WATER RESOURCES RESEARCH, VOL. 20, NO. 4, PAGES 463–470, APRIL 1984

Optimal Estimation of the Average Areal Rainfall and Optimal Selection of Rain Gauge Locations

G. BASTIN, B. LORENT, C. DUQUÉ, AND M. GEVERS

Laboratoire d'Automatique et d'Analyse des Systèmes, Louvain University

We propose a simple procedure for the real-time estimation of the average rainfall over a catchment area. The rainfall is modeled as a two-dimensional random field. The average areal rainfall is computed by a linear unbiased minimum variance estimation method (kriging) which requires knowledge of the variogram of the random field. We propose a time-varying estimator for the variogram which takes into account the influences of both the seasonal variations and the rainfall intensity. Our average areal rainfall estimator has been implemented in practice. We illustrate its application to real data in two river basins in Belgium. Finally, it is shown how the method can be used for the optimal selection of the rain gauge locations in a basin.

INTRODUCTION

We propose a simple procedure for the real-time estimation of the average rainfall over a catchment area from rainfall measurements made at a few measurement stations in that area. The estimation of such areal rainfall is an important step in many hydrological applications, such as evaluation of hydraulic balances, management of surface water resources, or real-time forecasting of river flows. For this last application the rainfall over the river basin is, of course, the main input to any rainfall–river flow forecasting model [*Lorent and Gevers*, 1976].

Following previous contributions [*Creutin and Obled*, 1982; *Rodriguez-Iturbe and Mejia*, 1974; *Chua and Bras*, 1982], the rainfall over a basin is modeled as a two-dimensional random field. This approach allows us to take into account, in a rigorous and systematic way, the seasonal and spatial variability of the rainfall process.

The estimator for the average areal rainfall is then a linear minimum variance unbiased estimator (also called BLUE) [*Papoulis*, 1965], which is obtained by a straightforward extension of the well-known kriging approach [*Delfiner and Delhomme*, 1975; *Journel and Huijbregts*, 1978; *Delhomme*, 1978]. The optimal estimator requires knowledge of the variogram of the rainfall random field as a function of space and time. In order to obtain realistic rainfall estimates, a theoretical variogram model must be chosen, and its parameters must be estimated. This is the most difficult step.

The main contribution of this paper is in the design of a procedure for the real-time estimation of a variogram model. The spatial variability of rainfall data has been analyzed under different sets of assumptions; the seasonal trends of the variogram and the influence of the rainfall intensity have been examined. This has led to the adoption of a simple variogram model, in which the time nonstationarity of the rainfall function is entirely concentrated in a time-varying scaling factor which can be adapted in real time. The advantage is that the weighting coefficients of the optimal rainfall estimator can now be computed once and for all, while the estimation variance is computed in real time using a very simple adaptive procedure.

The objective of our research was to design an adaptive estimator for the average areal rainfall which is simple enough to be used in real time and which does not have to rely on delicate meteorological interpretations. We believe that the proposed procedure achieves these objectives. Our estimator has been practically implemented; we present an application to real data in two river basins.

Finally, as an interesting by-product, we show how the optimal estimation method developed in this paper can also be used to optimally select the location of rainfall gauges in the catchment area.

2. DEFINITIONS AND NOTATION

The point rainfall depth is denoted $p(k, z)$, with $z = (x, y) \in \mathbf{R}^2$, a Cartesian space coordinate, and $k \in \mathbf{N}_+$, an integer index. We consider the discrete sequence (indexed by k):

$$\{p(k, z) \mid k = 1, 2, \cdots, K\}$$

of K nonzero point rainfall depths during K (not necessarily successive) time intervals, each one of duration T_S. In line with previous works [*Creutin and Obled*, 1982; *Rodriguez-Iturbe and Mejia*, 1974; *Chua and Bras*, 1982; *Delfiner and Delhomme*, 1975], for a fixed k, $p(k, z)$ is viewed as a realization of a two-dimensional random field (RF) on $\mathbf{R}^2$ denoted $P(k, z)$.

The mean and the variogram of this field are written as

$$m(k, z) = E[P(k, z)] \tag{1}$$

$$\gamma(k, z_i, z_j) = \tfrac{1}{2} E[\{P(k, z_i) - P(k, z_j)\}^2] \tag{2}$$

with (z_i, z_j) a pair of current points in $\mathbf{R}^2$. It will be assumed, in this paper, that for any k the field $P(k, z)$ is isotropic and fulfills the "intrinsic assumption": (1) the mean is space stationary (independent of z):

$$m(k, z) = m(k) \tag{3}$$

and (2) the variogram is isotropic and space-stationary (it depends only on the Euclidean distance d_{ij} between z_i and z_j):

$$\gamma(k, z_i, z_j) = \gamma(k, d_{ij}) \tag{4}$$

Consider a catchment area (it is most often a river basin) $\Omega \subset \mathbf{R}^2$ with rainfall measurement stations numbered 1 to N. For each value of the index k, the measurements are thus specific numerical values of the function $p(k, z)$:

$$p(k, z_1), p(k, z_2), \cdots, p(k, z_N) \tag{5}$$

Paper number 3W1935.
0043-1397/84/003W-1935$05.00

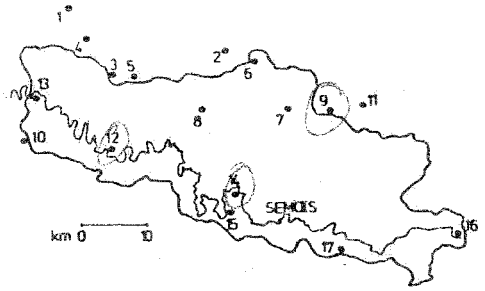

Fig. 1. Semois river basin: rain gauge locations.

