Groundwater Data Analysis

M.Tech Thesis Stage I Report

Submitted in partial fulfillment of the requirements for the degree of

Master of Technology

by

Lalit Kumar
Roll No:10305073

Ravi Sagar P
Roll No:10305037

under the guidance of

Prof. Milind Sohoni

Department of Computer Science and Engineering
Indian Institute of Technology Bombay
October 2011

Abstract

In this project work we did groundwater data analysis of Thane district in Maharashtra with the help of Ground Water Survey and development Agency(GSDA), Pune. We have analyzed groundwater data of 120 observation wells over the period of 35 years and developed yearly seasonal models. Our goal is to develop a regional seasonal model for Thane district that will help in better understanding of Thane ground water system and helps GSDA in developing the water budgets smoothly.

Acknowledgement

We express our deep gratitude and regard for our guide **Prof. Milind Sohoni**, for guiding us through this project work. We thank him for answering our queries and clearing our doubts with utmost patience. It was due to his guidance and support the project could move ahead in the correct direction. His encouraging words always motivated us towards the work. We also thank **Prof. Om Damani** and **Prof. Purushottam Kulkarni** for their valuable suggestions and inputs. We would also thank Rahul Gokhale for supporting us in this project work.

Lalit Kumar
(10305073)
Ravi Sagar P
(10305037)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

When rainfall occurs some of the water flows along the surface to water bodies such as lakes and streams, this water is called run-off, some part is evaporated in the atmosphere, and the remaining sinks into the ground which fills in the spaces between the particles of the sand and the fractured rocks. Below a certain depth in the ground all the spaces in soil particles are filled with water. [5]This zone is called the saturated zone and the depth at which it starts underneath the surface is known as water table. The water below the water table is known as groundwater. A scene depicting the same is shown in figure 1.1. The groundwater moves at a very slow rate below the ground and joins the stream. Groundwater is available nearly everywhere, though the quantity available varies depending on the geologic materials, rainfall and climate.

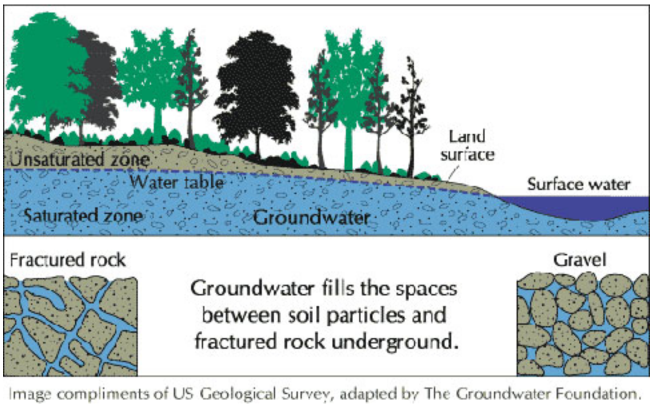

Figure 1.1: Groundwater

1.1 Importance of Groundwater

Groundwater contribution to world's fresh water supply is about 20% , which is 0.61% of entire world's water, including frozen ice and oceans. It is one of the most important natural resources which is often not recognized. Some of the important points for groundwater are listed below:

- Groundwater is the major source of drinking water in both urban and rural India.

- It is an important source of water for the agricultural and the industrial sector.
- The flow in the streams and the water in lake is sustained by the discharge of groundwater into them.
- Groundwater constitutes the underground part of the natural water cycle.
- In most cases the groundwater is cleaner then surface water as it is protected against contamination from the surface by soils and covering rock layers.

1.2 Motivation

Growth in population, urbanization and standard of living has resulted in increased demand of water for diverse purposes of irrigation, domestic and industrial uses. On the other hand resources of accessible water are decreasing, due to overuse and pollution. There is a huge gap in the demand and supply of water. In order to meet this increased demand surface water is not sufficient alone, so groundwater is being used. Wells are the main medium through which groundwater is taken out from the ground. During rainy season the wells are enriched with water, but these wells are over extracted either for domestic use or commercial purpose as a result wells go dry within a few months. This creates a problem for drinking water in non-monsoon period and the dependency on other sources such as water tankers increases. With continuous extraction water table drops and the rate of recharge of groundwater decreases. To properly utilize groundwater one needs to understand its spatial aspects and variation through time in groundwater behavior. This can be achieved by organizing spatial data on maps and through building mathematical models. These mathematical models would allow us to understand groundwater in many ways. The complexity of models could vary, we could have a complex model which predicts effect of various hydrological and geological changes on groundwater, whereas on the other hand a simple model may just predict groundwater variation depending on period of the year.

In this work we have studied the observation well data of Thane district for 35 years. The complete framework of project work is given in appendix-A. Chapter 2 defines the problem statement and what are our objectives in the future. Chapter 3 describes cleaning process of the observation well data(i.e ground water level in observation wells) on which we have carried out the work. It also provides detailed discrepancy analysis. Chapter 4 describes the single well seasonal models with different interpolation techniques and we also discuss the proof of correctness of our model. In Chapter 5 we discusses trend lines, their importance and GSDA approach for developing trend lines. In Chapter 6 we discuss need for spatial models, spatial correlation between rain fall data, voronoi diagrams and krigging interpolation technique. Finally in Chapter 7 we conclude and discuss the future work to be done in the project.

1.3 Some Terminology

- **GSDA-:**Groundwater Surveys and Development Agency (GSDA) is a Government of Maharashtra organization which deals with groundwater exploration, monitoring,development, and management.

- **Observation Wells-:**These are dedicated monitoring wells, selected to detect potential changes in groundwater flow and quality.

- **Watershed-:**A watershed is an area of land enclosed within mountain ridges from which water drains to a particular point along a stream. An image of watershed is shown in figure 1.2

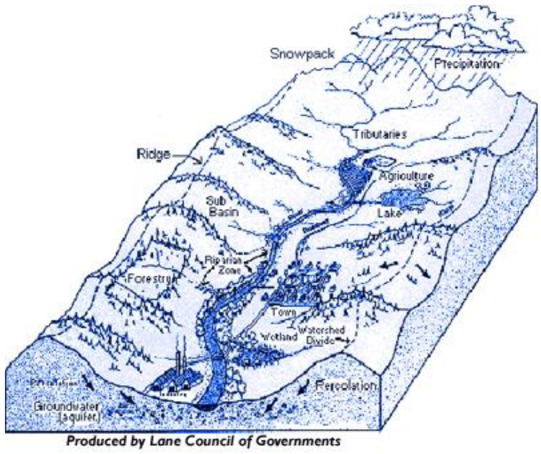

Figure 1.2: Watershed

- **Aquifer-:**It is that portion of layer underneath the ground which allows storage and movement of water through it.

Chapter 2

Problem Statement

GSDA has been recording groundwater data for all districts of Maharashtra for over 30 years. This repository of data must be used for assessing groundwater potential and usage, and for planning drinking water schemes. Techniques such as data mining, mathematical skills along with GIS representation techniques has to be used to develop better understanding of seasonality and water table, connections and relationships between different regions vis-a-vis groundwater. The objectives of the project is to study the groundwater behavior for the Thane district. This can be used in the planning for drinking water schemes in the district. Later on the work can be extended to other districts. Initially only observation well data is being used for the analysis purpose.

2.1 Building Yearly Seasonal Models

To understand the behavior of the groundwater level variation throughout the year at a observation well, we intend to make mathematical models for each well, it the yearly seasonal model. These models will help in predicting the groundwater level at specific locations. Initially these models will be build at the level of observation well and later extended to regional and district level. For building the initial models only observation well data of 25-35 years will be used. In subsequent models other factors such as rainfall, elevation and soil porosity, will also be considered to make these mathematical models so that one can have a better understanding of the groundwater behavior.

2.2 Understanding Water Budget

GSDA prepares a water budget for each district every year. This documents contains information about the requirement of water in the district and how this requirement is to be fulfilled. We need to understand this water budget and the method for its creation. We need to known the following aspects-:

- What all data is taken as input for the preparation of the water budget.
- What role does observation well data plays in the preparation of the water budget.
- What are the outputs of the water budget and how they are used for meeting water requirements .

- Who actually uses the output of the water budget. This will be done in the future stages of project.

2.3 Feedback for 163 Tanker dependent Wadis

There are 163 wadis in Thane district identified by GSDA which have severe scarcity of water in non-monsoon periods. These wadis are mainly located in the Jawahar, Mokhada, Shahapur and Vikramgad tahsils. GSDA has employed various schemes such as piped water system in 19 of these wadis. But previous schemes have not been successful in solving the water crisis situation. Now these wadis are being supplied water through tankers. We aim to understand the reasons for such situation in these wadis by studying the groundwater behavior and provide solutions to GSDA for solving the water crisis.

Chapter 3

Properties of Data

We had received observation well data from GSDA, Maharashtra for Thane district for the period 1975-2010(March) in an excel sheet format. The data showed depth of water in 120 observation wells measured at different times in a year.

3.1 Spatial Distribution

Total geographic area of Thane is about 9,558 km^2. A map of Thane is shown in figure 3.1. There are a total of 120 observation well i.e. around 1 observation well every 79.6 km^2. The observation wells are spread out in 13 tahsil covering 115 villages. The observation wells are chosen according to the watershed. The 120 observation wells are contained in 34 watersheds in Thane district. These 120 observation wells were plotted on Google Earth as shown in fig 3.1 All the 34 watersheds are westernly flowing watersheds. Typically a watershed is divided into three parts depending on the drainage of the water and usually each part has one observation well. First part is the run-off area in which around 90% of water runs off to the water bodies, this is the hilly area where water runs along the slope. The second part is the recharge part, in this 20-30% of rainfall will be captured and will percolate into the ground and the remaining will be run-off. The last part is the storage part in which 50-60% of water will percolate into the ground. But since in Thane the type of rock formation is hard rock therefore almost no watershed has the storage part. The following points have to considered while choosing a well as an observation well.

- The well should be a non-pumping well.
- The depth of the observation well is to be considered.
- The usage of pattern (if any) of well has to be looked into.

3.2 Data Description

The key facts about the data is given below:

- The data is for 120 observation wells, 92 of these are dug well and remaining 28 are bore wells.
- A total of 11682 observations were there in the received data.

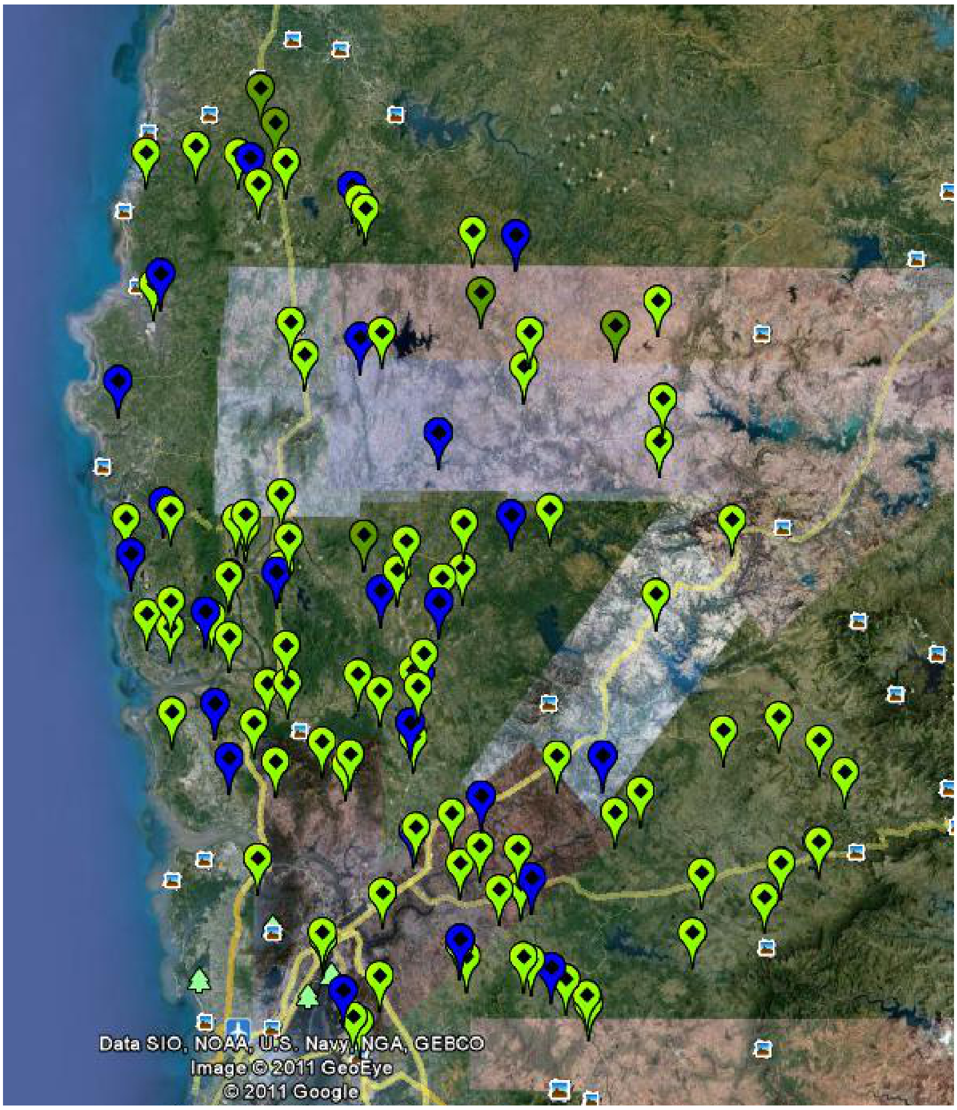

Figure 3.1: Observation wells plotted on Google Earth
Green balloons: Dug wells
Blue balloons: Bore wells

- The data for some dug well is available from as early as 1975 till 2010
- The data for bore well is available from year 2000 to 2010.
- The observation in a dug well is taken 4 times year mostly in JAN,MAR,MAY,OCT
- The observation in a bore well is taken more or less throughout the year.

