THE DESIGN OF RAINFALL NETWORKS IN TIME AND SPACE

by

Ignacio Rodríguez-Iturbe

and

José M. Mejía

MIT RALPH M. PARSONS LABORATORY
FOR WATER RESOURCES AND HYDRODYNAMICS

Report No. 176

Prepared under the support of
U.S. Department of Commerce, National Oceanic and Atmospheric Administration,
National Weather Service, Contract No. 2-36235
and with the cooperation of the Instituto Venezolano de Investigaciones Cientificas,
Caracas, Venezuela

November 1973

MIT

DEPARTMENT
OF
CIVIL
ENGINEERING

SCHOOL OF ENGINEERING
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Cambridge, Massachusetts 02139

R74-6

DSR 80495

RALPH M. PARSONS LABORATORY

FOR WATER RESOURCES AND HYDRODYNAMICS

Department of Civil Engineering

Massachusetts Institute of Technology

THE DESIGN OF RAINFALL NETWORKS IN TIME AND SPACE

by

Ignacio Rodríguez-Iturbe

and

José M. Mejía

Report N° 176

November 1973

Prepared under the support of U.S.
Department of Commerce, National Oceanic and
Atmospheric Administration, National Weather Service
Contract No. 2-36235
and with the cooperation of the Instituto Venezolano
de Investigaciones Cientificas, Caracas,
Venezuela

ABSTRACT

A methodology for the design of precipitation networks is formulated. The network problem is discussed in its general conception and then focus is made in networks to provide background information for the design of more specific gaging systems. The rainfall process is described in terms of its correlation structure in time and space. A general framework is developed to estimate the variance of the sample long-term mean areal precipitation and mean areal rainfall of a storm event. The variance is expressed as a function of correlation in time, correlation in space, length of operation of the network and geometry of the gaging array. The trade of time-vs-space is quantitatively developed and realistic examples are worked out showing the influence of the network design scheme in the variance of the estimated values.

ACKNOWLEDGEMENTS

This report is part of the final report for a study entitled "Rainfall Network Design". Financial support for this study was provided by the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service Contract No. 2-36235. Administrative support has been provided by the M.I.T. Division of Sponsored Research through DSR 80495.

The work was carried out by Dr. Ignacio Rodríguez-Iturbe, Associate Professor of Civil Engineering, M.I.T., and Dr. José M. Mejía, Associate Researcher, I.V.I.C., Venezuela.

The cooperation of the Instituto Venezolano de Investigaciones Cientificas, Caracas, is gratefully acknowledged as well as that of the Dirección de Obras Hidráulicas, Ministerio de Obras Públicas, Venezuela, which provided part of the data used in this study.

The authors gratefully acknowledge the suggestions of Prof. Peter S. Eagleson, Head of the M.I.T. Civil Engineering Department.

Finally the authors wish to thank Ms. Erika Babcock who performed all the necessary typing.

TABLE OF CONTENTS

1. THE NETWORK PROBLEM

The planning, development and management of water resources requires information about a large quantity of physical, economical and social factors. There is always an amount of uncertainty associated with the data which engineers and planners have to use for water resource problems, and it is this uncertainty what causes the question of how much information is enough and what kind of data do we need dealing with real life problems. The answer will always depend on the particular objectives which are being pursued and this is why it is so difficult to provide guidelines for the design of data collection programs. Data sampling activity should never be divorced from a preplanned program of interpretation and modelling of the system at hand. The choice of the variables to be measured, their sampling locations and sampling rates all depend upon the objectives of the programs, the type of models to be used to synthesize or to represent the system and upon the sensitivity of the decision making process to errors in the input information. Moreover, the validity of the information itself is determined by the conditions of sampling, the accuracy of the measurements and the time-space variability of the phenomena being sampled.

The U.S. Office of Water Data Coordination has defined three levels of information in regard to Data Network Design. Level 1 provides a base level of information for wide regional or national planning to be used for resource inventory and as background information for the design of more intensive and specific network systems. Level 2 concerns networks called to provide general water resources planning data and Level 3 is restricted to data collection programs for specific planning and managing activities (Rodda et al, 1969).

Levels 1 and 2 of information can be classified in what we call systems to provide "Regional Estimation" type of data. These networks provide the necessary data which are included in reconnaissance and preliminary exploration type of programs. Reconnaissance involves a general review of the hydrologic characteristics of large areas which are considered as a single major unit in a first stage of development planning. Specifically, reconnaissance should provide qualitative information in regard to characteristics like:

1. Amount of rainfall over the area.
2. Character of the streams existing in the region.
3. Existance of possible sources of groundwater.

Preliminary exploration, on the other hand, is oriented towards providing the first quantitative data about the hydrologic characteristics of the region like:

1. Monthly averages of rainfall over watersheds and subregions of the general area.
2. Gaging of the major streams detected in the reconnaissance stage.

3. Major trends existing in precipitation and ground water resources.

We call these Regional Estimation type of data because it will provide the planner with spatial rainfall averages and primary type of streamflow gaging stations from which a regionalization approach can be undertaken to hedge against the development of unanticipated needs.

Level 3, because of its same nature, is oriented towards the collection of accurate information regarding hydrologic characteristics influencing a preplanned program of economic development for the area. At this level of data collection we are interested not only in "Regional Estimation" type of problems but also in the "local estimation" of engineering variables at specific sites and subregions.

There exists a clear difference in the design process for data collection systems corresponding to Levels 1 and 2 and those under Level 3. When working in Level 3 type networks we have available a forecast of the economic development and relationships between the errors of estimate of the hydrological characteristics and the losses incurred in the developments envisaged. On the basis of this information it is possible to use mathem. programming techniques to obtain the optimum variation of the network characteristics during the planning period.

It is much more difficult to set up an economic optimization objective for the natural regime network, in this case the information is only loosely related to economic factors and the marriage between the sampling activity and the preplanned program of interpretation may be a difficult one to accomplish.

Two basic questions are involved in the design of data collections systems:

1. How much effort should be expended in the network and how should it be allocated to return the most information?
2. What inferences should be drawn from data collected by the network and what are the uncertainties in those inferences?

As Baecher (1972) discusses comprehensively, the first question involves statistical decision models; the second involves statistical inference. To approach the first question in a quantitative manner it is necessary to develop an analytical decision model of the network design process.

It is clear nevertheless, that inference and strategy decisions should not be made independent processes. Optimal actions will depend on the inferences which can be drawn from the collected data.

When designing a network to collect Level 3 type of information it is not very difficult to set up economic optimization objectives which usually come to the fact that marginal benefits produced by additional data should never be less than the cost to collect the data itself. There is a fair amount of work related to this topic in the hydrologic literature (Tschannerl, 1970; Dawdy et al., 1970; Moss, 1970; Duckstein and Kisiel, 1971; Davis and Dvoranchik, 1971;

etc.). The appropriate line of attack in these cases is based on Bayesian Decision Theory. This approach consists of the following steps (Davis and Dvoranchik, 1971):

1. Define the decision to be made and the possible alternatives.
2. Select the utility function.
    a) define goals
        i) select state variables (arguments of goal function).
        ii) develop stochastic properties of state variables.
    b) establish time preference
    c) include risk aversion
3. Making the decision.
    a) evaluate present knowledge (outcomes of alternatives and statistics of these outcomes).
    b) expected value of goal function for each alternative.
    c) select alternative with largest b).
4. Analysis of uncertainties.
    a) determine expected opportunity loss due to uncertainty
    b) evaluate data collection programs
        i) determine expected reduction in expected opportunity loss with additional data.
        ii) determine cost of obtaining additional data.

As pointed out by Davis and Dvoranchik, information is valuable only if its possession may cause a change of decision or action and its value is measured by the economic gain associated with the change of action.

A much less defined problem is the network design for areas where economic development is not foreseen - at least with enough clarity- within the planning horizon: level 1 and part of level 2 types of information. When we try to pursue similar lines of attack as used for level 3 of information, a dilemma arises: In most cases we know neither a decision space (for design) nor a utility function. The many uses to which level 1 of hydrologic information will be applied can not be foreseen at the time the data is being collected and even if they could be foreseen, there would be so many uses and so varied in character that a decision formulation would be unfeasible to construct.

Thus, optimizing allocations in a well defined Bayesian framework is not possible when designing networks for the base level of information. The planner then is left with two basic principles for level 1 data collection programs:

1. A network operated on a fixed budget should be designed to minimize the error of estimation of the hydrologic variables involved.
2. A network operated on a criteria of minimum acceptable accuracy should be designed for a minimum cost.

Any decision formulation which will lead us to optimizing allocations in this type of network will have to include either one of two parameters:

1. An estimate of the error made by the data collection sys-

tems, or

2. A measure of the amount of information collected by the system.

The approach to study those two parameters is different if we wish to set up a network which either

1. will allow the engineer to construct point estimates of a hydrologic characteristic at an ungaged site, or
2. will allow the engineer to estimate mean values of a hydrologic variable over a whole region.

The first case is that of stream gaging where for ungaged sites the selection of those that are to be gaged can be approached by means of a regional analysis of the information at the stations in the existing network or in similar systems. This type of analysis has been set forth by Matalas (1969) and Matalas and Gilroy (1968) and is based on a regression scheme that relates the means of the existing stations to physiographic and meteorological parameters. Using the same regression it is possible to obtain estimates of the mean values at ungaged sites.

The particular sites to be gaged are then those for which the variances of the estimated means are the largest. There are two main difficulties with this approach:

A. There is uncertainty about the transferability of the regression relation from gaged sites to ungaged sites. There is little we can do about this except to use the best physical criteria to justify the transfer.

B. The variance of the means of the ungaged sites consists of two components, namely, the spatial variation among and the time variation within the means of the gaged sites. The relative magnitudes of these components have not been found but if we could express them mathematically it would be possible to set forth an optimal scheme for stream gaging.

The second case we mentioned previously is that where the engineer needs a network to estimate mean values of a hydrologic variable over a certain region. This is typically the case of precipitation networks.

2. THE RAINFALL PROCESS

Precipitation is normally the most variable hydrologic element over a territory and its characterization is most commonly needed for water-balance studies and for flood forecasting.

In water-balance studies what is needed is the long-term mean areal precipitation during a certain interval of time. The interval of time depends mainly on the variability of rainfall, and it may be monthly, seasonal or annual depending of how representative are the measures in the time scale; in other words, it depends on the time scale of stationarity of the phenomenon and the use that is going to be made of the data.

In flood forecasting studies, precipitation data is commonly used for the construction of area-depth-duration curves and as input to rainfall-runoff types of models. In these cases what is needed is an estimation of the contribution of a particular type of storm to the area in consideration; this will be called mean areal rainfall for an event.

Long-Term Mean Areal Rainfall. - We will consider the rainfall process as a multidimensional random field, $f(x_i, t)$. function of the spatial coordinates x_i and the time t. For the determination of the mean value of this process, the correlation structure of the field both in space and time is of capital importance. For this particular problem it will be assumed that the process is stationary and furthermore that its correlation function is separable in terms of its spatial and temporal structure. The second assumption means that the covariance structure of $f(x_i, t)$ can be written in the form

$$\mathrm{cov}[f(x_i,t),\ f(x_{i'},t')] = \sigma_p^2 r(x_i - x_{i'}) \cdot r^*(t-t') \qquad (1)$$

where

σ_p^2 = point variance of $f(x_i, t)$

$r(x_i - x_{i'})$ = spatial correlation structure

$r^*(t-t')$ = temporal correlation structure

This assumption seems a normal one when examining that long term areal mean values are estimated by first forming a spatial average during each interval of time considered (years, months, etc.) and then adding them up in a discrete fashion.

The assumption of weakly stationarity means that

a) the expectation is a constant, and

b) the covariance function exists and is only a function of the difference between the spatial and temporal coordinates and not of the position or time itself.

This limits the applicability of our techniques to not too large areas and furthermore to regions which can be considered fairly uniform in what respect to hydrologic behavior ruling out or neglecting important factors like orographic effects.

Matheron (1965) has shown that the assumption of stationarity is in effect stronger than necessary and we may change it to what he calls an intrinsic hypothesis:

The increments of the random function $f(x)$, e.g. $f(x+H) - f(x)$, are weakly stationary but not necessarily the function $f(x)$ itself.

The importance of this change of hypothesis is, we believe, two-branched:

1) We may consider and adequately deal with a change of the mean value that can be considered as lineal drifts of the type

$$E[f(x+h) - f(x)] = ah = \sum_{i=1}^{n} a_i h_i \qquad (2)$$

(As a matter of fact, Yaglom (1962) gives Equation 2 as the definition itself of a process with stationary increments).

2) As shown by Yaglom (1962) it is appropriate to choose as the basic characteristic of $f(x)$ a function of the type $E\,|f(u) - f(v)|^2$ and <u>NOT</u> the correlation function which may not even exist.

The theory of processes with stationary increments is quite developed and up to now has mostly been used in turbulence. In the hydrologic problem we are dealing with, the importance of the first point (Equation 2) is self-evident, e.g. we may consider linear changes of mean rainfall with elevation. The second point is also very important but not self evident in respect to hydrology. We will now present a brief and general discussion of its importance.

If our spatial process $f(x)$ is not stationary and we act as if it were we will estimate from the data "something" that will be named correlation function and which may conduct us to non-sensical results. A good example of this is presented by Matheron in a recent paper (1970):

Let us denote by $f(x)$ a Brownian motion on a straight line $-\infty < t < \infty$, a realization of which is known as interval $0 \leq t \leq L$. $f(t)$ is a process with stationary increments which DOES NOT have a stationary covariance. If we proceed mechanically with our common estimating procedures we will:

1) Estimate $m = E\,[f(t)]$ by means of

$$\overline{f} = \frac{1}{L} \int_0^L f(t)\, dt, \quad \text{and}$$

2) Estimate the autocovariance by means of

$$C^*(h) = \frac{1}{L-h} \int_0^{L-h} C^*(t+h,t)\, dt$$

where

$$C^*(x,y) = [f(x) - \bar{f}] \cdot [f(y) - \bar{f}]$$

The problem is that neither m nor the covariance function really exist!

