STATE OF THE ART IN GROUNDWATER MONITORING NETWORK

This chapter introduces general groundwater monitoring problem in section 2.1. The section 2.2 reviews in brief the earlier work on spatial sampling analysis i.e., to determine the number and location of sampling points and for temporal sampling analysis i.e., to determine sampling frequency.

2.1 Groundwater monitoring - a desire or a necessity

In areas where groundwater for consumption is being abstracted, it is usually stipulated that continuous observations regarding the groundwater quantity and quality must be made. Observations and evaluation of the observations are costly, so it is of vital importance to make the observation program as optimal as possible. Optimization of an observation network means that the total number of observations is as minimal as possible, while the objectives of the monitoring are fulfilled. Before a groundwater resource (aquifer) can be used for abstraction, the geologic formation from which the groundwater is supposed to be abstracted must be investigated. Relevant questions to be answered are:

- Is this geologic formation valuable as a groundwater resource in the area of interest?
- Which are the potential areas for locating abstraction wells and infiltration basins?
- What are the best or optimal rates of abstraction or the artificial recharge?

To answer these questions a number of boreholes are established to investigate the geologic and hydrogeologic properties of the geologic formation. In these boreholes, tubes or pipes for observations of groundwater levels and groundwater quality are installed. It is also possible to use dug wells for these observations. Thus an observation network is established in
the formation, which can be used for continuous observations of groundwater quantity and quality.

The observations are performed as point measurements, but since the observed variable almost always has a distribution in space there is therefore need, during the evaluation of the observations, to make estimates of the observed variable at locations, where no observations have been made for example when mapping a variable over an area. All estimates however have a certain degree of uncertainty which is important to estimate. One way to determine the quality of an estimate is to compare the estimated value with a measured value at the same point. The difference between the estimated and the measured value is an indicator of how good the estimate is. By repeating this procedure for a great number of points it is possible to determine the ability of an estimation method. When observed data are sparse such a procedure is not possible to follow, and there is a need for using some interpolation method that can produce a measure of the estimation accuracy without measured data to compare with.

In some cases, the total annual number of groundwater observations may vary over time. Initially the total number of observation points in the network is kept maximum. After some time, perhaps several years, the observation program usually must be reduced. The reason for such a reduction could be:
- wells or tubes have dried out.
- tubes have disappeared and/or are difficult to find.
- impossible to perform measurements because of sabotage to the tube.
- less need for additional information.
- budget restrictions.
- lack of competent personal to perform and evaluate investigations.

*Certain evaluation is performed in order to reduce the number of observation points according to the reasons previously mentioned. Removing*
one or several observation points at a time from the network would reduce the observation program. It is important to decide which points are to be removed without losing too much information. A number of problems have thus been identified, and as far as estimation is concerned there are two issues that often must be addressed, these are:

- Which interpolation (estimation) method is the best regarding the monitoring objectives and considering the fact that, in the case of an existing observation network, that the number of observed data are sparse?
- Are there any interpolation methods which are able to produce some measure of the estimation accuracy together with the interpolated (estimated) value?

Concerning the problem of network design the following four issues, partially overlapping each other need to be addressed:

- How many observation points are needed, and which are the preferable spatial locations of these points to meet the monitoring objectives?
- How to determine an observation frequency to meet the monitoring objectives.
- How to define a simple criterion, for example the estimation accuracy, to be used when an observation program is to be reduced, or when it is developed.
- How to rank the different observation points in a network and to use this ranking when an observation network is being reduced i.e. points with the lowest rank are to be deleted from the network.

India is a vast country having diversity in almost all the field, viz., geology, hydrogeology, climate, etc. Although India occupies only 2.4% area of the worlds land, but supports over 17.5% of the world’s population i.e., about 1.13 billion peoples (estimate for March 10, 2008). (http://en.wikipedia.org/wiki/Demographics_of_India). Since India is stepping
towards the development, the pressure on the available water resources is increasing exponentially as the consumption of water per person is increasing and further the development of new technologies for exploration of water for all purposes added fuel to the fire. Available surface water resources have either become scanty or else polluted due to improper management, thus adding whole pressure on the groundwater resources. Now a day, groundwater had become a major source of water supply in various sectors and is getting depleted in these regions due to over-exploitation, because of high population growth as well as extensive agricultural uses and thus urgently requires a reliable management option. It is essential to have a thorough understanding of complex processes viz., physical, chemical and/or biological occurring in the system for groundwater assessment and management. To understand these complexities, groundwater models play an important role. Groundwater models are simplified, conceptual representations of a part of the hydrologic cycle. They are primarily used for hydrologic prediction (hydraulic head, flow rates and solute concentration) and for understanding hydrologic processes (contaminant migration, solute transport etc.).

