Journal of Hydrology, 111 (1989) 321–360
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

[1]

AN APPROXIMATE ANALYSIS OF SURFACE RUNOFF MODEL UNCERTAINTY

T.V. HROMADKA II and R.H. McCUEN

Water Resources Engineering, Williamson and Schmid, Irvine, CA 92714, and
Department of Applied Mathematics, California State University, Fullerton, CA (U.S.A.)
Department of Civil Engineering, University of Maryland, College Park, MD 20742 (U.S.A.)

(Received July 30, 1988; accepted after revision December 5, 1988)

ABSTRACT

Hromadka, T.V. and McCuen, R.H., 1989. An approximate analysis of surface runoff model uncertainty. J. Hydrol., 111: 321–360.

The recent work of Schilling and Fuchs (1986) is useful in evaluating the potential success of a surface runoff model in reliably predicting the runoff response at a stream gage from a catchment, given rainfall data from a single rain gage. Besides demonstrating the relative influences of the various components utilized in surface runoff models (i.e. routing, loss rates, subwatershed response functions, etc.), the cited paper provides a useful indication as to the magnitude of the uncertainty in surface runoff model predictions due to the unknown boundary conditions of the problem.

In this paper, the uncertainty in runoff predictions is approximated by coupling to the rainfall – runoff model a stochastic model which represents the error between measured runoff data and model estimates of runoff.

The stochastic model is developed using a multilinear model of either the catchment runoff itself, or a multilinear equivalent to some particular surface runoff modeling approach. From this coupled model, distributions of the predicted outcomes of criterion variables (e.g., peak flow rate, detention basin maximum volume, etc.) can be obtained, and confidence intervals can be subsequently estimated and used for flood control planning purposes.

Once the distribution of the predicted outcomes of the criterion variable is developed, T-year estimates of the criterion variable are evaluated.

Questions are considered regarding the optimum use of available rainfall–runoff data, and the optimum selection of the effective rainfall estimator (i.e., rainfall less losses, rainfall excess) used to develop the base input to the surface runoff model, when estimating T-year values of a criterion variable.

The main goal of this paper is to provide a mathematical review of methods which would allow designers and researchers to investigate the effect of uncertainty in analysis and design. New notation is introduced which provides a unification to the theory of uncertainty applied to rainfall–runoff models.

INTRODUCTION

In this paper, a multilinear rainfall–runoff model is used to develop uncertainty distributions for runoff hydrographs in the frequently occurring case

0022-1694/89/$03.50

where the uncertainty in the effective rainfall (rainfall less losses, rainfall excess) over the catchment dominates all other sources of modeling uncertainty. Indeed, just the uncertainty in the precipitation over the catchment appears to be a major obstacle in the successful development, calibration, and application of all surface runoff hydrologic models (e.g., Beard and Chang, 1979; Troutman, 1982; Loague and Freeze, 1985; Schilling and Fuchs, 1986). The coupling of the uncertainty in both the rainfall and loss rates results in an important source of uncertainty that should be included in runoff estimates.

Schilling and Fuchs (1986) write "that the spatial resolution of rain data is of paramount importance to the accuracy of the simulated hydrograph" due to "the high spatial variability of storms" and "the amplification of rainfall sampling errors by the nonlinear transformation" of rainfall into runoff. They recommend that a model should employ a simplified surface flow algorithm if there are many subbasins; a simple runoff coefficient loss rate; and a diffusion (zero inertia) or storage channel routing technique.

In their study, Schilling and Fuchs (1986) reduced the rainfall data set resolution from a grid of 81 gages to a single catchment-centered gage in an 1800 acre catchment. They noted that variations in runoff volumes and peak flows "is well above 100 percent over the entire range of storms implying that the spatial resolution of rainfall has a dominant influence on the reliability of computed runoff." It is also noted that "errors in the rainfall input are amplified by the rainfall-runoff transformation" so that "a rainfall depth error of 30 percent results in a volume error of 60 percent and peak flow error of 80 percent". They also write that"it is inappropriate to use a sophisticated runoff model to achieve a desired level of modeling accuracy if the spatial resolution of rain input is low" (in their study, the rain gage densities considered for the 1800 acre catchment are 81, 9, and a single centered gage). Similarly, Beard and Chang (1979) write that in their study of fourteen urban catchments, complex models such as continuous simulation typically have 20 to 40 parameters and functions that must be derived from recorded rainfall-runoff data; and "Inasmuch as rainfall data are for scattered point locations and storm rainfall is highly variable in time and space, available data are generally inadequate... for reliably calibrating the various interrelated functions of these complex models".

Garen and Burges (1981) noted the importance in the effective rainfall estimated for use in the Stanford Watershed Model, because the K1 parameter (rainfall adjustment factor) and UZSN parameter (upper level storage) had the dominant impact on the model sensitivity.

In the extensive study by Loague and Freeze (1985), three event-based rainfall-runoff models (a regression model, a unit hydrograph model, and a kinematic wave quasi-physically based (QPB) model) were used on three data sets of 269 events from three small upland catchments. In that paper, the term "quasi-physically based", or QPB, is used for the kinematic wave model. The three catchments were 25 acres, 2.8 mi^2 and 35 acres in size, and were extensively monitored with rain gage, stream gage, neutron probe, and soil

parameter site testing. For example, the 25 acre site contained 35 neutron probe access sites, 26 soil parameter sites (all equally spaced), an on-site rain gage, and a stream gage. The QPB model utilized 22 overland flow planes and four channel segments. In comparative tests between the three modeling approaches to measured rainfall- runoff data it was concluded that all models performed poorly and that the QPB performance was only slightly improved by calibration of its most sensitive parameter, hydraulic conductivity. They write that the "conclusion one is forced to draw...is that the QPB model does not represent reality very well; in other words, there is considerable model error present. We suspect this is the case with most, if not all conceptual models currently in use". Additionally, "the fact that simpler, less data intensive models provided as good or better predictions than a QPB is food for thought".

Troutman (1982) also discusses the often cited difficulties with the error in precipitation measurements "due to the spatial variability of precipitation". This source of error can result in "serious errors in runoff prediction and large biases in parameter estimated by calibration of the model".

Based on the literature (see Hromadka et al., 1987), a major difficulty in the use, calibration, and development, of surface runoff models appears to be the lack of precise effective rainfall data and the high model sensitivity to (and magnification of) rainfall distribution estimate errors. Nash and Sutcliff (1970) write that "As there is little point in applying exact laws to approximate boundary condition, this, and the limited ranges of the variables encountered, suggest the use of simplified empirical relations".

While surface runoff hydrologic models continue to be developed in technical component complexity, typically including additional algorithms for hydraulic routing effects and continuous soil moisture accounting, the problem setting continues to be poorly posed in a mathematical approximation sense in that the problem boundary conditions, (i.e., the effective rainfall over the catchment) remain unknown. Indeed, the usual case in studying catchment runoff response is to have only a single rain gage and stream gage available for data analysis purposes; and oftentimes, neither gage is within the study catchment. As a result, the rainfall distribution over the catchment remains unknown; hence, the problem's boundary conditions must be approximated as part of the problem solution. The fact that the uncertainty in the effective rainfall distribution over the catchment has a major impact on the success of any hydrologic model's performance and accuracy (e.g., Loague and Freeze, 1985; Schilling and Fuchs, 1986) indicates that the underlying assumption used to specify effective rainfall over the catchment must necessarily be a major factor in the development, calibration, and application, of any hydrologic model.

Some measured magnitudes of the spatial distribution of rainfall rates versus distance obtained from Illinois warm-season storms are given by Huff (1970), see also Table 1.

By coupling to the precipitation variation the variation in loss rates due to nonhomogeneity and other factors, the resulting variation in effective rainfall

TABLE 1

Average variation of point rainfall rates with distance (from Huff, 1970)

Starting point rate (in h^{-1})	Average difference (% for given distance mi)					
	1	2	4	5	8	10
0.1	64	74	81	87	90	93
0.2	61	69	78	82	85	87
0.5	56	63	71	75	77	79
1.0	52	58	65	69	72	74
2.0	49	54	60	64	67	69
5.0	45	49	55	58	60	62

rates versus distance over the catchment motivates the assumption that the effective rainfall distribution over the catchment is a random variable with respect to both space and time.

The Schilling and Fuchs (1986) study provides a good test case in the investigation as to the major sources of surface runoff modeling uncertainty. They write that "errors in simulation occur for a number of reasons, among them: (1) the input data, consisting of rainfall and antecedent conditions vary throughout the watershed and cannot be precisely measured; (2) the physical laws of fluid motion are simplified; and (3) model parameter estimates may be in error". From their detailed sensitivity study of a state-of-the-art surface runoff model, they concluded that for a distributed model with correct parameters, modeling accuracy is not seriously affected by: (1) simplifying the surface flow model if there are many subwatersheds; (2) simplifying the loss rate model with a correct overall runoff coefficient; and (3) using a simplified routing scheme such as the diffusion or storage routing approaches" (e.g., Muskingum routing).

This information is useful in developing a stochastic model to generate uncertainty distributions to be coupled with surface runoff predictions.

