REAL TIME ESTIMATION OF AREAL RAINFALL WITH LOW DENSITY TELEMETERED NETWORKS.

Thierry. LEBEL
Institut Français de Recherche Scientifique pour le Développement en Coopération(ORSTOM).
B 5045 - 34032 Montpellier Cedex- France.
G. BASTIN.
Laboratoire d'Automatique, Dynamique et Analyse des Systèmes
Université de Louvain. B 1348 Louvain la neuve. Belgique.

ABSTRACT The ability of telemetered networks to provide accurate estimates of rainfall-runoff model inputs is undermined by the high cost of automated stations and telemetering facilities. It is therefore of great importance to assess the resulting accuracy loss when areal rainfall is estimated using low density telemetered networks as compared to conventional raingage networks of greater density. Such an assessment has been carried out on a catchment area located in southern France, within the region covered by the flash flood warning system described in an accompanying paper presented at this symposium. The behaviours of two linear estimators are compared, using various network densities, mainly from the estimation error variance standpoint. For this comparison, the mean areal rainfall computed from a non-telemetered high density network is used as a reference. The two linear estimators considered are Thiessen polygons and climatological kriging.

INTRODUCTION

Real time flash flood forecasting requires as a first step the real time knowledge of the inputs (mainly rainfall) of the hydrological system. In the flash flood warning pilot system presented in a companion paper by Leoussov and Lebel this requirement is reached through telemetering facilities, the high cost of which results in a relatively low density observation network, thereby reducing the accuracy of the input estimation. The system is presently fully operational in terms of technology but improvement is needed with regard to accurate and timely flood forecasts. Along with a few other meteorological parameters rainfall and runoff data are telemetered to a central site where the flood forecasting is carried out. After studying a few hydrological models, an ARMAX model (Delleur and Obled, 1985) was developed which enables to forecast the runoff one hour in advance for several

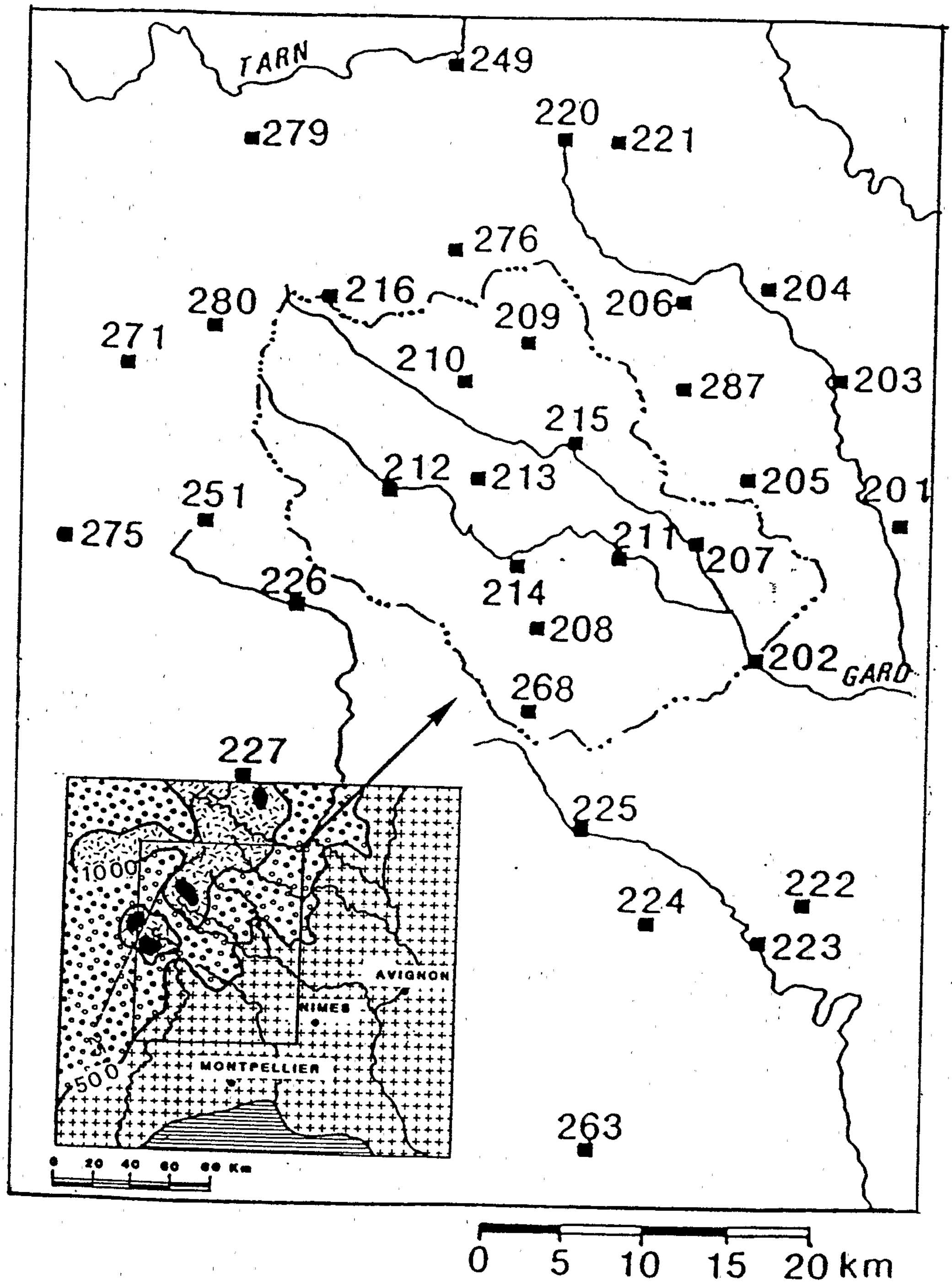

Figure 1 Recording raingage network over the area of study.

watersheds, the area of which range from 300 to 600 Km^2. Today active researchs are still carried out to increase the forecasting lead time up to three hours. It appears nevertheless that a good estimation of the input remains a critical requirement to improve the accuracy of the forecasting. It was consequently decided to compare several areal rainfall estimators (Lebel, 1984) in order to select the most accurate and to assess the loss of information involved when using the telemetered network for ral time computation of areal rainfalls.

