Soil Parameters and Sampling Scheme for Characterizing Soil Hydraulic Properties of a Watershed

Watershed modeling, which incorporates the stochastic nature of the hydraulic properties of the land surface and rainfall, requires a mathematical description of watershed variability, including the frequency distribution of key hydrologic parameters and the spatial structure of variances. Heterogenous watersheds require extensive sampling to characterize the spatial distribution of a property, such as hydraulic conductivity, which is frequently required as input to model calculations of infiltration and runoff. Since hydraulic conductivity, $K$, varies with water content, $\theta$, and soil water pressure, $h$, the $K(\theta)$ and $K(h)$ relationships can be conveniently represented by parameters in mathematical expressions relating these variables. The parameters of three different equations are examined as indices of the hydraulic properties of Oxisol soils in Hawaii's Pearl Harbor watershed. When what to measure and how to mathematically express the results is decided, the number and location of field measurement sites in a particular watershed are determined. Geostatistical concepts are applied to design a sampling scheme for a specific watershed in which the measured value of a hydrologic property or index at a given point is correlated with other measured values of the property that is dependent on the distance between sampling points. Required statistical parameters for the geostatistical approach are the mean, variance, and autocorrelation function or variogram. Criteria are specified for selecting the location and smallest possible number of observation points to best estimate the statistical parameters. The results suggest that 30 measurement sites are the minimum sample size for estimating the parameters required for stochastic modeling. The proposed sampling procedure is illustrated with a sampling strategy for a portion of the Pearl Harbor watershed.
AUTHORS:

Dr. Richard E. Green
Professor
Dept. of Agronomy and Soil Science
College of Tropical Agriculture
    and Human Resources
University of Hawaii at Manoa
3190 Maile Way, St. John 203F
Honolulu, Hawaii 96822

Dr. Eshel Bresler
Soil Physicist and Head
Division of Soil Physics
Institute of Soils and Water
Volcani Center
P.O. Box 6
Bet Dagan 50-250, Israel

Contents of this publication do not necessarily reflect the views and policies of the United States Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement or recommendation for use by the United States Government

$8.00/copy
Checks payable to: Research Corporation, University of Hawaii

Mail to: University of Hawaii at Manoa
    Water Resources Research Center
    2540 Dole St., Holmes Hall 283
    Honolulu, Hawaii 96822
    Tel.: (808) 948-7847 or -7848
ABSTRACT

Watershed modeling, which incorporates the stochastic nature of the hydraulic properties of the land surface and rainfall, requires a mathematical description of watershed variability, including the frequency distribution of key hydrologic parameters and the spatial structure of variances. Heterogeneous watersheds require extensive sampling to characterize the spatial distribution of a property, such as hydraulic conductivity, which is frequently required as input to model calculations of infiltration and runoff.

Since hydraulic conductivity, $K$, varies with water content, $\theta$, and soil water pressure, $h$, the $K(\theta)$ and $K(h)$ relationships can be conveniently represented by parameters in mathematical expressions relating these variables. The parameters of three different equations are examined as indices of the hydraulic properties of Oxisols soils in Hawai'i's Pearl Harbor watershed.

When what to measure and how to mathematically express the results is decided, the number and location of field-measurement sites in a particular watershed are determined. Geostatistical concepts are applied to design a sampling scheme for a specific watershed in which the measured value of a hydrologic property or index at a given point is correlated with other measured values of the property that is dependent on the distance between sampling points. Required statistical parameters for the geostatistical approach are the mean, variance, and autocorrelation function or variogram. Criteria are specified for selecting the location and smallest possible number of observation points to best estimate the statistical parameters. The results suggest that 30 measurement sites are the minimum sample size for estimating the parameters required for stochastic modeling. The proposed sampling procedure is illustrated with a sampling strategy for a portion of the Pearl Harbor watershed.
CONTENTS

ABSTRACT ......................................................... v
INTRODUCTION ..................................................... 1

SOIL HYDRAULIC PROPERTY INDICES
FOR WATERSHED CHARACTERIZATION ................................ 3
  HYDRAULIC CONDUCTIVITY DATA .................................. 3
  ASSESSMENT OF EQUATION PARAMETERS AS POSSIBLE
  INDICES OF SOIL HYDRAULIC PROPERTIES ...................... 5

SAMPLING SCHEME DESIGN FOR DETERMINING SOIL
HYDRAULIC PROPERTIES OF A HETEROGENEOUS WATERSHED .......... 11
  STATISTICAL CONCEPTS ........................................... 12
  STOCHASTIC MODEL UTILIZING STATISTICAL CONCEPTS .......... 14
  ESTIMATIONS OF STATISTICAL PARAMETERS ..................... 16
  SELECTING LOCATIONS OF RANDOM SAMPLING POINTS
  FOR AUTOCORRELATION AND VARIOGRAM ESTIMATES .............. 19
  RESULTS OF RANDOM SAMPLING POINT LOCATIONS ............... 22
  IMPROVING THE SETTING OF OBSERVATION POINTS
  BY CONDITIONAL STATISTICS ................................... 31

SUMMARY AND CONCLUSIONS ....................................... 38
REFERENCES ....................................................... 40

Figures

1. Schematic of Watershed Grid Showing the Sampling Points
   Generating Domain and Boundary of Actual Domain ............ 15
2. Mean Square Deviation Computed from Equation (25) for
   27 Realizations and Six N Sample Sizes ....................... 23
3. Average (Cumulative) Mean Square Deviation of
   1 through 27 Realizations (Runs) ............................ 24
4. Mean Square Deviation as a Function of Sample Size (N) .... 25
5. Average of 27 Realizations of Relative Maximum Lag
   and Relative Minimum Lag .................................... 26
6. Frequency of Number of Pairs of Points in Best Realization
   (Run No. 16) as a Function of Relative Lag ................. 27
7. Frequency of Number of Pairs of Points in Worst Realization
   (Run No. 20) as a Function of Relative Lag ............... 28
8. Cumulative Frequency of Number of Pairs of Points in Best and Worst Realizations.......................... 30
9. Schematic Representation of Best Locations in Watershed Domain................................................. 32
10. Schematic Presentation of Worst Realization.................................................. 32
11. Watershed Domain with Arbitrary 30 Regular Points from which Locations of Additional Sampling Points were Selected............................. 35
12. Watershed Domain with 30 Generated Sampling Points from which Locations of Additional Observation Points were Selected........................................... 35
13. Cumulative Frequency of Pairs of Points (n[τ]/M) as a Function of Relative Lag (τ/R)............................. 37
14. Watershed Domain with Locations of N = 20 + 20 Sampling Points from Figures 9 and 11..................... 39
15. Watershed Domain with Locations of N = 20 + 10 Sampling Points from Figures 9 and 12 ..................... 39

Tables

1. Equations Relating Hydraulic Conductivity to Soil-Water Pressure Head or Volumetric Water Content............. 4
2. Parameters for 3 Hydraulic Conductivity Equations for Lahaina, Molokai, and Wahiawa Soils....................... 5
3. Hydraulic Conductivity Parameters by Permeability Classes for O'ahu Oxisols....................................... 8
4. Order by Hydraulic Conductivity Parameters of Seventeen Oxisol Soils Sites, O'ahu................................. 9
5. Values of Mean Square Deviation, τ_max/R, τ_min/R of N + 10 Observation Points for N Taken from Best and Worst Realizations and from 10 Best Additional points................ 36
INTRODUCTION

Increasing demands for water in Hawai'i, coupled with a diminishing groundwater supply (Liu, Lau, and Mink 1981), require detailed analyses of all aspects of the hydrologic cycle to provide a rational basis for future water use and water resource development. Mathematical simulation is an important analytical tool which allows quantification of various hydrologic processes and provides a means of predicting the impact of natural and man-imposed environmental changes on surface waters and groundwaters. Simulation models have been recently applied in Hawai'i to study rainfall-runoff relationships in Moanalua Valley, O'ahu (Shade 1981), and groundwater movement in the Pearl Harbor aquifer (Liu, Lau, and Mink 1981). Deterministic models involving water movement in the soil, i.e., in the uppermost part of the vadose zone, require mathematical descriptions of (1) the hydraulic conductivity, \( K \), as a function of either volumetric water content, \( \theta \), or soil-water pressure head, \( h \), that is, \( K(\theta) \) or \( K(h) \); and (2) the soil-water retentivity function, \( \theta(h) \). The in situ measurement of these properties was the objective of previous research in Hawai'i (Ahuja et al. 1980; Chong, Green, and Ahuja 1981; Green et al. 1982).

The present study is a logical sequel to the previous work on methodology. To be most useful in modeling, hydraulic conductivity data should be summarized by mathematical functions, although tabular data may be used in numerical solutions. Thus, we have sought to identify the mathematical forms of the \( K(h) \) or \( K(\theta) \) relationship which are likely to be most useful in future modeling efforts. A primary objective of this study is to identify hydrologic indices for soils which would provide rational criteria for identifying soils that are hydrologically similar. A recent study of some Oxisols on O'ahu revealed that existing soil survey mapping units did not satisfactorily differentiate soil areas with regard to the hydrologic properties of these soils (Green et al. 1982). Thus, most Hawai'i watersheds must be evaluated from a hydrologic perspective, which may involve determination of the spatial distribution of hydraulic conductivity if such information is required in a model. If one of the recently developed simplified methods can be used to measure hydraulic conductivity and if the resulting \( K(\theta) \) or \( K(h) \) data are expressed in suitable equations, then the parameters of these equations may be the most appropriate indices for describing spa-
tial variability and for designating areas of relative homogeneity with re­spect to hydraulic conductivity. Given suitable measurement methods and
hydrologic indices for characterizing the soils in a given watershed, the
number and location of sites to be characterized must be determined—a sam­pling problem which constitutes the second objective of this study.

