Nordic Hydrology, 19, 1989, 77-88

Estimation of Catchment Rainfall Uncertainty and its Influence on Runoff Prediction

B. Storm and K. Høgh Jensen
Technical University of Denmark, Lyngby

J. C. Refsgaard
Danish Hydraulic Institute, Hørsholm

Interpolation of spatially varying point precipitation depths introduces uncertainties in the estimated mean areal precipitation (MAP). This paper describes a geostatistical approach – the Kriging method – to calculate the daily MAP on real-time basis. The procedure provides a linear unbiased estimate with minimum estimation variance. The structural analysis of the random precipitation field is automatized by relating the time-varying semivariogram model to the sample variance. This is illustrated on data from a Danish IHD catchment.

The conceptual rainfall-runoff model NAM incorporated into a Kalman-filter algortithm is applied to investigate the effects of uncertainties in MAP on the runoff predictions. Measurement and processing errors are not included in the investigation.

Introduction

The daily mean areal precipitation (MAP) is frequently used in hydrological applications, *e.g.* calculation of regional crop water requirement or as input to lumped hydrological models for predictions of floods or water yields. The MAP is usualy estimated from point values sampled at a limited number of points located in the catchment. Since the precipitation varies in space, it is impossible to achieve an exact estimate of the total precipitation depth. Consequently, it is important to

know the accuracy of the estimate. Neglecting measurement and processing errors, the precision of MAP depends on the observation network as well as the spatial autocorrelation of the rainfall event.

Several methods have been applied in hydrological studies, among which the Thiessen polygonal method, Thiessen (1911), is perhaps the method most often used. Frequently these estimations have been made without any structural analysis of the phenomenon. One reason might be that an analysis of the spatial structure of rainfall events is time-consuming and long records make this impossible without some operational procedure.

Several authors have studied the spatial variations of the precipitation. Among Nordic contributions, Gottschalk and Jutman (1982) presented a method using information about the correlation structure in space, and Simeonides (1984) has made an extensive research including a review of earlier works in this field. Finally, Allerup and Madsen (1981) have proposed a method which specifically takes the topographic information into account.

A widely applied method for interpolation of spatially distributed variables is the kriging method, based on "Theory of Regionalized Variables" developed by Matheron (1963, 1971). This method uses the autocorrelation between the samples and estimates values at ungauged points or regions without bias and with minimum estimation variance. In hydrology and related areas, several applications have been reported, *e.g.* Delhomme (1978, 1978), Virdee and Kottegoda (1984), and Veira *et al.* (1983). In rainfall applications, Delfiner and Delhomme (1975) show an example of MAP estimation for a catchment in Chad, Chua and Bras (1982) use universal kriging technique for estimations in mountainous regions, and recently Bastin *et al.* (1984) have proposed a procedure for real-time estimation of the MAP. This work is particularly interesting because it uses the kriging technique for a large quantity of time-varying data in a convenient manner. They utilize that if the time-varying semi-variogram models are proportional, then the weights associated to each precipitation station become time invariant, whereas the estimation variance depends on the individual time-varying semivariogram. Seasonal semivariograms related to the rainfall intensity are used.

The present paper follows in the line of Bastin *et al.* (1984), but proposes a procedure in which daily semivariograms are identified and relates these to the daily sample variance S^2 of the rainfall event. In future predictions, only S^2 needs to be calculated in order to obtain the estimation variance of the mean areal precipitation (MAP). The procedure is implemented to data from the Danish Tryggevælde Aa catchment. It is assumed that the data are not subject to any kind of drift.

The predicted MAP and estimation variance is used as input to a lumped conceptual rainfall-runoff model which has been reformulated in a state space form. The uncertainties in the MAP input is transferred to an uncertainty in the run-off hydrograph. The paper illustrates how this uncertainty is reduced by the evapotranspiration processes and the smoothing effect of the flow processes in the catchment.