However, in groundwater models, the area is discretised into finer grids or blocks, and the size of these grids or blocks are decided assuming that there is no variation in aquifer properties over the grid or block. It is a proved fact that finer the grid, the faster the iterations and the better the results. Although it is needed to have an observation at each grid or block to be fed as an input in the model, but practically it is not possible to have an observation at each grid or block. Further in countries like India which are not only vast in size but developing as well, the monitoring is decided on the basis of budget and other technical and non technical constraints. Thus there is always a need of an optimal monitoring network which will be able to reproduce the true variability of the parameter.
2.2 Groundwater Monitoring Network Design

The design of a groundwater monitoring network is a multi-objective design problem (Gangopadhyay et al., 2001). According to the design purpose of groundwater monitoring, the monitoring of groundwater can be separated into the groundwater quality-monitoring network and groundwater level-monitoring network (Loaiciga, 1989; Loaiciga et al., 1992; Wu, 1992; Zidek et al., 2000). Processing the information on the groundwater level and the solute concentration can provide an answer to all issues pertaining to groundwater management. Although, substantial literature has been generated through years to identify spatio-temporal attributes to wells that satisfy these objectives, however, the focus has been on the design of a groundwater quality monitoring network. The philosophy behind the design of a groundwater level observation is essentially the same as that for a groundwater quality monitoring network. Loaiciga et al., (1992) categorized the general approaches to network design as hydrogeologic or qualitative, when no advanced geostatistical method is applied and statistical or quantitative otherwise. This classification is somewhat limiting, because it excludes promising advanced statistical techniques such as those derived from information theory (Harmancioglu et al., 1998). The examples of optimizing the monitoring network design, based on information theory are in articles such as those by Amorocho and Espildora (1973), Caselton and Husain (1980), Caselton and Zidek (1984), Harmancioglu and Yevjevich (1987), Husain (1989), and Harmancioglu and Alpaslan (1992).

2.2.1 Hydrogeologic or Qualitative Approach

As the name suggests, this approach is purely based on the judgment of quantitative and qualitative hydrogeologic information, without the use of a formal quantitative mathematical method (Loaiciga et al., 1992). The number and locations of sampling sites are strictly determined by the hydrogeologic conditions. In general, the factors that may be considered in qualitative
approach includes geology, geomorphology, climatology, structural variations, and accessibility and feasibility. The hydrogeologic approach is better suited for site-specific studies or where there is a requirement of following some kind of guideline. To be able to use this method, the investigator needs to have enough site information to establish a good understanding of the groundwater system and how it influences the movement of contaminants and the resulting contaminant distribution. The investigator also needs to have the knowledge or a guideline on what constitutes an “optimal” monitoring-well configuration for a given probable contaminant migration pathway. Although this method is adequate, it is not necessarily to be optimal.

2.2.2 Quantitative Approach

The quantitative approaches used for determining the groundwater monitoring network design can be broadly grouped into three different types viz., Simulation approach, variance-based approach, and probability-based techniques. The main difference between these methods lies in the formulation of the objective function to be optimized, and the difference between different approaches to the constraints and in the way optimization is undertaken.