This report is divided into two sections which are closely associated
with the two objectives stated above. The sections are (1) Soil Hydraulic
Property Indices for Watershed Characterization and (2) Sampling Scheme De­
sign for Determining Soil Hydraulic Properties of a Heterogeneous Watershed.
A number of soil-water properties or related physical properties could probably be used as indices to evaluate the lateral heterogeneity of soils in a watershed. Saturated hydraulic conductivity, porosity, and bulk density of the surface horizon are reasonable candidates. Among these, the saturated conductivity, $K_s$, is the property most directly related to hydrologic processes. The U.S. Department of Agriculture, Soil Conservation Service (USDA, SCS) has used "permeability"—roughly equivalent to saturated conductivity—as one of several soil properties defining the nature of soil horizons. No single "best" hydrologic index for defining spatial variability of soils probably exists; the best choice will be governed by the intended use of the information and the financial and other resources required to accomplish the necessary measurements. In the present work we are assuming that it is desirable to assess the spatial distribution of soil properties which control water flow in soils in the unsaturated and saturated conditions. Since field methods are available to estimate $K(\theta)$ or $K(h)$ in the field without requiring excessive time and expense, indices derived from these data can be utilized.

A number of mathematical equations have been used by various investigators to relate $K$ to $h$ and $\theta$. Most of the proposed equations are empirical in nature; therefore, the choice of which functional form to use depends principally on the goodness of fit between measured data and the mathematical expression and on the intended use of the equations. The three equations given in Table 1 have been examined. The exponential form of $K(h)$ (eq. [1]) and the power form (eq. [2]) are easily transformed to linear forms. Also, the air-entry pressure can be roughly estimated by the intercept of the linearized functions $1/b$ and $2/b$. Equation (3) gives $K$ as a function of $\theta$, with $K_s$, $\theta_s$, and $\theta_p$ specified independently. Note that $\theta(h)$ is readily obtained from a combination of equations (2) and (3).

HYDRAULIC CONDUCTIVITY DATA

The data used in this study were obtained in previous research concerned with the development of simplified field methods of measuring hydrau
TABLE 1. EQUATIONS RELATING HYDRAULIC CONDUCTIVITY (K) TO SOIL-WATER PRESSURE HEAD (h) OR VOLUMETRIC WATER CONTENT (θ)

<table>
<thead>
<tr>
<th>Eq. No.</th>
<th>Independent Variable</th>
<th>a. Equation</th>
<th>b. Linearized Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>h</td>
<td>( \frac{K}{K_s} = \exp[\alpha(h - h_a)] )</td>
<td>( \ln\left(\frac{K}{K_s}\right) = \alpha h - \alpha h_a )</td>
</tr>
<tr>
<td>(2)</td>
<td>h</td>
<td>( \frac{K}{K_s} = \left(\frac{h_a}{h}\right)^n )</td>
<td>( \ln\left(\frac{K}{K_s}\right) = n \ln h_a - n \ln h )</td>
</tr>
<tr>
<td>(3)</td>
<td>θ</td>
<td>( \frac{K}{K_s} = \left(\frac{(\theta - \theta_p)}{(\theta_s - \theta_p)}\right)^\gamma )</td>
<td>( \ln\left(\frac{K}{K_s}\right) = \gamma \ln\left(\frac{(\theta - \theta_p)}{(\theta_s - \theta_p)}\right) )</td>
</tr>
</tbody>
</table>

**NOTE:** \( K_s \), the saturated hydraulic conductivity, \( \theta_s \), the saturated water content, and \( \theta_p \), the residual water content, are independently specified; the least-squares fitting yields the parameters \( \alpha \), \( n \), and \( \gamma \); \( h_a \), the air-entry soil-water pressure head is roughly estimated in the fitting procedure.

Hydraulic conductivity was measured by a transient drainage-flux method following steady infiltration on a number of sites in central O'ahu (Green et al. 1982). The soil series included the Lahaina (Tropeptic Haplustox), Molokai (Typic Torrox), and Wahiawa (Tropeptic Eustrustox) soils of the Oxisols soils order but from different suborders. These soils are all well-drained so that the drainage-flux method was appropriate. While conductivities were obtained for various profile depths, data for the 30-cm depth were available for the most sites and are, therefore, used in the present analysis. Site designations for each measurement site are given in the first column of Table 2. The designation L1-1, for example, refers to the Lahaina soil series, location 1, replicate 1. In all but three of the 17 sites listed in Table 2, conductivities were calculated from flux and gradient data which were obtained by a detailed analysis of transient soil-water pressure-head profiles, with profile water contents being derived from soil water-retention measurements on soil cores (Green et al. 1982). This analysis was not accomplished on sites M4-1, M4-2, and M4-3 for which conductivities were obtained by the simplified analysis of Chong, Green, and Ahuja (1981). This difference is relevant in our analysis here because the simplified method assumes power function relationships for \( \theta \) versus time and \( h \) versus time during the drainage process. This is tantamount to specifying the power-function form of the \( K(\theta) \) and \( K(h) \) relationships, as
### TABLE 2. PARAMETERS FOR THREE HYDRAULIC CONDUCTIVITY EQUATIONS FITTED TO K(h) AND K(θ) DATA FOR LAHAINA (L), MOLOKAI (M), AND WAHIAWA (W) SOILS, 0- TO 30-CM DEPTH

<table>
<thead>
<tr>
<th>Soil Site</th>
<th>Eq. (1)b</th>
<th>Eq. (2)b</th>
<th>Eq. (3)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1-1</td>
<td>0.0517</td>
<td>1.97</td>
<td>7.24</td>
</tr>
<tr>
<td>1-2</td>
<td>0.0629</td>
<td>1.79</td>
<td>8.15</td>
</tr>
<tr>
<td>L2-1</td>
<td>0.0315</td>
<td>1.21</td>
<td>5.10</td>
</tr>
<tr>
<td>2-2</td>
<td>0.0283</td>
<td>1.35</td>
<td>5.05</td>
</tr>
<tr>
<td>L3-2</td>
<td>0.0444</td>
<td>2.26</td>
<td>7.51</td>
</tr>
<tr>
<td>M1-1</td>
<td>0.0267</td>
<td>1.35</td>
<td>6.65</td>
</tr>
<tr>
<td>1-2</td>
<td>0.0367</td>
<td>1.58</td>
<td>6.04</td>
</tr>
<tr>
<td>M3-1</td>
<td>0.0427</td>
<td>2.39</td>
<td>7.55</td>
</tr>
<tr>
<td>3-2</td>
<td>0.0494</td>
<td>1.50</td>
<td>7.01</td>
</tr>
<tr>
<td>M4-1</td>
<td>0.0313</td>
<td>3.25</td>
<td>4.55</td>
</tr>
<tr>
<td>4-2</td>
<td>0.0232</td>
<td>1.93</td>
<td>4.12</td>
</tr>
<tr>
<td>4-3</td>
<td>0.0268</td>
<td>2.79</td>
<td>3.09</td>
</tr>
<tr>
<td>W1-1</td>
<td>0.0638</td>
<td>1.68</td>
<td>9.91</td>
</tr>
<tr>
<td>1-2</td>
<td>0.0785</td>
<td>1.57</td>
<td>9.00</td>
</tr>
<tr>
<td>W3-2</td>
<td>0.0510</td>
<td>1.93</td>
<td>9.33</td>
</tr>
<tr>
<td>3-3</td>
<td>0.0378</td>
<td>2.07</td>
<td>7.57</td>
</tr>
<tr>
<td>3-4</td>
<td>0.0378</td>
<td>2.05</td>
<td>5.67</td>
</tr>
</tbody>
</table>

**NOTE:** Hydraulic conductivity data and descriptions of each soil are provided elsewhere (Green et al. 1982). 

*Equations (1)b, (2)b, and (3)b are defined in Table 1.*

Evidenced by the perfect correlation ($r^2$ = 1) for the fit of equation (2), the power function, to $K(h)$ data for the M4 sites (Table 2).

**ASSESSMENT OF EQUATION PARAMETERS AS POSSIBLE INDICES OF SOIL HYDRAULIC PROPERTIES**

The linearized equations (1)b, (2)b, and (3)b in Table 1 were fitted by least-squares analysis to $K(h)$ and $K(θ)$ data to obtain the results in Table 2. All three equations provided a reasonably good fit, as evidenced by the $r^2$ values. The average $r^2$ values for equations (1)b, (2)b, and (3)b...
were respectively 0.93, 0.90, and 0.94. The choice of which equation to select for application purposes depends on four principal criteria:

1. The equation must well represent the experimental data
2. The equation parameters should be useful in discriminating between soils of dissimilar hydraulic properties
3. The equation should be a form which is most useful in subsequent applications of theories to practical situations
4. Previous use of the equation by other investigators on other soils allows a comparison of results and perhaps generalization of conclusions.

Equation (1), the exponential form, is probably the most widely used in recent years, primarily because of its usefulness in linearizing the nonlinear partial-differential equation for water flow in soils, particularly flow in two or three dimensions (Gardner 1958; Wooding 1968; Philip 1968). The linearization procedure involves transforming the nonlinear Richard's equation to a linear form with the aid of the integral transform, $\phi(h)$, sometimes called the "matrix flux potential,"

$$\phi(h) = \int_{h_0}^{h} K(h) dh.$$ (4)

Here $h_0$ corresponds to $\theta_0$, a reference water content usually taken at the dry end of the range of practical interest. The combination of equation (4), with $K(h)$ expressed by equation (1)a, accomplishes the linearization of the equation for steady flow, and allows the use of analytical or quasi-analytical solutions for certain practical boundary conditions. A variety of more recent applications of the transformed flow equations have been made for steady and transient flow regimes, including applications concerned with drip irrigation (Warrick 1974; Bresler 1977, 1978). Linearization of the flow equation for transient cases requires assumptions that may be quite unrealistic, but interesting applications abound (Warrick, Lomen, and Tonellato 1981).

With regard to the appropriateness of equation (1), the following statement by Philip (1968) is worth quoting:

Even though it cannot be claimed that [the] equation...is universally exact, it does model in a reasonably convincing and adaptable way the generally observed rapid and nonlinear decrease of $K$ with $\psi$ in unsaturated materials. $\alpha$ has the dimensions [length]$^{-1}$ and is conveniently expressed in cm$^{-1}$. Typically, $\alpha$ is about 0.01 cm$^{-1}$, and the range 0.05 cm$^{-1}$ to 0.002 cm$^{-1}$ seems to cover
most applications. $\alpha$ is a measure of the relative importance of gravity and capillarity for water movement in the particular material. Fine-textured materials, where capillarity tends to dominate, have small $\alpha$-values, and coarse-textured ones, where gravity effects manifest themselves most readily, have large $\alpha$-values.