Kriging Interpolation Technique

The MAP for a catchment of area A is defined as

$$p_A(k) = \frac{1}{A} \int_A p(k, z_i)\, dz \tag{1}$$

where $p(k,z_i)$ denotes the point precipitation depth in day k at point $z_i = (x_i, y_i) \in R^2$. Given a set of N observations randomly scattered in space, an optimal estimate of p_A is obtained by

$$\hat{p}_A(k) \equiv \sum_{i=1}^{N} \lambda_i(k)\, p(k, z_i) \tag{2}$$

where $\lambda_i(k)$ is the weight associated to the point z_i. For a fixed day k, each $p(k,z_i)$ can be considered as a realization of a two-dimensional random field (RF) on R^2. The weights $\lambda_i(k)$ are determined such that the RF estimator $\hat{p}_A(k)$ is unbiased and with minimum variance. This is mathematically written as

$$E\{\hat{p}_A(k) - p_A(k)\} = 0 \tag{3}$$

and

$$E\{(\hat{p}_A(k) - p_A(k))^2\} \equiv \text{minimum} \tag{4}$$

The spatial dependency of the RF is described by the autocovariance function $C(k,z_i,z_j)$ or alternatively by the semivariogram, which is defined as

$$\gamma(k, z_i, z_j) \equiv \frac{1}{2} E\{[p(k, z_i) - p(k, z_j)]^2\} \tag{5}$$

where (z_i,z_j) is a pair of points in the region of interest. In simple kriging, which is applied in the present paper, it is assumed that the random function $p(k,z_i)$ is weakly stationary of second order, *i.e.* the expectation $E(p(k,z_i))$ exists and independent of z_i; and for each pair (z_i,z_j), the covariance function $C(k,z_i,z_j)$ exists and depends only on the distance h_{ij} between the points. This implies, however, that the RF has a finite variance. To avoid this rather restrictive demand, a hypothesis of 'intrinsic' stationarity is usually applied in kriging. This weaker assumption requires a finite variance on the increments of RF and not on RF itself, which is expressed by using the semivariogram $\gamma(k,z_i,z_j)$ instead of the covariance $C(k,z_i,z_j)$.

It is additionally assumed that $p(k,z_i)$ is isotropic so that the semivariogram depends only on the distance h_{ij} between z_i and z_j and not on the direction.

By formulating Eqs. (3) and (4) in terms of $\gamma(k,h)$, the weights $\lambda_i(k)$ are found as the solution to the following kriging system

$$\sum_{j=1}^{N} \lambda_i(k)\gamma(k, h_{ij}) - \mu(k) \equiv \bar{\gamma}(k, z_i, A) \qquad i = 1 \ldots, N \tag{6a}$$

$$\sum_{i=1}^{N} \lambda_i(k) \equiv 1 \tag{6b}$$

where $\mu(k)$ is the Lagrange parameter. The minimum estimation variance $\sigma^2_e(k)$ is

$$\sigma^2_e(k) = \mu(k) + \sum_{i=1}^{N} \lambda_i(k)\, \bar{\gamma}(k, z_i, A) - \bar{\gamma}(k, A, A) \tag{7}$$

$\bar{\gamma}(k, z_i, A)$ is the average semivariogram between the data points z_i and all points within the cathment area A, and $\bar{\gamma}(k, A, A)$ is the average semivariogram between all points within A, the within-block variance of classical statistics (Burgess and Webster 1980). The average semivariograms are in practice calculated by discretizing the catchment in a large number of grid squares.

Semivariogram Model

The semivariogram in Eq. (5) is estimated from the data by

$$\hat{\gamma}(k, h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (p(k, z_i) - p(k, z_i + h))^2 \tag{8}$$

where $N(h)$ is the number of sampling pairs separated by a distance within the interval $[h-\Delta h;\ h+\Delta h]$. It displays a series of discrete points corresponding to each value of h.