This constitutes a realization of the random N vector:

$$\Pi(k) = \{P(k, z_1), P(k, z_2), \cdots, P(k, z_N)\} \tag{6}$$

To illustrate, later in the paper we present two applications: the Semois river basin (Figure 1; 1230 km^2) with $N = 17$ stations, $T_S = 1$ day, and $K = 2557$ daily observations; and the Dyle river basin (Figure 2; 600 km^2) with $N = 16$ stations, $T_S = 6$ hours, and $K = 1425$ six-hourly observations.

3. Optimal Estimation of the Average Areal Rainfall

The "average areal rainfall" $A(k)$ is defined as follows:

$$A(k) = \frac{1}{|\Omega|} \int_\Omega P(k, z)\, dz \tag{7}$$

where $|\Omega|$ is the area of the considered river basin. Clearly, $A(k)$ is a discrete-time scalar random sequence indexed by $k = 1, \cdots, K$. As is well known [see *Journel and Huijbregts*, 1978], an optimal (linear, unbiased, minimum variance) estimator of $A(k)$ can be computed for each k from the set of rainfall observations $\Pi(k)$, by

$$\hat{A}(k) = \sum_{i=1}^{N} \lambda_i(k) P(k, z_i) \tag{8}$$

where the coefficients $\lambda_i(k)$ are the solution of the kriging system:

$$\sum_{j=1}^{N} \lambda_j(k) \gamma(k, d_{ij}) + \mu(k) = \frac{1}{|\Omega|} \int_\Omega \gamma(k, z_i, \zeta)\, d\zeta \qquad i = 1, \cdots, N \tag{9a}$$

$$\sum_{i=1}^{N} \lambda_i(k) = 1 \tag{9b}$$

with $\mu(k)$ a Lagrange parameter. For practical computer implementation, a discretization square grid of M nodes is superimposed on Ω. The nodes are numbered $N + 1$ to $N + M$. The following numerical approximation is used:

$$\frac{1}{|\Omega|} \int_\Omega \gamma(k, z_i, \zeta)\, d\zeta \simeq \frac{1}{M} \sum_{j=1}^{M} \gamma(k, z_i, z_{N+j}) \tag{10}$$

Notice that it is equivalent to replace the right-hand side of equation (9a) by the numerical approximation (10) or to directly adopt a discrete space definition of $A(k)$ in lieu of (7):

$$A(k) = \frac{1}{M} \sum_{j=1}^{M} P(k, z_{N+j}) \tag{11}$$

With the numerical approximation (10), the estimation variance is then given by

$$\sigma_E^2(k) = \mu(k) + \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_i(k) \gamma(k, z_i, z_{N+j}) - \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \gamma(k, z_{N+i}, z_{N+j}) \tag{12}$$

4. Identification of a Variogram Model

The optimal λ_i are computed by the linear system (9) from the knowledge of the variogram $\gamma(k, d_{ij})$. In practice, however, the variogram is not given and must be inferred from the available data. This is the topic of the present section, where we shall study the estimation of the variogram under different sets of assumptions.

Estimation of a Global Mean Variogram

It is well known, of course, that the variogram is a function of the time index k, but as a first step we shall compute a

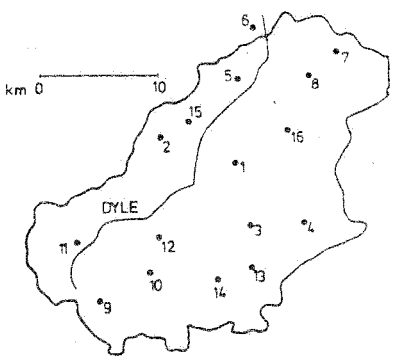

Fig. 2. Dyle river basin: rain gauge locations.

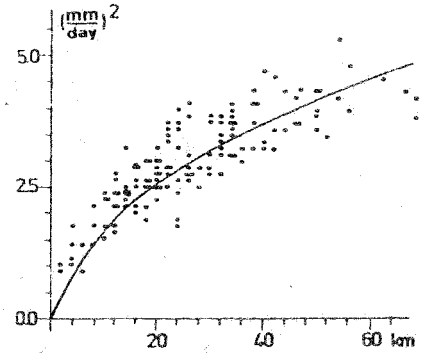

Fig. 3. Semois river: experimental variogram.