In this paper, the uncertainty problem is addressed by providing a methodology which can be incorporated into almost all surface runoff models. The methodology is based upon the standard theory of stochastic integral equations which has been successfully applied to several problems in the life sciences and chemical engineering (e.g., Tsokos and Padgett, 1974, provide a thorough development). The stochastic integral formulation is used to represent the total error between a record of measured rainfall–runoff data and the model estimates, and provides an answer to the question: "based upon the historic rainfall–runoff data record and the model's accuracy in estimating the measured runoff, what is the distribution of probable values of the subject criterion variable given a hypothetical rainfall event?"

Using the stochastic integral formulation, two important problems involving rainfall–runoff modeling predictions are addressed: (1) the distribution of criterion variable probable values given a hypothetical rainfall

event; and (2) the T-year return frequency distribution of a criterion variable given a rainfall–runoff record, and the development of a T-year design storm algorithm for the rainfall–runoff model.

STOCHASTIC FORMULATION OF RUNOFF ESTIMATES

Let M be a surface runoff model which transforms rainfall data for some storm event, i, noted by $P^i(\cdot)$, into runoff data, $M^i(\cdot)$, by:

$$M : P^i(\cdot) \rightarrow M^i(\cdot) \tag{1}$$

The operator M may include loss rate and flow routing parameters, memory of prior storm event effects, and other factors. It is assumed that M is a surface runoff model of complexity comparable to that utilized in the surface runoff model error analysis of Schilling and Fuchs (1986).

Consider the case of having only one rain gage and one stream gage for data purposes, and let $P^i(\cdot)$ be the rainfall measured for storm event i, and $Q_g^i(\cdot)$ be the runoff measured at the stream gage. The problem being studied in this paper is the prediction of runoff quantities at the stream gage site, using the surface runoff model, M.

Various error (or uncertainty) terms are now defined such that for arbitrary storm event i:

$$Q_g^i(\cdot) = M^i(\cdot) + E_a^i(\cdot) + E_d^i(\cdot) + E_r^i(\cdot) \tag{2}$$

where: $E_a^i(\cdot)$ is the modeling error due to inaccurate approximations of the physical processes (spatially and temporally); $E_d^i(\cdot)$ is the error in data measurements of $P^i(\cdot)$ and $Q_g^i(\cdot)$ (which is assumed hereafter to be of negligible significance in the analysis); $E_r^i(\cdot)$ is the remaining "inexplicable" error, such as due to the unknown variation of effective rainfall (i.e., rainfall less losses; rainfall excess) over the catchment, among other factors. The case study involving the highly detailed, fully dynamic, link-node model of Schilling and Fuchs (1986) indicates that $E_r^i(\cdot)$ is the dominant source of uncertainty in eqn. (2).

Let $E^i(\cdot)$ be defined to equal the total error:

$$E^i(\cdot) = E_a^i(\cdot) + E_d^i(\cdot) + E_r^i(\cdot) \tag{3}$$

where $E^i(\cdot)$ is closely related to $E_r^i(\cdot)$ due to the given assumptions. Because $E^i(\cdot)$ depends on the model M used in eqn. (1), then eqns. (2) and (3) are combined as:

$$Q_g^i(\cdot) = M^i(\cdot) + E_M^i(\cdot) \tag{4}$$

where $E_M^i(\cdot)$ is a conditional notation for $E^i(\cdot)$, given model type M.

The term, $E_M^i(\cdot)$, is a realization of a stochastic process when M is used on prediction. That is, for a future storm event D, the $E_M^D(\cdot)$ is not known precisely, but rather is a realization of a stochastic process distributed as $[E_M^D(\cdot)]$ where:

$$[Q_M^D(\cdot)] = M^D(\cdot) + [E_M^D(\cdot)] \tag{5}$$

In eqn. (5), $[Q_M^D(\cdot)]$ and $[E_M^D(\cdot)]$ are the stochastic processes of possible runoff and total modeling error, respectively, associated with model M, for storm event D. Hence in prediction, the model output of eqn. (5) is not a single outcome, but instead is a stochastic distribution of outcomes, distributed as $[Q_M^D(\cdot)]$. Should $\mathscr{A}$ be some functional operator on the possible outcome (e.g., detention basin volume; peak flow rate; median flow velocity, etc.) of storm event D, then the possible outcome of $\mathscr{A}$ for storm event D is distributed as $[A_M^D]$ where:

$$[A_M^D] = \mathscr{A}[Q_M^D(\cdot)] \tag{6}$$

STOCHASTIC INTEGRAL EQUATION METHOD (SIEM)

As with any hydrologic model, the need for rainfall–runoff data is of paramount importance, and the data is seldom available in sufficient quantity. The distribution, $[E_M^D(\cdot)]$, must then be estimated by using the available set of realizations of the individual stochastic process:

$$\{E_M^i(\cdot)\} = \{Q_g^i(\cdot) - M^i(\cdot);\ i = 1, 2, \ldots\} \tag{7}$$

where $\{E_M^i(\cdot)\}$ is the set of $E_M^i(\cdot)$ outcomes. Assuming elements in $\{E_M^i(\cdot)\}$ to be dependent upon some measure of $Q_g^i(\cdot)$, one may partition $\{E_M^i(\cdot)\}$ into classes of storms such as {mild, major, flood}, or equivalent, should ample storm data be available to develop significant distributions for the resulting subclasses. For simplifying development purposes, $[E_M^D(\cdot)]$ will be based on the entire set $\{E_M^i(\cdot)\}$ with the underlying assumption that all storms are of "equivalent" error. That is, at this point of our analysis it is assumed that the selected rainfall–runoff model performs as well for major storms as it does for minor storms (this assumption is not valid for all rainfall–runoff models).

The second assumption involved is to assume each $E_M^i(\cdot)$ is closely related to some function of precipitation, $F^i(\cdot) = F[P^i(\cdot)]$, where F is an operator which includes parameters, memory of antecedent rainfall, and other factors. Assuming that $E_M^i(t_0)$ depends only on the values of $F^i(\cdot)$ for time $t < t_0$, then $E_M^i(\cdot)$ is expressed as a causal linear filter [for only mild conditions imposed on $F^i(\cdot)$], given by the stochastic integral equation (see Tsokos and Padget, 1974), the uncertainty is focused to $F^i(\cdot)$ realizations by:

$$E_M^i(t_0) = \int_{s=0}^{t_0} F^i(t_0 - s)\, \cong_M^i(s)\, ds \tag{8}$$

where $\cong_M^i(\cdot)$ is the transfer function between $E_M^i(\cdot)$ and $F^i(\cdot)$. Other convenient candidates to be used instead of $F^i(\cdot)$, are the rainfall, $P^i(\cdot)$, and the model itself, $M^i(\cdot)$.

Given a significant set of storm data, an underlying distribution $[\cong_M(\cdot)]$ of the $\{\cong_M^i(\cdot)\}$ may be identified, or the $\{\cong_M^i(\cdot)\}$ may be used directly as in the

case of having a discrete distribution of equally-likely realizations. Using $[\cong_M(\cdot)]$ as notation for both cases of distributions stated above, the predicted response from M for future storm event D is:

$$[Q_M^D(\cdot)] = M^D(\cdot) + [E_M^D(\cdot)] \tag{9}$$

Combining eqns. (8) and (9), the SIEM results in:

$$[Q_M^D(t)] = M^D(t) + \int_{s=0}^{t} F^D(t - s)\,[\cong_M(s)]\,ds \tag{10}$$

and for the functional operator $\mathscr{A}$, on the possible set of outcomes (e.g., detention basin volume, peak flow rate, pipe size, etc.) on storm event D, eqn. (6) is rewritten as:

$$[A_M^D] = \mathscr{A}[Q_M^D(t)] = \mathscr{A}\left(M^D(t) + \int_{s=0}^{t} F^D(t - s)[\cong_M(s)]\,ds\right) \tag{11}$$

Confidence intervals for the A_M^D can now be directly obtained from the distribution $[A_M^D]$. It is noted that $[A_M^D]$ is necessarily a random variable distribution which depends on the model structure, M. In order to demonstrate the utility of the SIEM, and also further develop the mathematical underpinnings of the methodology as applied to rainfall–runoff modeling, a suite of three applications is presented. The applications not only illustrate the approach to including modeling error in criterion variable predictions, but also introduce additional mathematical analysis which is used in subsequent SIEM simplifications.

Application 1

Surface runoff model description and data forms

A link-node model of a 3 mi^2 catchment was developed analogous to the test catchment used by Schilling and Fuchs (1986). The link-node model utilized 81 equally sized subareas, with parameters defined for both the loss rates and open channel flow hydraulics. The diffusion (i.e., zero inertia) routing algorithm of Akan and Yen (1981) was used for the unsteady flow routing (this procedure was also noted as adequate in the subject Schilling and Fuchs (1986) study). For loss rates, a Horton model was used in each subarea, R_j, for the previous area fraction by

$$f_j(t) = \left[f_0 + (f_\infty - f_0)e^{-kt}\right]_j \tag{12}$$

where f_0 is an initial loss rate; f_∞ is the ultimate loss rate; k is a timing coefficient; and j refers to subarea R_j. Values for f_0 varied between 0.8–1.2 in h^{-1}; f_∞ varied between 0.4 and 0.9 in h^{-1}; and k was chosen such that $f_j(t)$ is within 5% of f_∞, at storm time of 30 min. For impervious areas, the infiltration rate was

assumed to be zero. An initial abstraction, Ia, of 0.1 and 0.25 in was used for impervious and pervious areas, respectively. The Ia returns to its maximum value over a duration of 24 h of no rainfall. All subareas are assumed to be 60% pervious.