Bresler (1978) presents a table of equation (1) parameters for a number of porous materials. Warrick, Lomen, and Tonellato (1981) present parameter values for additional soils. Values of $\alpha$ presented in these references are generally within the range specified by Philip (1968), although an upper limit of $\alpha = 0.10 \text{ cm}^{-1}$ might be more appropriate than $0.05 \text{ cm}^{-1}$.

An examination of $\alpha$ values in Table 2 for the Oxisols in this study reveals a range from 0.023 to 0.078, where the mean value of $\alpha$ is 0.043. These values are at the upper end of the range suggested by Philip and thus indicate soils in which "gravity effects manifest themselves most readily." The high values of $\alpha$ for the A horizon of these Oxisols is consistent with their being highly aggregated, with a relatively high percentage of large pores which drain readily at high water contents.

The saturated conductivity, $K_s$, for these soils tends to be positively correlated with $\alpha$, as indicated by the linear correlation coefficient $r = 0.71$. To relate the magnitude of $K_s$ values for these soils to the USDA SCS permeability classes, the $K_s$ data are presented in increasing order and grouped according to permeability classes in Table 3. It is interesting to observe that at least one soil site fits in each class, with the exception of the "very slow" class which is omitted from the table. Thus, these Oxisols have a wide range of $K_s$ values, of which most are within the conductivity ranges of the "moderate" and "moderately rapid" classes. Also, the $K_s$ values are not neatly ordered in groups corresponding to soil series, although the Molokai soils have the lowest $K_s$ values, and Wahiawa and Lahaina have the highest values. Even the replicate sites for a given location of a soil series have widely different $K_s$ values in several cases (compare, for example, M1-1 with M1-2, L2-1 with L2-2, and W1-1 with W1-2). These results suggest that $K_s$ is probably not the best index for characterizing differences of soils with respect to some hydrologic processes. $K_s$ is highly sensitive to the presence of large pores or cracks which may not be representative of a given soil location. It seems likely that values of $\alpha$ would be less subject to the influence of very large pores. Values of $\alpha$ for each soil site are shown with the corresponding $K_s$ value in Table 3. Mean values
### TABLE 3. HYDRAULIC CONDUCTIVITY PARAMETERS BY PERMEABILITY CLASSES FOR O'AHU OXISOLS

<table>
<thead>
<tr>
<th>USDA SCS* Permeability Class</th>
<th>Soil-Site Designation</th>
<th>( K_0 ) (cm/min)</th>
<th>( \bar{K}_0 ) (cm/min)</th>
<th>( \alpha ) (em/min)</th>
<th>( \bar{\alpha} ) (em/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow (0.0021-0.0085 cm/min)</td>
<td>M4-3</td>
<td>0.008</td>
<td>0.008</td>
<td>0.0268</td>
<td>0.0268</td>
</tr>
<tr>
<td>Moderately Slow (0.0085-0.034 cm/min)</td>
<td>M4-2</td>
<td>0.022</td>
<td>0.025</td>
<td>0.0232</td>
<td>0.0300</td>
</tr>
<tr>
<td></td>
<td>M1-2</td>
<td>0.028</td>
<td></td>
<td>0.0367</td>
<td></td>
</tr>
<tr>
<td>Moderate (0.034-0.106 cm/min)</td>
<td>M4-1</td>
<td>0.041</td>
<td>0.062</td>
<td>0.0313</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W3-4</td>
<td>0.049</td>
<td></td>
<td>0.0378</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L2-2</td>
<td>0.050</td>
<td></td>
<td>0.0283</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M1-1</td>
<td>0.055</td>
<td></td>
<td>0.0267</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W1-1</td>
<td>0.082</td>
<td></td>
<td>0.0638</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W3-3</td>
<td>0.096</td>
<td></td>
<td>0.0378</td>
<td></td>
</tr>
<tr>
<td>Moderately Rapid (0.106-0.212 cm/min)</td>
<td>M3-2</td>
<td>0.114</td>
<td>0.133</td>
<td>0.0494</td>
<td>0.0451</td>
</tr>
<tr>
<td></td>
<td>L2-1</td>
<td>0.120</td>
<td></td>
<td>0.0315</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L3-2</td>
<td>0.120</td>
<td></td>
<td>0.0444</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M3-1</td>
<td>0.121</td>
<td></td>
<td>0.0427</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W3-2</td>
<td>0.146</td>
<td></td>
<td>0.0510</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L1-1</td>
<td>0.180</td>
<td></td>
<td>0.0517</td>
<td></td>
</tr>
<tr>
<td>Rapid (0.212-0.423 cm/min)</td>
<td>W1-2</td>
<td>0.300</td>
<td>0.300</td>
<td>0.0785</td>
<td>0.0785</td>
</tr>
<tr>
<td>Very Rapid (&gt;0.423 cm/min)</td>
<td>L1-2</td>
<td>0.44</td>
<td>0.440</td>
<td>0.0629</td>
<td>0.0629</td>
</tr>
</tbody>
</table>

*U.S. Department of Agriculture, Soil Conservation Service.

of \( \alpha \) for each permeability class tend to increase with an increase in \( K_0 \). To examine the relative position of the seventeen soil sites relative to their respective \( \alpha \) values, the \( \alpha \) data from Table 2 are presented in increasing order in Table 4, with the soil sites arranged in columns for each of the three soil series. The ranges of \( \alpha \) values for the three series are seen to overlap, but \( \alpha \) values are more well ordered in the series groups than are \( K_0 \) values. Perhaps of even greater significance is the fact that replicates of each soil site have \( \alpha \) values which are close neighbors in the list of ordered \( \alpha \) values. With the exception of M1-1 and M1-2, which differ by 0.01 cm\(^{-1}\), and W3-2, which is 0.013 cm\(^{-1}\) greater than its nearest replicate W3-4, most replicate values of \( \alpha \) are close. Since differences in the hydraulic properties of soils should be less for replicates at a given location than between locations, the Table 4 groupings suggest that \( \alpha \)
TABLE 4. ORDER BY HYDRAULIC CONDUCTIVITY PARAMETER $\alpha$
OF SEVENTEEN OXISOL SOIL SITES, O'AHU

<table>
<thead>
<tr>
<th>$\alpha$ (cm$^{-1}$)</th>
<th>Molokai</th>
<th>Lahaina</th>
<th>Wahiawa</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0232 M4-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0267 M1-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0268 M4-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0283 L2-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0313 M4-1</td>
<td></td>
<td>L2-2</td>
<td></td>
</tr>
<tr>
<td>0.0315 M4-1</td>
<td></td>
<td>L2-1</td>
<td></td>
</tr>
<tr>
<td>0.0367 M1-2</td>
<td></td>
<td></td>
<td>W3-3</td>
</tr>
<tr>
<td>0.0378 M1-2</td>
<td></td>
<td></td>
<td>W3-4</td>
</tr>
<tr>
<td>0.0427 M3-1</td>
<td></td>
<td>L3-2</td>
<td></td>
</tr>
<tr>
<td>0.0444 M3-2</td>
<td></td>
<td></td>
<td>W3-2</td>
</tr>
<tr>
<td>0.0510 L1-1</td>
<td></td>
<td></td>
<td>W1-1</td>
</tr>
<tr>
<td>0.0517 L1-2</td>
<td></td>
<td></td>
<td>W1-2</td>
</tr>
<tr>
<td>0.0629 L1-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0638 W3-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0785 W3-4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Eq. (1) of Table 1.

may provide the discrimination that is desirable in a hydrologic index for soils. Thus, all four criteria stated earlier seem to be satisfied by equation (1).

A similar assessment of $\eta$ and $\gamma$ in equations (2) and (3), respectively (see Tables 1, 2), reveals that these parameters are less correlated with soil sites and are poorly ordered in comparison with $\alpha$, and thus may not be as desirable as $\alpha$ as hydraulic conductivity indices. This conclusion is based on too limited a data set, however, to be considered valid for all soils. The examination of replicated conductivity data for a broader range of soils would be worthwhile to determine whether $\alpha$ would be of more universal value as an index for assessing the differences between soils.

Another possibility is the combination of $K_s$ and $\alpha$ incorporated into a single index value. Theoretical justification for such a combination exists in that the substitution of equation (1) into (4) and integration between the reference pressure $h_0$ (e.g., $h_0 = -1000$ cm water) and the upper limit, $h = 0$, yields a value for the matrix flux potential, $\phi(h)$, of $\phi(0) = K_s/\alpha$. This combination index, however, is slightly dominated by variation in $K_s$, as $K_s$ values in Table 3 have a range of 5.5-fold, while $\alpha$ values vary by only 3.4-fold; thus, $K_s/\alpha$ is ordered among the seventeen Oxisols soil sites...
only slightly better than is $K_s$.

Although the parameters $\eta$ and $Y$ in equations (2) and (3) do not appear to be as useful as $\alpha$ to represent the hydraulic conductivity of the soils included in this study, equations (2) and (3), the power functions, have some merits which are worth considering. We mentioned previously that the method of Chong, Green, and Ahuja (1981) gives $K$ as a power function of either $h$ or $\theta$. Thus, if this method is used to determine $K(h)$ or $K(\theta)$, the superior fit of equations (2) and (3) may be an important consideration. Also, the power functions have been used in numerous studies of flow in coarse-textured materials by A.T. Corey and associates, such as Brooks and Corey (1964) and in studies of spatial variability of hydraulic properties by Russo and Bresler (1980a, 1980b), so that an extensive body of literature exists supporting the use of these functions. Additionally, an equation similar to equation (3) has been effectively used by Campbell (1974) in a method which derives $K(\theta)$ from the soil-water retention curve of a soil. The equation $K(\theta) = K_s (\theta/\theta_s)^{2b+3}$, which resulted from Campbell's analysis, is equivalent to equation (3) with $\theta_p = 0$ and $Y = 2b + 3$. Our experience suggests that if one is interested in the whole range of $K(\theta)$ between $\theta_s$ and $\theta_p$, then the power functions (eqn. [2], [3]) are superior. On the other hand, for the cases in which the soil remains close to saturation (e.g., during intermittent, high-frequency irrigation), equation (1) may be preferred. Equations (2) and (3) may also be superior for theoretical analyses of water flow and solute transport in spatially variable soils (Dagan and Bresler 1979; Bresler and Dagan 1981).