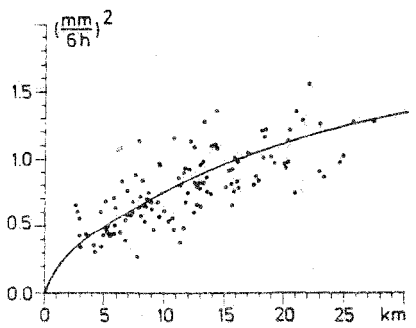

Fig. 4. Dyle river: experimental variogram.

TABLE 1. Global Time-Invariant Estimates of Variogram Parameter

	α	β
Semois river	0.56	0.51
Dyle river	0.204	0.56

global time-invariant estimate of the variogram in the form of a time average over the K time intervals with nonzero rainfalls. The reason for doing this will become apparent later. From the observations in the rain gauges located at points z_i and z_j, the following unbiased estimate is obtained:

$$\hat{\gamma}(d_{ij}) = \frac{1}{2K} \sum_{k=1}^{K} \{p(k, z_i) - p(k, z_j)\}^2 \quad (13)$$

Such an estimate has been computed for every pair of rain gauges in the Semois river basin ($K = 2557$) and in the Dyle river basin ($K = 1425$). The results are graphically presented in Figures 3 and 4.

The experimental variogram takes the form of a somewhat extended cluster of points. On the basis of many experimental results such as those of Figures 3 and 4, and in line with common practice in the geostatistical literature, we shall fit the following very simple model to the experimental variogram:

$$\gamma(d_{ij}) = \alpha {d_{ij}}^{\beta} \quad (14)$$

By a least squares fit, the values of Table 1 are obtained (for d_{ij} expressed in kilometers). The reason why this global model is of interest is that the parameter β of this time-invariant model will be used later in a time-dependent model.

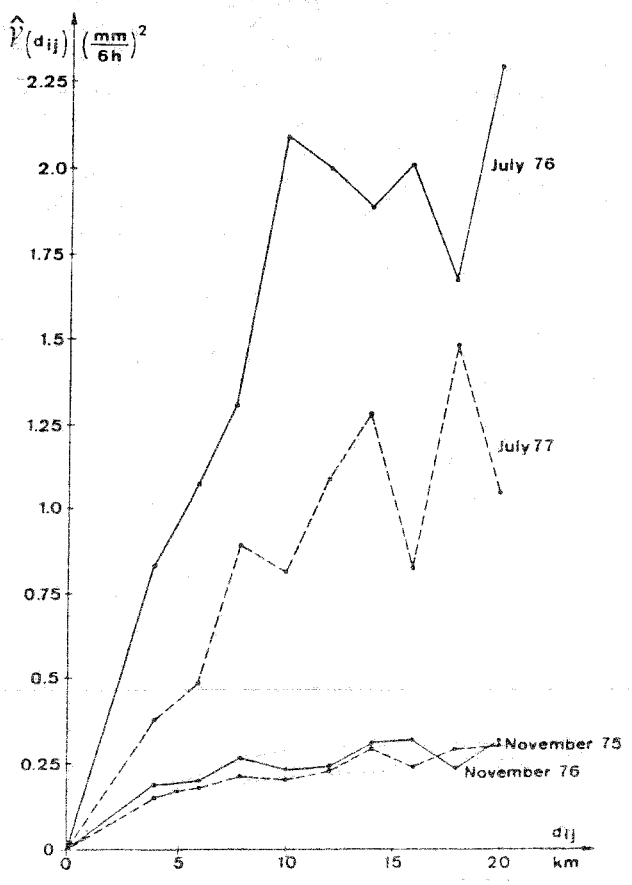

Fig. 5. Experimental monthly variograms for the Dyle river.

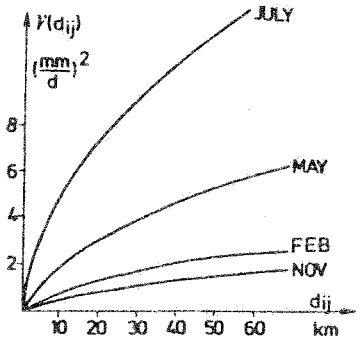

Fig. 6. Semois river: estimated monthly variograms.

Note that other forms of theoretical variograms (such as the Gaussian, exponential, or spherical model) could also be used. It is difficult to validate a particular variogram model experimentally, since the estimate of the average areal rainfall computed with each variogram cannot be compared with a "true" average rainfall, because the latter is unknown. In addition, it has been shown [*Bastin and Gevers*, 1984] that the variogram (in the least squares sense) does not necessarily lead to the best kriging estimates. In any case, the methodology developed further in this paper applies almost unchanged to a general class of variograms, provided they have the form $\alpha\gamma^*(d_{ij}, \beta)$.

Seasonal Trend of the Rainfall RF

It is, of course, unrealistic to assume a time-invariant model for the variogram because (1) it does not take into account the potential seasonal trends of the phenomenon and (2) it would yield a unique estimation variance σ_E^2 of the average areal rainfall (see (12)) for all rainfall events, whatever the meteorological conditions and the rainfall intensity. This is not very plausible.