The received data contained the following fields:

1. **District-:** This field contains the name of the district to which the observation well belonged. Since the data was for only Thane district all the rows had the value Thane

2. **Tahsil-:** This field contains the name of the tahsil to which the observation well belongs.

3. **Watershed-:** This field indicates in which watershed does the observation well lie.

4. **Village-:** This field contains the name of the village to which the observation well belongs.

5. **Site_Type-:** The value in this field indicates whether the observation well is a dug well or bore well.

6. **Depth-:** In this field the depth of the observation well is indicated.

7. **Wls_date-:** This field has the value of date on which the observation was taken.

8. **Wls-:** This field indicates the depth from the ground level at which water is found in the observation well.

9. **Site_id-:** This is a unique id assigned to each observation well. It was concatenation of latitude and longitude of the observation well.

A sample of 5 rows of data is shown in table 3.1

Table 3.1: 5 sample rows from data

Tahsil	Village	Watershed	Site_Type	wls_date	wls	Depth	Site_id
BHIWANDI	Akoli	WF-27	Dug Well	1991-01-31	3.5	5.5	W192915073024001
BHIWANDI	Akoli	WF-27	Dug Well	1991-03-31	3.7	5.5	W192915073024001
DAHANU	Kalamdevi	WF-09	Dug Well	1988-08-0	0.4	5.5	W200446073004501
DAHANU	Kalamdevi	WF-09	Dug Well	1988-09-30	0.5	5.5	W200446073004501
BHIWANDI	Padgha	WF-31	Bore Well	2002-10-08	3.46	30	W192143073103001

3.3 Implicit Errors and Added Field

In the initial analysis certain very obvious errors were seen in the data. Following are the types of errors found-:

- **Duplicate Entries-:** There were two entries of water level for a observation well on the same date. Most of them also showed the same water level. From the two entries the second one was retained and first one was deleted. A total of 366 entries were deleted by this approach.

- **Negative Depth-:** There were 12 entries which showed negative depth of water. These entries were also deleted from the data.

- **Water Level greater than Depth-:** Two entries in the data showed that depth of water is greater than depth of the well.

The value of field site_id for a observation well was a string consisting of latitude and longitude value of that observation well. Two columns(latitude,longitude) were added to the data set to explicitly store the latitude and longitude values. With these latitude and longitude values we tried to locate exact location of observation wells in Google Earth. One sample snapshot of Satiwali dug well is shown in fig 3.2

Figure 3.2: Observation wells marked on Google Earth

3.4 Flagging Errors

Before using the data for making mathematical models or doing some sort of analysis , it was necessary to identify the errors in the data. The following two types of discrepancies were flagged in the data set.

1. **Gaps in Reading-:**The observations are supposed to be taken while maintaining the intervals between them as decided by GSDA. There were certain observation which were found violating these constraints. We had flagged these observations. Readings which were taken than 210 days apart were marked.

2. **More Increase/Decrease-:**A normal trend for observation well is that water level should increase(depth decrease) in monsoon period and decrease (depth increase) in non-monsoon period. We decided to flag all those readings which were not in accordance with it. We assumed the monsoon period starts from June 01 and ends at October 31. Now if an observation in non-monsoon shows decrease in depth of water as compared to their preceding observation respectively then these observations were flagged. e.g. consider the observations shown in table 3.2 In the above two observations the second observation is in the month of May so the depth of water indicated by this observation should be more than the preceding observation. But instead the depth of water indicated by the observation is less than the preceding observation, therefore it is flagged. Similarly if

Table 3.2: Flagging example in non-monsoon

Village	Site_Type	Wls_Date	Wls	Depth	Flag
Khodala	Dug Well	2000-04-06	5	5.8	0
Khodala	Dug Well	2000-05-25	2.6	5.8	1

any observation showed decrease in depth in monsoon period as compared to preceding observation then that observation was also flagged. Sample of such flagging is shown in table 3.3

Table 3.3: Flagging example in monsoon

Village	Site_Type	Wls_Date	Wls	Depth	Flag
Saravali	Bore Well	2006-07-24	1	24	0
Saravali	Bore Well	2006-08-22	2.1	24	1

A total of 1230 observations out of 11302 observations were flagged.Out of these 697 were in monsoon period and remaining 533 are in non-monsoon period.

The above approach of flagging discrepancy was found to be not so good. The approach had the following problems:

- An observation was being compared to its preceding observation irrespective of the gap between the two readings.
- The hard deadline for the start and end of monsoon was not the correct approach. Suppose we had a observation on June,01 then according to our assumption this observation is in monsoon, so it should have less depth(more water) as compared to preceding observation. But being on June,01 it is not necessary that rain would have happened on that day.

After this task one more set of observations were deleted. These were those observations which showed water at depth 0(well is full) in non-monsoon period. There were 36 such observations. For the discrepancy analysis it was not correct to mark readings across the monsoon period because there was no information about the rain. So we decided to flag the observations only in non-monsoon(Nov-May) period. Let a observation be denoted as O_i and the observation immediately preceding O_i in time is called O_{i-1}. Any observation O_i in the period November to May should show increase in water depth(as no rains) as compared to its preceding observation O_{i-1} provided the O_{i-1} is not before October. A total of 471 observations which violated this were flagged. The observations flagged by this new approach were checked against the rainfall data. For this purpose rainfall data for the period 1986-2007 was procured from GISE lab, IIT Bombay. The rainfall data was taken for points(called as rain gauge) shown with red balloons in the figure 3.3. Discrepancies only till year 2007(373 out of 471) was checked against rainfall, as rainfall data was available only till year 2007. The total rainfall as per the raingauge nearest to observation well for days between observation O_i and

Figure 3.3: Rainfall points in red marks

O_{i-1} was calculated. After analysis we found that out of 373 observation marked for increase in water, 102 had non-zero rainfall whereas there was no rainfall for other 271 observation. If we neglect changes smaller than 0.2m then out of these 271 observation, 189 showed increased level of water by 0.2m or more. The observation well which occurs maximum number of time in these 189 observation are Satiwali_Bore_Well(10), Ghol_Dug_Well(9), Vasar_Bore_Well(8). Plots showing the groundwater depth over the years for these wells is shown below in fig 3.4, fig 3.5 and fig 3.6. The circles on the graph indicates the discrepant readings. These wells should be looked into as they show increase in water level quite a few times even when there is no rain. A table showing top ten increases in water level in non-monsoon is shown in table 3.4.

There were observation in data which showed huge variation in few days. To flag these errors slope was used. The rate of change of groundwater depth per day was plotted between two observations. Now the observation with huge variation in few days had very high value and were the outliers in slope values. The graph for two villages indicating slopes is shown in fig 3.7 and 3.8 The peak in the figure 3.7 is an outlier whereas the graph in figure 3.8 has no such observation. The values due to which their is an peak or high slope value in table 3.5 are shown below-:

To detect such outliers in the slope values the interquartile range was used. If Q_1 and Q_3 are the lower and upper quartiles then any observation outside the range [Q_1-k(Q_3-Q_1),Q_3+k(Q_3-Q_1)] was decided as an outlier where Q_1 is the lowest 25% values of the data and Q_3 is the highest 25% values of the data and k is the constant. In our case the k was chosen to be 3 as with k=3 all the extreme outliers were rejected. The observation for which outlier was detected were flagged. These flagged entries were later discarded for making mathematical models using splines, which would be explained in Chapter 4.

Table 3.4: Top 10 increase in water level in non-monsoon

Village _Site_type	well depth(m)	gauge (km)	prev _date	prev _depth(m)	date	depth	diff (m)	rain (mm)
Kudan Bore_Well	30	42.492	2002-05-03	13.68	2002-05-22	3.75	-9.93	0
Kajali Dug_Well	14	40.377	2003-10-08	7	2004-01-31	1.4	-5.6	1
Talasari Dug_Well	8	40.672	1988-05-03	8	1988-05-30	2.6	-5.4	0
Awale Dug_Well	7.35	38.117	1988-01-30	7.35	1988-02-03	2.1	-5.25	0
Talasarimal Dug_Well	8.2	49.393	2001-04-06	8	2001-05-30	3.1	-4.9	0
Zhari Dug_Well	7.4	39.333	1997-10-31	6.4	1998-01-31	2	-4.4	69
Mahim Bore_Well	20	25.001	2002-04-02	9.55	2002-04-03	5.18	-4.37	0
Safale Bore_Well	25.9	28.896	2000-10-03	10	2000-12-13	5.69	-4.31	25
Palghar_kolgaon Bore_Well	30	31.751	2007-12-18	7.15	2007-12-28	3.2	-3.95	0
Veyour Dug_Well	10.1	31.586	1988-04-22	7.7	1988-05-30	4.15	-3.55	0