Consider the problems involved in characterizing the hydraulic properties of soils in large areas of land, portions of which are inaccessible for conventional vehicles and which may even lack a supply of water for field measurement of $K$. The field method of Chong, Green, and Ahuja (1981) would be advantageous for relatively level, accessible areas; however, Campbell's calculation method is preferable for inaccessible areas. Since both methods utilize power functions to describe $K(\theta)$, the resulting coefficients can be equated to express all of the data with the same parameters.
Physically based deterministic conceptual approaches have been traditionally used in mathematically modeling. In the application of these models to actual situations, the watershed, or a specific hill slope within its boundaries, is generally assumed to be a fictitious homogeneous medium under deterministic or, in some cases stochastic, storm and rainfall events. In the conceptual modeling of watershed processes, stochastic components are necessary because real watershed are generally heterogeneous, and the effective hydraulic properties that govern flow mechanisms cannot always be defined or are meaningless in a spatially variable watershed (Bresler and Dagan 1983). In addition, spatial heterogeneity of the field is generally coupled with a stochastic pattern of rainfall intensity and duration in time and in space.

Because of the stochastic nature of both land and meteorological properties in a given watershed, a tremendous amount of data must be collected to estimate heterogeneity and uncertainty in terms of the frequency distribution of its parameters, amount of variability, and spatial structure. The numerical data must be processed to provide information and to estimate the various statistical parameters and the structural features of the variables pertinent to stochastic conceptual watershed modeling. The model is generally composed of a deterministic component, a correlated component, and a random component, or of a superimposition of a correlated component on a random component. The sample size and location of the observation points are determined by the exact model, its parametric data, allocation of resources, and expected accuracy.

To determine the model parameters, samples are taken, their relevant properties analyzed, and the model parameters inferred by statistical methods. In the simplistic and classical statistical approach, the sample values in the watershed or hill-slope domain are assumed to have an equal probability of being selected. The fact that two samples close to each other are more likely to have closer values than if they are far apart, is not considered. On the other hand, the sample values in the spatial statistical approach are considered to be realizations of random functions; thus, because the value of a sample is also a function of its position in the watershed, the exact location of the sampling is also considered. The simi-
larity between sample values is quantified as a function of the distance between the samples. In the classical statistical approach, the average values and the variances of the parameters of interest can be estimated from the samples. The spatial statistical or the geostatistical approach is used to estimate the similarity between values as a function of distance or, in other words, to estimate the spatial structure parameters from the values of the sample and from the locations of the observation points.

The purpose of this section of the report is to propose an independent random sampling scheme of determining the statistical parameters required for watershed modeling. The stochastic-conceptual model of rainfall-runoff processes on a hill slope is used to characterize the statistical parameters. A sampling scheme is then designed to estimate the needed parameters for the watershed that contributes runoff to West Loch, Pearl Harbor (Green et al. 1977). The results are given in terms of location and number of sampling points for these selections by using various criteria. The value of additional sampling points and their preferred locations are also determined, and the "best" sampling scheme in the watershed indicated.

STATISTICAL CONCEPTS

Soil hydraulic properties and hydrological conditions at a given depth \( z \) of the watershed and at time \( t \) are functions of the two-dimensional space coordinates \( x,y \) in a horizontal plane. These watershed variables may vary in an irregular manner in space; therefore, the prediction of their values at a point in the watershed on the basis of measurements at other points in the same watershed is subject to uncertainty. The space function watershed variables are regarded as random variables of the space coordinates \( x,y \) or, for brevity, of the position vector \( \mathbf{r} \) of components \( x,y \). Denoting by \( u \) any hydrologic and soil hydraulic variable, a "sufficient" statistical characterization of \( u \) at a set of points \( \mathbf{r}_i \) (\( i = 1, 2, \ldots, N \)) in the watershed is defined by the probability density function (PDF), \( f(u) \). In the particular case of a stationary (statistically homogeneous) and isotropic process, the autocovariance \( C_{uu} \) of \( u \) can be defined by

\[
C_{uu}(\mathbf{r}_1, \mathbf{r}_2) = \sigma_u^2 \rho_{uu}(r_{1,2})
\]  

(5)
where $\sigma^2_u$ is the constant variance of $u$, $\rho_{uu} \leq 1$ is the autocorrelation coefficient, and $\tau$ is the lag defined by

$$\tau_{1,2} = |r_1 - r_2|.$$  

The autocorrelation $\rho_{uu}$ has a value of $\rho_{uu}(0) = 1$ and it is assumed that $\rho_{uu} \rightarrow 0$ for sufficiently large $\tau$. When $\rho_{uu} = 0$, the variables at points $r_1$ and $r_2$ are uncorrelated.

In a one-dimensional process, e.g., time series, the integral scale of $u$ (designated $J_u$), which roughly characterizes $\rho_{uu}$, is defined by Lumley and Panofsky (1964) as

$$J_u = \int_0^\infty \rho_{uu}(\tau) \, d\tau.$$  

Referring to our two-dimensional stationary and isotropic process $u$, the integral scale which can be interpreted as the space scale of $u$ in the watershed is similarly defined by

$$J_u = [2 \int_0^\infty \rho_{uu}(\tau) \, \tau \, d\tau]^{1/2}.$$  

Another point of principle is the exchange between ensemble averages ($u$, $\sigma^2_u$) and space averages ($\tilde{u}$, in eq. [16]; $S^2$, in eq. [17]) which are determined from a single realization. By the "ergodic" assumption, ensemble and space averages are equal. This equality is valid, provided that $u$ is stationary and the equivalent length scale of the field $R$ is much larger than the integral scale $J_u$. If $u$ is sampled on a finite field, then $\tilde{u}$ is subject to an estimated error which depends on $\sigma^2_u$, $\mu_u$ and on the ratio $J/R$ (eq. [30]).

The autocovariance of the stationary random function $u(x, y)$—characterized by its constant average—can also be expressed by the isotropic variogram of $u$ defined by

$$\gamma_u(\tau) = 1/2 \langle [u(r_1) - u(r_2)]^2 \rangle = \sigma^2_u - C_{uu}(r_1, r_2)$$  

where $\langle \rangle$ stands for average or expectation. In any case whenever $\sigma^2_u$ and $\rho_{uu}(\tau)$ exist, the variogram is uniquely determined by

$$\gamma_u(\tau) = \sigma^2_u \left[ 1 - \rho_{uu}(\tau) \right].$$  

Because a stationary variogram does not imply the existence of a "sill" at $\sigma^2$, the autocorrelation is not uniquely determined by equation (10) unless the variogram becomes constant for $\tau > \lambda$ (i.e., has a sill at $\tau > \lambda$). The range $\lambda$ corresponds to the separation distance $\tau$ for which $\rho_{uu}(\tau) = 0$ and, therefore, for $\tau > \lambda$ the variogram $\gamma_u = \sigma^2$. 
STOCHASTIC MODEL UTILIZING STATISTICAL CONCEPTS

To calculate streamflow hydrographs and aquifer recharge in the watershed, the watershed (Fig. 1) which contributes runoff to West Loch of Pearl Harbor is considered as a hill-slope unit. The recharge rate and streamflow hydrographs can be calculated in the watershed and for each storm event similar to Freeze (1980). To achieve these goals, Monte Carlo simulation experiments are used. These experiments involve the generation of three classes of parameters and conditions: (1) time-independent watershed parameters, such as soil hydraulic properties, topographic elevation, and surface runoff travel time; (2) external storm properties and internal rainfall intensity pattern; and (3) initial soil and hydrologic conditions, such as initial soil-water content, depth of the unsaturated zone, and water table elevation. Following the simulation of any number of events, a statistical analysis is carried out to calculate streamflow hydrographs and groundwater recharge.

The core of the stochastic model is the generation of a two-dimensional field of autocorrelated parameter values. Any parameter or a variable \( u \) to be generated is defined at a point \((x_i, y_j)\) in the horizontal \((x, y)\) plane by \( u_{ij} \) for \( i = 0, 1, \ldots, I; j = 1, 2, \ldots, J \). The value of \( u_{ij} \) is selected from a normally distributed population with mean \( \mu_u \) and variance \( \sigma_u^2 \). In addition, the two-dimensional process that controls the spatial distribution of \( u \) was assumed to be stationary and isotropic, with exponential type autocorrelation structure given by

\[
\rho_{uu}(\tau) = \exp(-\alpha_u |\tau|) \quad (11)
\]

where \( \alpha_u \) may be called the "autocorrelation parameter" and the lag \( \tau \) between any two points \((x_i, y_j)\) and \((x_k, y_l)\) for \( i, k = 1, 2, \ldots, I; j, l = 1, 2, \ldots, J \) is calculated from

\[
\tau = [(x_i - x_k)^2 + (y_j - y_l)^2]^{\frac{1}{2}} \quad (12)
\]

Although equation (11) is a very popular relationship for the correlation structure, a few other types of the function \( \rho_{uu}(\tau) \) have been adopted in the literature. Among those are the linear function (Smith and Freeze 1979), spherical, Gaussian, and monomial (Delhomme 1978).

