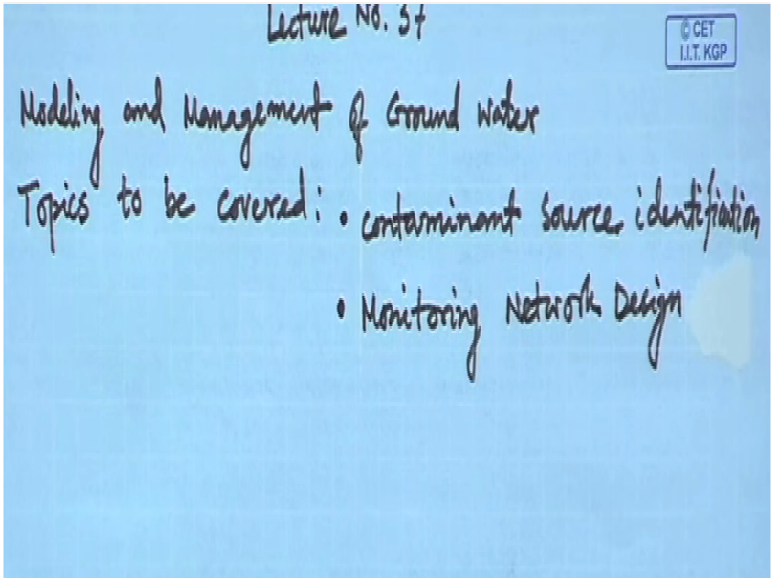
Ground Water Hydrology
Prof. Anirban Dhar
Department of Civil Engineering
Indian Institute of Technology – Kharaghpur

Module No # 08
Lecture No # 37
So Modeling and Management of Ground Water: Contaminant Source, Identification, Monitoring Network Design

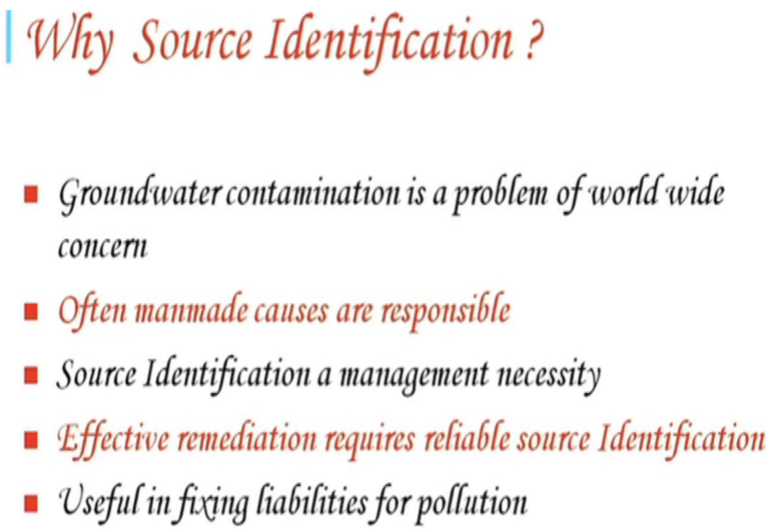
Welcome to this lecture number 37 of ground water hydrology course and in this particular lecture class cover this modeling.

(Refer Slide Time: 00:29)



Management of groundwater, under this topic to be covered is contaminant source identification and monitoring network design. So first topic is this contaminant source identification.

(Refer Slide Time: 01:43)

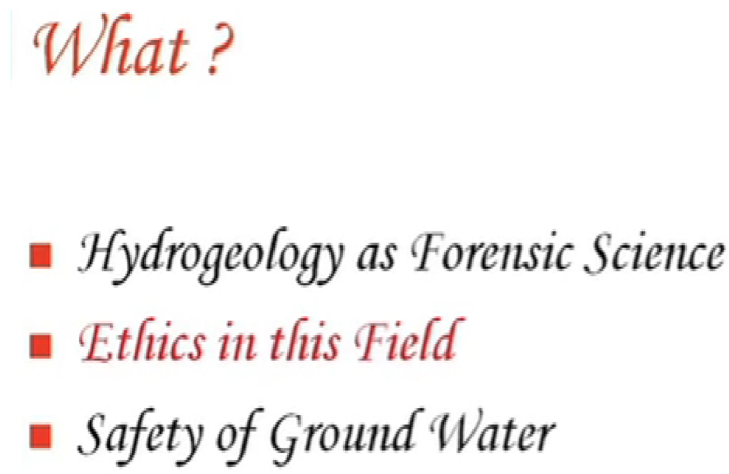


Why this contaminant source identification groundwater contamination is a problem of worldwide concern and this is often manmade causes are responsible. Either we can have some kind of geogenic source or we can have anthropogenic thing like arsenic problem is geogenic thing but contamination. There can be dumping of some pollutant in unlined pones that is some kind of man made cause. So source identification its management is necessity, why it is necessary?

It is necessary from groundwater management point of view, if we can manage to identify the source, we can have some kind of remediation strategy for that particular aquifer and we can decontaminate. We can start the remediation process in that particular aquifer. Next is effective remediation requires reliable source identification. So this is the point that remediation thing we need some kind of proper estimate about the source both its space and time.

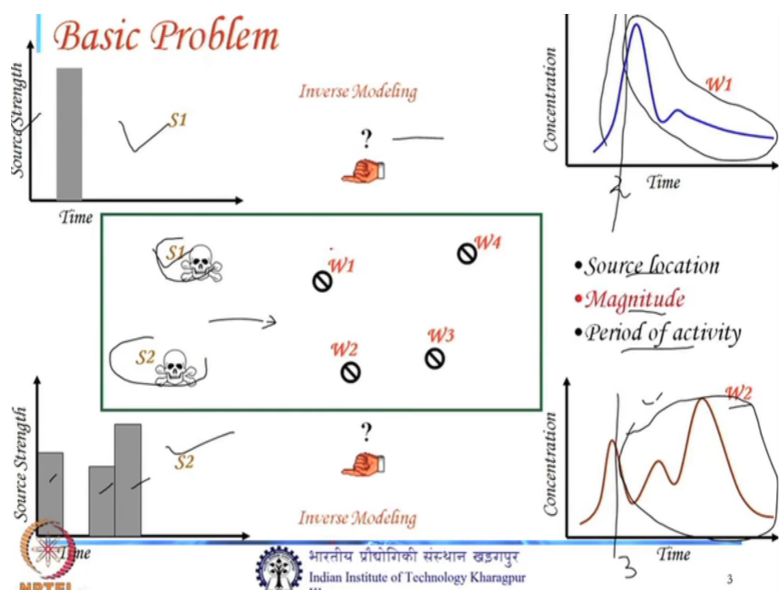
And its strength also like key parameter for identification. So useful in fixing liabilities for pollution, let us say for manmade causes we cannot do anything with the geogenic causes but for manmade causes it is important that we should fix the liabilities. So that if 1, 2 or 3 or 4 defaulters are present within that groundwater juristic area, then we can fix the share of their responsibilities for the contamination or some kind of prize that they need to pay for the health related spending of that locality.

(Refer Slide Time: 05:04)



So what is this hydrogeology as forensic science? And forensic because, we need to identify the things properly without identification it is a difficult thing and ethics in this field. So which should have some kind of ethics and safety of groundwater, that is the most important thing without proper management strategy or identification strategy we cannot protect our groundwater.

(Refer Slide Time: 05:48)



So what is the basic problem? Basic problem is that, let us say we have a source S1, another source S2 and this rectangular part is one aquifer. Let us say our wells W1, W2 wells, W3 and W4. These are in the down gradient portion of the aquifer. Let us say this is your flow direction, so with this aquifer configuration and these many observations well or monitoring wells.

We can identify S1 or S2 in terms of its location whether S1 is responsible, whether S2 is responsible in terms of this its strength, whether for the first year, second year, third year fourth year may be S1 is responsible for the contamination only in the second year. But in case of S2 they are responsible for the contamination of first, third and fourth year that is also important to find out the activity period.

So three things are important here source location, source strength and their activity period. So these three points are most important, so in this case forward modeling will give us source and strength. This source and this strength will give us some kind of breakthrough curve that is time versus concentration curve for W1 location like this and for more forward modeling, we can get W2 or breakthrough curve for W2.

As this multiple peaks due to different strength and their activity periods but the complicated problem is that we may not have a proper observation or management plan in place and we do not have any proper monitoring strategy for that contaminated area. So it may so happen that the contamination may be noticed after 10 or 20 years after it has started in that particular area. So the problem is that inverse modeling.

Inverse modeling means we have the breakthrough curve, maybe we can find the breakthrough curve in trunk ended sense. Trunk ended means, let us say that we are starting the monitoring and the end of second year. So we will get this part of the breakthrough curve only. So same for this part, let us say we have started this W2 well observation after third year.

So we may get this kind of breakthrough curve or trunk ended breakthrough curve. So it is important that with its complete or trunk ended or limited information, we need to have some kind of proper estimate of the source in terms of its strength, its location and its activity period.

(Refer Slide Time: 11:05)

Inverse problem: types

- *Backward or retrospective problem*
  - *The initial conditions are to be found.*
- *Coefficient inverse problem*
  - *This is the classical parameter estimation problem where a constant multiplier in a governing equation is to be found.*
- *Boundary inverse problem*
  - *Some missing information at the boundary of a domain is to be found.*

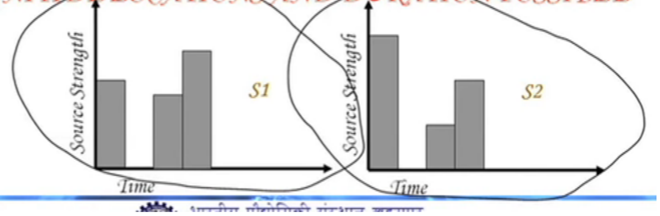
So source location magnitude and period of activity are three important aspects now what are these inverse problems? So first type is backward or retrospective problem, the initial conditions are to be found. Second one is coefficient inverse problem in this one classical parameter estimation problem where a constant multiplier in the governing equation is to be found out and boundary inverse problem.