By definition, $\hat{\gamma}(k,h) = 0$ when $h = 0$, and it increases as h increases. Some models assume that the function reaches a maximum value C (denoted the sill which is approximately equal to the sample variance) at some distance a (called the range). Beyond the range a, it remains constant and the data are not correlated.

Other models may show an ever increasing behaviour, *e.g.* the model of the power type which is presented later. It may also be found that $\hat{\gamma}(k,h)$ approaches a positive definite value C_o called the nugget value as h approaches zero. This discontinuity is interpreted as an existence of variability at scales smaller than the sampling distance or as measurement errors (Journel and Huijbregts 1978). Finally, if the experimental semivariogram shows a constant behaviour C_o (pure nugget) for all k, the data are completely random and no spatial dependence exists. It is important to emphasize what effect this latter behaviour has on the kriging system, Eqs. (6a) to (7). The weights $\lambda_i(k)$ will all be equal to $1/N$ which means that mean areal precipitation is found simply as the arithmetic mean, and σ^2_e is equal to $S^2(k)/N$ where $S^2(k)$ is the sample variance.

Several semivariogram models may represent $\hat{\gamma}(k,h)$ in Eq. (8). Bastin and Gevers (1985) present examples on models particularly suitable in real-time estimations. In the present study, a power type model has been selected

$$\gamma(k,h) = \alpha(k)\ h^{\beta(k)} \tag{9}$$

where α and β are two time-dependent parameters which can be estimated from the experimental semivariogram $\hat{\gamma}(k,h)$. A reasonable and convenient approximation is to assume $\beta(k) = \beta_0$ as time invariant (Bastin *et al.* 1984) which simplifes Eq. (9) to

$$\gamma(k,h) \equiv \alpha(k)\ h^{\beta_0} = \alpha(k)\ \gamma^*(h) \tag{10}$$

Two obvious benefits arise from this simplification. Firstly that the solution of the kriging system, Eqs. (6a) and (6b), does not depend on the time-varying parameter $\alpha(k)$. This means that the kriging system only needs to be solved once (*i.e.* $\lambda_i(k) \equiv \lambda_i$ is time independent). This saves a large amount of computer time. Secondly, the estimation variance σ^2_e may be written as

$$\sigma^2_e(k) = \alpha(k)\ \sigma^{*2}_e \tag{11}$$

where σ^{*2}_e is calculated from Eq. (7) by substituting $\gamma(k,h)$ and $\mu(k)$ with $\gamma^*(h)$ and μ^* respectively. Also, σ^{*2}_e needs to be calculated only once.

As mentioned earlier, the semivariogram is closely related to the sampling variance $S^2(k)$, and it is therefore natural to find a unique relation between the time-varying parameter $\alpha(k)$ and $S^2(k)$. Eq. (11) may be rewritten as

$$\sigma^2_e(k) = \alpha_0 S^2(k)\ \sigma^{*2}_e \tag{12}$$

where α_0 is the average slope on the $\alpha(k)$-$S^2(k)$ curves. It is seen from Eq. (12) that once α_0 and σ^{*2}_e have been identified, the estimation error $\sigma^2_e(k)$ may be directly inferred from the sampling variance, which makes it possible to calculate $\sigma^2_e(k)$ on real-time basis.

Parameter Estimation

The theoretical semivariogram model is found from a so-called structural analysis. By fitting the model to the daily experimental semivariogram $\hat{\gamma}(k,h)$, the parameters $\alpha(k)$ and $\beta(k)$ are estimated. A least square procedure is used in the fitting process.

Daily precipitation data for the period 1980-1984 from the Danish Tryggevælde Aa catchment, see Fig. 1, have been avarilable for the identification. Only days showing an arithmetric sample mean larger than 1 mm/day are included in the analysis.

