Published November, 1988

Disjunctive Kriging as an Approach to Management Decision Making

S. R. YATES* AND MARYLYNN V. YATES

ABSTRACT

Disjunctive kriging is a nonlinear estimation technique that allows the conditional probability that the value of a spatially variable management parameter is greater than a cutoff level to be calculated. The method can be used in management decision making to help determine when some reclamation action is necessary. Two input parameters are required to use the technique: a cutoff level and the critical probability level. The use of disjunctive kriging as a decision making tool is illustrated using the regulation of septic tank setback distance as a means for minimizing the contamination of groundwater by viruses. Two examples are described: given a setback distance, the spatial distribution of the conditional probability that the virus concentration will be greater than acceptable levels is calculated; and given a critical probability level, calculating the spatial distribution of setback distances which satisfy that probability level. The study showed that, to be 90% confident that virus concentrations would be within acceptable limits, in many areas the setback distance would have to be severalfold higher than prescribed by current regulations.

GEOSTATISTICAL METHODS have been used in the analysis of a variety of agricultural problems. Typically, geostatistical analyses are used to determine various quantities that describe the variability of a parameter in space. In particular, quantities such as the semivariogram and correlation scales are the ultimate goal of a geostatistical investigation.

Geostatistical techniques can also be used for management decision making. Examples include Russo (1984a,b), who described a method for using geostatistics to aid in managing the salinity of a heterogeneous field, and Zirschky (1985), Zirschky and Harris (1986) and Zirschky et al. (1985), who investigated the use of geostatistics for determining reclamation strategies for the cleanup of hazardous waste sites.

A relatively new technique to soil and hydrologic sciences, disjunctive kriging (DK) has several advantages over linear estimation methods. First, since DK is a nonlinear estimator, in general, it provides a more accurate estimate of the property of interest. Second, and more importantly, an estimate of the conditional probability for the property of interest can easily be obtained. This conditional probability can be used as an input to a management decision making model to provide a quantitative means for determining whether management actions are necessary.

Two pieces of information must be available a priori to use DK for management decision making. First, the level or quantity of the property that is considered hazardous or undesirable must be known. This value is called the cutoff or critical level and values of the property that are larger than this level represent the undesirable event. The next necessary information is the probability level that spurs management action. This is the critical probability level where the undesirable levels of the property being investigated will no longer be tolerated. Whenever the value of a property in a region is larger than the cutoff level at a probability equal to or greater than the critical probability level, some corrective action will be applied to lower the value of the property to acceptable levels.

Distinctive kriging, unlike other geostatistical methods such as ordinary kriging, can be used as a quantitative method for making management decisions since the crucial piece of information, the conditional probability, is available. Without this piece of information, regions where the level of a property is greater than a cutoff level can be found, but since the region is demarked only from estimates, it remains unknown whether the region is significantly over the cutoff level or not.

The purpose of this paper is to demonstrate how DK can be used as a management decision making tool. This research was undertaken to develop a means whereby septic tank placement could be optimized to protect drinking water from viral contamination. A brief description of the DK method is given and followed by an example that illustrates how DK can be used to manage viral contamination in the subsurface, and especially groundwater, due to septic tank effluent. This example illustrates how geostatistics can be used to minimize the risks of drinking virus contaminated water and shows how the method can be used for other contamination problems. Knowledge of the spatial distribution of viruses in soil and groundwater is important since viruses are responsible for a large percentage of the waterborne disease outbreaks that occur in the USA (Keswick et al., 1985). One major source of viruses in the subsurface is septic tank effluent, of which approximately 3.8 trillion liters is introduced into the soil each year (Office of Technology Assessment, 1984).

MATERIALS AND METHODS

Water samples were collected from 71 municipal drinking water wells in the Tucson Basin (shown in Fig. 1) and were used in the determination of the virus inactivation rates. Viruses were added to the water samples, which were incubated at the in situ groundwater temperature. Periodically, subsamples were analyzed to determine the number of viruses remaining in the water samples. The virus inactivation rates were calculated based on these analyses. The procedures used to obtain the inactivation rate are described in greater detail by Yates et al. (1986a) and Yates and Yates (1987).