For the channel links, impervious rectangular channels were assumed throughout the link–node model with channel slopes varying between 0.0040 and 0.0080 (see Hromadka and Yen, 1986), and Manning's friction factors between 0.015 and 0.022. The entire channel system drains to the single stream gage, where critical depth is assumed as the hydraulic control.

Runoff from each subarea was developed by convoluting the resulting effective rainfall for each subarea with the standard SCS unit hydrograph. Subarea time-to-peak values were specified between 30 and 45 min. (Again, the Schilling and Fuchs study indicated that a surface runoff model of this level of detail is relatively insensitive to the subarea runoff modeling technique, and the uncertainty in runoff model predictions due to these type of errors is small in comparison to other factors.)

The "true" runoff hydrograph for storm event i is defined to be given by the above link–node model where precipitation data is supplied on a subarea-by-subarea basis, $P_j^i(\cdot)$. That is, rainfall data are assumed available on a subarea basis. The resulting runoff at the stream gage is defined to be the "true" runoff, $Q_g^i(\cdot)$. Thus, the generation of $Q_g^i(\cdot)$ data is analogous to the Schilling and Fuchs (1986) study where rainfall data is supplied on a subarea-by-subarea basis.

For the model M, the above link–node model [used to develop the "true" runoff, $Q_g^i(\cdot)$] is again utilized, but with rainfall data, $P^i(\cdot)$, supplied only at the midpoint of the catchment; i.e., not on a subarea basis. That is, $P^i(\cdot)$ is the rainfall data "made available" for use in model M to develop the $M^i(\cdot)$ for each storm event, i. It is noted that a rain gage centered with a square-shaped, 3 mi^2 catchment provides much better rainfall data resolution than available for most studies.

For the development of $Q_g^i(\cdot)$, subarea rainfalls, $P_j^i(\cdot)$, are all assumed to be similar in shape to $P^i(\cdot)$, but with random fluctuations in uniform magnitude and timing:

$$P_j^i(t) = \lambda_j^i P^i(t - \theta_j^i) \tag{13}$$

where λ_j^i is a positive constant for subarea j and storm i; and θ_j^i is a constant which provides for a simple translation in time. Both λ_j^i and θ_j^i vary between subareas, and also vary for each storm i. For each subarea, the λ_j^i and θ_j^i are assumed to be distributed as $[\lambda_j]$ and $[\theta_j]$ where:

$$[\lambda_j] = U[0.4d, 1.6d] \tag{14}$$

$$[\theta_j] = U[-0.2d, +0.2d] \tag{15}$$

where eqns. (14) and (15), U is the uniform distribution; d is distance in miles from the catchment centroid; and in eqn. (15), ± 0.2 are in units of h mi^{-1}.

Choices for the constants used in eqn. (14) are motivated by the work of Huff (1970). Rainfall data for $P_i(\cdot)$ were obtained by using 250 storm patterns collected from three rain gage stations located in Los Angeles, California. (The choice of the underlying probability density functions used in eqns. (14) and (15) are for convenience only. Other density functions may be selected at this point of the analysis, and further research is needed as to the best choice used in the model.)

In summary, the "true" runoff, $Q_g^i(\cdot)$, is developed by using the 81-subarea link–node model with subarea rainfall defined by use of eqns. (14) and (15) coupled with $P^i(\cdot)$. The rainfall–runoff model runoff, $M^i(\cdot)$, is developed by using the 81-subarea link–node model, but with rainfall in each subarea defined by just $P^i(\cdot)$. For this ideal test case, the terms $E_s^i(\cdot)$ and $E_d^i(\cdot)$ are negligible in eqn. (3), and $E_M^i(\cdot) = E_r^i(\cdot)$.

Development of the distribution, $[\cong_M(\cdot)]$

Each of the considered 250 storms were run in time sequence by simply joining the continuous records from three rain gages located in Los Angeles, California. For each storm event, i, a new set of parameters $\{\hat{\lambda}_j^i, \hat{\theta}_j^i\}$ was generated for $j = 1, 2, ..., 80$ and the $Q^i(\cdot)$ generated using the link–node model. Similarly, $M^i(\cdot)$ is developed using the link–node model with $P^i(\cdot)$ defined on each subarea. To develop the $E_M^i(\cdot)$ realizations, the error, $E_M^i(\cdot)$, was assumed to be correlated to the effective rainfall estimate, $F^i(\cdot)$, (i.e., a base input), where $F^i(\cdot)$, is defined to be the area-averaged effective rainfall rate for the total catchment. For this example, $F^i(t) = P^i(t) - \bar{f}(t)$, where $\bar{f}(t)$ is a mean loss rate function defined by use of a Horton equation with $f_0 = 1.0, f_\infty = 0.65$, and an initial abstraction of 0.19 in.

The set $\{E_M^i(\cdot)\}$ was determined directly from eqn. (7), and the realization $\cong_M^i(\cdot)$ is determined from eqn. (8) by a least squares matrix solution. Thus for the 250 storm events, 250 realizations of $\{\cong_M^i(\cdot), i = 1, 2, ..., 250\}$ were developed. This set of realizations forms but a small sample of the outcomes from the stochastic process and is therefore subject to the usual sampling error difficulties. However, $\{\cong_M^i(\cdot)\}$ can still be used to make inferences regarding the uncertainty in predictions using the model M.

Functional operator distributions

When predicting the runoff response from the catchment to a hypothetical or design storm event, D, use of the model M provides the single outcome, $M^D(\cdot)$. When coupling the uncertainty information contained in $[\cong_M(\cdot)]$ the predicted outcome is the stochastic process given by eqn. (5). Because of our limited data, $[\cong_M(\cdot)]$ is assumed to be uniformly distributed over $\{\cong_M^i(\cdot), i = 1, 2, ..., 250\}$, and therefore $[Q_M^D(\cdot)]$ is approximately distributed ($\simeq^D$) by the SIEM model (again, the choice of the uniform distribution is made for convenience only and due to limited data for the catchment under study):

$$[Q_M^D(\cdot)](\simeq^D)\ M^D(\cdot) + \left\{ \int_{s=0}^{t} F^D(t - s) \cong i_M(s)\ ds,\ i = 1, 2, ..., 250 \right\} \tag{16}$$

When given the functional operator, $\mathscr{A}$, then the distribution of outcomes, $[A_M^D]$, is approximately distributed as the frequency distribution given by:

$$[A_M^D] \simeq^D \left\{ \mathscr{A}\ (M^D(t) + \int_{s=0}^{t} F^D(t - s) \cong i_M(s)\ ds),\ i = 1, 2, ..., 250 \right\} \tag{17}$$

For example, a storm of key interest in flood control planning is the 100-yr return frequency 24-h design storm described in HEC TD-15 of the U.S. Army Corps of Engineers (1982). Using this 24-h storm pattern and the function F to define $F^D(\cdot)$, the distribution of runoff model outcomes is approximately developed in eqn. (16). For a particular operator, $\mathscr{A}$, eqn. (17) can be used to develop the predicted outcome distribution for this design input, $F^D(\cdot)$.

For the test 3 mi^2 watershed, the peak flow rate distribution is shown in Fig. 1a. The frequency distribution of Fig. 1a is obtained by use of eqn. (17) where the operator $\mathscr{A}$ is simply the identification of the peak flow rate for each element in $\{Q_M^D(\cdot),\ i = 1, 2, ..., 250\}$. That is, a search for max $Q_M^D(\cdot)$ for each sampled realization estimated from eqn. (16).

Another frequently occurring problem is the sizing of the detention basin capacity (volume) with a fixed discharge-rating curve at the basin outlet. For a detention basin with a pipe outlet, the operator $\mathscr{A}$ is now the maximum volume demand on the basin for each element in $\{Q_M^D(\cdot),\ i = 1, 2, ..., 250\}$. A typical frequency distribution of detention basin volumes needed for $F^D(\cdot)$ is shown in Fig. 1b.

From Figs. 1a and 1b, confidence interval estimates can be developed for flood protection given the specified design storm.

