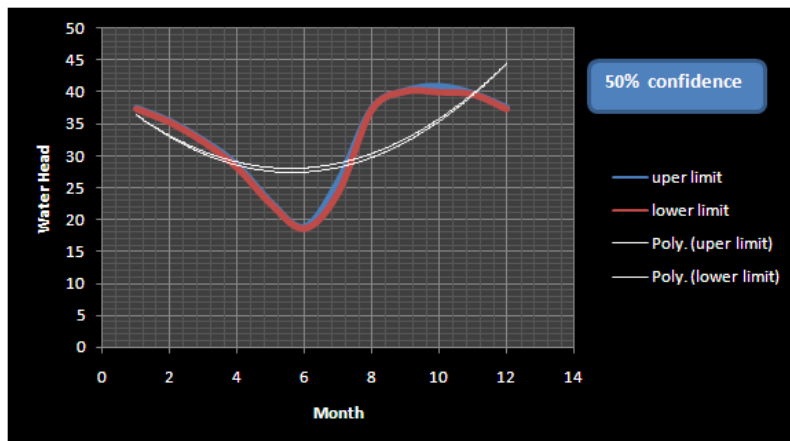

Fig. 4.4.2.1 5% confidence (Year 2009)

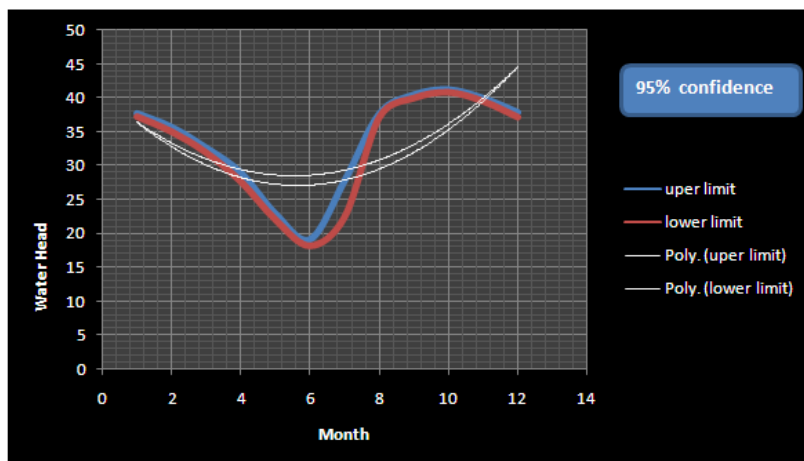Fig. 4.4..2.2 50% confidence (Year 2009)

Fig. 4.4..2.3 95% confidence (Year 2009)

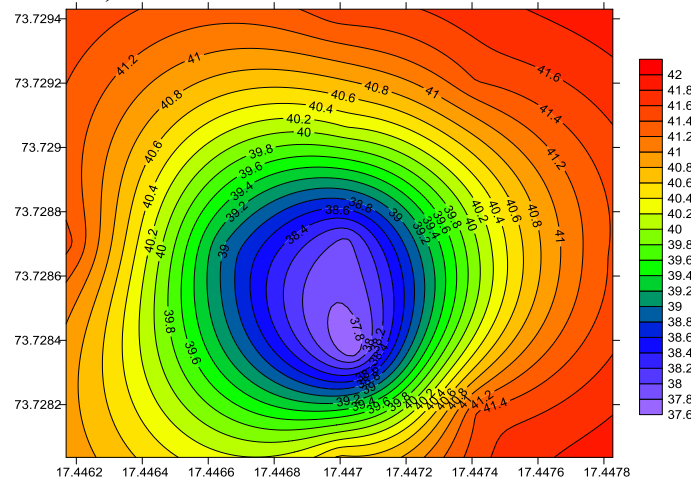**4.4 Yearly Results:**
4.4.1 Contour map (2000 - 2010)

Fig. 4.4.1.1 contour map of 2000-2010.