The value of any spatial variable \( u \) at a point \((x_i, y_j)\) is generated as a three-parameter representation of the normal distribution, i.e., \( u \sim \mathcal{N} (\mu_u, \sigma_u, \alpha_u) \).
Figure 1. Schematic of watershed grid showing the sampling points generating domain \((X = Y = 20 \text{ km})\) and boundary of actual watershed domain, O'ahu, Hawai'i.
The point value of $u_{i,j}$ at $(x_i, y_j)$ is obtained from a set of residuals $\varepsilon_{i,j}$, which are taken from a standard normal population $\varepsilon \sim N(0, 1; \sigma_u)$, from

$$u_{i,j} = u + \varepsilon_{i,j} \sigma_u \cdot$$

The algorithm used by Freeze (1980) to generate $\varepsilon_{i,j}$ is based on that of Mejia and Rodriguez-Iturbe (1974). For $N \gg 50$, $\varepsilon_{i,j}$ is calculated from

$$\varepsilon_{i,j} = \sqrt{(2/N)} \sum_{m=1}^{N} \cos[w_m(x_i \sin \beta_m + y_i \cos \beta_m + \phi_m)$$

where $\beta_m$ and $\phi_m$ are taken from a uniform distribution population over the range 0 to $2\pi$ (i.e., $\beta, \phi \sim U[0, 2\pi]$) and $w_m$ is related to the autocorrelation function by

$$w_m = \alpha_u \left[1/(1 - G_m)^2 - 1\right]^{1/2}$$

with $G_m$ taken from $G_m \sim U(0,1)$. Note that any generated set of $u_{i,j}$ values $(i = 1, 2, \ldots, I; j = 1, 2, \ldots, J)$ constitutes one realization of the stochastic process $u \sim N(\mu_u, \sigma_u; \alpha_u)$ from which a sample mean $\bar{u}$, a sample standard deviation $S_u$ and a sample autocorrelation parameter $\alpha_u$ can be calculated. A few alternative methods for generating two dimensional fields of autocorrelated parameters are available (Freeze 1980; Dagan 1982). For example, assuming that the variable of interest $u$ with values $u(x_i)$, $i = 1, 2, \ldots, N$ has an $N$-variate normal PDF, then $\varepsilon_{i,j}$ can be generated by utilizing the ready-made $N$-variate normal generator program using the autocovariance matrix calculated from equation (5) using (11) and (12) or estimated from (18).

**ESTIMATIONS OF STATISTICAL PARAMETERS**

To generate at the $(x_i, y_j)$ points a realization of values derived from a stochastic process of a normally distributed spatial variable $u$, with an autocorrelated parameter value $\alpha_u$ (eq. [11]), an estimate is necessary of the two first statistical moments $\mu_u$, $\sigma_u$, and the autocorrelation parameter $\alpha_u$. The sample mean and variance of one realization (a watershed) are simply estimated from one realization field data by

$$\bar{u} = \frac{1}{N} \sum_{i=1}^{N} u_i$$

$$S_u^2 = \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \bar{u})^2$$

with $\alpha_u$ calculated from the autocorrelation function $\alpha_u = \frac{1}{(1 - G_m)^2 - 1}$. A few alternative methods for estimating $\alpha_u$ are available (Freeze 1980; Dagan 1982).
and

\[ S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \bar{u})^2 \]  \hspace{1cm} (17)

where \( N \) (sample size) is the number of observation (sampling) points in that realization. The autocorrelation coefficient estimate \( \hat{\rho}_{uu}(\tau) \) is calculated as a function of the lag, \( \tau \), from equation (5) with estimated values of the autocovariance \( \hat{C}_{uu} \) calculated by

\[ \hat{C}_{uu}(\tau) = \frac{1}{n(\tau) - 1} \sum_{m=1}^{n(\tau)} [u(r_m) - \bar{u}_1][u(r_m + \tau) - \bar{u}_2] \]  \hspace{1cm} (18)

and estimated values of the variance calculated from

\[ S_1^2 = \frac{1}{n(\tau) - 1} \sum_{m=1}^{n(\tau)} [u(r_m) - \bar{u}_1]^2 \]  \hspace{1cm} (19)

and

\[ S_2^2 = \frac{1}{n(\tau) - 1} \sum_{m=1}^{n(\tau)} [u(r_m + \tau) - \bar{u}_2]^2 \]  \hspace{1cm} (20)

Here \( n(\tau) \) is the number of pairs of observation points (samples) separated by a distance \( \tau \) in the watershed domain, \( u(r_m) \) is the value of \( u \) at the location \( r_m \), and \( u(r_m + \tau) \) is the value of \( u \) at the position \( r_m + \tau \), where \( r_m \) and \( r_m + \tau \) are the positions of the two samples in the \( m \)th pair. The averages \( \bar{u}_1 \) and \( \bar{u}_2 \) in equations (18) through (20) are defined by

\[ \bar{u}_1 = \frac{1}{n(\tau)} \sum_{m=1}^{n(\tau)} u(r_m) \]  \hspace{1cm} (21)

and

\[ \bar{u}_2 = \frac{1}{n(\tau)} \sum_{m=1}^{n(\tau)} u(r_m + \tau) \]  \hspace{1cm} (22)

To calculate the quantities given in equations (16) through (22) to obtain the best estimates of \( \mu_u \), \( \sigma_u \), and \( \rho_{uu}(\tau) \) or \( \sigma_u \), which are essential for the stochastic model, the location and number of sampling points from a single watershed realization must be selected in such a way that a few criteria are met. As no information is available regarding the variability and integral scale of each parameter \( u \) in the finite watershed, the practical number of sampling points for a given acceptable error in estimating \( \mu_u \) (Russo and Bresler 1982, eq. [7]) could not be estimated. Obviously, the larger the number of observation points, the better \( \mu_u \) and \( \sigma_u \) are respectively estimated by \( \bar{u} \) and \( S \). To estimate the values of \( \sigma_u \) from equation (11) or
to estimate $\rho_{uu}(\tau)$ from any other type of relationship, the autocorrelation coefficient estimate as a function of $\tau$ is calculated from

$$r_{uu}(\tau) = \hat{C}_{uu}(\tau)/[S_1(r_1) S_2(r_2)]$$

(23)

where $\tau$ is obtained from equation (12). It is probable that the best estimate of $\alpha_u$ from $\rho_{uu}(\tau)$ is obtained from equations (18) through (23) for the case in which $n(\tau)$ is uniformly distributed for sufficiently large $n$, and that when the range of $\tau$ from $\tau \to 0$ to $\tau \to R$ is maximum. The problem of a sampling scheme design is therefore boiled down to the question of how to best divide evenly the equivalent field length $R$ into a sufficiently large number of lag classes, each of which have an approximately equal number of pairs of sampling points. The lag $\tau$ of the first (smallest $\tau$) class should be as small as possible, while the largest one would best be as close to $R$ as possible. Obviously, the sampling points using the above-mentioned criteria must be selected without violating any sampling principles in the statistical sense.

Randomness seems to be a desirable requirement, or criterion, for sampling in order to evaluate unconditional statistics, especially when information regarding the statistical structure is not available. The term random indicates that each point in the $x,y$ plane has an equal chance of being selected as a member of the sample. This also insures that the laws of probability and their mathematical treatment apply. A personal bias which is caused by too subjective sampling is also prevented. Subjective judgement is needed, however, to select from a set of random samples the one which permits best estimates of $\rho_{uu}$ or of $\alpha_u$, i.e., sufficiently large $n(\tau)$ and equal weight to each $\tau$ to ensure uniform distribution of $n(\tau)$. In addition, if the other alternatives to equation (11) are considered, then to decide which is the best fitted equation to $r_{uu}(\tau)$—linear, spherical, Gaussian, monomial, or exponential (eq. [11])—the behavior of the autocorrelation function close to the origin ($\tau \to 0$) and its behavior at large lags, i.e., $\tau \to R$, are important. Time, personnel, and budget are, in many circumstances, limited; therefore, the number of sampling points must always be as small as possible.

It should be noted here that the sampling scheme which best fits the estimation of the autocorrelation function may also serve as the best scheme for the estimation of the variogram from

$$\hat{\gamma}(\tau) = \frac{1}{2\gamma(\tau)} \frac{n(\tau)}{\sum_{m=1}^{\gamma(\tau)} [u(r_m) - u(r_m + \tau)]^2}. $$

(24)
The variogram thus calculated may be used as an alternative method for generating two-dimensional fields of autocorrelated parameters, such as the simple simulation method described by Dagan (1982) and the turning-bands approach described by Journel (1974). An estimate of the variogram is also needed for improving the locations of the sampling points using conditional simulation techniques, as will be discussed later.

SELECTING LOCATIONS OF RANDOM SAMPLING POINTS
FOR AUTOCORRELATION AND VARIOGRAM ESTIMATES

To estimate the statistical parameters (eqs. [16] through [24]) for unconditional statistics, an optimal setting of sampling points must be obtained. The limited purpose of this section is simply to select a distribution of \((x_i, y_i)\) points which yield a uniform distribution of lags in a given interval between \(\tau_{\text{min}}\) (i.e., \(\tau \to 0\)) and \(\tau_{\text{max}}\) (i.e., \(\tau \to R\)). The problem may have a unique solution, or a class of exact solutions, but only for a regular setting of points. An empirical Monte Carlo simulation method for the problem of a random sampling setting is proposed in this section. In the following section an improvement of the sampling setting selection is proposed by employing conditional statistics techniques.

The population from which the random samples are drawn consists of all the points of the surface plane \((x, y; z = 0)\) in West Loch of the Pearl Harbor watershed (Fig. 1). The autocorrelation and the variogram are then functions of the distance between any pair of points and the direction along its two components. For simplicity and consistency with Freeze's (1980) model, the isotropic case in which the autocorrelation and the variogram are functions of the modulus \(|\tau|\) are considered. The simulation procedure for selecting the locations of the sample observation points for the "best" estimate of \(\gamma(\tau)\) and \(\rho(\tau)\) in the isotropic case is outlined below.

Procedure

The general procedure is to draw from a random number generation realizations of the coordinate \((x_i, y_j)\), for \(i = 1, 2, \ldots, I; j = 1, 2, \ldots, J, I = J > N\), in a square domain with sides \(X = Y\) which include the watershed boundaries (Fig. 1). The square domain is needed to prevent anisotropy
which may be induced by the sampling selection technique in the case of unevenness of \( X \) and \( Y \) (see steps 2 through 4 below). A sample of size \( N \), consisting of those values of the coordinates \((x_i, y_j)\) which reside within the boundaries of the watershed domain, is selected in each realization. The sample size is allowed to change between its minimum value, \( N = N_{\text{min}} \), and its maximum value, \( N = N_{\text{max}} \), with intermediate values of \( N \) between \( N_{\text{max}} \) and \( N_{\text{min}} \). For each sample \( N \), the number of pairs of points \( n(\tau) \) in each class interval \( \tau + \Delta\tau \) is counted. The mean square deviation of number of pairs per class from average number of pairs per class is calculated for each realization and each sample size from

\[
\text{MSD} = \frac{\sum_{k=1}^{K} (n_k - \bar{n})^2}{\bar{n}^2 (25)}
\]

where \( k \) is class lag or distance index, \( K \) is total number of lag or distance classes, \( \bar{n} \) is average number of pairs in a class (eq. [28]) and \( n_k \) is number of pairs in the \( k \)th class. Ideally, \( \text{MSD} = 0 \) in the case that has the same number of pairs in all distance (lag) classes. Unfortunately, this ideal case of uniform distribution of lags in the interval between \( \tau_{\text{min}} \) and \( \tau_{\text{max}} \) cannot be obtained. One can, however, repeat the procedure and select the best realization with the lowest \( \text{MSD} \) in the entire simulated ensemble.

