TIME AVERAGED AREAL MEAN OF PRECIPITATION:

ESTIMATION AND NETWORK DESIGN

Project A-050-PR.

by

Rafael L. Bras
Raúl Colón

Faculty of Engineering
Mayaguez Campus
University of Puerto Rico

from

Water Resources Research Institute
Research Center
School of Engineering
Mayaguez Campus
University of Puerto Rico

to

Office of Water Research and Technology
U. S. Department of the Interior
Washington, D. C.

"The work upon which this publication is based was supported in part by funds provided by the United States Department of the Interior as authorized under the Water Resources Research Act of 1964, Public Law 88-379".

February 1977

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (CONT'D)

TIME AVERAGED AREAL MEAN OF PRECIPITATION: ESTIMATION AND NETWORK DESIGN

by

Dr. Rafael L. Bras

Massachusetts Institute of Technology

and

Raul Colon

University of Puerto Rico

ABSTRACT

Rainfall is recognized as a random process in time and space. With this in mind, data collection is treated as an estimation problem where discrete, noisy and incomplete information is used to estimate the true rainfall process. The estimation of the unknown time averaged areal mean of precipitation is carried through a state augmentation procedure and the use of multivariate linear estimation concepts, in particular, the Kalman-Bucy filter. A technique results which can be used to analyze existing data networks; design new networks; and process data from existing networks. The procedure can handle any network configuration, and explicitly accounts for the number of stations, their particular location, the duration of observations and the measurement errors. Results are presented.

INTRODUCTION

Up to the present, the design of hydrologic data collection networks has been mostly based on minimizing cost, which translates to a problem of accessibility and maintenance of observation stations. Present-day hydrologic capabilities, though, demand high levels of accuracy in the statistics obtained from collected data. Furthermore, the scientist must know the degree of accuracy of used information and its effects on performed calculations. Fortunately, recent advances in probability and estimation theory provide the tools to systematically analyze the accuracy of alternative network designs.

The objective of this work was to develop a technique to determine the accuracy in estimating the time-averaged areal mean of precipitation from a given network configuration. The procedure explicitly considers the existing trade-off between the number and location of observations and the length of record.

The above objective is accomplished by first recognizing that rainfall is a random process in time and space. Once this fact is established, the data collection network analysis is treated as a problem of estimation where discrete, noisy and incomplete information is used to estimate the true rainfall process and the desired statistic. The multivariate estimator used is the Kalman-Bucy filter.

The authors are aware of only two previous works analytically dealing with network design to obtain the time-averaged areal mean precipitation. Eagleson [1967] used spectral analysis to arrive at his conclusions. He transformed the two-dimensional rainfall process into one-dimensional in order to apply simplified spectral theory. His results provided a measure of accuracy as a function of the number of stations, number of years of record, size of area of interest, and rainfall correlation radius. Rodríguez and Mejía [1974] worked on the time domain and obtained the mean square error of estimation, again as a function of the number of stations, area, length of record, and correlation coefficient. Station location was not considered in any of the works. Rodríguez and

Mejía provided solutions under assumptions of random and stratified random sampling.

Recently, Bras and Rodríguez [1976a, 1976b] presented a methodology for designing rainfall data collection networks for obtaining the areal mean of an event and for minimizing discharge prediction error from runoff models. The presented techniques considered not only the number of stations, but their location and possible observation errors. The reader is referred to the above two references for complete literature review of the network design problem and for an introduction to the techniques used in this work.

PROBLEM FORMULATION

The goal is to estimate the areal rainfall averaged in time. For example, the user may want to find the yearly areal mean rainfall for the purpose of yield studies. The desired statistic is defined as:

$$P = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{A} \int_{A} f(\underline{x}, t) \, d\underline{x} \tag{1}$$

where $f(\underline{x}, t)$ = rainfall accumulation at point with coordinates vector $\underline{x}$ and during period t, i,e., one year

A = area

T = number of observed time periods.

Notice that total rainfall accumulation at different time periods may be correlated.

Rodríguez and Mejía [1974] proved that for a stationary process in space and time P (Equation 1) has zero variance, therefore P is a constant. Define the total rainfall accumulation, at any point in time and space as

$$f(\underline{x}_i,\ t) = P + \varepsilon(\underline{x}_i,\ t) \tag{2}$$

where P is the mean areal time average, of the process, as defined in Equation 1, and $\varepsilon(\underline{x}_i, t)$ is a deviation from the mean at the point with coordinate vector $\underline{x}_i$ and time t. The mean of $\varepsilon(\underline{x}_i, t)$ is zero.

Discretizing at N points in space, define a vector

$$f(t) = \begin{vmatrix} f(x_1,\ t) \\ \vdots \\ f(x_N,\ t) \end{vmatrix} = \underline{1}\ P + \underline{\varepsilon}(t) \tag{3}$$

where $\underline{1}$ = Nx1 column vector with elements equal to 1.

$$\underline{\varepsilon}(t) = \begin{vmatrix} \varepsilon(x_1,\ t) \\ \vdots \\ \varepsilon(x_N,\ t) \end{vmatrix}$$

In their work, Rodríguez and Mejía argued that the covariance function of the process $f(\underline{x}, t)$ could be adequately represented by a general separable form:

$$E[\varepsilon(\underline{x}_1,\ t_1)\ \varepsilon(\underline{x}_2,\ t_2)] = \sigma^2\ \rho^{|\tau|}\ r(\tilde{x}) \tag{4}$$

where $\tau = t_2 - t_1$

$\tilde{x} = |\underline{x}_2 - \underline{x}_1|$

= distance between points 2 and 1

$r(\cdot)$ = spatial covariance function

ρ = lag one correlation coefficient

σ^2 = point variance

The above equation, together with Equation 3, implies a homogeneous and isotropic process in space as well as stationary in time.

The form of Equation 4 leads to a stationary multivariate autoregressive process of the form,

$$f(t) = \underline{1}\ P + A\{f(t-1) - \underline{1}\ P\} + B\ W(t)$$

$$= 1\ P + A\ \underline{\varepsilon}(t-1) + B\ W(t)$$

$$= [I - A]\ \underline{1}\ P + A\ f(t-1) + B\ W(t) \qquad (5)$$

where I = identity matrix

A = NxN matrix

B = NxN matrix

$W(t)$ = Nx1 vector of random variables

$$E[W(t_1)\ W^T(t_2)] = \begin{cases} I & t_1 = t_2 \\ \underline{0} & t_1 \neq t_2 \end{cases}$$

Parameters A and B of Equation 5 are given by

$$\begin{aligned} A &= S_{21}\ S_{11}^{-1} \\ BB^T &= S_{22} - S_{21}\ S_{11}^{-1}\ S_{12} \end{aligned} \qquad (6)$$

S_{21} = NxN matrix

$$= \sigma^2 \rho\{r(\underline{x}_j - \underline{x}_i)\} \qquad \forall_{j,i}$$

$$S_{11} = \sigma^2\{r(x_j - x_i)\} \qquad \forall_{j,i}$$

= NxN matrix

$$S_{12} = S_{21}^T$$

$$S_{22} = S_{11}$$

For the unique form of Equation 4, which suggests separability in time and space

$$A = \begin{bmatrix} \rho & & 0 \\ & \ddots & \\ 0 & & \rho \end{bmatrix}$$

The methodology is not limited to this type of separable covariance function. Nevertheless, whatever the covariance function, the autoregressive model given by Equation 5 must be a good representation to maintain the validity of the following development.

The spatial covariance function, $r(\cdot)$, can take many different forms, the most common being Bessel or exponential variations.

The reader is referred to Bras and Rodríguez [1976c] for a more extensive discussion of the forms of $r(\cdot)$.

A network of rainfall gages at N points in space can be represented (see Bras and Rodríguez, 1976a, b) as,

$$Z(t) = H\, f(t) + V(t) \tag{7}$$

where $Z(t)$ = Nx1 vector of observations at time t

$f(t)$ = Nx1 state vector of rainfall at time t as given by Equation 5

$V(t)$ = Nx1 vector of measurement error

$$E[V(t_1)V^T(t_2)] = \begin{cases} R & t_1 = t_2 \\ 0 & t_1 \neq t_2 \end{cases}$$

R = error covariance matrix, assumed of diagonal form in this work. This is a simplifying assumption, not a limiting one.

$E[V(t)] = 0$

H = NxN matrix defining the raingage network.

In this work, H is an identity matrix, because observations are being made at all N points where the state vector is defined. Also, raingages are assumed to measure rainfall directly, with no losses. Values of the elements of H less or greater than 1 could be used to account for persistent under or over-estimation of rainfall by any given gage.

Equation 7 defines observations of the state vector f(t) which is composed of two components, P and $\underline{\varepsilon}(t)$. Represent the constancy of P by a "no dynamics" equation

$$P(t + 1) = P(t) \qquad (8)$$

Combining 5 and 8 results in an augmented state-space formulation

$$\begin{bmatrix} f(t) - \underline{1}\,P(t) \\ \cdots\cdots \\ P(t) \end{bmatrix} = \begin{bmatrix} A & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & 1 \end{bmatrix} \begin{bmatrix} f(t-1) - \underline{1}\,P(t-1) \\ \cdots\cdots \\ P(t-1) \end{bmatrix} + \begin{bmatrix} B \\ \cdots \\ 0 \end{bmatrix} W(t) \qquad (9)$$

or

$$f'(t) = A'\,f'(t-1) + B'\,W(t) \qquad (10)$$

where $f'(t) = (N + 1) \times 1$ new state vector

$$= \begin{bmatrix} f(t) - 1\, P(t) \\ \cdots\cdots\cdots \\ P(t) \end{bmatrix} = \begin{bmatrix} \underline{\varepsilon}(t) \\ \cdots \\ P(t) \end{bmatrix}$$

$A' = (N + 1) \times (N + 1)$ matrix

$B' = (N + 1) \times (N + 1)$ matrix

Equation 7 can be transformed to observations on the new state vector $f'(t)$ as:

$$Z = H[I \,\vdots\, \underline{1}] \begin{bmatrix} \underline{\varepsilon}(t) \\ \cdots \\ P(t) \end{bmatrix} + V(t)$$

$$= H'\, f'(t) + V(t) \tag{11}$$

where I = NxN identity matrix

$\underline{1}$ = Nx1 column vector of 1's

The linear system of equations 10 and 11 can be studied under a Kalman-Bucy filter formulation [Schweppe, 1973] to obtain linear estimates of the state vector $f'(t)$ given observations $Z(t)$. Furthermore, the linear solution yields the mean square error of estimating $f'(t)$ which will, therefore, provide the error of estimating P for a given network, in terms of number and location of observations, and a given number of time periods.

The Kalman filter formulation is:

$$\hat{f}'(t/t) = \begin{bmatrix} \hat{\underline{\varepsilon}}(t/t) \\ \cdots\cdots \\ \hat{P}(t/t) \end{bmatrix} = A'\hat{f}'(t-1/t-1) + K(t)\{Z(t) - H'A'\hat{f}'(t-1/t-1)\} \tag{12}$$

$$K(t) = \sum (t/t)\ H'^T\ R^{-1} \tag{12b}$$

$$\sum(t/t) = \sum(t/t-1) - \sum(t/t-1)\ H'^T\{R + H'\sum(t/t-1)\ H'^T\}^{-1}\ H'\sum(t/t-1) \tag{12c}$$

$$\sum(t/t-1) = A'\ \sum(t-1/t-1)\ A'^T + B'\ B'^T \tag{12d}$$

where $\hat{f}'(t/t)$ = estimate of state vector $f'(t)$ given observations up to time t

$K(t)$ = $(N + 1) \times (N + 1)$ (in this case) gain matrix

$\sum(t/t) = E[\{f'(t) - \hat{f}'(t/t)\}\{f'(t) - \hat{f}'(t/t)\}^T]$

= mean square error of estimation matrix at time t given observations up to time t.