In order to verify whether a seasonal trend is present, monthly experimental variograms have been computed. Typical examples for the Dyle river basin are presented in Figure 5. For graphical clarity the clusters of points have been approximated by a dashed line which is obtained by dividing the d_{ij} axis into a number of classes and by computing the means of $\hat{\gamma}(d_{ij})$ for all points d_{ij} which belong to the same class. This figure clearly shows the seasonal behavior of the spatial variability of the rainfall process: the variogram appears much larger in summer than in autumn. Similar trends have been observed all through every year for which we had data (with

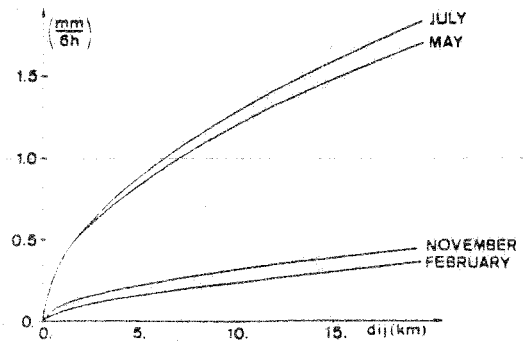

Fig. 7. Dyle river: estimated monthly variograms.

TABLE 2. Estimated Values of α and β for the Monthly Variograms of the Form $\alpha(m)d_{ij}^{\beta(m)}$

	Semois Basin		Dyle Basin	
m	$\alpha(m)$	$\beta(m)$	$\alpha(m)$	$\beta(m)$
January	0.15	0.63	0.067	0.44
February	0.28	0.54	0.063	0.59
March	0.23	0.62	0.072	0.60
April	0.35	0.47	0.221	0.29
May	0.51	0.59	0.362	0.52
June	1.57	0.40	0.673	0.30
July	1.59	0.51	0.368	0.54
August	1.11	0.51	0.505	0.51
September	0.53	0.53	0.144	0.54
October	0.21	0.74	0.042	0.56
November	0.13	0.62	0.105	0.49
December	0.32	0.53	0.090	0.61
Global	0.56	0.51	0.204	0.56

rare exceptions due to very special meteorological conditions such as the drought of 1976 in western Europe).

One could think of explaining these seasonal variations in meteorological terms, such as convective rainfalls in summer or frontal rainfalls in winter. However, since we want to propose a simple procedure for the real-time estimation of the average areal rainfall, we shall include all the seasonal behavior of the rainfall in the variogram parameters, so as to avoid the delicate problem of meteorological interpretation.

The results of Figure 5 suggest that a first way of incorporating the seasonal variations is to assume a piecewise stationary seasonal trend (on a monthly basis) for the RF. More precisely, we assume that the variogram $\gamma(k, z_i, z_j)$ is time-invariant during a month, but not necessarily from one month to another. The theoretical variogram model is now written as

$$\gamma(k, z_i, z_j) = \gamma(m, d_{ij}) = \alpha(m)d_{ij}^{\beta(m)} \tag{15}$$

where m is the index of the month to which the day k belongs ($m = 1, 2, \cdots, 12$).

We have computed a theoretical variogram of the form $\alpha(m)d_{ij}^{\beta(m)}$ for each month by least squares fitting to all the available data for that month; for example, the data of November 1975 and November 1976 are taken in the same class and processed together.

The results are shown in Figures 6 and 7 and in Table 2. Again, they clearly show the seasonal patterns of the variograms. It follows from Table 2 that the monthly variograms differ much more in the coefficient $\alpha(m)$ than in the coefficient $\beta(m)$. This observation suggests a further simplification of the variogram model: The coefficient $\beta(m)$ is assumed time-invariant and set to a fixed value β^* while all the time non-stationarity of the RF is concentrated in the scaling factor $\alpha(m)$. The variogram model is then written as

$$\gamma(k, z_i, z_j) = \alpha(m)\gamma^*(d_{ij}) \tag{16}$$

with

$$\gamma^*(d_{ij}) = d_{ij}^{\beta^*} \tag{17}$$

where m is the index of the month to which day k belongs. It is a kind of "separation of variables"; the variogram model is separated into two factors: $\alpha(m)$, which is time-varying but space-invariant, and $\gamma^*(d_{ij})$, which is time-invariant but space-dependent. In the case of the Dyle river basin, for example, one can take $\beta^* = 0.56$, which is the value obtained by a least squares fit over all the available data points for this basin (see Figure 4 and Table 1). Then $\alpha(m)$ can be computed by fitting the model (16) to the cluster of points corresponding to each month. When the value of β^* is fixed a priori, it is actually better to use a weighted least squares fitting for the estimation of $\alpha(m)$, where the weighting matrix takes into account the geometrical location of the rain gauges in the basin. For more details, see *Bastin and Gevers* [1984]. The results are shown in Table 3 and Figure 8. Figure 8 shows the estimated monthly variograms of the form $\alpha(m)d_{ij}^{\beta(m)}$ and the simplified monthly variograms of the form $\alpha(m)d_{ij}^{\beta^*}$ for four typical months. The figure shows that the simplification is certainly justified: the curves are so close that they can hardly be distinguished.