Once the virus inactivation rates were determined, a setback distance, defined herein as the distance between a water supply well and a potential source of contamination such as a septic tank, was determined by using Darcy's law (Freeze and Cherry, 1979) as

$$\text{setback distance} \simeq tKi/\eta_e \qquad [1]$$

where t, K, i, and η_e, respectively, are the travel time (d), the

S.R. Yates, U.S. Salinity Lab., Univ. of California, Riverside, CA 92521; M.V. Yates, Cooperative Extension, Univ. of California, Riverside, CA 92521. Contribution of the U.S. Salinity Lab., 4500 Glenwood Dr., Riverside, CA 92501 and the Dep. of Soils and Environmental Sciences, Univ. of California, Riverside, CA 92521. Received 9 Feb. 1988. *Corresponding author.

Published in Soil Sci. Soc. Am. J. 52:1554–1558 (1988).

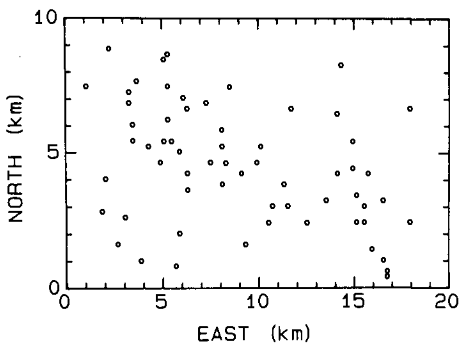

Fig. 1. Location of groundwater sample collection sites in the Tucson Basin.

hydraulic conductivity (m d^{-1}), the hydraulic gradient (m m^{-1}), and the effective porosity (m^3 m^{-3}).

Hydraulic gradients were calculated from a water table elevation map obtained from the city of Tucson. Hydraulic conductivity values were calculated on the basis of transmissivity values provided by the State of Arizona Department of Water Resources. Travel times were calculated using the virus inactivation rate as the number of days required to achieve a 7 log decrease in virus numbers. These values were used in Eq. [1] to obtain a setback distance at each of the sample locations.

In using Eq. [1] it has been assumed that 10^7 viruses are removed from the groundwater in time t, the travel time depends only on the regional flow characteristics, that only horizontal flow occurs, and the local values for the gradient, hydraulic conductivity, and effective porosity remain constant. Although the assumptions used to determine the setback distances are somewhat restrictive, the proposed management decision making method can be illustrated equally well using this simple model as compared to more complex models. If the proposed method is to be used by a municipality for regulatory purposes, however, it is suggested that a more comprehensive technique for determining the setback distances be used, which includes all the important factors affecting the travel time, and therefore, setback distance.

Geostatistical Analysis

Since several reviews of the DK method are available in the literature (Matheron, 1976; Journel and Huijbregts, 1978; Yates et al., 1986b,c; Yates, 1986), only a brief description will be provided herein.

The overall objective when using kriging methods is to obtain a moving average of a property that is distributed in space and/or time. For ordinary kriging (OK) this amounts to determining the constant weighting coefficients. For DK, on the other hand, unknown functions must be determined that may or may not be linear. When these functions are linear and the random function is multivariate normal, the DK method is the same as the OK method and, therefore, OK can be considered a special case of the more general DK method.

Consider a second order stationary random function $Z(\mathbf{x})$ that has been sampled over two dimensions at n locations: $\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_n$. It is assumed that $Z(\mathbf{x})$ is spatially correlated and this correlation can be described by a semivariogram under a second order stationarity hypothesis.

The DK method utilizes the autocorrelation function in determining the weighting coefficients for a series of Hermite polynomials. Therefore, the second order stationarity conditions are required so that the variance exists, in which case the autocorrelation function can be written in terms of the semivariogram

$$\rho(\mathbf{h}) = 1 - \gamma(\mathbf{h})/\gamma(\infty) \qquad [2]$$

where $\rho(\mathbf{h})$ is the autocorrelation function, $\gamma(\mathbf{h})$ is the semivariogram, $\gamma(\infty)$ is the sill value of the semivariogram, and $\mathbf{h}$ is the distance vector.