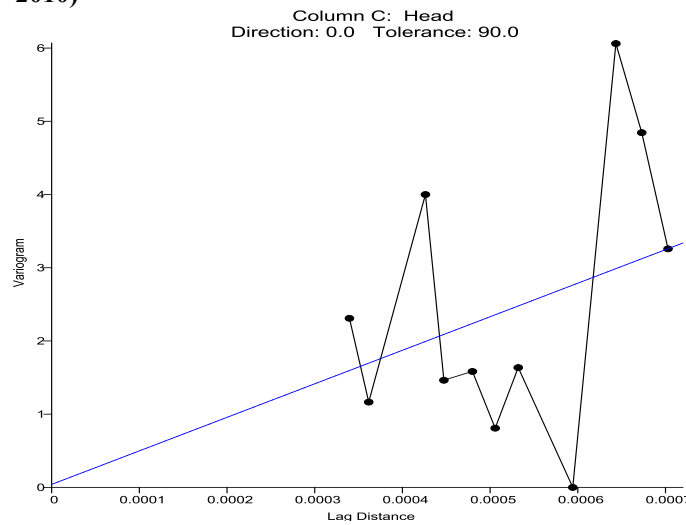**4.4.2 variogram (2000 - 2010)**

Fig. 4.4.1.2 Variogram; 2000-2010

Yearly Representation of Confidence interval for 2000-2010:-
Maximum values are taken from that maximum values mathematical modeling is done to take out the results :-
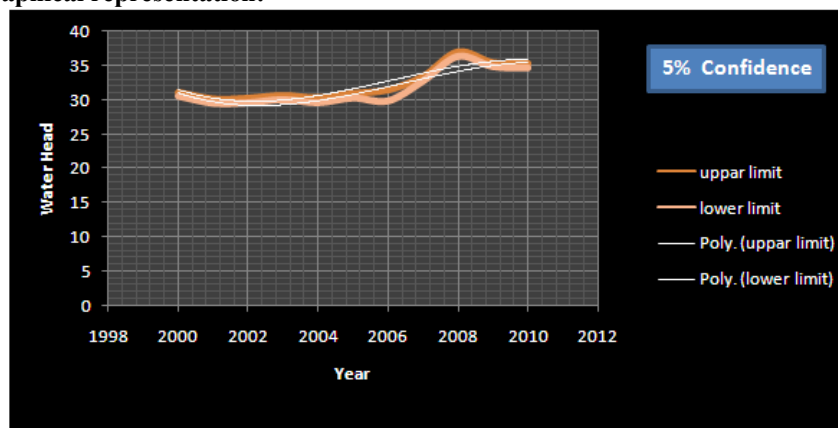
  4.4.3.1 Graphical representation:

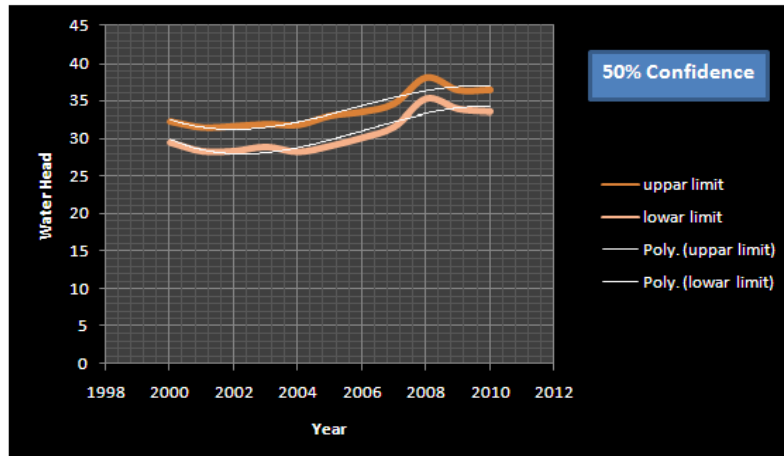**Fig. 4.4.3.1.1 5% confidence (Year 2000-2010)**

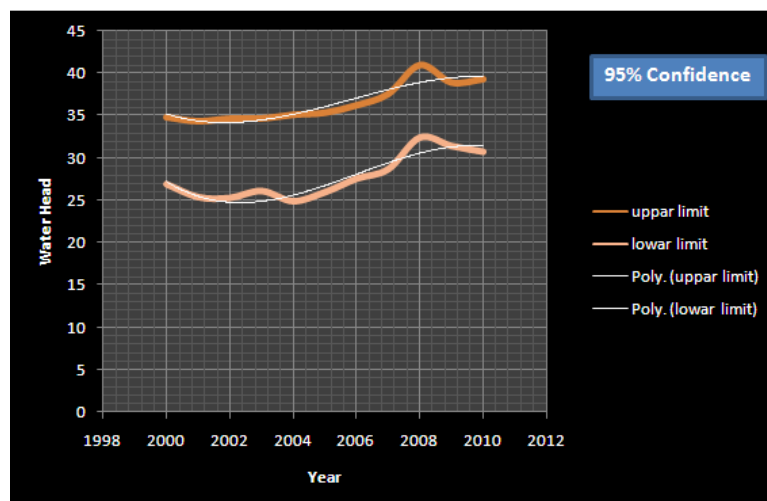**Fig. 4.4.3.1.2 50% confidence (Year 2000-2010)**

Fig. 4.4.3.1.3 95% confidence (Year 2000-2010)

Graphical Comparison of monthly variation of confidence interval of year 2000, 2009 studied as well as yearly comparison for max. head values can be studied from 2000-2010.