Simulation methods consider uncertainty in the hydraulic conductivity field and therefore uncertain head distribution and velocity fields. In this method realization of hydraulic conductivity, which is treated as a Random field are generated and used as an input for groundwater flow and transport simulation. Hydraulic conductivity fields are usually estimated by conditional simulations, given a spatial covariance model obtained with the existing field data. The resulting differences in the estimated velocity fields will affect the optimal distribution of monitoring stations according to some objective function (e.g., depending on the probability of failing to detect contamination). Meyer and Downey (1988) proposed a method for determining the best location for the monitoring wells, after the work of Massmann and Freeze (1987a & b) in a
risk-cost -benefit analysis for waste facilities. The method's intention was to select the networks that maximize the probability of detection in the face of uncertainty. However, its practical applicability is hindered by the extreme simplicity of the analytical model used by Meyer and Downey (1998). Other examples of this method are given by Ahlfeld and Pinder (1988); Meyer and Brill (1988); Wagner and Gorelick (1989); Meyer et al., (1989); Lee and Kitadinis (1991); Shafike et al., (1992); Woldt and Bogardi (1992); Tiedman and Gorelick (1993) and Reed et al., (2000). Grabow et al., (1993) proposed a method for network reduction without the need to simulate mass transport, stated as being applicable for both point and diffuse sources, through only used by the authors for a point source. Simulation approach involves two main computations. The first one is the simulation of the hydraulic conductivity field and the other one is the solution of the groundwater flow and transport problem. The computational effort required is proportional to the number of realizations needed to correctly represent statistically the flow and concentration field. So the selection of a random field generator directly affects the effectiveness of this method, and thus the attractiveness of the method. Application of neural networks and genetic algorithm optimization integrated with flow and transport simulation models are found in Rogers and Dowla (1994) and Mckinney and Lin (1994). A review of optimization and decision analysis for aquifer restoration and contaminant migration-control through pump-and-treat was made in Freeze and Gorelick (1999). These methods also have high potential in the design of regional or local networks for reference level monitoring.

Variance-based method is a group of techniques using statistical properties of the estimated value mainly variance to quantify the uncertainties associated with groundwater system. Variance reduction techniques use the variance of the estimation error as an indicator of the accuracy of the estimated values. In geostatistics, the mathematical definition of variance of estimation error means that its value does not depend on the actual values of the measured variables, but on the relative spatial distribution of the
measuring locations. Therefore, one may use variance of estimation error as an indicator of which spatial distribution is best for a sampling network by testing all the combinations between available sampling locations and selecting the combination that minimizes variance of estimation error. Two often used methods are the geostatistical method and the Kalman filter. The geostatistical method offers variance estimation algorithms, so it is used to give a prior estimate of the covariance (or variance) of the interested variable. The Kalman filter is an optimal estimator that combines measurement with existing system information. So it is used to update the covariance after samples are taken.

The variance reduction approach (Rouhani, 1985), is an extension of kriging technique. Kriging provides a possible spatial interpolation technique for time-independent variables. The values of the variable under consideration at a location are estimated from a local weighted average of the observations. The weights are determined in such a way that the estimation is optimal in the sense of the least estimation error variance. In addition to the spatial estimate of the variable, the variance of the corresponding estimation error is also calculated. The estimation error associated with values interpolated from a set of measured hydrogeological variable values can be used in network design to reduce the uncertainty of the interpolated values. Two types of methods have been suggested in which interpolation error serves as a measure of network performance. In first type, which is a trial and error method, those sampling site are added that contributes most to the reduction of the variance of estimation error of the variable of interest, associated with the set of established sampling locations. Additional sampling sites are added, one at a time, until the variance of estimation cannot be further reduced, or when the additional gain in statistical accuracy is outweighed by other constraints e.g., economic (Journel and Huijbregts., 1978; Marsily., 1986). It is also possible to do ranking of potential sites using an information ranking function, based on the largest variance reduction. Few authors have used multivariate techniques of principal component analysis (PCA) for ranking of wells. PCA was used in
surface water hydrology to identify the important geomorphological parameters that contribute to runoff from a catchment (Haan, 1977). Gangapadhyay et al., (2001) proposed a method to evaluate a groundwater level monitoring network using PCA to discriminate against the value of information collected from monitoring network. The aim of the work was that during the budget constraints, municipality managers can prioritize the sampling from the monitoring network.