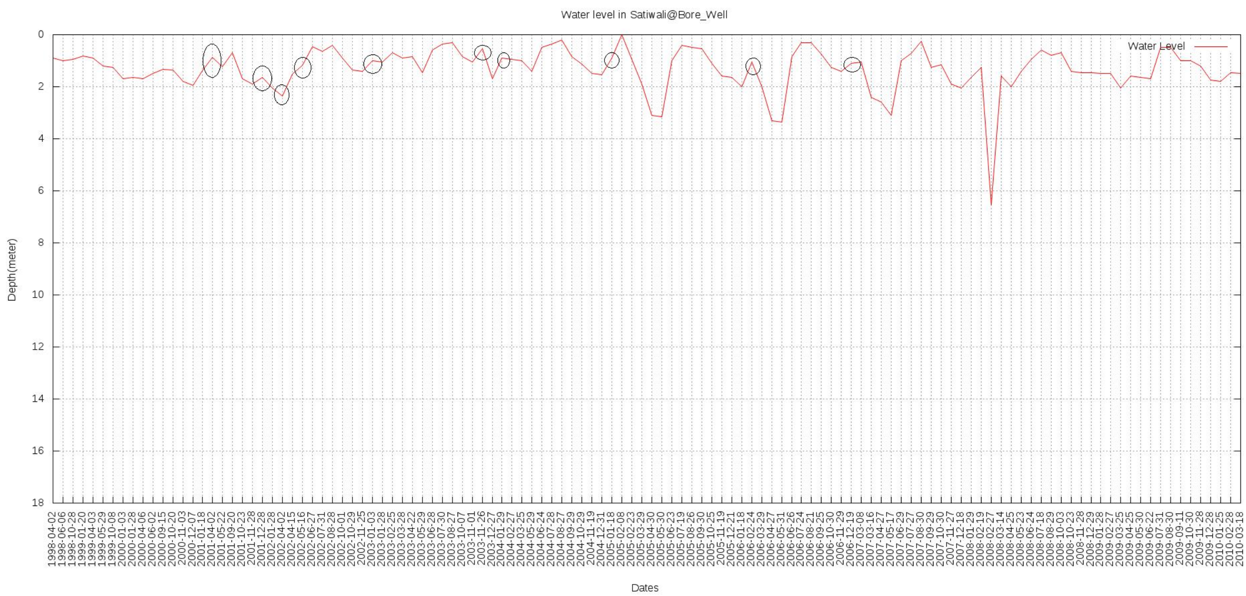

Figure 3.4: Satiwali Bore Well with marked discrepancy

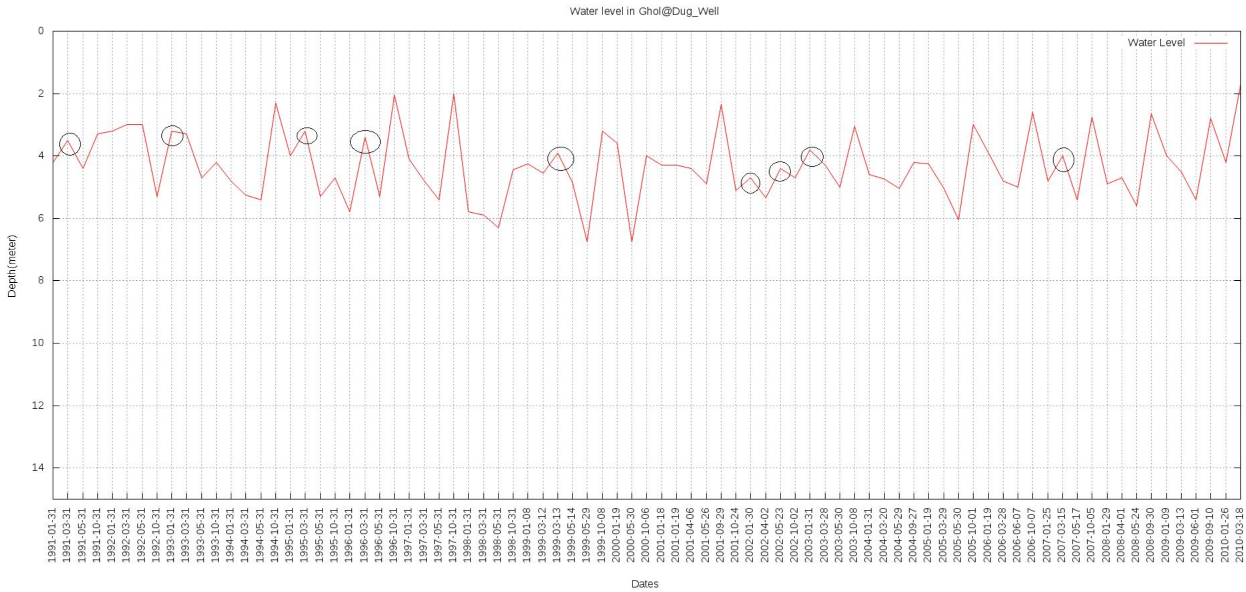

Figure 3.5: Ghol Dug Well with marked discrepancy

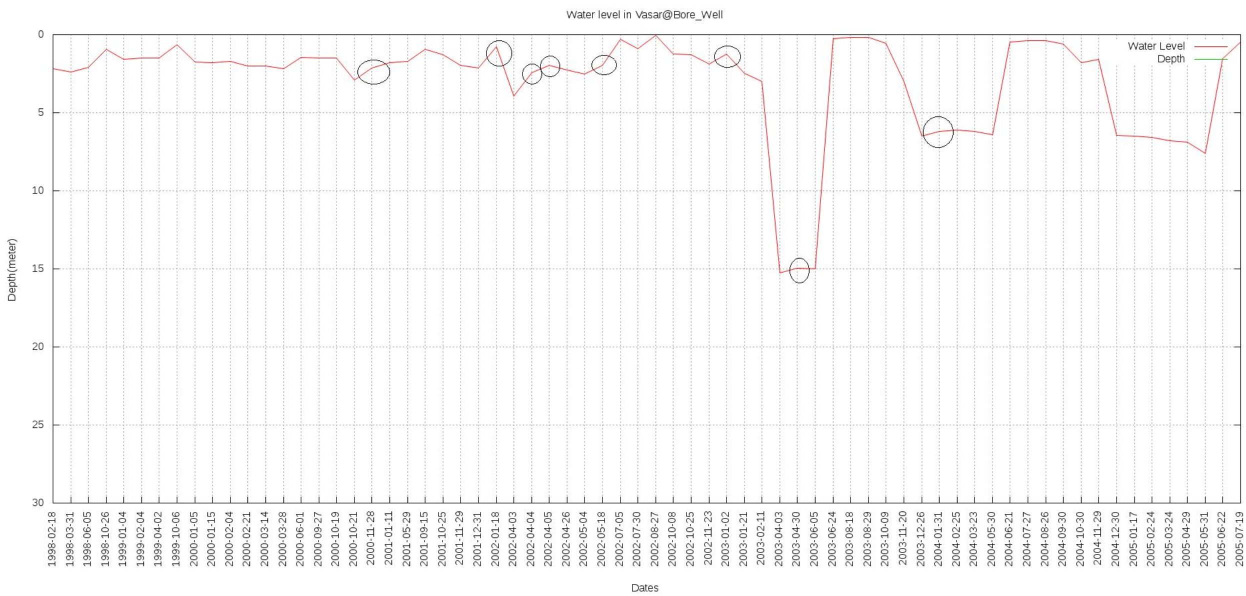

Figure 3.6: Vasar Bore Well with marked discrepancy

Table 3.5: Sample value showing high slope

village	site_type	depth	wls_date	wls
Agashi_Boling	Dug Well	10	2001-01-17	3.1
Agashi_Boling	Dug Well	10	2001-04-02	4.3
Agashi_Boling	Dug Well	10	2001-05-16	4.75
Agashi_Boling	Dug Well	10	2001-09-28	0.9
Agashi_Boling	Dug Well	10	2001-10-04	5
Agashi_Boling	Dug Well	10	2001-10-23	2.2

Figure 3.7: Slope values for a observation with outlier

Figure 3.8: Slope values for a observation well without outlier

Chapter 4

The Single Well Seasonal Models

4.1 Seasonal Model Objective

Seasonal model is a mathematical model that shows the seasonal behavior of the ground water level in a particular well over the period. The model is developed using the sample data of ground water for last 30 years. The main objective of this model is to predict the water level of a particular well at particular day in a year using the available samples. Using the sample data over 30 years of 120 observation wells we have developed the seasonal model of each individual observation well.

4.2 The Mathematical Model and The Need for Interpolation

In order to develop seasonal model of a particular well we first taken all the available observations(or sample data) for that well. These observations are in the form of depth of water level and date on which reading is taken. We first converted the date in to the day of the year by considering Jan 1^{st} is the 1^{st} day and Dec 31^{st} as 365^{th} day of the year. For example a reading on 1993-05-31 is converted in to reading on a 151 day of the year. Now all these readings will have a day value varying from 1 to 365. Thus we have scaled down the 30 years of sampling data window in to a single year sampling data window.

Our aim of developing the seasonal model is to predict water level trend over the period, we can do this by fitting a curve to the data that best suits to the given data. While fitting a data set we first choose the model to fit. Wrong choice of model will affect the result of analysis. Since the ground water is obviously seasonal we decided to build periodic model. The model space was thus a linear space of periodic functions.

4.2.1 Linear Curve Fitting:

[3], [4] Consider a data set $(x_1, y_1), (x_2, y_2), ...(x_n, y_n)$. Let us fit a function $y = F(x)$ to the given set using linear fit method. Where $F(x) = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_n f_n(x)$.

Now the linear fit model tries to find the values of the constants $a_1, a_2 \cdots a_n$ by minimizing the square of the error(i.e difference between the actual data value y_i and the value obtained by fitting the function $F(x_i)$. This is known as Least square error minimization).

$$\begin{aligned} (y-F(x)^2 &= \textstyle\sum_{i=1}^{n}(y_i - F(x_i))^2 \\ err(a,y) &= \textstyle\sum_{i=1}^{n}(y_i - F(x_i))^2 ----------->Eq(1) \end{aligned}$$

To get the values of constants $a_1, a_2, \cdots, a_n$.

$$\frac{\partial(err)}{\partial(a_r)} = 0 \text{ where r = 1,2,...n.}$$

That is equal to $\sum_{i=1}^{n}((y_i - a_1f_1(x_i) - a_2f_2(x_i)\cdots - a_nf_n(x_i)).f_r(x_i)) = 0$

If we expand the above equation we will get

$$\begin{bmatrix} \sum_i(f_1(x_i))^2 & \sum_i f_1(x_i)f_2(x_i) & \cdots & \sum_i f_1(x_i)f_n(x_i) \\ \sum_i f_2(x_i)f_1(x_i) & \sum_i(f_2(x_i))^2 & \cdots & \sum_i f_2(x_i)f_n(x_i) \\ \vdots & \vdots & \cdots & \vdots \\ \sum_i f_n(x_i)f_1(x_i) & \sum_i f_n(x_i)f_2(x_i) & \cdots & \sum_i(f_n(x_i))^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_i y_if_1(x_i) \\ \sum_i y_if_2(x_i) \\ \vdots \\ \sum_i y_if_n(x_i) \end{bmatrix}$$

By solving the above matrix equation we will get the constants a_1, a_2 and a_3 which minimizes the error in fitting the above function $y = F(x)$. We can apply this linear curve fitting to our sample data by considering day of measurement as x parameter and depth of water level as y parameter. Since the ground water data shows periodic behavior(Depth of water level increases in monsoon period and decreases in non monsoon period) a periodic function would fit the data properly. Hence we have chosen the periodic function $F(y) = a_1\, sinx + a_2\, cosx + a_3$ to fit the data. Here we are using the trigonometric functions we have converted range of x from 1 to 365 to 0 to 2π. Our final expression is

$$\begin{bmatrix} \sum_i sin(x_i)^2 & \sum_i sin(x_i)cos(x_i) & \sum_i sin(x_i) \\ \sum_i cos(x_i)sin(x_i) & \sum_i cos(x_i)^2 & \sum_i cos(x_i) \\ \sum_i sin(x_i) & \sum_i cos(x_i) & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \sum_i y_isin(x_i) \\ \sum_i y_icos(x_i) \\ \sum_i y_i \end{bmatrix}$$

by solving the above equation we get the constants a_1, a_2 and a_3. Thus by substituting the constant values in the function $f(y) = a_1\, sinx + a_2\, cosx + a_3$ we derived the seasonal model of every observation well. We also computed the square error of the function using Eq(1).