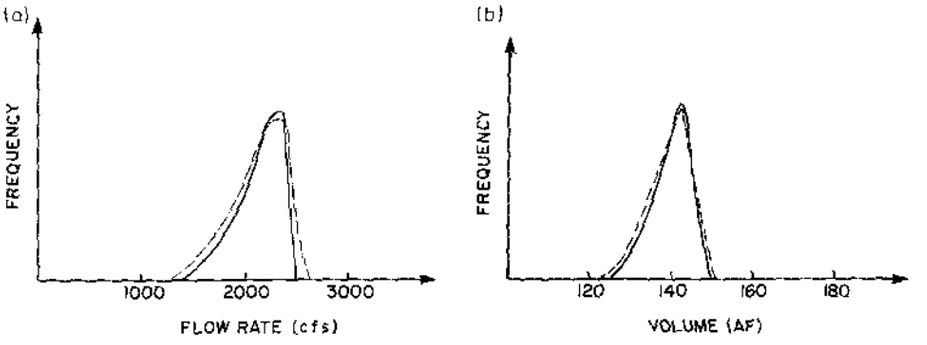


Fig. 1. Frequency distribution of (a) peak flow rates and (b) detention basin maximum volumes for Application 1. Dashed line is true distribution, eqn. (18).

Sensitivity of functional operator distributions to sampling error

The previous application problem utilized a set of 250 storm events with random fluctuations of the λ_j^i and θ_j^i defined on a storm-by-storm basis. Assuming that the distributions of $[\lambda_j]$ and $[\theta_j]$ are correct in eqns. (14) and (15), then the "true" distribution of any functional operator, $\mathscr{A}$, operating on $[Q_M^D(\cdot)]$ is obtained by generating $[A_M^D]$ directly from $M^D(\cdot)$. That is, generate the distribution of possible outcomes of $\mathscr{A}[M^D(\cdot)]$ according the set of distributions, $\{[\lambda_j], [\theta_j], j = 1, 2, ..., 80\}$. The "true" distribution $[A^D]$ is given by (for the "true" catchment response given by the model structure, M):

$$[A^D] = \mathscr{A}[M(P^D(\cdot), \{[\lambda_j], [\theta_j], j = 1, 2, ..., 80\})] \tag{18}$$

It is noted that for mountainous regions, or areas where storm tracks and patterns are known, the random variables used in eqn.(18) may not be uniform, and may include a deterministic component that can be modeled by expected values.

For A being the criterion variable of peak flow rate, or the maximum detention basin volume associated with storm event $P^D(\cdot)$, the corresponding distributions are shown as dashed lines on Figs. 1a and b. From the figures it is seen that for this example problem (which is similar to the test considered in Schilling and Fuchs (1986)), use of eqn. (11) to generate $[A_M^D]$ is a reasonable approximation of the true distribution, $[A^D]$, obtained in eqn. (18).

Application 2

Example development of total error distributions

The previous concepts are now utilized to directly develop the total error distributions, $[E_M(\cdot)]$, for a set of three idealized catchment responses. Besides providing a set of applications, additional notation and concepts are introduced, leading to the introduction of storm classes. It is stressed that the main objective in the following analysis is to develop the background mathematics identifying the source of uncertainty in runoff predictions, and how the uncertainty is propagated through the considered model structures.

Let F be a functional which operates on rainfall data from a single rain gage, $P^i(\cdot)$, to produce a base input for storm i, $F^i(\cdot)$, by:

$$F{:}P^i(\cdot) \rightarrow F^i(\cdot) \tag{19}$$

F may include parameters, memory, and other effects. For example, F may be the operator utilized in a simple phi-index loss model, or F may be the total integration process utilized in the continuous simulation approach to soil moisture accounting.

The catchment R is subdivided into m homogeneous subareas, $R = UR_j$, (see Fig. 2; in which $m = 9$), such that in each R_j, the effective rainfall, $e_j^i(\cdot)$, is assumed given by:

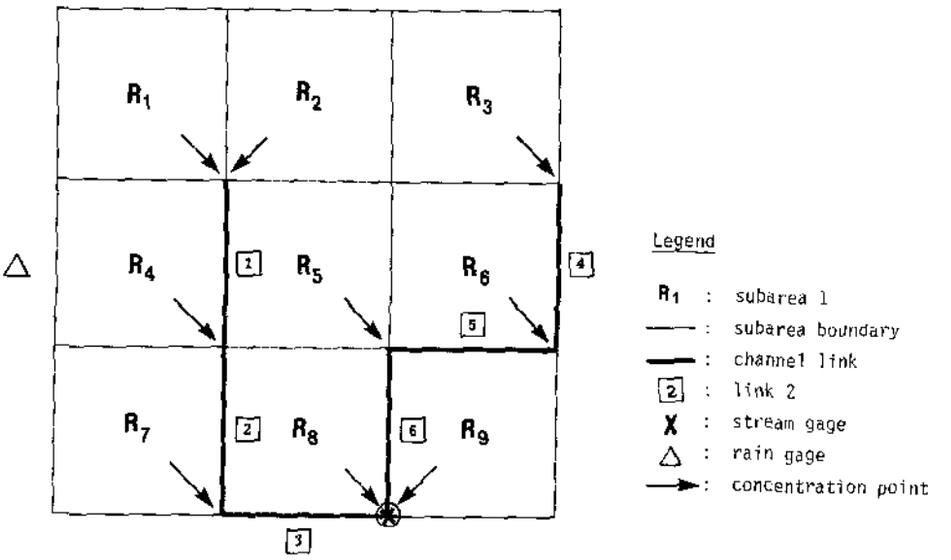


Fig. 2. Example problem watershed schematic (9 subareas).

$$e_j^i(\cdot) = \lambda_j(1 + X_j^i)\, F^i(\cdot) \tag{20}$$

where λ_j is a simple proportion factor; and where X_j^i is a random variable which is constant for storm event i (i.e., the unknown variability in the actual effective rainfall over R is lumped into the single random variable, X_i). The parameter λ_j is defined for subarea R_j and represents the relative runoff response of R_j in comparison to $F^i(\cdot)$, and is a constant for all storms, whereas X_j^i is a sample of the random variable distributed as $[X_j]$, where the set $\{[X_j], j = 1,2,...,m\}$ may be mutually dependent. (Again, further partitioning of the random variables may be necessary such as due to different storm track patterns.)

The subarea runoff is:

$$q_j^i(t) = \int_{s=0}^{t} e_j^i(t - s)\, \phi_j^i(s)\, ds = \int_{s=0}^{t} \lambda_j(1 + X_j^i)\, F^i(t - s)\, \phi_j^i(s)\, ds \tag{21}$$

where $\phi_j^i(\cdot)$ is analogous to the subarea unit hydrograph (UH) for storm i.

At this stage of development, unsteady flow routing along channel links (see Fig. 2) is assumed to be pure translation, (e.g. see Sarikelle et al., 1978). Thus each channel link, L_k, has the constant translation time, T_k. Hence from Fig. 2, the total runoff response at the stream gage for storm event i, $Q_g^i(\cdot)$, is the sum of subarea runoffs, each translated by the sum of associated link travel times:

$$Q_g^i(t) = \sum_{j=1}^{9} q_j^i(t - \tau_j) \tag{22}$$

where $q_j^i(t - \tau_j)$ is defined to be zero for negative arguments; and τ_j is the sum of link travel times (e.g., $\tau_1 = T_1 + T_2 + T_3$; $\tau_6 = T_5 + T_6$; $\tau_9 = 0$).

Then for the above particular assumptions, the stochastic integral equation is developed as:

$$\begin{aligned} Q_g^i(t) &= \sum_{j=1}^{9} \int_{s=0}^{t} \lambda_j (1 + X_j^i)\, F^i(t - s)\, \phi_j^i(s - \tau_j)\, ds \\ &= \int_{s=0}^{t} F^i(t - s) \left[\sum_{j=1}^{9} \lambda_j (1 + X_j^i)\, \phi_j^i(s - \tau_j) \right] ds \end{aligned} \tag{23}$$

In a final form, the runoff response for the given assumptions is:

$$Q_g^i(t) = \int_{s=0}^{t} F^i(t - s) \sum_{j=1}^{9} \lambda_j \phi_j^i(s - \tau_j) ds + \int_{s=0}^{t} F^i(t - s) \sum_{j=1}^{9} \lambda_j X_j^i \phi_j^i(s - \tau_j)\, ds \tag{24}$$

In the above equations, the $\{X_j^i\}$ are unknown to the modeler for any storm event i. However, the work of Schilling and Fuchs (1986) indicates that the variability lumped into these X_j^i is significant and therefore should not be ignored by arbitrarily setting all the $X_j^i = 0$. Thus, $\{X_j^i\}$ are but samples of the individual subarea distributions $\{[X_j], j = 1,2,...,m\}$ where m is the total number of subareas. In design practice, rainfall–runoff models are typically used in an expected value sense which ignores the variabilities in the $[X_j]$. Consequently, the rainfall–runoff model estimate for storm event i is the single outcome, $M^i(\cdot)$, where:

$$M^i(t) = \int_{s=0}^{t} F^i(t - s) \sum_{j} \lambda_j \phi_j^i(s - \tau_j) ds \tag{25}$$

where each $[X_j] \equiv 0$ in eqn. (24).

Then, $Q_g^i(\cdot) = M^i(\cdot) + E_M^i(\cdot)$ where:

$$E_M^i(t) = \int_{s=0}^{t} F^i(t - s)\, \cong_M^i(s)\, ds \tag{26}$$

where $\cong_M^i(s)$ follows directly from eqns. (24) and (25).