Matheron (1970) shows that we will further get

$$E\,[C^*(h)] = \frac{1}{3} L - \frac{4}{3} h + \frac{2}{3} \frac{h^2}{L}$$

which is absolute nonsense, as seen by the value of variance

$$E[C^*(0)] = \frac{1}{3} L$$

which is truly infinite and which we have "found" to depend only on the length of the sample L.

The conclusion is that if our process is non-stationary but can be considered one of stationary increments we should not try to characterize it with the covariance function but with a function of the type proposed by Yaglom (1962).

Matheron (1965) uses as characterizing function the half-variogram:

$$\gamma \ (h) = \frac{1}{2} \text{Var} \ [f(x+h) - f(x)] \tag{3}$$

Research needs to be performed to quantitatively assess the potentiality of this approach for hydrologic sampling problems. Its promise is great because it will allow the engineer to deal with data processes that are changing in space.

The temporal correlation structure of rainfall in terms of years, months or weeks appears to be quite weak and can be approximated by a simple Markovian scheme:

$$r^*(t-t') = \rho^{|t'-t|} \tag{4}$$

Here ρ denotes the first autocorrelation coefficient which is practically always less than 0.250.

The spatial correlation structure, $r(x_i - x_{i'})$ in Equation 1 poses a different problem. The available data is not sufficient for statistical discrimination between different kinds of functions which could represent the spatial correlation structure. Two problems may be distinguished in this aspect:

1) what kind of function do we use to represent the spatial correlation? and

2) how do we fit the parameters of that function with unevenly spaced areal data?

Both questions are very important and we will deal with the second one later in this paper. With respect to the first question it has been common to look for correlation functions which decay as a function of distance; one of this type is the exponentially decaying function,

$$r(x,y) \quad = \quad e^{-h(x^2+y^2)^{1/2}} \tag{5}$$

which for an isotropic process can be written as

$$r(v) \quad = \quad e^{-hv} \tag{6}$$

where v represents the distance between any two points. From a conceptual point of view it can be shown (Matern, 1947) that Equation 5 corresponds to a process which in its simplest form can be written as

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 - \alpha^2\right]^{3/4} \cdot f(x,y) \quad = \quad \varepsilon(x,y) \tag{7}$$

where $\varepsilon(x,y)$ represents an uncorrelated two-dimensional process. Whittle (1954) points out that it is difficult to visualize a physical mechanism which would lead to such a relation.

If we are interested for example in relating rainfall at a point in space $f(x,y)$ symmetrically to rainfall at all points around (x,y), the equation to use is

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 - k^2\right] f(x,y) = \varepsilon(x,y)$$

which leads to the covariance function

$$r(v) = b\, v\, K_1\,(bv) \tag{8}$$

where K_1 denotes a modified Bessel function, b is a constant, and v represents the distance between points.

The correlation function (8) may be regarded as the "elementary correlation in two dimensions, similar to the exponential $e^{-\alpha|x|}$ in one dimension. Both correlation curves are monotone decreasing, but Equation (8) differs in that it is flat at the origin, and that "its rate of decay is slower than exponential" (Whittle, 1954).

Figure 1 shows a comparison between the covariance functions e^{-v} and $v\,K_1(v)$. The process with covariance function e^{-v^2} is computationally easy but it is thought to be "too continuous" to be realistic (Matern, 1960). In fact, it is deterministic along any straight line in a plane (Karhunen, 1952).

All the developments of this paper will be made for both correlation functions given by Equations 6 and 8. It is important in the design of rainfall networks to develop a feeling for the range of variation existing in the parameters of the correlation function. Eagleson (1967) presents the correlation structure as a function of

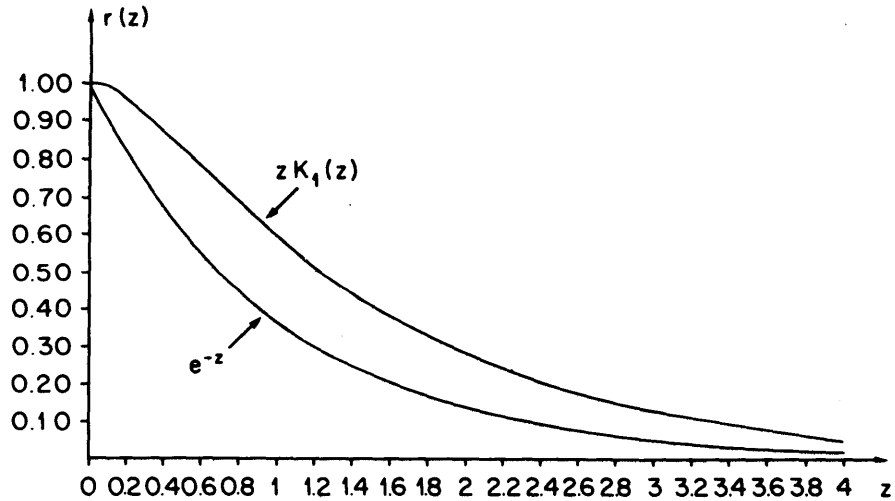

Figure 1 - Correlation Functions $r(z) = e^{-z}$ and $r(z) = z\,K_1(z)$

distance for a 1250 mi^2 catchment in Australia. Using annual data with thirteen raingages and 17 years of data he finds a correlation of 0.5 for a distance of 19 miles. This will correspond to values

$$h = 0.0365 \text{ mi}^{-1} \quad \text{and} \quad b = 0.0684 \text{ mi}^{-1}$$

in Equations 6 and 8 respectively.

Stol (1972) working with monthly data in The Netherlands shows correlation decays which are much faster for July than for January.

$$h \text{ (January)} = 0.0010 \text{ Km}^{-1} \quad h \text{ (July)} = 0.0096 \text{ Km}^{-1}$$

Hendrick and Comer (1970) using daily precipitation data during the months of June-August (1961-1966) in northern Vermont show a strong decay in correlation which produces $r = 0.51$ at a distance of 5 miles. This corresponds to

$$h = 0.135 \text{ mi}^{-1} \quad \text{and} \quad b = 0.26 \text{ mi}^{-1}$$

These examples are shown as kind of smoothly, well-fitted types of decay but reality is not as simple as this. A homogeneous area of 30.000 Km^2 with 26 raingages was chosen in central Venezuela in order to investigate correlation decay as function of distance. The region is shown in Fig. 2. It was found that although it is true that smooth correlation structures can be "fitted" to the data, it makes a great difference from the point of view of network design <u>how</u> the actual fitting was performed. Using a sophisticated scheme - to be described later in the paper - which takes into account both the relative position of the stations and the length of the records, it was estimated that

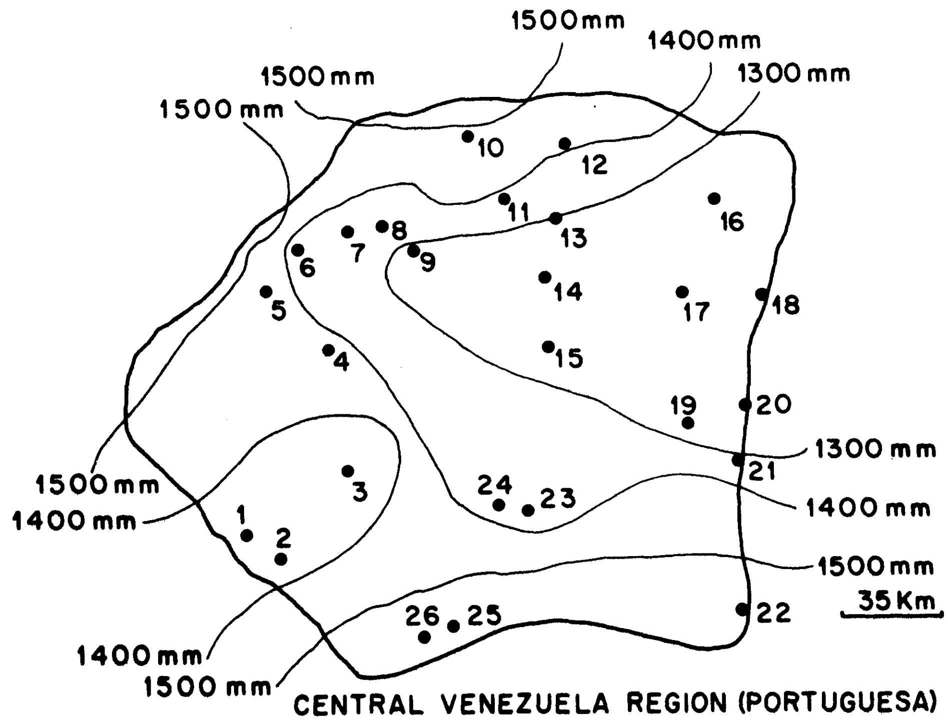

Figure 2 - Central Venezuela Region (Portuguesa State) Used as Example in the Paper.

$$r(10\ \text{Km}) = 0.942 \quad \text{and} \quad r(50\ \text{Km}) = 0.533$$

Those values will render exponentially decaying factors (Equation 6) of

$$h(10\ \text{Km}) = 0.0060\ \text{Km}^{-1}, \quad h(50\ \text{Km}) = 0.012\ \text{Km}^{-1}$$

which are quite different, and as will be seen later would require very different network densities when maintaining constant the required precision and the length of operation of the network. The question is then, what h, or b, to use?

The answer intrinsically depends on the size of the area being analyzed. Thus, if the area is small, the correlation in space should be fitted with the criterion of preserving estimated correlation coefficients for a short distance. The opposite will be true when we deal with a large region. A typical distance which characterizes the size and shape of the area being analyzed is the mean distance between two randomly chosen points in the region; this we will define as the "characteristic correlation distance".

Gosh (1951) derives the distribution of the distance between two points chosen at random in a plane convex region. The region is denoted by S, and its area and perimeter by A and P respectively. When S is a rectangle with sides A_1 and A_2, the frequency function can be written as

$$\frac{1}{\sqrt{A}}\ f\ (v/\sqrt{A},\ \sqrt{A_1/A_2}) \qquad (9)$$

where v is the distance among the points and

$$f(w,a) = 2w[f_1(w,a) + f_2(wa,a) + f_2(w/a,1/a)]$$

with

$$f_1(w,a) = \begin{cases} \pi + w^2 - 2w(a+1/a) & 0 < w < \sqrt{a^2+a^{-2}} \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(w,a) = \begin{cases} 2\sqrt{w^2-1} - 2\cos^{-1}(1/w) - a^{-2}(w-1)^2 & 1 < w < \sqrt{1+a^4} \\ 0 & \text{otherwise} \end{cases}$$

Matern (1960) has used Equation 9 to compute the mean distance between two randomly chosen points in seven regions of area 1:

Circle	0.5108	Rectangle $\alpha = 2$	0.5691
Hexagon	0.5126	Rectangle $\alpha = 4$	0.7137
Square	0.5214	Rectangle $\alpha = 16$	1.3426

Equilateral triangle 0.5544.

When working with a region of area A it is only necessary to adjust those factors with proportionality coefficients made up of the ratio of two corresponding distances in the figure of area A and the figure of unit area. Thus in the region of Central Venezuela the zone can be approximated by a rectangle with sides ratio equal to 2; with an area of 30.000 Km^2 this gives a diagonal of 268 Km. A unit area rectangle of the same shape has a diagonal of 1.58 and thus the characteristic correlation distance is in this case

$$\frac{0.5691 \times 268}{1.58} = 97 \text{ Km.}$$

We need then to fit the correlation structure in space preserving estimated correlation coefficients for distances of the order of 100 Km, which were found to be equal to 0.21. This in turn produces

$$h = 0.0156 \text{ Km}^{-1} \qquad b = 0.0234 \text{ Km}^{-1}$$

which will be the correlation parameters to be used later in this paper when designing a network for this particular region. The watershed analyzed by Eagleson (1967) has an area of 1250 mi^2 and can also be approximated by a rectangle with $\alpha = 2$. This gives a diagonal of 56 mi. and a characteristic correlation distance of 20 mi., which in this case agrees with Eagleson's correlation radius r_o of 19 miles, defined to be the distance at which the correlation function drops to 0.5.

Areal Mean for Rainfall Event. Generally speaking, the difference between the mean rainfall P_a over an area about the storm center (Eagleson, 1970)

1. increases with decrease in the total rainfall depth.
2. decreases with increasing duration.
3. is greater for convective and orographic precipitation than for cyclonic.
4. increases with increasing area.

For convective storms (in Arizona), Woolhiser and Schwalen (1959) have fitted the average areal rainfall distribution with the function

$$P_a/P_t(0) = 1 - [0.14/P_t(0)]\, A_s^{0.6} \qquad (10)$$

where $P_t(0)$ is the total depth in inches at the storm center, P_a is the average depth over the circular area A_s (in square miles) surrounding the center, and radial symmetry is assumed. Since

$$A_s = \pi r^2 \quad \text{and} \quad P_a = \frac{1}{\pi r^2} \int_o^r 2\pi r \; P_t(r) \; dr$$

Eagleson (1967) gives Equation 10 in the form

$$P_t(r)/P_t(0) \;=\; 1 - 0.72 \; (r/r_o) \qquad (11)$$

where

$$r_o \;=\; 1.73 \; P_t(0)$$

is the correlation radius already defined. The shape of the functions given by Equations 6 and 8 suggests that they may be appropriate for the description of the spatial correlation structure of a rainfall event. It is necessary nevertheless to get an idea of the range of variation of the parameters.

Fogel and Duckstein (1969) present data which show storm center depths for convective rainfall in Arizona varying from 0.75 to 5 inches. This is equivalent to correlation radii from 1.30 miles for the weaker storms up to 9 miles for the more intensive ones. These values in turn produce correlation structures of the form

$$r(v) \;=\; e^{-0.533v} \quad \text{or} \quad r(v) \;=\; 0.93 \; v \; K_1(0.93 \; v) \qquad (12)$$

for the storm with center depth of 0.75 inches and

$$r(v) \;=\; e^{-0.080v} \quad \text{or} \quad r(v) \;=\; 0.13 \; v \; K_1(0.13 \; v) \qquad (13)$$

for the storm with center depth of 5 inches.