This paper intends to summarize the main results of this comparison for two widely used linear estimators : a classical method, the Thiessen polygon technique, and a more recent, statistically based method, kriging. Using the geostatistical approach on which is based the kriging technique, a theoretical accuracy criterion is developed : the scaled variance of the estimation error. This criterion is used as a tool to assess the accuracy of the different estimators and the influence of the network density.

THIESSEN POLYGONS AND CLIMATOLOGICAL KRIGING

The areal rainfall, over an area S, is commonly defined as :

$$Z_k^s = \frac{1}{s} \int_S Z_k(x,y)\, dx dy \qquad (1)$$

where $Z_k\ (x,y)$ denotes the point rainfall depth at the point (x,y), for the k^{th} time interval of duration θ, and $s = |S|$

This quantity Z^s_k is obviously unknown since the rainfall depth is accessible only at a finite number (say n) of scattered pointwise observations. It is therefore common practice in hydrology to estimate Z^s_k using linear estimators of the form :

$$Z_k^s = \sum_{i=1}^{n} \lambda_i Z_k^i \qquad (2)$$

i.e. as a weighted mean of the random variables $Z^1{}_k, Z^2{}_k, \ldots, Z^n{}_k$ observed at the raingages. Thiessen polygons and Kriging differ from one another in the way of computing the coefficients λ_i.

Thiessen polygons : In this method (Thiessen, 1911), the watershed S is divided into n zones of influence S_i, one for each raingage. The zone of influence of a raingage is defined by those points which are closer to that gage than to any other station.

The weighting coefficients λ_i are then computed as :

$$\lambda_i = s_i/s \qquad i = 1,\ldots,n$$

where $s_i = |S_i|$

Climatological kriging : Climatological kriging is an extension of the kriging technique as applied to rainfall fields by Delfiner and Delhomme(1973), Creutin and Obled(1982), Bastin et al.(1984) among others. Kriging is a linear minimum variance estimation method, and as such it requires knowledge of the statistical spatial structure of the random field. This structure is characterized by a structural function, the variogram, which is often identified separately for each realization of the rainfall field. In a real time context this approach has two main drawbacks: i) most often, a large number of field realizations have been observed and are available for the inference of the covariance function. By treating each realization separately, one makes only very partial use of the global statistical information contained in the whole data set; ii) a careful determination of the random field structure function at each time step may be too time consuming for real time operation with short time steps. Furthermore, reliable values of the model parameters cannot be obtained from a small number of data points (less than around 15-20).

Therefore on the basis of several previous investigations (Creutin and Obled, [1982]; Bastin et al., [1984]; Lebel and Bastin, [1985]), it appears that a reasonable trade off is to adopt an analytical variogram model of the form :

$$\gamma(h;k) = \alpha(k)\,\gamma^*(h,\beta) \qquad (3)$$

where h is the euclidian distance, $\alpha(k)$ a scaling parameter and β a shape parameter. With this structure, all time non-stationarity (i.e. dependence on the time index k) is concentrated in the scale factor $\alpha(k)$, while the component $\gamma^*(h,\beta)$ (which we call the "scaled climatological variogram") is time invariant.

In a region of relatively regular weather patterns, Bastin et al (1984) successfully used a single climatological variogram γ^* for the estimation of areal rainfall throughout the year. In such a case the scaling factor $\alpha(k)$ mainly reflects the seasonal variation of the spatial structure of the rainfall field. In regions where the climatic variability is stronger (as it is the case in the Cevennes region), a unique climatological variogram $\gamma^*(h,\beta)$ is used only for storms issuing from the same kind of weather conditions. The parameter $\alpha(k)$ then mainly accounts for the scale effect due to the variation in time of the mean rainfall intensity. When the variogram is bounded we can impose, without loss of generality, that:

$$\lim_{h\to\infty} \gamma^*(h,\beta) = 1$$

$\alpha(k)$ being the variance of the k^{th} field and $\gamma^*(h,\beta)$ the unique variogram of all the scaled random fields defined by :

$$Z^*_k = Z_k/\sqrt{\alpha(k)} \qquad k=1,\ldots,K \qquad (4)$$

The scaled experimental variogram of the process Z^* is obtained from the accumulation of K scaled field realizations :

$$\gamma^*(h_{ij}) = \tfrac{1}{2}\sum_{k=1}^{K}\left\{(z^i_k)^* - (z^j_k)^*\right\}^2$$

where $(z^i_k)^* = z^i_k/\sqrt{\alpha(k)}$,
and z^i_k denotes the value taken by the random variable Z^s_k.

It can thus be proceeded to the identification of a model of the climatological variogram γ^* of Z^*. This inference is performed using a "mean squared interpolation error" (MSIE) criterion (Lebel and Bastin, 1985) and is based on a much larger data set than the one which would have been used in a single realization context.

An example of the scaled climatological variogram model corresponding to the time-step $\theta = 1$ h and using the entire recording raingage network (fig.1) over the Cevennes region is shown in figure.2.

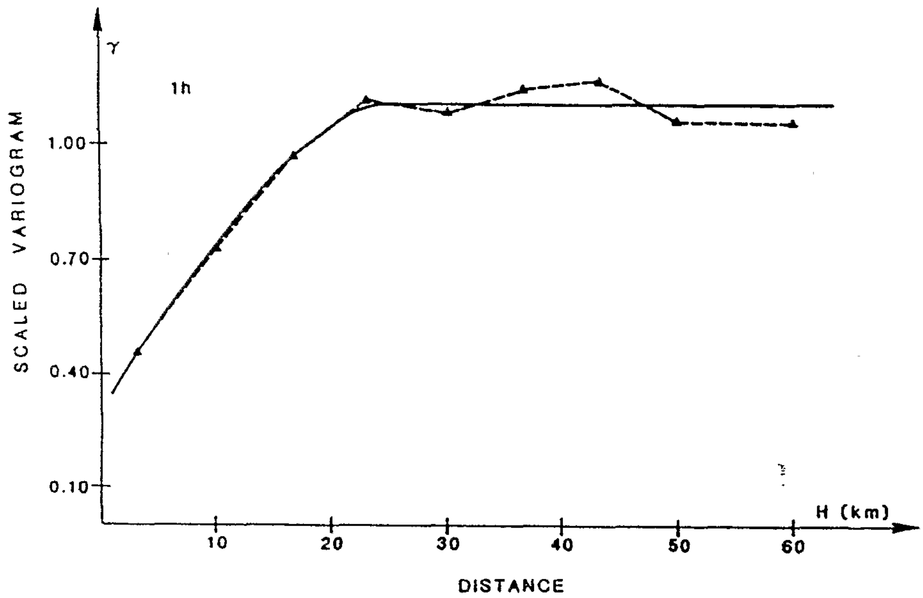

Figure 2 Scaled climatological variogram of hourly rainfall.