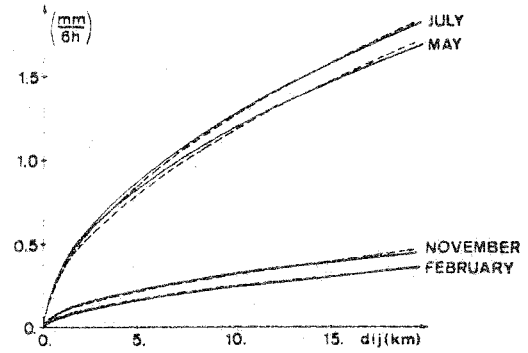

Fig. 8. Dyle river: estimated monthly variograms (solid curves), identical to Figure 7, and simplified monthly variograms (dashed curves).

The simplified form (16) of the variogram model has some important implications for the computation of the optimal estimate of the average rainfall:

1. It is easy to show that the system (9) with the approximation (10) can be rewritten as

$$\sum_{j=1}^{N} \lambda_j \gamma^*(d_{ij}) + \mu^* = \frac{1}{M}\sum_{j=1}^{M} \gamma^*(d_{i,N+j}) \qquad i = 1, \cdots, N \tag{18a}$$

$$\sum_{i=1}^{N} \lambda_i = 1 \tag{18b}$$

TABLE 3. Estimated Values of α and β for the Monthly Variogram $\alpha(m)d_{ij}^{\beta(m)}$ and the Simplified Monthly Variogram $\alpha(m)d_{ij}^{\beta^*}$ for Four Typical Months in the Dyle River Basin

	Original Variogram Model		Simplified Variogram Model	
	α	β	α	β^*
February	0.063	0.59	0.068	0.56
May	0.362	0.52	0.326	0.56
July	0.368	0.54	0.349	0.56
November	0.105	0.49	0.088	0.56

TABLE 4. Number of Rainfall Events in Each Class and for Each Season in the Dyle River Basin

	Intensity Range for Each Class, mm/6 hours					
	0.0–0.2	0.2–0.4	0.4–0.8	0.8–4.0	4.0–7.2	7.2–12.8
Winter	78	62	68	133	20	2
Spring	67	47	61	122	18	12
Summer	100	53	45	114	30	14
Fall	103	62	48	137	22	7

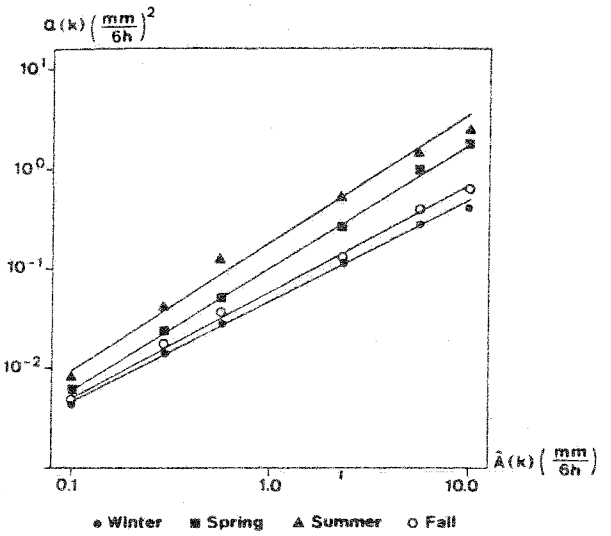

Fig. 9. Seasonal relationship between the estimated average rainfall $\hat{A}(k)$ and the variogram scale factor $\alpha(k)$.

Hence the λ_i are independent of $\alpha(k)$, and consequently they are time-invariant. Therefore the optimal estimator

$$\hat{A}(k) = \sum_{i=1}^{N} \lambda_i P(k, z_i) \tag{19}$$

turns out to be a unique time-invariant weighted sum of the observations $P(k, z_i)$, where the weighting coefficients λ_i depend only upon the geometrical location of the rain gauges and can be computed once and for all.

2. One can also show that the estimation error variance can be written as

$$\sigma_E^2(k) = \alpha(k) V_E^* \tag{20}$$

with

$$V_E^* = I^* + \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{M} \lambda_i \gamma^*(d_{i,N+j}) - \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \gamma^*(d_{N+i,N+j}) \tag{21}$$

TABLE 5. Weighting Coefficients λ_i

Semois River Basin ($V_E^* = 0.48$)		Dyle River Basin ($V_E^* = 0.63$)	
Rain Gauge	λ_i, %	Rain Gauge	λ_i, %
1	0.8	1	6.6
2	2.0	2	5.4
3	3.8	3	5.6
4	2.0	4	8.2
5	3.6	5	4.7
6	3.3	6	2.9
7	8.5	7	5.6
8	9.6	8	5.6
9	6.5	9	6.4
10	3.9	10	8.4
11	5.2	11	8.8
12	8.9	12	6.8
13	4.0	13	5.9
14	9.8	14	6.8
15	7.4	15	4.6
16	8.6	16	7.4
17	12.1		