The result of the structural analysis is shown in Table 1. Approximately one-third of the data record is included in the analysis. In 15% of these cases, the experimental semivariograms $\hat{\gamma}(k,h)$ show no spatial dependence (Pure Nugget

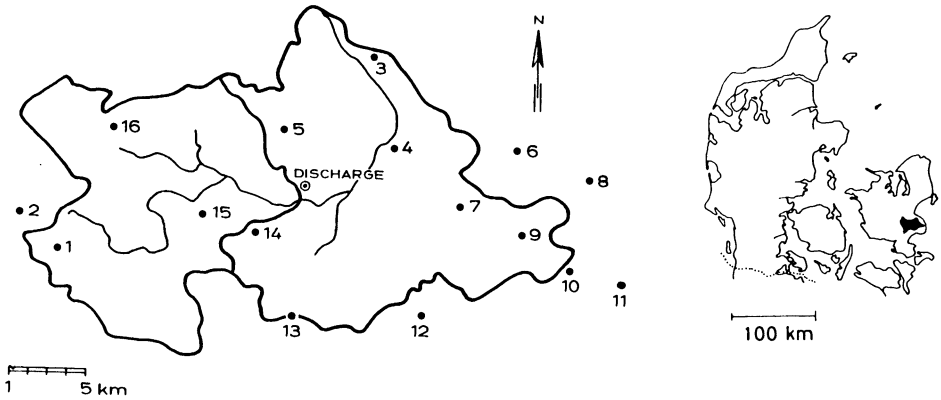

Fig. 1. Tryggevælde Aa catchment, location of precipitation and discharge gauge stations.

effect, $\beta(k) = 0$ and $\gamma(k,h) = \alpha(k)$). Table 1 shows the mean monthly values of $\bar{\alpha}(k)$ and $\bar{\beta}(k)$ from the first fits. As it appears from Table 1, the variations in $\bar{\beta}(k)$ are less pronounced than for $\bar{\alpha}(k)$. In a repeated least square fitting with a fixed $\bar{\beta}(k) = \beta_o = .689$, new values of $\bar{\alpha}(k)$ and α_o are found.

Two examples of daily semivariograms are given in Fig. 2. The theoretical models are the best fits of $\alpha(k)$ with $\beta(k) = \beta_o$. It is seen that the experimental data from April 4 indicates a pure nugget effect without any spatial dependency at all. In a real-time estimation, a constant semivariogram model $\gamma(k,h) = S^2(k)$ would be

Table 1 – Mean Monthly Values of $\alpha(k)$ and $\beta(k)$ Calculated on Daily Basis

Month	Total no. of days	No. of days with Pure Nugget	$\bar{\alpha}(k)$	$\bar{\beta}(k)$	α_o
Jan.	62	16	1.45	.62	.202
Feb.	37	8	2.31	.56	.175
Mar.	53	8	1.36	.57	.175
Apr.	39	4	.93	.67	.195
May	46	7	1.40	.82	.194
Jun.	55	2	1.98	.79	.188
Jul.	41	6	4.04	.67	.184
Aug.	42	4	2.67	.62	.202
Sep.	53	5	1.35	.80	.190
Oct.	79	5	1.43	.68	.191
Nov.	63	10	1.36	.67	.198
Dec.	55	11	1.56	.80	.192
Total	595	86			
Mean				.689 ($= \beta_o$)	.191 ($= \bar{\alpha}_o$)