The density of the classical raingage network used to identify the variogram model is much larger than the density of the telemetered network presented in the companion paper by Leoussov and Lebel. The basic data were the strongest hourly autumn rainfalls between 1971 and 1980. These rainfalls were accumulated to obtain 2, 4, 6, 12 and 24 hour rainfalls. The identification of the scaled variogram for each of the six time steps was carried out using the MSIE method. The main results are as follow:

a) A locally constant drift model and a spherical isotropic scaled climatological variogram model were selected : it should be emphasized that this structural choice (stationary drift, bounded and isotropic variogram...) is not arbitrary, but results from a careful analysis.

b) The spherical variogram model is of the form :

$$\gamma^*(h,\beta) = \frac{1}{2}\left\{(3\,\frac{h}{\beta}) - (\frac{h}{\beta})^3\right\} \quad for \;\; 0 \leq h \leq \beta$$

$$\gamma^*(h,\beta) = 1 \qquad for \;\; h>\beta$$

where β is the range of the variogram. The optimal estimates of β arising from the identification study are illustrated in figure 3, along with the empirical relation :

$$\beta(\theta) = 25 \;\; \theta^{0.3}$$

that has been derived from these values. This relation allows computation of β (and hence of the areal rainfall) at time steps for which the variogram was not inferred.

c) The parameter $\alpha(k)$ is estimated as the experimental variance of the k^{th} realization of the rainfall field.

For further details see Lebel and Bastin (1985).

An advantage of the climatological approach is that the coefficients λ_i of the Kriging estimation are independent of $\alpha(k)$. Hence, they depend only on the scaled climatological variogram $\gamma^*(h,\beta)$. and can be computed once and for all, as for the Thiessen estimator. This makes the climatological kriging technique very well suited to real time applications.,

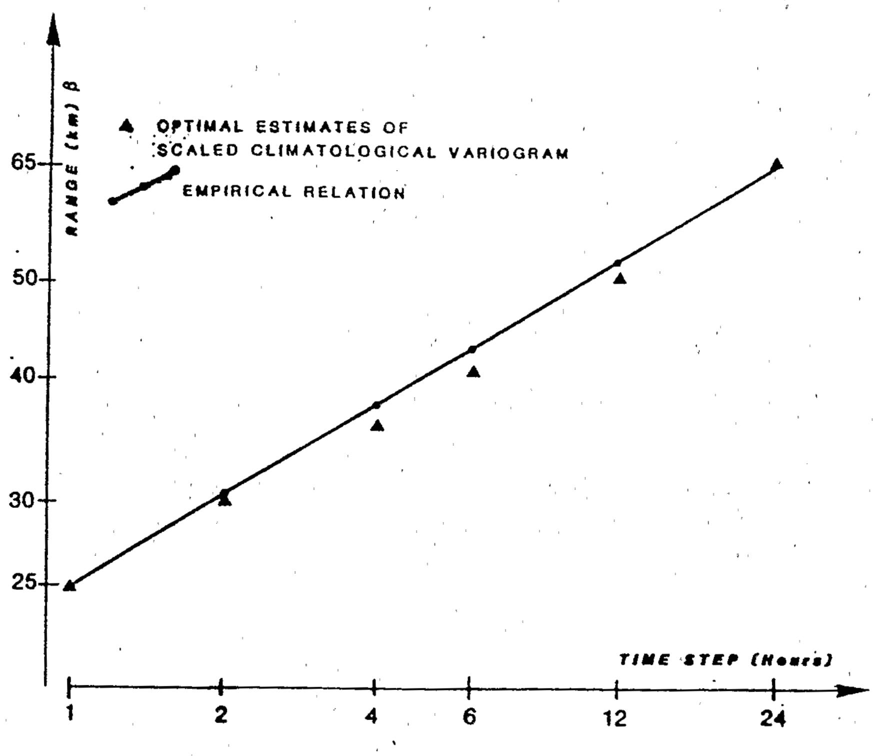

Figure 3 Relationship between the duration of rainfall accumulation and the range of the spherical variogram model.

MEASURING THE AREAL RAINFALL ACCURACY WITH THE SCALED VARIANCE OF THE ESTIMATION ERROR.

The variance of the estimation error of any linear estimator is defined as:

$$(\sigma^s_k)^2 = Var\left(Z^s_k - Z^s_k\right) = Var\left(\sum \lambda_i Z^i_k - Z^s_k\right) \qquad (5)$$

A byproduct of the climatological variogram approach is the possibility of computing a scaled variance of estimation error $(\sigma^s_u)^2$, that can be used as a global (i.e. not relevant to a single event) comparative index of the accuracy of the areal rainfall estimation for various network densities. As a matter of fact, using the climatological variogram defined in section 2,it is easy to show (Lebel and Bastin, 1985) that the estimation error variance (5) of any linear estimator can be written :

$$\left(\sigma^s_k\right)^2 = \alpha(k).\left(\sigma^s_u\right)^2 \qquad (6)$$

with :

$$\left(\sigma^s_u\right)^2 = \frac{1}{s}\sum_{i=1}^{n}\lambda_i\int_S \gamma^*(\overline{u_i u},\beta)\, du - \frac{1}{s^2}\int_S\int_S \gamma^*(\overline{uu'},\beta)\, du\, du' + \mu \qquad (7a)$$

where u_i is the location of raingage i, u and u' are current points in S, $u_i u$ is the euclidian distance between u_i and u, and μ is the Lagrange multiplier (in the case of zero order drift).

In practice, the integrals in expression (7a) are computed using the following discrete approximation :

$$\left(\sigma^s_u\right)^2 = \frac{1}{s}\sum_i \lambda_i \sum_l \gamma^*\left(\overline{u_i u_l},\beta\right) - \frac{1}{s^2}\sum_{l=1}^{L}\sum_{m=1}^{M}\gamma^*\left(\overline{u_l u_m},\beta\right) + \mu \qquad (7b)$$

Given expression (6), $(\sigma^s_u)^2$ is the ratio between the areal rainfall estimation error variance and the field variance.