Some missing information at the boundary of a domain is to be found out, so our problem is basically backward or retrospective problem, inverse problems are mostly ill posed problems because most of the cases. We may not have a unique solution for a unique inverse problem.

(Refer Slide Time: 12:11)

Difficulties in source identification

- *Sparsely distributed observation wells*
- *Sparsity of observation data*
- *Inaccurate prediction of contaminant transport processes-modeling errors, measurement errors etc.*
- *BILLIONS of Possible Discrete COMBINATIONS OF MAGNITUDE LOCATIONS AND DURATION POSSIBLE*

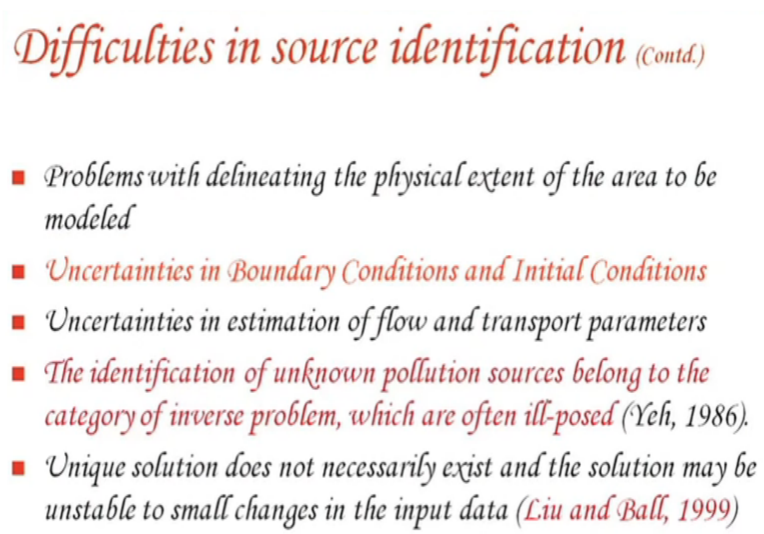


So difficulties in source identification sparsely distributed observation wells that is the most important thing because if there is no proper observation or monitoring network is there then it is a problem and sparsity of observation. Data observation also sparsed in nature, inaccurate prediction of contaminant, transport processes, modeling errors, measurement errors. So one kind of error that may be there is related to modeling error.

Another one is measurement error, so sparsity of data error in measurement and error in modeling. These three are the most important things for source identification. So billions of possible discrete combination of magnitude locations and duration is possible. So there may be multiple combinations that will give you the same state of break breakthrough curves, that is available for a particular monitoring well.

So it is important that a proper strategy should be followed for monitoring, otherwise there will be difficulties in identification of sources. So we can have situations where, let us say this is first kind of combination in our previous problem. We were having S1 and S2 these two are the sources, so we have two things that this kind of S1 combination and the second S2 combination may give the same results in the down gradient monitoring wells in terms of breakthrough curves.

(Refer Slide Time: 14:58)



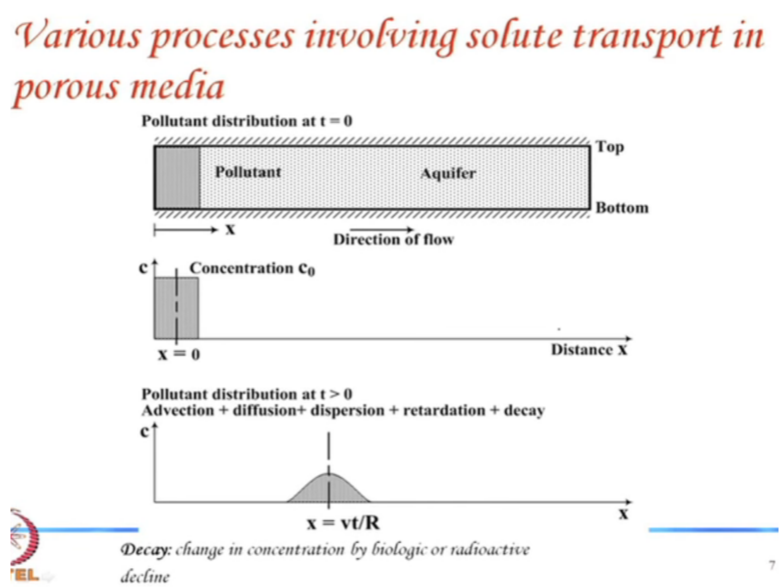
So other difficulty problems with delineating the physical extent of the area to be modeled because we need to have certain kind of limitation in terms of delineating the physical extent of the area uncertainties in the boundary conditions and initial conditions. So the problem is that although in terms of our classification, we are interested in finding out the initial conditions but boundary conditions are also important.

Because physical extent is important and if we are not considering a proper physical extent of the area that we model then boundary condition is a critical thing for modeling uncertainties in estimation of flow transport parameters. So uncertainties are also there in terms of estimation of flow transport parameters like hydraulic conductivity, longitudinal dispensability, transport dispensability.

These can play important tool in modeling and identification of sources and identification of unknown pollution sources belongs to the category of inverse problem which are often ill posed because we do not have a proper system in place for which we can say that this is our proper source that we have identified from our problem.

But the problem is that we can have a multiple combination for which there will be same breakthrough curve in the monitoring wells and unique solution does not necessarily exist and solution may be unstable to small changes in the input data. So sensitivity of the solution approach is that another important issue.

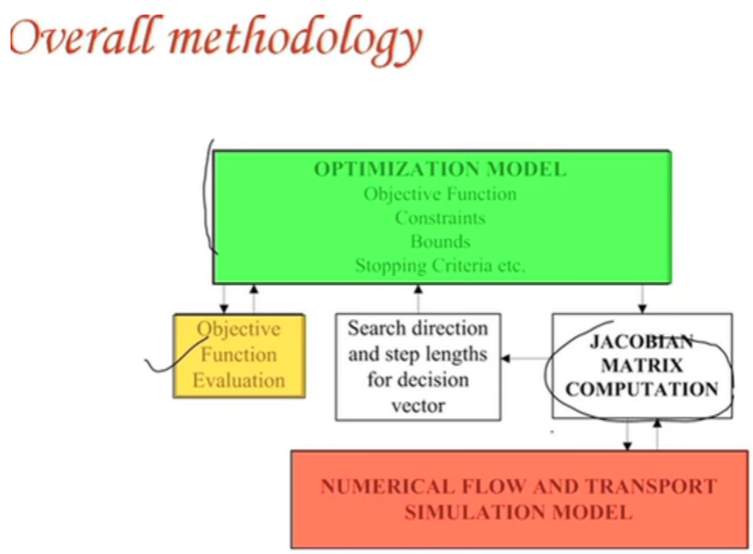
(Refer Slide Time: 17:57)



So various process involving solute transport in the porous media. So we can have advection processes where groundwater flow is caused by gravity. Next we can have diffusion molecular process where constituents are spread due to differences in concentration. Next we can have dispersion mixing process caused by differences in velocity in magnitude and direction of water particles.

Another one is adsorption processes where certain constituents are attached to grain material and final thing is that decay. So there will be combination of these processes which will dictate the source identification. So proper transport process identification is the first part of any source identification problem.

(Refer Slide Time: 19:08)

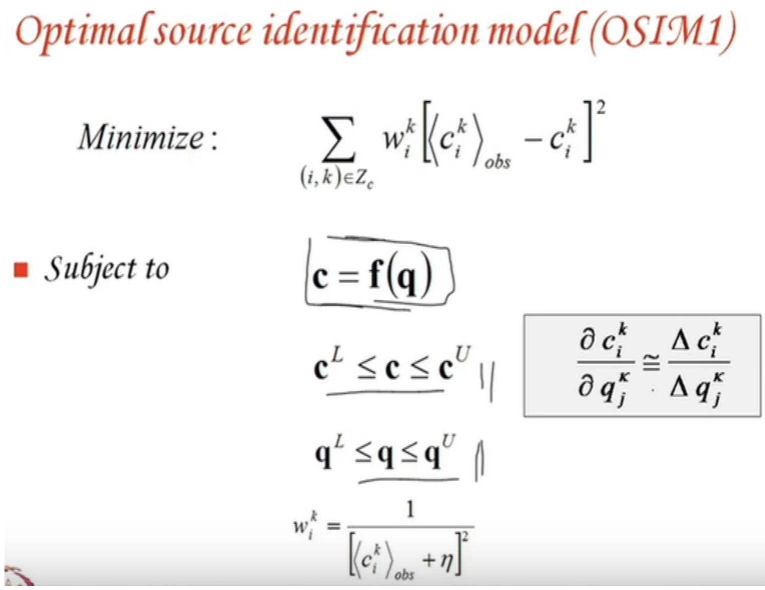


So overall methodology is that, we can have optimization model, objective function, and we can have Jacobian matrix and with this we can have some kind of search direction step length for decision vector that can be determined and the most important thing is the flow transport stimulation model. So we can use or flow transport stimulation model as linked stimulation optimization model as external module and we can calculate our objective functions.

Also our search directions are basically dependent on the Jacobian matrix and with this Jacobian matrix it will give the proper direction. But the problem is that if we have Jacobian matrix best approach, then you may or may not get a proper global optimal solution. So it is important that your selection of optimization model is also important thing.

So first is identification of proper transport process that is identification of proper numerical flow and transport simulation model, simulate the complex hydrogeological system. Next is identification of proper optimization model and intermediate things are your linking things these are intermediate calculations.

(Refer Slide Time: 21:17)



So in this case we can have through optimization problems. First one is the observed concentration in monitoring well and this is our estimated concentration. So the square of this distance and this is weighted one. So weight is calculated like this is observed concentration plus ETA value. ETA is a small value which gives some kind of support for the weight. This C, FQ, FQ is again or simulation model.

This is restriction in terms of concentration in terms of injection from our sources. Let us say that there will be some kind of physical limit for the injection rate that is maintained with this particular constraint and there is some physical limit for concentration that is maintained with this particular constraint and this is one part typical component of our jacobian matrix. That is being calculated based on final difference approach or difference approach.