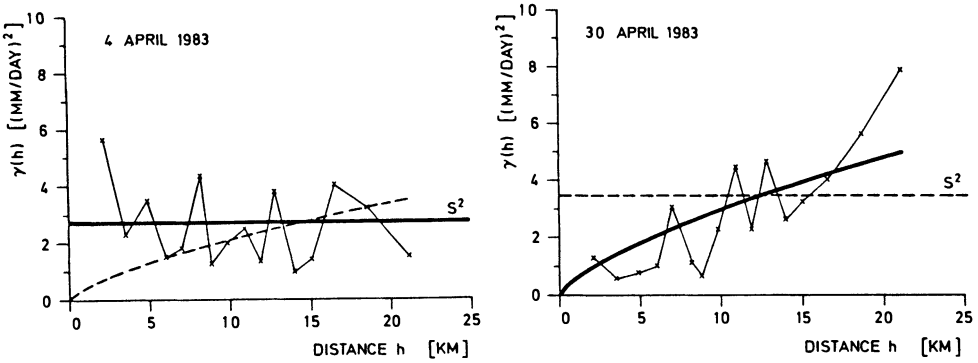

Fig. 2. Experimental and theoretical semivariograms for April 4 and 30, 1983.

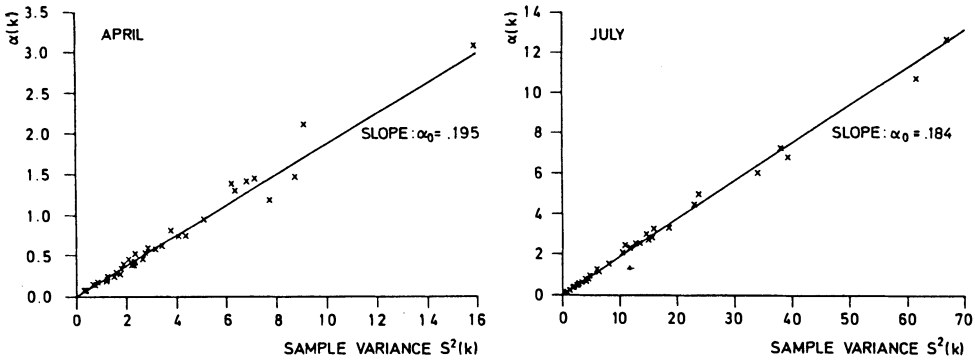

Fig. 3. Relation between sample variance $S^2(k)$ and $\alpha(k)$ for the months April and July.

adopted. A pure nugget effect would in a prediction be identified in the least square procedure by giving a parameter value of β less than or equal to zero. The second example from April 30 shows a clear spatial dependency, and the fitted model is used, but Fig. 2 shows larger S^2 for the dependent situation (April 30) than for the independent (April 4).

Fig. 3 shows the relation between $\alpha(k)$ and the sampling variances $S^2(k)$ for April and July. The slopes α_o for each month are given in Table 1. They show a relative constant value without any clear seasonal variation. It is therefore reasonable to use a constant mean value of $\bar{\alpha}_o = .191$ in Eq. (12).

Runoff Simulation

In order to investigate to what extent the estimation error in the MAP gives rise to errors in the runoff hydrograph, the NAM-model (Nielsen and Hansen 1973) has been used to simulate the runoff from the catchment. NAM is a deterministic

rainfall-runoff model of the lumped, conceptual type. In the present study, the model has first been calibrated for the period 1980-1984 on a subcatchment where the only discharge station in the catchment is situated. A reformulated state-space version, in which the NAM-model is incorporated in a Kalman filter algorithm, was introduced (Refsgaard *et al.* 1983). In this version, the uncertainties of the various states in the model are calculated for every time step. The daily MAP and its standard error are necessary input data to the model. These figures calculated from kriging, are utilized directly as input to the NAM-model/Kalman filter. In this way, it is possible to simulate both the runoff and the uncertainties as a result of various input uncertainties. The model was applied for prediction of the runoff from the whole catchment for the two years 1983 and 1984.

Several sources of errors exist in rainfall-runoff modelling. Besides errors in the input data, not only the MAP but also other climatological input, the resulting error in the simulated runoff may arise from errors in inappropriate model structure and non-optimal model parameter. It is emphasized that these errors also play an important role on the size of the uncertainty of the simulated hydrograph. However, these sources of errors have not been included in the following results.