TABLE 6. Semois River Basin Estimation Results for Some Chosen Days in 1971

	Date, 1971				
	Jan. 26	April 26	June 18	Aug. 8	Dec. 19
Rain Gauge					
1	33.7	14.4	27.0	14.2	11.4
2	34.7	17.2	26.0	11.4	11.8
3	32.3	16.3	31.6	24.0	12.9
4	29.0	15.5	34.2	14.2	16.3
5	33.8	16.6	33.6	20.0	13.2
6	32.0	16.8	35.2	9.1	9.8
7	30.3	20.0	26.4	7.7	9.5
8	35.8	17.8	32.3	12.3	11.2
9	31.5	15.7	19.3	8.2	10.0
10	28.4	15.6	41.8	23.2	7.3
11	33.9	20.7	28.2	7.0	12.4
12	39.5	19.6	39.6	22.8	9.2
13	35.3	12.8	29.3	24.5	12.8
14	28.2	24.5	32.9	11.0	12.0
15	24.5	23.1	38.4	11.1	8.0
16	31.0	21.4	24.3	7.8	11.1
17	30.9	21.5	29.9	8.0	13.3
Estimation results					
$\hat{A}$	31.9	19.5	30.9	12.7	11.1
$\hat{\alpha}$	2.09	2.08	6.39	8.30	0.85
σ_E	0.99	0.98	1.73	1.97	0.63
$\sigma_E/\hat{A}$	3.1%	5.0%	5.6%	15.5%	5.7%

Values are in millimeters per day.

Here also, the variance V_E^* is time-invariant and can be computed once and for all; the time dependence of $\sigma_E^2(k)$ is only through $\alpha(k)$.

Influence of the Rainfall Intensity

For both the Semois and the Dyle river basin we have pointed out the important seasonal variations of the variogram. One might wonder, however, whether these seasonal variations in the variogram are not greatly amplified by the differences between the mean rainfall intensity in summer and in winter. More specifically, are the larger values of the variogram in summer not caused by the higher intensity of the rainfall during that season rather than by a truly larger spatial variability?

In order to study the potential relation between the rainfall intensities and the variogram, we have performed the following analysis:

1. The estimates $\hat{A}(k)$ of the average areal rainfall for all the available rainfall events are computed, using (19) with the λ_i computed by (18).

2. The rainfall data are partitioned into four seasons defined as winter: January, February, March; spring: April, May, June; summer: July, August, September; and fall: October, November, December.

3. For each season the rainfalls are subdivided into six classes according to their intensity levels measured by $\hat{A}(k)$ computed in part 1 of the analysis. For the Dyle river basin the number of available data in each class for each season is given in Table 4.

4. For each class in each season the coefficient $\alpha(k)$ of the variogram model is computed by a least squares fitting as in the previous sections.

The result of this procedure is illustrated for the Dyle river by the chart presented in Figure 9, which shows $\alpha(k)$ versus $\hat{A}(k)$ (with a bilogarithmic scale) for each of the four seasons.

TABLE 7. Dyle River Basin Estimation Results for Some Chosen 6-hour Periods in 1976, 1977, and 1978

	Dec. 6, 1978 1200–1800	April 28, 1978 1200–1800	May 16, 1977 1200–1800	July 21, 1976 0600–1200	Nov. 1, 1976 1800–2400	Nov. 12, 1976 1800–2400
Rain Gauge						
1	1.8	10.3	7.0	6.3	0.8	1.5
2				14.6	1.0	0.0
3	2.6	9.3	7.8	3.6	1.1	2.4
4	0.0	4.7	6.1	2.3	1.3	2.3
5	2.8	6.8	10.3	10.4	1.1	2.6
6	1.3	12.4	9.3	14.0	1.2	2.8
7	1.9	4.5	10.3	11.6	1.2	
8	1.1	7.1	14.7	11.6	0.9	2.5
9	1.6	2.6	9.6	6.9	1.4	2.4
10	1.2	0.6	12.0	6.2	1.8	2.2
11	1.4	4.3	14.0	9.3	1.8	
12	1.7	5.5	10.0	8.0	1.7	2.2
13	1.9	7.4	0.0		1.6	2.2
14	1.3	6.2	4.6	4.0	1.9	2.7
15	2.6	4.3	11.0	18.0		2.6
16	2.7	11.2	12.0	7.1	0.9	2.2
Estimation results						
$\hat{A}$	1.76	5.83	9.40	7.98	1.34	2.17
$\hat{x}$	0.077	1.093	1.783	3.156	0.051	0.107
σ_E	0.22	0.83	1.06	1.41	0.18	0.26
$\sigma_E/\hat{A}$	12.7%	14.3%	11.2%	17.7%	13.3%	11.8%

Values are in millimeters per 6 hours.