The figure 4.1 shows the seasonal model of a Chahade dug well. Instead of using a single function we have used the four periodic functions to know which function best suits to the periodicity of sample data. The four functions are

$k = 1 \quad f(y) = a_1sinx + a_2cosx + a_3$
$k = 2 \quad f(y) = a_1sinx + a_2cosx + a_3sin2x + a_4cos2x + a_5$
$k = 3 \quad f(y) = a_1sinx + a_2cosx + a_3sin2x + a_4cos2x + a_5sin3x + a_6cos3x + a_7$
$k = 4 \quad f(y) = a_1sinx + a_2cosx + a_3sin2x + a_4cos2x + a_5sin3x + a_6cos3x + a_7sin4x + a_8cos4x + a_9$

We calculated the error value for each of the above function and compared. We found that k=2 or 3 will best represent the given sample data with least error value. The models are showing bad behavior in between. The reason is large gap between sampling points and high density at some points. This is because lot of our sample data is collected in Jan, March, May/June and Oct, very few readings are taken in remaining months. The clusters in the

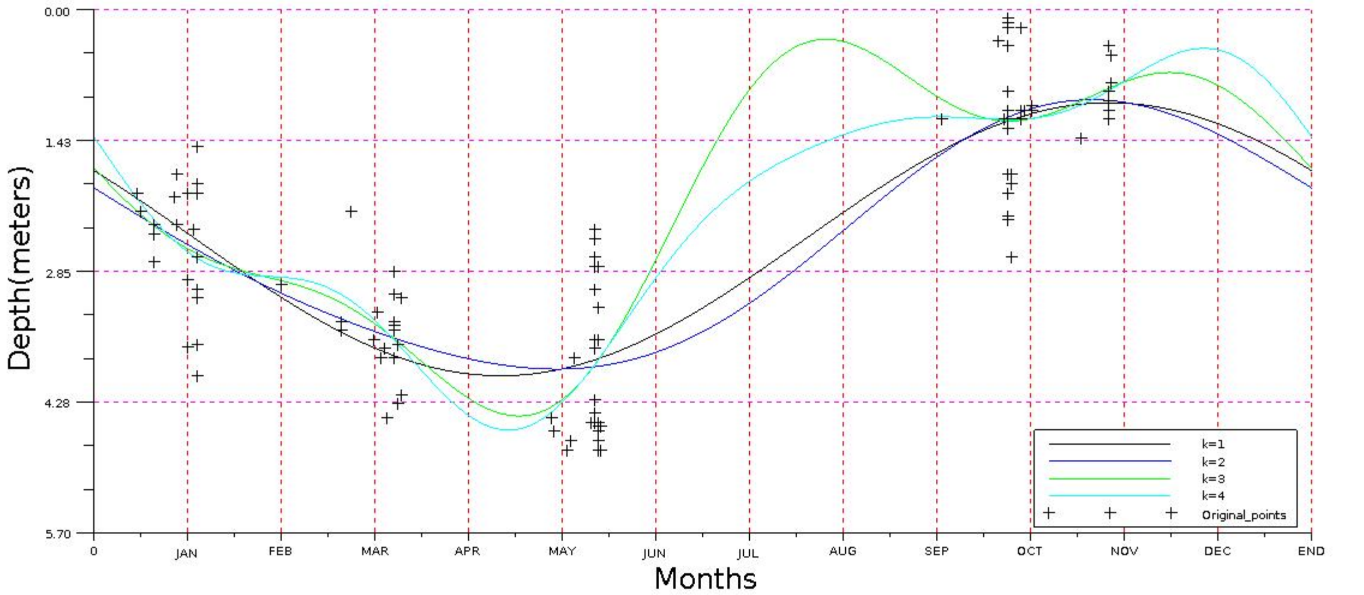

Figure 4.1: Seasonal model of Chahade dug well

graphs indicates this. Where there is no adequate data the model behaved unpredictably. To avoid this problem we need more sample data in less dense area i.e we need sample data with less gaps. This problem can be solved by interpolation.

4.3 Interpolation and Results

[12] Interpolation is process of estimating the intermediate values of an independent random variable given some sample values of that random variable. The depth of water level is also resembles the property of a random variable we can use interpolation to estimate intermediate values. Our sample data has 4 or 5 readings per year with 3 months gap which is not sufficient. We decided to interpolate the sample data with 15 days interval. For this purpose we have chosen linear interpolation which is one of the simplest interpolation method.

4.3.1 Linear Interpolation:

[12], [6] Linear interpolation takes two data points,(x_a, y_a) and (x_b, y_b), and gives the any interpolation point between them as,

$$y = y_a + (y_b - y_a)\frac{(x-x_a)}{(x_b-x_a)} \text{ at point (x,y)}$$

That is it takes two successive sample points and joins them using a straight line and gives the values on that line as interpolation points.

By using this linear interpolation we have generated large sample data where the gap between two samples is 15 days and used this data to generate the seasonal model as discussed in the above section. figure 4.2 shows the linear interpolated values of Chahade dug well with

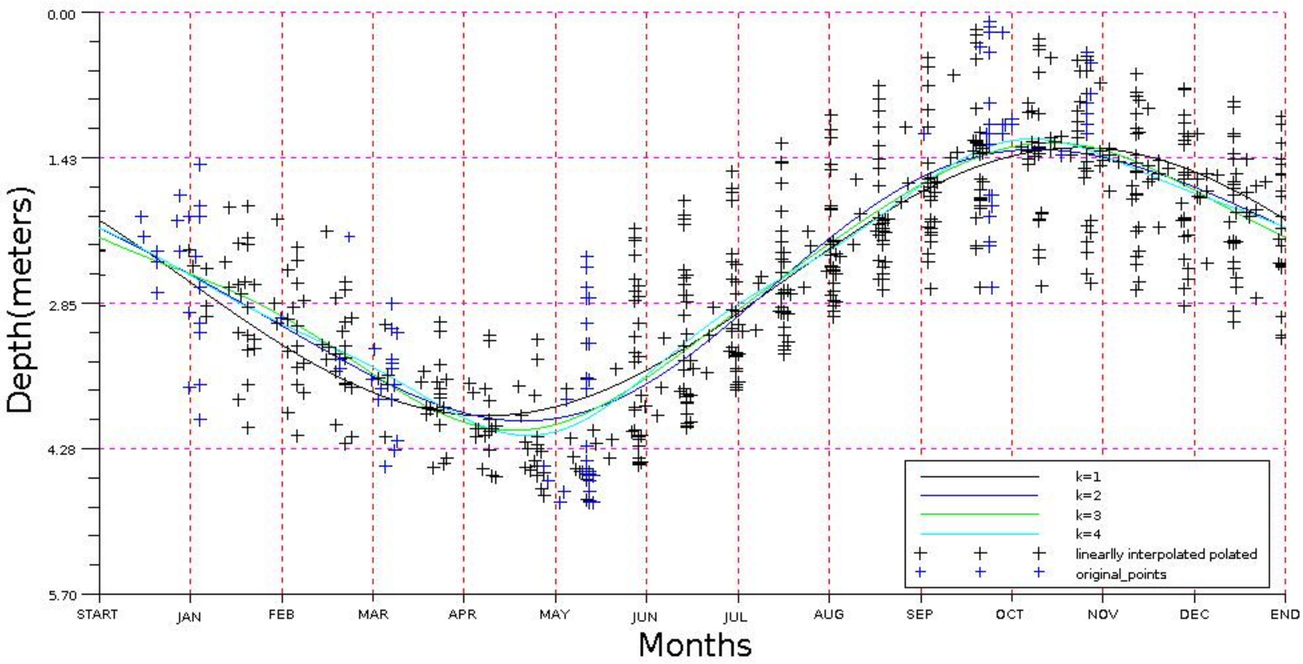

Figure 4.2: Linear interpolated values of Chahade dug well

15 days interval and the seasonal model that is generated using this data. After rigorous examination on above generated graphs we found that linear interpolation method does not suits to our sample data. To understand this let us consider that we have a sample data on Aug 10^{th} as 4m, and another reading on Dec 4^{th} is also 4m. This does not mean that depth of water level remained constant during that period. The depth of water level might have decreased till the end of monsoon(i.e end of September) and started increasing because of usage and would have reached to 4m again. When we interpolate the points between these two readings using linear interpolation it gives the points on the line that connects these two readings which is a straight line. As we discussed above the water level may not follows straight line thus we are estimating the wrong values between those two points. So we need a better interpolation method that gives us proper values.

4.3.2 Spline Interpolation:

[7], [8] Linear interpolation uses a linear function(straight line) for each of intervals $[x_k, x_k+1]$. Where spline interpolation uses low-degree polynomials in each of the intervals, and chooses the polynomial pieces between successive intervals such that they fit smoothly together. Finally the resulting function is called a spline. The basic idea is

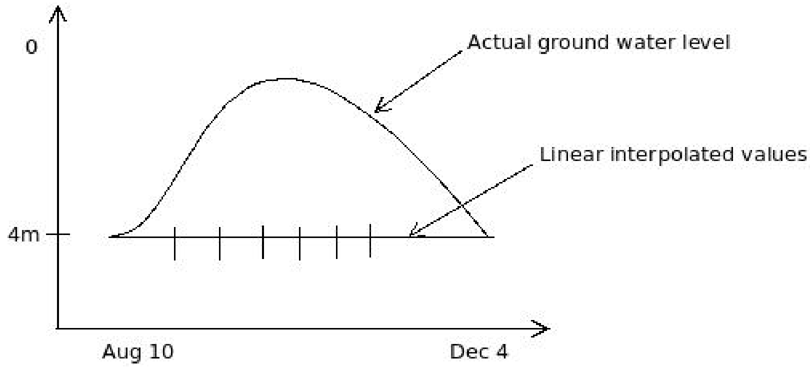

Figure 4.3: Drawback of linear interpolation

Table 4.1: Value of Errors

Interpolation Used	K=1	K=2	K=3	K=4
No Interpolation	7.176594	7.090292	6.996131	6.984524
Linear Interpolation	15.916068	15.629721	15.547885	15.479055
Spline Interpolation	20.994736	20.665765	20.658505	20.655860

Let $s(x) = a(i) + b(i) * x + c(i) * x^2 + d(i) * x^3$ for $x(i-1) \leq x \leq x(i)$, i = 1,..,n
This gives 4*n unknown coefficients. So, we need 4*n constraints to solve the problem. We require

$$\begin{array}{lrl} (i) & s(x(i)) & = y(i) \\ (ii) & s(x(i)+0) & = s(x(i)-0) \\ (iii) & s'(x(i)+0) & = s'(x(i)-0) \\ (iv) & s''(x(i)+0) & = s''(x(i)-0) \end{array}$$

for i = 1,...,n-1

Thus we get n+1 constraints from equation (i) and 3(n-1) constraints from remaining equations. Now we have 4n-2 constraints leaving 2 degrees of freedom for choosing the coefficients. If we make the second derivative of the 1st node and the last node zero then it is called "natural"cubic spline. If we make $s'''(x)$ to be continuous at x(1) and x(n-1) then it is called the "not-a-knot "condition.

Thus spline function will best interpolates our sample data. We have used "not-a-knot"condition to interpolate the sample data. Before computing the interpolated values we first eliminated the discrepancies(mentioned in chapter 3) from the sample data and then we applied cubic spline interpolation. Finally we build our seasonal model using this data. The final graph of the model is shown in figure 4.4. The following table shows the computed error values of original data(i.e with out interpolation), interpolated(Linear ans Spline) data with different functions(k=1,2,3,4. See section 4.2.1).

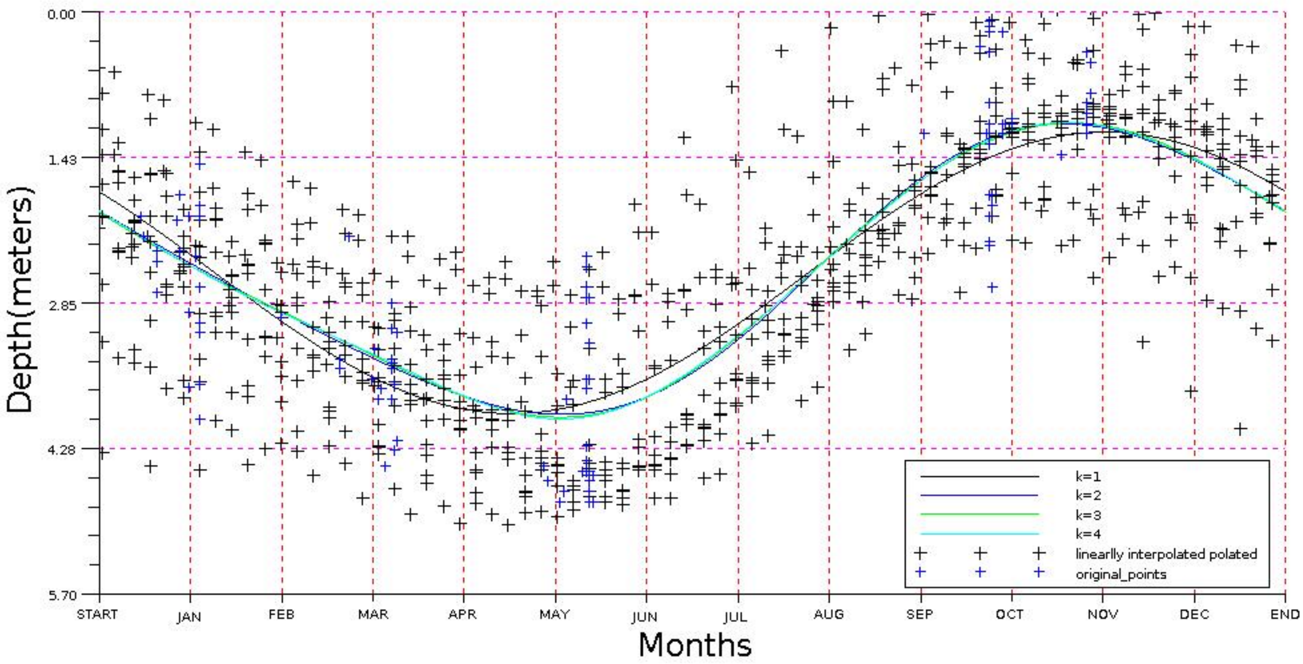

Figure 4.4: Seasonal model of Chahade dug well with spline interpolation.