The simulated hydrograph with a 95% uncertainty band is shown in Fig. 4. It appears from the figure that the error band in general is largest in the wet period, both in absolute and in relative terms. This is confirmed by the monthly average values in Table 2. In the summer period with a small runoff, all the precipitation evapotranspirates and the uncertainty with it. Only large summer thunder storms which result in runoff may give rise to large uncertainties.

The differences in the standard errors between the MAP and the runoff are

Table 2 – Standard Error of Simulated Hydrograph σ_Q, Coefficient of Variation of Simulated Hydrograph CV_Q and Percentage Reduction in Standard Error σ_Q/σ_e between Runoff and MAP.

	σ_Q (mm/d)		CV_Q (%)		σ_Q/σ_e (%)	
	1983	1984	1983	1984	1983	1984
Jan.	.045	.084	3.4	4.5	29	28
Feb.	.044	.033	3.2	2.4	32	37
Mar.	.041	.030	3.8	6.2	29	33
Apr.	.066	.009	4.2	3.9	33	10
May	.081	.002	5.6	2.2	32	2
Jun.	.018	.088	3.0	11.8	16	23
Jul.	.003	.013	2.1	7.1	10	9
Aug.	.002	.006	2.1	7.1	8	5
Sep.	.002	.022	2.1	10.0	1	11
Oct.	.001	.047	2.1	7.0	1	23
Nov.	.011	.025	4.0	4.0	6	33
Dec.	.059	.027	6.9	3.5	39	37

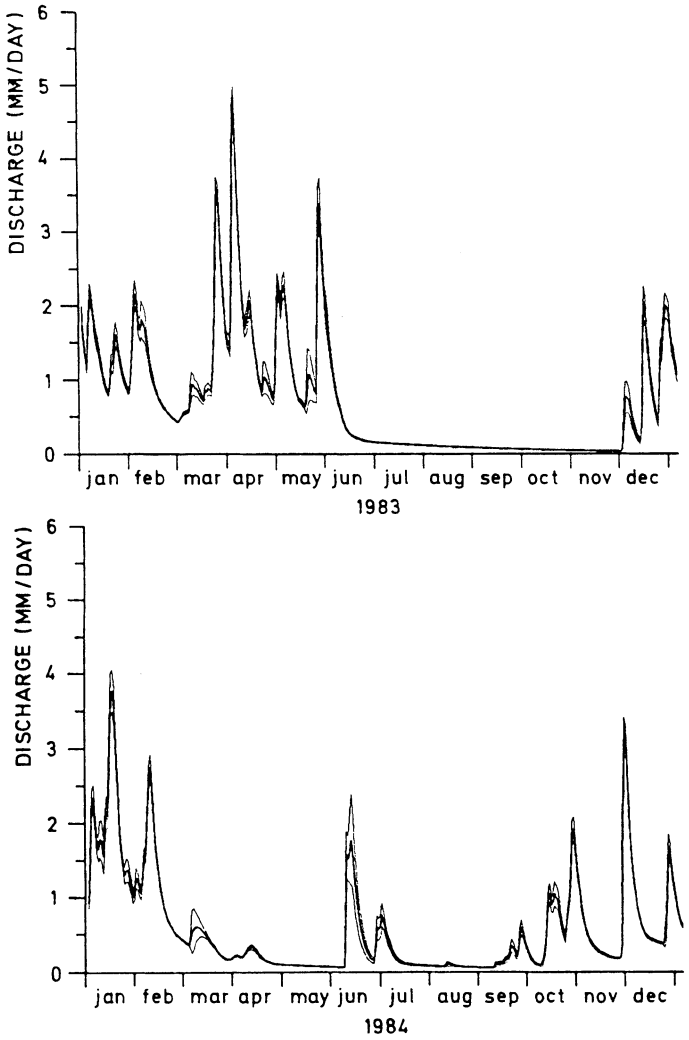

Fig. 4. Simulated daily runoff, 1983-1984, with 95% confidence limits.

illustrated in Fig. 5 and the right column of Table 2. The instantaneous error is strongly damped as a consequence of the filter effect in the model. In the period from late autumn to early spring, even relatively small errors on the MAP result in runoff uncertainties, whereas large isolated rainfall events with large estimation errors during the summer may not show up on the runoff hydrograph. The mean monthly reduction in the standard error is approximately 30% in the wet season and only a few percent in the summer. This seasonal pattern corresponds well with the proportion of rainfall that runs off.