Once the climatological variogram model has been chosen (i.e. once the value of β has been chosen), the scaled estimation error variance $(\sigma^s_u)^2$ can be viewed as depending exclusively on the number and the locations of the raingages. Therefore, $(\sigma^s_u)^2$ is an efficient tool for solving raingage network optimization problems such as the optimal choice of raingage locations (Bastin et al., 1984). In this paper it will be used as the basic criterion for analysing the infuence of the network density on the estimation and for comparing the two estimators . Of course, the validity of this criterion must be checked since by definition the kriging estimation error variance is lower than that of the other two methods.

INFLUENCE OF THE NETWORK DENSITY ON THE ESTIMATION ACCURACY.

In order to compare several estimates of the areal rainfall over a given watershed (i.e two estimators for various network densities), a reference value has to be chosen since no direct measurement of the areal rainfall is available. The Thiessen estimate based on the network of 34 classical raingages was selected to provide for such a reference value; it is noted $(z^s_k)^t$. In fact any linear estimator using such a dense network could have been selected since they yield equivalent results (Lebel,1984).

Because the density of the telemetered network is fairly low compared to that of the entire network, three other intermediate density networks were considered (figure 4). The network density is defined as the area (1450 km²) of a circle containing the main watershed, divided by the number of gages located inside this circle.

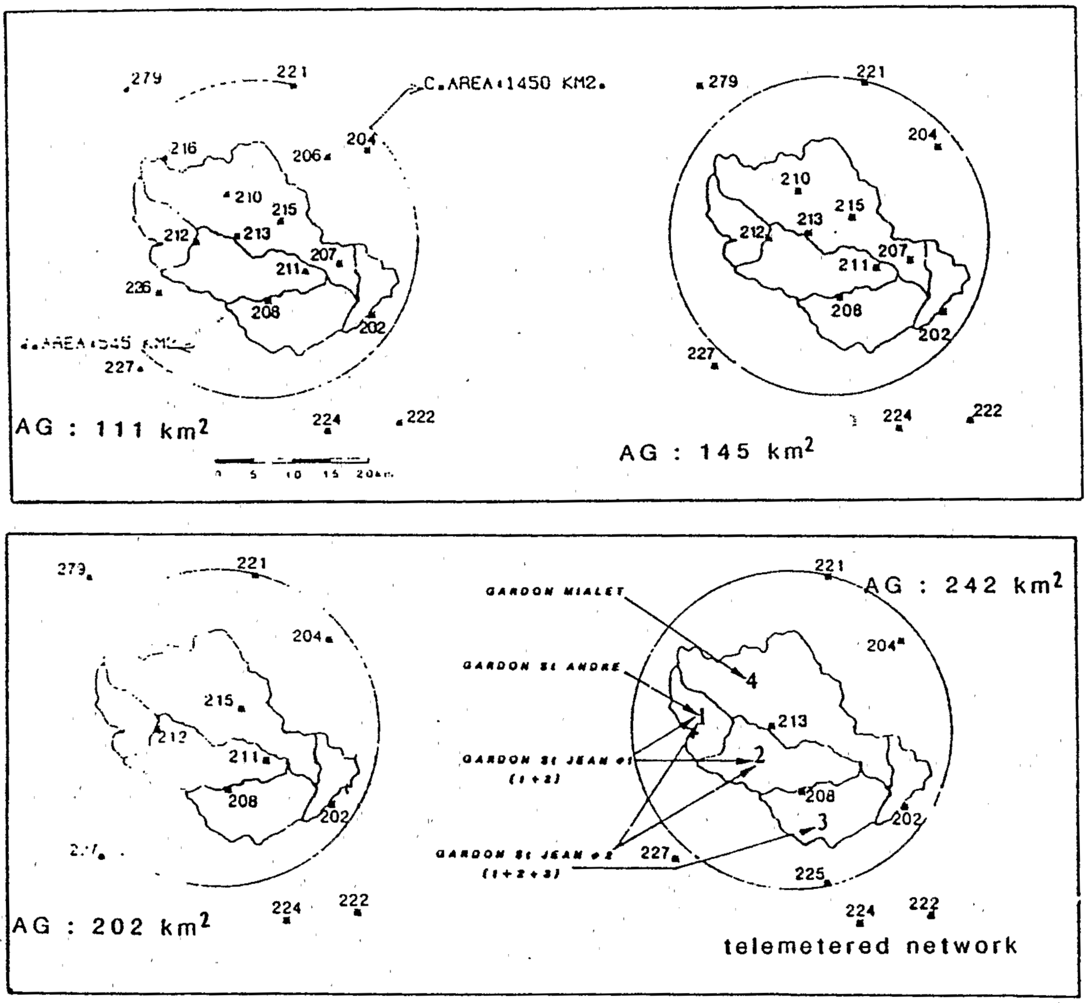

Figure 4 Telemetered network and networks of intermediate density. Ag is the area per gage.

Theoretical variances of estimation error.
The scaled variances of estimation error computed using expression (7b), provide overall comparison criteria, irrespective of the magnitude of a given event. The results of these computations are summarized in table 1 and illustrated in figure 5. It can be seen that:

Because the density of the telemetered network is fairly low compared to that of the entire network, three other intermediate density networks were considered (figure 4). The network density is defined as the area (1450 km^2) of a circle containing the main watershed, divided by the number of gages located inside this circle.