To obtain the DK estimator, the original data must be transformed into a new variable, $Y(\mathbf{x})$, with a standard normal distribution where pairs of sample values are bivariate normal. The function, $\phi[Y(\mathbf{x})]$, which describes this transformation is

$$\phi[Y(\mathbf{x})] = Z(\mathbf{x}) = \sum_{k=0}^{\infty} C_k\, H_k[Y(\mathbf{x})] \qquad [3]$$

where the values for $Y(\mathbf{x})$ are obtained by taking the inverse, $Y(\mathbf{x}) = \phi^{-1}[Z(\mathbf{x})]$ and $H_k[Y(\mathbf{x})]$ is a Hermite polynomial of order k. The C_k's are the Hermitian coefficients, which are determined using the properties of orthogonality, and are generally determined using numerical integration, as follows

$$C_k = \frac{1}{k!\sqrt{(2\pi)}} \sum_{i=1}^{J} \omega_i\, \phi(v_i) H_k(v_i) \exp[-v_i^2/2] \qquad [4]$$

where v_i and ω_i are the abscissa and weight factors for Hermite integration (Hochstrasser, 1965). If $\phi(s)$ is a standard normal random function, the mean and variance of the data can be found from the coefficients, C_k. The mean value is equal to C_o and variance of the data is $\Sigma_{k=1}^{\infty} k! C_k^2$.

The DK estimator is found from a sum of unknown functions of the transformed sample values, $Y(\mathbf{x}_i)$. It is required that each unknown function, $f_i[Y(\mathbf{x}_i)]$, depend on only one transformed value, $Y(\mathbf{x}_i)$. This gives the DK estimator

$$Z_{\mathrm{DK}}^*(\mathbf{x}_o) = \sum_{i=1}^{n} f_i[Y(\mathbf{x}_i)] = \sum_{i=1}^{n} \sum_{k=0}^{\infty} f_{ik} H_k[Y(\mathbf{x}_i)] \qquad [5]$$

where f_i is the unknown function with respect to the transformed variable, $Y(\mathbf{x}_i)$, which is to be determined. In the right-most part of Eq. [5], the unknown function has been written in terms of a Hermite polynomial with coefficients, f_{ik}.

In an analogous manner to OK, an unbiased estimator with minimum estimation variance is sought, which results in the following system of equations

$$Z_{\mathrm{DK}}^*(\mathbf{x}_o) = \sum_{k=0}^{K} C_k\, H_k^*[Y(\mathbf{x}_o)] \qquad [6]$$

$$H_k^*[Y(\mathbf{x}_o)] = \sum_{i=1}^{n} b_{ik}\, H_k[Y(\mathbf{x}_i)] \qquad [7]$$

where the series in Eq. [6] has been truncated to K terms and b_{ik} are the DK weights. The $H_k^*[Y(\mathbf{x}_o)]$ represents the estimated value of the kth Hermite polynomial at the estimation site. The sum of these estimates multiplied by the coefficient, C_k, [which transforms $Y(\mathbf{x})$ into $Z(\mathbf{x})$] makes up the DK estimate at $\mathbf{x}_o$. To obtain an estimated value for the Hermite polynomial, the DK weights, b_{ik}, must be found by solving the linear kriging equation for each k

$$\sum_{i=1}^{n} b_{ik} (\rho_{ij})^k = (\rho_{oj})^k; \qquad j = 1,2,3,\ldots,n. \qquad [8]$$

When $k = 0$, Eq. [8] represents the unbiasedness condition, that is, that the sum of the weights equals unity. The disjunctive kriging variance is

$$\sigma_{\mathrm{DK}}^2 = \sum_{k=1}^{K} k! C_k^2 \left[1 - \sum_{i=1}^{n} b_{ik} (\rho_{oi})^k\right]. \qquad [9]$$

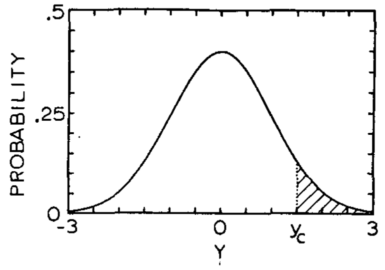

Fig. 2. Hypothetical conditional probability density function. The shaded area represents the conditional probability of exceeding the cutoff level, y_c.

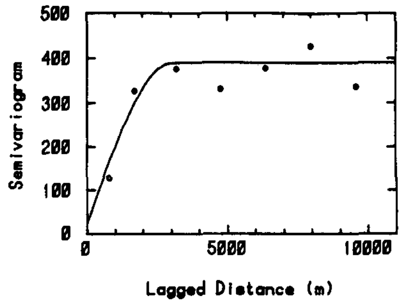

Fig. 3. Semivariogram for the septic tank setback distances.