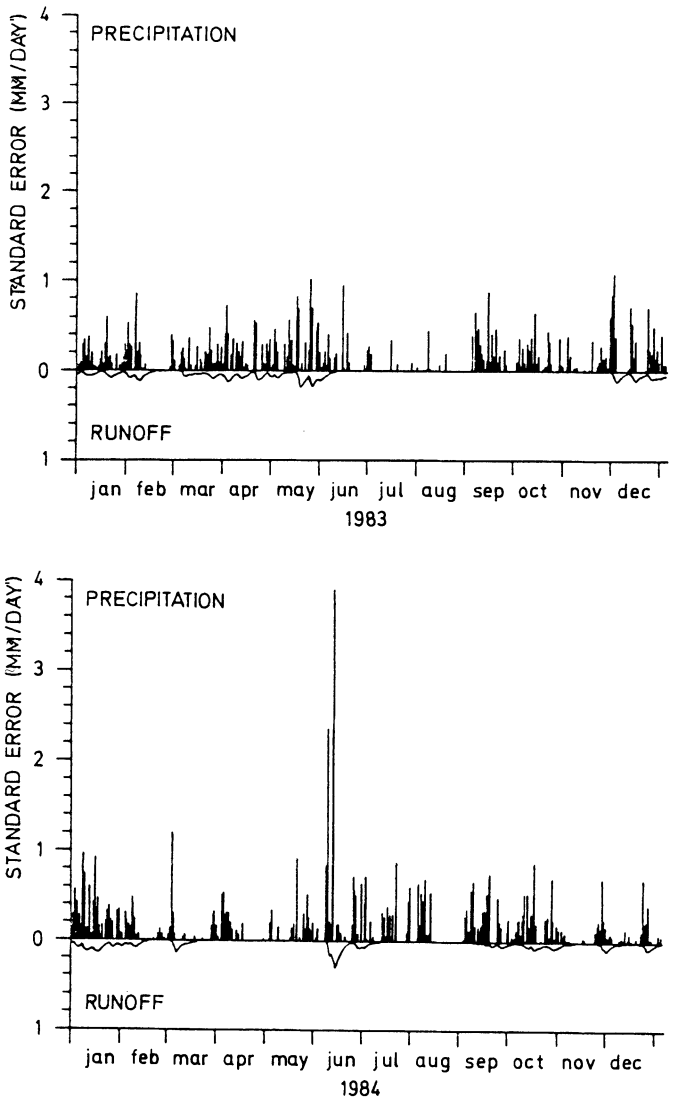

Fig. 5. Daily standard error and mean areal precipitation and runoff for 1983-1984.

Conclusions

The paper applies the kriging interpolation technique for a real-time prediction of the daily mean areal precipitation and the associated estimation variance $\sigma^2_e(k)$. The procedure relates $\sigma^2_e(k)$ directly to the sample variance $S^2(k)$. This relation is found from a structural analysis of 5 years daily data in the catchment. The relation

between $\sigma^2_e(k)$ and $S^2(k)$ is catchment specific, and before predictions in other regions, a similar structural analysis is necessary.

The analysis reveals that in approximately 15 % of days investigated (with MAP larger than 1 mm/day), there is no sign of spatial correlation. At days where a correlation is found, the estimation variance on the MAP is reduced approximately 60 % in relation to an estimation of purely independent data.