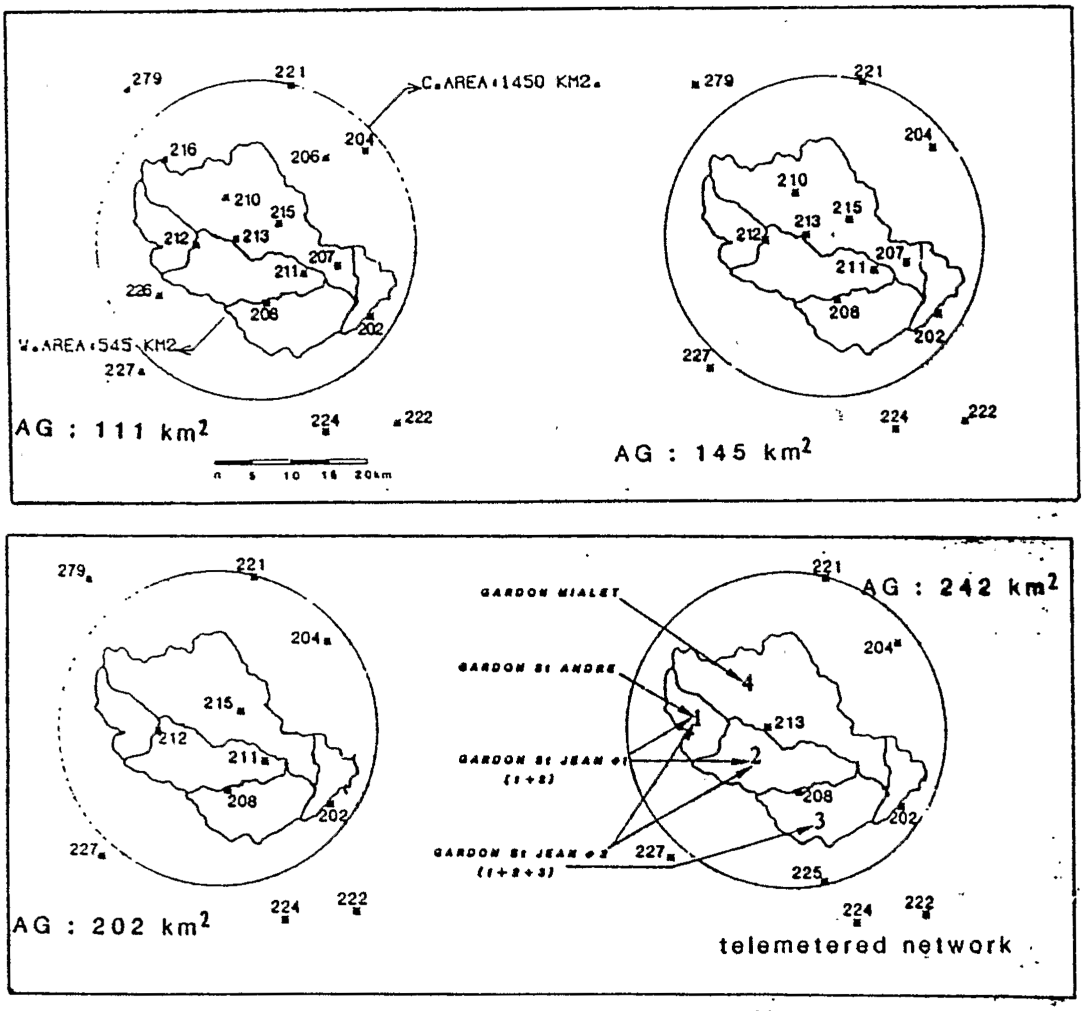

Figure 4 Telemetered network and networks of intermediate density. Ag is the area per gage.

Theoretical variances of estimation error.

The scaled variances of estimation error computed using expression (7b), provide overall comparison criteria, irrespective of the magnitude of a given event. The results of these computations are summarized in table 1 and illustrated in figure 5. It can be seen that:

1) Whatever the method considered, the estimation error variance increases in a fairly regular way as the network density decreases. It tops at 81% of the scaled field variance when the Thiessen estimate is used with the telemetered network on the smallest watershed.
2) By contrast, one can observe the large differences that always exist between kriging and the other estimates, especially when it comes to low density networks.

Nevertheless the conclusion that kriging is more accurate (in a way that can be evaluated with table 1) holds only if the theoretical variance $(\sigma^{*}_{u})^2$ computed from the variogram model is a realistic measure of the actual estimation variance (whatever the estimator considered). This is the reason why we believe that an experimental validation analysis based on the "true" reference values defined above is needed to check the validity of the results of table 1.

Table 1 : Scaled variance of hourly areal rainfall estimation error

Number of stations (Ag km²)*	17 (112)	14 (145)	11 (207)	Telemetered (242)
GARDON A St ANDRE (53 km²)				
Kriging	152	.221	230	563
Thiessen	194	.327	327	814
GARDON A St JEAN (N° 1 - 165 km²)				
Kriging	.040	.054	.063	.148
Thiessen	.047	.067	.078	.210
GARDON DE MIALET (237 km²)				
Kriging	.028	.040	113	.185
Thiessen	.035	.063	.153	289
GARDON DE St JEAN (n°2 - 265 km²)				
Kriging	.033	.039	.042	.080
Thiessen	.039	.049	.054	.110
GARDON D'ANDUZE (545 km²)				
Kriging	.015	020	.039	.074
Thiessen	020	031	055	.108

* Ag is the gage area in square kilometers

Experimental validation.

Experimental validation was carried out in two ways :
1) by computing correlations between the "true" reference values $(z^{s}_{k})^{\dagger}$ and the low density network estimates z^{s}_{k}, on a sample of hourly rain events, and 2) by counting the number of times that the reference value belongs to the theoretical confidence