It will be shown later in this paper that Equations (12) and (13) lead to quite different types of network, posing then the question of what type of storm should be the commanding one when designing a network for the determination of areal mean of rainfall events. This question is tied up to the considerations made in the section entitled "The Network Problem," where we noted the importance of relating the network to the economic criteria involved with the problem at hand. This point will be developed at length in a future paper by the authors.

It is also necessary to get an idea of the areal extension of convective storms when in the process of designing a network. This has been done using the relationship presented by Fogel and Duckstein (1969)

$$P_t(r) = P_t(0)\, e^{-\pi r^2 t} \tag{14}$$

where t is a dispersion parameter given by

$$t = 0.27\, e^{-0.67\, P_t(0)}$$

We will arbitrarily fix the limit of the storm at a depth of 0.1 inches, obtaining in this manner areas of

$$A = 12 \text{ mi.}^2 \quad \text{for} \quad P_t(0) = 0.75 \text{ in.} \qquad \text{and}$$

$$A = 435 \text{ mi.}^2 \quad \text{for} \quad P_t(0) = 5.0 \text{ in.}$$

For great <u>cyclonic storms</u> in the United States, Boyer (1957) fitted the average areal rainfall distribution with the storm centered function

$$P_t(r) \ / \ P_t(0) \ = \ e^{-ar} \tag{15}$$

where the parameter a is given by Eagleson (1967) as

$$a \ = \ 1.68/r_o \tag{16}$$

A typical isohyetal pattern for this type of storms is shown in Figure 3 which shows the rainfall produced on August 12/13, 1955, in the Baltimore area by Hurricane Connie. This particular example can be described with

$$P_t(0) \ = \ 8.50 \text{ in.} \quad \text{and} \quad P_t(8.3 \text{ mi}) \ = \ 7.0 \text{ in.}$$

Using Equation 15 to compute "a" and Equation 16 to estimate the correlation radius, a value of 73 miles is obtained for r_o. This in turn corresponds to correlation structures of the type

$$r(v) \ = \ e^{-0.009\, v} \quad \text{and} \quad r(v) \ = \ 0.016\, v\, K_1(0.016\, v) \tag{17}$$

The areal extension of this type of storms is so large - $P_t(r) = 0.5$ in. correspond to $A = 18.000 \text{ mi}^2$ - that it will cover any region which we may consider homogeneous for the purpose of network design.

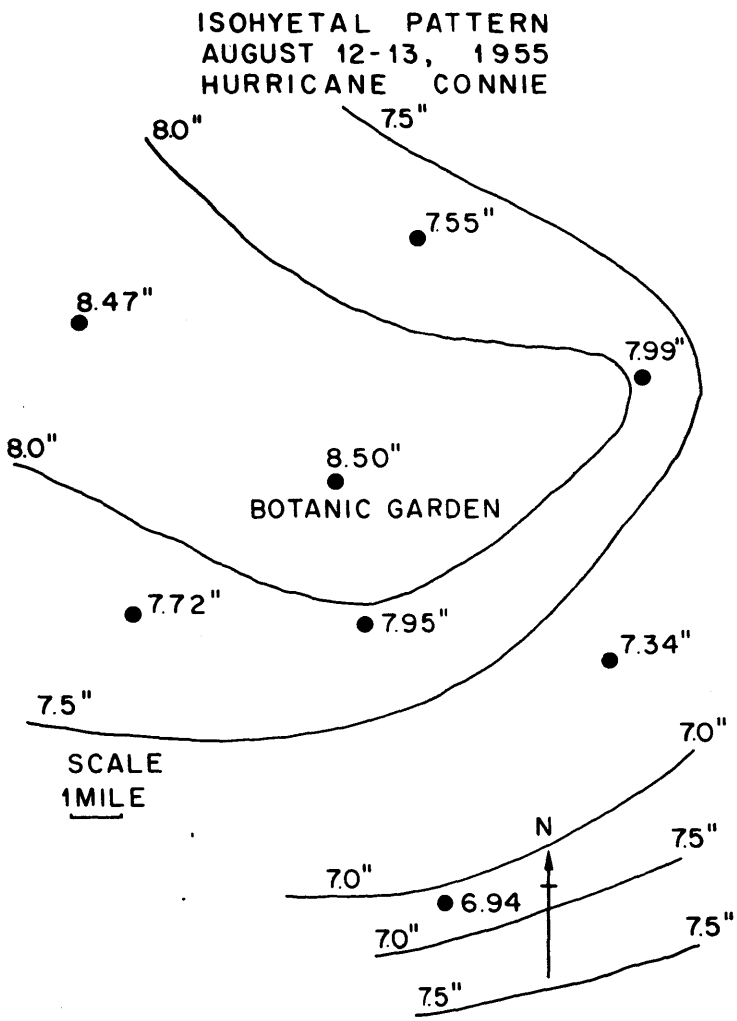

Figure 3 - Isohyetal Pattern of Cyclonic Storm

3. GENERAL FRAMEWORK FOR ESTIMATING LONG-TERM AREAL MEAN RAINFALL

Let us establish the following notation:

$f(x_i,t)$ difference between rainfall depth at the point of spatial coordinates x_i during year, month, season t, and the mean of the process

N number of stations in the network

T, number of years, months, seasons, the network is in operation

A, area in consideration.

The hydrologist wants to estimate

$$\lim_{T'\to\infty} \frac{1}{AT'} \sum_{t=1}^{T'} \int_A f\,(x_i,t)\,dx_i \tag{17}$$

by means of

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f\,(x_i,t) = \bar{P} \tag{18}$$

The precision of the estimation is measured by the variance of $\bar{P}$:

$$E\left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f(x_i,t) - \lim_{T'\to\infty} \frac{1}{AT'} \sum_{t=1}^{T'} \int_A f(x_i,t)\,dx_i\right]^2 \tag{19}$$

We will prove now that the mean value given by Equation 17 has zero variance and therefore can be considered as a constant.

$$E\left[\lim_{T'\to\infty} \frac{1}{AT'} \sum_{t=1}^{T'} \int_A f(x_i,t)\,dx_i\right]^2 =$$

$$= E\left[\lim_{T'\to\infty} \frac{1}{A^2T'^2} \sum_{t=1}^{T'} \sum_{t'=1}^{T'} \int_A \int_A f(x_i,t)\, f(x_{i'},t')\,dx_i\,dx_{i'}\right] =$$

$$\lim_{T'\to\infty} \frac{1}{A^2T'^2}\, \sigma_p^2 \sum_{t=1}^{T'} \sum_{t'=1}^{T'} \int_A \int_A r(x_i - x_{i'})\, \rho^{|t-t'|}\,dx_i\,dx_{i'}$$

where use has been made of Equations 1 and 4 of the second section of the paper. We can now write the previous expression as

$$\lim_{T\to\infty} \frac{1}{A^2T'^2} \sigma_p^2 \left[\sum_{t=1}^{T'} \rho^{|o|} + 2 \sum_{t=1}^{T'-1} \sum_{t'=t+1}^{T'} \rho^{|t'-t|} \right] \int_A \int_A r(x_i - x_{i'})\, dx_i dx_{i'} \tag{20}$$

We have now that,

$$\sum_{t=1}^{T'} \rho^{|o|} = T' \quad \text{and} \quad \sum_{t=1}^{T'-1} \sum_{t'=t+1}^{T'} \rho^{t'-t} =$$

$$= \sum_{t=1}^{T'-1} \rho+\rho^2 + \ldots + \rho^{T'-t} \tag{21}$$

Calling

$$S = 1 + \rho + \rho^2 + \ldots + \rho^{T'-t-1}$$

and subtracting $S - \rho S$ we can write

$$S = \frac{1 - \rho^{T'-t}}{1 - \rho}$$

and Equation 21 is equal to

$$\sum_{t=1}^{T'-1} \frac{\rho(1-\rho^{T'-t})}{1 - \rho} = \frac{\rho}{1-\rho}\left[(T'-1) - \frac{\rho}{1-\rho}(1 - \rho^{T'-1})\right] \tag{22}$$

Substituting Equation 22 in Equation 20 we obtain

$$\lim_{T'\to\infty} \frac{1}{A^2T'^2} \sigma_p^2 \left\{ T'+2 \frac{\rho}{1-\rho} \left[(T'-1) - \frac{\rho}{1-\rho} (1-\rho^{T'-1})\right] \right\} \cdot \int_A \int_A r(x_i - x_{i'}) dx_i dx_{i'}$$

$$= 0 .$$

In this manner without loss of generality we can consider

$$E[f(x_i,t)] = 0 \quad \text{and} \quad E[f^2(x_i,t)] = \sigma_p^{\ 2}$$

The variance of the regional mean $\bar{P}$ (Equation 19) is then given by

$$\mathrm{Var}[\bar{P}] = \frac{1}{N^2T^2}\, E\left[\sum_{t=1}^{T}\ \sum_{i=1}^{N} f\ (x_i,t)\right]^2 \tag{23}$$

The problem we face now is to evaluate Var $\bar{P}$ as a function of the correlation structure of the process both in space and time, the number of stations in the network, the sampling geometry of the network and the length of operation of the stations. To this end we write

$$\mathrm{Var}[\bar{P}] = \frac{1}{N^2T^2}\, E\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\ \sum_{i'=1}^{N} f(x_i,t)\ \ f(x_{i'},t)\ + \right.$$

$$\left. +\ 2\sum_{t=1}^{T-1}\ \sum_{t'=t+1}^{T}\ \sum_{i=1}^{N}\ \sum_{i'=1}^{N} f(x_i,t)\ \ f(x_{i'},t')\right] =$$

$$= \frac{1}{N^2T^2}\,\sigma_p^{\ 2}\left\{\left[\sum_{i=1}^{N}\sum_{i'=1}^{N} r(x_i - x_{i'})\right]\left[\sum_{t=1}^{T} 1 + 2\sum_{t=1}^{T-1}\ \sum_{t'=t+1}^{T}\rho^{\,t'-t}\right]\right\} \tag{24}$$

Let us now call

$$F_2(N) = \frac{\sum_{i=1}^{N}\sum_{i'=1}^{N} r(x_i - x_{i'})}{N^2} \tag{25}$$

and

$$F_1(T) = \frac{1}{T^2}\left[\sum_{t=1}^{T} 1 + 2\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\rho^{t'-t}\right] =$$

$$= \frac{T + 2\sum_{t=1}^{T-1}\frac{\rho}{1-\rho}(1-\rho^{T-t})}{T^2} =$$

$$= \frac{T + 2\,\frac{\rho}{1-\rho}\left[(T-1) - \frac{\rho}{1-\rho}(1-\rho^{T-1})\right]}{T^2} \qquad (26)$$

Equation 24 can then be written as

$$\mathrm{Var}[\bar{P}] = \sigma_p^{\,2}\,[F_1(T)]\;[F_2(N)] \qquad (27)$$

where the variance of the regional mean is expressed as a function of the point variance of the process multiplied by two reduction factors, one of them $F_2(N)$ due to sampling in space and the other $F_1(T)$ due to sampling in time.

$F_1(T)$ is independent of the number of stations and the spatial properties of the process, it is only a function of the correlation in time and the length of time the network has been in operation. Figure 4 shows $F_1(T)$ as function of ρ and T; when $T = 1$ year (month, season), the variance reduction due to temporal sampling is equal to 1 meaning that there is no reduction at all in the variance of the long-term areal mean with respect to the point variance of the process.

The variance reduction due to spatial sampling $F_2(N)$ depends on the correlation structure in space, the sampling geometry and the number of stations. Three types of sampling schemes can be considered.

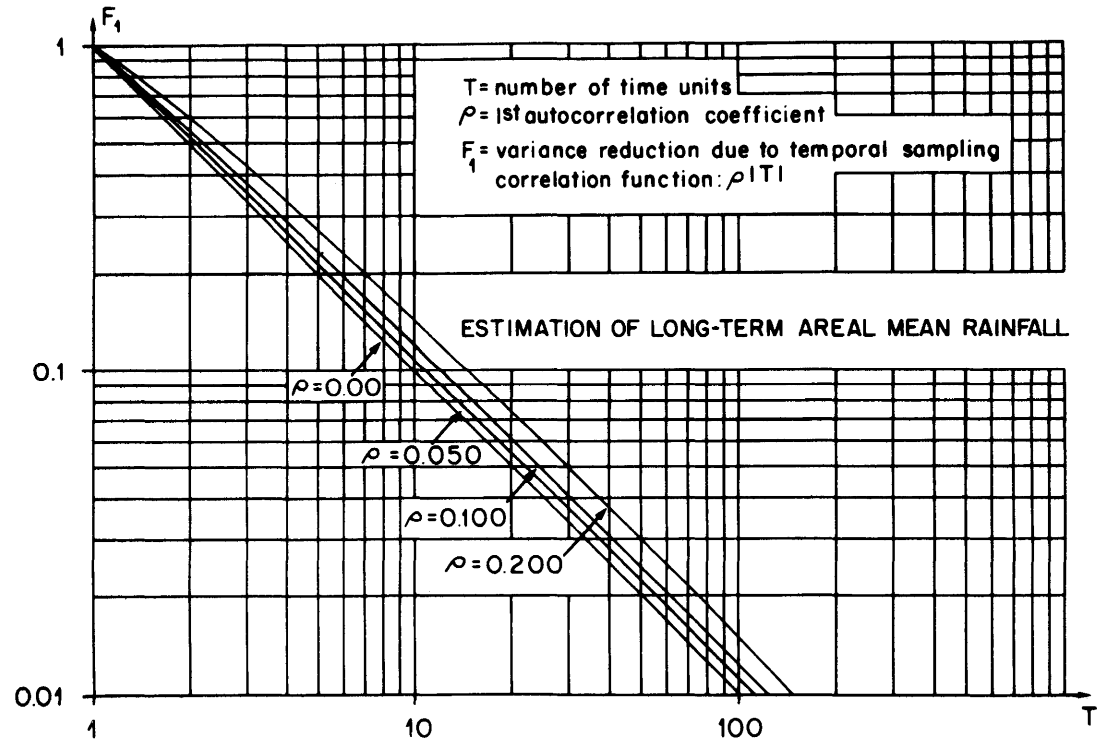

Figure 4 - Variance Reduction Factor due to Temporal Sampling. Used in the Estimation of Long-Term Mean Areal Rainfall.