Notice that the mean square error of estimation matrix, $\sum(t/t)$ is not a function of the observations $Z(t)$. You can, therefore, have an a priori measure of the accuracy of the data collection network. By definition (see Equation 12), the bottom diagonal value of $\sum(t/t)$ corresponds to the mean square error of estimating the long-term areal average of the rainfall process with a given number of stations N, at particular locations, and for t observations.

The solution of Equation 12 requires the definition of initial conditions $\hat{f}'(0/0)$ and $\sum(0/0)$.

In this particular case, it may be argued that there is no initial knowledge of the last element, P, in the state vector. In Bayesian jargon, only "diffuse prior" information on P is available. Fortunately, there is prior information on $\underline{\varepsilon}(t)$, it has mean 0 and covariance matrix given by Equation 4. This information is inherent in the parameters A and B of the autoregressive model.

It is useful to discuss further the implications of the above statements. Assume absolutely no information on $\underline{\varepsilon}(t)$ and P, that is the complete state vector, of dimensions N + 1, is unknown. Theoretically then the prior covariance matrix is given by

$$\sum(0/0) = \infty\ I \tag{13}$$

The above implies,

$$\sum^{-1}(1/0) = 0 \tag{14}$$

which in the limit results in (see Schweppe, 1973),

$$\sum^{-1}(1/1) = [H'^{T}\ R^{-1}\ H^{T}] \tag{15}$$

The above equation is singular for cases where the dimensions of the observations, Z, are less than the dimensions of the state vector. This singularity will prevent the inversion of Equation 15 and make impossible the solution of Equation 12.

In the hypothetical case of complete diffuse information on the state vector f'(t), the above will be the situation. The state vector has dimensions N + 1 and only N observations are available at any given time. For the case of no time and space correlations, the system will be undetermined at all time steps and will, therefore, have no unique solution. In the case of finite time and space correlations, the state dynamics equation will provide additional information. As such, the system will become determined after a given number of observations (function of the extent of correlation), when enough knowledge is available to account for the extra degree of freedom.

As previously mentioned, the problem at hand has completely diffuse information on P and knowledge of $\underline{\varepsilon}(t)$. Equations 12 could be rederived to obtain the diffuse prior, or Fisher model, estimator [Schweppe, 1973]. The solution is insured since at any one time there are more observations than completely unknown variables, except for the case when N = 1 and t = 1.

Instead of rederiving Equations 12, $\sum(0/0)$ can be assumed to be a very large diagonal matrix. $\sum(0/0)$ is chosen as large as necessary to make the solutions of Equation 12, at a desired time, t, insensitive to the initial condition. The above procedure was chosen for a) simplicity, b) preference in the form and general acceptance of Equation 12, c) insight into the system behavior inherent in the use of a large $\sum(0/0)$, and d) generality of Equation 12 for the case where some prior knowledge is available (i.e., an operating network).

In data collection network design and analysis, clearly an initial estimate of the state vector, $\hat{f}'(0/0)$, is not required. Interest lies exclusively in the accuracy measure $\sum(0/0)$. Equations 12 can nevertheless be also used in an operating framework, given an existing network and corresponding observations. In that case, the goal is to obtain estimates of the rainfall accumulation during any time period at the raingage locations as well as an updated estimate of the long term areal mean. Equations 12 are applicable in such an operating framework. A value of $\hat{f}'(0/0)$ must then be given with a corresponding $\sum(0/0)$ which is proportional to the users confidence on the selected $\hat{f}'(0/0)$.

EXAMPLES AND RESULTS

The technique developed in the previous section can be used for network design purposes, analysis of existing networks, or estimation of rainfall totals and long term areal mean precipitation based on filtered noisy observations. The formulation allows observations at any given location, this is not a random sampling technique, and explicitly considers measurement or instrument error.

Due to the generality of the formulation, the following example is necessarily limited. In order to present results comparable to Rodríguez and Mejía [1974], network analysis was performed on a series of uniformly spaced configurations of stations over square areas. The number of stations used ranged from 2 to 32.

For brevity, only results using a spatial covariance function $r(\cdot)$ (see Equation 4), of the single exponential type are presented.

Following Rodríguez and Mejia's example on a region of central Venezuela, a point variance of 0.0544 m^2 (5.44×10^4 mm^2) was used. Instrument error variances (values of the diagonal matrix R) were kept an order of magnitude less than the point variance, taking values of 0.005.

The initial mean square error matrix $\sum(0/0)$ was chosen to be of the form 50 I. This is 3 orders of magnitude larger than the point variance. The above form of $\sum(0/0)$ was selected after studying the effect of various initial conditions. Table 1 shows some of the results of such experiments. The table gives the mean square error of estimating the long term areal mean value (given by the lower diagonal of

Table 1

TESTS ON EFFECTS OF INITIAL MEAN SQUARE ERROR MATRIX FORM

$A = 4100\ km^2$ $\alpha = 0.0156$ $A\alpha^2 \approx 1$ $\rho = 0.25$

	$\Sigma(0/0) = 100\ I$	$\Sigma(0/0) = 50\ I$
t = 1		
N	M.S.E.	M.S.E.
2	3.071	1.556
4	1.573	0.804
6	1.065	0.549
8	0.808	0.421
t = 2		
N	M.S.E.	M.S.E.
2	0.076	0.074
4	0.062	0.059
6	0.057	0.054
8	0.054	0.051
t = 3		
N	M.S.E.	M.S.E.
2	0.038	0.037
4	0.031	0.031
6	0.029	0.028
8	0.028	0.027
t = 4		
N	M.S.E.	M.S.E.
2	0.025	0.025
4	0.021	0.021
6	0.019	0.019
8	0.019	0.018
t = 5		
N	M.S.E.	M.S.E.
2	0.019	0.019
4	0.016	0.016
6	0.015	0.014
8	0.014	0.014

Table 1 (Continued)

TESTS ON EFFECTS OF INITIAL MEAN SQUARE ERROR MATRIX FORM

$A = 40000\ km^2$ | $\alpha = 0.0156$ | $A\alpha^2 \approx 9.7$ | $\rho = 0.25$

	$\Sigma(0/0) = 100\ I$	$\Sigma(0/0) = 50\ I$	$\Sigma(0/0) = 25\ I$	$\Sigma(0/0) = 5\ I$
t = 1				
N	M.S.E.	M.S.E.	M.S.E.	M.S.E.
2	3.062	1.546	0.789	0.183
4	1.559	0.785	0.405	0.097
6	1.049	0.533	0.276	0.069
8	0.791	0.404	0.211	0.055
t = 2				
N	M.S.E.	M.S.E.	M.S.E.	M.S.E.
2	0.058	0.057	0.053	0.047
4	0.037	0.036	0.034	0.027
6	0.031	0.030	0.029	0.022
8	0.028	0.027	0.026	0.019
t = 3				
N	M.S.E.	M.S.E.	M.S.E.	M.S.E.
2	0.029	0.029	0.028	0.025
4	0.018	0.018	0.018	0.015
6	0.016	0.015	0.015	0.013
8	0.014	0.014	0.014	0.012
t = 4				
N	M.S.E.	M.S.E.	M.S.E.	M.S.E.
2	0.019	0.019	0.019	0.018
4	0.012	0.012	0.012	0.011
6	0.010	0.010	0.010	0.009
8	0.010	0.009	0.009	0.008
t = 5				
N	M.S.E.	M.S.E.	M.S.E.	M.S.E.
2	0.014	0.014	0.014	0.013
4	0.009	0.009	0.009	0.008
6	0.008	0.008	0.008	0.007
8	0.007	0.007	0.007	0.006

$\sum(t/t)$ in Equation 12) as a function of N, the number of stations (uniformly distributed), and the number of time samples, t, for different values of $\sum(0/0)$.

The spatial covariance function used was of the single exponential type with parameter $\alpha = 0.0156$ which corresponds to that used in Central Venezuela by Rodríguez and Mejía. Two square areas, 4100 Km^2 and 40,000 Km^2 were used yielding non-dimensional area parameters, $A\alpha^2$, of 1 and 9.7, respectively.

As the table shows, at time step 3, results are already fairly insensitive to initial conditions. Only when $\sum(0/0) = 5$ I does the effect of initial conditions persist up to $t = 5$ after which it disappears. The selection of $\sum(0/0) = 50$ I insures satisfactory representation of our diffuse prior knowledge and at the same time yields results fairly insensitive to initial conditions at $t = 3$, which is acceptable to our needs. Diffuse knowledge of $\underline{\varepsilon}(t)$ was not required since, in fact, prior knowledge of this part of the state vector is available. The mechanics of the filter "learns" this fact very fast, irrelevant of the size of the prior variance used.

The use of a finite $\sum(0/0)$ can also be justified by arguing that the hydrologist can certainly establish a range for the expected values of precipitation and its mean.

Rodríguez and Mejía [1974] and Bras and Rodríguez [1976a] prove that for random sampling, stratified random sampling and uniformly distributed systematic sampling, the mean square error of estimation is only a function of the non-dimensional area $A\alpha^2$ and not of the area, A, and the spatial correlation parameter, α, individually.

The examples of this work are uniformly distributed networks so the results are given in terms of values of $A\alpha^2$ equal to 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0 and 10000.0. The reader is cautioned that for the general case of any network distribution, the solution must be obtained for individual values of A and α .

The lag 1 correlation factor, ρ, took values of 0.0, 0.25 and 0.50. Results are presented in Tables 2 through 10 based on a point variance of 0.0544 m^2. A complete set of tables is available on microfilm.

Figures 1 to 9 show other results normalized by dividing the Mean Square Error by the point variance of the example, 0.0544 m^2. This makes them comparable to Rodríguez and Mejía's [1974] results. The reader can corroborate that both methodologies give similar results but keep in mind that the problems are slightly different. In their work, Rodríguez and Mejía dealt with random and stratified random sampling with no measurement error. This work investigates non-random sampling networks with allowance for measurement error.

The plots of figures 10 through 14 show the mean square error of estimating long term areal mean precipitation as it varies with time (duration of sampling) and number of uniformly spaced stations for different areas. Figures 10 to 14 attempt to explicitly show the trade-off between duration of sampling (time) and number of sampling stations (space) with curves of constant mean square error.

The tables and figures are self-explanatory. It is clear that mean square estimation error goes down as the number of stations and sampling duration increase. (Other figures available on microfilm.)