One advantage the DK method has over OK is that an estimate of the conditional probability that the value at an estimation site is greater than an arbitrary critical value, y_c, can be calculated. This is shown diagrammatically in Fig. 2 where the conditional probability is the shaded area under the curve and to the right of the critical value, y_c. This conditional probability is a useful means for determining the risk of various management alternatives. The conditional probability is obtained by defining an indicator variable that is equal to unity if $Y(\mathbf{x}_i) \geq y_c$ and zero otherwise (see Yates et al., 1986b).

This allows the conditional probability to be written in terms of the conditional expectation and gives the estimator of the conditional probability as

$$P^*_{\mathrm{DK}}(\mathbf{x}_o) = 1 - G(y_c) + g(y_c) \sum_{k=1}^{K} H_{k-1}(y_c) H^*_k[Y(\mathbf{x}_o)]/k! \qquad [10]$$

where $G(y_c)$ and $g(y_c)$ are the cumulative and probability density functions, respectively, for a standard normal variable, and $H^*_k[Y(\mathbf{x}_o)]$ is found using Eq. [7]. The estimated conditional probability density function, $\mathrm{pdf}^*_{\mathrm{DK}}(\mathbf{x}_o)$, is found by taking the derivative of Eq. [10] with respect to y_c and is

$$\mathrm{pdf}^*_{\mathrm{DK}}(\mathbf{x}_o) = g(u)\left\{1 + \sum_{k=1}^{K} H_k(u) H^*_k[Y(\mathbf{x}_o)]/k!\right\}. \qquad [11]$$

RESULTS AND DISCUSSION

The length of time viruses remain infective in groundwater was measured in 71 groundwater well samples in the Tucson Basin. The setback distance was defined as the distance the virally contaminated groundwater would travel under regional groundwater flow conditions from the source (i.e., a septic tank system) to a well during the time interval required for 10^7 viruses to be inactivated. The setback distance at each of the 71 locations was determined using Eq. [1]. The experimental semivariogram shown in Fig. 3 was modeled using a spherical equation with a nugget of 24.0, a sill of 389.0, and a range of 3.0 km. The semivariogram, which was "validated" using the jackknifing method described by Vauclin et al. (1983), produced a reduced mean and variance of 0.027 and 1.11, respectively. A "valid" semivariogram should have a reduced mean and variance of approximately 0.0 and 1.0, respectively, and gives an indication of the differences between the actual and estimated values at the sample locations. The integral scale, λ^*, (Russo and Bresler, 1981; Yates et al., 1988) was found to be 1.3 km. The correlation function was determined using Eq. [2].

Table 1. Hermite coefficients, C_k, which define the transform function ϕ (x).

k		C_k	
0		14.34	
1		15.23	
2		6.666	
3		0.824 9	
4		−0.494 8	
5		−0.192 3	
6		0.001 283	
7		0.111 1	
8		0.000 574	
9		−0.000 754	
Mean (data)	14.35	Mean (Hermite)	14.34
Variance (data)	336.49	Variance (Hermite)	336.19

The next step was to determine the coefficients, C_k, which define the transform function given in Eq. [3]. The method used to obtain the C_k's is described in greater detail by Yates et al. (1986c) and uses a numerical integration scheme in Eq. [4] where the functional values for $\phi(\mathbf{x}_i)$ at the abscissa points were found using 10th order polynomial interpolation. The Hermite coefficients along with the mean and variance determined using the coefficients, C_k, are given in Table 1.

After determining the Hermite coefficients, C_k, DK estimates and estimation variances were obtained using Eq. [6] through [9]. A DK computer program (Yates et al., 1986d) was used to estimate the setback distances that are shown in Fig. 4A along with the estimation variance in Fig. 4B. Comparing the results of Fig. 4A with Fig. 3, 7, and 10 of Yates and Yates (1987) shows that in terms of making estimates of the setback distances in the Tucson Basin, linear and nonlinear estimators produce approximately the same results, although in theoretical terms, the nonlinear estimator (i.e., DK) is a better estimator. This observation suggests that, at least for this data set, the additional complexity of the nonlinear DK estimator is not justified if only spatial estimates of the septic tank setback distance are required.