Simple Random Sampling.- In this type of network, each station is located with a uniform probability distribution over the whole space, A, independently of the other stations.

Stratified Random Sampling.- In this case we assume our space a is divided into a number of non-overlapping congruent strata α_i. From each stratum k points are chosen randomly where the raingages will be located.

Systematic Sampling. - In this kind of scheme, the cluster of sampling units forms some regular geometric pattern.

Simple random sampling and stratified sampling can be realistic schemes for practical hydrologic purposes where we either distribute the stations more or less randomly or divide the region in several sub-areas of similar sizes where the position of the stations in each sub-area is determined by conditions like accessibility, nearby population centers, etc. Systematic sampling because of the practical problems and expenses it involves does not qualify as a realistic sampling scheme in real world hydrology. We will therefore skip its study.

Variance Reduction Factor Due to Spatial Random Sampling.- We need to evaluate

$$F_2(N) = \frac{1}{N^2}\left[\sum_{i=1}^{N}\sum_{i'=1}^{N} r\,(x_i - x_{i'})\right] =$$

$$= \frac{1}{N^2}\, E\left[\sum_{i=1}^{N} r\,(o) + 2\sum_{i=1}^{N-1}\sum_{i'=i+1}^{N} r(x_i - x_{i'})\right]$$

since the 2nd term is a random variable, we will work with its expected value

$$F_2(N) = \frac{1}{N^2}\left\{N + N\ (N-1)\ E[r(x_i - x_{i'})\ |A]\right\} \tag{28}$$

where $E[r(x_i - x_{i'})|A]$ represents the expected value of the correlation between two points randomly located on an area of size A. It depends both on the shape of the area and the type of spatial correlation structure which characterizes the process.

$$E[r(x_i - x_{i'})|A] = \int_o^R r(v)\ f\ (v)\ dv \tag{29}$$

where $r(v)$ represents the spatial correlation assumed to be isotropic and $f(v)$ is the frequency function of the distance v between two randomly chosen points in the area A. R represents the largest distance existing in A. It was previously seen that $f(v)$ is a function of the shape of the area under study - see Equation 9 - but fortunately it varies little for the shapes normally found in nature; this can be seen in the values of the mean distance between two randomly chosen points given before for seven regions of area 1. Only for the rectangle with $\alpha = 16$ we can notice sizeable difference in values and this is because in this case we are moving from a two-dimensional case to a transect or one-dimensional space. Because of the previous reason we will perform all our computations for a region of square shape which is the simplest one to evaluate. Furthermore, we need a scheme to generalize the computations for different combinations of areas and correlation parameters.

Let us consider an area of size A on which we superimpose a process with correlation parameter h (or b) equal to one. Equation 29 takes then the form

$$\int_{o}^{d} r(v) \quad f\ (v)\ dv \tag{30}$$

where d is the length of the diagonal of the square shaped region.

Let us analyze a similar region with area A/h^2 where h is the magnitude of the parameter of the correlation structure. Equation 29 comes now in the form

$$\int_{o}^{d/h} r_1(v) \quad f_1(v)\ dv \tag{31}$$

where d/h is the length of the new diagonal and we have to analyze the form of $r_1(v)$ and $f_1(v)$.

$r_1(v)$ is simply given by

$$r_1(v) \quad = \quad r(hv) \tag{32}$$

and $f_1(v)$ will be of the same form of $f(v)$ but it is affected by a factor of proportionality equal to 1/h which reflects the change made when going from area A to A/h^2.

$$\frac{1}{h}\, f_1\ (v/h) \quad = \quad f(v)$$

or

$$f_1(v) \quad = \quad h\ f\ (vh) \tag{33}$$

Equation 33 can then be written as

$$\int_{o}^{d/h} r(hv)\ h\ f\ (hv)\ dv$$

and making $vh = v'$ we obtain

$$\int_{o}^{d} r(v') \cdot f(v') \cdot dv' \tag{34}$$

which is identically equal to Equation 30. In this manner we have proved that an area of size A with a process with correlation parameter equal to 1 has the same expected value of the correlation between two randomly chosen points as a homologous area of size A/h^2 over which is acting a process with correlation parameter h.

Thus, what remains constant is the product Ah^2 if we want to obtain the same value of

$$E[r(x_i - x_{i'}) \mid A]$$

Equation 28 was evaluated by calculating the integral given by Equation 29 for a large range of values of Ah^2 maintaining N fixed and then varying N and repeating the procedure. The evaluation of Equation 29 was performed numerically with the use of the expression for $f(v)$ given by Equation 9. Two sets of values were obtained, one for each of the correlation structures given by Equations 6 and 8.

The results are presented in Figure 5 and 6 which will be studied in detail later in their application to practical cases.

<u>Variance Reduction Factor Due to Spatial Stratified Sampling.-</u> We need to evaluate

$$F_2(N) = \frac{1}{N^2}\left[\sum_{i=1}^{N}\sum_{i'=1}^{N} r(x_i - x_{i'})\right] =$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N}\left\{r(o) + \sum_{\substack{i'=1 \\ i'\neq i}}^{N} r(x_i - x_{i'})\right\}\right] \tag{35}$$

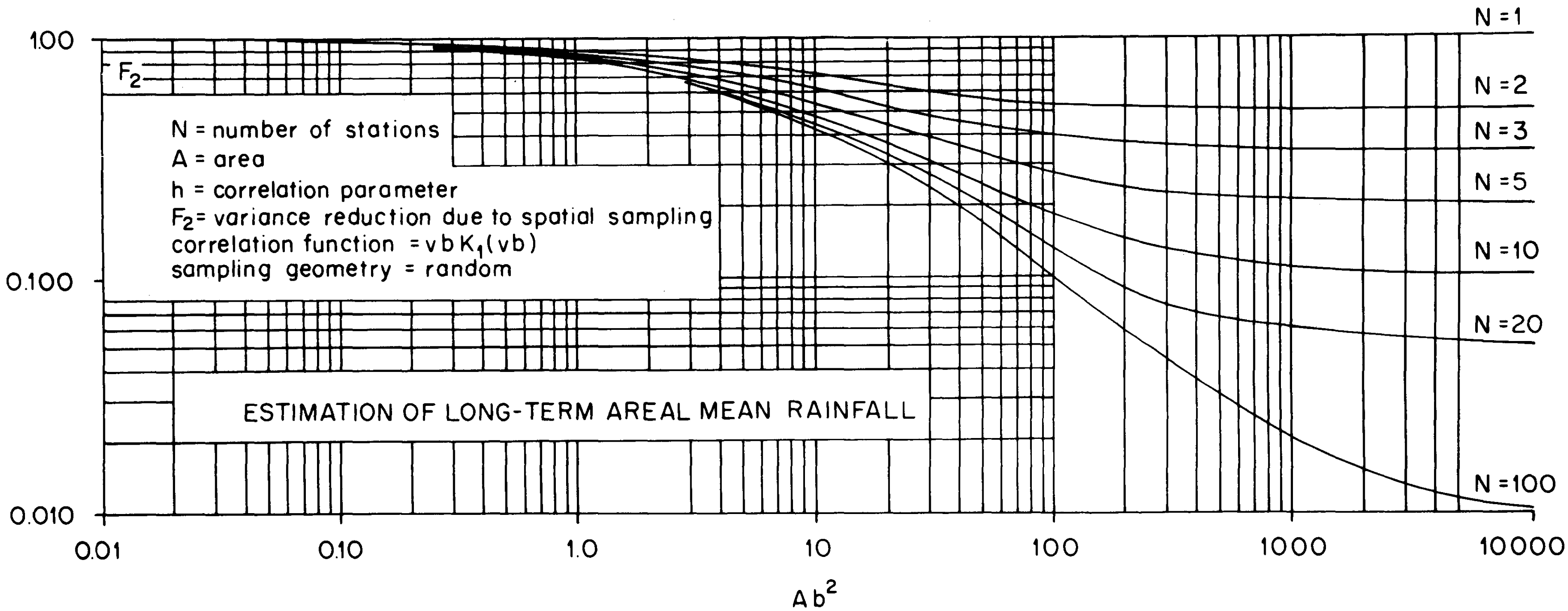

Figure 5 - Variance Reduction Factor due to Spatial Sampling with Random Design. Used in the Estimation of Long-Term Mean Areal Rainfall with $r(v) = bv\,K_1(bv)$.

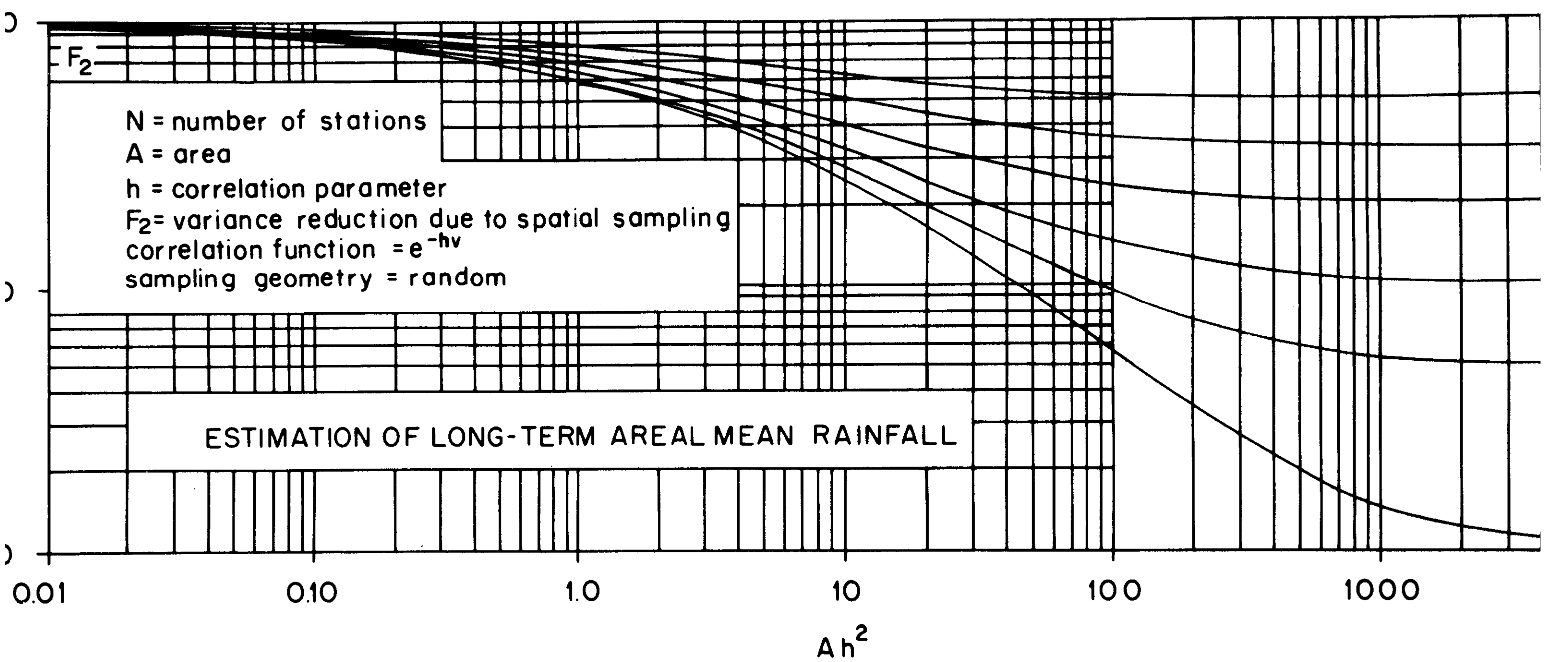

Figure 6 - Variance Reduction Factor due to Spatial Sampling with Random Design. in the Estimation of Long-Term Mean Areal Rainfall with $r(v) = \epsilon^{-hv}$.

We will call now

x_i, y_i two different points randomly located in the same stratum

x_i, $y_{i'}$ two different points randomly located in different strata

Equation (35) becomes

$$F_2(N) = \frac{1}{N^2}\left\{ N + \sum_{i=1}^{N}\left[\sum_{i'=1}^{N} r(x_i - y_{i'}) - r(x_i - y_i)\right]\right\} \tag{36}$$

where the points in the 2nd summation are now divided into points in different strata (1st term), and points in the same stratum (2nd term). We can further write

$$F_2\ (N) = \frac{1}{N^2}\ [N + W_1 - W_2]$$

where, taking expected values as before

$$W_1 = E\left[\sum_{i=1}^{N}\ \sum_{i'=1}^{N} r\ (x_i - y_{i'})\right]$$

Since x_i and $y_{i'}$ are random variables with uniform distributions $\frac{1}{a}$ and $\frac{1}{a'}$,

$$W_1 = \sum_{i=1}^{N}\ \sum_{i'=1}^{N} \int_{a_i}\int_{a_{i'}} \frac{r(x_i - y_{i'})}{a^2}\ dx_i\ dy_{i'} \tag{37}$$

a_i and $a_{i'}$ represent the area of the strata i and i' assumed to be equal in size. Equation 36 is easily simplified when we notice that the double summation in i and i' of the integrals over a_i and $a_{i'}$ gives in effect two integrals over the whole area A,

$$W_1 = \int_A \int_A \frac{r(x_i - y_{i'})dx_i \, dy_{i'}}{a^2} = \frac{A^2 E[r(x_i - y_{i'})|A]}{a^2} \tag{38}$$

We will assume now one station per stratum and the area of the stratum will be adjusted according to the number of stations in order to meet this assumption. In this manner

$$W_1 = N^2 E\ [r(x_i - y_{i'})|A] \tag{39}$$

Equation 39 can now be evaluated with the same procedure for Equation 29 and previously described in detail. The term W_2 in Equation 36 is equal to

$$W_2 = \sum_{i=1}^{N} r(x_i - y_i)$$

And taking expected values as before,

$$W_2 = N\ E\ [r(x_i - x_{i'})\ |\ (A/N)] \tag{40}$$

where $E[r(x_i - x_{i'})|(A/N)]$ represents the expected value of the correlation between two points randomly chosen in the area (A/N). Equation 40 can also be evaluated with the same procedure already described for Equation 29.

