Groundwater Monitoring Network Design

H.A. Loaiciga
Department of Geological Sciences, Wright State University, Dayton, Ohio, 45435, USA

BACKGROUND AND PROBLEM OVERVIEW

The design of a groundwater quality monitoring network is an important aspect in aquifer restoration and prevention of groundwater pollution. By network design it is understood in this study the selection of the number and locations of sampling wells at a site where there exists contamination. The problem of network design is described in simple terms by classical sampling theory (Cochran, 1977). Suppose a set of independent, identically distributed data are sampled with the purpose of estimating the sample mean. Under further assumptions the optimal sample size, n*, is given by the expression

$$n^* = (\Psi \sigma^2 / C)^{1/2} \qquad (1)$$

in which Ψ = a factor related to the expected loss incurred in estimating the population mean; σ^2 = the (assumed known) population variance and C = the additional cost per sampling well. Eq. (1) contains the three factors of overriding importance in network design: (1) a measure of risk aversion or loss related to the fact that a practical decision will be made based on the results from the sample and that a quantifiable loss is expected due to errors in estimates derived from a data set; (2) the variance, a measure of spread, of the population or phenomenon under study; (3) the cost of sampling. Equation (1) indicates that, other things being constant, the sample size increases as the coefficient associated with expected loss becomes larger. The same increasing trend applies as the variance of the process or population increases. Intuitively, the larger the cost of adding one more sampling well, the more pronounced the trade-off between expected loss (due to errors in results) and the cost of sampling will be, resulting in a smaller sample size that best balances these competing objectives. Within the context of a temporal/spatial

process such as subsurface contamination, knowledge of the number of sampling points is not sufficient. The locations of wells are required as well. The purpose of this paper is to develop a network design procedure that considers the spatial nature of contaminant plumes.

NETWORK DESIGN CONSIDERATIONS

The problem of network design involves some background information. Foremost is hydrogeologic data such as: (1) geologic setting (lithology and stratigraphy); (2) groundwater flow patterns and volumes; (3) recharge areas and rates; (4) aquifer characteristics (e.g., hydraulic conductivity, dispersion coefficients); (5) existing monitoring wells and their locations. A description of general geomorphoclimatic characteristics of a site is also important. This might help to understand the subsurface environment, and could also influence site accessibility, monitoring costs and potential threats of contamination to the surrounding environment. Data on precipitation, temperature, evapotranspiration, topography, accessibility, site size, proximity to ecosystems/population centers would be most valuable. In other words, network design is not a purely statistical optimization exercise. Rather, it has a broader context, and fortunately, such context may be incorporated, to some extent, in the network design approach given herein.

PLUME DETECTION

Prior to undertaking the network design problem, one must have a predetermined set of N locations at which new wells could be developed for monitoring purposes. In addition, the grid layout containing the sampling points and the geometry of such layouts (i.e., shape and distance between sampling points) must be known. Evidently, the sampling grid must be such that, with probability one, one or more wells hit or intersect the contaminant plume. However, there may be instances in which a grid layout that would hit the "hot spot" is not readily obvious unless some information is available about the contaminant plume geometry. This section explains how to use some basic results of hydrodynamic dispersion to approximately assess the probability of failing to detect a plume with some regular arrays of sampling wells.

For large Peclet numbers, contaminant plumes developed by hydrodynamic dispersion (under suitable hydrogeological heterogeneity and anisotropic conditions) are approximately ellipsoidal in shape. The semi-axes of the ellipsoidal plume may be obtained from the square roots of the variances of the plume particles displacements, σ^2_{ij}. The variances are related to macrodispersivity parameters, A_{ij}, by the expression (see Bear, 1972, pp. 610).

$$\sigma^2_{ij} = 2LA_{ij} \qquad (2)$$

in which L = expected distance of the plume's center of mass to

the source of contamination along the direction of average groundwater velocity. A_{ij} may be inferred from hydrogeological data. For example, along the major principle axis (Neuman et al., 1987) of macrodispersivity,

$$A_{11} = (3/8)\sigma_y^2 \ell \tag{3}$$

where σ_y^2 = variance of log-hydraulic conductivity; and ℓ = correlation length of the log-hydraulic conductivity. The major principal axis is assumed to be oriented in the direction of average groundwater velocity. For the other principal axes $A_{22} = A_{33} = f \cdot A_{11}$, where f is a fraction between 0 and 1. There is some controversy in the literature as to what f should be. Neuman et al. (1987) suggest that f = 0 (for large Peclet numbers), while Gelhar and Axness (1983) proposed a nonzero f. For practical purposes and $1/10 \leq f \leq 3/10$ seems reasonable based on empirical evidence reported in the literature. In order to illustrate how to use Eq. (2) for plume geometry delineation assume a normal plume distribution