The estimation error also decreases as the area increases for a fixed sampling duration with a given number of stations. This behavior is not necessarily obvious. Particularly, it is opposite to the

Table 2

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 0.1$, $\rho = 0.0$

Number of Stations

Time	2	4	8	16	28	32
1	.0529	.0489	.0472	.0455	.0448	.0446
2	.0264	.0245	.0236	.0228	.0224	.0223
3	.0176	.0163	.0157	.0152	.0149	.0149
4	.0132	.0122	.0118	.0114	.0112	.0112
5	.0106	.0098	.0094	.0091	.0090	.0090
6	.0088	.0082	.0079	.0076	.0075	.0074
7	.0076	.0070	.0067	.0065	.0064	.0064
8	.0066	.0061	.0059	.0057	.0056	.0056
9	.0059	.0054	.0052	.0051	.0050	.0050
10	.0053	.0049	.0047	.0046	.0045	.0045
11	.0048	.0044	.0043	.0041	.0041	.0041
12	.0044	.0041	.0039	.0038	.0037	.0037
13	.0041	.0038	.0036	.0035	.0034	.0034
14	.0038	.0035	.0034	.0033	.0032	.0032
15	.0035	.0033	.0031	.0030	.0030	.0030
16	.0033	.0031	.0030	.0028	.0028	.0028
17	.0031	.0029	.0028	.0027	.0026	.0026
18	.0029	.0027	.0026	.0025	.0025	.0025
19	.0028	.0026	.0025	.0024	.0024	.0024
20	.0026	.0024	.0024	.0023	.0022	.0022

Table 3

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 100$, $\rho = 0.0$

	Number of Stations					
Time	2	4	8	16	28	32
1	.0299	.0150	.0084	.0048	.0037	.0036
2	.0150	.0075	.0042	.0024	.0019	.0018
3	.0100	.0050	.0028	.0016	.0012	.0012
4	.0075	.0038	.0021	.0012	.0009	.0009
5	.0060	.0030	.0017	.0010	.0007	.0007
6	.0050	.0025	.0014	.0008	.0006	.0006
7	.0043	.0021	.0012	.0007	.0005	.0005
8	.0037	.0019	.0010	.0006	.0005	.0004
9	.0033	.0017	.0009	.0005	.0004	.0004
10	.0030	.0015	.0008	.0005	.0004	.0004
11	.0027	.0014	.0008	.0004	.0003	.0003
12	.0025	.0013	.0007	.0004	.0003	.0003
13	.0023	.0012	.0006	.0004	.0003	.0003
14	.0021	.0011	.0006	.0003	.0003	.0003
15	.0020	.0010	.0006	.0003	.0002	.0002
16	.0019	.0009	.0005	.0003	.0002	.0002
17	.0018	.0009	.0005	.0003	.0002	.0002
18	.0017	.0008	.0005	.0003	.0002	.0002
19	.0016	.0008	.0004	.0003	.0002	.0002
20	.0015	.0008	.0004	.0002	.0002	.0002

Table 4

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 10000$, $\rho = 0.0$

	Number of Stations					
Time	2	4	8	16	28	32
1	.0297	.0148	.0074	.0037	.0021	.0019
2	.0148	.0074	.0037	.0019	.0011	.0009
3	.0099	.0050	.0025	.0012	.0007	.0006
4	.0074	.0037	.0019	.0009	.0005	.0005
5	.0059	.0030	.0015	.0007	.0004	.0004
6	.0050	.0025	.0012	.0006	.0004	.0003
7	.0042	.0021	.0011	.0005	.0003	.0003
8	.0037	.0019	.0010	.0005	.0003	.0002
9	.0033	.0017	.0008	.0004	.0002	.0002
10	.0030	.0015	.0007	.0004	.0002	.0002
11	.0027	.0014	.0007	.0003	.0002	.0002
12	.0025	.0012	.0006	.0003	.0002	.0002
13	.0023	.0011	.0006	.0003	.0002	.0001
14	.0021	.0011	.0005	.0003	.0002	.0001
15	.0020	.0010	.0005	.0002	.0001	.0001
16	.0019	.0009	.0005	.0002	.0001	.0001
17	.0017	.0009	.0004	.0002	.0001	.0001
18	.0017	.0008	.0004	.0002	.0001	.0001
19	.0016	.0008	.0004	.0002	.0001	.0001
20	.0015	.0007	.0004	.0002	.0001	.0001

Table 5

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 0.10$, $\rho = 0.25$

	Number of Stations					
Time	2	4	8	16	28	32
1	1.5619	.8139	.4316	.2381	.1548	.1409
2	.0839	.0743	.0667	.0577	.0504	.0487
3	.0426	.0387	.0360	.0327	.0300	.0294
4	.0286	.0262	.0246	.0228	.0214	.0210
5	.0215	.0198	.0187	.0175	.0166	.0164
6	.0172	.0159	.0151	.0142	.0135	.0134
7	.0143	.0133	.0127	.0120	.0115	.0113
8	.0123	.0114	.0109	.0103	.0099	.0098
9	.0108	.0100	.0096	.0091	.0088	.0087
10	.0096	.0089	.0085	.0081	.0078	.0078
11	.0086	.0080	.0077	.0073	.0071	.0070
12	.0078	.0073	.0070	.0067	.0065	.0064
13	.0072	.0067	.0064	.0061	.0059	.0059
14	.0066	.0062	.0059	.0057	.0055	.0055
15	.0062	.0057	.0055	.0053	.0051	.0051
16	.0058	.0054	.0051	.0049	.0048	.0048
17	.0054	.0050	.0048	.0046	.0045	.0045
18	.0051	.0047	.0045	.0044	.0042	.0042
19	.0048	.0045	.0043	.0041	.0040	.0040
20	.0045	.0042	.0041	.0039	.0038	.0038

Table 6

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 100$, $\rho = 0.25$

	Number of Stations					
Time	2	4	8	16	28	32
1	1.5416	.7830	.3954	.1991	.1149	.1009
2	.0487	.0245	.0136	.0078	.0059	.0057
3	.0242	.0122	.0068	.0039	.0030	.0029
4	.0161	.0081	.0045	.0026	.0020	.0019
5	.0121	.0061	.0034	.0020	.0015	.0015
6	.0097	.0049	.0027	.0016	.0012	.0012
7	.0080	.0040	.0023	.0013	.0010	.0010
8	.0069	.0035	.0019	.0011	.0009	.0008
9	.0060	.0030	.0017	.0010	.0008	.0007
10	.0054	.0027	.0015	.0009	.0007	.0007
11	.0048	.0024	.0014	.0008	.0006	.0006
12	.0044	.0022	.0012	.0007	.0006	.0005
13	.0040	.0020	.0011	.0007	.0005	.0005
14	.0037	.0019	.0010	.0006	.0005	.0005
15	.0034	.0017	.0010	.0006	.0004	.0004
16	.0032	.0016	.0009	.0005	.0004	.0004
17	.0030	.0015	.0008	.0005	.0004	.0004
18	.0028	.0014	.0008	.0005	.0004	.0003
19	.0027	.0013	.0008	.0004	.0003	.0003
20	.0025	.0013	.0007	.0004	.0003	.0003

Table 7

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 10000$, $\rho = 0.25$

Time	Number of Stations 2	4	8	16	28	32
1	1.5415	.7828	.3945	.1980	.1134	.0992
2	.0484	.0242	.0121	.0061	.0035	.0030
3	.0241	.0120	.0060	.0030	.0017	.0015
4	.0160	.0080	.0040	.0020	.0011	.0010
5	.0120	.0060	.0030	.0015	.0009	.0008
6	.0096	.0048	.0024	.0012	.0007	.0006
7	.0080	.0040	.0020	.0010	.0006	.0005
8	.0068	.0034	.0017	.0009	.0005	.0004
9	.0060	.0030	.0015	.0007	.0004	.0004
10	.0053	.0027	.0013	.0007	.0004	.0003
11	.0048	.0024	.0012	.0006	.0003	.0003
12	.0044	.0022	.0011	.0005	.0003	.0003
13	.0040	.0020	.0010	.0005	.0003	.0002
14	.0037	.0018	.0009	.0005	.0003	.0002
15	.0034	.0017	.0009	.0004	.0002	.0002
16	.0032	.0016	.0008	.0004	.0002	.0002
17	.0030	.0015	.0007	.0004	.0002	.0002
18	.0028	.0014	.0007	.0004	.0002	.0002
19	.0027	.0013	.0007	.0003	.0002	.0002
20	.0025	.0013	.0006	.0003	.0002	.0002

Table 8

Time	Number of Stations 2	4	8	16	28	32
1	5.5874	2.9739	1.5489	.8034	.4769	.4220
2	.1590	.1421	.1311	.1178	.1055	.1023
3	.0782	.0716	.0677	.0630	.0589	.0578
4	.0519	.0479	.0456	.0429	.0408	.0402
5	.0388	.0359	.0344	.0326	.0312	.0308
6	.0310	.0288	.0276	.0262	.0253	.0250
7	.0258	.0240	.0230	.0220	.0212	.0210
8	.0221	.0206	.0198	.0189	.0183	.0182
9	.0193	.0180	.0173	.0166	.0161	.0160
10	.0172	.0160	.0154	.0147	.0143	.0142
11	.0154	.0144	.0139	.0133	.0129	.0129
12	.0140	.0131	.0126	.0121	.0118	.0117
13	.0129	.0120	.0116	.0111	.0108	.0108
14	.0119	.0111	.0107	.0103	.0100	.0100
15	.0110	.0103	.0099	.0095	.0093	.0093
16	.0103	.0096	.0093	.0089	.0087	.0087
17	.0096	.0090	.0087	.0084	.0082	.0081
18	.0091	.0085	.0082	.0079	.0077	.0077
19	.0086	.0080	.0077	.0074	.0073	.0072
20	.0081	.0076	.0073	.0070	.0069	.0069

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 0.10$, $\rho = 0.5$

Table 9

Time	Number of Stations 2	4	8	16	28	32
1	5.5738	2.9514	1.5212	.7728	.4453	.3903
2	.0930	.0468	.0260	.0148	.0113	.0108
3	.0445	.0224	.0125	.0072	.0055	.0053
4	.0291	.0147	.0082	.0047	.0037	.0035
5	.0217	.0109	.0061	.0035	.0027	.0026
6	.0173	.0087	.0049	.0028	.0022	.0021
7	.0143	.0072	.0040	.0023	.0018	.0018
8	.0122	.0062	.0035	.0020	.0016	.0015
9	.0107	.0054	.0030	.0018	.0014	.0013
10	.0095	.0048	.0027	.0016	.0012	.0012
11	.0085	.0043	.0024	.0014	.0011	.0010
12	.0078	.0039	.0022	.0013	.0010	.0010
13	.0071	.0036	.0020	.0012	.0009	.0009
14	.0066	.0033	.0019	.0011	.0008	.0008
15	.0061	.0031	.0017	.0010	.0008	.0007
16	.0057	.0029	.0016	.0009	.0007	.0007
17	.0053	.0027	.0015	.0009	.0007	.0007
18	.0050	.0025	.0014	.0008	.0006	.0006
19	.0047	.0024	.0013	.0008	.0006	.0006
20	.0045	.0023	.0013	.0007	.0006	.0006

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE
OF PRECIPITATION FOR $A\alpha^2 = 100$, $\rho = 0.50$

Table 10

	Number of Stations					
Time	2	4	8	16	28	32
1	5.5736	2.9513	1.5205	.7720	.4441	.3890
2	.0925	.0463	.0231	.0116	.0066	.0058
3	.0442	.0221	.0111	.0055	.0032	.0028
4	.0290	.0145	.0072	.0036	.0021	.0018
5	.0215	.0108	.0054	.0027	.0015	.0013
6	.0172	.0086	.0043	.0021	.0012	.0011
7	.0142	.0071	.0036	.0018	.0010	.0009
8	.0122	.0061	.0030	.0015	.0009	.0008
9	.0106	.0053	.0027	.0013	.0008	.0007
10	.0094	.0047	.0024	.0012	.0007	.0006
11	.0095	.0042	.0021	.0011	.0006	.0005
12	.0077	.0039	.0019	.0010	.0006	.0005
13	.0071	.0035	.0018	.0009	.0005	.0004
14	.0065	.0033	.0016	.0008	.0005	.0004
15	.0060	.0030	.0015	.0008	.0004	.0004
16	.0056	.0028	.0014	.0001	.0004	.0004
17	.0053	.0026	.0013	.0007	.0004	.0003
18	.0050	.0025	.0012	.0006	.0004	.0003
19	.0047	.0023	.0012	.0006	.0003	.0003
20	.0044	.0022	.0011	.0006	.0003	.0003