(Refer Slide Time: 23:19)

Optimal source identification model (OSIM2)

$$\text{Minimize}: \quad \Delta_\tau$$

- *Subject to*

$$\Delta_\tau - \sum_{(i,k)\in Z_c} w_i^k \left[\left\langle c_i^k \right\rangle_{obs} - c_i^k\right]^2 = 0$$

$$\mathbf{c} = \mathbf{f}(\mathbf{q})$$

$$\mathbf{c}^L \leq \mathbf{c} \leq \mathbf{c}^U$$

$$\mathbf{q}^L \leq \mathbf{q} \leq \mathbf{q}^U$$

Second model is that your objective function is linear in nature but our constraints are non linear. Previously we have seen that our objective function was non linear in OISM 2 or OISM 1 and this is also non linear constraint and these are our linear constraint. So some literatures suggest that there is advantage in placing any linear objective function in state of non linear objective function.

So it is being converted like this is considered as equality constraint within the optimization problem. So we need to incorporate some kind of errors. So for a particular hypothetical illustrative problem, we have considered that this is a simulated value. We have added some kind of error this it represents the measurements are assumed error free it may represent a special case where the value is 0 and standard normal random variant for concentration.

So we can introduce some kind of error with the simulated values and we can create some observed value to test our true objective function or our true formulations OSIM 1 and OSIM 2.

(Refer Slide Time: 25:46)

Performance Evaluation Criteria

- *Normalized error estimate for source fluxes*

$$NEE_f(\%) = \frac{\sum_{\kappa=1}^{N_{dp}} \sum_{j=1}^{N_{dl}} \left| \langle q_j^{\kappa} \rangle_{est} - \langle q_j^{\kappa} \rangle_{act} \right|}{\sum_{\kappa=1}^{N_{dp}} \sum_{j=1}^{N_{dl}} \langle q_j^{\kappa} \rangle_{act}} \times 100$$

- *Standard deviation of estimated source flux*

$$\langle q_j^{\kappa} \rangle_{S.D.} = \sqrt{\frac{\sum_{l=1}^{N_r} \left[\langle q_j^{\kappa} \rangle_l - \langle q_j^{\kappa} \rangle_{avg} \right]^2}{N_r - 1}}$$

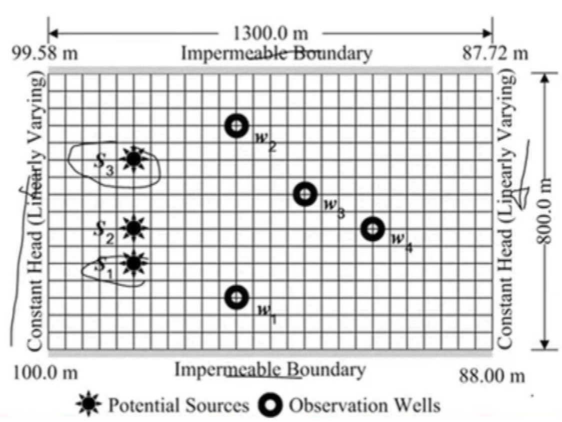
with average estimated value of $\langle q_j^{\kappa} \rangle_{avg} = \frac{1}{N_r} \sum_{l=1}^{N_r} \langle q_j^{\kappa} \rangle_l$

So what is our evaluation criteria? We have used this normalized errors estimate for source fluxes. This is actual value and this is estimated value, so this is average one for this is for a particular case so this is number of realizations. So for multiple realization, what is the difference? So from this we can guess the standard deviation of the estimated source flux and this is the average estimated value of the source strength.

(Refer Slide Time: 26:41)

Illustrative application (ISA-I)

50 m × 50 m × 2 months



So let us consider one illustrative problem where we have three sources one S1, S2 and S3 and we have 3, 4 monitoring wells W1, W2, W3 and W4 out of this. This thing is having two impermeable boundaries and two constant head boundaries. These are linearly varied, so in this direction and this is final direction.

(Refer Slide Time: 27:32)

Solution results

Disposal Period	Source Locations	Source Fluxes (gm/sec) Actual Flux	Estimated Source Fluxes Mahar and Datta (1997)	MINOS OSIM1	MINOS OSIM2	NPSOL OSIM1	NPSOL OSIM2
1	S_1	47.00	46.75	47.00	46.99	47.00	46.99
	S_2	0.00	0.00	0.00	0.00	0.00	0.00
	S_3	30.00	29.92	30.00	29.99	30.00	30.00
2	S_1	15.00	15.38	15.00	15.00	15.00	15.00
	S_2	0.00	0.00	0.00	0.00	0.00	0.00
	S_3	58.80	54.86	58.80	58.80	58.80	58.80
3	S_1	37.00	36.20	36.99	36.99	37.00	36.99
	S_2	0.00	0.00	0.00	0.00	0.00	0.00
	S_3	0.00	0.00	0.00	0.00	0.00	0.00
4	S_1	0.00	0.00	0.00	0.00	0.00	0.00
	S_2	0.00	0.00	0.00	0.00	0.00	0.00
	S_3	35.00	33.63	35.00	35.0	34.99	35.00
NEE_f			3.06 %	≈ 0 %	≈ 0 %	≈ 0 %	≈ 0 %

So this is the pollution result for disposal period 1 it has been found that this is the actual flux and this is the estimated flux wells. This is 47 from OSIM this is giving better result compared to OSIM 2 using this minus and NPSOL and this is actual flux 0 for all the cases. This is 30 the strength and in case of MAHAR and DATTA.

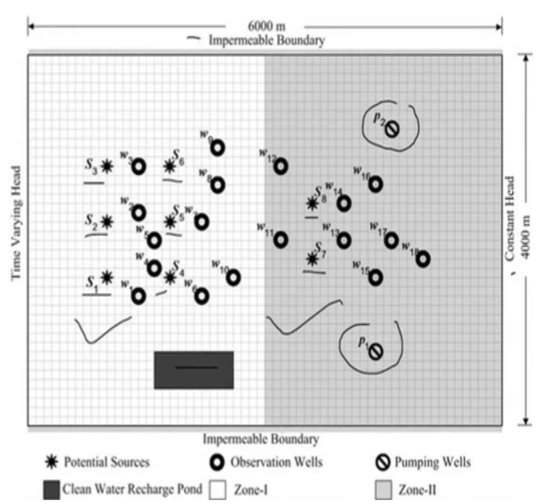
This was 39.92, this is 30, this is 29.99, this is 30 again. This is 30 and for period 2 also the things are almost matching period 3, it is almost matching period four. This both methods are matching in case of NPSOL.

(Refer Slide Time: 29:19)

Illustrative application (ISA-II)

$$h(t) = 120\left(1 + \frac{\sin(\omega t_y)}{60}\right), \forall t \in [t_y, t_y + 1]$$

$$\omega = \pi/10$$



This OSIM 2 is performing better and in this case we have 1, 2, 3, 4, 5, 6, 7, 8 sources or potential sources and this many pumping or observation wells and these two are basically pumping wells. We have two zone cases here and we have clean recharge pond. We have zone 1, then zone 2 and this HT is time varying it which is defined like this and this is a constant head and these two boundaries are impermeable boundaries.

(Refer Slide Time: 30:16)

Pumping rates in l/s

	Year									
Well	1	2	3	4	5	6	7	8	9	10
p_1	5.60	4.00	4.50	6.65	6.40	3.80	5.20	5.50	6.20	5.00
p_2	8.00	6.50	7.50	5.00	7.00	5.60	6.70	4.80	5.50	4.60

So with this kind of pumping rates, these are the pumping rates for location one P1 and P2

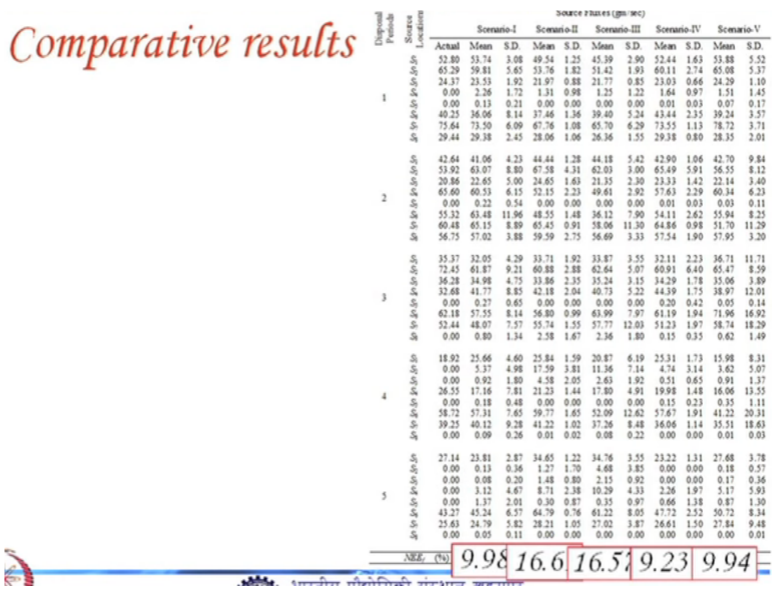
(Refer Slide Time: 30:31)

Different scenarios

- *Scenario-I: incorporating error factor (ξ) of 0.10 in the measurement data*
- *Scenario-II: 5 % increase in hydraulic conductivity*
- *Scenario-III: 5% decrease in porosity value*
- *Scenario-IV: 5 % increase in longitudinal dispersivity*
- *Scenario-V: missing data with error factor (ξ) of 0.10*
  - *concentration data during the first five years are assumed to be missing for the observation wells w4, w5, w6, and w8*
  - *observation wells w11, w13, w16, and w18, the concentration data during the first ten years*

Have considered different scenario one that is error factor is theta, that is .1 in the measurement data 5 % increase in hydraulic conductivity, 5% decrease in the porosity values, 5% increase in longitudinal dispersivity missing data with error factor concentration data during the first 5 years are assumed to be missing for the observation wells 4, 5, 6 and 8 and observation wells W1, W13, W16, W18.