V.   conclusions

- Results showing the behavior of surface water available over the site according to which seasonal variation is determined.
- By using results uncertainty can be predicted
- Methodology is simple and it is mainly based on logical nature of probability theory (CLT)i.e. confidence interval
- Results provide the undulating nature of ground which gives silting and scouring nature of site.
- Three things impact the width of a confidence interval   Confidence level: This is the 95% part of the 95% confidence interval and also typically takes values of 90%, 99%, 80% and 85%. Confidence levels are the "advertised coverage" of a confidence interval. If we were to sample from the same user population 100 times, we'd expect the average to fall within the interval 95,90etc., timesoutof100.Variability:as measured by the standard deviation. Populations (and samples) with morevariabilitygeneratewiderconfidenceintervals. Sample Size: Smaller sample sizes generate wider intervals. There is an inverse square root relationship between confidence intervals and sample sizes. If you want to cut your margin of error in half, you need to approximately quadruple your sample size.
- For confidence intervals using task times you should perform a log transformation on the raw values,
- We are 95% confident in the method of computing confidence intervals, not in any given interval.
- You can use the overlap in confidence intervals as a quick way to check for statistical significance
- Confidence intervals tell you the most likely range of the unknown population average or percentage such as the average completion rate, the average level of satisfaction or the percentage of users likely

to recommend your product. Note: Confidence intervals don't tell you the likely range of all values, just how much the average value is likely to fluctuate.

- Confidence intervals provide both the location and precision of a measure: For example, the graphs below show the 95% confidence intervals around the average System Usability Scale (SUS) score
.

VI. Significances

1) Future forecasting of head of water at the location of lake tapping in Koyna Dam
2) Helps to design of lake tapping projects at other reservoir.
3) Prediction of Seasonal variation – Domestic, Industrial & Irrigation management.
4) CIs not providing only the reader an easy-to-interpret range estimate of the population mean, they also give information about the "precision" of an estimate.
5) A CI provides the necessary information needed when conducting a meta-analysis and most importantly, allows a researcher to immediately compare a current result with CIs from previous studies.
6) CIs can also be used to test hypotheses when necessary.
7) Location of spillway:-
a) Which type of spillway use in dam?
b) Maximum free board provision.
c) Design of vertical convex curve (Y = polynomial).
d) For regulatory works, Design of spillway gates.

VII. Future Work

1) By applying methodology, In future we can apply fuzzy systems in order to achieve the head variation which helps in research work.

2) In INDIA Lake Tapping projects are very less and in future need will be more so the methodology helps in lake tapping projects at another site.

3) Technique mostly emphasize on mathematical modeling which helps in future forecasting.

4) Water Scarcity and Drought Conditions can be handled by using confidence interval data by knowing the head variation.

5) Comparison of actual data with graphical data as well as mathematical modeling data can be done in order to check the Statistical Technique.

6) Also Spillway, Sluice gates, Opening gates can be design by using data.

7) From rainfall variation over the site and prediction of head we can interpret the storage and how much water should be needed for domestic, irrigation and industrial purpose.

8) Proper watershed management can be achieved.