The detailed simulation procedure for selecting the location of \( N \) observation points in the watershed, the boundaries of which are given in Figure 1, consists of the following steps.

1. Let the origin \((0, 0)\) be at the lowest left corner, leaving small margins outside the watershed, and the points \((X, Y)\) be at the point of max \((x\) or \(y)\), leaving margins, within the boundaries of the watershed. Set \( Y = \max (x) \), as in Figure 1, or \( Y = \max (y) \) if \( \max (y) > \max (x) \), and then set \( X = Y \) (Fig. 1) to prevent anisotropy induced by the sampling selection technique due to the unevenness of \( X \) and \( Y \) (see steps 2-4).

2. Draw a number from \( U (0, 1) \) and multiply it by \( X \) to obtain \( x_1 \), a temporary coordinate value of point No. 1 in the \( x \) direction. Repeat to get \( x_2, x_3, \ldots, x_i, \ldots, x_I \).

3. Repeat step 2 but multiply each random number drawn from \( U (0, 1) \) by \( Y \) to get \( y_1, y_2, \ldots, y_i, \ldots, y_I \), temporary coordinate values of the points 1, 2, \ldots, \( i \), \ldots, \( I \) in the \( y \) direction.

4. Combine the results of steps 2 and 3 to obtain the \((x, y)\) coordi-
nates of a large number of temporary points 1, 2, ..., as \((x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_I, y_I)\). The term temporary is used here to distinguish between the points \((x, y)\) which are in the \(X, Y\) domain and those points which are within the boundaries of the watershed (Fig. 1).

5. Check if the coordinates \((x_i, y_i), i = 1, 2, \ldots, I\) are within the boundaries of the watershed domain given in Figure 1. Identify as \((x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_L, y_L)\) coordinates of the new ensemble of random points (with \(L = J N_{\text{max}}\), and \(J\) is number of realizations in the simulated ensemble) within the boundaries of the watershed domain. The locations of the actual \(N_{\text{min}} \leq N \leq N_{\text{max}}\) observation points in each realization are now obtained from this ensemble.

6. Set the realization index to \(j = 1\) (i.e., first realization); the space coordinate index to \(i = 0\), and the sampling size index to \(p = 0\). Set also \(\Delta N\) as an arbitrary increment of the number of observation points so that the actual sample size \((N)\) is given by \(N = N_{\text{min}} + p\Delta N\).

7. Set the sample size \(N = N_{\text{min}} + p\Delta N\) and then the counting index \(L = j \cdot N\) and identify the points with the coordinates \((x_{i+1}, y_{i+1}), (x_{i+2}, y_{i+2}), \ldots, (x_{i+L}, y_{i+L})\). Note that the index \(L = N, N + 1, \ldots, L\) counts all the points in the ensemble.

8. Calculate total number of pairs of points in each realization from

\[
M = N \cdot (N - 1)/2. \tag{26}
\]

9. Calculate \(\tau_m, m = 1, 2, \ldots, M\) from equation (12), and arrange the \(M\) values of \(\tau_m\) into \(K\) lag classes, with class interval \(\Delta \tau\), determined from the distance

\[
\Delta \tau = \left\{\left[\max(x_i)\right]^2 + \left[\max(y_i)\right]^2\right\}^{1/2}/K \tag{27}
\]

Note that the lag of the distance \(\tau\) is given by \(\tau_k = k\Delta \tau\) for \(k = 1, 2, \ldots, K\).

10. Count \(n_k\) the number of pairs of points in the \(k\)th class for \(k = 1, 2, \ldots, K\).

11. Calculate average number of pairs in a class by

\[
\bar{n} = M/K. \tag{28}
\]

12. Calculate mean square deviation of number of pairs per class \((n_k)\) from average number of pairs per class \((\bar{n})\) by equation (25). Note
that for the ideal but unobtainable case that has the same number of pairs in all distance classes, the value of MSD is zero.

13. Compare \( N \) with \( N_{\text{max}} \). If \( N \) is larger than \( N_{\text{max}} \), go to the next step (step 14); otherwise (\( N \) smaller than \( N_{\text{max}} \)) set \( i = \ell \) and \( p = 1 \) and return to step 7 above to calculate new \( N \) with the new value of \( p \) and to calculate the new \( \ell \) with the new value of \( N \). Repeat with \( p = 1, 2, \ldots, P \) when \( P = (N_{\text{max}} - N_{\text{min}})/\Delta N + 1 \).

14. Set \( j = 2 \) (the second realization), \( p = 0 \) and \( i = \ell \) and return to step 7 above to calculate new \( N \) with the new value of \( p \) and to calculate new \( \ell \) with the new values of \( j \) and \( N \). Repeat for \( j = 3, 4, \ldots, J \) where \( J \) is the total number of realizations in the ensemble. To decide the value of \( J \) go to next step (step 15).

15. Calculate the average of MSD for all samplings from

\[
\overline{\text{MSD}}_j = \frac{1}{J} \sum_{j=1}^{J} \text{MSD}_j
\]

for \( j = 1, 2, \ldots, J \). Compare \( \overline{\text{MSD}}_j \) with \( \overline{\text{MSD}}_{j-1} \) and with \( \overline{\text{MSD}}_{j-2} \). If the differences are within a tolerance error, stop; otherwise return to step 7 with \( j = J + 1 \), \( p = 0 \), and \( i = \ell \).

It should be pointed out here that the main purpose of step 13 above is to obtain samples of sizes varying in the range between \( N_{\text{min}} \) and \( N_{\text{max}} \) for the same simulated realization. The procedure enables one to use smaller sample sizes for variables requiring more manpower or other resources for their determination and larger sample sizes for variables which cost less to measure. But still all the variables consist of the same realizations which have been drawn from the same population.

RESULTS OF RANDOM SAMPLING POINT LOCATIONS

Results of the locations of the random sampling points obtained by the previously described procedure are demonstrated here for the watershed contributing runoff to West Loch, Pearl Harbor. The relevant watershed domain with its outside margins and size along the \( x \) and \( y \) coordinates is shown on the map (Fig. 1). Thirty lag (distance) classes \((K = 30)\) have been chosen to demonstrate the results since a range of \( K \) between 25 and 60 gave essentially the same results. The distribution of number of pairs of points per
lag class \( n(\tau) \) is obviously more uniform as the number of lags \( (K) \) is decreased (uniformity of \( n(\tau) \) distribution is increased as \( K \) is decreased). However, if the number of classes of pair of points \( n(\tau) \) is too small, i.e., there are too few distance classes, an accurate estimate of \( \alpha \) from equation (11) by the best-fit function \( \rho_{\mu \alpha}(\tau) \) or any other accurate descriptions of \( \rho_{\mu \alpha}(\tau) \) may not be possible.

Results of mean-square deviation (MSD) of number of pairs as computed from equation (25) with \( K = 30 \) for 27 realizations (sampling runs) and six groups \( (P = 9, \text{but only six groups are given in the examples}) \) of number of observation points (sample sizes \( N = 10, 20, 30, 50, 70, 90 \text{ with } N_{\min} = 10 \text{ and } N_{\max} = 90 \) ) are given in Figure 2. Note in Figure 2 that the ordinate

![Figure 2](image_url)

**Figure 2.** Mean square deviation computed from eq. (25) for 27 realizations (runs) and 6 sample sizes \( N \).
(MSD scale) is changed for MSD >40). As one would expect, the group of $N = 10$ sample points has, without exception, much larger MSD than all the other $N$ groups. The next nonuniform $n(\tau)$ distribution is calculated for the group of sample size $N = 20$ sampling points. Here the calculated MSD is, with a few exceptions, larger than the $N = 30$ group which is then in turn, with more exceptions, larger than the sample size $N = 50$ group. The order of MSD is not as clear for the rest of the $p$ groups. It should be noted, however, that in some cases (e.g., run no. 22) the order of MSD is in the opposite direction (except for $N = 10$). The picture is clearer in Figure 3 where average MSD values (MSD), calculated from equation (29) are plotted as a function of realization (run) no. $j = 1, 2, \ldots, J$ with $J = 27$. Here the order of decreasing MSD coincides with the order of increasing number of sampling points $N$ from 10 to 90. Also, the differences of MSD are relatively large when $N$ increases from 10 to 30, after which they are quite

![Figure 3. Average (cumulative) mean square deviation of 1 through 27 realizations (runs)](image-url)
small. In Figure 3 the ordinate scale is also changed for \( \text{MSD} > 40 \) because \( \text{MSD} \) for \( N = 10 \) is tremendously larger than for \( N > 20 \).

Although 27 runs may be an insufficient number to represent the entire ensemble of samples of size \( N \) from which the "best" realization has to be selected, it was decided to stop at this run because \( \text{MSD} \) did not change much beyond run no. 22. Based on the data of the 27 runs given in Figure 2, run no. 16 was selected as the best realization representing the locations of the sampling points of any size \( N \) in the Pearl Harbor watershed within the boundaries given in Figure 1. Run no. 20 was selected as the realization representing the "worst" situation of the ensemble of all possible samples of size \( N \), for \( N \) ranging between \( N_{\text{min}} = 10 \) and \( N_{\text{max}} = 90 \).

The solid line of Figure 4 presents "ensemble" average \( \text{MSD}_J \) of the \( J = 27 \) realizations (\( \text{MSD}_{27} \)) as a function of number of sampling points (sample size \( N \)). It can be seen that \( \text{MSD} \) drops sharply when sample size increases from 10 to 20 observation points, then decreases gradually when \( N \) changes from 20 to 30, with very small decrease of \( \text{MSD} \) when \( N \) further increases from 30 to 90. Similar results are observed in Figure 4 for run no. 16 (best realization [dashed line]) and run no. 20 (worst realization [dot-dashed line]).