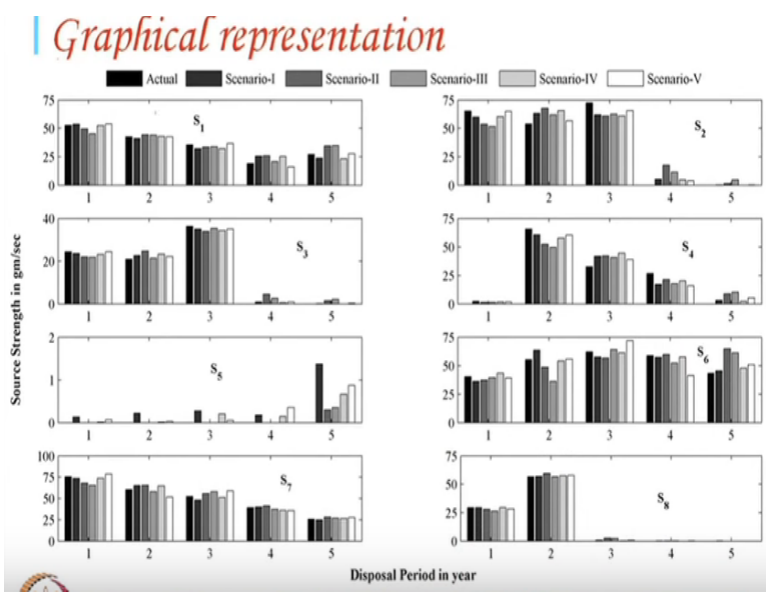
(Refer Slide Time: 31:38)



Disposal Periods	Source Location	Actual	Scenario-I Mean	Scenario-I S.D.	Scenario-II Mean	Scenario-II S.D.	Scenario-III Mean	Scenario-III S.D.	Scenario-IV Mean	Scenario-IV S.D.	Scenario-V Mean	Scenario-V S.D.
1	S_1	52.80	53.74	3.08	49.54	1.25	45.39	2.90	52.44	1.63	53.88	5.52
	S_2	65.29	59.81	5.65	53.76	1.82	51.42	1.93	60.11	2.74	65.08	5.37
	S_3	24.37	23.53	1.92	21.97	0.88	21.77	0.85	23.03	0.66	24.29	1.10
	S_4	0.00	2.26	1.72	1.31	0.98	1.25	1.22	1.64	0.97	1.51	1.45
	S_5	0.00	0.13	0.21	0.00	0.00	0.00	0.00	0.01	0.03	0.07	0.17
	S_6	40.25	36.06	8.14	37.46	1.36	39.40	5.24	43.44	2.35	39.24	3.57
	S_7	75.64	73.50	6.09	67.76	1.08	65.70	6.29	73.55	1.13	78.72	3.71
	S_8	29.44	29.38	2.45	28.06	1.06	26.36	1.55	29.38	0.80	28.35	2.01
2	S_1	42.64	41.06	4.23	44.44	1.28	44.18	5.42	42.90	1.06	42.70	9.84
	S_2	53.92	63.07	8.80	67.58	4.31	62.03	3.00	65.49	5.91	56.55	8.12
	S_3	20.86	22.65	5.00	24.65	1.63	21.35	2.30	23.33	1.42	22.14	3.40
	S_4	65.60	60.53	6.15	52.15	2.23	49.61	2.92	57.63	2.29	60.34	6.23
	S_5	0.00	0.22	0.54	0.00	0.00	0.00	0.00	0.01	0.03	0.03	0.11
	S_6	55.32	63.48	11.96	48.55	1.48	36.12	7.90	54.11	2.62	55.94	8.25
	S_7	60.48	65.15	8.89	65.45	0.91	58.06	11.30	64.86	0.98	51.70	11.29
	S_8	56.75	57.02	3.88	59.59	2.75	56.69	3.33	57.54	1.90	57.95	3.20
3	S_1	35.37	32.05	4.29	33.71	1.92	33.87	3.55	32.11	2.23	36.71	11.71
	S_2	72.45	61.87	9.21	60.88	2.88	62.64	5.07	60.91	6.40	65.47	8.59
	S_3	36.28	34.98	4.75	33.86	2.35	35.24	3.15	34.29	1.78	35.06	3.89
	S_4	32.68	41.77	8.85	42.18	2.04	40.73	5.22	44.39	1.75	38.97	12.01
	S_5	0.00	0.27	0.65	0.00	0.00	0.00	0.00	0.20	0.42	0.05	0.14
	S_6	62.18	57.55	8.14	56.80	0.99	63.99	7.97	61.19	1.94	71.96	16.92
	S_7	52.44	48.07	7.57	55.74	1.55	57.77	12.03	51.23	1.97	58.74	18.29
	S_8	0.00	0.80	1.34	2.58	1.67	2.36	1.80	0.15	0.35	0.62	1.49
4	S_1	18.92	25.66	4.60	25.84	1.59	20.87	6.19	25.31	1.73	15.98	8.31
	S_2	0.00	5.37	4.98	17.59	3.81	11.36	7.14	4.74	3.14	3.62	5.07
	S_3	0.00	0.92	1.80	4.58	2.05	2.63	1.92	0.51	0.65	0.91	1.37
	S_4	26.55	17.16	7.81	21.23	1.44	17.80	4.91	19.98	1.48	16.06	13.55
	S_5	0.00	0.18	0.48	0.00	0.00	0.00	0.00	0.15	0.23	0.35	1.11
	S_6	58.72	57.31	7.65	59.77	1.65	52.09	12.62	57.67	1.91	41.22	20.31
	S_7	39.25	40.12	9.28	41.22	1.02	37.26	8.48	36.06	1.14	35.51	18.63
	S_8	0.00	0.09	0.26	0.01	0.02	0.08	0.22	0.00	0.00	0.01	0.03
5	S_1	27.14	23.81	2.87	34.65	1.22	34.76	3.55	23.22	1.31	27.68	3.78
	S_2	0.00	0.13	0.36	1.27	1.70	4.68	3.85	0.00	0.00	0.18	0.57
	S_3	0.00	0.08	0.20	1.48	0.80	2.15	0.92	0.00	0.00	0.17	0.36
	S_4	0.00	3.12	4.67	8.71	2.38	10.29	4.33	2.26	1.97	5.17	5.93
	S_5	0.00	1.37	2.01	0.30	0.87	0.35	0.97	0.66	1.38	0.87	1.30
	S_6	43.27	45.24	6.57	64.79	0.76	61.22	8.05	47.72	2.52	50.72	8.34
	S_7	25.63	24.79	5.82	28.21	1.05	27.02	3.87	26.61	1.50	27.84	9.48
	S_8	0.00	0.05	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
NEE_r (%)			9.98		16.6		16.57		9.23		9.94	

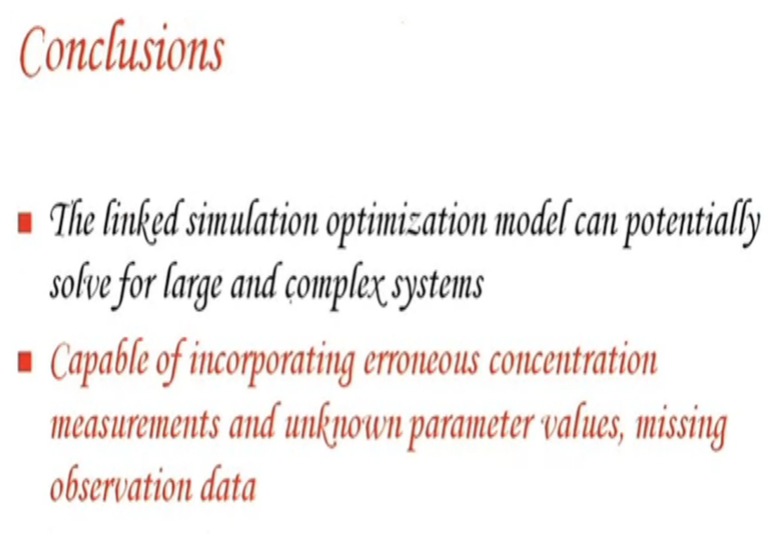
The concentration data during first 10 years are considered to be missing so in this case we can see that for different scenarios the inner percentage is 9.98%.

(Refer Slide Time: 32:00)



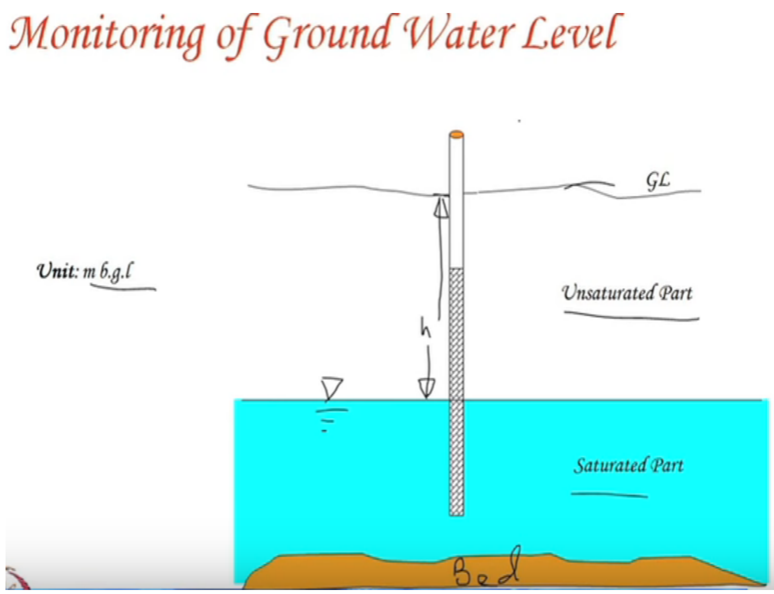
This is 16.68, 9.23, 9.94 %. so graphical representation of different scenarios, this is actual scenario and with actual scenario now things are varying here.

(Refer Slide Time: 32:11)



So conclusion is that the linked simulation optimization model can potentially solve the large and complex system. Capable of incorporating erroneous concentration measurements and unknown parameter values missing observation data.