MEAN SQUARE ERROR OF ESTIMATING THE LONG TERM AREAL AVERAGE OF PRECIPITATION FOR $A\alpha^2 = 10000$, $\rho = 0.50$

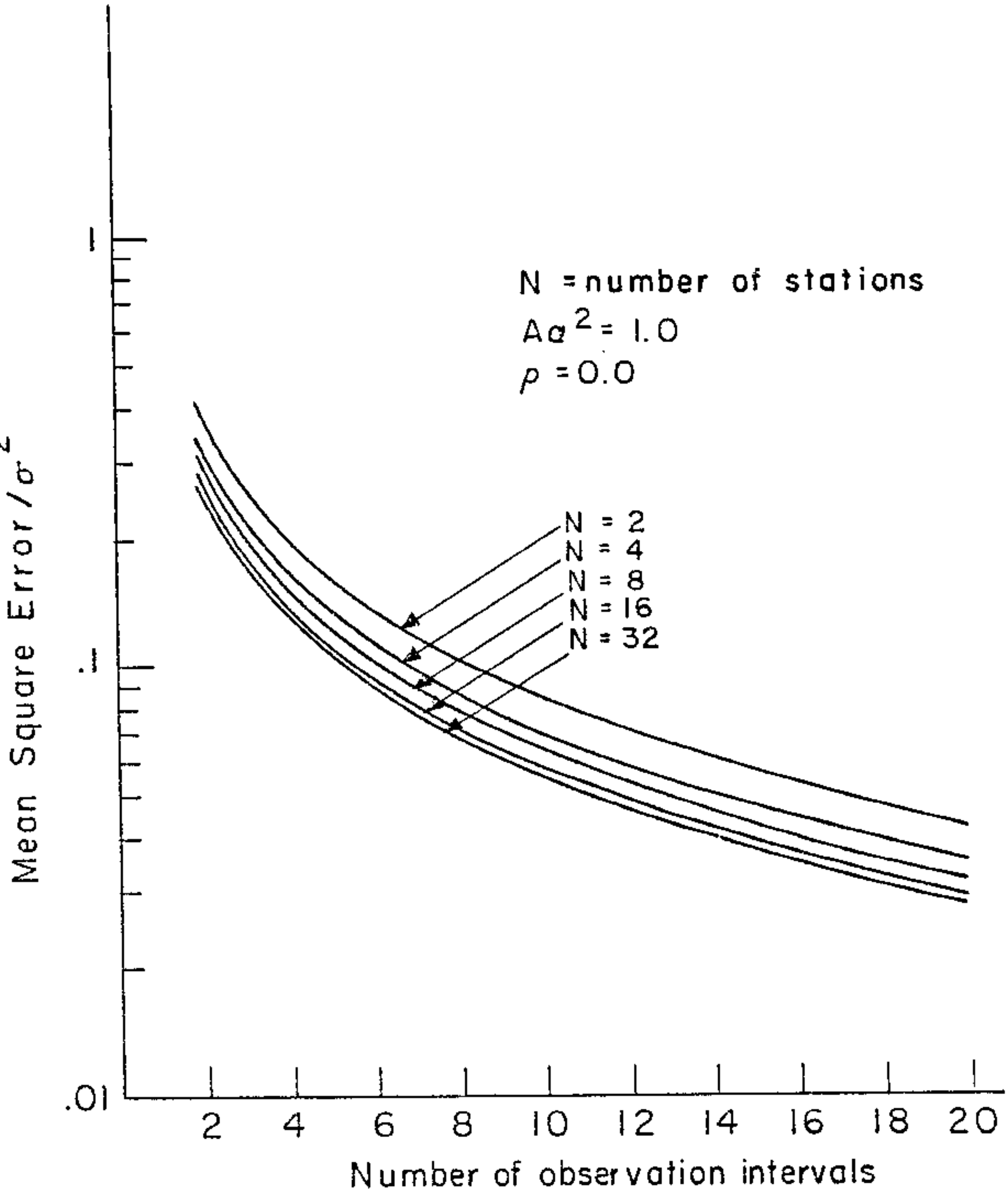

Figure 1: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

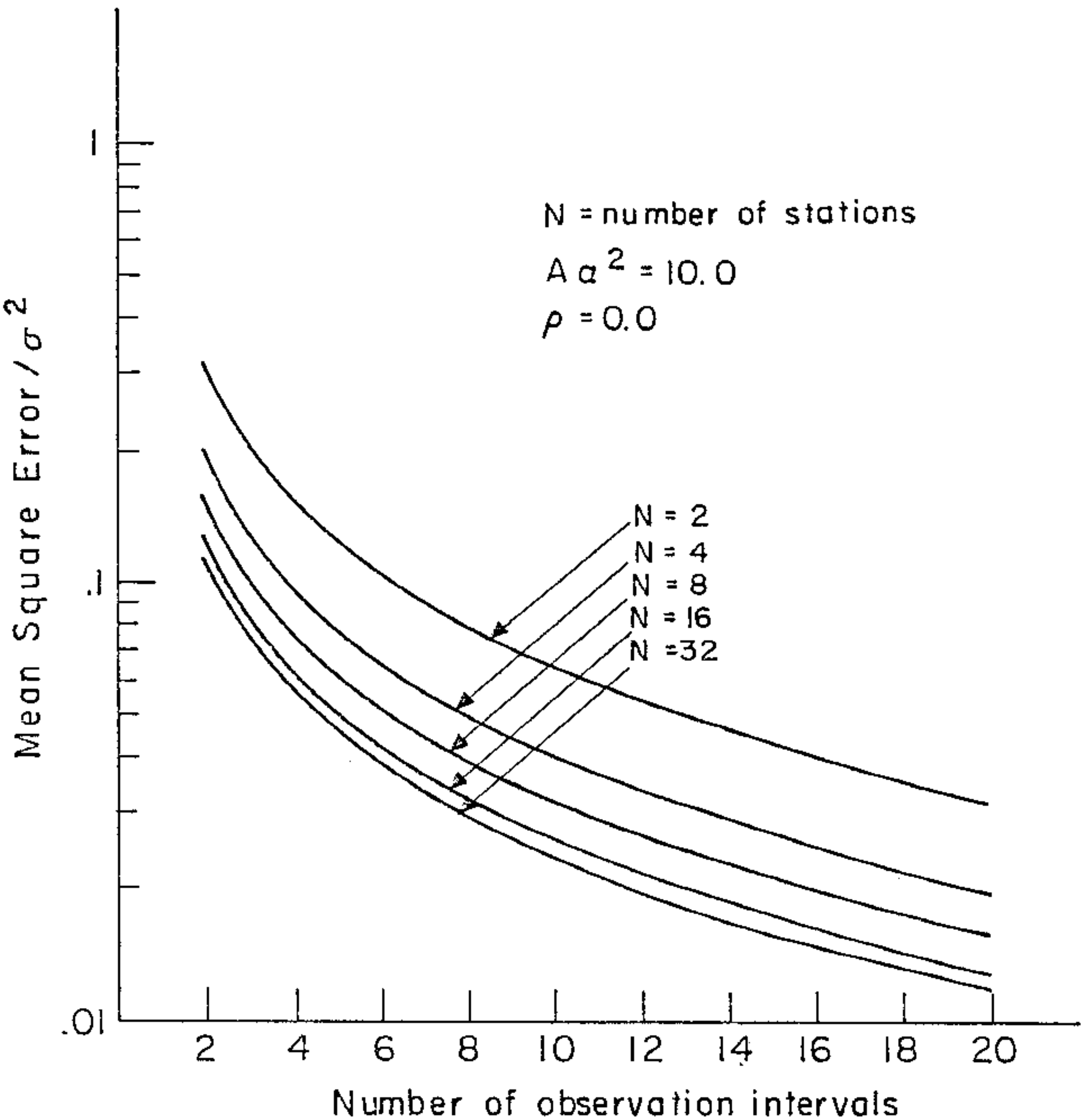

Figure 2: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

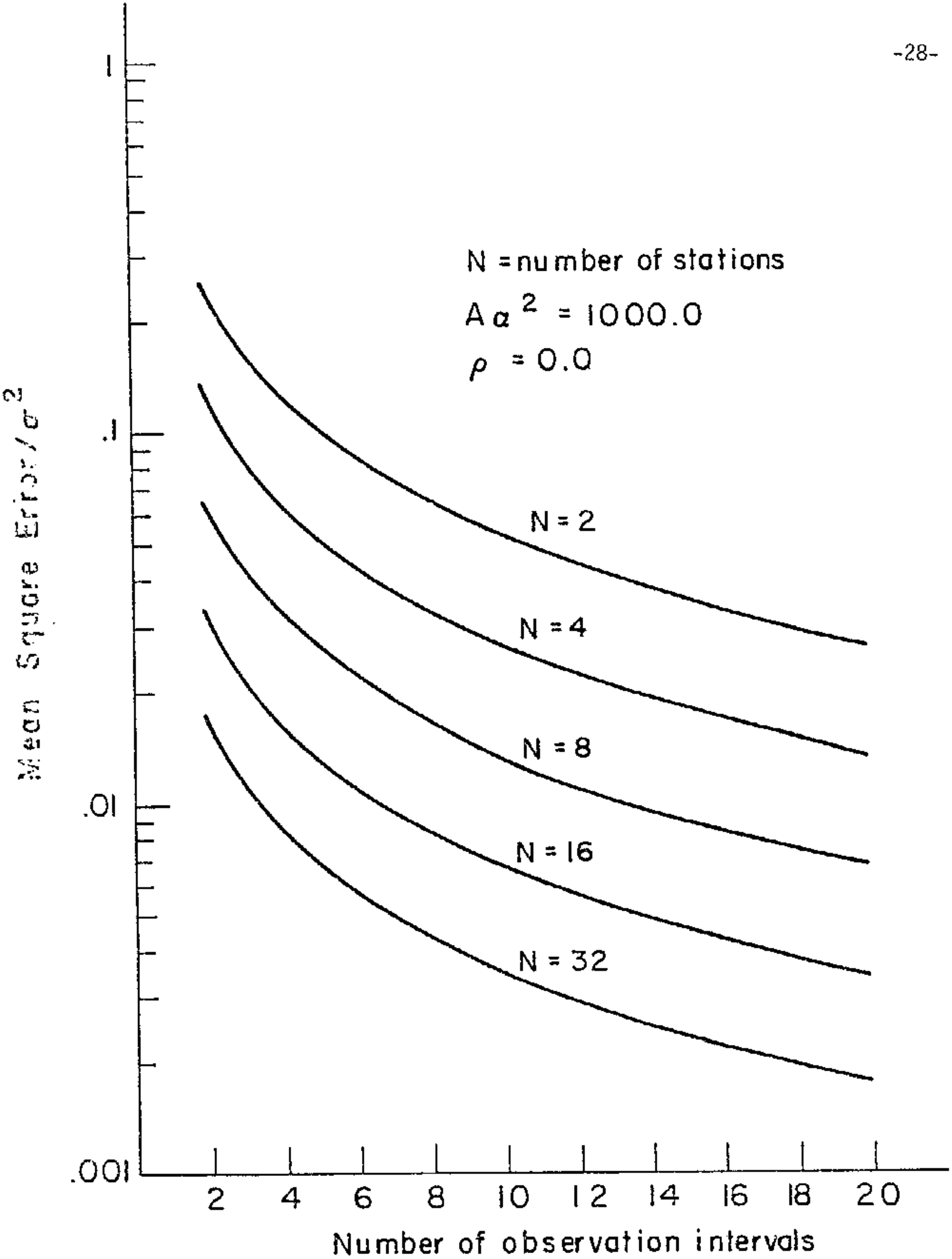