Should the variability in $\phi_j^i(\cdot)$ be minor (with respect to i) such $\phi_j^i(\cdot) \cong \phi_j(\cdot)$ for all i (see Schilling and Fuchs, 1986), then the above equations can be further simplified as:

$$M^i(t) = \int_{s=0}^{t} F^i(t - s)\, \eta(s)\, ds \tag{27}$$

where $\eta(s) = \sum_{j=1}^{9} \lambda_j \phi_j(s - \tau_j)$. Additionally, the distribution of the stochastic process $[\cong_M(\cdot)]$ is readily determined for this simple example by:

$$[\cong_M(\cdot)] = \sum_{j=1}^{9} [X_j]\lambda_j \phi_j(s - \tau_j) \tag{28}$$

where $[\cong_M(\cdot)]$ is directly correlated to the nine random variables, $\{X_j; j = 1,2,...,9\}$. It is again noted that the random variables, X_j, may be all mutually dependent.

In this example problem, the distribution of $[\cong_M(\cdot)]$ is directly evaluated due to the particular simplifying assumptions and the knowledge of $\{[X_j]\}$. In practice, the several distributions $\{[X_j]\}$ are unknown to the modeler, but the summed net effect of the $\{[X_j]\}$ can be estimated by the development of $[\cong_M(\cdot)]$ from the several samples, $E_M^i(\cdot)$.

For the important problem of prediction, the SIEM provides an estimate of runoff as the distribution $[Q_M^D(\cdot)]$ where $[Q_M^D(\cdot)] = M^D(\cdot) + [E_M^D(\cdot)]$, and M refers to the particular model type used.

For this example problem, the SIEM formulation is:

$$[Q_M^D(t)] = \int_{s=0}^{t} F^D(t - s)\, \eta(s)\, ds + \int_{s=0}^{t} F^D(t - s)\, [\cong_M(s)]\, ds \tag{29}$$

where the error distribution, $[E_M^D(\cdot)]$, is assumed to be correlated to the base input, $F^D(\cdot)$, as provided in eqns. (26) and (28).

Multilinear routing and storm classes

The above example problem is now extended to include the additional assumption that the channel link travel times are strongly correlated to some set of characteristic descriptions of the hydrograph being routed, such as some weighted mean flow rate of the associated hydrograph. For example, the widely used Convex Routing technique (Mockus, 1972) often utilized the 85-percentile of all flows in excess of one-half of the peak flow rate as a statistic used to estimate the routing parameters. But by the previous example problem definition of $e_j^i(t)$, all runoff hydrographs in the link–node channel system would be highly correlated to an equivalent weighting of the base input, $F^i(\cdot)$. Hence, storm classes, $[\![\xi_z]\!]$, of "*equivalent*" $F^i(\cdot)$ *realizations could be defined* where all elements of $[\![\xi_z]\!]$ have the same characteristic parameter set, $\mathscr{C}[F^i(\cdot)]$:

$$[\![\xi_z]\!] = \{F^i(\cdot) | \mathscr{C}[F^i(\cdot)] = z\} \tag{30}$$

and for all $F^i(\cdot)\ \varepsilon [\![\xi_z]\!]$, each representive channel link travel time, T_k, is identical, that is, $T_k = T_{k_z}$ for all $F^i(\cdot)\ \varepsilon [\![\xi_z]\!]$. In the above definition of storm class, z is a characteristic parameter set in vector form.

This extension of the translation routing algorithm to a multilinear formulation (involving a set of link translation times) modifies the previous runoff equations to be:

$$M^i(t) = \int_{s=0}^{t} F^i(t - s) \sum_j \lambda_j \phi_j(s - \tau_j^z) \mathrm{d}s = \int_{s=0}^{t} F^i(t - s)\eta_z(s)\, \mathrm{d}s;\ F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{31}$$

where:

$$\eta_z(\cdot) = \sum_j \lambda_j \phi_j(s - \tau_j^z)$$

and:

$$E_M^i(t) = \int_{s=0}^{t} F^i(t - s) \cong_{M_z^i}(s)\, \mathrm{d}s;\ F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{32}$$

The structure of the new set of equations motivates an obvious extension of the definition of the subarea UH, the subarea λ_j proportion factor, and the subarea random variable distribution $[X_j]$, to all be also defined on the storm class basis of $[\![\xi_z]\!]$. Thus in final form:

$$M^i(t) = \int_{s=0}^{t} F^i(t - s) \sum_j \lambda_j^z \phi_j^z\ (s - \tau_j^z)\, \mathrm{d}s = \int_{s=0}^{t} F^i(t - s)\eta_z(s)\, \mathrm{d}s;$$
$$F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{33}$$

$$[\cong_{M_Z}(\cdot)] = \sum_j [X_j^z]\lambda_j^z \phi_j^z(s - \tau_j^z);\ F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{34}$$

And in prediction, the SIEM model is:

$$[Q_M^D(t)] = M^D(t) + [E_M^D(t)];\ F^D(\cdot)\varepsilon[\![\xi_D]\!] \tag{35}$$

where:

$$[E_M^D(t)] = \int_{s=0}^{t} F^D(t - s)\, [\cong_{M_D}(s)]\, \mathrm{d}s;\ F^D(\cdot)\varepsilon[\![\xi_D]\!] \tag{36}$$

Multilinear hydrologic routing

The cited Schilling and Fuchs study concluded that in their model structure, a fully dynamic unsteady flow routing algorithm employed for all the channel links could be simplified to a hydrologic-type storage routing model without a significant loss of accuracy in overall modeling accuracy (e.g., Muskingum).

The Muskingum and Convex hydrologic storage routing methods are both widely used techniques. For each channel link, either hydrologic method can be calibrated should flow data be available (e.g., Hromadka et al., 1987). It is tacitly assumed in our analysis/development that the parameters used in either of the considered storage routing methods can be locally calibrated for each link used in the runoff model (although only one stream gage and one rain gage are available to the modeler in our example problem).

Because $0 < C < 1$, $(1 - C)^n$ decreases in significance as the exponent n increases, and a finite number of terms may be adequate in eqn. (42) for all runoff hydrographs with the equivalent routing parameter, C.

Thus for all runoff hydrographs with similar C values, eqn. (37) reduces to a convolution method with a constant transfer function.

The Muskingum technique of eqn. (40) is seen to be similarly rewritten as:

$$O_{T-dT} = (k_3)\, I_{T+dT} + (k_1 + k_2k_3)\, I_T + k_2(k_1 + k_2k_3)\, I_{T-dT} + k_2^2(k_1 + k_2k_3)\, I_{T-2dT} + k_2^3(k_1 + k_2k_3)\, I_{T-3dT} + \ldots . \quad (43)$$

and, for all runoff hydrographs with similar routing parameters, the Muskingham technique is again a convolution method with a constant transfer function. Additionally, the coefficients' influence decreases with increasing time span to time T.

From the above, either of the considered hydrologic routing methods are actually linear analogs when given constant routing parameters. In the use of storm classes of the base input, $F^i(\cdot)$, noted by $[\![\xi_Z]\!]$, all the links in M are assigned the identical set of routing parameters (on a link basis).

Once again, the eqns. (33)–(36) apply except that the previous multilinear translation-routing model is now replaced by either of the above hydrologic storage routing methods with constant routing parameters on a storm class basis (i.e., a multilinear hydrologic flow routing model).

Example

An 18-subarea catchment model is developed with channel links, shown schematically in Fig. 3. Also shown in the figure are subarea area, runoff proportion, λ_j, and the time-to-peak T_p (h) used with the standard SCS unit hydrograph (all of this data is assumed to apply for a specific storm class). The SCS unit hydrograph and a runoff coefficient are specified for each subarea. Convex channel routing is used with constant C-values listed in Fig. 3. Hence, only one storm class is being defined for all base inputs.

To demonstrate the model's linearity, one rainfall event $P^1(\cdot)$, and the link–node model-produced $O^1(\cdot)$ was used to determine the equivalent unit hydrograph (Fig. 4) and overall runoff coefficient (0.75). Using this new UH model data, the single UH model equates to the 18-subarea link–node model, for all storms, $P^i(\cdot)$.

A multilinear surface runoff model

Application 2 demonstrated the direct development of the total error distribution, $[E_M(\cdot)]$, for a set of three particular model structures. In this section, the results of application 2 are generalized to include a wide range of possibilities.