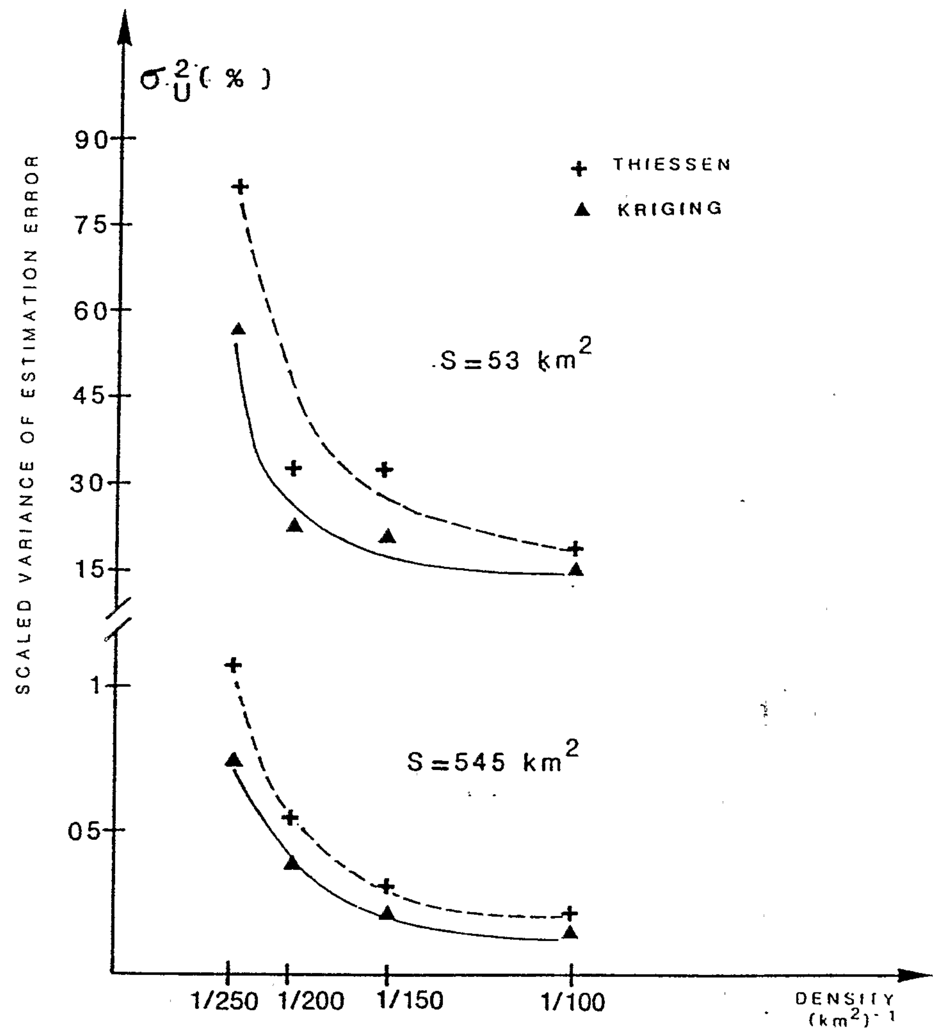

Figure 5 Scaled variance of estimation error as a function of the network density and the surface area.

interval (C.I) of the estimate.Computation of correlation coefficients allows an a posteriori assessment of the cofluctuation of tested estimations and reference values.

It is worth nothing that these "true" values are independent of the variogram model used to computed the estimation error variance since Thiessen estimates were taken as the reference.

The data set used for the computation of the correlation coefficients between the reference values and the various areal estimates was enlarged to include 200 hourly events, i.e.97 events not used in the variogram model inference were taken into account. The Thiessen estimations using the dense network were computed for the 200 events, making up five reference data sets (one for each subwatershed). Thiessen and kriging values were computed using four other networks of decrasing densities thus making up twelve data sets for each subwatershed to be compared to the corresponding reference data set

The variation of correlation coefficients (table 2) with the network density and the watershed area is very similar to the variation of the scaled estimation variances (table 1). The interest of this test is that it applies to the actual rainfall process Z_k. and not to the scaled random field Z^*_k.

Table 2 : Correlation coefficients between reference areal rainfalls computed with the dense network of 34 stations and various estimates (200 events).

Number of stations	17 (112)*	14 (145)	11 (207)	Telemetered (242)
GARDON A St ANDRE (53 km²)				
Kriging	.98	.96	91	49
Thiessen	96	84	.84	14
GARDON A St JEAN (N°1 - 165 km²)				
Kriging	.96	94	.93	78
Thiessen	.92	91	86	64
GARDON DE MIALET (237 km²)				
Kriging	99	96	87	85
Thiessen	.98	93	.76	76
GARDON DE St JEAN (n°2 - 265 km²)				
Kriging	99	96	.87	85
Thiessen	98	.93	.76	76
GARDON D'ANDUZE (545 km²)				
Kriging	98	.96	93	87
Thiessen	97	94	.91	.83

*Area per gage (km²)

An experimental procedure was then derived to test both the accuracy of the estimators and the reliability of the theoretical variances of estimation error. This procedure is as follows :

- Computation of $(\sigma'_u)^2$. and the sampling variance of the k^{th} field $\alpha(k)$ allows computation of the <u>theoretical</u> unscaled variance of the estimation error $(\sigma'_k)^2$ using expression (6). For each subwatershed, the value of $(\sigma'_u)^2$ is given in table 1, while the value $\alpha(k)$ is the same for every watershed since it is a characteristic statistical parameter of the field. $\alpha(k)$ is computed with the dense network in order to provide an accurate as possible estimation of the true variability of the field.

- Once this has been done, the <u>theoretical</u> confidence interval of the estimation is expressed as :

$$\left(z_k^s\right)^{\dagger} \pm c.\sigma_k^s \qquad (8)$$

where c is a constant whose value defines the amplitude of the confidence interval, and σ^s_k is the theoretical unscaled standard deviation of estimation error.
- Next it is determined whether or not the reference value z^s_k belongs to the theoretical confidence interval.
If the distribution of errors is assumed to be Gaussian, z^s_k should belong to the confidence interval 68 times out of a hundred for $c = 1$, and 95 times out of a hundred for $c = 2$.

The test was performed using the reference data sets of 200 hourly events set up to compute the correlation coefficients above. The scores of table 3 are the average over the 200 events for each watershed and each network. It can be seen that, except for the smallest watershed, the proportion of "true" values belonging to the theoretical one standard deviation confidence interval remains around the expected value of .68. Concerning the two standard deviation intervals, the experimental proportions of "hits" are very close to the expected theoretical value of .95. As a consequence, the curves of Figure 5 may be deemed relevant in assessing the performance of one of the network considered herein, with respect to the area of the watershed and the network density.