$$C(x,y,z) = [C_0/(2\pi)^{3/2}\ \sigma_{11}\sigma_{22}\sigma_{33}] \cdot \exp\left[-(1/2)(m_1^2/\sigma^2_{11} + m_2^2/\sigma^2_{22} + m^2_3/\sigma^2_{33})\right] \tag{4}$$

in which $m_1 = x - (x_0 + L)$; $m_2 = y - y_0$, and $m_3 = z - z_0$ are the coordinates relative to the contaminant plume center of mass and (x_0, y_0, z_0) represents the location of the contaminant source; C_0 is the concentration at the source (in theory the source should be a point source). The axes x and m_1 are aligned with the direction of the mean groundwater flow velocity, which in turn is taken as the direction of the major principal axis of the dispersivity tensor. By taking the logarithm on both sides of Eq. (4), solving for the quadratic expression under the exponent on its right-hand side, and normalizing coordinates to obtain the standard ellipse equation (4) yields

$$m_1'^2/\sigma^2_{11} + m_2'^2/\sigma^2_{22} + m_3'^2/\sigma^2_{33} = 1 \tag{5}$$

where $m_i' = m_i/\sqrt{k}$ $(i = 1,2,3)$ and $k = -2\ell n(C/C_0)\ (2\pi)^{3/2}\ (\sigma_{11}\sigma_{22}\sigma_{33})$. Letting the concentration level C(x,y,z) be a fraction p of C_0, Eq. (5) represents the ellipsoid at a concentration pC_0 with semiaxes equal to σ_{11}, σ_{22}, and σ_{33} (see Eq. 2). By selecting p sufficiently small so that pC_0 will fall below detection limits Eq. (5) would represent the plume configuration of interest. Certainly, the ellipsoid semiaxes must be sufficiently large relative to the distance between sampling points so that various wells will intersect the contaminant plume. Otherwise, i.e., if the dimensions of the ellipsoidal plume are relatively small compared to the distance between grid points, then the probability of the plume not being hit by any sampling well becomes an issue.

NETWORK DESIGN

Expected Loss

One approach to determining the number and location of sampling wells is to minimize the total monetary cost involved in the development of the sampling network plus the expected loss incurred in a decision through an estimation error υ on the variable of interest. It is being assumed that groundwater quality data serves as a basis for practical decisions and that losses from such decisions are quantifiable in monetary terms. Sampling theory enables us to find the frequency distribution $g(\upsilon,n)$ of υ which, for a specified sampling method will depend on the sample size n. Under classical assumptions (i.e., independent, identically distributed observations), the expected loss for a sample of size n is

$$R(n) = \int s(\upsilon)g(\upsilon,n)d\upsilon \qquad (6)$$

where $s(\upsilon)$ is the loss incurred through an error of estimate υ. When the space dimension becomes an issue, such as in groundwater networks, Eq. (6) is further complicated by the fact that there are multiple possible arrangements of n wells and each arrangement may produce a different error of estimation υ. Matters are simplified considerably if one assumes a quadratic loss, $s(\upsilon) = \Psi\upsilon^2$ where Ψ is a loss coefficient. If the estimation error has zero mean, then the expected loss is $R(\upsilon) = E(s(\upsilon)) = \Psi V(\upsilon)$ where $V(\upsilon)$ denotes the variance of estimation error. The coefficient Ψ may be estimated by utility analysis within a mean-variance framework.

Geostatistical Estimation

Suppose that groundwater quality data is used in calculating an average concentration, C_0, over a domain. Such domain could be the vicinity of a well, or any other spatial section of aquifer suitable defined by the nature of groundwater investigations. The geostatistical approach is quite suitable for the purpose of concentration estimation based on data collected at nearby sampling wells (control points). Let us assume that n wells provide concentration data ($n \leq N$, where N is the total number of possible wells, see section on PLUME DETECTION). The concentration (it may be a spatial average) estimate is $C^* = \Sigma \lambda_i C_i$ $(i = 1,2,\ldots,n)$, where the λ's are suitable weights imposed on the measured concentrations C_i. It is well known that the geostatistical approach (see, e.g., Journel and Huijbregts, 1978) yields optimal weights λ_i that minimize the variance of estimation error $\upsilon = C_0 - C^*$ and produce an unbiased estimate of C_0 (i.e., the expected value of C^* equals the mean value of C_0). The application of either simple or universal kriging (or any other related variation of these) presumes that the covariance of the concentration field is known. In addition, for spatially variable fields, such as concentration, the mean concentration is variable, i.e., there is a trend, introducing additional parameters. Therefore,

we see that network design has imbedded into it the estimation of correlation and trend parameters. There is an ample body of literature addressing this estimation problem and due to space limitations this subject may not be pursued further herein. It will be seen, though, that in the process of network design one must solve a parameter estimation problem for each network configuration considered.