As before, let F be a functional defined on the (single) rain gage data, $F: P^i(\cdot) \rightarrow F^i(\cdot)$. The catchment R is subdivided into m subareas, $\{R_j, j = 1,2, \ldots, m\}$

A simple hydrologic storage routing model in the Convex method which results (in a hydrologic sense) in both translation and peak flow attenuation of the runoff hydrograph due to channel storage effects. The governing relationship used in this approach is (Hromadka et al., 1987):

$$O_{T+dT} = (1 - C)O_T + CI_T \tag{37}$$

where: I_T = hydrograph inflow at time T; O_T = channel outflow at time T; O_{T+dT} = channel outflow at time $T + dT$; and C = a routing coefficient (where C is between 0 and 1). Rearranging eqn. (37) gives the explicit statement:

$$O_{T+dT} = O_T + C(I_T - O_T) \tag{38}$$

and solving for C yields:

$$C = (O_{T+dT} - O_T)/(I_T - O_T) \tag{39}$$

The Muskingum channel routing method is an alternative to the Convex routing method. It is somewhat more flexible than the Convex method since the value of O_{T+dT} depends on I_{T+dT}, I_T, and O_T, and not just I_T and O_T. As an alternative to eqn. (37), the Muskingum method uses the following equation for routing:

$$O_{T+dT} = C_0 I_{T+dT} + C_1 I_T + C_2 O_T \tag{40}$$

where the coefficients are given by:

$$C_0 = -(Kx - 0.5dT)/(K - Kx + 0.5dT) \tag{41a}$$

$$C_1 = (Kx + 0.5dT)/(K - Kx + 0.5dT) \tag{41b}$$

$$C_2 = (K - Kx - 0.5dT)/(K - Kx + 0.5dT) \tag{41c}$$

The sum of the three coefficients C_0, C_1, and C_2 equals 1.0. In eqns. (40) and (41), K and x are the routing parameters. While K and x can be estimated from hydrographs measured at the upstream and downstream channel sections, the value of K is taken to be equal to the reach travel time. Methods of assigning a value of x are less respected. Some suggest that a value for x of 0.2 is reasonable, while others suggest 0.4; hence, another random variable is involved. As the value of x decreases from the upper limit of 0.5, the peak of the downstream hydrograph becomes more attenuated. To ensure non-negative ordinates for O_{T+dT} the following limits should be respected:

$$x \leqslant dT/2K \leqslant (1 - x); \qquad x \leqslant 0.5 \tag{41}$$

The hydrologic storage routing methods considered in the above can be formulated into a convolution technique, where the transfer functions are parameter dependent.

The Convex technique transfer functions are derived from eqn.(37) as:

$$O_{T+dT} = CI_T + (1 - C)\,CI_{T-dT} + (1 - C)^2\,CI_{T-2dT} + \ldots . \tag{42}$$

where C is the routing parameter.

linked together by unsteady flow routing models. For simplicity, R is assumed to be free-draining, without dams nor detention basins. The entire link-node model drains freely to the single stream gage where the data, $Q_g^i(\cdot)$, is measured. The problem is to predict the runoff response at the stream gage for a hypothetical storm event, $P^D(\cdot)$, defined at the single available rain gage. Baseflow considerations are assumed to be minor.

Each subarea's effective rainfall (rainfall less losses), $e_j^i(\cdot)$, is now defined to be the sum of proportions of $F^i(\cdot)$-translates by:

$$e_j^i(t) = \sum_k \lambda_{jk}(1 + X_{jk}^i)F^i(t - \theta_{jk}^i); \qquad F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{44}$$

where X_{jk}^i and θ_{jk}^i are samples of the random variables distributed as $[X_{jk}]$ and $[\theta_{jk}]$, respectively. In the above equation and all equations that follow, it is assumed that a storm class system is defined, $[\![\xi_z]\!]$, such that for $F^i(\cdot)\varepsilon[\![\xi_z]\!]$ all parameters and probabilistic distributions are uniquely defined, and there is no loss in understanding by omitting the additional notation needed to indicate the storm class (see Application 2).

The subarea runoff is:

$$q_j^i(t) = \int_{s=0}^{t} \sum_k \lambda_{jk}(1 + X_{jk}^i)\, F^i(t - \theta_{jk}^i - s)\, \phi_j(s)\mathrm{d}s; \qquad F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{45}$$

or a simpler form:

$$q_j^i(t) = \int_{s=0}^{t} F^i(t - s) \sum_k \lambda_{jk}(1 + X_{jk}^i)\, \phi_j(s - \theta_{jk}^i)\mathrm{d}s; \qquad F^i(\cdot)\varepsilon[\![\xi_z]\!] \tag{46}$$

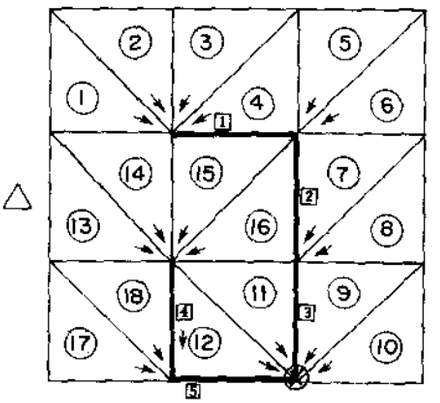
It is assumed that the unsteady flow channel routing effects are related to the magnitude of runoff in each channel link, which is additionally correlated to the magnitude of the base input realization, $F^i(\cdot)$. On a storm class basis, each channel link is assumed to respond linearly in that (e.g., Doyle et al., 1983):

$$O_1^i(t) = \sum_l a_l I_1^i(t - \alpha_l) \tag{47}$$

where $O_1^i(t)$ and $I_1^i(t)$ are the outflow and inflow hydrographs for link 1, and storm event i; and $\{a_l\}$ and $\{\alpha_l\}$ are constants which are defined on a storm class basis which is also used for the base input, $F^i(\cdot)$. Thus, the channel link flow routing algorithm is multilinear with routing parameters defined according to the storm class, $[\![\xi_z]\!]$ (see Becker and Kundzewicz, 1987, for an analogy based on multilinear approximation of nonlinear routing).

Should the above outflow hydrograph, $O_1(t)$, now be routed through another link (number 2), then $I_2(t) = O_1(t)$ and from the above:

$$O_2(t) = \sum_{k_2=1}^{n_2} a_{k_2} I_2(t - \alpha_{k_2})$$



Subarea #	Area Acres	λ_j	T_p (hours)
1	400	.94	.7
2	400	.80	.85
3	600	.90	.4
4	600	.70	.6
5	550	.80	.8
6	550	.40	.9
7	350	.94	.6
8	350	.82	.7
9	450	.80	.7
10	450	.60	.8
11	650	.95	.4
12	650	.85	.6
13	500	.94	.5
14	500	.84	.7
15	360	.75	.9
16	360	.25	.9
17	640	.80	.7
18	640	.40	.9

Channel Information:

Length	= 1000 ft
Base	= 5 ft
n	= 0.015
So	= .0050

Link	C
1	0.7
2	0.8
3	0.9
4	0.85
5	0.9

Total Area = 9000 Acres

Legend

- subarea boundary
- link
- ⊗ stream gage
- △ rain gage
- ① Subarea 1
- [1] Link 1

Fig. 3. Example problem watershed schematic (18 subareas).

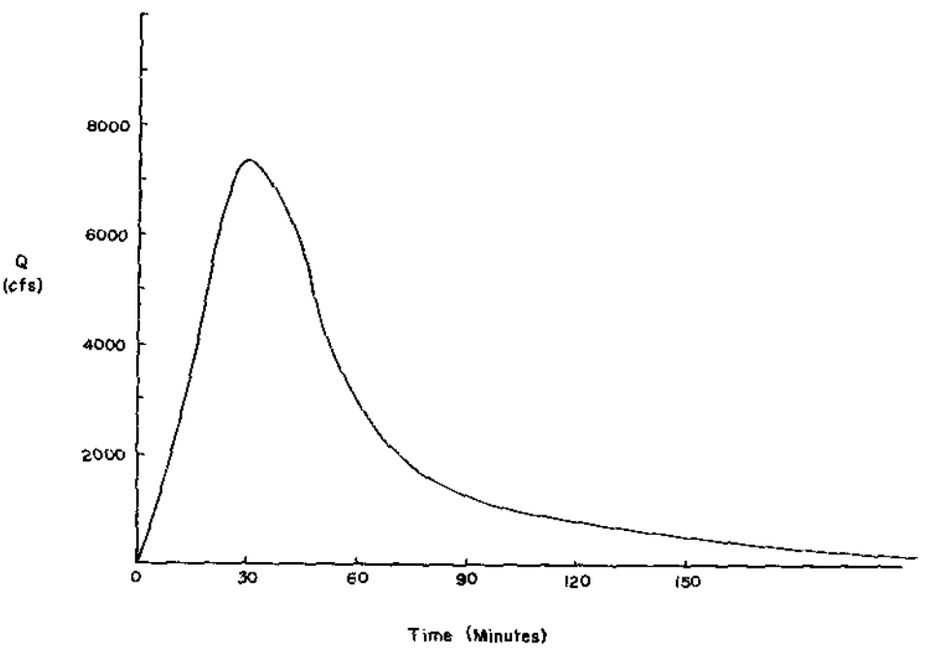


Fig. 4. Unit hydrograph equivalent to example problem link–node model (3-min interval).

input is used.) Consequently, the above multilinear surface runoff model structure represents a highly detailed and distributed parameter model of the surface runoff process which not only can be used to represent the catchment runoff response itself, but also can be used to approximate the response of most other hydrologic modeling approaches (e.g., Schilling and Fuchs, 1986). But of major concern is the effect on the runoff prediction (at the stream gage) from the model M, due to the randomness exhibited by the mutually dependent set of random variables, $\{X_{jk}, \theta_{jk}\}$. Hence for any operator, $\mathscr{A}$, on the predicted runoff response of eqn. (52), the outcome for $\mathscr{A}$ for storm event $P^D(\cdot)$ is the distribution $[A_M^D]$, where for all model parameters defined:

$$[A_M^D] = \mathscr{A}[Q_g^D(\cdot)] = \mathscr{A}(\{[X_{jk}], [\theta_{jk}]\}) \tag{53}$$

The multilinear surface runoff model will be used in the following analysis of uncertainty. (It is noted that although only the simple multilinear model of eqn.(52) is carried forward, use of other rainfall–runoff models in the SIEM formulation will result in analogous results in developing the necessary stochastic distributions.)