Delhomme (1978) applied the geostatistical fictitious point method (usually used to assess the quality of covariance models when estimating with kriging) to determine the optimal location of rain gauges. If the number of stations is large, then the dimension of the combinatorial problem may be exhaustively intractable. Olea (1984) presented a method to select the best pattern and sample density that would meet a specified average standard error or maximum standard error of estimation of standard variables. The method was based on a nearest-neighbor analysis of two dimensional point distributions of spatial variables. Rouhani and Hall (1988) proposed the incorporation of risk, defined as the weighted sum of the expected value and the estimation variance, in order in order to correct the blindness of the estimation variance to extreme values. The method also considers temporal changes in the hydrologic variables. Loaiciga (1989) also proposed a variance reduction method using time-dependent spatial models (based on space-time means and covariances), with good results. This writer used the mixed integer programming model of Hsu and Yeh (1989), originally developed for optimum design for parameter identification. Maximum periodicity was allowed in this study, but did not include trade-off analysis between sampling periodicity and further increasing the number of stations. Later developments in the solution of time-dependent models were proposed by Pardo-Iguzquiza (1998), with the inclusion of the climatological variogram (Bastin et al., 1984; Lebel et al., 1987).
Hudak and Loaiciga (1992) presented a heuristic approach for groundwater monitoring network augmentation. In this augmentation problem, monitoring wells were added to a pre-existing network for the purpose of contaminant plume characterization. Information on aquifer properties and contaminant concentrations obtained from pre-existing wells, and the locations of these wells, were considered in the augmentation process. The flow domain was discretized into a network of potential sampling sites. Each site was assigned a weight which was equal to the contaminant concentration calculated by a numerical mass transport model. The objective was to maximize the total coverage of sites. If a site was within a certain distance of a well, it was covered by this well. The main idea of this method was to place more wells in high-concentration areas. The solution was obtained by running the LINDO (Linear, Interactive, and Discrete Optimizer) mathematical programming package. This method was easy to understand, and easy to implement.

Later Hudak and Loaiciga (1993) extended their work to multilayered groundwater flow system monitoring-network design. The primary objective of the selection process was preferential location of monitoring wells at points of high contamination susceptibility. The way of assigning nodal weight was different than the previous work. The weights, instead of being determined from existing contaminant concentrations, were derived by evaluating the location of a site relative to the contaminant source and the likely contaminant migration pathways. Integer programming techniques (such as branch and bound) were used to solve the problem.

The second type is an optimization method in which estimation error at a point or set of points is the objective to be minimized, and potential measurement locations are the decision variables (Bogardi et al., 1985; Carrera and Szidarovzsky., 1985; Knopman and Voss., 1989; Solomatine., 1999; Hsu and Cheng., 2000; Pinder et al., 2002; Reed., 2002; Dorn and Ranjithan., 2003). The Kriging technique, without the flow equations and time
dimension but with a polynomial approximation of the drift, has been applied in the design of monitoring networks for groundwater, such as for optimum selection of sites for monitoring groundwater level (Carrera et al., 1984; Rouhani., 1985; Spruill and Candela., 1990; Prakash and Singh., 2000). Gao et al., (1996) presented a simple algorithm to fast compute the revised kriging estimation variance when new sample locations are added. Ahmed (2004) has evolved the optimal monitoring network for air temperature firstly by removing the redundant points and the added new measurement points were the variance of estimation error was high. Kumar et al., (2005) have used universal kriging for water level monitoring network and have found that the same accuracy in terms of variance of estimation error can be achieved with the reduction of 12% of monitoring well. Devi et al., (2006) have designed the monitoring network for water level using ordinary kriging in ghatiya watershed Madhya Pradesh, India. Ahmed et al., (2007) used ordinary kriging to reduce the number of water level monitoring observation based of pre-decided limits for the variance of estimation error. They also used the result of cross-validation test to rank the observation wells for fluoride contamination, so that the wells can be chosen according to the rank of priority under budget constraint.

The above-mentioned methods all attempt to minimize hydrogeological variable estimation error solely on the basis of the interpolation procedure used for estimation. Wood and McLaughlin (1984) combined groundwater simulation and Kriging to reduce the size of an existing measurement network. Modeling errors, the difference between measured and predicted state variables, were kriged to obtain modeling errors and uncertainty estimates at unmeasured locations.

Agnihotri and Ahmed (1997) analyzed the ambiguities in the data collection network design mainly based on the geostatistical estimation variance reduction method using a few numerical examples. They have observed that the network designs based on kriging variance reduction
Chapter 2 – State of the Art in Groundwater Monitoring Network

approach is having an ambiguity as the maximum value allowed for the variance or standard deviation of estimation error is decided on an ad-hoc basis. In the absence of an objective function directly involving the location of measurement points, it is difficult to minimize the variance of the estimation error. Either this value is arbitrarily chosen or optimization of a data collection network may be terminated if the corresponding change in variance or standard deviation of estimation error is negligible. Further they have proved with few case studies that the method of optimizing measurement points by reducing estimation variance at the central point of the considered area has an ambiguity that the same network will be valid for an even much larger area, if the center remains unchanged. Also kriging variance calculated for an estimation over the entire area/block increases if the area/block is increased keeping the center of the area unchanged for the same network. However, a few ambiguities were found while considering an area whose center point was also shifted with the origin. Further, the method of estimating the mean of the parameter by ordinary kriging was applied to calculate the kriging variance for different networks. Although the ambiguity that the same network representing different areas remains, this method provides a comparatively lesser kriging variance.