The Criterion: Network Configuration

Let x_i denote a binary variable that can take the values 1 or 0 depending on whether a sampling well is developed or not at the ith location. The ith location is any of the possible well sites. Then, the estimator of C_0 is $C^* = \Sigma \lambda_i X_i C_i$ $(i = 1,2,\ldots,N)$. The expected loss is proportional to the variance of estimation error as proposed previously. Consequently, the network design criterion is to

$$\underset{\lambda_i, X_i}{\text{Minimize}} \sum_{i=1}^{N} K_i X_i + \Psi \left[\sigma^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j X_i X_j \sigma_{ij}^2 - 2 \sum_{i=1}^{N} \lambda_i X_i \sigma_{i0}^2 \right) \tag{7}$$

in which K_i is the cost of a sampling well at the ith location; σ^2 is the variance of concentration C_0; σ_{ij}^2 and σ_{i0}^2 are the covariances between the ith and jth and the ith and 0 location respectively. There is a budget (B) constraint appended to Eq. (7), i.e.,

$$\sum_{i=1}^{N} K_i X_i \leq B \tag{8}$$

If the estimator C^* is unbiased, then $EC^* = m_0$ where m_0 is the expected value of the estimated concentration C_0, that is $\Sigma \lambda_i x_i m_i = m_0$ $(m_i = E(C_i))$. Assume that the trend at the ith location is

$$m_i = \sum_{k=1}^{Q} b_k^i \beta_k$$

in which β_k are (usually unknown) parameters and b_k^i are functions of the location coordinates of the ith sampling well. Then, the unbiasedness condition is easily shown to be

$$\sum_{i=1}^{N} \lambda_i X_i b_k^i = b_k^0, \quad k = 1,2,\ldots,Q \tag{9}$$

Equations (7) - (9), along with the binary condition X_i = [1 or 0] depending on whether a site is developed or not, respectively, constitute the mathematical network design problem.

On the Solution of the Network Design Problem

The design problem as outlined above is rather complex to solve for the following reasons: (i) the decision variables X_i are binary; (ij) the problem is nonlinear in the objective function as well as in the constraints (see Eq. (9)); (iii) the covariance structure (determined by the hydrogeologic setting) of the concentration field (see σ_2, σ^2_{ij} and σ^2_{i0} in Eq. (7)) are most likely to be unknown implying that there must be a parameter estimation module imbedded in the optimization program. The author believes that with the availability of high-speed "super" computers the solution of the optimization problem is more efficiently approached with a random, combinatorial, search for the optimal number and location of sampling wells.

CONCLUSIONS

The design of groundwater monitoring networks is a function of (1) the statistical heterogeneity and geologic anisotropy of the aquifer; (2) the hydrodynamics of plume migration; (3) the practical decisions that result from a data acquisition program; (4) budgetary constraints imposed on available resources. The joint treatment of such factors and their incorporation on a mathematical (combinatorial) formulation of the problem to determine the best well locations was given in this paper. The next step should be the application to an actual design. Efforts are in progress in this direction. However, the mathematical conceptualization and commanding factors in groundwater network design has been set in this work.

REFERENCES

Bear, J. (1972). Dynamics of Fluids in Porous Media, American Elseiver, New York, N.Y., pp. 764.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed., John Wiley, New York, N.Y., pp. 428.

Gelhar, L.W., and Axness, C.L. (1983). "Three-dimensional stochastic analysis of macrodispersion in aquifers," Water Resour. Res., 19(1), 161-180.

Journel, A.G., and Huijbregts, Ch. J. (1978). Mining Geostatistics, Academic Press, p. 600.

Neuman, S.O., Winter, C.L., and Newman, C.M. (1987). "Stochastic theory of field-scale Fickian dispersion in anisotropic porous media," Water Resour. Res., 23(3), 453-466.