Figure 3: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

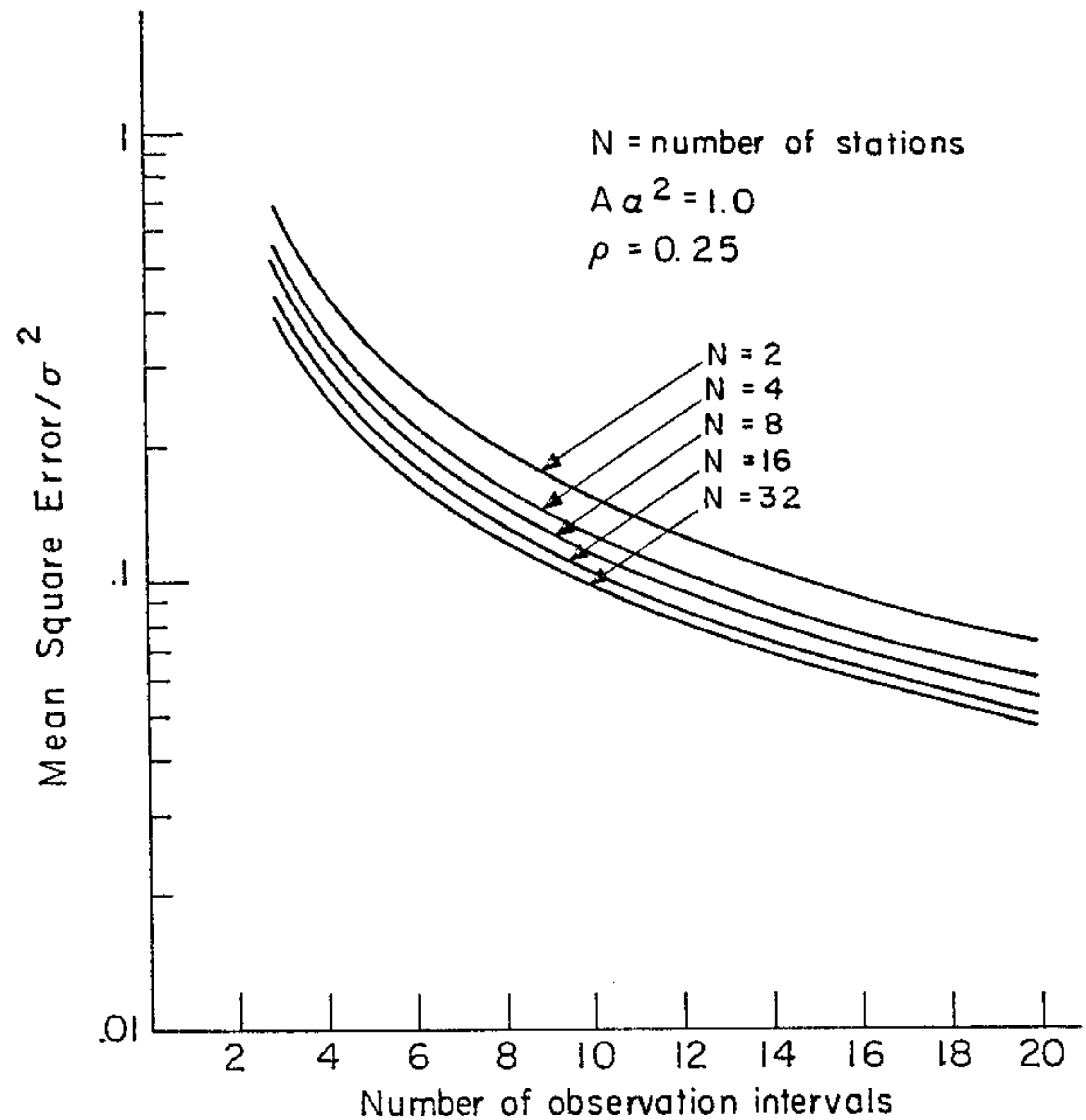

Figure 4: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

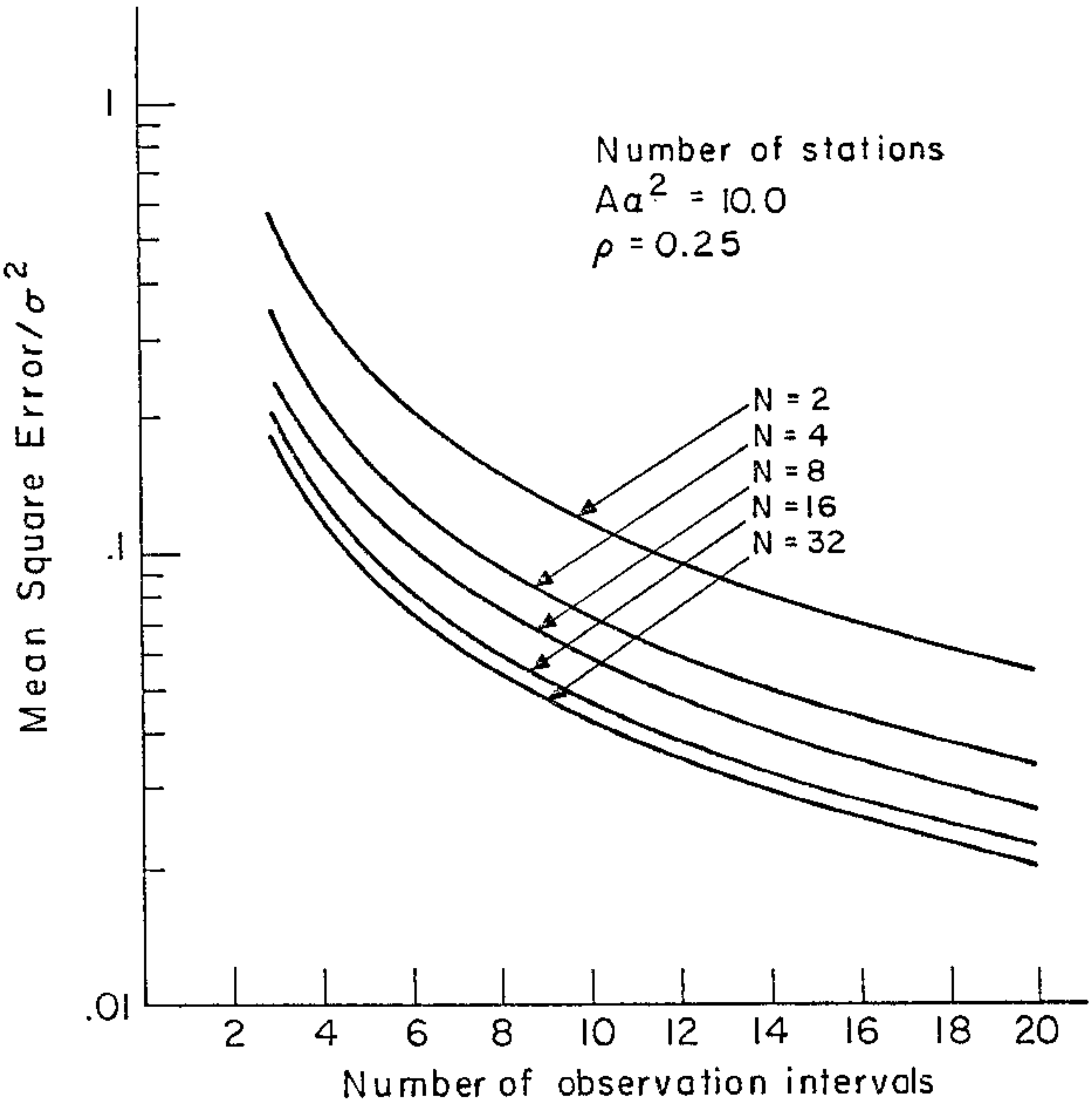

Figure 5: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

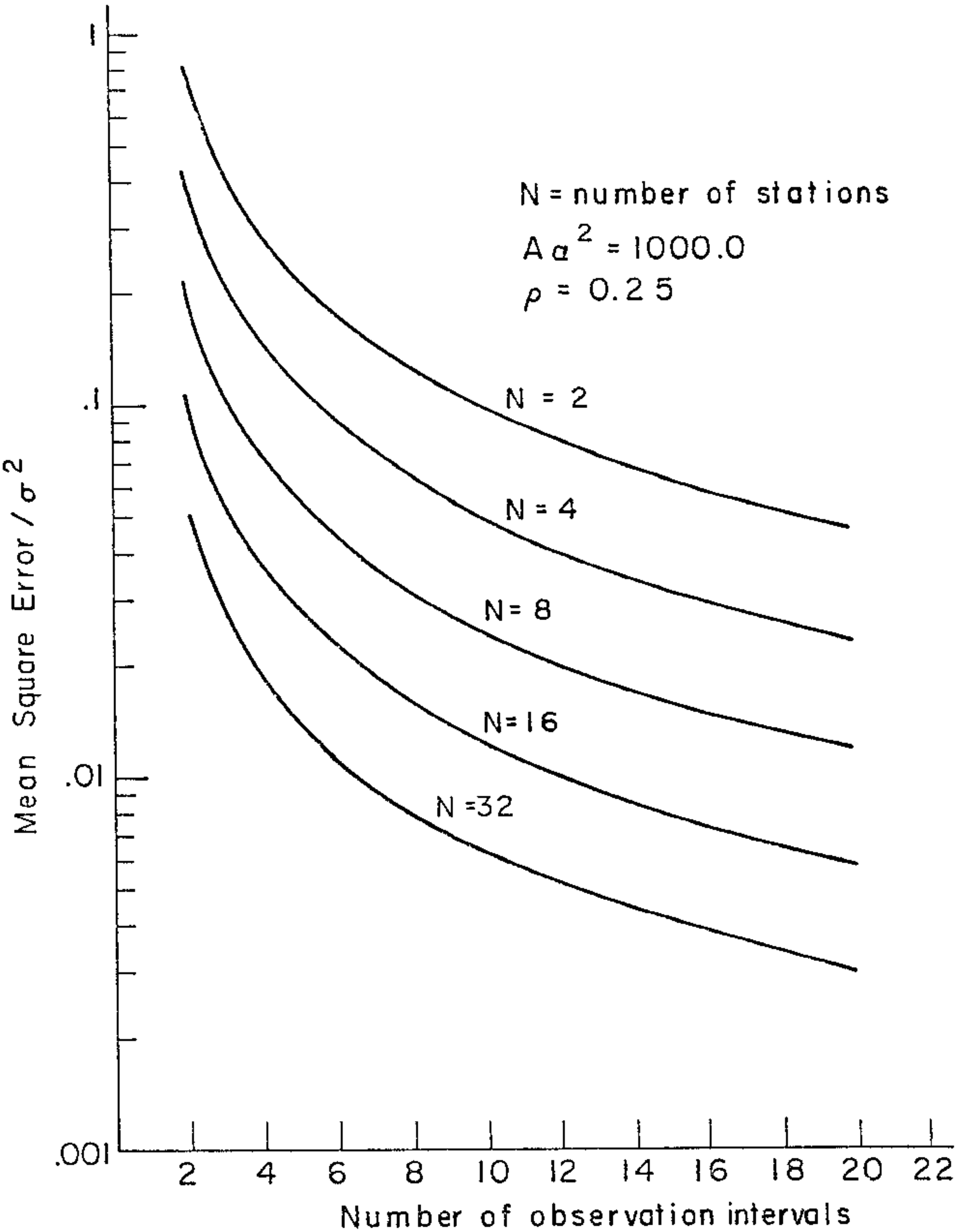

Figure 6: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