---

**Figure 4.** Mean square deviation as a function of sample size.
line]. Thus, if MSD is a good representation of the distribution uniformity of \( n(\tau) \), then it seems that \( N = 30 \) is the minimum sample size that meets the criteria of uniform \( n(\tau) \) distribution. Obviously, a larger number of observation points would give greater confidence in the results and lower errors in estimating means of the necessary hydraulic properties; it would also provide a larger number of pairs in each of the \( K \) lag classes.

A similar conclusion can be drawn from Figure 5 which gives for the \( J = 27 \) realizations the averages of maximum and minimum lags \( (\tau_{\text{max}} \text{ and } \tau_{\text{min}} \) in terms of their values relative to equivalent field length \( R \)) calculated from equation (12) and plotted as a function of \( N \) by the respective solid and dashed lines. Since the behavior of \( \rho(\tau) \) and \( \gamma(\tau) \) near the origin (i.e., at \( \tau \rightarrow 0 \)) and at large \( \tau \) (i.e., at \( \tau \rightarrow R \)), are important for purposes of best fitting equations, such as equation (11), it seems from Figure 5 that sample size \( N \) must be as large as possible but, again, should include at least 30 or more observation points.

Figures 6 and 7 give the frequency, i.e., the number of pairs of data points \( n(\tau) \) relative to the total number of pair of points \( (M) \), as a function of the relative lag \( (\tau) \), i.e., the lag relative to the equivalent watershed length. The data are plotted for the six sample sizes: (a) \( N = \)

![Figure 5. Average of 27 realizations of relative maximum lag \((\tau_{\text{max}}/R)\) and relative minimum lag \((\tau_{\text{min}}/R)\)](image)
NOTE: Sample sizes $N$ for (a) $N = 10$ ($M = 45$ pairs of points), (b) $N = 20$ ($M = 190$), (c) $N = 30$ ($M = 435$), (d) $N = 50$ ($M = 1225$), (e) $N = 70$ ($M = 2415$), (f) $N = 90$ ($M = 4005$).

NOTE: Dashed rectangles indicate uniform $n(t)$ distribution.

Figure 6. Frequency of number of pairs of points in best realization (run no. 16) as a function of relative lag
NOTE: Sample sizes N for (a) N = 10 (M = 45 pairs of points), (b) N = 20 (M = 190), (c) N = 30 (M = 435), (d) N = 50 (M = 1225), (e) N = 70 (M = 2415), (f) N = 90 (M = 4005).

NOTE: Dashed rectangles indicate uniform $\nu(T)$ distribution.

Figure 7. Frequency of number of pairs of points in worst realization (run no. 20) as a function of relative lag.
10 (so that \( M = 45 \)), (b) \( N = 20 \) \( (M = 190) \), (c) \( N = 30 \) \( (M = 435) \), (d) \( N = 50 \) \( (M = 1225) \), (e) \( N = 70 \) \( (M = 2415) \), and (f) \( N = 90 \) \( (M = 4005) \). Figure 6 presents the \( n(\tau) \) distribution data for the best realization (run no. 16), while in Figure 7 the data are plotted for the worst realization (run no. 20). Since the number of pairs \( n(\tau) \) in each lag class \( \tau \) has been obtained for the number of pairs in the interval of \( \tau \) between \( \tau = k\Delta \tau \) and \( \tau = (k + 1)\Delta \tau \), \( k = 1, 2, \ldots, K - 1 \), the exact points of \( \tau \) that have been inserted for \( \tau/R \) in Figures 6 and 7 were obtained by considering the relative weight of number of pairs of data points in each lag class \( \tau_k \) of the \( K = 30 \) classes. It should be remembered here that although the smaller number of classes \( (K < 30) \) tends to make \( n(\tau) \) more uniform, details would be eliminated in the \( \rho(\tau) \) function and would obscure the basic pattern of the data for the \( \rho(\tau) \) estimates. Too many classes \( (K > 30) \), on the other hand, may result in erratic patterns of alternating high and low values for the estimated \( \rho(\tau) \) function, which would yield a relatively small number of pairs \( n(\tau) \) in each \( \tau \) and would make the \( n(\tau) \) distribution less uniform.

For orientation purposes, the frequency of the ideal uniform distribution is drawn as dashed rectangles in Figures 6 and 7. Thus, for a completely uniform distribution of \( n(\tau) \), all the data points in Figures 6 and 7 would have to be concentrated along the uniform \( n(\tau) \) line extending from \( \tau/R \to 0 \) to \( \tau/R \to 1 \) (this dashed line at \( n(\tau)/M \) of 3.33 is indicated by the rectangles shown in Figs. 6, 7). Obviously, such an ideal situation would never be achieved, but it is much better approached by the data points given in Figure 6, representing the best sample realization, than those given in Figure 7 which represents the worst realization. Scrutiny of these figures also suggests that the differences between \( N = 50 \) to \( N = 90 \) (sample sizes [d] through [f], Figs. 6, 7) are negligibly small, despite the expectation that the larger the sample size \( (N) \)—and therefore, the total number of pairs of points \( (M) \)—the better the \( n(\tau) \) uniformity. Also, results for the case of \( N = 30 \) (c) do not differ much from \( N = 50 \), but differ considerably from \( N = 20 \) (b) and \( N = 10 \) (a). Since the data of Figures 6 and 7 are given in terms of frequencies, i.e., \( n(\tau) \), relative to the total number of pairs \( (M) \), the results of differences between (a) and (f) in the absolute values of \( n(\tau) \) are thus much more pronounced.

To visually compare the best to the worst realization and the two realizations to the ideal uniform \( n(\tau) \) distribution, the cumulative frequency of
NOTE: Dashed line representing uniform distribution is given for orientation.

Figure 8. Cumulative frequency of number of pairs of points \( n|\tau|/M \) in the best (run no. 16) and worst (run no. 20) realizations as a function of the relative lag \( \tau/R \).
n(τ)/M data are plotted in Figure 8 as a function of the relative lag (τ/R) for the same six sample sizes as in Figures 6 and 7. Figure 8 clearly demonstrates that although the best sample realization is far from ideal, it is a much better selection than the worst one. As indicated in Figure 8, a sample of size less than N = 30 is not, in any case, recommended. The actual size for N ≥ 30 depends on the tolerable level of error, which in turn depends on (1) the coefficient of variation of the property in question; (2) its integral scale; (3) the equivalent length of the field; (4) type of statistical distribution; and (5) the selected level of significance (Russo and Bresler 1982). The availability of manpower and budget to perform the measurements are of course additional practical considerations.

Schemes representing the watershed and the selected locations of the 20, 30, and 50 observation points for the best and the worst realization are respectively given in Figures 9 and 10. Here again the superiority of the best over the worst sample is visualized for any sample size. Also, a sample size of N = 30 seems to be the minimum size for spatial variability measurements.

IMPROVING THE SETTING OF OBSERVATION POINTS
BY CONDITIONAL STATISTICS

After the scheme for the locations of the N observation points has been established, the data on the hydraulic property of interest can be collected to estimate μu, σu, ρuu, and/or αu needed for the watershed simulation. The same data can also be used to calculate variograms from equation (24) and then to improve the location of additional observation points for conditional statistics by using the fictitious point method and the kriging technique, as suggested by Russo and Bresler (1982), or by any other conditional simulation method. This is a more complex case because it involves determination of the variance as a function of number of data points and their setting. Also, the variance of estimation of the variances has to be determined because conditional values depend on variances of the estimation of ρuu and on the statistical structure and setting of the observation points. Unfortunately, such data are not yet available for the present case of the Pearl Harbor watershed. However, to improve the points setting determined in the previous section, it is possible to generate the data of any
Figure 9. Schematic of best locations in watershed domain as generated in run no. 16 for $N = 20, 30, \text{ and } 50$.

Figure 10. Schematic of worst locations in watershed domain as generated in run no. 20 for $N = 20, 30, \text{ and } 50$. 
hydraulic parameter of interest at the observation points, such as those
given in Figures 9 and 10 by equations (13) through (15), or any other con-
ditional simulation methods (compare, for example, Dagan 1982).

To determine the "optimal" location of the additional observation
points, Delhomme's (1978) fictitious point method was employed. This method
makes use of the fact that the kriging variance of a spatially variable
field depends only on the spatial structure and not on the values of the
"additional" observation points. Since the kriging variance does not depend
on the actual values of the measurements, the generated values of the data
at the already selected sampling points (Figs. 9, 10) would not affect the
estimated kriging variance of the fictitious points to be added to the set
of observation points. The procedure is therefore to assume values for \( \alpha_u, \mu_u, \) and \( \sigma_u \) and to generate values for \( u \) at the already "existing" observation points \((x_i', y_i') \ i = 1, 2, ..., N \) (Figs. 9, 10). Considering the location of an additional sampling point \((x_{N+1}', y_{N+1}')\), the kriging system corresponding to the \( N + 1 \) sampling points \((x_i', y_i') \ i = 1, 2, ..., N + 1\), is constructed and the corresponding kriging variance is calculated. Another
additional fictitious point is located and the kriging variance considering
the \( N + 2 \) sampling points is calculated. The location of the additional
sampling points is determined according to the order of reducing the relative
kriging variance.