Application 3

Application to rainfall–runoff data

The multilinear surface runoff model provides a convenient approximation of other surface runoff model's responses. From eqn.(52), the SIEM gives:

$$[Q_g^D(t)] = \int_{s=0}^{t} F^D(t - s)[\eta(s)]ds; \qquad F^D(\cdot)\varepsilon[\![\xi_D]\!] \tag{54}$$

where now $[\eta(s)]$ is the distribution of the stochastic process representing the random variations from the set of mutually dependent random variables, $\{X_{jk}, \theta_{jk}\}$, defined on a storm class basis. The parameters utilized in the $[\eta(\cdot)]$ used in eqn. (54) are $\boldsymbol{P} \equiv \{a_{\langle k\rangle}, \alpha_{\langle k\rangle}, \lambda_{jk}, \phi_j(\cdot)\}$. If $\boldsymbol{P}$ is defined correctly (for the associated storm class), the mean of the random variables are $E[X_{jk}] = 0$ and $E[\theta_{jk}] = \theta_{jk}$; otherwise, errors in the estimates of $\boldsymbol{P}$ are transferred to the expected value of the random variables X_{jk} and θ_{jk}. Then in prediction:

$$E[Q_g^D(t)] = \int_{s=0}^{t} F^D(t - s)E[\eta(s)]ds; \qquad F^D(\cdot)\varepsilon[\![\xi_D]\!] \tag{55}$$

which is a multilinear version of the classic unit hydrograph method.

Then the model M structure of eqn. (51) is given by:

$$M^D(\cdot) = E[Q_g^D(\cdot)] \tag{56}$$

with the total error distribution given by:

$$[E^D(\cdot)] = [Q_g^D(\cdot)] - E[Q_g^d(\cdot)] \tag{57}$$

where all equations are defined on a storm class basis. Given sufficient rain-

$$= \sum_{k_2=1}^{n_2} a_{k_2} \sum_{k_1=1}^{n_1} a_{k_1} I_1(t - \alpha_{k_1} - \alpha_{k_2}) \tag{48}$$

For L links, each with their own respective stream gage routing data, the above linear routing technique results in the outflow hydrograph for link number L, $O_L(t)$, being given by:

$$O_L(t) = \sum_{k_L=1}^{n_L} a_{k_L} \sum_{k_{L-1}=1}^{n_{L-1}} a_{k_{L-1}} \cdots \sum_{k_2=1}^{n_2} a_{k_2} \sum_{k_1=1}^{n_1} a_{k_1} \times I_1(t - \alpha_{k_1} - \alpha_{k_2} - \cdots - \alpha_{k_{L-1}} - \alpha_{k_L}) \tag{49}$$

Using vector notation, the above $O_L(t)$ is written as:

$$O_L(t) = \sum_{\langle k \rangle} a_{\langle k \rangle} I_1(t - \alpha_{\langle k \rangle}) \tag{50}$$

For subarea R_j, the runoff hydrograph for storm i, $q_j^i(\cdot)$, flows through L_j links before arriving at the stream gage and contributing to the total modeled runoff hydrograph, $M^i(\cdot)$. All of the parameters $a^i_{\langle k \rangle}$ and $\alpha^i_{\langle k \rangle}$ are constants on a storm class basis. Consequently from the linearity of the routing technique, the m-subarea link node model is given by the sum of the m, $q_j^i(\cdot)$ contributions:

$$M^i(t) = \sum_{j=1}^{m} \sum_{\langle k \rangle_j} a^i_{\langle k \rangle_j} q_j^i(t - \alpha^i_{\langle k \rangle_j}) \tag{51}$$

In final form, the predicted runoff response for storm event D is the SIEM formulation:

$$[Q_g^D(t)] = \int_{s=0}^{t} F^D(t - s) \left(\sum_{j=1}^{m} \sum_{\langle k \rangle_j} a^i_{\langle k \rangle_j} \sum \lambda_{jk}(1 + [X_{jk}]) \phi_j(s - [\theta_{jk}] - \alpha^i_{\langle k \rangle_j}) \right) ds;\ F^D(\cdot) \varepsilon [\![\xi_D]\!] \tag{52}$$

Given $F^i(\cdot) \varepsilon [\![\xi_z]\!]$, all subarea runoff parameters $\{\lambda_{jk}, \phi_j(\cdot)\}$ and distributions $\{[X_{jk}], [\theta_{jk}]\}$ are uniquely defined for $j = 1, 2, \ldots, m$; and all link routing parameters $\{a_l, \alpha_l\}$ are also uniquely defined. Then the entire link–node model is linear on a storm class basis and once more eqns. (33)–(36) apply without modification.

This last result is significant because except for the few reported fully dynamic routing surface runoff models, all hydrologic models in use today ignore backwater effects in the channel routing algorithms and in most cases can be adequately approximated by the multilinear routing algorithm presented above (e.g., Becker and Kundzewicz, 1987). Additionally, all surface runoff models utilize some average effective rainfall estimation algorithm, F, which provides the base unit, $F^i(\cdot)$, for storm event i, which can subsequently be used to approximate the subarea effective rainfall on a storm class basis. (When the catchment is highly nonhomogeneous, then more than one base

of central Los Angeles, California, Table 2 summarizes the considered basin characteristics. Table 3 lists the available rain gage sites and the storm dates of events used in the analysis. All storms considered in Table 3 are assumed to be elements of the same storm class considered important for flood control.

TABLE 3a

Precipitation gauges used in Los Angeles County flood reconstitutions

Streamgauge location	Storm reconstitution	LACFCD raingauge No.*
Alhambra Wash near	Feb 78	191, 303, 1114B
Klingerman Street	Mar 78	191, 303, 1114
	Feb 80	191B, 235, 280C, 1014
Compton Creek near	Feb 78	116, 291
Greenleaf Drive	Mar 78	116, 291
	Feb 80	116, 291, 716
Limekiln Creek above	Feb 78	57A, 446
Aliso Creek	Feb 80	259, 446
San Jose Creek Channel	Feb 78	92, 1078, 1088X
above Workman Mill Rd.	Feb 80	96CE, 347E, 1088
Sepulveda Dam (inflow)	Feb 78	57A, 292DE, 446, 735H
	Mar 78	57A, 435, 762
	Feb 80	292, 446, 735
Verdugo Wash at Estelle Ave.	Feb 78	280C, 373C, 498, 758

No.*	Station name	Lat.	Long.	Elev.	Type
L057A	Camp Hi Hill (OPIDS)	34-15-18	118-05-41	4240	SR
L0092	Claremont--Pomona College	34-05-48	117-42-33	1185	SR
L0096CE	Puddingstone Dam	34-05-31	117-48-24	1030	SR
L0116	Inglewood Fire Station	33-47-53	118-21-22	153	SR
L0191(B)	Los Angeles Alcazar	34-03-46	118-11-54	400	SR
L0235	Henninger Flats	43-11-38	118-05-17	2550	SR
L0259	Chatsworth–Twin Lakes	34-16-43	118-35-41	1275	SR
L0280C	Sacred Heart Academy	34-10-54	118-11-08	1600	R
L0291	Los Angeles–96th & Central	33-56-56	118-15-17	121	R
L0292(DE)	Encino Reservoir	34-08-56	118-30-57	1075	SR
L0303	Pasadena–Cal Tech	34-08-14	118-07-25	800	SR
L0347E	Baldwin Park–Exp. Station	34-05-56	117-57-40	384	SR
L0373C	Briggs Terrace	34-14-17	118-13-27	2200	SR
L0435	Monte Nido	34-04-41	118-41-35	600	SR
L0446	Aliso Canyon–Oat Canyon	34-18-53	118-33-25	2367	SR
L0498	Angeles Crest Hwy–Drk Cny Tr	34-15-21	118-11-45	2800	R
L0716	Los Angeles–Ducommun Street	34-03-09	118-14-13	306	SR
L0735(H)	Bell Canyon	34-11-40	118-39-23	895	R
L0758	Griffith Park -Lower Spr Cyn	34-08-02	118-17-27	600	R
L0762	Upperstone Canyon	34-07-27	118-27-15	943	R
L1014	Rio Hondo Spreading	33-59-57	118-06-04	170	SR
L1078	Covina–Griffith	34-04-10	117-50-47	975	SR
L1088(X)	La Habra Hgts–Mut Water Co	33-56-55	117-57-51	445	SR
L1114(B)	Whittier Narrows Dam	34-01-29	118-05-02	239	SR

S = standard 8" raingauge (non recording); R = recording raingauge.

fall–runoff data, the total error distribution can be approximately developed by use of eqn. (57). Should another surface runoff model structure be used, then $E[Q_g^D(\cdot)]$ is replaced by the alternative model, and the set of realizations of $[E^D(\cdot)]$ is obtained from eqn. (57).