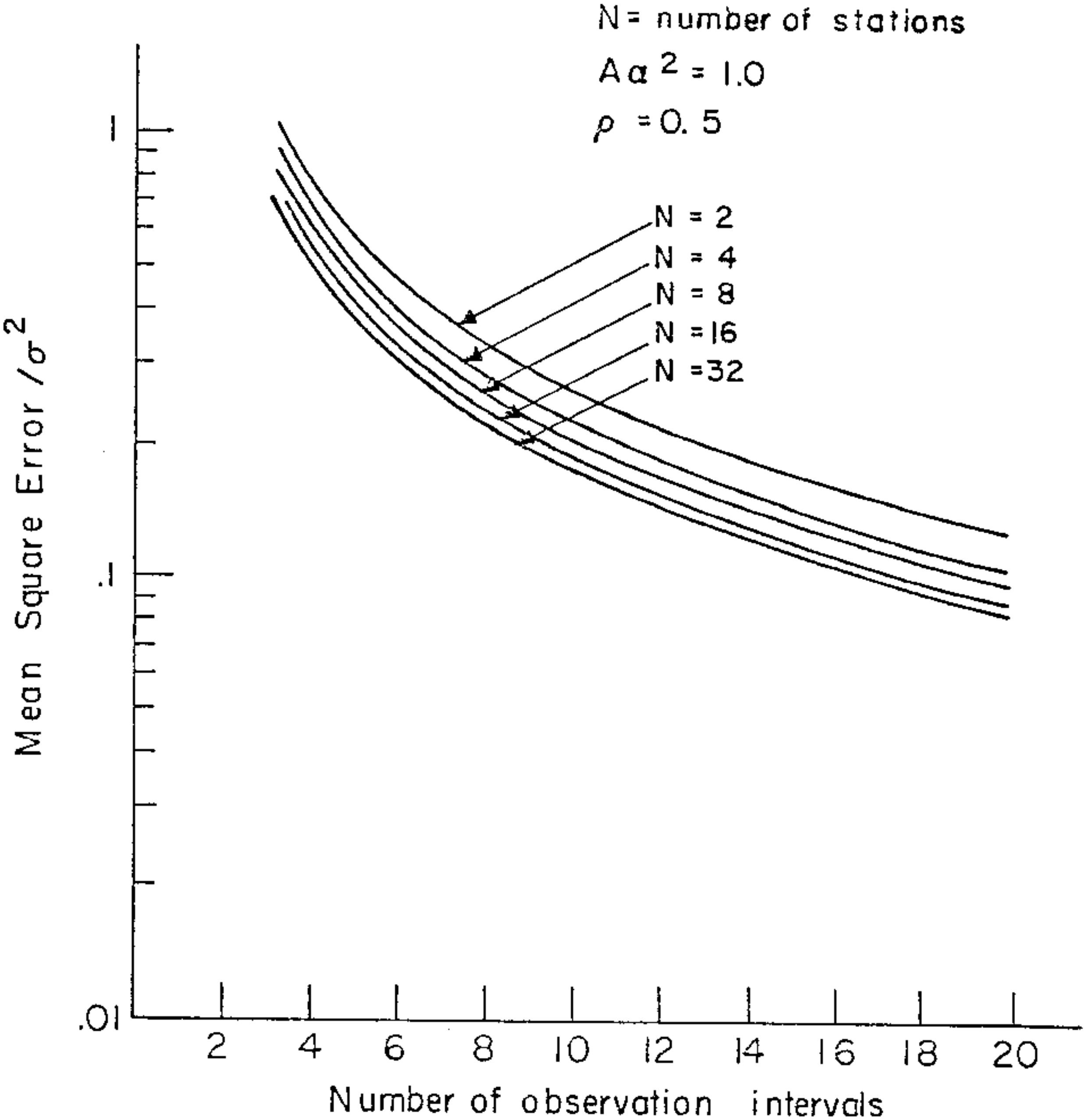

Figure 7: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

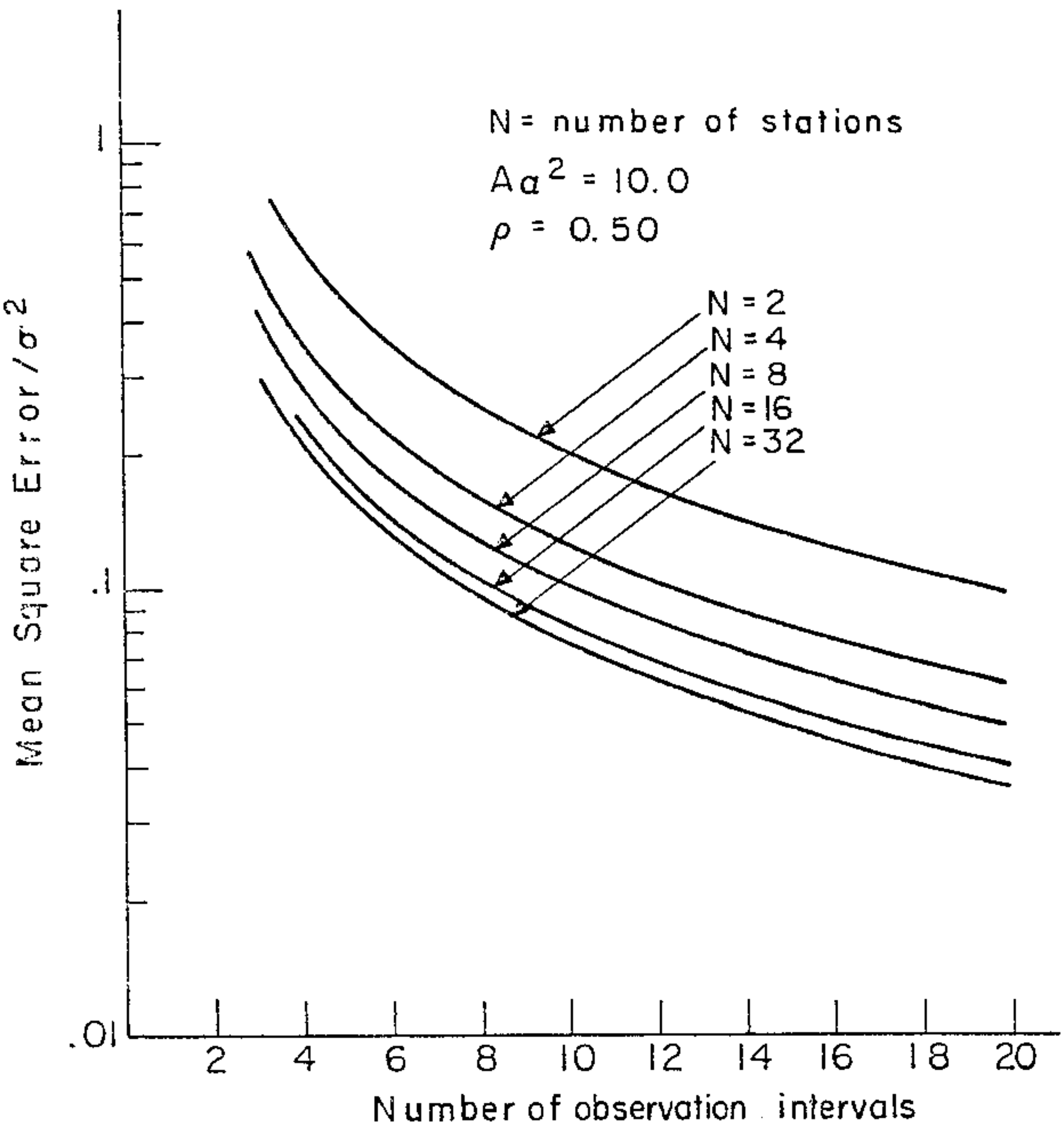

Figure 8: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

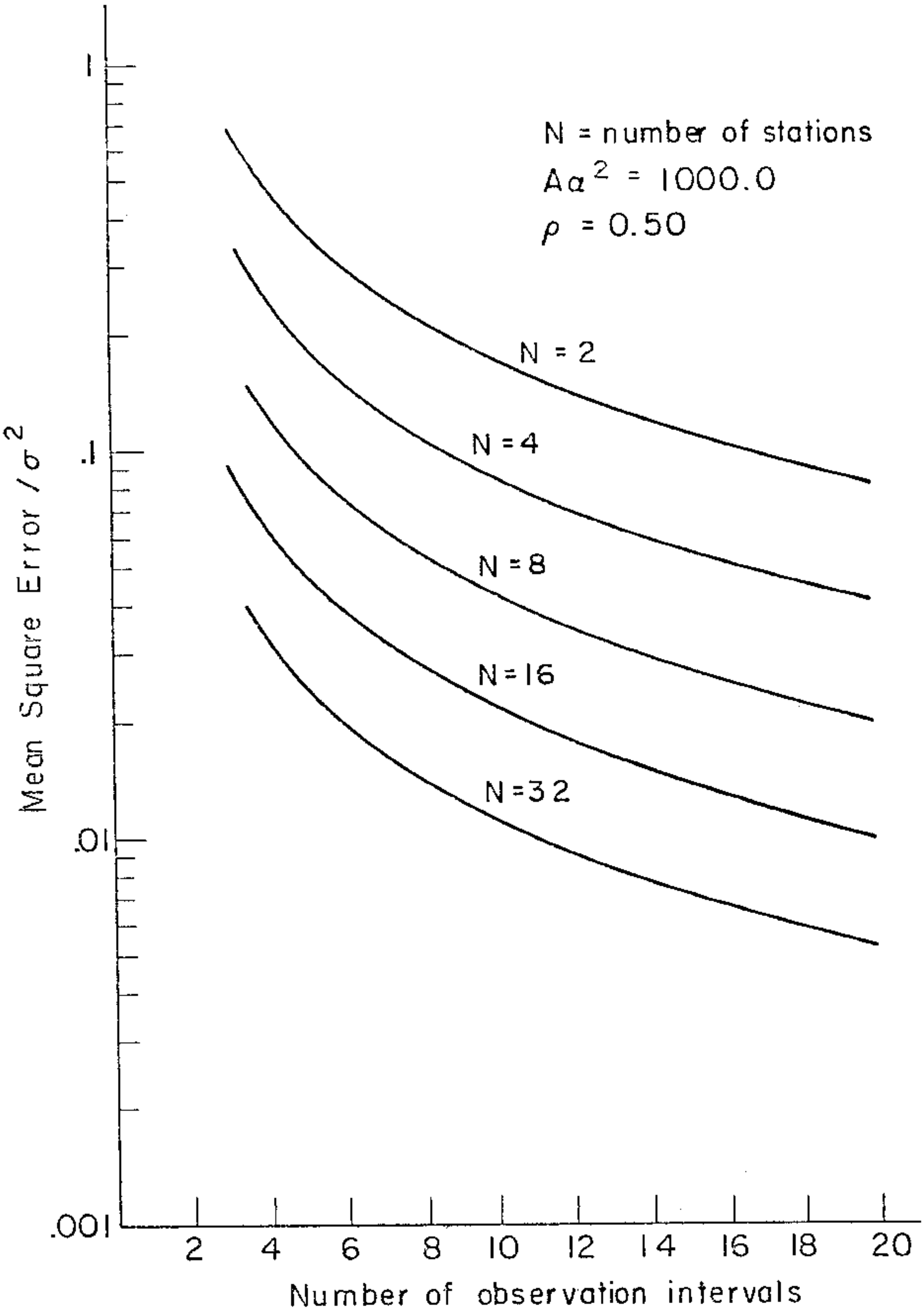

Figure 9: Normalized Mean Square Error as a function of observation intervals and number of stations for a uniformly distributed Network