Similarly then for the case of spatial random sampling, the study of stratified sampling was made for a square region and moreover the strata were assumed of square shape in order to avoid boundary or frontier problems.

The results of this part of the analysis showing the variance reduction factor due to spatial stratified sampling - Equation 36 - are shown graphically in Figures 7 and 8.

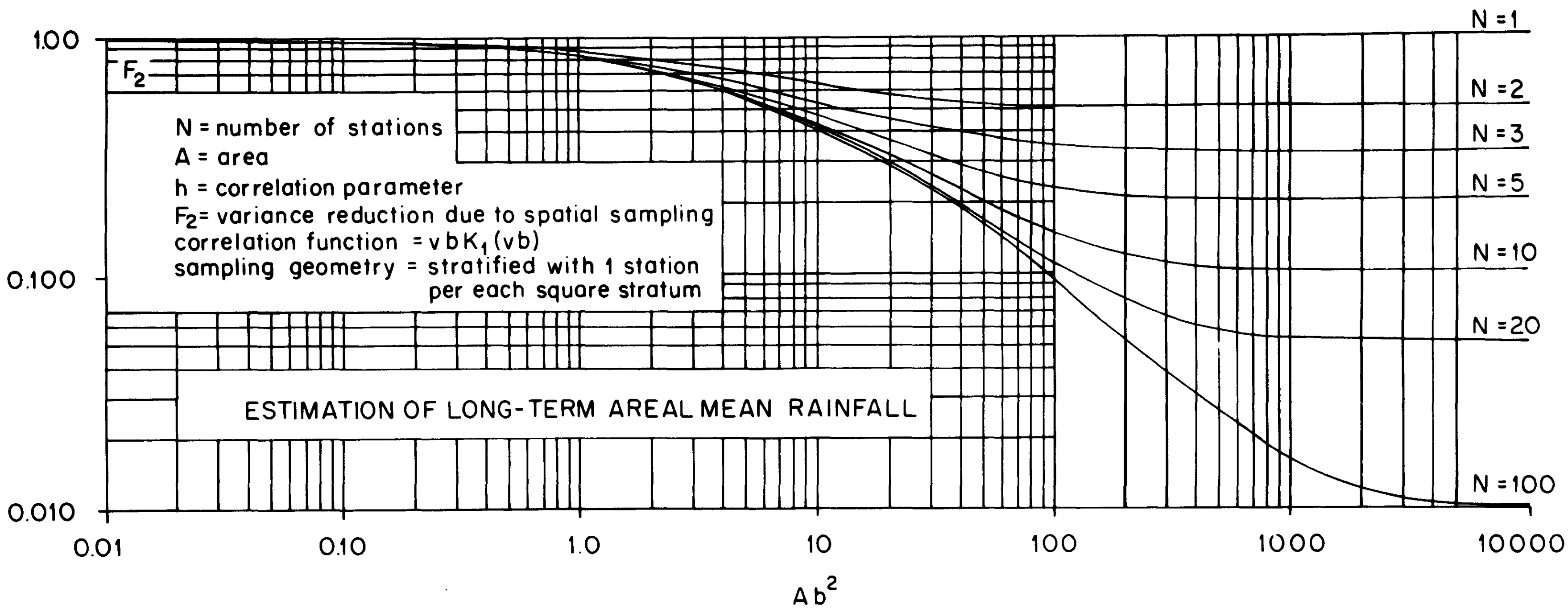

Figure 7 - Variance Reduction Factor due to Spatial Sampling with Stratified Design. Used in the Estimation of Long Term Mean Areal Rainfall with $r(v) = vb\,K_1(vb)$.

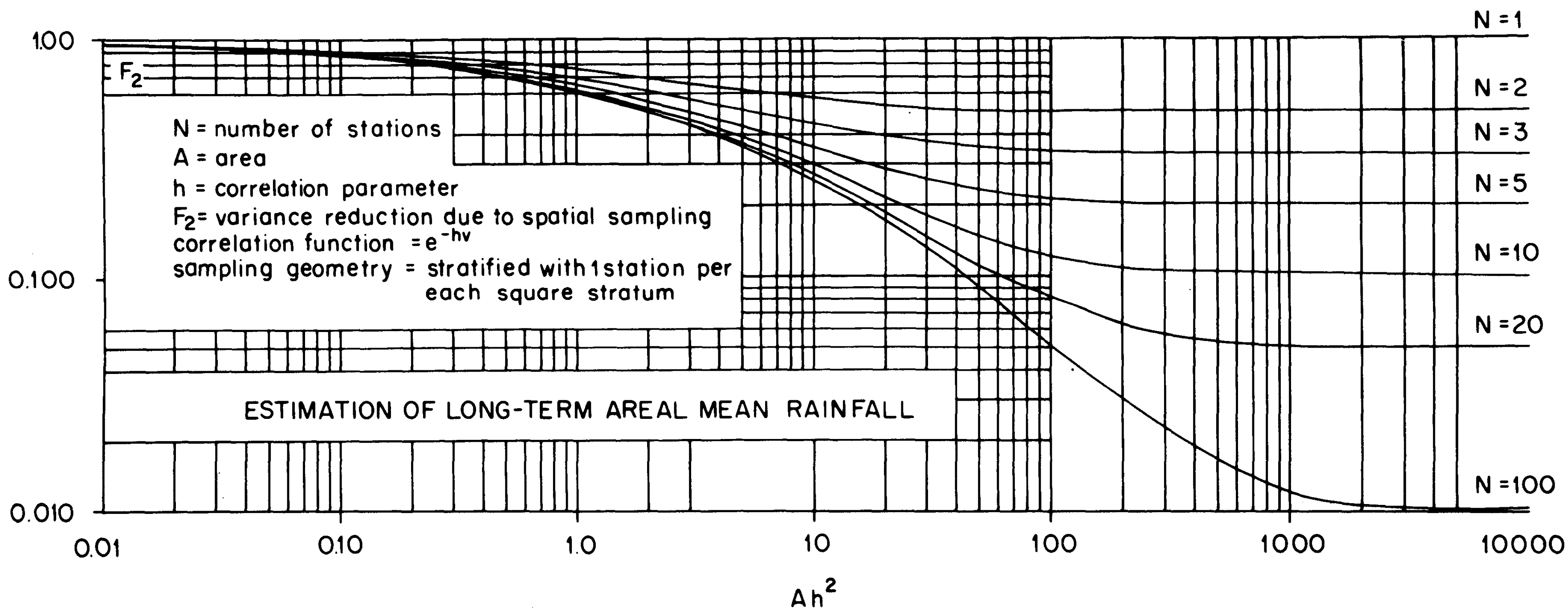

Figure 8 - Variance Reduction Factor due to Spatial Sampling with Stratified Design. Used in the Estimation of Long Term Mean Areal Rainfall with $r(v) = e^{hv}$.

Figure 5, 6, 7, and 8 provide an analytical tool for trading time-vs-space in the estimation of long-term spatial averages of precipitation. Their use in network design will be illustrated with two examples.

4. TRADING TIME-VS-SPACE.

1) Central Venezuela Region. The area of approximately $30{,}000\ Km^2$ is shown in Figure 2. Its 26 raingages, the stations with their mean annual precipitation values, the standard deviations and the length of the records are given in Table 1. To start with, it is necessary to adjust a correlation structure both in time and space which will be representative of the whole region. The parameter b to use in the correlation function $vb\ K_1(vb)$ was estimated from the equation

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=k_i(i,j)}^{k_f(i,j)} f_{i,k}\, f_{j,k} = \sum_{i=1}^{N} \sum_{j=1}^{N} [k_f(i,j) - k_i(i,j) + 1]\; v_{i,j}\, b\, K_1(v_{i,j} b) \qquad (41)$$

where

$k_i(i,j)$ represents the first - or initial - year for which the records of both stations i and j exist.

$k_f(i,j)$ represents the final year for which the records of both stations i and j exist.

$f_{i,k}$ is the standardized amount of rainfall during year k at station i (could be also a monthly or seasonal amount).

TABLE N° 1

Station N°	Years or Record	Mean mm	St.Dev. mm	Station N°	Years of Record	Mean mm	St.Dev. mm	Station N°	Years of Record	Mean mm	St.Dev. mm
1	58-71	1445	193	10	43-71	1440	194	19	52-64	1325	221
2	55-71	1412	131	11	56-71	1341	223	20	58-71	1252	276
3	55-65	1269	234	12	50-71	1427	261	21	52-71	1269	246
4	51-71	1404	127	13	44-71	1318	200	22	52-71	1514	318
5	56-71	1500	156	14	58-71	1296	210	23	61-71	1462	283
6	55-71	1342	189	15	53-71	1308	225	24	48-65	1370	194
7	53-71	1328	234	16	61-71	1240	228	25	52-65	1429	191
8	48-71	1294	248	17	61-71	1255	215	26	46-65	1452	199
9	54-71	1144	195	18	54-71	1155	206				

Table 1 - Description of Rainfall Data used in the Central Venezuela Example.

$v_{i,j}$ stands for the distance between stations i and j.

N is the number of stations.

Equation 41 will yield of value of b where we have weighted all the stations according to the length of their record. In this example, a value of b was obtained equal to

$$b = 0.0234 \text{ Km}^{-1}$$

As it was discussed previously in this paper, the characteristic correlation distance is in this case about 100 Km which yields

$$r(100) = 0.0234 \times 100 \times k_1 (0.0234 \times 100) = 0.21$$

Fitting now an exponential decay at this distance, we get

$$r(100) = e^{-100h} = 0.21; \quad h = 0.0156$$

Thus the two equations to be used for describing the spatial correlation structure are

$$r(v) = 0.0234 \, v \, K_1 (0.0234 \, v) \qquad (42)$$

and

$$r(v) = e^{-0.0156v} \qquad (43)$$

where v is in kilometers.

The variance reduction factors due to spatial sampling $F_2(N)$ are given in Table 2. We can see that $F_2(N)$ decreases much more when going from 1 station to 5 than when the gages are increased from 5 to 100.

Before reaching any practical conclusions we need to estimate the variance reduction factor due to temporal sampling $F_1(T)$. For this

TABLE N° 2

	$F_2(N)$ Bessel Type of Correlation		$F_2(N)$ Exponential Correlation	
N	Random	Stratified	Random	Stratified
1	1.00	1.00	1.00	1.00
2	0.66	0.60	0.65	0.60
3	0.55	0.48	0.54	0.48
5	0.46	0.40	0.43	0.39
10	0.40	0.36	0.37	0.34
20	0.37	0.34	0.33	0.31
100	0.34	0.32	0.31	0.29

Table 2 - Variance Reduction Factor due to Spatial Sampling with $Ab^2 = 16.43$ and $Ah^2 = 7.30$ (Central Venezuela Example).

we will have to estimate first the autocorrelation coefficient representative in time for the whole area.

ρ can be estimated as the solution of

$$\sum_{i=1}^{N} \sum_{i=1}^{N} \sum_{k=k'_i(i,j)}^{k'_f(i,j)} f_{i,k} \; f_{j,k+1} = \rho \sum_{i=1}^{N} \sum_{j=1}^{N} [k'_f(i,j) - k'_i(i,j) + 1] \; v_{i,j} \; b \; K_1(v_{i,j} b) \qquad (44)$$

where $k'_i(i,j)$ and $k'_f(i,j)$ represent the initial and final year for which both the record of station i and the record of station j in the following year exist. The other terms are the same as in Equation 41 where b has been estimated. For the annual data in this example, the obtained ρ was 0.00.

For $\rho = 0.00$, the values of the variance reduction factor due to temporal sampling are given in Table 3.

Combining Tables 2 and 3, we can estimate the efficiency of different network schemes for the area considered. In the case of one station in operation during 20 years we can expect a <u>total variance reduction factor</u> of

$$F_1(T) \cdot F_2(N) = 0.050 \times 1 = 0.050$$

In other words, this network will produce an estimate of the long-term areal mean with a variance of the order of 5% of the variance of the point rainfall process (Equation 27). If we wish to accomplish that type of precision in a lapse of 10 years we will need

$$F_2(N) = \frac{0.050}{F_1(10)} = 0.50$$

TABLE N° 3

T	$F_1(T)$	T	$F_1(T)$
1	1.000	15	0.067
2	0.500	20	0.050
5	0.333	30	0.033
5	0.20	50	0.020
7	0.140	75	0.013
10	0.100	100	0.010

Table 3.- Variance Reduction Factor due to Temporal Sampling with $\rho = 0.0$

This corresponds to $N = 4$ stations in the case of random sampling for both correlation structures given by Equations 42 and 43 and to $N = 3$ stations when the network is stratified.

It is interesting to observe that the same precision of 0.050 can not be obtained in a lapse of 5 years because it will be needed that

$$F_2(N) = \frac{0.050}{0.20} = 0.25$$

which is a value smaller than the asymptotic one of Equation 28 when N goes to infinity. From the graphs it can be seen that with $Ah^2 - 7.30$ and $F_2(N) = 0.25$, the corresponding value of N is still larger than 100. We thus have the important conclusion that trading time-vs-space in hydrologic data collection can be done when we do not reduce the time interval too much, but no "miracles" can be expected in short times even from the most dense of all possible networks.

Table 4 presents the combined factors $F_1(T) \times F_2(N)$ for the example under consideration. This product represents the total reduction in variance relative to variance of point rainfall when estimating the long-term areal mean with N stations during T years. The table in question was constructed for the Bessel-type correlation function given by Equation 42 but the use of Equation 43 gives practically identical results. It can be seen that even for quite a small number of years - like 2, 5 or 10 - 5 stations will accomplish most of the possible reduction in variance and there is little justification in going over this number. There is indeed a general result as will be seen in the next example. It can also be observed that $F_1(T)$ weights more than $F_2(N)$ in the reduction of the variance of the long-term areal mean: when $T = 5$ years,

TABLE N° 4

$F_1(T) \times F_2(N)$

N	T = 2	T = 5	T = 10
1	0.500	0.200	0.100
2	0.320	0.132	0.066
3	0.275	0.110	0.055
5	0.230	0.092	0.046
10	0.200	0.080	0.040
20	0.185	0.074	0.037
10	0.170	0.068	0.034

Table 4.- Total Factor of Variance Reduction due to Temporal and Spatial Sampling in the Central Venezuela Region. Bessel Type of Correlation with a Randomly Designed Network.