References

[1]. Groundwater quality assessment of the shallow aquifers west of the Nile Delta (Egypt) using multivariate statistical andgeostatistical techniquesOriginalResearchArticl
[2]. JournalofAfricanEarthSciences, Volume95, July2014, Pages123-137Alaa A. Masoud
[3]. Influence of the Response Delay on the Assembling Technique Based on the Statistical Feed-forwardControlModel OriginalResearchArticleProcedi Engineering, Volume63, 2013, Pages 98-107C. Hernández, R. Tutsch
[4]. Performance loss rate of twelve photovoltaic technologies under field conditions using statistical techniques OriginalResearchArticle SolarEnergy, Volume103, May2014, Pages28-42George Makrides, Bastian Zinsser, Markus Schubert, George E. Georghiou
[5]. Elemental analysis of ancient potteries of Vellore Dist, Tamil Nadu, India by ED-XRF technique with statistical approach Beni-SuefUniversity Journal of Basic and Applied Sciences, In Press, Corrected Proof, Available online6March2014. A. Naseerutheen, A. Chandrasekaran, A. Rajalakshmi, R. Ravisankar
[6]. Confidence intervals for quantiles in finite populations with randomized nomination sampling
[7]. Mohammad Nourmohammadi, Mohammad Jafari Jozani ∗, Brad C. Johnson
[8]. Computing confidence intervals for log-concave densities
[9]. Mahdis Azadbakhsh, Hanna Jankowski ∗, Xin Gao
[10]. Department of Mathematics and Statistics, York University, Toronto, Canada
[11]. David, H.A., Nagaraja, H.N., 2003. Order Statistics, third ed. John Wiley and Sons.Deshpande, J.V., Frey, J., Ozturk, O., 2006. Nonparametric ranked-set sampling confidence intervals for quantiles of a finite population. Environ. Ecol. Stat. 13, 25–40.
[12]. Frey, J., 2007a. Distribution-free statistical intervals via ranked-set sampling. Canad. J. Statist. 35, 585–596.
[13]. Frey, J., 2007b. A note on a probability involving independent order statistics. J. Stat. Comput. Simul. 77, 969–975.
[14]. Gemayel, N.M., Stasny, E.A., Wolfe, D.A., 2010. Optimal ranked set sampling estimation based on medians from multiple set sizes. J. Nonparametr. Stat. 22,517–527.
[15]. Hettmansperger, T.P., Sheather, S.J., 1986. Confidence intervals based on interpolated order statistics. Statist. Probab. Lett. 4, 45–79.
[16]. Jafari Jozani, M., Johnson, B.C., 2012. Randomized nomination sampling for finite populations. J. Statist. Plann. Inference 142, 2103–2115.
[17]. Jafari Jozani, M., Mirkamali, S.J., 2010. Improved attribute acceptance sampling plans based on maxima nomination sampling. J. Statist. Plann. Inference 140, 2448–2460.

[18]. Jafari Jozani, M., Mirkamali, S.J., 2011. Control charts for attributes with maxima nominated samples. J. Statist. Plann. Inference 141, 2386–2398.
[19]. Kim, S.H., 1993. Stochastic comparisons of order statistics. J. Korean Stat. Soc. 22, 13–25.
[20]. Kvam, P.H., Samaniego, F.J., 1993a. On estimating distribution functions using nomination samples. J. Amer. Statist. Assoc. 88, 1317–1322.
[21]. Kvam, P.H., Samaniego, F.J., 1993b. On the inadmissibility of empirical averages as estimators in ranked set sampling. J. Statist. Plann. Inference 36, 39–55.
[22]. Nourmohammadi, M., Jafari Jozani, M., Johnson, B.C., 2013. Distribution-free tolerance intervals with randomized nominated samples (submitted).
[23]. Nyblom, J., 1992. Note on interpolated order statistics. Statist. Probab. Lett. 14, 129–131.
[24]. Ozturk, O., 2012. Quantile inference based on partially rank-ordered set samples. J. Statist. Plann. Inference 147, 2116–2127.
[25]. Higgins M, Keller J. Familial occurrence of chronic respiratory disease and familial resemblance in ventilatory capacity. J Chronic Dis 1975;28:
[26]. 239–51.
[27]. Shoukri MM, Ward RH. Use of linear models to estimate genetic parameters and measures of familial resemblance in man. J R Stat Soc
[28]. Ser C 1989;3:467–79.
[29]. Bartko JJ. The intraclass correlation reliability coefficients. Psychol Bull1966;83:762–5.
[30]. Shoukri MM, Ward RH. Estimation of intraclass correlation. Commun Stat Theory Methods 1985;13(10):1239–55.
[31]. Donner A. A review of inference procedures for the intraclass correlation in the one-way random effects model. Int Stat Rev 1986;54(1):67–82.
[32]. Efron B, Tibshirani RJ. An introduction to bootstrap. Chapman and Hall;1993.
[33]. Shao J, Tu D. The jackknife and bootstrap. NY: Springer; 1996. Raudenbush SW, Bryk SA. Hierarchical linear model: applications and data analysis methods. 2nd ed. London, UK: Sage Publications; 2002.
[34]. Donner A,Koval JJ. The large sample variance of intraclass correlation.Biometrika 1981;67:719–22.
[35]. WWW.google.com.
[36]. Science Direct, Springer etc.
[37]. Irrigation Department, Government of Maharashtra.INDIA.