The chart shows that even when rainfalls of the same intensity levels are considered, the seasonal trend is still clearly present in $x(k)$. For example, when $\hat{A}(k) = 1$, the value of $x(k)$ is twice as large in summer as in winter. However, the chart also shows that the estimation error variance $\sigma_E^2(k) = x(k)V_E^*$ is a function of the rainfall intensity and that choosing a unique variogram model within a given season would lead to a systematic underevaluation of σ_E^2 for high-intensity rainfalls and to an overevaluation for low-intensity rainfalls. Therefore the procedure that we have adopted ultimately for the computation of the average areal rainfall is as follows:

1. For each basin considered, using all available data, compute once and for all the coefficient β^*, the weighting factors λ_i, the normalized error variance, and the chart $(x(k)\text{-}\hat{A}(k))$ just described and illustrated in Figure 9.
2. Then, for each period k, compute

$$\hat{A}(k) = \sum_{i=1}^{N} \lambda_i p(k, z_i)$$

pick the value of $x(k)$ corresponding to $\hat{A}(k)$ on the chart, and compute $\sigma_E^2(k) = x(k)V_E^*$.

Notice that with this procedure the time nonstationarity of the random field is concentrated in the time-varying scalar parameter $x(k)$, which is computed in real time and which takes into account both the seasonal variations and the effects of the rainfall intensity through the use of the chart $(x(k)\text{-}\hat{A}(k))$.

Tables 5, 6, and 7 show some typical results of this estimation procedure for the two basins considered here.

5. Optimal Selection of the Rain Gauge Locations

As we have pointed out in the previous section, the normalized estimation error variance V_E^* depends only on the geometrical location of the measured points. Obviously, the choice of the variogram model and of the parameters is conditioned by the particular set of available data. But once the variogram model is chosen, the variance V_E^* can be viewed as depending exclusively on the location of the rain gauges. Hence it becomes possible to compute the error variance V_E^* associated with any set of hypothetical data points without getting actual data at these points. Therefore the normalized variance V_E^* is an efficient tool for solving rain gauge allocation problems. For example, the variance V_E^* can be used for (1) choosing the location of an additional rain gauge in order to improve the estimation accuracy as much as possible, (2) selecting the best locations of additional rain gauges among a set of admissible locations, and (3) selecting the most representative subset of M rain gauges from a set of N available ones.

We shall now illustrate these points in the case of the Semois river described in the previous section.

Iterative Selection of the Most Representative Rain Gauges

Two potential supplementary rain gauge locations (numbered 18 and 19) are added to the 17 existing ones (see Figure 10). For each of the 19 locations we can compute the variance V_E^* as if each of them was the only one available and select the one that leads to the smallest V_E^*.

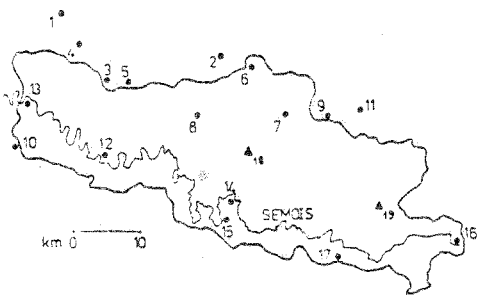

Fig. 10. Two potential supplementary rain gauges (triangles).

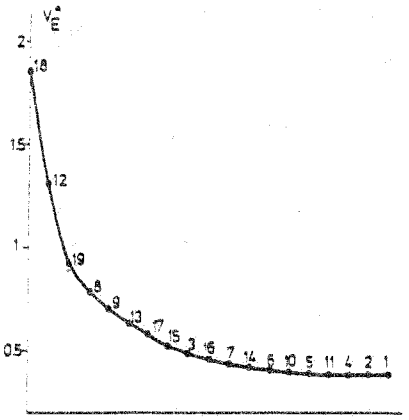

Fig. 11. Iterative selection of the rain gauge locations in the Semois river basin.

Next we can add to this first gauge a second station which, combined with the first one, leads to a minimum V_E^* again. This procedure can be continued, adding more stations and monitoring the decrease of the normalized variance V_E^* of the estimation error, until the obtained precision is judged satisfactory. For the Semois river basin the result of this successive selection is illustrated in Figure 11. We notice that (1) the last seven rain gauges chosen (6, 10, 5, 11, 4, 2, and 1) are obviously superfluous, since including them in the optimal estimator does not result in any significant decrease of V_E^* and (2) the two potential supplementary locations (18 and 19) are among the three "best" ones in order to improve the average rainfall variance estimation.

Selection of the Most Representative Subset of Three Rain Gauges

At some point, the Belgian Ministry of Public Works, in charge of waterways, decided to equip three existing rain gauges in the Semois river basin with telemeasurement facilities. It was therefore desirable to choose the best subset of three rain gauges among the 17 existing ones. The normalized variance V_E^* was computed for all possible combinations of three locations. The best configuration was found to be (9, 12, 14). Notice, however, that with this configuration the estimation variance V_E^* is slightly larger than with the configuration (12, 18, 19) involving the two hypothetical locations 18 and 19 (see Table 8).