$F_1(T) = 0.200$ and a value $F_2(N) = 0.200$ can not be obtained in this example. This shows again that trading time-vs-space, although possible and in some instances necessary, is an expensive proposition.

It is important to emphasize that the form of the correlation function-Bessel type or exponential decay- does not seem to matter if the fitting has been done with good criteria. Thus Equations 42 and 43 give practically the same results, but if after obtaining $b = 0.0234$ we had tried to fit an exponential decay using distances far from the characteristic correlation distance (100 Km), the results would have been non-sensical.

For example, if we make $v = 10$ Km in equation

$$r(v) = 0.0234\, v\, K_1\, (0.0234\, v)$$

we obtain $r\,(10) = 0.942$, and if we now adjust

$$r(v) = e^{-hv}$$

with $r(10) = 0.942$, the value of h becomes equal to $0.0060\ Km^{-1}$ and $Ah^2 = 1.08$, in contrast with $Ah^2 = 7.30$ corresponding to Equation 43. With $Ah^2 = 1.08$ it is impossible-even with the most dense network- to obtain a reduction $F_1(T) \cdot F_2(N) = 0.050$ (equivalent to one station during 20 years) in a period of 10 years; on the other hand, with $Ah^2 = 7.30$ it can be accomplished with 4 stations during 10 years.

Another important conclusion is then that the functional shape of the correlation in space does not affect the results provided the fitting of the parameters has been done properly. On the contrary, if the parameters of the correlation structure are evaluated by fitting the functional form to r's corresponding to distances very different from the characteristic correlation distance of the region, then the results will be totally misleading.

We have been considering the evaluation of the variance of the estimate of long-term areal mean precipitation. This variance was expressed as

$$\text{Var}\ [\bar{P}] = \sigma_p^2 \cdot F_1(T) \cdot F_2(N)$$

and a scheme was given to evaluate $F_1(T)$ and $F_2(N)$. This scheme also provides a method for quantitatively trading time-vs-space in network design. Our work in this section would not be complete without a discussion of the estimation of $\bar{P}$, the regional precipitation mean value, and σ_p^2, the variance of the point rainfall process.

<u>Evaluation of $\bar{P}$ and σ_p^2.</u> - σ_p^2 can be estimated from point records at each station. Because the process has been assumed stationary, σ_p^2 has to be the same in all stations and thus we can put back to back all the individual records and compute the variance of that series of data. This variance will be an estimate of σ_p^2. For the region in Central Venezuela, the obtained result was $5.44 \times 10^4\ mm^2$ of rainfall, and thus the total reduction of $F_1(T) \cdot F_2(N)$ should be applied to this value in order to obtain the variance of our estimated long-term areal mean. In the case when $F_1(T) \cdot F_2(N) = 0.50$, which we discuss previously, we are obtaining an estimate of the longterm areal mean that has a variance of

$$\text{Var}\ [\bar{P}] = 0.050 \times 5.44 \times 10^4\ mm^2$$

or equivalently

$$\text{St.Dev.}(\bar{P}) = 52.16\ mm.$$

In order to judge the magnitude of Var $(\bar{P})$, an estimate of $\bar{P}$ is needed. Being this an important point and truly the justification itself of a network, it seems convenient to present a scheme which will optimize the estimation of $\bar{P}$ taking into account the different lengths of records at the stations, the spatial correlation among the gages and the correlation in time existing at each record. Depending on these factors we can conceive the application of a weight at each station, such that

$$\bar{P} = \sum_{i=1}^{N} \alpha_i \frac{1}{T_i} \sum_{t=T_{i,i}}^{T_{f,i}} f(x_i, t) \qquad (45)$$

where

$$\sum_{i=1}^{N} \alpha_i = 1$$

and

$f(x_i,t)$ represents the amount of rainfall at the station of spatial coordinates x_i during the year (month or season) t.

N is the number of stations

$T_{i,i}$ stands for the initial year of record at station i

$T_{f,i}$ is the final year of record at station i

and

$$T_i = T_{f,i} - T_{i,i} + 1.$$

Without loss of generality we can assume $T_{i,i} \geq T_{i,j}$ for $i > j$.

The problem is to find the α_i which minimize the variance of $\bar{P}$. We can write

$$E\left[\sum_{i=1}^{N}\alpha_i \frac{1}{T_i}\sum_{t=T_{i,i}}^{T_{f,i}} f(x_i,t)\right]^2 = E\left\{\sum_{i=1}^{N}\frac{\alpha_i^{\,2}}{T_i^{\,2}}\left[\sum_{t=T_{i,i}}^{T_{f,i}} f^2(x_{i,t}) + \right.\right.$$

$$\left. + 2\sum_{t=T_{i,i}}^{T_{f,i}-1}\ \sum_{t'=t+1}^{T_{f,i}} f(x_i,t)\ f(x_i,t')\right] + 2\sum_{i=1}^{N-1}\ \sum_{i'=i+1}^{N}\frac{\alpha_i\alpha_{i'}}{T_iT_{i'}}$$

$$\left.\left[\sum_{t=T_{i,i}}^{T_{f,i}}\ \sum_{t'=T_{i,i'}}^{T_{f,i'}} f(x_i,t)\ f(x_i',t')\right]\right\} = \sigma_p^{\,2}\left\{\sum_{i=1}^{N}\frac{\alpha_i^{\,2}}{T_i^{\,2}}\left[\sum_{t=T_{i,i}}^{T_{f,i}} 1 + \right.\right.$$

$$\left.\left. + 2\sum_{t=T_{i,i}}^{T_{f,i}-1}\ \sum_{t'=t+1}^{T_{f,i}}\rho^{t'-t}\right] + 2\sum_{i=1}^{N-1}\ \sum_{i'=i+1}^{N}\frac{\alpha_i\alpha_{i'}}{T_iT_{i'}}\, r(x_i-x_{i'})\sum_{t=T_{i,i}}^{T_{f,i}}\ \sum_{t'=T_{i,i'}}^{T_{f,i'}}\rho^{|t'-t|}\right\}$$

(46)

Now let

$$S_{ii'} = \sum_{t=T_{i,i}}^{T_{f,i}}\ \sum_{t'=T_{i,i'}}^{T_{f,i'}}\rho^{|t'-t|} \qquad (47)$$

we can write Equation 46 as

$$\mathrm{Var}(\bar{P}) = \sigma_p^{\,2}\left[\sum_{i=1}^{N}\frac{\alpha_i^{\,2}}{T_i^{\,2}} S_{ii} + 2\sum_{i=1}^{N-1}\ \sum_{i'=i+1}^{N}\frac{\alpha_i\alpha_{i'}}{T_iT_{i'}}\, r(x_i-x_{i'})\ S_{ii'}\right] =$$

$$= \sigma_p^{\,2}\left[\sum_{i=1}^{N}\ \sum_{i'=1}^{N}\frac{\alpha_i\alpha_{i'}}{T_iT_{i'}}\, r(x_i-x_{i'})\ S_{ii'}\right]$$

Making up now the Lagrangian with the purpose of minimizing $\mathrm{Var}(\bar{P})$ we write

$$L = \sum_{i=1}^{N} \sum_{i'=1}^{N} \frac{\alpha_i \alpha_{i'}}{T_i T_{i'}} r(x_i - x_{i'}) S_{ii'} + \lambda \left(\sum_{i=1}^{N} \alpha_i - 1 \right)$$

from where we get

$$\frac{\partial L}{\partial \alpha_i} = \left[\frac{2\alpha_i}{T_i^2} S_{ii} + 2 \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{\alpha_j}{T_i T_j} r(x_i - x_j) S_{ij} \right] + \lambda = 0 =$$

$$= \sum_{j=1}^{N} \frac{\alpha_j}{T_i T_j} r(x_i - x_j) S_{ij} + \lambda \quad \text{for} \quad i = 1,2,\ldots,N \qquad (48)$$

and

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{N} \alpha_i - 1 = 0 \qquad (49)$$

Equations 48 and 49 represent a system of N+1 equations to get the N+1 unknowns given by the α_i's and λ. This scheme was applied to the Central Venezuela region; the weights α_i for each of the 26 stations are given in Table 5. The mean value $\bar{P}$ obtained from Equation 45 was 1403 mm. The negative weights of some stations can be interpreted as a subtraction of information which has been incorporated by other stations in order to produce a minimum variance in the regional rainfall $\bar{P}$ when using the records analyzed. Table 6 shows the weights when only 15 stations instead of all 26 are used in the estimation of $\bar{P}$, it is interesting to note the changes in weights revealing the relative increase-or decrease-in the importance of some stations. This kind of procedure can be used to investigate which are important-or superfluous-sites when analyzing an existing network.

TABLE N° 5

Station	Weight $\alpha_i \times 10^3$	Station	Weight $\alpha_i \times 10^3$	Station	Weight $\alpha_i \times 10^3$
1	79.56	10	166.50	19	-29.52
2	65.09	11	-51.54	20	17.54
3	-46.57	12	32.29	21	50.29
4	12.73	13	111.54	22	137.56
5	54.99	14	-16.07	23	20.81
6	13.38	15	- 2.17	24	51.31
7	3.44	16	70.79	25	12.71
8	79.83	17	-25.24	26	126.83
9	-35.21	18	99.08		

Table 5.- Weights of Stations for Estimation of Mean Value with Minimum Variance. Central Venezuela Example with 26 Stations.

TABLE N° 6

Station	Weight $\alpha_i \times 10^3$	Station	Weight $\alpha_i \times 10^3$	Station	Weight $\alpha_i \times 10^3$
1	82.79	10	138.42	17	-21.20
2	65.43	13	127.25	18	111.82
3	-34.22	14	-35.17	22	170.31
4	63.53	15	- 2.83	24	57.69
8	78.94	16	68.74	26	128.48

Table 6.- Weights of Stations for Estimation of Mean Value with Minimum Variance. Central Venezuela Example with 15 Stations.

2) New South Wales (Australia). - This second example considers a catchment near Lismore along the northern coast of New South Wales as described by Eagleson (1967). It was seen before in this paper that this region has a characteristic correlation distance of 20 miles which will be used in the fitting of the spatial correlation structure.

From Eagleson's data we have that

$$r(20 \text{ mi}) = 0.47$$

which in turn yields

$$h = 0.037 \text{ mi}^{-1} \quad \text{and} \quad b = 0.0684 \text{ mi}^{-1}$$

in Equations 6 and 8 respectively. Thus

$$r(v) = e^{-0.037v} \quad ; \quad Ah^2 = 1.71$$

$$r(v) = 0.0684\, v\, K_1\, (0.0684\, v) \quad ; \quad Ab^2 = 5.85$$

will be used as spatial correlation functions in this example.

Table 7 gives the values of $F_2(N)$ for a randomly designed network. It is observed that after 3 stations - or even 2 - there is a very small decrease in the variance reduction factor due to spatial sampling. The relative importance of this decrease will be even slighter when $F_1(T)$ is brought into action. Thus for 10 years of data and assuming $\rho(1)$ in time equal to zero, we get $F_2 = 0.62$. Fixing $N = 5$,

$$F_1(10) \cdot F_2(5) = 0.062$$

Similarly,

$$F_1(10) \cdot F_2(100) = 0.053.$$

TABLE N° 7

N	$F_2(N)$ Bessel Correlation	$F_2(N)$ Exponential Correlation
1	1.00	1.00
2	0.76	0.76
3	0.69	0.69
5	0.62	0.62
10	0.56	0.56
20	0.54	0.54
100	0.53	0.53

Table 7.- Variance Reduction Factor due to Spatial Sampling with a Random Design for $Ab^2 = 5.85$ and $Ah^2 = 1.71$ (New South Wales Example).

There is only a decrease of 1% in the variance of the estimated long-term areal mean when increasing the number of stations from 5 to 100, hardly economical.

5. GENERAL FRAMEWORK FOR ESTIMATING AREAL MEAN FOR RAINFALL EVENTS

In this case we want to estimate mean rainfall over a certain area A

$$Z(A) = \frac{1}{A} \int_A f(x_i)\, dx_i \qquad (50)$$

where $f(x_i)$ denotes rainfall depth on the point of spatial coordinates x_i in the space A. In practice, the hydrologist estimates Z(A) by the arithmetic mean of N point samples represented by the raingage stations

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^{N} f\,(x_i) \qquad (51)$$

The performance of the network can be characterized by the variance

$$E\,[\bar{Z} - Z\,(A)]^2 = \sigma_N^{\,2} \qquad (52)$$

where it is important to understand that the expectation is taken over all possible outcomes which the rainfall $f(x_i)$ may produce over the space A. Z(A) is now a random variable and this is an important difference with Equation (19) when estimating long-term areal mean rainfall. Notice also that the time element does not play now an explicit role.

We can write Equation 52 as

$$E\left[\frac{1}{N}\sum_{i=1}^{N} f(x_i) - \frac{1}{A}\int_A f(x_i)\,dx_i\right]^2 =$$

$$= \frac{\sigma_p^2}{N^2}\left[\sum_{i=1}^{N} 1 + 2\sum_{i=1}^{N-1}\sum_{i'=i+1}^{N} r(x_i - x_{i'})\right] - \frac{2}{NA}\,\sigma_p^2$$

$$\sum_{i=1}^{N}\int_A r(x_i - x_{i'})\,dx_{i'} + \frac{\sigma_p^2}{A^2}\int_A\int_A r(x_i - x_{i'})\,dx_i\,dx_{i'} \tag{53}$$

where σ_p^2 represents the point variance of the process. Equation 53 will now be evaluated for two types of network designs: simple random sampling and stratified sampling.