4.4 The Normal Model and Maximum Likelihood Estimator(MLE)

It is clear that the seasonal models are giving us the average water level at particular period of time. We did this using linear curve fitting model. How far this is correct. We should know the correctness of our proposed model. We will argue for the correctness of our model in following section.

4.4.1 Maximum Likelihood Estimator(MLE):

[11] The MLE is used for estimating an unknown parameter of a probability distribution. Let us made some assumptions.

- Our sample data i.e depth of water level as a normal random variable Y with mean μ and variance σ .
- Sample data that is observation values as $y_1, y_2, ...y_n$ (We have assumed the depth of water level as y in our previous discussion we will continue the representation).

The density function of Y is f(y) = $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}$ Now we try to estimate the mean(or average) of random variable Y using the maximum likelihood estimator. For this purpose we assume that variance σ is constant. According to MLE in order to estimate the μ we first maximize the f(y).

$$\begin{aligned}\prod_{i=1}^{n} f(y_i) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_i-\mu}{\sigma})^2} \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_1-\mu}{\sigma})^2} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_2-\mu}{\sigma})^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_n-\mu}{\sigma})^2} \\ &= (\frac{1}{\sigma\sqrt{2\pi}})^n \times e^{-\frac{1}{2}(\frac{y_1-\mu}{\sigma})^2} \times e^{-\frac{1}{2}(\frac{y_2-\mu}{\sigma})^2} \times \ldots. \times e^{-\frac{1}{2}(\frac{y_n-\mu}{\sigma})^2} \\ &= (\frac{1}{\sigma\sqrt{2\pi}})^n \times e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n}(y_i-\mu))^2}\end{aligned}$$

Apply log function on both sides.

$$\begin{aligned}\log\left(\prod_{i=1}^{n} f(y_i)\right) &= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i-\mu)\right)^2 \\ &= -n\log\sigma - n\log\sqrt{2\pi} - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i-\mu)\right)^2 - - - - - - - - - > Eq(2)\end{aligned}$$

in order to find μ value that maximizes we differentiate the above equation with μ. Finally we get

$$\mu = \sum_{i=1}^{n}(y_i-\mu))^2$$

The above equation is similar to Eq(1) in section 4.1. where $\mu = a_1\, sinx + a_2\, cosx + a_3$. That is we can estimate mean μ of a random variable that maximizes the value of f(y) by using the least square error minimization method. This is exactly what we did in our linear curve fitting model.

4.5 Seasonal Models With Dry Readings

While taking the sample readings if the well is dry the depth of the well is taken as the reading. That is when the well is dry we are considering that water exists at that depth which is wrong. When the well is dry the ground water level may go beyond the depth of the well. Instead of finding the exact value ,we are simply taking the depth of the well as reading. This indicates that any reading that is more than the depth of the well is normalized to its depth.This is known as one sided error.

Our current model does not handle this problem. So we need a model that deals with dry readings of a well. Prof. Milind Sohoni and his Ph.D student Rahul.B.Gokhale have designed a model that deals with dry readings of a well. We will implement this model in our future work.

4.6 Computation Of Variance and Its Significance

In general variance is the measure of how far the sample data is spread out from sample mean. The variance of depth of water level gives us how much is the water level in a observation well varied over the period. Large value of variance indicates that water level has huge variations which makes that observation well suspect. That is the well might be in usage or there may be a river(or canal) near to the observation well. So the readings from these wells are not reliable. We have calculated the variance of all the observation wells and listed out the top few to report to GSDA.

4.6.1 Calculating Variance using MLE

To calculate the value of variance we again use the Maximum Likelihood Estimator method. We can calculate it by differentiating the Eq(2) in section 4.4 with σ. Then we get.

$$\sigma \quad = \{ \frac{\sum_{i=1}^{n} (y_i - \mu)^2}{n} \}^{\frac{1}{2}}$$

Here we are calculating the σ value with an average μ which gives an incorrect value. Because when we folded 30 years of data in to single year there are lot of sample data at each observation date with different values. So we first computed the μ_i at each observation date. Now we calculated the variance using

$$\sigma \quad = \{ \frac{\sum_{i=1}^{n} (y_i - \mu_i)^2}{n} \}^{\frac{1}{2}}$$

Where μ_i is calculated using $\mu_i = asin(x_i) + bcos(x_i) + c$. Finally we normalize the variance w.r.t depth, by dividing it by the depth of well. We call this as normalized variance. The top 3 wells having maximum variance are listed below in table 4.2. The complete list for all the villages is given in appendix C. Plots for each of these three wells showing their water level over the years are shown in fig 4.5, fig 4.6 and fig 4.7 below:

Table 4.2: Top 3 Wells with high variance

Well_name	Normalized Variance	Depth
Washind_1_Dug_Well	0.361308	7.000000
Talasari_Dug_Well	0.286490	8.000000
Satiwali_Dug_Well	0.244649	7.200000

4.7 Dug-well vs. Bore-well

In this section we will discuss the correlation between dug well and bore well. The reason why we did is to know how much spatial correlation does exist between observation wells. In general the depth of a bore well is higher than the dug well, so there exists water in bore well even the near by dug well is dry.If the water level in a dug well shows high correlation with near by bore well, when that dug well is dry we can use the water level of of near by high correlated bore well to estimate water level in that dug well.

As basic step we first computed the correlation between dug well and bore well that are located in same village. We have computed the correlation between water levels of dug well and bore well in the same month, i.e we compared the water level of dug well in particular month with the water level in bore well of the same month. The 4.3 table shows the results.

The results are fairly acceptable. We will extend this to the all dug wells in our future work. That is for all dug wells we are interested in finding correlation between a dug well and its near by bore well. So that we can divide the observation area(Thane district) in to regions where each region contains a single bore well and multiple dug wells that have high correlation with that bore well. It is similar to voronoi regions with bore wells as voronoi sites. We discuss about voronoi diagrams in section 6.2.

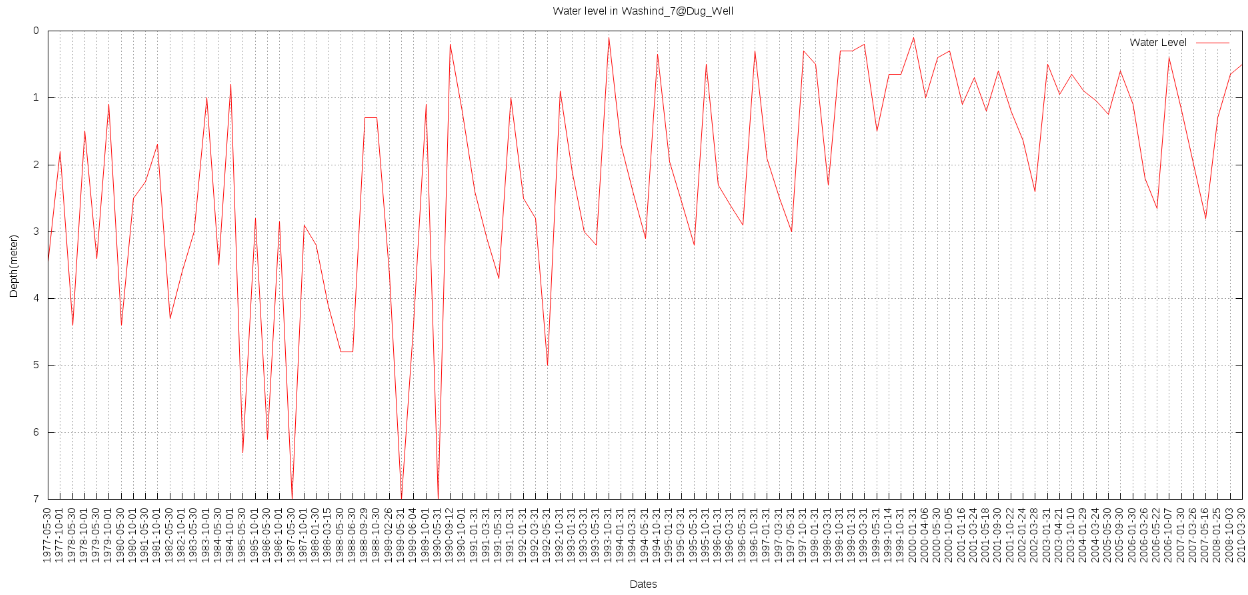

Figure 4.5: Water level in Washind_1_Dug_Well over the years.

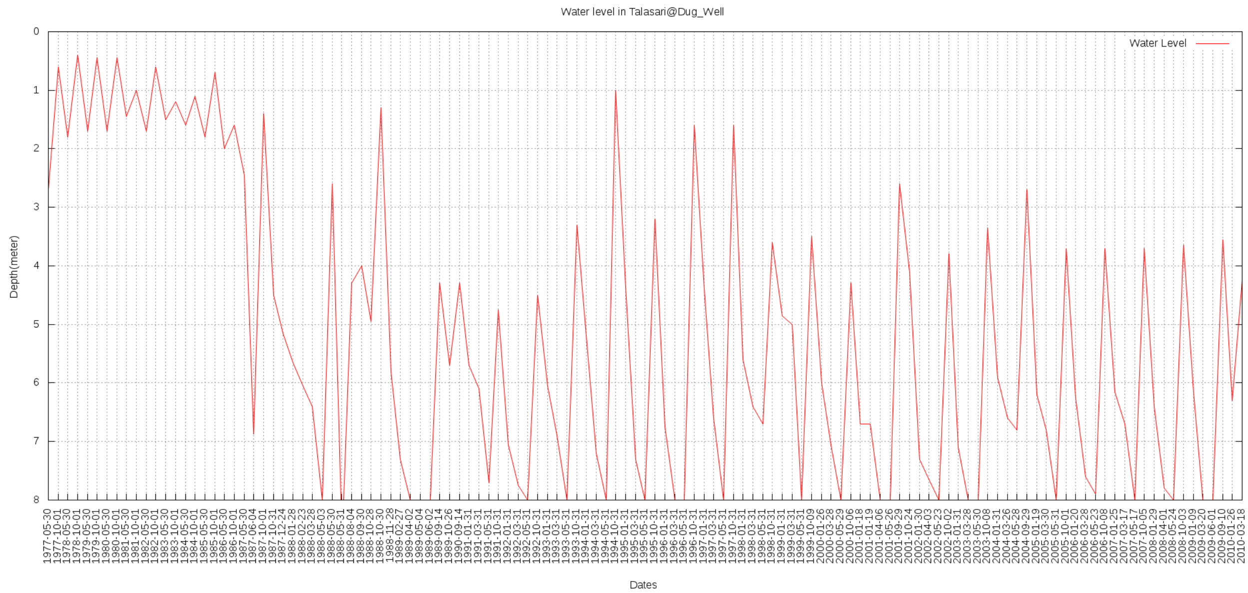

Figure 4.6: Water level in Talasari_Dug_Well over the years.

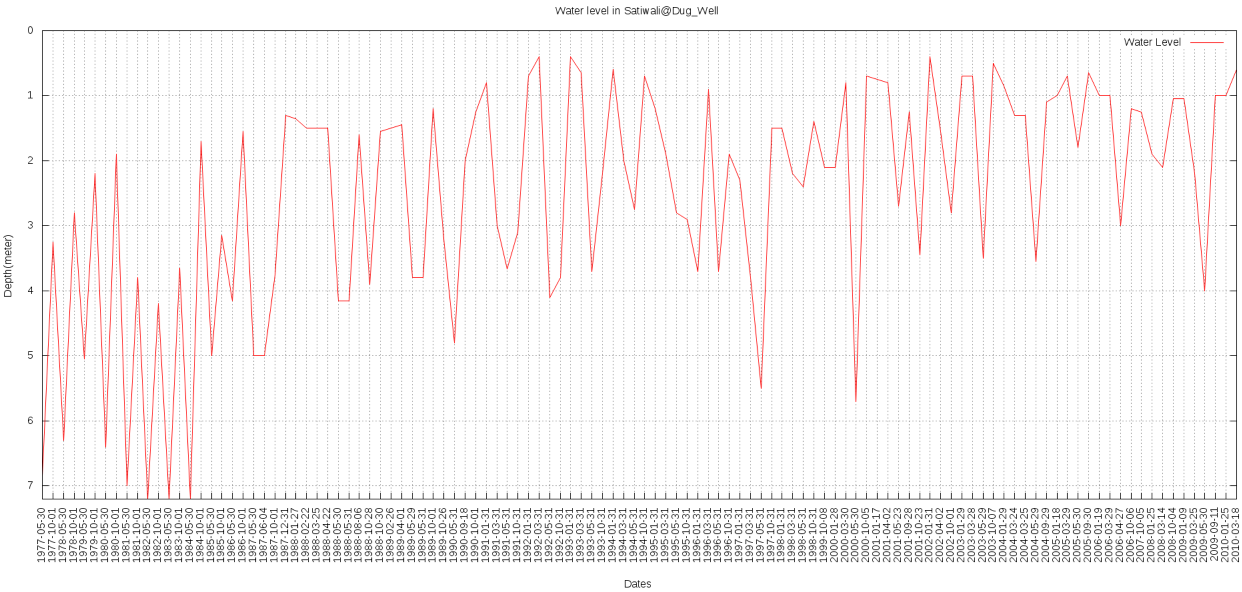

Figure 4.7: Water level in Satiwali_Dug_well over the years.