An advantage DK has over linear kriging methods is that the conditional probability that the setback distance is greater than a critical setback distance can be obtained. Depending on the management model, this information can be expressed in two ways: (i) estimating the spatial distribution of the conditional probability given a critical setback distance; or (ii) given a critical probability level, determining the areal distribution of setback distances that satisfy that probability level.

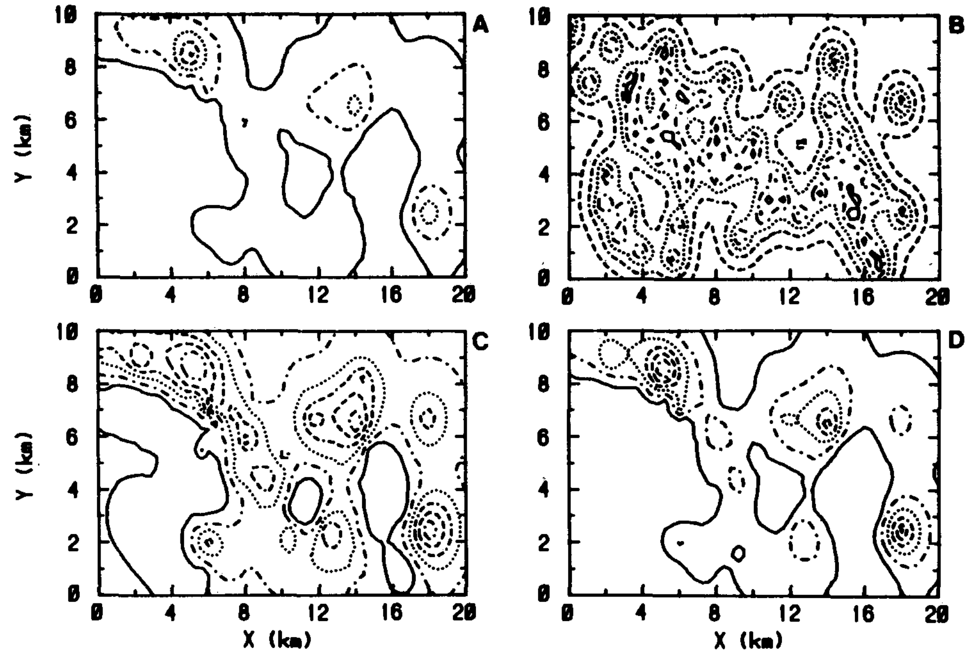

Fig. 4. Contour diagrams for the disjunctive kriging estimates (A), estimation variance (B), and conditional probabilities that the setback distances are > 15 m (C) and > 30 m (D). Contour line types ——, -·-·-, ·····, --·--, -----, and -··-·· are: (A) 15, 30, 45, 60, 75, and 90 m; (B) 100, 150, 200, 250, 300, and 350 m; (C and D) 0.15, 0.30, 0.45, 0.60, 0.75, and 0.9, respectively.

Shown in Fig. 4C and 4D are the conditional probability levels that the setback distance is greater than the critical setback distance of 15 and 30 m, respectively. In this example, if the maximum allowed probability that the setback distance is greater than the cutoff level (i.e., 15 or 30 m) is known, then Fig. 4C and 4D can be used to delineate the areas where septic tanks should not be allowed, which provides a means for groundwater managers to make more informed decisions. In this way the allowing or not allowing the placement of a septic tank or a water supply well (if septic tanks are already present) can be based on quantitative information that accounts for the environmental conditions. The examples shown in Fig. 4C and 4D would be useful in aiding management decision making for the situation where zoning laws specify the setback distances. In most parts of the USA the mandatory setback distance is about 15 to 30 m (Plews, 1977). Figure 4 shows that in many parts of the Tucson Basin a setback distance greater than the 30 m required by law would be necessary to insure 7 logs of virus inactivation, however. Therefore, for lo-