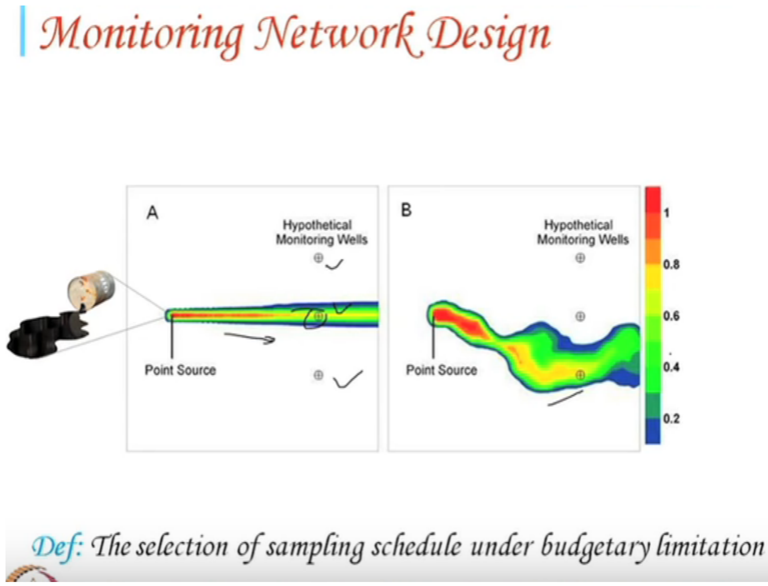
(Refer Slide Time: 32:37)



So monitoring is most important thing for any management problem or source identification problem. Let us say this is our GL value GL and this is unsaturated part of the aquifer, this is saturated part and this is our basically bed rock and unit of and this is our groundwater level in the aquifer. So unit of groundwater level that is generally considered as age is BGL value meter.

BGL is below ground level, this is a typical piezometer and this is the screen or well screen.

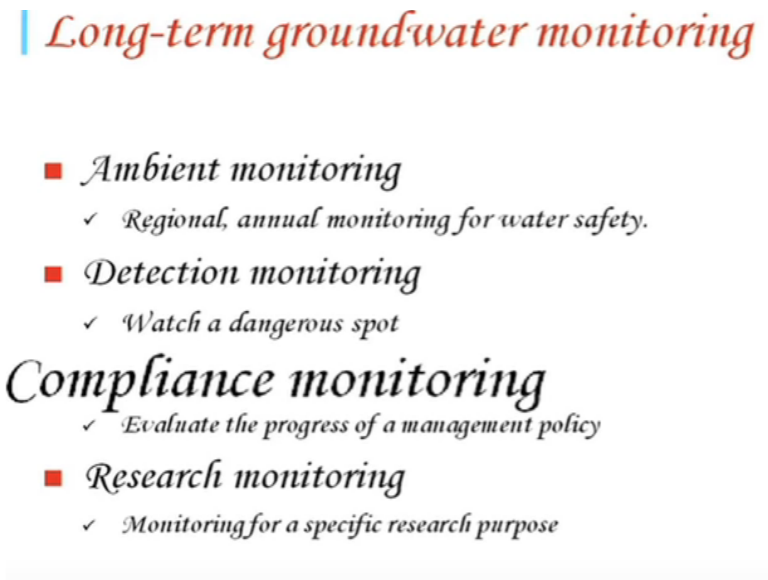
(Refer Slide Time: 34:06)



So with that we can monitor the groundwater level in any area so this is about the water length. What about the contamination? Let us say have a point source and we had three down gradient locations. This is the direction of hydraulic gradient, so ideally this point source should be detected by this particular monitoring well and if we estimate the concentration.

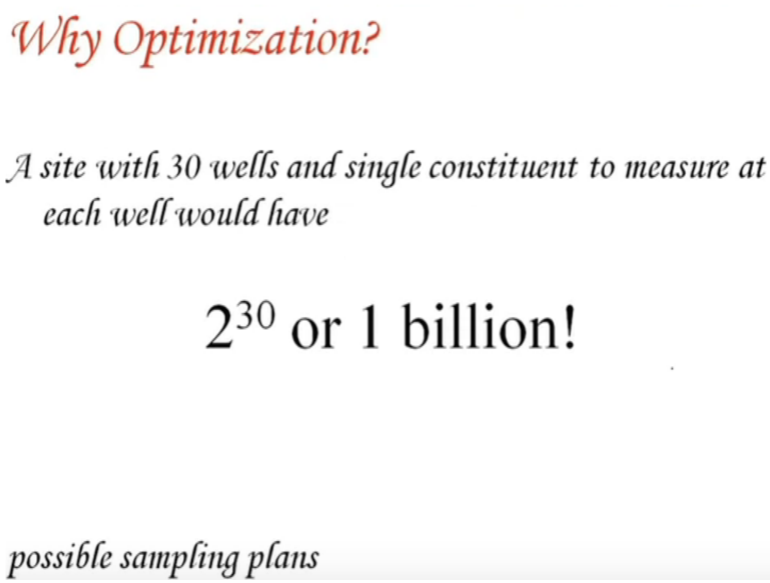
So we can correctly identify that point source but in reality the situation is different or soil is highly heterogeneous and due to that heterogeneity the point source may be detected by the third well which is another well. So the selection of sampling schedule under budgetary limitation. That is one most important thing so monitoring network is basically that finding out that sampling schedule under cost constraints.

(Refer Slide Time: 35:45)



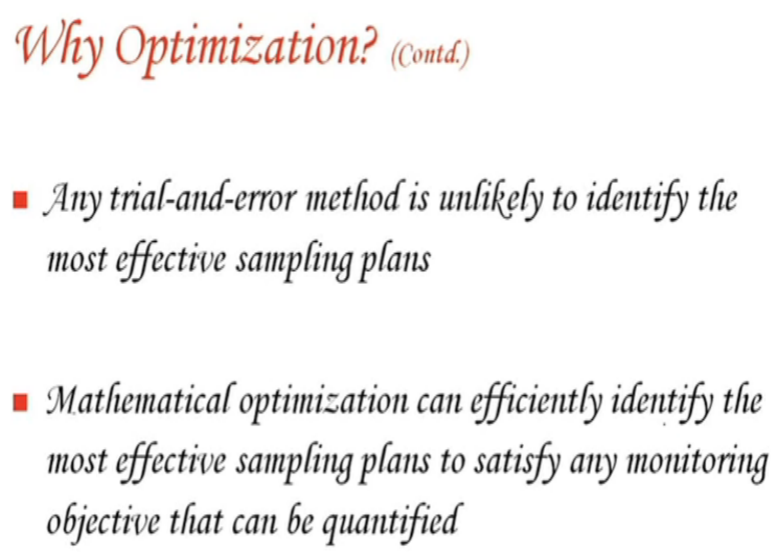
So long term groundwater monitoring is important. So first thing is ambient monitoring which is basically regional annual monitoring for water safety. Detection monitoring watch a dangerous spot for detection. Compliance monitoring evaluate the progress of a management policy or remediation process. Research monitoring is that monitoring for a specific research purpose. So out of that most of the cases compliance monitoring is the important one.

(Refer Slide Time: 36:33)



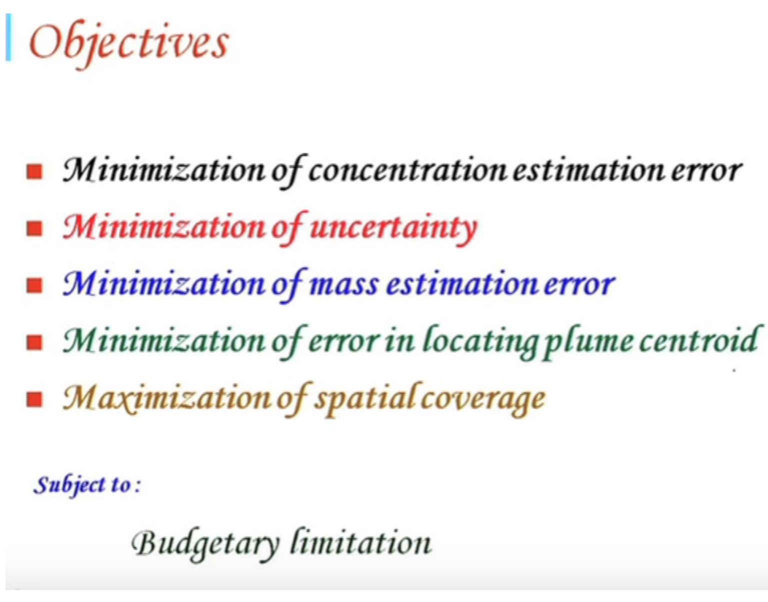
So a site with 30 wells and single constituent single chemical constituent to measure at each well would have 2 to the power 30 or 1 billion possible sampling plans. Either 0 or 1 either 0 or 1 whether to monitor or not to monitor that way we have two to the power thirty solutions.

(Refer Slide Time: 36:59)



So any trial and error method is unlikely to identify the most effective sampling plans. So mathematical optimization can effectively identify the most effective sampling plans to satisfy any monitoring objective that can be quantified.

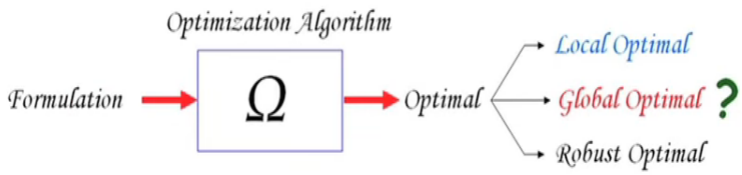
(Refer Slide Time: 37:19)



So objectives for monitoring are minimization of concentration estimation error. Second one is minimization of uncertainty. Third one is mass estimation error minimization of error in locating plume centroid. Maximization of spatial coverage and all are subject to budgetary limitation.

(Refer Slide Time: 37:52)

Basic Approach



$$\mathbf{x} \in \Re^n, \left\| \mathbf{x} - \mathbf{x}^* \right\|_2 \rightarrow 0 \text{ as } N_{iter} \rightarrow \infty$$

Linearity + Convexity + Proper Optimization Algorithm = Guaranteed Global Optimality

So basic approach is that we have the formulation. We have the optimization problem, we can find out the optimal solution at the local global or robust optimal ideally. If we have a large number of idealation then it should reach to this X should reach to the ideal global optimal solution and linearity or convexity of the constraints space. Those are important things and we need to select proper optimization algorithm for guaranteed global optimality.