The MAP and estimation variances $\sigma^2(k)$ have been used for prediction of a runoff hydrograph with an associated uncertainty band. This was simulated by use of the state-space version of the rainfall-runoff model NAM incorporated in a Kalman-filter algorithm. The study shows a clear filtering effect of the MAP and a reduction in estimation errors on the runoff compared to the estimation errors on the MAP.

References

Allerup, P., Madsen, H., and Riis, J. (1982) Methods for calculating areal precipitation, *Nordic Hydrology, Vol. 13*, pp. 263-278.

Bastin, G., Lorent, B., Duqué, C. and Gevers, M. (1984) Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations, *Water Resour. Rs., Vol. 20(4),* pp. 463-470.

Bastin, G. and Gevers, M. (1985) Identification and optimal estimation of random fields from scattered point-wise data, *Automatica, Vol. 21(2),* pp. 139-155.

Burgess, F.M., and Webster, R. (1980) Optimal interpolation and isarithmic mapping of soil properties. II Block Kriging, *Jour. of Soil Science, Vol. 31,* pp. 333-341.

Chua, S-H. and Bras, R.L. (1982) Optimal estimations of mean areal precipitation in regions of orographic influence, *J. Hydrol, Vol. 57*, pp. 23-48.

Delfiner, P., and Delhomme, J.P. (1975) Optimum interpolation by kriging in: *Display and Analysis of Spatial Data,* Ed. by J. C. Davies and M. J. McCullagh, John Wiley, New York, pp. 96-114.

Delhomme, J.P. (1978) Kriging in the hydrosciences, *Advances in Water Resour, Vol. 1(5),* pp. 257-266.

Delhomme, J.P. (1979) Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach, *Water Resour. Res., Vol. 15(2),* pp. 269-280.

Gottschalk, L., and Jutman, T. (1982) Calculation of areal means of meteorologic variables for watersheds, 7th Nordic Hydrological Conference, NHP-report No. 2, pp. 720-736.

Journel, A.G., and Huijbregts, C.J. (1978) *Mining geostatics,* Academic, New York.

Matheron, G. (1963) Principles of geostatics, *Economic Geology, Vol. 58,* pp. 1246-1266.

Matheron, G. (1971) The theory of regionalized variables and its application. Les Cahrers du Centre de Morphologie Mathematique, Fas. 5, C.G. Fontainebleau.

Nielsen, S.A., and Hansen, E. (1973) Numerical simulation of the rainfall runoff process on a daily basis, *Nordic Hydrology, Vol. 4,* pp. 171-190.

Refsgaard, J.C., Rosbjerg, D., and Markussen, L.M. (1983) Application of the Kalman filter to real-time operation and to uncertainty analyses in hydrological modelling. IUGG/IAHS Hamborg Symposium, IAHS Publ. No. 147, pp. 273-282.

Simeonidis, A. (1984) Precision in estimating mean areal precipitation for small watersheds. UNGI Rapport Nr. 60, Naturgeografiska Institutionen, Uppsala Universitet.

Thiessen, A.H. (1911) Precipitation averages for large areas, *Monthly Weather Rep., Vol. 39 (7),* pp. 1082-1084.

Vieira, S.R., Hartfield, J.L., Nielsen, D.R., and Biggar, J.W. (1983) Geostatistical theory and application to variability of some agronomical properties, *Hilgardia, Vol. 51 (3).*

Virdee, T.S., and Kottegoda, N.T. (1984) A brief review of kriging and its application to optimal interpolation and observation well selection, *Jour. of Hydro Sciences, Vol. 29(4),* pp. 367-387.

Received: 14 October, 1987

Address:
B. Storm and K. Høgh Jensen,
Institute of Hydrodynamics and Hydraulic Engineering, ISVA,
Technical University of Denmark,
Building 115,
DK-2800 Lyngby,
Denmark.

J. C. Refsgaard,
Danish Hydraulic Institute,
Agern Allé 5,
DK-2970 Hørsholm,
Denmark.