Table 4.3: Correlation of water depth between bore well and dug well in same village

Name of the village	Correlation	Distance in km
Kanchad	0.953883	0
Gokhiware	0.820901	6.21
Zhari	0.801363	1.73
Badlapur	0.553062	2.81

Chapter 5

Single Well Trend lines

5.1 Trend lines

To comprehensively understand the behavior of groundwater at a location we need to know what has been the trend of groundwater behavior in that location over a long period. Trend-lines are drawn to see these trends. The benefits of detecting the correct trend is as follows-:

- It will shed some light on the effect of rainfall on groundwater over the years.
- It will enable us to predict the future degree of threat to groundwater.
- It will enable us to predict the extraction capacity of groundwater.

5.2 GSDA Method

GSDA draws hydro graph for each block to represent the long term trends in groundwater depth. A sample of such hydro graph is shown in the figure 5.1.

They make two trend lines to predict the behavior of groundwater. One is pre-monsoon trend line and other is a post monsoon trend line. For determining the pre-monsoon trendline the average value of readings for pre-monsoon in each year is plotted and then a straight line is fitted to these points. Similarly a straight line is fitted to the average values of post monsoon period to get the post monsoon trendline. This approach does not give good trends as groundwater behavior is not linear. We intend to build more accurate trendline for each observation well so that better understanding of groundwater could be had. The approach for doing will be worked out in the subsequent stages.

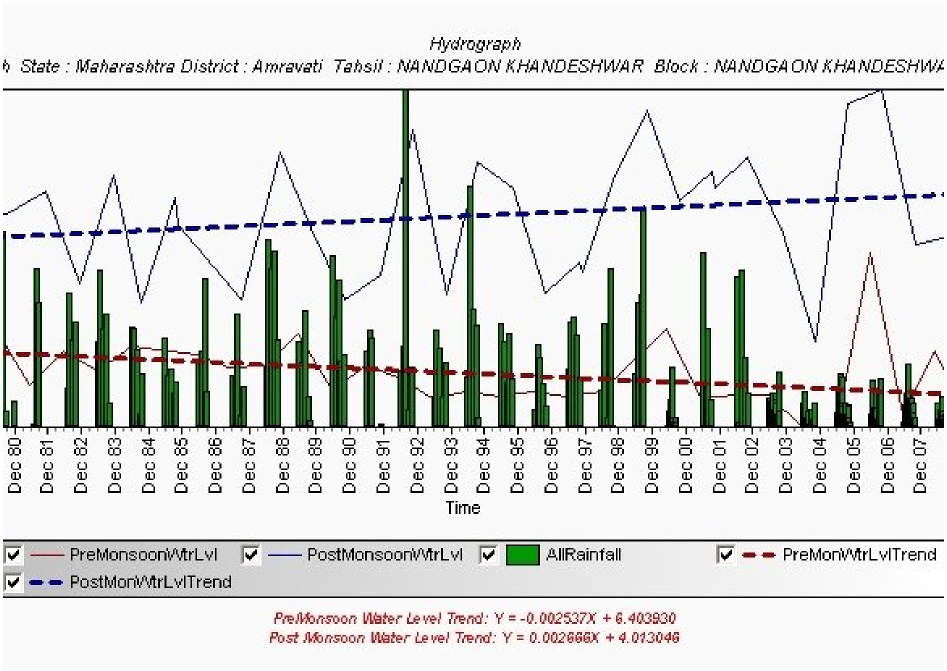

Figure 5.1: Hydro graph made by GSDA
source:http://www.mahagsda.org/gsda/web/METADATA.html

Chapter 6

Spatial Models

[10] Spatial model is a form of dividing spatial area in to number of grids or polygons of similar type. In general the output of a spatial model is a map that is subdivided in to number of regions. Where all the area that share a similar value of a particular property is grouped in to a single region. The process of developing regional seasonal models is nothing but developing a spatial model with ground water level as dividing property. So we thought better understanding of spatial model will help us in developing good regional seasonal model. As a basic step in we first did spatial analysis on rainfall data. The following section describes the spatial correlation of rainfall data that is measured at one degree interval.

6.1 Rainfall Data Correlation

We had received rainfall for the period 1986-2007 from Geographical Information Science and Engineering(GISE) Advanced Research Laboratory, IIT Bombay. This data was for area bounded by the latitude 15.5 N to 22.5 N and longitude 73.5 E to 80.5 E which covers more than entire Maharashtra. Consider this bounding box is divided by horizontal and vertical lines separated by interval of 1 degree. Henceforth we refer to points of intersections of these lines as *locations in bounding box* . The rainfall data was available for each day of every year for locations in the bounding box. The rainfall data for location at 15.5 N, 73.5 E was correlated to rainfall data at all the other locations in the bounding box for the monsoon period(June-October). For everyday in this period we calculated the following correlations-:

- Correlation on the same day e.g. data for 15^{th} June for location 15.5 N, 73.5 E was correlated to data for 15^{th} June for all the other locations in the bounding box. The table below shows the correlation value obtained for every location.

22.5	0.0111	0.0537	0.0931	0.0596	0.0690	0.0384	0.0461	0.0351
21.5	0.0678	0.0673	0.0662	0.0756	0.0831	0.0976	0.0716	0.0331
20.5	0.0844	0.1130	0.1076	0.1235	0.08300	0.0575	0.0706	0.0475
19.5	0.1768	0.0887	0.1245	0.1258	0.0968	0.0907	0.1077	0.0863
18.5	0.2222	0.0990	0.1220	0.1387	0.1166	0.1272	0.1152	0.1097
17.5	0.3218	0.1056	0.0536	0.1142	0.1194	0.0877	0.0942	0.1181
16.5	0.6259	0.1643	0.0752	0.0887	0.0974	0.0660	0.0704	0.0877
15.5	1.0000	0.4227	0.1608	0.1020	0.0877	0.1071	0.0948	0.0831
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

- Correlation with one day difference e.g (i) data for 15^{th} June for location 15.5 N, 73.5 E was correlated to data for 16^{th} June for all the other locations in the bounding box (ii) data for 15^{th} June for location 15.5 N, 73.5 E was correlated to data for 14^{th} June for all the other locations in the bounding box.

Similarly the correlation was computed for 2,3,4 and 5 day differences. The result for these correlation can be found in Appendix-B. Below are shown three tables. Each table shows on how much difference of days was the highest correlation in rainfall was found between the reference point(R) and all other locations in the bounding box.

22.5	5	5	3	4	5	5	4	5
21.5	5	2	3	4	4	4	5	5
20.5	5	0	3	3	3	4	4	4
19.5	4	1	3	3	3	3	3	3
18.5	4	0	0	0	2	2	3	2
17.5	0	0	1	0	0	1	1	4
16.5	0	0	4	1	0	1	1	1
15.5	R	0	0	0	1	1	1	1
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

22.5	2	1	1	1	1	0	0	5
21.5	2	1	1	1	0	0	0	0
20.5	1	1	1	0	0	0	0	0
19.5	1	1	0	0	R	0	0	-1
18.5	1	0	0	0	0	0	0	-1
17.5	0	0	1	0	0	0	-1	-1
16.5	0	1	0	-1	-1	-1	-1	-1
15.5	-3	-2	-2	-2	-2	-4	-1	-1
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

22.5	2	1	1	1	1	0	0	R
21.5	2	1	1	1	0	0	0	0
20.5	2	-5	-5	0	0	0	0	0
19.5	-4	-5	-5	-5	-5	0	-1	-1
18.5	-5	-5	-5	-5	-5	-5	-1	-1
17.5	-4	-5	-5	-5	-5	-5	-5	-5
16.5	-5	-5	-5	-5	-5	-5	-5	-3
15.5	-5	0	-3	-5	-5	5	-3	-3
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

6.2 Voronoi Diagram

[1], [2] Voronoi diagram is one of the spatial model in which decomposition of given space determined by distance to specified family of objects in the space. These objects are usually called the sites or the generators and to each such an object one associates a corresponding voronoi cell, namely the set of all points in the given space whose distance to the the given object is not grater than their distance to other objects. It also called as voronoi tessellation or voronoi decomposition.

Definition: Let P be a set of n distinct points (sites) in the plane.The Voronoi diagram of P is the subdivision of the plane into n cells, one for each site. A point q lies in the cell corresponding to a site $p_i \in P$ iff Euclidean distance(q, p_i) $<$ Euclidean distance(q, p_i), for each $p_i \in P$, $j \neq i$.

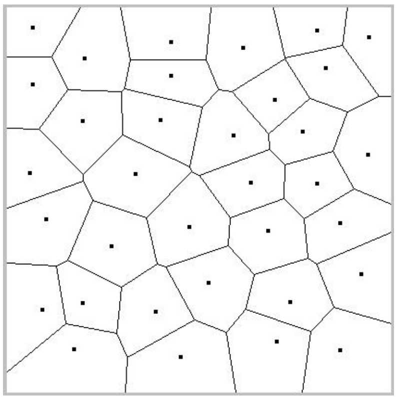

Figure 6.1: Voronoi diagram

It is useful application in many fields.Some of its applications are

- In databases the nearest neighbor queries can be answered using voronoi diagrams.
- Useful in polymer physics.

- In wireless networks.
- In climatology to estimate the rain fall of an area using series of measurement points.
- In mining to estimate the reserve of valuable minerals and materials.e.t.c.

How these voronoi diagrams are useful in our experiment. As we are interested in developing a regional spatial models like dividing a spatial region into sub regions, where each region contains wells that have high correlation between them. We are thinking that dividing the region into vornoi regions will give a better spatial model.

6.3 Krigging Interpolation

While developing our single well seasonal models we just considered the depth of ground water level as our main factor. But there are some other factors that we need to consider while computing models.They are

- Depth of the observation well.
- rainfall data.
- rainy days.
- geographical locations of observation well.

Among all these factors rainfall has huge impact on models because the water level in well is greatly effected by rainfall. So if we have exact rain fall information at observation well we can build effective seasonal model. We also need rainfall data for better discrepancy analysis(as discussed in chapter 3). Since rain gauges are usually not present at observation locations, we estimate this rainfall from rainfall at known locations. This can be done by using **Krigging Interpolation** which is a spatial interpolation technique. In krigging while estimating an interpolated value at some point it considers all the sample data with appropriate weights associated with them. We can also use krigging to extrapolate the ground water level in observation wells to any arbitrary point and predict the water level at that point. Let us discuss the mathematics behind the simple krigging.

6.3.1 Definition:

The basic problem that krigging tackles is to arrive at an estimate for a function value at a point x_0, given its observations at the points $(x_i)_{i=1}^n$. If Z is to be this function, the krigging method constructs (λ_i) (which depend on the point x_0 and the sequence (x_i)) and an estimator for $Z(x_0)$, viz.,$\sum_{i=1}^n \lambda_i Z(x_i)$. The krigging technique prescribes these λ_i under certain assumptions.

[9] Let Z be the random function of stationary model. A random function satisfying the following conditions is said to be the stationary model.