(Refer Slide time: 38:41)

Spatial Interpolation of Concentration

So spatial interpolation of concentration is important because we can have some kind of monitoring information for re-selected location. For other locations we can get some kind of estimate about the concentration from some kind of interpolation spatial interpolation technique distance.

(Refer Slide Time: 39:19)

Inverse distance weighting (IDW)

$$w_l = \left(\frac{1}{d_{l,\mathbf{x}}}\right)^p$$

$$d_{l,\mathbf{x}} = \|\mathbf{x} - \mathbf{x}_l\|_2$$

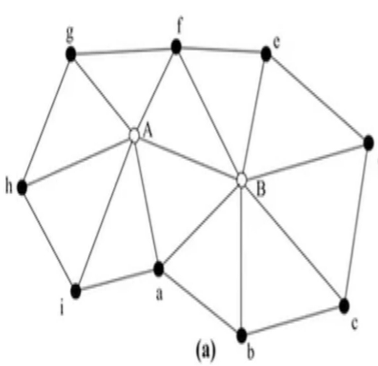
$$\langle f_j(\mathbf{x})\rangle_{est} = \sum_{l\in N_j^b} w_l f_l \Big/ \sum_{l\in N_j^b} w_l$$

Inverse distance weighting or IDW is the most common method where this WL is the weight 1 by WLX. This is the distance between two points that is L and X to the power P. P is the exponent, DL is the distance or this estimated value for any particular attribute or parameter is estimated summation of WL into FL, L is in the neighborhood points of J.

So J is estimating the value for J and NB is the neighborhood set then L is in the neighborhood set of that neighborhood set of J. This is weight, the actual value is again summation of total weight.

(Refer Slide Time: 40:33)

Illustration



$$\langle c_A\rangle_{est} = \frac{\sum_{l\in\{a,f,g,h,i\}} \left(\frac{1}{d_{l,A}}\right)^p \langle c_l\rangle_{act} + \left(\frac{1}{d_{B,A}}\right)^p \langle c_B\rangle_{est}}{\sum_{j\in\{a,B,f,g,h,i\}} \left(\frac{1}{d_{j,A}}\right)^p}$$

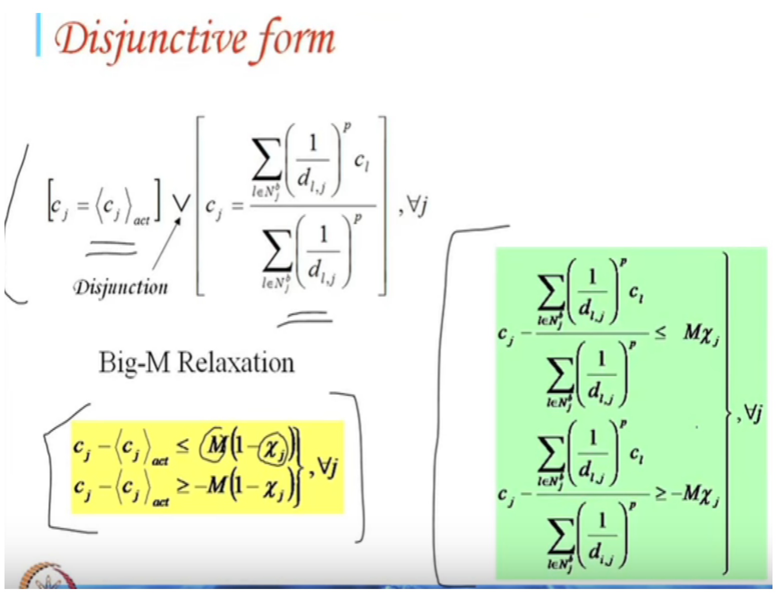
$$\langle c_B\rangle_{est} = \frac{\sum_{l\in\{a,b,c,d,e,f\}} \left(\frac{1}{d_{l,B}}\right)^p \langle c_l\rangle_{act} + \left(\frac{1}{d_{A,B}}\right)^p \langle c_A\rangle_{est}}{\sum_{l\in\{a,b,c,d,e,f,A\}} \left(\frac{1}{d_{l,B}}\right)^p}$$

So one illustrative thing is that let us say we have this configuration where A and B two wells are unmonitored locations. But other locations we have monitored situation, so if we are drawing this triangular neighborhood locations. So for A we will see that E is the neighborhood location for A and in case of B F will be the neighborhood location for B. But the problem is that A distance is far compared to A B.

So that this information of B should be utilized while calculating A and information of A should be utilized while calculating the value of B. So this is estimated value, so we should have some kind of actual values in the neighborhood plus estimated value of B while calculating or estimating the concentration.

That is while calculating or estimating the value at B we need to use the concentration values which will be estimated using our previous equation. So these two equations will act as constraints in our optimization model because these two variables are unknown.

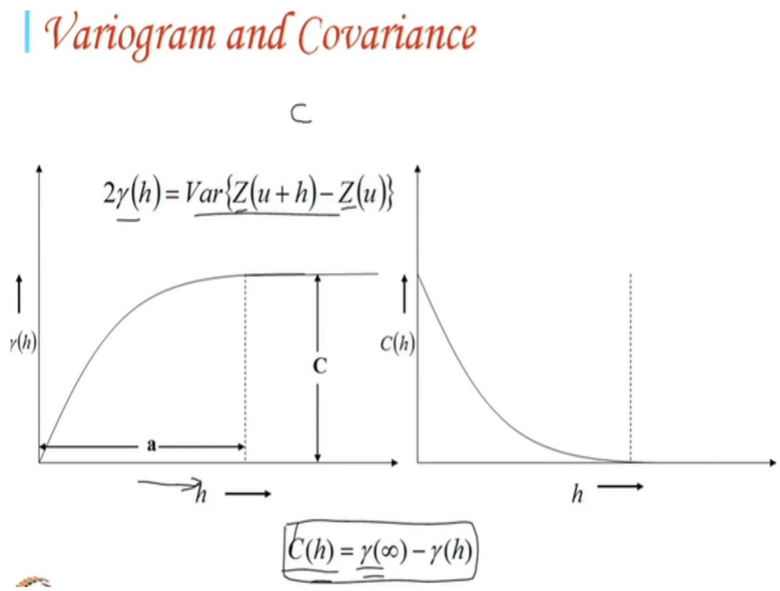
(Refer Slide Time: 42:48)



So we can write it in disjunctive form that is if a well is monitored then we have actual value, otherwise it should be based on the neighborhood concept or we should get the estimation from the neighboring points. If we use big M relaxation, so this is a large value of M and this XI thing has got one. If a particular location is monitored otherwise it is 0.

If particular location is not monitored, so for this particular form of constraints we can use this to get two sets of constraints that is if it is monitored then this goes to one means this is less than equals to 0. And this is again greater than equals to 0, that means ideally they should be both the equation will converged to the equality constraint and the CJ will be calculated based on this actual value. Otherwise it will be calculated from this particular equation.

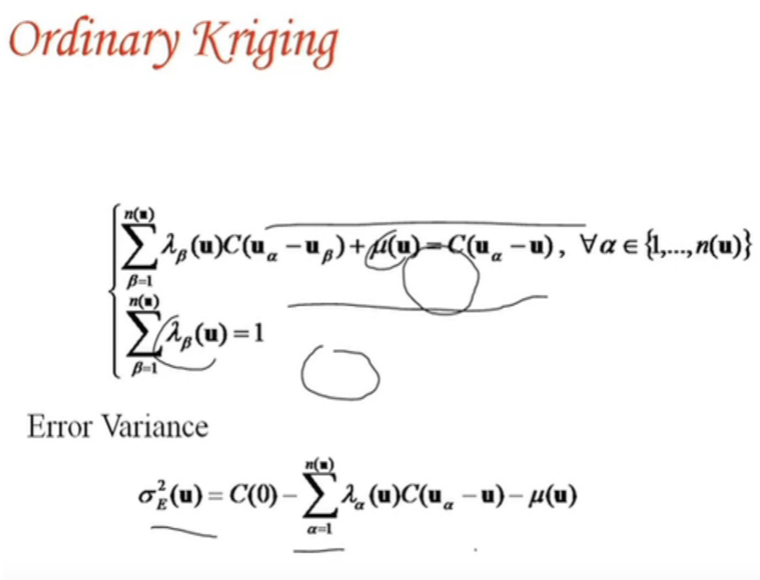
(Refer Slide Time: 44:39)



So variogram is important thing. So it basically gives some kind of idea about the spatial variability of any attribute. So this is the age or this lack distance between two different spatial points. This is the variogram value and variogram is calculated. This variance half of the variance between U + Z U + age and Z.

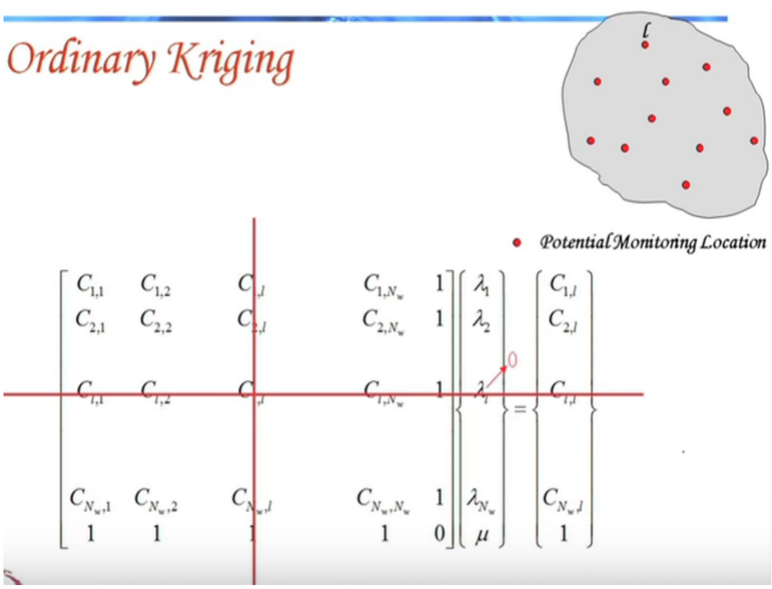
Z is any particular attribute for our case we can calculate concentration value using this approach and this C is covariance and covariance is related to this gamma infinity. This gamma infinity is basically constant value minus gamma age. So we can use this expression for our calculation.