Table 3 : Percentage of reference values belonging to the theoretical confidence interval computed with expression (8).
a : one standard deviation interval
b : two " " "

	St ANDRE (53 km²)		St JEAN n°1 (165 km²)		St JEAN n°2 (265 km²)		MIALET (237 km²)		ANDUZE (545 km²)	
	a	b	a	b	a	b	a	b	a	b
Kriging										
17	.88	1.00	.60	.79	.72	88	81	96	69	.97
14	82	.96	.58	.69	69	90	67	92	68	.93
11	.85	.96	65	.87	.69	.91	70	96	72	.94
tel	82	.98	.68	90	70	95	69	.95	73	.96
Thiessen										
17	87	.98	.67	87	80	94	78	96	75	.96
14	79	.94	.66	91	77	92	70	92	.71	.92
11	79	96	68	.86	72	92	.56	95	63	.95
tel	81	97	.69	94	.79	96	65	.96	70	97

*Theoretical percentages, assuming a Gaussian distribution of errors are : a : 682
b : 954

COMMENTS AND CONCLUSIONS

The experimental confirmation of the theoretical values of the estimation error variance proves the robustness of the climatological approach and of the variogram inference process. Among the few variogram models available the spherical model is a convenient tool, for it provides a value of the decorrelation distance. In the Cevennes region, the relation between this distance and the time step of rainfall accumulation seems to be well approximated by a power type function. This allows computation of the range (decorrelation distance) for any time step between one hour (25 km) and 24 hours (65 km). Since the procedure of theoretical estimation error variance computation was validated experimentally on hourly data, extension of the method to other time steps appeared founded. The Kriging scaled variances were computed for six time steps (1, 2, 4, 6, 12, 24 hours) with the variogram model inferred from experimental variograms (see values of the range in figure 3). An accuracy measure of the Kriging interpolation process was thus available for various network densities, watershed areas and time steps of rainfall accumulation. This information is summarized in figure 6. For clarity, only a few values are marked in the chart, but other values are easy to infer because the distance between the parallel straight lines is relatively small.

Using the chart of fig. 6, it is possible to evaluate whether or not the density of any network meets a desired accuracy with respect to a given time step and a given watershed area. This is of interest in assessing existing networks as well as in designing a future network. Furthermore, values of the estimation error variance may be derived for time steps at which no data were collected, providing that a good estimation of the correlation range is possible.

This is of particular interest since the time step desired is often dependent on the watershed response to the rainfall input. In many cases, no data are available at that time step, but the chart allows an a priori assessment of the expected variance of estimation error, before any further study or investment are considered . Although the general patterns of this chart are similar to those given by Huff (1970) in Illinois and Woodley et al. (1975) in Florida, it would be of course unwise to extrapolate the results to other areas.

Concerning the accuracy of areal rainfall estimation with the telemetered network, note that the scaled Kriging variance of estimation error remains lower than ten per cent for watersheds of areas greater than 100 km^2. This indicates that ground based networks provide a sufficiently accurate estimation for hourly areal rainfall in this region, since for smaller watersheds one would probably have to work at a time step smaller than one hour. However, it must be kept in mind that this region is very well instrumented, and conversely this approach has proved that in many other French region an accurate real time estimation of areal rainfall is still impossible with currently available data. Given the results of this paper, the increase in accuracy that can be expected from the use of meteorological radar, compared with relatively dense telemetered networks, should be studied carefully to help hydrologists decide which solution is best suited to a given problem.

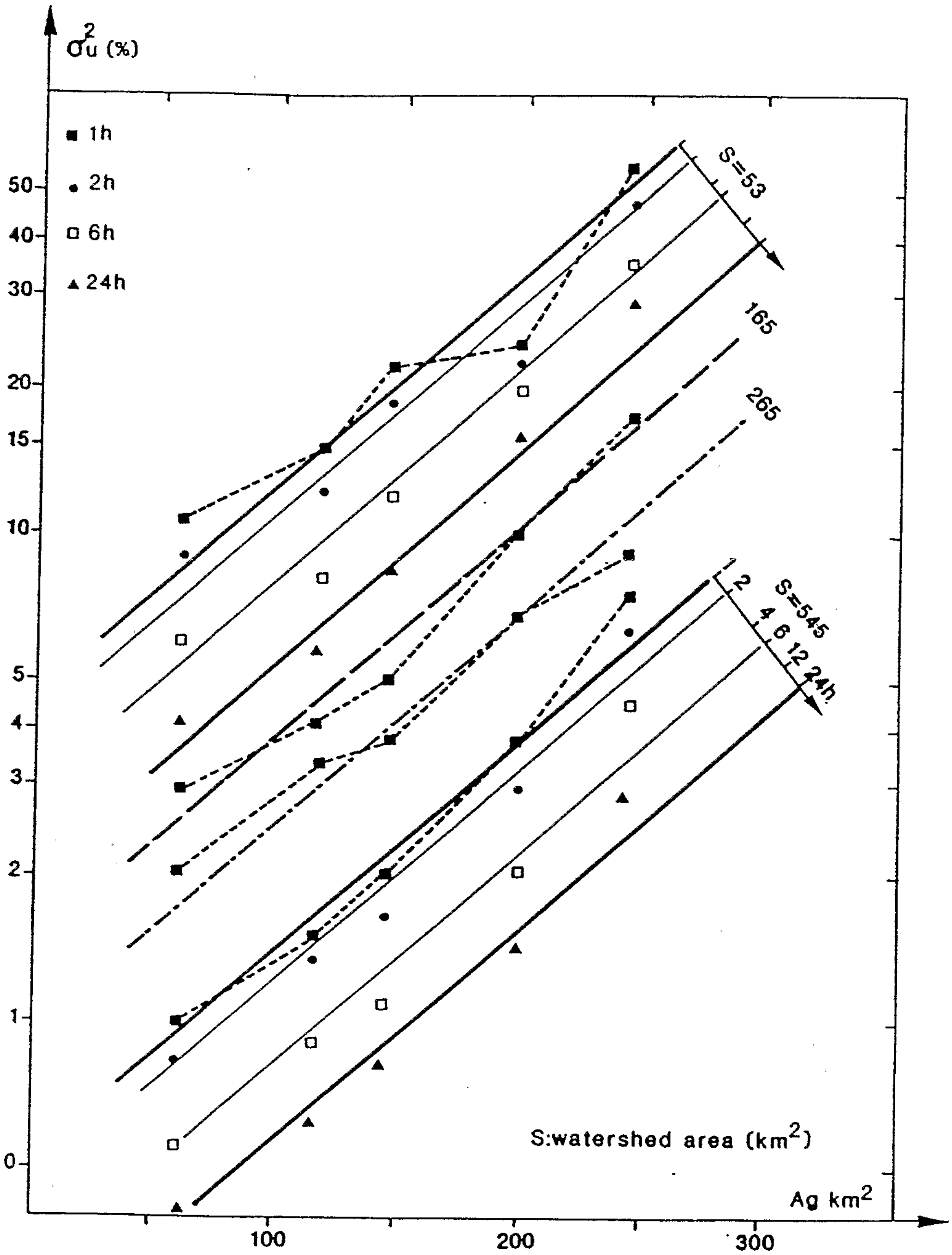

Figure 8 : Theoretical scaled estimation variance of areal rainfall as a function of watershed area, gage area and time step.