Although the kriging variance is independent of the data values at the
sampling points \((x_i', y_i') \ i = 1, 2, ..., N\). the error of the data generated
for each realization by equations (13) through (15) depends on the values of
\( \sigma_u, \mu_u, \) and \( \alpha_u \) used for the generation. The error, in terms of \( \sigma_u^2/\mu_u^2\) in-
volved in such an estimate for our finite watershed with equivalent length \( R \)
is given by Russo and Bresler (1982, eq. [7]) as

\[
\frac{(\sigma_u/\mu_u)^2}{\mu_u} = \left(\frac{\bar{u}}{S_u}\right)^2 \left[\frac{2\beta^2}{N} + 2J/R + 2Z_\beta \sqrt{2J/R}\right] \tag{30}
\]

where \( \sigma_u^2 \) and \( \mu_u \) are the values of the variances and means used to generate
the data, \( S_u \) and \( \bar{u} \) are the corresponding estimates (eqq. [16], [17]) from
the generated \( N \) data, \( Z_\beta \) is the value of \( Z \) in the standard normal distri-
bution curve at probability level \( \beta \), and \( J_u \) the integral scale which is re-
lated to \( \alpha_u \) through equations (8) and (11) by

\[
J_u = \left[2\int_0^\infty \tau \exp(-\alpha_u \tau) \, d\tau\right]^\frac{1}{2} = \frac{\sqrt{2}}{\alpha_u} . \tag{31}
\]

Equations (30) and (31) indicate that the error involved in each simu-
lation run depends (for a fixed $R$, $N$, and $\beta$) on the value of $\alpha_0$ which has been selected to simulate the data at the locations given in Figures 9 and 10. Thus, in simulating a wide range of $\sigma_0$, $\mu_0$, and $\alpha_0$, the number of runs to generate data for the same $\sigma_0/\mu_0$ ratio have been increased as the value of $\alpha_0$ becomes smaller.

Four sets of $\sigma_0$ and $\mu_0$ have been used to simulate the data to be used in the "improved location" technique. These were $\sigma^2 = 0.16$, $\mu = -0.1$ (equivalent to Bet Dagan soil); $\sigma^2 = 1.346$, $\mu = -0.616$ (Panoche soil); $\sigma^2 = 0.01$, $\mu = 1$ (to simulate small coefficient of variance, $CV = 0.1$); $\sigma^2 = 16$, $\mu = 1$ (to simulate large $CV = 4$). Each of these $\mu$ and $\sigma$ sets was combined with values of $\beta/R$ ranging from $2 \times 10^{-6}$ to 10 corresponding to the range of $\alpha_0$ between 400 and 0.01 and less. Two sets of $(x_i, y_i)$ coordinates with $N = 20$ and $N = 30$ (Figs. 9, 10) have been used to simulate the data of the arbitrary hydraulic parameters.

The locations of the sample points, in addition to those already existing, $N = 20$ or $N = 30$ (Figs. 9, 10), were selected as the points in the watershed which yield maximum reduction in the "estimated" relative kriging variance, taking into account the spatial structure of the simulated hydraulic property and the position of the additional points with respect to the sample points. Actual selection of an additional sampling point utilized the kriging technique coupled with the fictitious point method (Russo and Bresler 1982) and the simulated data of the $N = 20$ or $N = 30$ sample points. The exponential variograms were then estimated from equation (10) using (11) with the same values of $\alpha_0$ as those used to simulate the arbitrary hydraulic property data. By using this variogram, a solution of the kriging system with the $N$ "original" observation points (Figs. 9, 10) was obtained and then used to calculate the kriging variance of the two sets of additional 30 points as illustrated in Figures 11 and 12. The $(x_i, y_i)$ coordinate $i = N + 1$, $N + 2$, ..., $N_{\text{Final}}$ for which the relative kriging variance reduction was calculated was then selected as the first-choice location for the $N + 1$ point. This procedure was repeated for $N + 2$ (second choice), $N + 3$ (third choice), and ..., $N + 30$ points ($N_{\text{Final}}$ choice). The first set of points from which the $(x_i, y_i)$ coordinates $(i > N)$ of the additional locations were selected (Fig. 11) consists of points chosen arbitrarily in a somewhat regular grid. The second set from which the additional points were selected (Fig. 12) was chosen as one of the better runs (using the MSD criterion).
Figure 11. Watershed domain with arbitrary 30 regular points from which additional sampling locations were selected (see Fig. 14)

Figure 12. Watershed domain with 30 generated sampling points (run no. 12) from which additional observation points were selected (see Fig. 15)
given in Figure 2. For the points given in Figure 12, the coordinates of the observation points generated for run no. 12 for \( N = 30 \) points (Fig. 2) were chosen.

The order of preference for the locations of the additional sampling points is independent of \( \sigma_u \) and \( \mu_u \) when \( J/R \) was larger than 0.01. The dependence of the order of preference on \( \sigma_u, \mu_u, \) and \( J/R \) for smaller values of \( J/R \) is a result of lack of correlation so that the differences in the kriging variance between the additional points is negligible. This may also be deduced from equations (30) and (31). Actually, when \( J/R \) is sufficiently small, say less than about 0.01, the kriging variance is the same for all additional points (equal to the kriging variance for all the points given in Figs. 11, 12) so that the selection of the additional points is indifferent, that is none of the 30 points in Figures 11 and 12 is preferred over the other one.

To determine whether the simulation method of selecting the location of additional observation points can be used to improve the procedure of selecting the sampling point locations as described in the previous section,

### TABLE 5. VALUES OF MEAN SQUARE DEVIATION, \( \tau_{\text{max}}/R, \tau_{\text{min}}/R \) OF \( N + 10 \) OBSERVATION POINTS FOR \( N \) TAKEN FROM BEST AND WORST REALIZATIONS AND FROM TEN BEST ADDITIONAL POINTS

<table>
<thead>
<tr>
<th>RUN NO.</th>
<th>( N = 20 + 10 )</th>
<th></th>
<th>( N = 30 + 10 )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best Realization</td>
<td>Worst Realization</td>
<td>Best Realization</td>
<td>Worst Realization</td>
</tr>
<tr>
<td></td>
<td>( F. 11^* )</td>
<td>( F. 12)</td>
<td>( F. 11 )</td>
<td>( F. 12 )</td>
</tr>
<tr>
<td>MSD(^t)</td>
<td>12 9.2 6.9 10.3 16.7</td>
<td>9.0 9.9 23.7 14.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 6.4 7.8 20.7 16.8</td>
<td>7.6 8.4 20.5 19.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 --- --- 12.7 ---</td>
<td>--- --- --- ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau_{\text{max}}/R )</td>
<td>12 9.9 1.00 1.00 0.86</td>
<td>1.00 1.00 0.83 0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 9.9 0.98 0.9 0.87</td>
<td>1.00 1.00 0.97 0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 --- --- 0.87 ---</td>
<td>--- --- --- ---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau_{\text{min}}/R )</td>
<td>12 0.017 0.018 0.025 0.25</td>
<td>0.02 0.02 0.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 0.017 0.018 0.052 0.053</td>
<td>0.019 0.021 0.028 0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 --- --- 0.052 ---</td>
<td>--- --- 0.027</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figures 11 and 12 from which 10 best additional points were taken.
\(^t\)Mean square deviation.
the two methods were combined by adding to the \( N \) original points (Figs. 9, 10) the first preferred ten observation points from Figures 11 and 12 which have the largest relative kriging variance reduction. After combining the points, the sampling field contains a sample size of \( N + 10 \) observation points that can be judged using the MSD, \( \tau_{\min} \), and \( \tau_{\max} \) criteria. The results for \( N = 20 \) and \( N = 30 \) plus the ten additional points are given in Table 5.

A comparison between the data given in Table 5 and the corresponding values given in Figure 2 (run nos. 12, 16, 20), Figure 4, and Figure 5 suggests that the method of improving sampling selection with the generation of arbitrary data, slightly improved the design of the exact locations of

![Figure 13. Cumulative frequency of pairs of points as a function of relative lag for the best realization with \( N = 30 \) (solid line) compared to 20 points from Fig. 8 plus 10 first selection of additional points in Fig. 12 (x).](image)

NOTE: Dashed line represents uniform distribution for orientation.
the observation points. This can also be seen in Figure 13 where the cumulative frequency of \( n(\tau)/M \) in the improved sampling selection is plotted as a function of relative lag \( (\tau/R) \) and compared with equivalent data taken from Figure 8.

Using the criterion of uniform distribution of the pairs of observation points, Figures 14 and 15 may represent the final recommendation for the locations of the observation points (including the additional ten points) for a sample of size 30 for the 20 points shown in Figure 9 and the additional ten points from Figures 11 and 12, respectively. The order of selecting the additional points is also given in these figures.

**SUMMARY AND CONCLUSIONS**

Stochastic components in the modeling of a watershed are necessary because watersheds are generally heterogeneous, and the effective properties that govern flow mechanisms do not exist (Bresler and Dagan 1983). Because of the stochastic nature of watersheds, large amounts of data must be collected to estimate the necessary statistical parameters and the structural features of the hydraulic properties of interest. To estimate the spatial structure parameters by using the spatial statistical approach, the sample values and the locations and number of the observation points must be known. An independent method of random sampling is proposed for determination of the spatial statistical parameters required for stochastic watershed modeling. The limited statistical parameters are the first two central moments (mean and variance) and the autocorrelation function or the variogram. The sampling scheme criteria for selecting the location and the smallest possible number of the observation points to best estimate the statistical parameters are randomness, uniform distribution of a number of pairs of observation points \( n(\tau) \) separated by a distance or lag \( \tau \) in the watershed domain, sufficiently large \( n(\tau) \) for each class of \( \tau \), and enough pairs of points at small \( \tau \) (i.e., \( \tau \rightarrow 0 \)) and large \( \tau \) (i.e., \( \tau \rightarrow R \)), where \( R \) is the equivalent length of the watershed. An ensemble of \( J \) realizations of \( x, y \) coordinates of the observation points within the boundaries of a given watershed domain have been randomly simulated. By using the aforementioned criteria, the best realization of the observation-point coordinates of sample size \( N \) can be selected, and the exact location of the observation points for the sample
Figure 14. Watershed domain with locations of $N = 20 + 10$ sampling points of which 20 points were taken from run no. 16 (Fig. 9), and 10 are additional points from Fig. 11 in numerical order of selection of sampling points.

Figure 15. Watershed domain with locations of $N = 20 + 10$ sampling points of which 20 points were taken from run no. 16 (Fig. 9), and 10 are additional points from Fig. 12 in numerical order of selection of sampling points.
size $N$ improved by the kriging technique and the fictitious point method. The overall results suggest that 30 observation points are the minimum sample size for estimating the parameters necessary for stochastic modeling, and that a larger sample size is always recommended, especially for easy-to-determine variables. The method enables the adjustment of the sample size according to sampling ability with respect to budget, manpower, and time.

REFERENCES