- Constant mean i.e $E[Z(x_i)] = \mu \; i = 1, 2, ..n$

- The two point covariance function should depend only on the distance between those two points.
  $R(\|x - x'\|) = R(h) = E[(Z(x) - \mu)(Z(x') - \mu)]$, where $\|x - x'\|$ is the distance between x, x'

Given n measurements of Z $x_1, x_2, ...x_n$ at different locations, Estimated value of Z at x_0 is

$$\hat{Z}_0 = \sum_{i=1}^{n} \lambda_i Z(x_i)$$

Where $Z(x_i)$ is sample data at x_i,$\hat{Z}_0$ is estimated value at x_0. Now the problem is reduced to select $\lambda_1, \lambda_2, ...\lambda_n$. The difference between actual value at x_0 i.e $Z(x_0)$ and estimated value $\hat{Z}_0$ is the estimation error.

$$\hat{Z}_0 - Z(x_0) \quad = (\sum_{i=1}^{n} \lambda_i Z(x_i)) - Z(x_0)$$

For a good estimator we should select the coefficients $\lambda_1, \lambda_2, ...\lambda_n$ in such a way that the estimator meets the following conditions.

- **Unbiasedness:** On the average the estimation error should be minimum. That is

  $$E[\hat{Z}_0 - Z(x_0)] \quad = \sum_{i=1}^{n} \lambda_i \mu - \mu \quad = (\sum_{i=1}^{n} \lambda_i - 1)\mu \quad = 0$$

  But the value of μ is not known. For any value of μ to make the estimator unbiased it is required that $\quad \sum_{i=1}^{n} \lambda_i = 1$

- **Minimum Variance:** The mean square estimation error must be minimum.

  $$\begin{aligned} E[(\hat{Z}_0 - Z(x_0))^2] &= -\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\|x_i - x_j\|) + 2\sum_{i=1}^{n} \lambda_i \gamma(\|x_i - x_0\|) \\ \gamma(\|x - x'\|) &= \frac{1}{2} E[(Z(\hat{x}) - Z(x'))^2] \end{aligned}$$

  Where γ is variogram or average of variance.

Now we have to find the values of coefficients $\lambda_1, \lambda_2, ...\lambda_n$ that minimizes the above variance expression. We can solve this problem by Lagrange multipliers. The necessary condition for the minimization are given by the linear krigging system of $n+1$ equations with $n+1$ unknowns.

$$-\sum_{j=1}^{n} \lambda_j \gamma(\|x_i - x_j\|) + v \quad = -\gamma(\|x_i - x_0\|) i = 1, 2, ..n$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

Where v is Lagrange multiplier. We now convert the above equation in to matrix form of $Ax = b$, where

$$x = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ 1 \end{bmatrix} \qquad b = \begin{bmatrix} -\gamma(\|x_1 - x_0\|) \\ -\gamma(\|x_2 - x_0\|) \\ \vdots \\ -\gamma(\|x_n - x_0\|) \\ 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & -\gamma(\|x_1 - x_2\|) & \cdots & -\gamma(\|x_1 - x_n\|) & 1 \\ -\gamma(\|x_2 - x_1\|) & 0 & \cdots & -\gamma(\|x_2 - x_n\|) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ -\gamma(\|x_n - x_1\|) & -\gamma(\|x_n - x_2\|) & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

By solving the above matrix equation we will get the coefficients $\lambda_1, \lambda_2, ...\lambda_n$. Using this coefficients we can estimate the Z at any location x_0.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this stage of the work the understanding of data, validation and flagging of discrepancies has been done. Yearly seasonal depicting the observation well behavior in a year depending on water depth observed during last 20-30 years have been done. The 120 observation wells were also plotted on the Google earth. The wells which have showed very high variance in the depth of water have been identified. We suspect that these well are not good observation wells.The rainfall correlation analysis has been done as a start up to the spatial analysis. With this we could get a rough pattern of monsoon movement. The data needed for the future work has been identified.

7.2 Future Work

The future work will be extension to the current work done. The seasonal models build will be made more reliable by incorporating other factors such as rainfall, elevation into the model. Other data such rainfall data, aquifer boundary map, watershed maps and elevation data will be procured from GSDA in the coming time. The elevation data,along with other spatial data such as latitude and longitude will be used to build spatial models. Through these spatial models we intend to know how does one well behaves as compared to other wells near its location. We aim to investigate is there some sort of pattern in observation wells which lie in the same aquifer. For this the aquifer boundary map will be used. By the end we aim to create a reference manual which could be used by GSDA in planning their drinking water scheme, water budget etc. The big picture of the project is to do similar work for other district as well.

Appendix A

Frame work

The complete framework for the project is as following-:

- Understanding and validation of data
  - Spatial distribution of wells
  - Deleting obvious errors
  - Flagging discrepancies
- Building seasonal models
  - Objectives of seasonal models
  - Mathematical formulation for models
  - Non-interpolated models.
  - Linearly interpolated models.
  - Models interpolated using spline
  - Computation of variance and its significance
- Making single well trendlines
  - Need for trendline
  - Trendline by GSDA
  - Our method of trendlines
- Building spatial models
  - Objective of spatial models
  - Spatial models at well and regional level
  - Voronoi Diagram
  - Rainfall correlation analysis
  - Krigging Method
- Feedback
  - Feedback towards better drinking water schemes.
  - Extend work to more district

Appendix B

Rainfall Correlation Results

Rainfall correlations with one 1 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.0560	0.0795	0.0879	0.0667	0.0425	0.0359	0.0306	0.0219
21.5	0.0989	0.0911	0.0885	0.0921	0.0986	0.0888	0.0724	0.0155
20.5	0.1101	0.0804	0.1000	0.1276	0.09178	0.0765	0.0633	0.0627
19.5	0.1884	0.0733	0.1131	0.1188	0.1065	0.0895	0.1128	0.0979
18.5	0.2309	0.0808	0.0868	0.1274	0.1082	0.1260	0.1144	0.1253
17.5	0.2975	0.0805	0.0102	0.0909	0.0887	0.0752	0.1143	0.1190
16.5	0.5053	0.1586	0.0145	0.0235	0.0309	0.0508	0.0465	0.0926
15.5	1.0000	0.3322	0.0885	0.0258	0.0261	0.0485	0.0293	0.0571
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with one 2 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.0667	0.0677	0.0944	0.0613	0.0543	0.0689	0.0328	0.0277
21.5	0.1310	0.1103	0.0795	0.0835	0.1125	0.1116	0.0958	0.03600
20.5	0.1017	0.0766	0.0820	0.1416	0.1170	0.0998	0.0928	0.0945
19.5	0.1938	0.0398	0.1117	0.1396	0.1539	0.1094	0.1281	0.1627
18.5	0.2309	0.0771	0.0570	0.1372	0.1365	0.1350	0.1406	0.1925
17.5	0.2681	0.0461	-0.0171	0.0337	0.0457	0.0538	0.0902	0.1221
16.5	0.3642	0.1107	-0.0104	-0.0101	0.0308	0.0401	0.0279	0.0861
15.5	1.0000	0.2953	0.0190	-0.0275	-0.0097	0.0160	0.0232	0.0388
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with one 3 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.05524	0.07634	0.10847	0.07144	0.05908	0.07871	0.05132	0.05433
21.5	0.11546	0.09432	0.10944	0.10815	0.12655	0.11510	0.11831	0.05206
20.5	0.11498	0.06772	0.11281	0.18020	0.17859	0.15336	0.10888	0.13191
19.5	0.20236	0.03752	0.12454	0.14741	0.18221	0.16201	0.14560	0.18891
18.5	0.22607	0.05574	0.05294	0.13026	0.11731	0.12998	0.15883	0.19075
17.5	0.23315	0.04520	-0.0218	0.01902	0.02556	0.04211	0.06718	0.11204
16.5	0.27804	0.08538	0.00385	0.00909	0.01954	0.03033	0.02009	0.06801
15.5	1.0000	0.24371	-0.0016	-0.0218	-0.0043	0.03088	0.01240	0.02039
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with one 4 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.07181	0.07772	0.09687	0.09440	0.06219	0.09924	0.09474	0.08730
21.5	0.13483	0.10502	0.10665	0.13075	0.12778	0.12351	0.14172	0.08323
20.5	0.13893	0.04763	0.10714	0.15973	0.16419	0.15382	0.12496	0.13416
19.5	0.20454	0.06800	0.08257	0.12955	0.15883	0.15054	0.11963	0.16679
18.5	0.24550	0.04899	0.03759	0.11834	0.11056	0.12805	0.13270	0.15122
17.5	0.23088	0.05614	-0.0094	0.02880	0.04639	0.04959	0.07104	0.12422
16.5	0.25589	0.08675	-0.0128	-0.0243	0.00125	0.01959	0.02222	0.07139
15.5	1.000	0.24463	-0.0067	-0.0428	-0.0146	0.00769	0.01597	0.01582
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with one 5 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.09944	0.09242	0.08954	0.07652	0.07788	0.10552	0.07863	0.09079
21.5	0.15986	0.08389	0.08508	0.10710	0.09599	0.10851	0.15721	0.08370
20.5	0.14094	0.02951	0.07475	0.11290	0.12070	0.12505	0.11237	0.11858
19.5	0.19072	0.02524	0.02321	0.07482	0.12349	0.12653	0.11016	0.15896
18.5	0.24080	0.04453	0.01209	0.07295	0.08253	0.10688	0.10708	0.15385
17.5	0.22998	0.05371	-0.0206	0.02778	0.03761	0.05374	0.05240	0.09813
16.5	0.21607	0.11156	-0.0199	-0.0424	-0.0118	0.00564	0.00098	0.03485
15.5	1.000	0.24261	0.00685	-0.0319	-0.0134	-0.0189	-0.0165	-0.0031
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with one -1 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.03044	0.02809	0.05466	0.03045	0.02664	0.01169	0.04365	0.03151
21.5	0.05587	0.04663	0.04663	0.06059	0.04651	0.05694	0.04605	0.05085
20.5	0.06505	0.10162	0.06641	0.10538	0.06786	0.03073	0.03267	0.04694
19.5	0.13049	0.09056	0.07852	0.07756	0.08608	0.06085	0.06228	0.07889
18.5	0.15854	0.07469	0.12177	0.12930	0.10309	0.10358	0.10332	0.10492
17.5	0.23573	0.09365	0.07768	0.07301	0.11640	0.10049	0.11770	0.09958
16.5	0.40123	0.09823	0.07477	0.09463	0.07940	0.08400	0.07495	0.12979
15.5	1.000	0.25523	0.08498	0.09605	0.11908	0.13487	0.11551	0.12917
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with -2 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.00654	0.00824	-0.0052	-0.0026	0.00727	-0.0070	-0.0036	-0.0020
21.5	0.03999	0.04427	0.03802	0.04171	0.01398	0.01682	0.00357	0.02317
20.5	0.03822	0.11200	0.06011	0.07659	0.05244	0.01126	0.00138	0.00573
19.5	0.09186	0.07150	0.07312	0.08692	0.06508	0.04753	0.05353	0.05456
18.5	0.09935	0.06396	0.08500	0.09678	0.07066	0.05875	0.06333	0.07031
17.5	0.15822	0.04805	0.05384	0.07014	0.08805	0.07528	0.07011	0.06406
16.5	0.26103	0.06814	0.05083	0.07172	0.05682	0.05121	0.06607	0.12188
15.5	1.000	0.26103	0.06814	0.05083	0.07172	0.05682	0.05121	0.06607
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with one -3 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.01123	0.00036	-0.0120	-0.0089	-0.0230	-0.0196	-0.0287	-0.0187
21.5	0.01937	0.01988	0.03183	0.01644	-0.0067	-0.0088	-0.0169	-0.0207
20.5	-0.0034	0.08913	0.06863	0.07031	0.03704	0.01205	-0.0154	-0.0096
19.5	0.04261	0.05042	0.07983	0.08198	0.07082	0.03780	0.03072	0.01723
18.5	0.03504	0.02255	0.05469	0.09409	0.05799	0.06506	0.04201	0.01995
17.5	0.09706	0.03184	0.04103	0.03953	0.07847	0.04554	0.07919	0.05042
16.5	0.17796	0.03540	0.05833	0.05286	0.05113	0.06923	0.04627	0.09799
15.5	1.000	0.08287	-0.0032	0.00469	0.07322	0.10952	0.08632	0.07977
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with -4 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.00531	0.01267	-0.0076	-0.0063	-0.0127	-0.0334	-0.0440	-0.0433
21.5	-0.0003	0.01022	0.01795	0.00293	-0.0190	-0.0182	-0.0335	-0.0464
20.5	-0.0029	0.04765	0.01713	0.02738	0.01164	-0.0029	-0.0309	-0.0324
19.5	0.02639	0.03310	0.05169	0.02955	0.02211	0.01054	0.01620	-0.0058
18.5	0.00521	0.00108	0.03294	0.04472	0.02692	0.03204	0.02650	0.01268
17.5	0.04252	0.01628	0.05552	0.07500	0.02632	0.03448	0.05167	0.03870
16.5	0.12656	0.05819	0.09761	0.05290	0.00382	0.03170	0.00528	0.06088
15.5	1.0000	0.06627	-0.0051	0.01274	0.05064	0.09231	0.07267	0.04684
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Rainfall correlations with -5 day difference between the point 15.5 N, 73.5 E and all other locations.