(Refer Slide Time: 45:50)



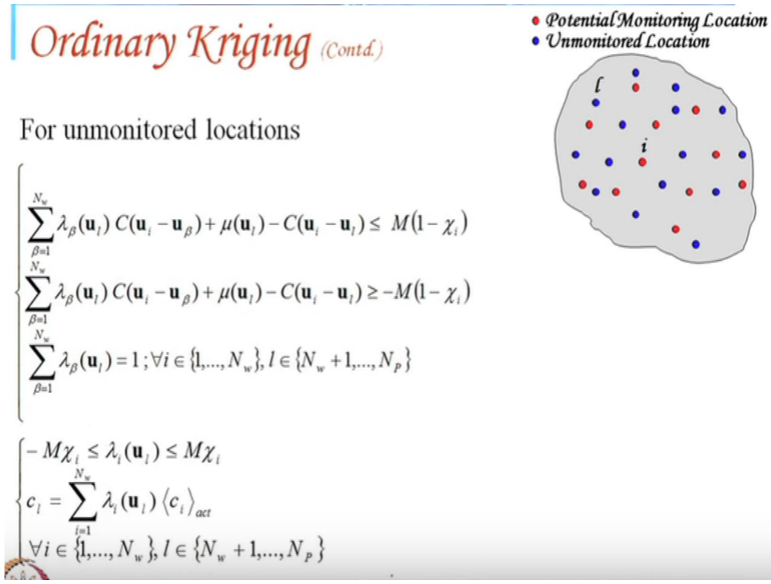
So ordinary kriging is basically minimization the estimation variance subject to our constraints. That is estimation in terms of this weights and should be such that summation of weights should be one. So finally we can get this particular set of equation and we can solve it and we can get the value of lambda and Mu for our particular problem and we can also estimate the variance for any particular problem for ordinary.

(Refer Slide Time: 46:38)



These are the potential monitoring location. So while it is monitored then this part is not required. This is lambda is basically 0.

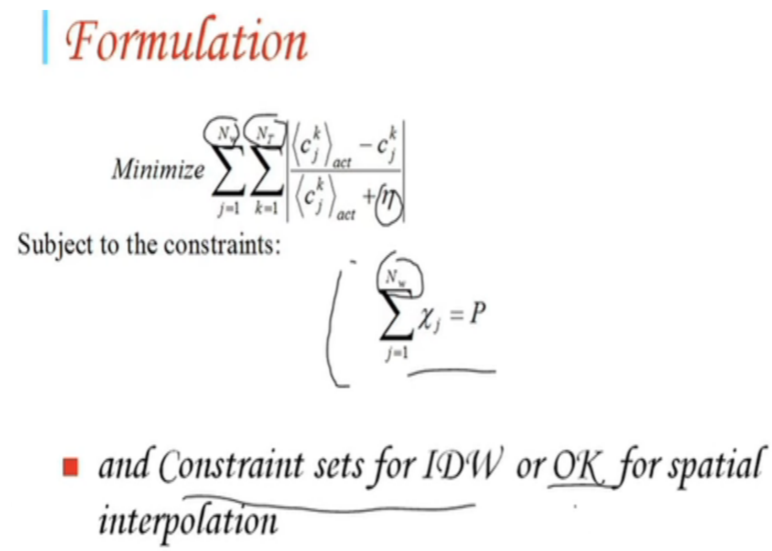
(Refer Slide Time: 47:02)



And based on the thing that potential monitoring location and unmonitored location, we can use the big M method to write our original Kriging equation like this and we can have situations where Kai I and Kai L. This is 1 1, 1 1 means this represents the situation like that particular row or coloumn whether it will be used or not that will be determined by the value of Kai I and Kai L.

And this left hand side this whole thing is represented as the Zai I L. So we can use this as our constraints for our optimization problem and we can finally get these constraints. So these are basically special constraints in case of IDW. We have seen we have got set of special constraints these are special constraints for our optimization problem related to groundwater monitoring.

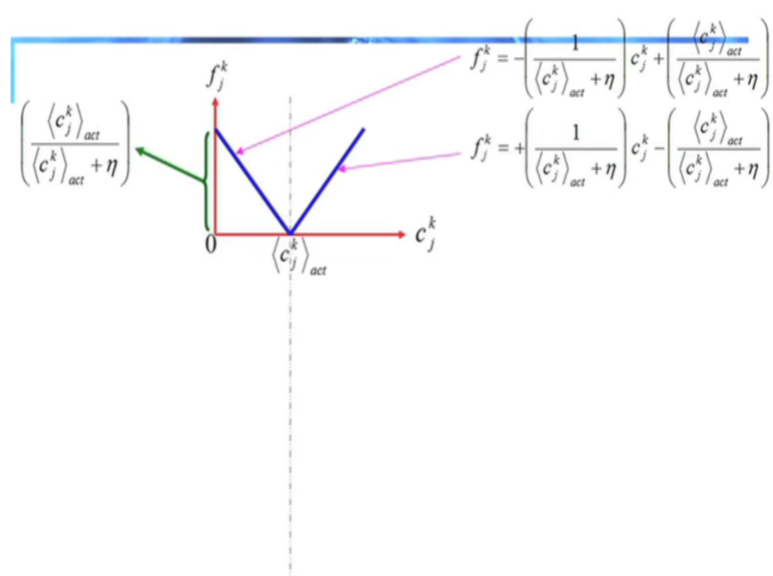
(Refer Slide Time: 48:35)



So now the formulation is basically, this is actual value. This is estimated one and actual value divided plus this is ETA. ETA is some number which varies between 0 to1 and this is number of wells and number of time periods. So at the end of any particular time period and for a particular well.

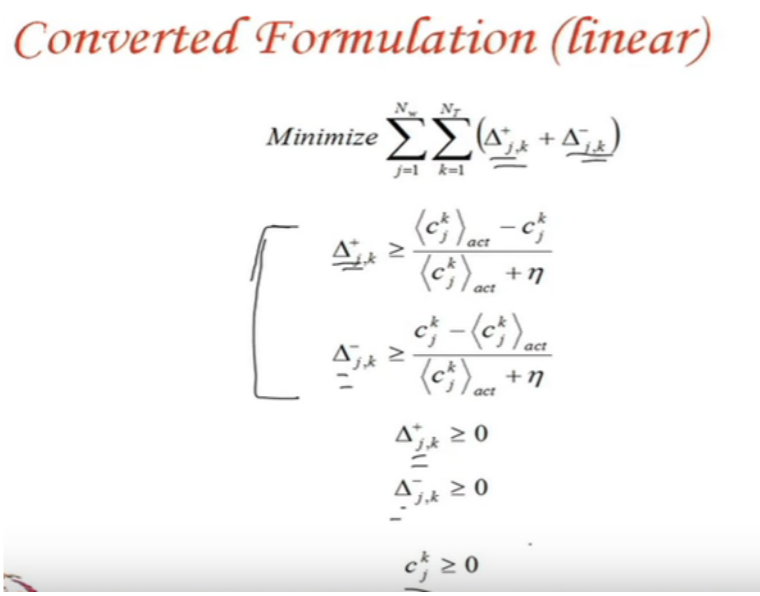
This is valid and we need to minimize the total deviation and this is basically normalized deviation and absolute of normalized deviation and this is the cost constraint. We can install only P number of monitoring wells. Out of this NW possible potential monitoring wells and we have that IDW or ordinary kriging for spatial interpolation.

(Refer Slide Time: 49:46)



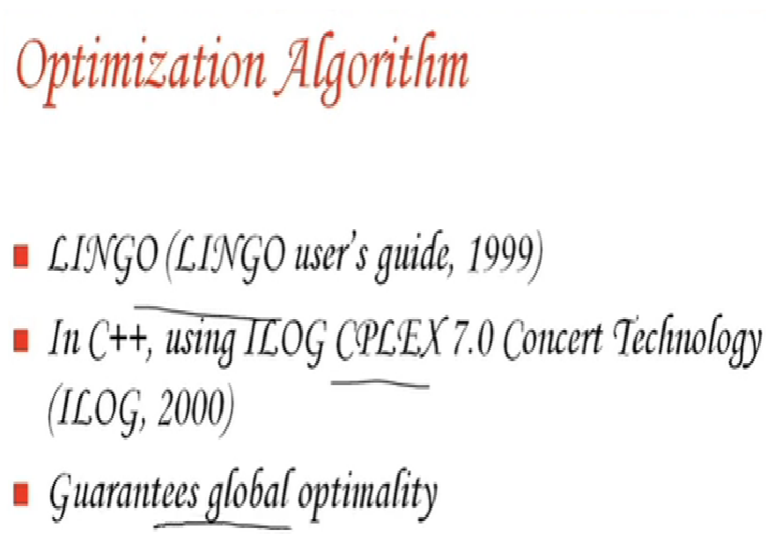
So if we see our objective function, this is having absolute updated and it is like this blue line and if we differentiate, it will be discontinuous function.

(Refer Slide Time: 50:05)



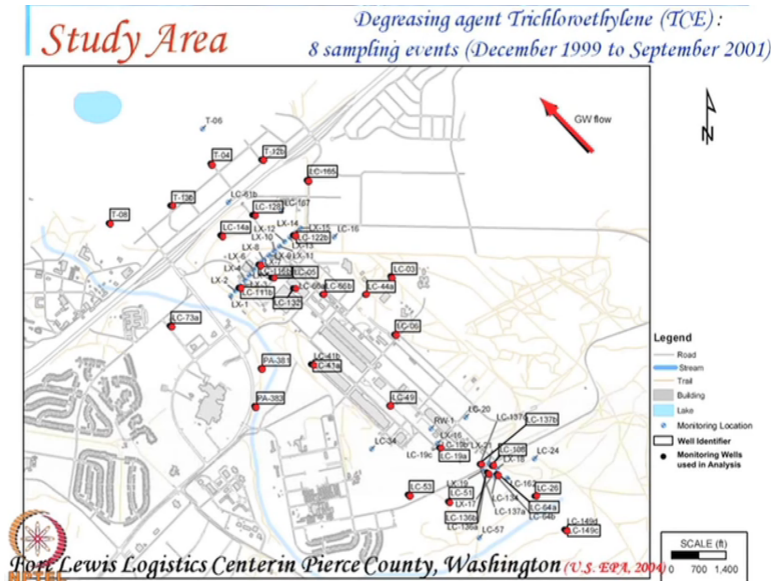
So it is better if we convert it into linear formulation, we have found out a linear equivalent thing of that. So this is the linear equivalent thing and positive and negative. So both are positive and we can minimize this if this is 0 that means, we have got 0 estimation error and this is basically to balance or to calculate the absolute operates and CJ, it should have a proper value which should be greater than equal to 0.