We have also compared the estimated values of $\hat{A}(k)$ and of $\sigma_E/\hat{A}(k)$ using either the three stations (9, 12, 14) or the 17 stations. Typical results for the Semois river basin are presented in Table 9.

River Flow Prediction

The average areal rainfall, estimated by the procedure presented in this paper, has been used as input of a rainfall–river flow model that is used for the short-term prediction of river flows. Despite the fact that with the configuration (9, 12, 14), $V_E^* = 1.02$ as compared with $V_E^* = 0.48$ when 17 rain gauges are used, it has been observed that this deterioration of V_E^* with the configuration (9, 12, 14) increases the river flow prediction error variance only slightly [*Gevers and Bastin*, 1982]. This observation is of interest for the implementation of a real-time telemeasuring network; it shows that river flows can be forecast in real time with good precision using only a very limited number of telemeasured gauges.

TABLE 8. Normalized Estimation Variance and Weighting Coefficients for Several Rain Gauge Configurations in the Semois River Basin

Rain Gauge Configurations	V_E^*	Weighting Coefficients λ_i
9, 12, 14	1.02	$\lambda_9 = 0.35$, $\lambda_{12} = 0.31$, $\lambda_{14} = 0.34$
12, 18, 19	0.93	$\lambda_{12} = 0.33$, $\lambda_{18} = 0.35$, $\lambda_{19} = 0.32$
17 stations	0.48	see Table 4

TABLE 9. Comparison Between the Estimated Values of $\hat{A}(k)$ and of $\sigma_E/\hat{A}(k)$ for Some Days in 1971 Using Either Three or 17 Stations, in the Semois River Basin

	Date, 1971				
	Jan. 26	April 26	June 18	Aug. 16	Dec. 19
Three-Station Rain Gauge Configuration (9, 12, 14)					
$\hat{A}(k)$, mm/d	32.9	19.9	30.4	13.7	10.4
$\sigma_E/\hat{A}(k)$, %	4.4	7.3	8.4	21.2	8.9
17-Station Rain Gauge Configuration					
$\hat{A}(k)$, mm/d	31.9	19.5	30.9	12.7	11.1
$\sigma_E/\hat{A}(k)$, %	3.1	5.0	5.6	15.5	5.7

6. CONCLUSIONS

We have presented a new procedure for the estimation of the average areal rainfall over a river basin. The procedure is simple and concise and can be implemented in real time: it does not require any meteorological interpretation of the rainfalls. The estimated average rainfall is a fixed linear combination of the measured point rainfall depths. The time nonstationarity is reflected in the variance of the estimate through an adaptive parameter $\alpha(k)$, which can be read off directly from a precomputed chart; this parameter takes into account both the seasonal variations and the effects of the rainfall intensity.

One of the by-products of our procedure is that it yields a simple method for the selection of "the most informative" rain gauges among a set of existing ones, or for the selection of an optimal location to install additional rain gauges. An application has been presented.

Our procedure has been used for the real-time estimation of the average areal rainfall on several river basins. The estimated rainfalls were then used as input of a river flow prediction model.

Acknowledgments. This work has been supported by the Fondation Universitaire Luxembourgeoise and by the Fonds de Développement Scientifique (UCL). The authors wish to thank the Institut Royal Météorologique de Belgique for providing the Semois river data, and the Département de Génie Rural (Louvain University) for providing the Dyle river data.

REFERENCES

Bastin, G., and M. Gevers, Identification and optimal estimation of random fields from scattered pointwise data, *Automatica*, in press, 1984.

Chua, S. H., and R. L. Bras, Optimal estimation of mean areal precipitation in regions of orographic influence, *J. Hydrol.*, *57*, 713–728, 1982.

Creutin, J. D., and C. Obled, Objective analysis and mapping techniques for rainfall fields: An objective comparison, *Water Resour. Res.*, *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. McCullagh, pp. 96–114, John Wiley, New York, 1975.
Delhomme, J. P., Kriging in the hydrosciences, *Adv. Water Resour.*, *1*(5), 251–266, 1978.
Gevers, M., and G. Bastin, What does system identification have to offer?, paper presented at 6th Congress on Identification and Systems Parameter Estimation, Int. Fed. of Automat. Control, Washington, D. C., June 7–11, 1982.
Journel, A. G., and C. J. Huijbregts, *Mining Geostatistics*, Academic, New York, 1978.
Lorent, B., and M. Gevers, Identification of a rainfall-runoff process, paper presented at 4th Congress on Identification and Systems Parameter Estimation, Int. Fed. of Automat. Control, Tbilisi, USSR, Sept. 21–27, 1976.
Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw Hill, New York, 1965.
Rodriguez-Iturbe, I., and J. M. Mejia, The design of rainfall networks in time and space, *Water Resour. Res.*, *10*(4), 713–728, 1974.

G. Bastin, C. Duqué, M. Gevers, and B. Lorent, Laboratoire d'Automatique et d'Analyse des Systèmes, Bâtiment Maxwell, B-1348 Louvain-la-Neuve, Belgium.

(Received July 24, 1983;
revised November 23, 1983;
accepted December 6, 1983.)