Graham and McLaughlin (1986) presented an algorithm to determine the sampling locations for characterizing a contaminant plume. The variance of concentration estimation errors was obtained by co-kriging measurements of hydraulic conductivity, hydraulic heads and contaminant concentration. New wells were chosen at the sites to produce the largest reduction in the variance of the estimation error.

The use of the Kalman filter for an evaluation of the uncertainty in hydraulic head, and its application to field problems, can be found in Van Geer and Van der Kloet (1986); Van Geer et al., (1991); and Zhou et al., (1991); Wu (1992); Andricevic (1993) used the Kalman filter to sequentially adjust the withdrawal rates in a groundwater supply model by accounting for
the observed field information. Another method is the coupling of the Kalman filter and the finite-difference method. The alternative network with the standard deviation of estimation error approximated a given threshold value at each interpolated location is accepted as the optimal network. But, the selection of threshold value of the standard deviation of estimation error is not described in previously published literature.

The reduction of the dimension of an existing monitoring network is an optimization combinatorial problem, which subset of stations should be kept in the new design. When the number of stations is large, the number of possible combinations is too great for all possible combinations to be tested, even in very fast computers. A common practical decision is to accept one good solution even if it is not the optimum one. There are some heuristic optimization algorithms that can deal with combinatorial problems. First there is a general class of strictly descending algorithms that includes sequential exchange with node-swap or with node substitution, downhill simplex, and search with multiple randomly generated starting solutions. A second general class of algorithms, not strictly descending, includes simulated annealing and tabu search, and combinations of them. Other algorithms based on nature, like genetic algorithms and ant colony optimization are also showing very good results (Cieniawski et al., 1995; Dorigo et al., 1996). Most of these methods have been applied to monitoring network design with varying degrees of success. However, simulated annealing is the technique that has been most frequently applied to monitoring network design, probably because it is the oldest of the nonstrictly descending methods and has shown to be very efficient (Lee and Ellis, 1996).

Kuo (1993) applied a Kalman filter combined with a groundwater finite element model to obtain a contaminant concentration profile. Simulated annealing was used to solve the optimization problem. Other related example is described in work of Wu and Bian (2003). Nunes et al., (2003) developed a method to determine the optimal subset of stations for groundwater
monitoring network using simulated annealing. They performed cast-benefit analysis to determine the number of stations to include in the new design versus loss of information and their result have shown that the relative reduction in exploration casts more than compensates for the relative loss in data representativeness. Nunes et al., (2004 a) proposed a method for designing groundwater monitoring network, well suited for reducing the existing network where data are missing from the time series records. Simulated annealing optimization algorithm has been used in this work to minimize the variance of the estimation error obtained by kriging in combinatorial problems, created by selecting an optimal subset of stations from the original stations. Nunes et al., (2004 b) compared three optimization model using simulated annealing based on i) one that maximizes spatial accuracy, ii) one that minimizes the temporal redundancy, and iii) one that tends to maximize the spatial accuracy as well as minimizes the temporal redundancy. They have found the inclusion of both temporal and spatial information in the optimization model contributes to selection of the most relevant stations. Wu (2004) proposed an algorithm coupled with Kalman filter and finite element method for network design of groundwater monitoring in regions with larger intensity of groundwater abstraction.

Delhomme (1979) studies the spatial variability and uncertainty of different parameters used in groundwater model suing geostatistical approach. Wilson et al., (1978) combined groundwater flow modeling and the Kalman filter for optimal design of the groundwater monitoring. Herrera (1998) developed a cost-effective method for the design of a groundwater monitoring network. The method combined a Kalman filter with a groundwater flow and transport model. The author proved that the model error had a strong correlation in time and that correlation should not be ignored in the sampling network designs. The sampling location was determined by the location where the total uncertainty was reduced most after a sample was taken at that location. Space-time cross-correlation of the concentration field was addressed.
The probability based approaches consider the probability of exceeding a certain level of the field variable as the criterion to be controlled in the network design problem (Rouhani and Hall, 1988).