Figure 6 Theoretical scaled estimation variance of areal rainfall as a function of watershed area, gage area and time step.

REFERENCES

BASTIN, G., B. LORENT, C. DUQUE, M. GEVERS. Optimal estimation of the average rainfall and optimal selection of raingauge locations. Water Resources Research, 20(4) 463-470, 1984.

CREUTIN, J.D. and C. OBLED. Objective analysis and mapping techniques for rainfall fields : an objective comparison. Water Resources Research, 18(2), 413-431, 1982.

DELFINER, P. and J.P. DELHOMME. Optimum interpolation by Kriging, in : Display and Analysis of spatial data, edited by J.C. DAVIS and M.J. Mc. CULLAGH, pp. 96-114, JOHNY WILEY, New York, 1973.

DELLEUR, J. and C. OBLED. Flash floods forecasting in cevennes region of France : a case study. IWRA 5th world congress on water ressources. June 1985. Brussels. Belgium. 871-881.

HUFF, F.A. Sampling errors in measurement of mean precipitation. J. Appl. Meteor, (9) 35-44, 1970.

LEBEL, T. Moyenne spatiale de la pluie sur un bassin versant : estimation optimale, génération stochastique et gradex des valeurs extrêmes. Thesis, Institut National Polytechnique de Grenoble, 1984.

LEBEL, T. and G. BASTIN. Variogram Identification by the Mean-square Interpolation Error Method with application to Hydrologic fields. Journal of Hydrology. Vol. 77, 31-56, 1985.

THIESSEN, A.H. Precipitation averages for large areas. Montly Weather Review, 39, 1082-1084, 1911.

WOODLEY, W.L., A.R. OLSEN, A. HERNDON and V. NIGGERT. Comparison of gage and radar methods of convective rain measurements. J. Appl. Meteor., 14, 909-928, 1975.

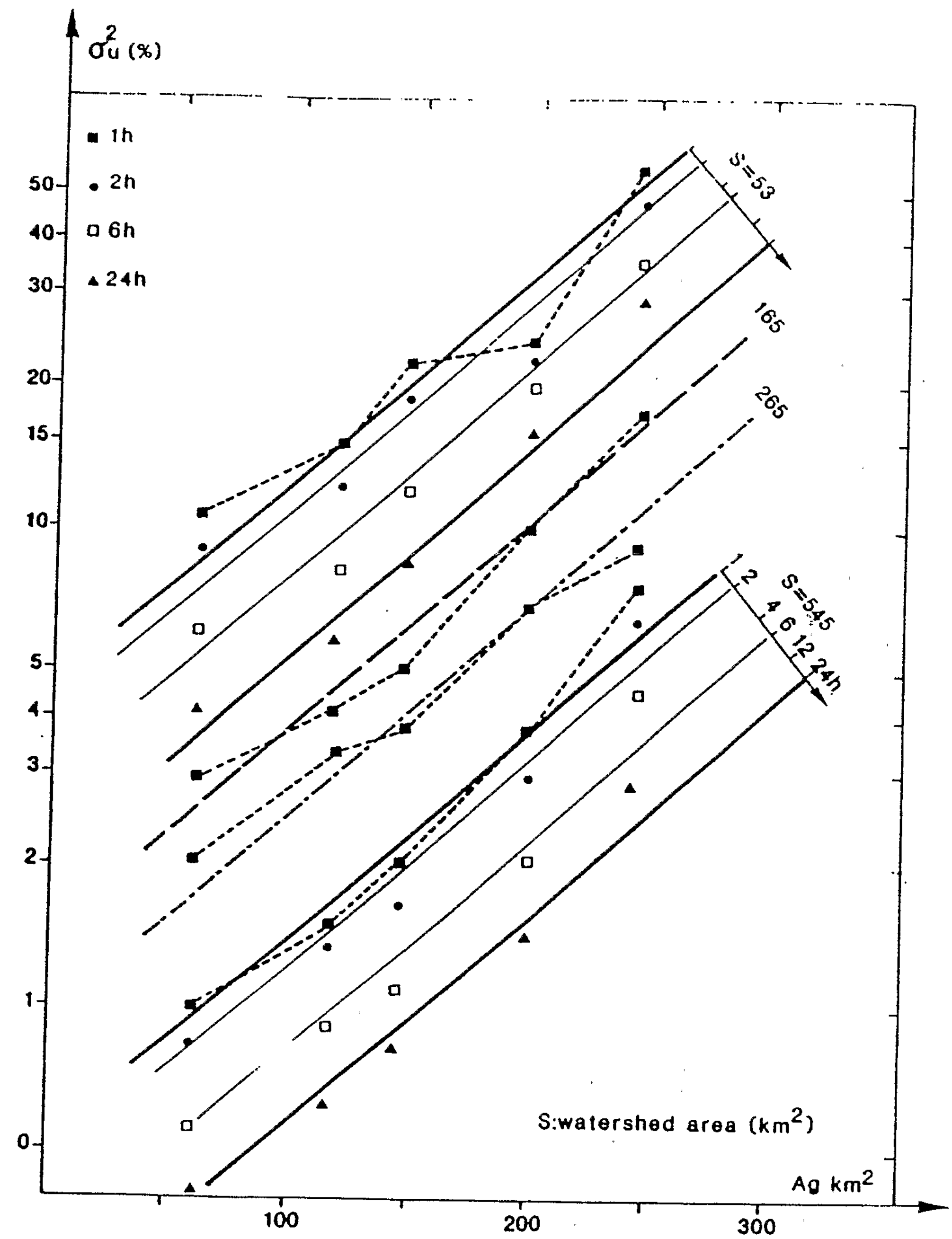

Figure 8 : Theoretical scaled estimation variance of areal rainfall as a function of watershed area, gage area and time step

Figure 6 Theoretical scaled estimation variance of areal rainfall as a function of watershed area, gage area and time step.