(Refer Slide Time: 50:55)



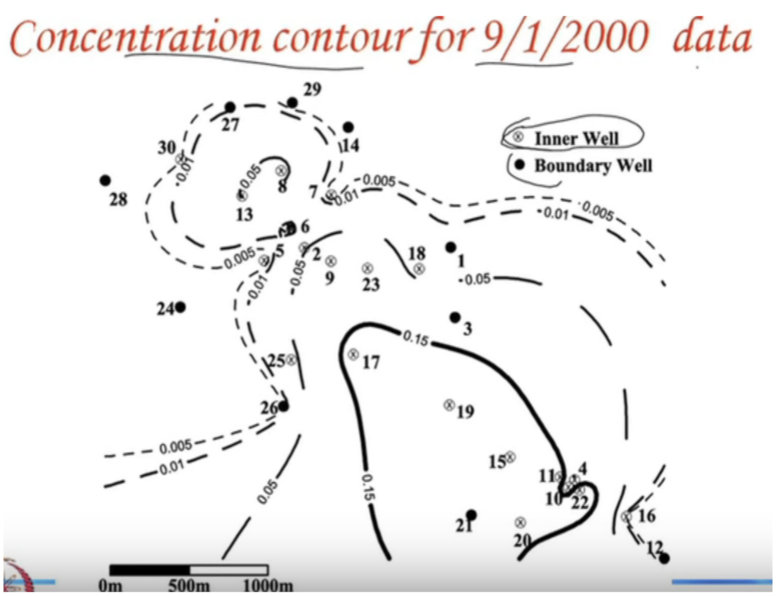
This can be solved in any algorithm for optimization algorithm. LINGO or CIPLEX, if you have linear programming then it will give guarantees global optimality.

(Refer Slide Time: 51:12)



So for a particular study area decreasing agent trichloroethylene TCE, eight sampling events from December 1999 to 2001. This methodology was just a fort Lewis logistics ceneterin pierce country Washington and for this.

(Refer Slide Time: 51:49)



On the boundary wells and these are the inner wells. So for this is the concentration contour for 2000 data and this a scale we can see that we have these many wells and out of that.

(Refer Slide Time: 52:13)

Scenarios

- *IDW-Scenario-I: Only for Sept, 2000*
- *IDW-Scenario-II: Only for Sept, 2000 with boundary well restriction*
- *IDW-Scenario-III: All 8 time periods*
- *IDW-Scenario-IV: All 8 time periods with boundary well restriction*
- *OK-Scenario-V: Only for Sept, 2000*

$$\gamma(h)=\begin{cases} C\left[1.5\left(\frac{h}{a}\right)-0.5\left(\frac{h}{a}\right)^{3}\right] & \text{if } h\le a \\ C & \text{if } h>a \end{cases}$$

We have IDW scenario where only September data 2000 was used. ID two scenario only September 2000 data with boundary well restriction was used, that all boundary wells should be selected and IDW scenario 3 that is all 8 time period data was used and scenario 4 all 8 time periods with boundary well restriction was used and ordinary kriging scenario 5 only for September data was used.

(Refer Slide Time: 52:52)

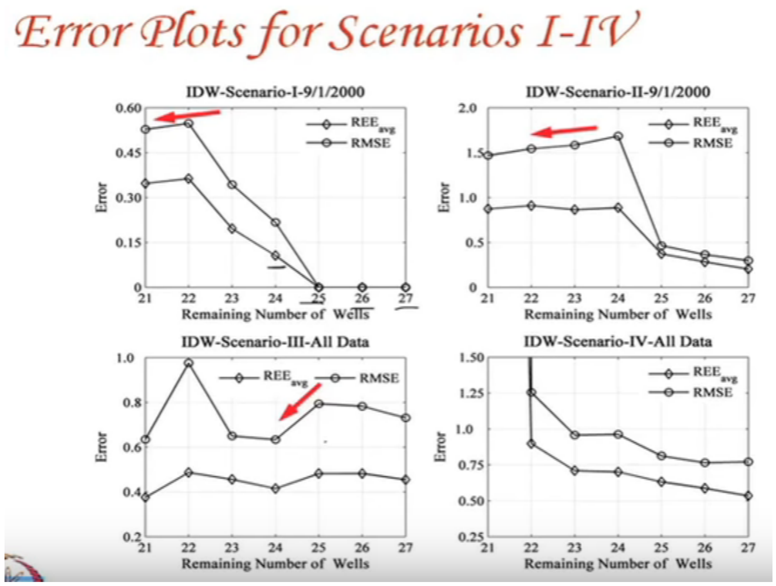
Performance Measures

$$REE_{avg}=\frac{1}{N_r N_T}\sum_{j=1}^{N_r}\sum_{k=1}^{N_T}\frac{\left|c_j^k-\left\langle c_j^k\right\rangle_{act}\right|}{\min\left(c_j^k,\left\langle c_j^k\right\rangle_{act}\right)}$$

$$RMSE=\sqrt{\frac{1}{N_r N_T}\sum_{j=1}^{N_r}\sum_{k=1}^{N_T}\left(\frac{c_j^k-\left\langle c_j^k\right\rangle_{act}}{\min\left(c_j^k,\left\langle c_j^k\right\rangle_{act}\right)}\right)^2}$$

So in this case the performance measures where this particular relative estimation error that is average value based on number of wells removed or eliminated and number of time periods same here. This is RMSE or route mean square error difference between the actual value and this is the estimated one.

(Refer Slide Time: 53:26)



And error plots shows some interesting results, so interestingly if we have out of these many wells if we have only 27, 26 or 25 wells. The error is almost negligible for scenario one but if we remove more number of wells. That is six wells if we remove then it will be a problem.

Out of this wells if we remove more number of wells then it will be a problem and interestingly if we remove more number of wells again this error is decreasing. So we can infer that not only number of wells, also configuration is important for any monitoring network. So particular configuration can give lesser error compared to a monitoring network we have more number of wells.

(Refer Slide Time: 54:49)

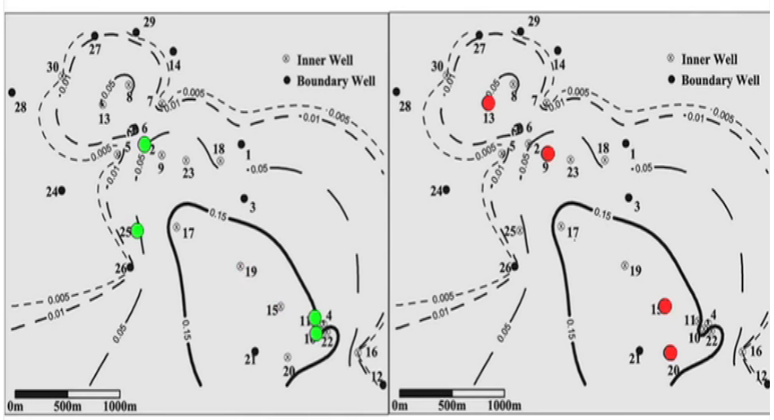
Comparison of Errors

P	% reduction in Wells	IDW-Scenario-I		IDW-Scenario-II		OK-Scenario-V		Li and Hilton (2005)	
		REE_{avg}	RMSE	REE_{avg}	RMSE	REE_{avg}	RMSE	REE_{avg}	RMSE
27	10 %	0.00000	0.00000	0.20431	0.29743	0.11425	0.12389	0.361	0.383
26	13 %	0.00000	0.00000	0.28272	0.36526	0.12856	0.13736	0.578	0.595
25	17 %	0.00000	0.00000	0.37332	0.46367	0.21662	0.27779	0.523	0.559
24	20 %	0.10625	0.21701	0.88695	1.68459	0.28371	0.33667	0.515	0.545
23	23 %	0.19617	0.34306	0.86535	1.58423	0.32319	0.36643	0.562	0.637
22	27 %	0.36316	0.54784	0.91040	1.54397	-	-	0.649	0.725
21	30 %	0.34730	0.52792	0.87409	1.46862	-	-	0.862	1.165

So IDW scenario have compared this with the existing solutions and this is our IDW scenario one is far better compare to existing results. Also for up to seventeen percent reduction in wells this is performing better as I have already told that configuration of monitoring wells. That is also important thing here and for ordinary krigings scenario it out performs the existing results.

(Refer Slide Time: 55:45)

Removal locations



Li and Hilton (2005) OK-Scenario-V

So these are the removal location. For both the cases we have found out the different configuration, that is why our results are better.

(Refer Slide Time: 55:56)

Number of variables

Scenario	Number of Variables	
	Real	Integer
IDW-Scenario-I	$3N_w$ =90	N_w =30
IDW-Scenario-II	$3N_w - N_B$ =80	$N_w - N_B$=20
IDW-Scenario-III	$3N_w N_T$ =720	N_w =30
IDW-Scenario-IV	$(3N_w - N_B)N_T$=640	$N_w - N_B$=20
OK-Scenario-V	$N_w(N_w + 4)$ =1020	N_w =30

So number of variables it is giving a guaranteed optimal solution because our optimization problem is linear in nature and our approach is also proper. So that is why it is giving guaranteed global optimal solution. So you can see that ordinary kriging scenario have 1020 and N W that is 30 integer variables in last case. We have 1 0, 2 0 real variables and 30 integer variables.

So this is all about optimization thing and in optimization basically we have covered this monitoring network design part and now we can see that what is the importance of optimization in monitoring network design methodology so with this lecture 37 ends.