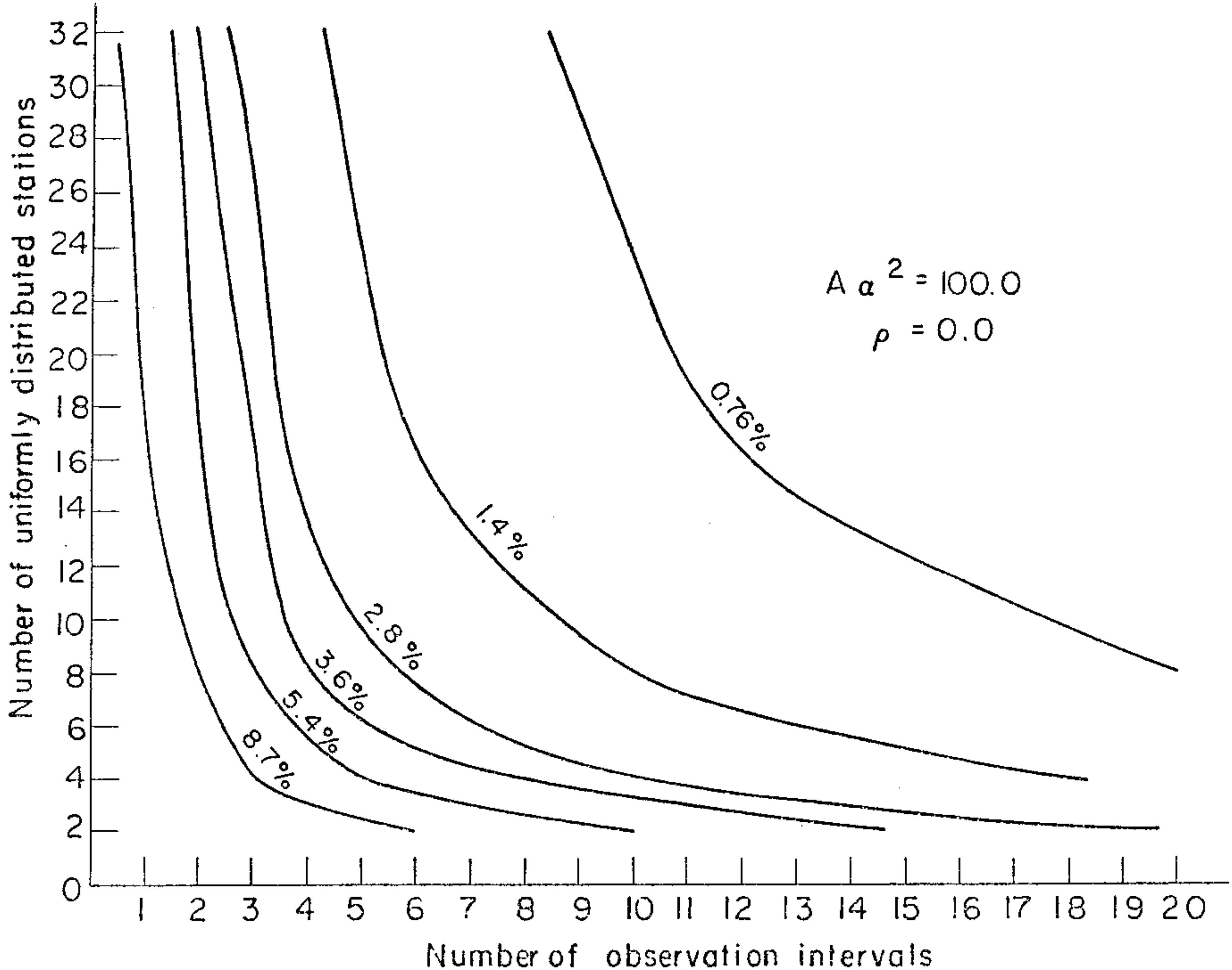

Figure 10: Trade-off curves between number of stations and intervals of observation for different mean square error as a percentage of point variance

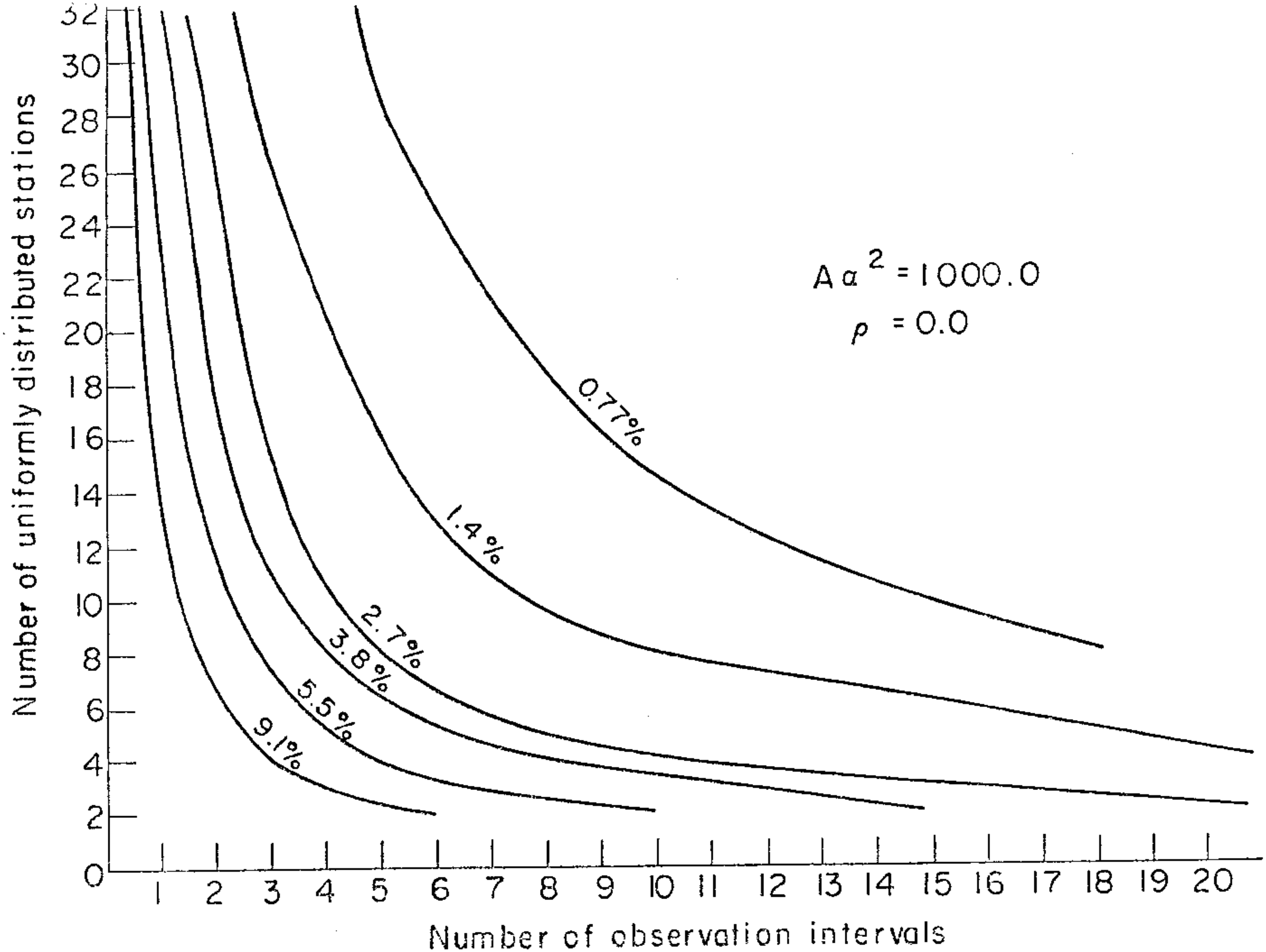

Figure 11: Trade-off curves between number of stations and intervals of observation for different mean square error as a percentage of point variance

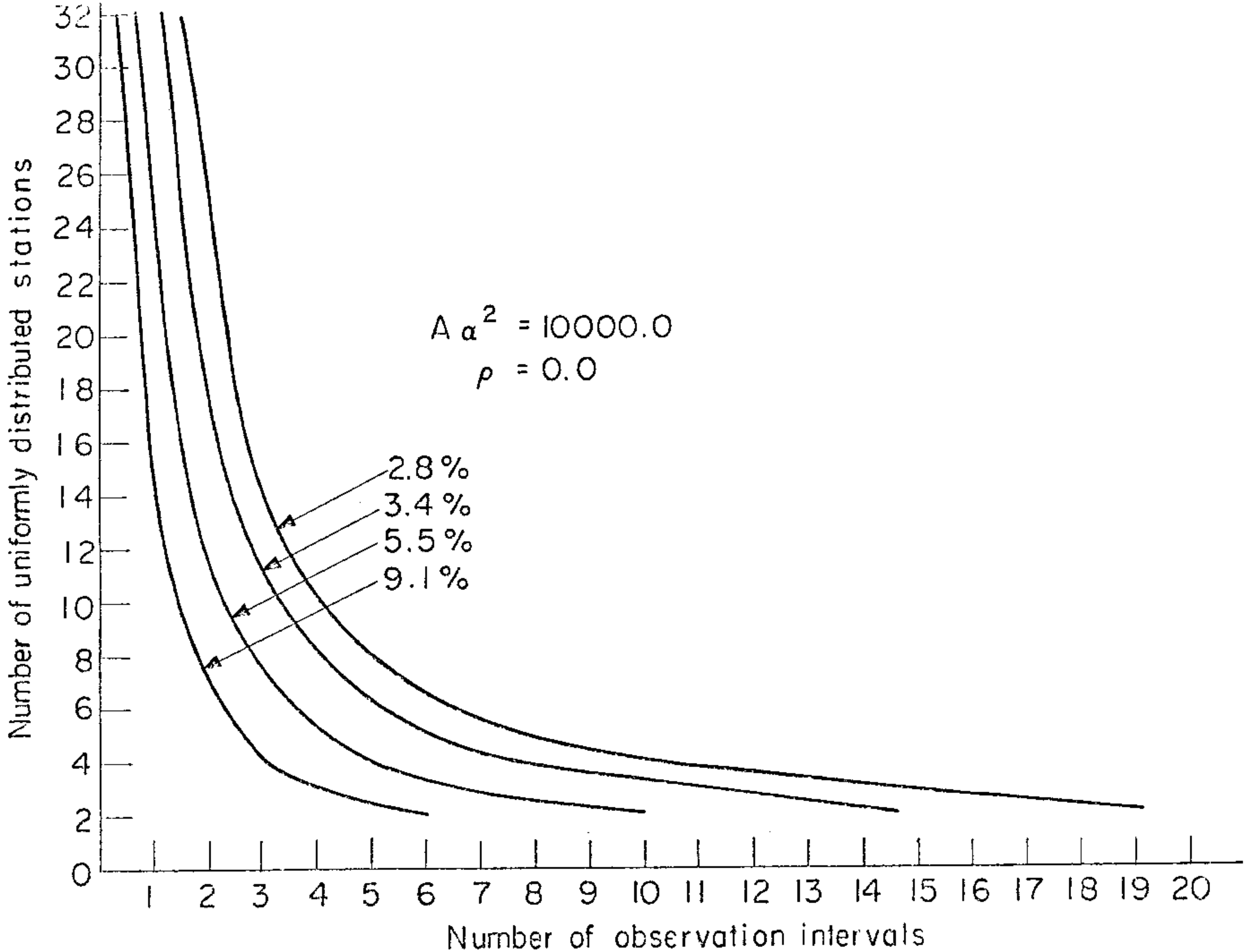

Figure 12: Trade-off curves between number of stations and intervals of observation for different mean square error as a percentage of point variance