22.5	0.00745	0.03078	0.00971	-0.0058	-0.0318	-0.0397	-0.0375	-0.0328
21.5	-0.0037	0.02477	-0.0051	-0.0013	-0.0070	-0.0058	-0.0299	-0.0374
20.5	-0.0062	0.04621	0.01175	0.01413	0.03955	0.00378	-0.0419	-0.0232
19.5	0.00385	0.02499	0.03015	0.00708	0.03141	-0.0029	-0.0140	-0.0323
18.5	0.00321	-0.0070	0.02229	0.02574	0.01670	0.01391	0.00879	-0.0071
17.5	0.02471	0.00616	0.04083	0.04542	0.03172	0.03997	0.03470	-0.0013
16.5	0.09000	0.02883	0.05605	0.04392	0.04658	0.05391	0.00999	0.04330
15.5	1.0000	0.04737	-0.0127	-0.0056	0.06300	0.04530	0.05042	0.00607
Latitude Longitude	73.5	74.5	75.5	76.5	77.5	78.5	79.5	80.5

Appendix C

Variance Vs Discrepancy count

Table C.1: Variation Vs Discrepancy for Bore Well

Village_Site_type	Normalized variance	Depth(m)	Discrepancy_count
Tokavde_Bore_Well	0.135601	24.000000	1
Mandawa_Bore_Well	0.118143	30.000000	2
Gokhiware_Bore_Well	0.090139	18.000000	4
Ambiste_kh_Bore_Well	0.088790	17.000000	4
Chavindra_Bore_Well	0.088093	13.500000	12
Sakharshet_chalatwad_Bore_Well	0.084797	22.500000	7
Safale_Bore_Well	0.081599	25.900000	3
Nare_Bore_Well	0.070050	18.000000	0
Ghansoli_Bore_Well	0.069164	12.700000	7
Kudan_Bore_Well	0.069139	30.000000	3
Saravali_Bore_Well	0.066113	24.000000	10
Goveli_Bore_Well	0.063398	17.250000	9
Talwada_Bore_Well	0.061150	30.000000	2
Waret_Bore_Well	0.056852	30.000000	1
Chndansar_Bore_Well	0.052540	24.000000	7
Palghar_kolgaon_Bore_Well	0.051322	30.000000	4
Kanchad_Bore_Well	0.051086	18.000000	5
Bhatsai_Bore_Well	0.050426	18.000000	17
Mahim_Bore_Well	0.048823	20.000000	6
Padgha_Bore_Well	0.045149	30.000000	5
Vasar_Bore_Well	0.043790	30.000000	20
Neharoli_Bore_Well	0.037457	24.000000	12
Suksale_Bore_Well	0.036618	30.000000	7
Zhari_Bore_Well	0.036371	30.000000	3
Satiwali_Bore_Well	0.033998	18.000000	23
Nimbavali_Bore_Well	0.032268	30.000000	3
Udawa_Bore_Well	0.029945	30.000000	6
Badlapur_Bore_Well	0.010982	30.000000	9

Table C.2: Variation Vs Discrepancy for Dug Well

Village_Site_type	Normalized variance	Depth(m)	Discrepancy_count
Washind_1_Dug_Well	0.361308	7.000000	7
Talasari_Dug_Well	0.286490	8.000000	5
Satiwali_Dug_Well	0.244649	7.200000	15
Mangrul_Dug_Well	0.236172	7.600000	3
Karav_Dug_Well	0.210794	8.000000	5
Dapode_Dug_Well	0.208616	5.250000	7
Zhai_Dug_Well	0.204683	7.700000	4
Pelhar_Dug_Well	0.195875	7.000000	3
Sakwar_Dug_Well	0.193537	6.000000	2
Kasa_bk_Dug_Well	0.193456	6.500000	4
Vevaji_Dug_Well	0.188279	7.600000	5
Kudus_Dug_Well	0.187964	6.000000	5
Shilphata_Dug_Well	0.185485	4.800000	2
Dahisar_Dug_Well	0.184878	9.500000	2
Awale_Dug_Well	0.184196	7.350000	6
Katrap_Dug_Well	0.181839	3.100000	7
Kambe_Dug_Well	0.181274	6.900000	4
Pimpalas_Dug_Well	0.179127	6.550000	5
Jawhar_Dug_Well	0.177788	7.650000	5
Tokawade_Dug_Well	0.176864	5.000000	5
Newale_Dug_Well	0.175254	8.200000	7
Pimpalshet_Dug_Well	0.168240	8.500000	6
Morhande_Dug_Well	0.166949	5.100000	2
Zhari_Dug_Well	0.166374	7.400000	2
Karvele_Dug_Well	0.164134	6.300000	6
Washind_2_Dug_Well	0.164020	3.050000	1
Gates_Bk_Dug_Well	0.163726	7.500000	2
Musarne_Dug_Well	0.163443	6.000000	1
Khodala_Dug_Well	0.160623	5.800000	5
Khaniwade_Dug_Well	0.160202	5.000000	3
Shivale_Dug_Well	0.158998	11.000000	4
Parli_Dug_Well	0.156835	5.100000	9
Dolhare_Dug_Well	0.155206	5.500000	2
Kanhor_Dug_Well	0.155049	8.500000	3
Vedhi_Dug_Well	0.154184	8.700000	2
Titwala_Dug_Well	0.150577	7.000000	10
Thilher_Dug_Well	0.149291	6.200000	2
Kanchad_Dug_Well	0.149217	7.500000	3
Kalamdevi_Dug_Well	0.147872	5.500000	4
Shil_t_chon_Dug_Well	0.147601	7.100000	3

Village_Site_type	Normalized variance	Depth(m)	Discrepancy_count
Talasarimal_Dug_Well	0.146322	8.200000	6
Thunepada_Dug_Well	0.146191	5.950000	5
Makunsar_Dug_Well	0.145569	9.800000	2
Shendrun_Dug_Well	0.144218	4.700000	3
Sasne_Dug_Well	0.143413	8.850000	3
Safala_Dug_Well	0.143257	10.500000	2
Dhanivri_Dug_Well	0.142133	5.500000	0
Kharade_Dug_Well	0.141822	8.200000	5
Rayta_Dug_Well	0.141264	4.000000	6
Dhuktan_Dug_Well	0.139684	6.100000	2
Vehaloli_Dug_Well	0.139429	5.100000	2
Bhinar_Dug_Well	0.137708	6.250000	3
Dhanoshi_Dug_Well	0.136474	6.500000	3
Veyour_Dug_Well	0.132438	10.100000	2
Chahade_Dug_Well	0.132308	5.700000	1
Mandvi_Dug_Well	0.131775	9.100000	1
Bursunge_Dug_Well	0.130789	8.650000	4
Manor_Dug_Well	0.129866	7.000000	9
Bapgaon_Dug_Well	0.129090	7.400000	1
Kogde_Dug_Well	0.129032	7.000000	3
Nihe_Dug_Well	0.127235	7.000000	2
Sange_Dug_Well	0.126788	4.700000	2
Tembhare_Dug_Well	0.126783	5.500000	6
Durves_Dug_Well	0.126400	9.600000	0
Pawane_Dug_Well	0.123983	5.000000	8
Badlapur_Dug_Well	0.122507	7.950000	9
Sawta_Dug_Well	0.119359	8.400000	4
Inde_Dug_Well	0.114526	7.800000	2
Vihigaon_Dug_Well	0.113316	7.500000	2
Lalthan_Dug_Well	0.112166	6.400000	2
Vadoli_Dug_Well	0.110972	5.600000	5
Warwade_Dug_Well	0.107835	7.600000	3
Ghodbandar_Dug_Well	0.105859	8.200000	4
Met_Dug_Well	0.105451	8.300000	2
Kopar_Karane_Dug_Well	0.104875	4.700000	5
Mokhada_Dug_Well	0.104028	9.000000	3
Varaskol_Dug_Well	0.099815	7.000000	2
Shelonde_Dug_Well	0.099336	12.500000	5
Borivali_T_Padgha_Dug_Well	0.098335	10.600000	8

Village_Site_type	Normalized variance	Depth(m)	Discrepancy_count
Talegaon_Dug_Well	0.097990	6.100000	3
Gokhiware_Dug_Well	0.095679	5.500000	3
Pali_Dug_Well	0.095415	6.000000	2
Govade_Dug_Well	0.093311	6.600000	0
Akoli_Dug_Well	0.093280	5.500000	1
Thane_Dug_Well	0.091938	7.050000	12
Ghol_Dug_Well	0.091636	10.400000	15
Deoli_Dug_Well	0.089111	6.200000	1
Kajali_Dug_Well	0.087745	14.000000	11
Kopari_Dug_Well	0.080330	7.550000	10
Wada_Dug_Well	0.078266	9.000000	6
Shirgaon_Dug_Well	0.069068	9.000000	1
Agashi_Boling_Dug_Well	0.060889	10.000000	4

Appendix D

GSDA Meeting

A meeting with GSDA officials was organized, the details of meeting are as follows-:

- Date : 11^{th} October, 2011
- Venue : GSDA office, Thane
- Members
  - Purushottam Kulkarni (Professor CSE, IIT Bombay)
  - A.G Hegde (Senior Geologist, GSDA Thane)
  - Lalit Kumar (M.Tech student, IIT Bombay)
  - Ravi Sagar (M.Tech student, IIT Bombay)
- Minutes of meeting
  - Selection of observation wells was discussed with Mr. Hegde. Observation are chosen keeping in mind the watersheds. Each watershed consist of three parts-runoff, recharge and storage and typically each part contains a observation well.
  - For better understanding of observation wells behavior their soil type,elevation and rainfall has to be considered.
  - The well depth may increase over the period due to de-silting but the well depth is not measured very frequently.
  - Water depth in dry wells can be estimated from the neighbouring wells, neighbour has to be defined on the basis of elevation, soil type and watershed..
  - The water level trends are calculated from average water levels in pre-monsoon and post-monsoon.
  - Mr. Hegde explained briefly the water budget assessment procedure. This is mostly based on population, drinking water requirement(40lpcd), rainfall and recharge.
  - Data needed from GSDA for the project ahead is:(i) Tahsil level rainfall data (ii) Elevation data (iii) Watershed maps (iv) Water budget document (v) Hydrographs generated by GSDA for Thane.
  - A formal request to get the data was made to GSDA Thane office.

Bibliography

[1] Voronoi Diagrams. http://en.wikipedia.org/wiki/voronoi_diagram.

[2] Voronoi Diagrams. http://www.cs.wustl.edu/ pless/546/lectures/l11.html.

[3] Linear Curve Fitting. http://en.wikipedia.org/wiki/linear_regression.

[4] Linear Curve Fitting. http://www.stat.yale.edu/courses/1997-98/101/linreg.htm.

[5] Ralph C Heath. *Basic Ground-Water Hydrology.* U .S . Geological Survey Water-Supply Paper 22200, 1983.

[6] Linear Interpolation. http://en.wikipedia.org/wiki/linear_interpolation.

[7] Spline Interpolation. http://en.wikipedia.org/wiki/spline_interpolation.

[8] Spline Interpolation. http://www.math.utah.edu/ wright/applets/spline/spline.html.

[9] P.K Kitanidis. *Introduction to GEOSTATISTICS Appllication in Hydrogeology.* Press Syndicate of tthe University of Camebridge, 1997.

[10] Spatial Models. http://en.wikipedia.org/wiki/spatial_analysis.

[11] Sheldon M. Ross. *INTRODUCTION TO PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS.* Elsevier Academic Press, third edition edition, 2004.

[12] Interpolation Techniques. http://en.wikipedia.org/wiki/interpolation.