<u>Variance Reduction Factor due to Random Sampling.</u>- In this case, Equation 53 can be written

$$\sigma_N^2 = \frac{\sigma_p^2}{N^2}\left\{N + E\,[r(x_i - x_{i'})|A]\cdot N\cdot(N-1)\right\} - \frac{2}{NA}\cdot\sigma_p^2\cdot$$

$$\cdot\frac{N}{A}\cdot\int_A\int_A r(x_i - x_{i'})\,dx_i\,dx_{i'} + E[r(x_i - x_{i'})\,|\,A]\,\sigma_p^2 =$$

$$= \frac{\sigma_p^2}{N^2}\left\{N + N(N-1)\cdot E[r(x_i - x_{i'})|A]\right\} - 2\sigma_p^2\,E[r(x_i - x_{i'})\,|\,A] +$$

$$+ E[r(x_i - x_{i'})|A]\,\sigma_p^2 = \frac{\sigma_p^2}{N} - \frac{\sigma_p^2}{N}\cdot E\,[r(x_i - x_{i'})|A] =$$

$$= \sigma_p^2\,F_2(N) \tag{54}$$

where

$$F_2(N) = \frac{1 - E[r(x_i - x_{i'})|A]}{N}$$

Variance-Reduction Factor due to Stratified Sampling.- Equation 53 is written now

$$\sigma_p^2 \left\{ \frac{1}{N^2} [N + \sum_{i=1}^{N} \sum_{i'=1}^{N} r(x_i - y_{i'}) - \sum_{i=1}^{N} r(x_i - y_i)] - \right.$$

$$- \sigma_p^2 \frac{2}{NA} \frac{N}{A} \sum_{i=1}^{N} \int_A \int_{a_i} r(x_i - y_{i'}) dx_i dy_{i'} + \frac{\sigma_p^2}{A^2} \int_A \int_A r(x_i - y_{i'}) dx_i dy_{i'}$$

whose notation has been explained in the previous section of this paper and which can be written as

$$\frac{\sigma_p^2}{N^2} \{N + N^2 E[r(x_i - y_{i'}) | A] - N E[r(x_i - y_{i'}) | (A/N)]\}$$

$$- \quad 2\sigma_p^2 E[r(x_i - y_{i'}) | A] + \sigma_p^2 E[r(x_i - y_{i'}) | A] =$$

$$= \quad \sigma_p^2 \; F_2 \; (N) \tag{55}$$

where

$$F_2(N) = \frac{1 - E[r(x_i - y_{i'}) | (A/N)]}{N}$$

and the assumption has been made of one station per stratum.

Figures 9, 10, 11 and 12 show Equations 54 and 55 for the two types of correlation functions used in this paper. The rationale behind this graphication is the same as the one for Figures 5, 6, 7 and 8 already explained. It is observed that for very large values of Ah^2, Ab^2.

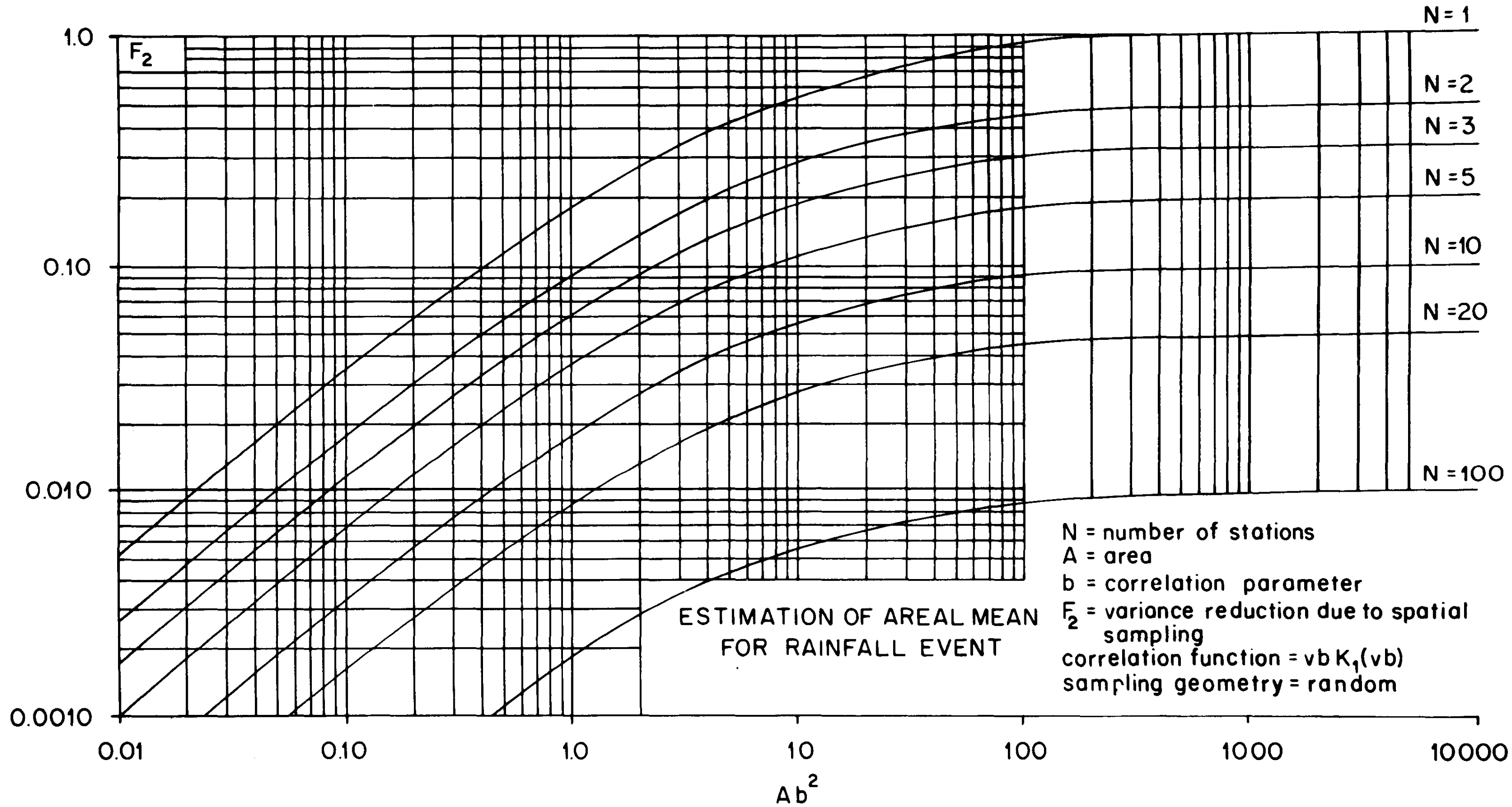

Figure 9 - Variance Reduction Factor due to Spatial Sampling with Random Design. Used in the Estimation of Areal Mean of Rainfall Event with $r(v) = bv\ K_1(bv)$.

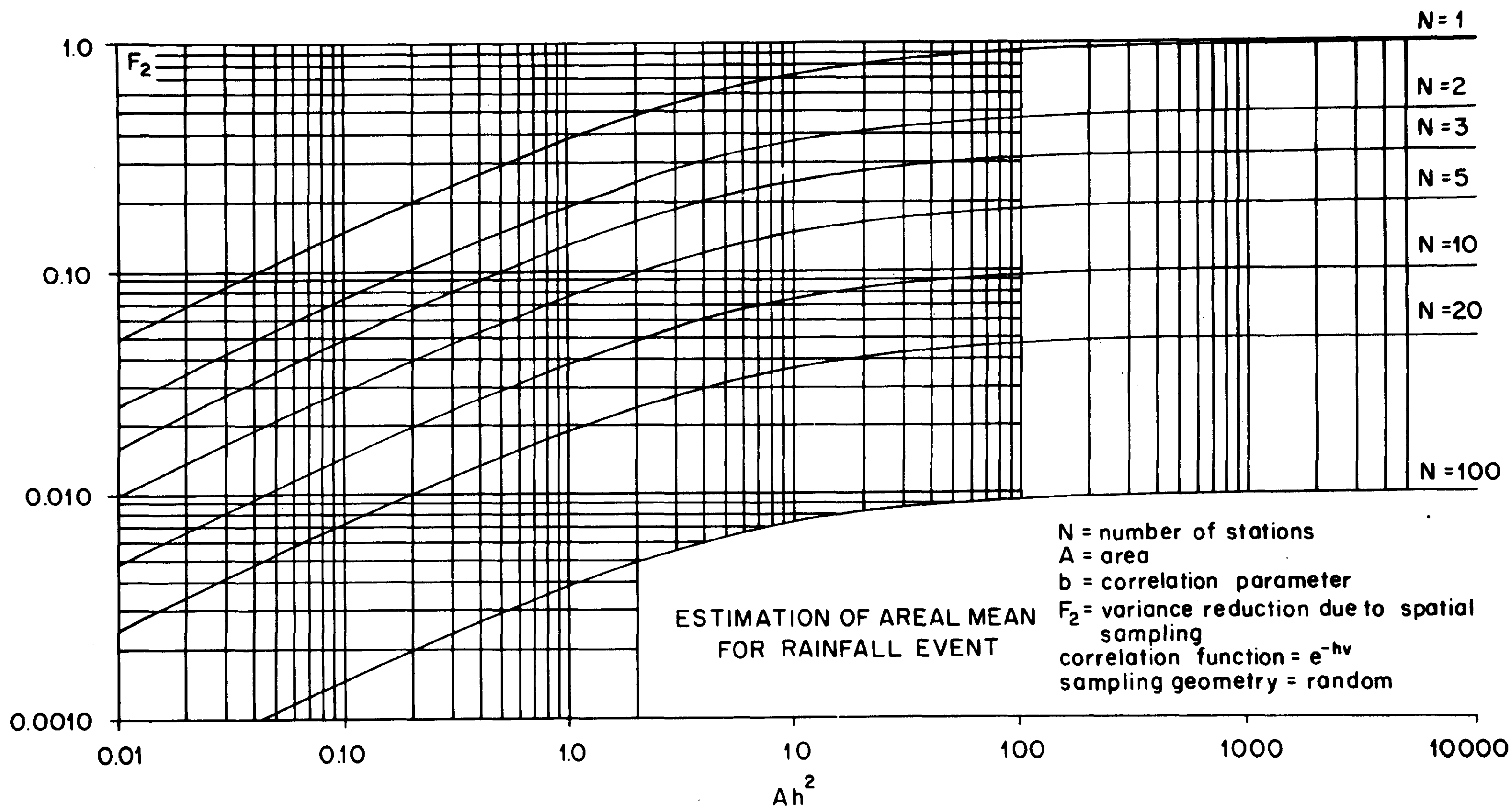

Figure 10 - Variance Reduction Factor due to Spatial Sampling with Random Design. Used in the Estimation of Areal Mean of Rainfall Event with $r(v) = e^{-hv}$.

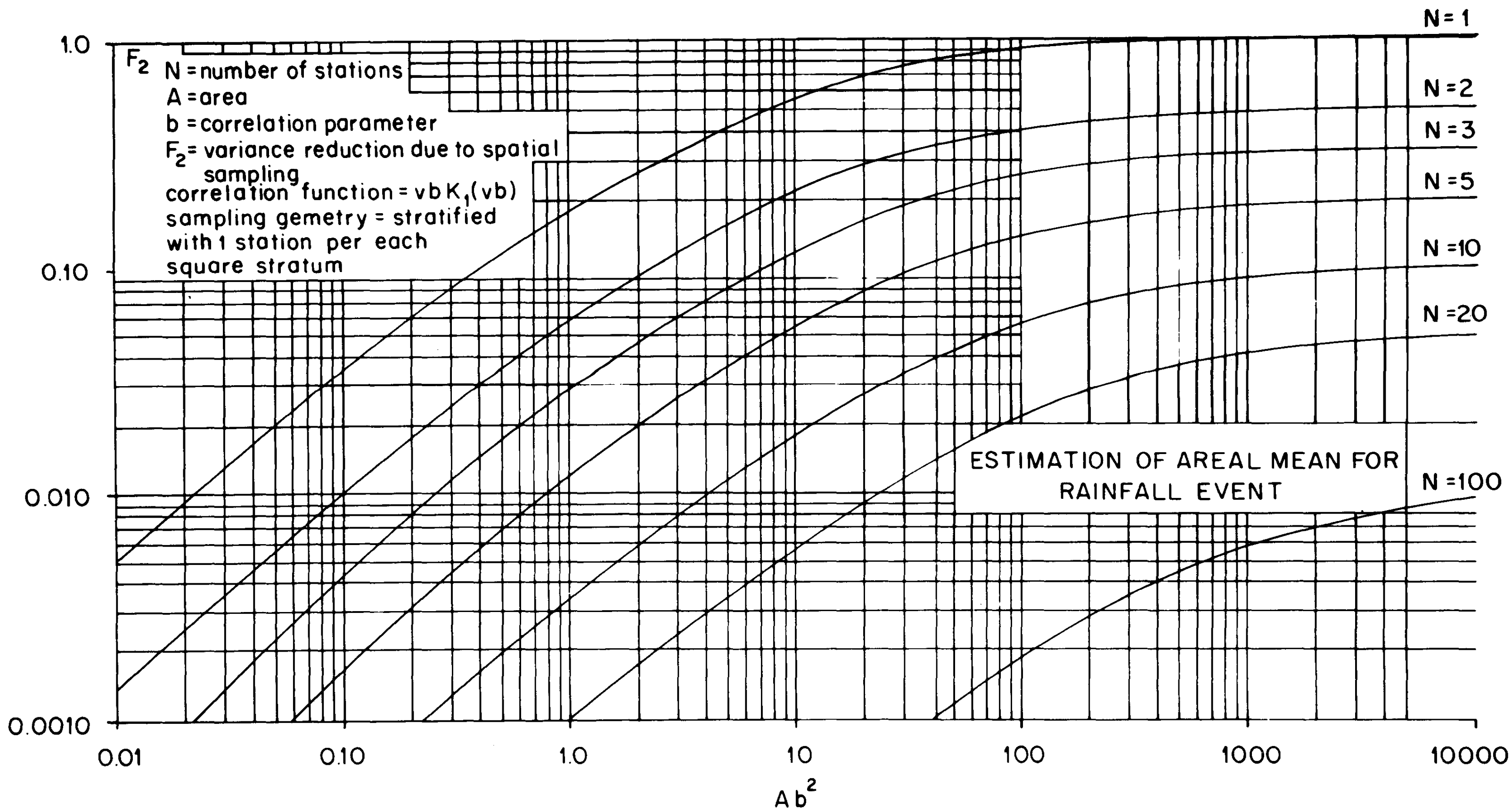

Figure 11 - Variance Reduction Factor due to Spatial Sampling with Stratified Design. Used in the Estimation of Areal Mean of Rainfall Event with $r(v) = bv\ K_1(bv)$.