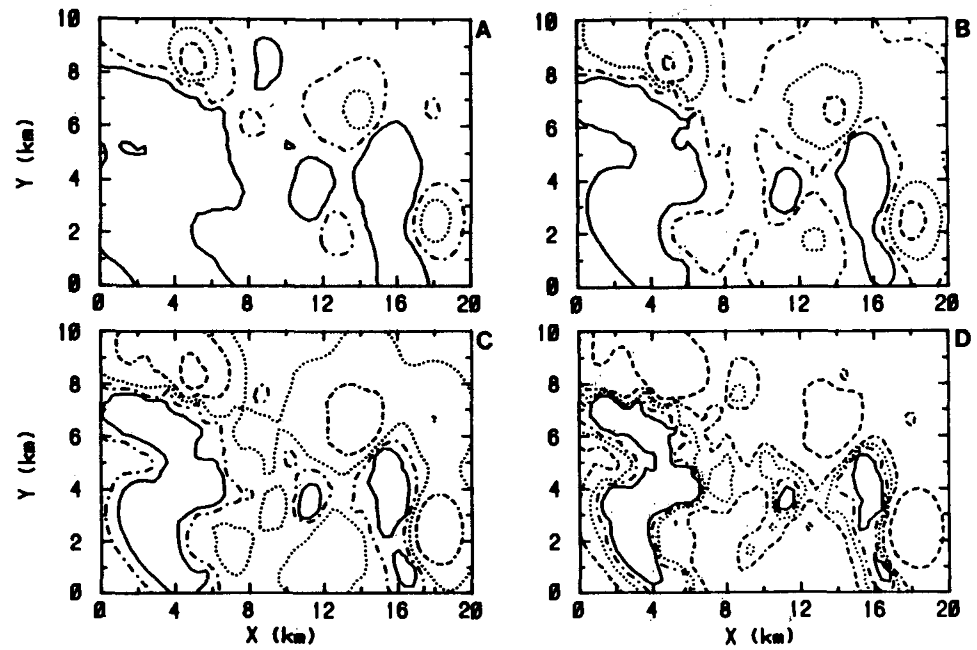

Fig. 5. Contour diagram for the critical setback distances given a minimum allowed probability level using disjunctive kriging. In A, B, C, and D, the conditional probability is 0.20, 0.10, 0.05, and 0.01, respectively. Contour levels: —— 20.0m, -·-·- 40.0 m, ····· 60.0 m, --·--·- - 80.0 m, ----- 100.0 m.

cations where the setback distance is regulated, it would be possible to use DK and the conditional probability to more accurately determine whether it would be safe to place a septic tank at a given location. It would also be possible to use a zonal approach in a management decision making model. For example, a regulation might state that if the probability is <0.25, no restrictions are necessary in that area; between 0.25 and 0.5, septic tanks could be placed on lots meeting certain requirements, such as soil percolation values and minimum lot size; and >0.5, an even larger lot size could be required, or septic tanks could be prohibited in those areas. The method can easily be manipulated to allow such flexibility.

Another approach to regulate septic tank placement would be to require that the probability be high that a prescribed level (i.e., 7 orders of magnitude) of virus inactivation will occur in the time that the water moved from the septic tank to the well. For this case, the regulations would provide a minimum probability level that must be achieved. Then DK could be used to determine the critical setback distances, that is, the minimum allowed setback distance that is associated with the required probability level. This has been done for the Tucson Basin for four probability (of contamination) levels: 0.20, 0.10, 0.05, and 0.01, and is shown in Fig. 5A through Fig. 5D, respectively. For this situation, and given a minimum allowed probability level, the contour levels denote the minimum setback distance necessary to assure adequate distance to achieve 7 orders of magnitude of virus inactivation as the groundwater travels that distance. As expected, the setback distances associated with the higher probability levels are greater than would be required in the same location compared to those calculated for a lower level. For example, at (18,1), a 40 m setback distance would be required to be 80% certain that the groundwater would be free of viral contamination at a well (i.e., 1 − probability of contamination). However, 100 m would be required at the same location if a 99% probability of freedom from contamination is required. Thus, the probability level chosen will have a profound effect on the calculated setback distance.

CONCLUSIONS

The DK method can be used to aid in management decision making of the placement of septic tanks to protect a municipality's drinking water supply. The method can be used provided that the regulations are based on a specified probability that the actual setback distance will be greater than a critical setback distance. The DK method can be used to provide this information in two forms: (i) the conditional probability can be determined given a known critical setback distance, or (ii) the critical setback distances can be determined given a known minimum allowed conditional probability level. Both approaches are described herein.