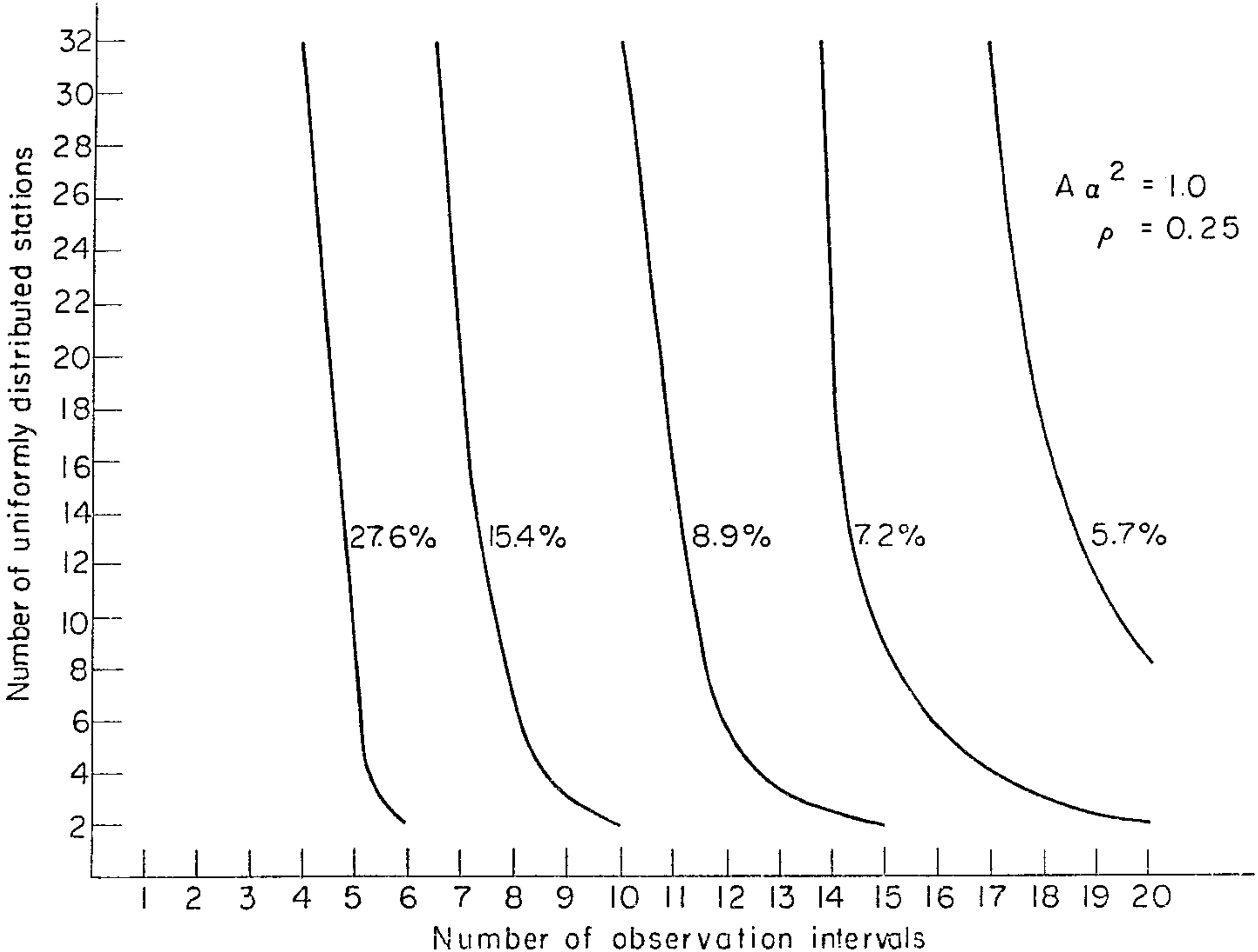

Figure 13: Trade-off curves between number of stations and intervals of observation for different mean square error as a percentage of point variance

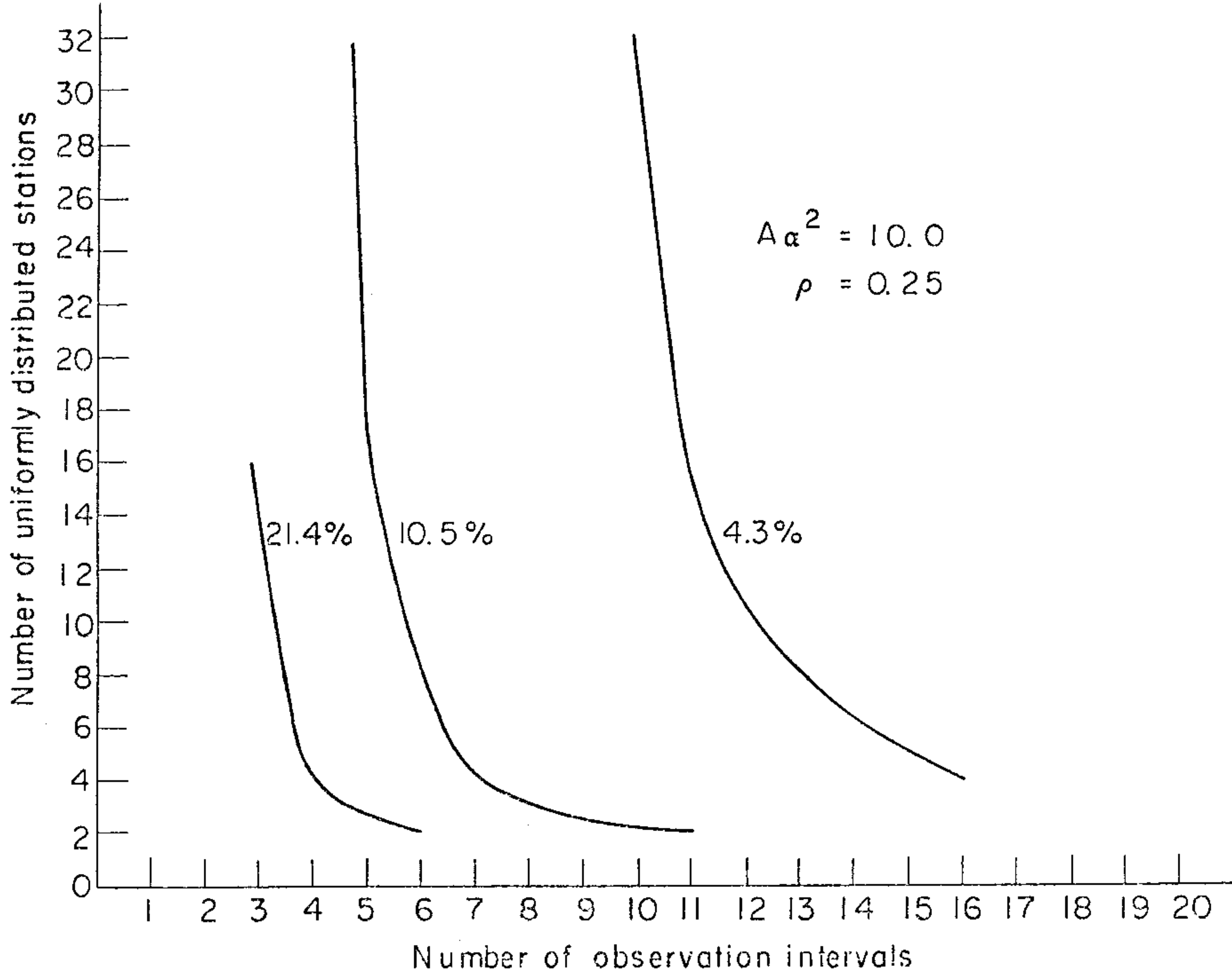

Figure 14: Trade-off curves between number of stations and intervals of observation for different mean square error as a percentage of point variance

behavior observed when estimating the areal average of a rainfall event (see Rodríguez and Mejia, 1974, and Bras and Rodríguez, 1976a). Intuitively, the explanation lies in the assumption of spatial homogeneity of the process. In the average of an event problem, each observed point in space is an estimate of what happened at that particular point. In the limit, where the area is the point, the areal average of the event is exactly known.

Assume now a finite area with N number of stations. This will result in a given estimation error. Increasing the number of stations for the same area defines the event better so the error of estimating the mean must decrease. Fixing the number of stations but uniformly spreading them over a now bigger area then results in a poorer definition of the event and a larger error in estimating its areal average.

The long term average of an event is, according to the statement of this work, an unknown but constant value in time and space (stationarity and homogeneity conditions). Each observation in time and space is an estimate of the long term areal mean. In the limit, one station at a point, for one or more time periods, is as good an estimate of the long term mean over an infinitely small area as over a very large area. Clearly, this is different from the case of the areal mean of an event. Imagine N stations over a small area, observing for a fixed number of time periods. Each observation in time and space is an estimate of the long term areal mean. Increasing N, fixing the area size, decreases the error of estimation due to more observations, but at a lesser rate, due to the fact that each new observation in space provides less new information because of increasing spatial correlation. Fixing N instead and increasing the area (spreading out the available number of stations) reduces correlation between stations and so augments information per spatial observation resulting in smaller error of estimating the constant long term areal mean.

The space-time trade-off curves shown in Figures 10 to 14 lead to a very important observation. Notice the sharp curvature of the curves. At any of the extremes, reduction in the number of stations (time) implies tremendous increases in the number of intervals (stations) required. Clearly then an efficient and logical design for a given accuracy level lies at the curve's elbow. Movements to either side of this point imply large changes in design parameters.

CONCLUSIONS

A methodology for designing and analyzing data collection networks with the goal of obtaining the best estimate of the time averaged areal mean of precipitation has been presented. Example results show similarity to those of the previous work by Rodríguez and Mejía [1974]. The technique is different, though, and improved in the sense that it explicitly considers measurement errors and particular placement of observation stations in a systematic, not random, manner. Any network configuration can be studied, with spatially varying measurement errors. The accuracy measure solved for in this work should be further combined wtih a cost objective to have a complete network design criteria. Network design will then be the search for the number and configuration of stations which for a given duration of observations minimizes an objective function of mean square estimation error and cost (see Bras and Rodríguez, 1976a).

Other advantages of the procedure are generality and simplicity. It is very general in the sense that it is not limited to areas of any shape (as other derivations have been) and it is not, theoretically, limited to isotropic and separable covariance structures like the example used here and as required by Rodríguez and Mejía's derivation. In theory, although doubtful in practice, the procedure could be even extended to a non-stationary and non-homogeneous situation.

Another convenient property is the fact that the procedure explicitly acknowledges an unknown mean for the process and allows introduction of prior information. In their statistical approach, Rodríguez and Mejía needed a zero mean assumption which, in fact, implies knowing the answer a priori! This became obvious to the authors when a similar approach was followed and the Kalman filter was used as an estimator. Under a zero mean assumption, as the number of stations increased, the error increased instead of decreased. The estimator was being forced to use noisy observations when, in fact, it knew the areal mean perfectly, it was zero. The reader is referred to Bras and Colón [1977] for this instructive mistake.

The reader is also referred to Bras and Colón for the particular network analysis of a basin in Central Venezuela. There it was found that 5 stations were very satisfactory. The technique was then used to specify which 5 station combination, of the existing ones, gave the most accurate measurement. The same reference also provides a more complete set of figures explaining the results.

In conclusion, a comment on the use of the Kalman filter and other estimation techniques in Hydrology and Water Resources problems is warranted. Certainly, this is a very powerful tool. It nevertheless requires very careful and complete understanding. Extensive use by the senior author constantly points to subtle issues which are usually ignored. Issues of observability, controllability, knowledge and form of prior information, accuracy of model parameters and numerical problems are constantly appearing in the use of Kalman filters. The careless user may end up with perfect looking results which are of little value and nothing but restatements of his assumptions.

ACKNOWLEDGEMENTS

This work was sponsored by the Water Resources Institute, University of Puerto Rico, School of Engineering, under O.W.R.T., Department of the Interior, Project Number A-050-PR. The authors are grateful to Dr. Roberto Vazquez, Director of the Water Resources Institute, during most of the project duration, for his constant cooperation. The principal author also acknowledges the very useful discussions with Prof. Daniele Veneziano of MIT and Dr. Mircea Grigoriu. Mr. Peter Kitanidis and Mr. Edward Johnson also cooperated in the final editing of the article.

REFERENCES

Bras, R. L. and I. Rodríguez-Iturbe, "Network Design for the Estimation of Areal Mean of Rainfall Events," Water Resources Research, Vol. 12, No. 6, Dec. 1976a.

Bras, R. L. and I. Rodríguez-Iturbe, "Rainfall Network Design for Runoff Prediction," Water Resources Research, Vol. 12, No. 6, Dec. 1976b.

Bras, R. L. and I. Rodríguez-Iturbe, "Evaluation of Mean Square Error Involved in Approximating the Areal Average of a Rainfall Event by a Discrete Summation," Water Resources Research, Vol. 12, No. 2, April 1976c.

Bras, R. L. and R. Colón, "The Design of Optimum Data Collection Networks to Estimate the Time Averaged Areal Mean of Precipitation in Puerto Rico," Technical Report, Water Resources Institute, University of Puerto Rico, Mayaguez, Puerto Rico, 1977.

Eagleson, P. S., "Optimum Density of Rainfall Networks," Water Resources Research, Vol. 4, No. 3, 1967.

Rodríguez-Iturbe, I. and J. M. Mejía, "The Design of Rainfall Networks in Time and Space," Water Resources Research, Vol. 10, No. 4, August 1974.

Schweppe, F. C., Uncertain Dynamic Systems, Prentice Hall, N. J., 1973.