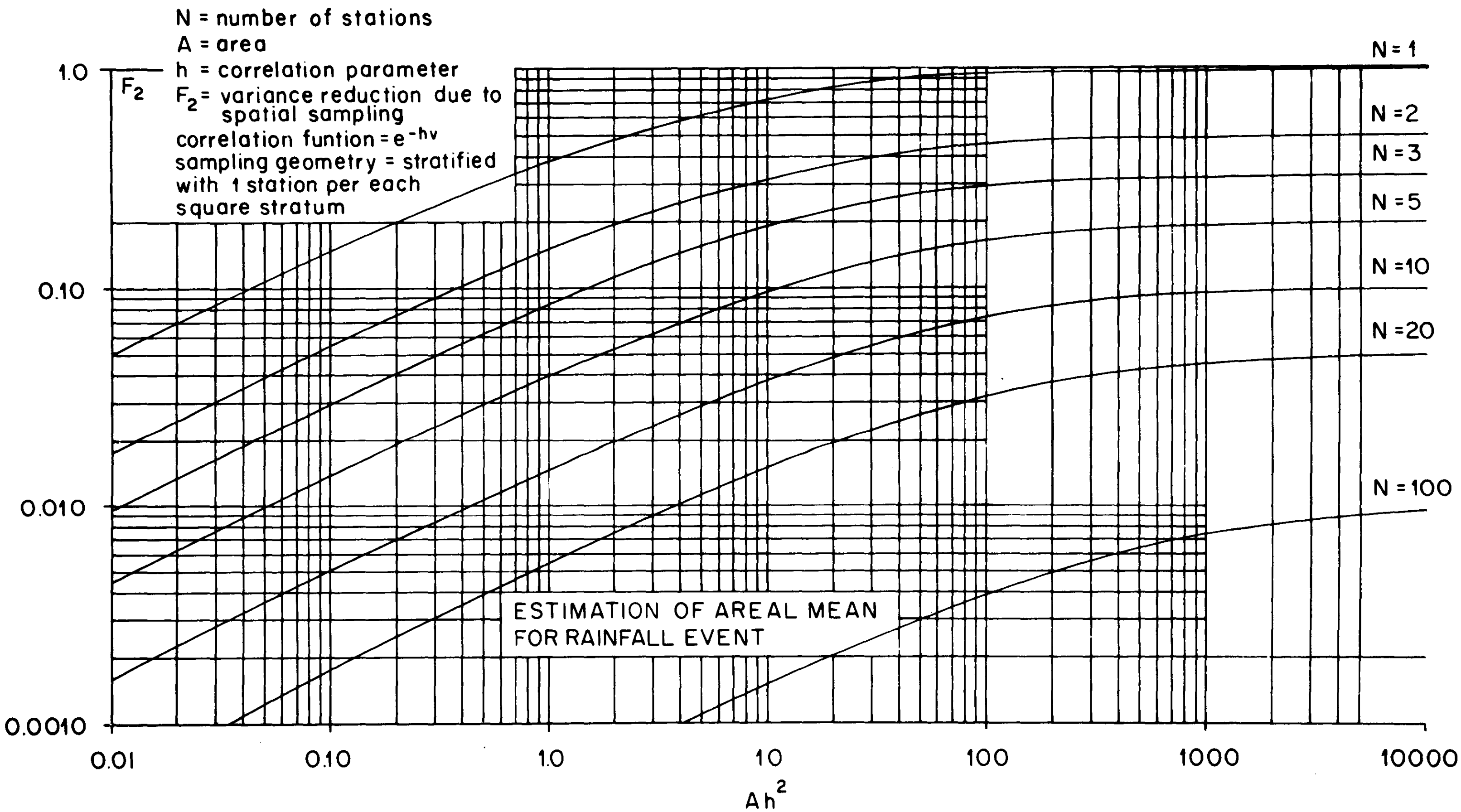

Figure 12 - Variance Reduction Factor due to Spatial Sampling with Stratified Design. Used in the Estimation of Areal Mean of Rainfall Event witn $r(v) = e^{-hv}$.

the curve corresponding to $N = 1$ goes to $F_2(1) = 1$, meaning that one gage alone will give an estimte of the areal mean with variance equal to the point variance of the process σ_p^2. Nevertheless for small values of Ah^2, Ab^2, even one gage alone will produce an estimate of the areal mean whose variance is considerably smaller than σ_p^2.

Equations 54 and 55 will be used now for the sampling in space of convective and cyclonic storms with the purpose of estimating the areal mean of these types of events.

Analysis of Convective Storms.- Under the section headed "The Rainfall Process" we studied sensible values of the correlation parameter - h or b - depending on the intensity of the storm. It was seen that realistic values are of the order

$P_t(0) = 5''$	$h = 0.080$,	$b = 0.130$
$P_t(0) = 2''$	$h = 0.200$,	$b = 0.355$
$P_t(0) = 0.75''$	$h = 0.533$,	$b = 0.930$

Tables 8 and 9 give the results of estimating the areal mean values of these three types of storms over areas of 1500 mi^2, 500 mi^2 and 50 mi^2 when a network of 3 gages is used by the engineer. It is seen that the variance reduction factor $F_2(N)$ is much smaller for the case of a heavy storm in all cases, meaning that the error made in these cases is of less importance. The area has a logical relation, the larger the area, the larger $F_2(N)$, meaning that more stations are needed to maintain the same precision in the estimate. It is also important to notice that stratification can significantly reduce the value of $F_2(N)$; with only 3 stations, reduction largely depends on the size of the area in consideration

TABLE N° 8

$F_2(N)$ for Bessel Correlation Function, N = 3

Area(mi)^2	Random Design			Stratified Design		
	$P_t(0)$=5"	$P_t(0)$=2"	$P_t(0)$=0.75"	$P_t(0)$=5"	$P_t(0)$=2"	$P_t(0)$=0.75"
1,500	0.240	0.320	0.320	0.180	0.280	0.320
500	0.175	0.280	0.320	0.110	0.230	0.310
50	0.055	0.160	0.260	0.025	0.094	0.210

Table 8 - Variance Reduction Factors for Different Combinations of Areas and Convective Storms with a Network of 3 Stations.

TABLE N° 9

$F_2(N)$ for Experimental Correlation Function, N = 3

	Random Design			Stratified Design		
Area mi	$P_t(0)=5''$	$P_t(0)=2''$	$P_t(0)=0.75''$	$P_t(0)=5''$	$P_t(0)=2''$	$P_t(0)=0.75''$
1,500	0.250	0.310	0.320	0.190	0.280	0.320
500	0.190	0.280	0.320	0.140	0.230	0.310
50	0.082	0.165	0.260	0.050	0.110	0.210

Table 9 - Variance Reduction Factors for Different Combinations of Areas and Convective Storms with a Network of 3 Stations

because the larger the area the more similar are the random and the stratified scheme.

From Figures 9 through 12 it is seen that the number of stations now plays a very important role in the estimation process, somewhat differently than for the estimation of long-term areal means.

The question again arises of what is an acceptable precision? Although the answer should include economic considerations as previously noted, an engineering idea can be obtained depending on the value of σ_p^2 in Equations 54 and 55. An idea of the value of σ_p^2 can be obtained by simple generation of random storm centers over the area; each of these centers represents a storm with an areal pattern as given by Equatio 14, and in this manner we can "measure" the amount of rainfall recorded by our network. This type of experiment was done for different combinations of storms, areas and networks; as an example 20 events with $P_t(0) = 5''$ were simulated over an area of 1250 mi^2 with a network of 3 gages randomly located. The actual mean areal depth of the event is approximately 0.603" over 1250 mi^2 which is obtained from Equation 14. The network on the other hand yields a mean depth of 0.530" for the 20 events analyzed, the variance obtained from the simulated records was $\sigma_p^2 = 1.32$ in. which is quite large considering the area involved.

Figures 9 through 12 should prove of help when evaluating the magnitude of the error made in the input estimation when using rainfall-runoff models.

Analysis of Cyclonic Storms.- Hurricane Connie in the Baltimore area was described with Equation 17 using $h = 0.009$ and $b = 0.016$. Assuming we want to estimate the areal mean depth of this event over an area of 1500 mi^2 with a network of 3 gages we obtain

$$Ah^2 = 0.12 \rightarrow F_2(3) = 0.054 \text{ (random)}; \quad F_2(3) = 0.032 \text{ (stratified)}$$

$$Ab^2 = 0.38 \rightarrow F_2(3) = 0.032 \text{ (random)}; \quad F_2(3) = 0.014 \text{ (stratified)}$$

Because of the areal coverage of the storm and its intensity there are large differences in $F_2(N)$ according to the type of correlation structure used in the analysis. From a conceptual point of view the authors would prefer to assume the Bessel-type of correlation structure rather than the exponential one as discussed in the first section of the paper.

CONCLUSIONS

The following conclusions appear to be in order from this study:

1) In the design of rainfall networks it is important to consider spatial correlation, time correlation, number of stations and network geometry.

2) When adjusting a spatial correlation structure it is important to do so at the "characteristic correlation distance."

3) For estimating long-term areal mean values of precipitation, the commanding factor is the length of time the network has been in operation.

4) Trading time-vs-space is possible in many cases when estimating long-term areal mean values. Nevertheless it is an expensive proposition.

5) It is possible to evaluate the variance of the areal mean for both the long-term case and the event case as function of the factors described in 1).

6) The functional form of the correlation in space seems to have importance only for the case of estimating mean areal precipitation from cyclonic storms.

REFERENCES

Boyer, M.C. A Correlation of the Characteristic of Great Storms, Trans.Am. Geophys. Union, 38 (2), 233-236, 1957.

Davis, D.R., and W.M. Dvoranchik, Evaluation of the Worth of Additional Data, Water Resources Bull., 7 (4), 700-707, 1971.

Dawdy, D.R., H.E. Kubik, and E.R. Close, Worth of Streamflow Data for Project Design - A Pilot Study, Water Resources Res., 6(4), 1045-1050, 1970.

Duckstein, L., and C.C. Kisiel, Efficiency of Hydrologic Data Collection Systems, Role of Type I and II Errors, Water Resources Bull., 7(3), 592-604, 1971.

Eagleson, P.S., Optimum Density of Rainfall Network, Water Resources Res., 4 (3), 1021-1033, 1967.

Eagleson, P.S., Dynamic Hydrology, 462 pp., McGraw-Hill, 1970.

Fogel, M.M., and L. Duckstein, Point Rainfall Frequencies in Convective Storms, Water Resources Res., 5 (6), 1229-1237, 1969.

Ghosh, B., Random Distances Within a Rectangle and Between Two Rectangles, Bull. Calcutta Math. Soc., 43, 17-24, 1951.

Hendrick, R.L., and G.H. Comer, Space Variations of Precipitation and Implications for Rain Gage Network Design, Journal of Hydrology, 10, 151-163, 1970.

Karhunen, K., Ueber ein Extrapolationproblem in dem Hilbertschen Raum, Compte Rendu Congres de Mathematiciens Scandinaves, Trondheim, 35-41, 1952.

Matalas, N.C., and E.J. Gilroy, Some Comments on Regionalization in Hydrologic Studies, Water Resources Res., 4 (6), 1361-1369, 1968.

Matalas, N.C., Optimum Gaging Station Location, in "The Progress of Hydrology," Proceedings 1st Inter. Seminar for Hydrology Professors, University of Illinois, 473-489, 1969.

Matern, B., Metoder att uppskatta noggrannheten vid linje och provytetaxering, Medd. fran statens skigsforskningsinst. 36, No. 1, 1947.

Matern, B., Spatial Variation, Medd. fran statens skigsforskninginst., 49, No.5, 1960.

Matheron, G., Les Variables Regionalisées et leur Estimation: Masson, Paris, 1965.

Matheron, G., Random Functions and their Application in Geology, in Geostatistics, edited by D.F. Merrian, Plenum Press, 1970.

Moss, M.M., Optimum Operating Procedure for a River Gaging Station Established to Provide Data for Design of a Water Supply Project, Water Resources Res., 6 (4), 1051-1061, 1970.

Rodda, J.C. et al., Hydrologic Network Design - Needs, Problems and Approaches, World Meteor. Organization, Report 12, 1969.

Rodriguez-Iturbe, I., General Report on Data Network Design, Proceedings International Symposium on Uncertainties in Hydrologic and Water Resources Systems, 18 p., Tucson, Arizona, 1972.

Rodriguez-Iturbe, I., E.H. Vanmarcke and J.C. Schaake, Problems of Analytical Methods in Hydrologic Data Collection, Proceedings International Symposium on Uncertainties in Hydrologic and Water Resources Systems, 4.8-1, 4.8-21, Tucson, Arizona, 1972.

Stol, P.T., The Relative Efficiency of the Density of Rain-Gage Networks, Journal of Hydrology, 15, 193-208, 1972.

Tschannerl, G., A Decision Framework for the Efficient Use of Data Sources, Center for Population Studies, No. 4, Harvard University, 1970.

Whittle, P., On Stationary Processes in the Plane, Biometrika, 41, 434-449, 1954.

Woolhiser, D.A., and H.C. Schwalen, Area-Depth-Frequency Relations for Thunderstorm Rainfall in Southern Arizona, Tech. Paper 527, Agr. Engr. Dept., Ariz. Agr. Expt. Sta., Univ. Arizona, 1959.

Yaglom, A.M., An Introduction to the Theory of Stationary Random Functions, Prentice-Hall, 1962.

TITLES FOR ILLUSTRATIONS

LIST OF TABLES