The second approach, that is, specifying a minimum allowed conditional probability and determining the necessary setback distance has advantages over the former in that the epidemiological considerations are the only requirement for the management decision making model, and this information can be determined without consideration of the hydrologic environment. This allows a municipality to determine to what extent they will go in describing the hydrologic environment and the extent of the hydrologic investigation. Also, the hydrologic environment will not be embedded in the regulation. Regulating the setback distance based on a critical distance (e.g., 15 m) does not take into account that at various places the hydrologic setting might allow safe distances of 5 m while other locations may require distances in excess of 200 m.

ACKNOWLEDGMENT

The research described in this article was supported in part by the USEPA, R.S. Kerr Environmental Research Laboratory, P.O. Box 1198, Ada, OK 74820, through an Interagency Cooperative Research Agreement no. DW 12932631O-01-0.

REFERENCES

Freeze, R.A., and J.A. Cherry. 1979. Groundwater. Prentice Hall, Englewood Cliffs, NJ.

Hochstrasser, U.W., 1965. Handbook of mathematical functions. M. Abramowitz and I.A. Stegun (ed.) Dover Publ., New York.

Journel, A.G., and Ch.J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York.

Keswick, B.H., T.K. Satterwhite, P.C. Johnson, H.C. Dupont, S.L. Secor, J.A. Bitsura, G.W. Gary, and J.C. Hoff. 1985. Inactivation of Norwalk virus in drinking water by chlorine. Appl. Environ. Microbiol. 50:261–264.

Matheron, G. 1976. A simple substitute for conditional expectation: The disjunctive kriging. *In* Proc. North Atlantic Treaty Organization Advanced Study Inst. "Geostat 75", D. Reidel, Hingham, MA.

Office of Technology Assessment. 1984. Protecting the nation's groundwater from contamination. OTA-0-233, U.S. Congress, Washington DC.

Plews, G. 1977. Management guidelines for conventional and alternative onsite sewage systems—Washington State. p. 187–193. *In* N.I. McClelland (ed.) Individual Onsite Wastewater Systems, Proc. 3rd Natl. Conf. Ann Arbor Science, Ann Arbor, MI.

Russo, D. 1984a. Statistical analysis of crop yield soil water relationships in heterogeneous soil under trickle irrigation. Soil Sci. Soc. Am. J. 48:1402–1410.

Russo, D. 1984b. A geostatistical approach to solute transport in heterogeneous fields and its applications to salinity management. Water Resour. Res. 20:1260–1270.

Russo, D., and E. Bresler. 1981. Soil hydraulic properties as stochastic processes: I. An analysis of field spatial variability. Soil Sci. Soc. Am. J. 45:682–687.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175–184.

Yates, M.V., and S.R. Yates. 1987. A comparison of geostatistical methods for estimating virus inactivation rates in ground water. Water Res. 21:1119–1125.

Yates, M.V., S.R. Yates, A.W. Warrick, and C.P. Gerba. 1986a. Use of geostatistics to predict virus decay rates for determination of septic tank setback distances. Appl. Environ. Microbiol. 52:479–483.

Yates, S.R. 1986. Disjunctive kriging. III. Cokriging. Water Resour. Res. 22:1371–1376.

Yates, S.R., A.W. Warrick, A.D. Matthias, and S. Musil. 1988. Spatial variability of remotely sensed surface temperatures at field scale. Soil Sci. Soc. Am. J. 52:40–45.

Yates, S.R., A.W. Warrick, and D.E. Myers. 1986b. Disjunctive kriging. I. Estimation and conditional probability. Water Resour. Res. 22:615–621.

Yates, S.R., A.W. Warrick, and D.E. Myers. 1986c. Disjunctive kriging. II. Examples. Water Resour. Res. 22:623–630.

Yates, S.R., A.W. Warrick, and D.E. Myers. 1986d. A disjunctive kriging program for two dimensions. Comput. Geosci. 12:281–313.

Zirschky, J. 1985. Geostatistics for environmental monitoring and survey design. Environ. Int. 11:515–524.

Zirschky, J., and D.J. Harris. 1986. Geostatistical analysis of hazardous waste site data. J. Environ. Eng. (NY) 112:770–784.

Zirschky, J., G.P. Keary, R.O. Gilbert, and E. J. Middlebrooks. 1985. Spatial estimation of hazardous waste site data. J. Environ. Eng. (NY) 111:777–789.