In this application, the base input functional F: $P^i(\cdot) \rightarrow F^i(\cdot)$ is specified as:

$$F\text{: } P^i(\cdot) \rightarrow \lambda P^{(i)} \tag{58}$$

where λ is a simple runoff coefficient (e.g., Scully and Bender, 1969; Schilling and Fuchs, 1986). The stochastic integral equation is:

$$Q_g^i(t) = \int_{s=0}^{t} P^i(t - s)\eta^i(s)\, ds \tag{59}$$

For this loss function, storm classes of base input are necessarily just storm classes of rainfall without any subclassification due to prior rainfall and soil wetness. In this application, storm classes are defined according to the 85-percentile of the rainfall intensity (z) in excess of one-half of the peak 5-min mean intensity. Storm classes are then assembled according to the characteristic z-value on 0.5-in increments. (Should another loss function be used, such as a continuous soil-moisture accounting algorithm, then F would be defined as the resulting effective rainfall.)

Because of the usual sparcity of rainfall–runoff data, several catchments are considered in order to regionalize the statistical results. For the study location

TABLE 2

Basin characteristics of watersheds used in flood reconstitutions

	Drainage area (mi^2)	L (mi)	L_{ca} (mi)	S (ft mi^{-1})	Percent impervious	
					Total	Effective
Los Angeles County watersheds						
Alhambra Wash near Kingerman	15.2	8.20	4.60	79	60	32
Compton Creek near Greenleaf Drive	22.6	11.0	4.00	11.0	65	45
Limekiln Creek above Aliso Creek	10.3	7.4	3.6	294.0	25	25
San Jose Channel above Workman Mill Road	83.4	23.0	8.5	60.0	18	18
Sepulveda Dam (inflow)	152.0	19.0	9.0	143.0	35	24
Verdugo Wash at Estelle Ave.	26.8	11.4	5.70	310.0	25	20
Orange County watersheds						
El Modena Irvine Channel at Myford Road	11.9	6.34	2.69	52	40	40
Santa Ana–Delhi Channel at Irvine Ave.	17.6	8.71	4.17	16.0	40	40
Westminster Channel at Beach Blvd.	6.7	5.65	1.39	13	40	40

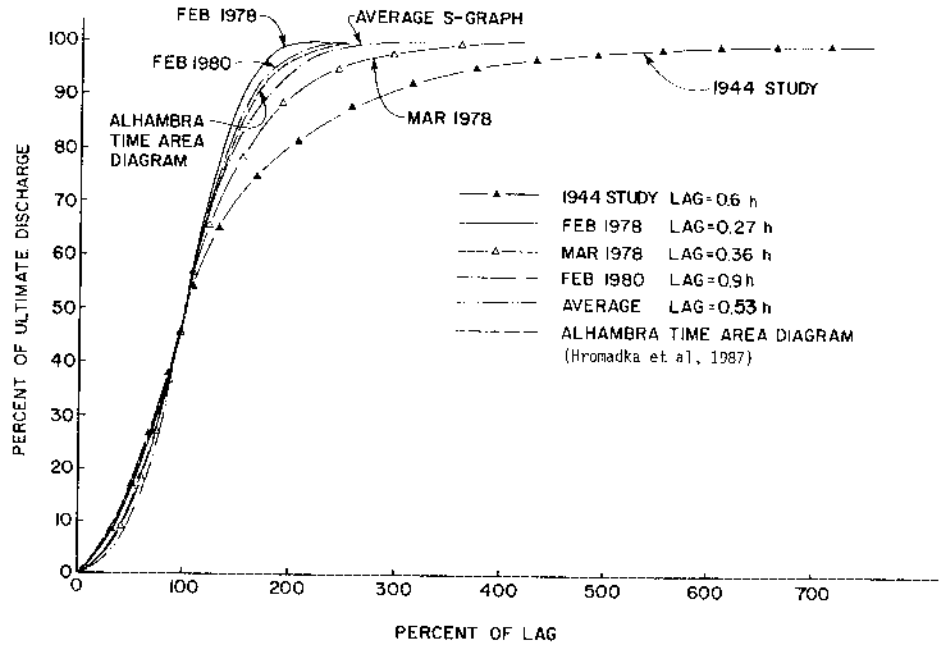


Fig. 5. Alhambra Wash best-fit S-graphs.

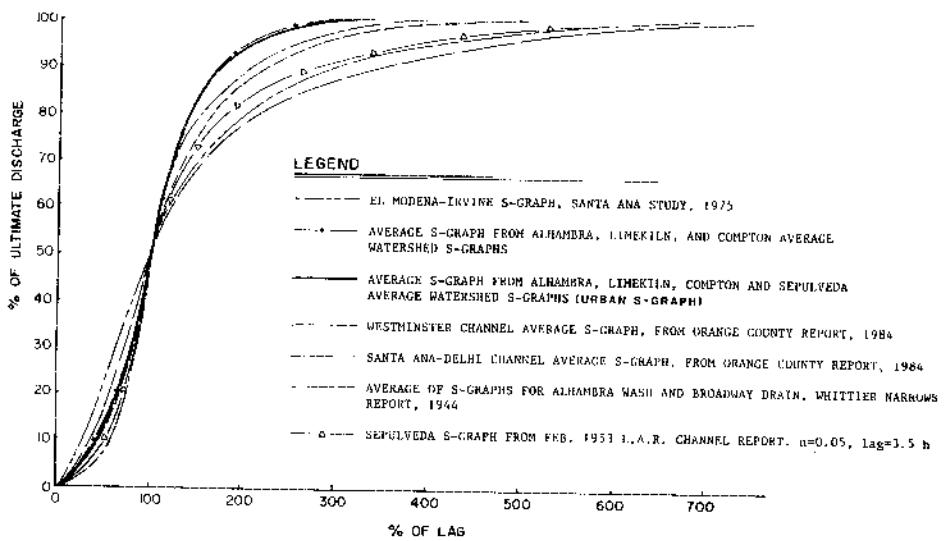


Fig. 6. Catchment averaged S-graphs.

TABLE 3b

Precipitation gauges used in Orange County flood reconstitutions

Streamgauge location	Storm reconstitution	OCEMA raingauge No.*
El Modena–Irvine Channel at Myford Rd.	Feb 69	61,121
Santa Ana–Delhi Channel at Irvine Ave.	Mar 79 Jan 80 Nov 78 Feb 80 Jan 79 Mar 80 Mar 79 Nov 82	121, 165
Westminster Channel at Beach Blvd.	Dec 74 Mar 78 Feb 80	162, 184

No.*	Station name	Lat.	Long.	Elev.	Type
OC-61	Tustin–Irvine Ranch	33-43-46	117-46-58	118	S
OC-121	OCEMA–Santa Ana	33-45-94	117-52-11	180	R
OC-162	Westminster	33-45-08	117-59-17	39	SR
OC-165	Costa Mesa	33-40-07	117-53-35	56	SR
OC-184	Garden Grove City Hall	34-46-35	117-55-49	121	S

S = standard 8" raingauge (non recording); R = recording raingauge.

For each storm event and catchment, the rainfall–runoff data are used to directly develop the $\{\eta^i(\cdot)\}$. On a catchment basis, the several $\eta^i(\cdot)$ are pointwise averaged together to determine an estimate for $E[\eta(\cdot)]$ for the prescribed storm class. Note that for this simple loss functional, the probabilistic effects of prior rainfall are transferred to the $\eta(\cdot)$ realizations.

Summation graphs of the $\{\eta^i(\cdot)\}$ indicated that normalizing could be performed by plotting mass along the y-axis from 0 to 100% of mass; and the x-axis as time with respect to the parameter "lag" where 100% of lag equals the time at 50% of total mass. Plots of normalized $\eta^i(\cdot)$ realizations for Alhambra Wash for several storms are shown in Fig. 5, and plots of the estimates of $E[\eta(\cdot)]$ for the several catchments are shown in Fig. 6. (Such normalization techniques are commonly used to relate unit hydrographs to catchment characteristics; see Hromadka et al., 1987).

A comparison of Figs. 5 and 6 shows that the variation in $E[\eta(\cdot)]$ among the several considered catchments is of a magnitude similar to the variation in the $\eta^i(\cdot)$ for Alhambra Wash alone. Therefore in order to regionalize the total error distributions, the variations among all the catchment $E[\eta(\cdot)]$ are assembled together as one regionalized distribution.

To further define the data, a scaling parameter Y is defined by plotting each $\eta^i(\cdot)$ realization on Fig. 7 and averaging the upper and lower reading for Y. The regionalized distribution for the parameter Y is shown in Fig. 8. That is, in the rescaling of reconstituted unit hydrographs into S-graph form, the variations between unit hydrographs are seen to be related to the variations in the scaling

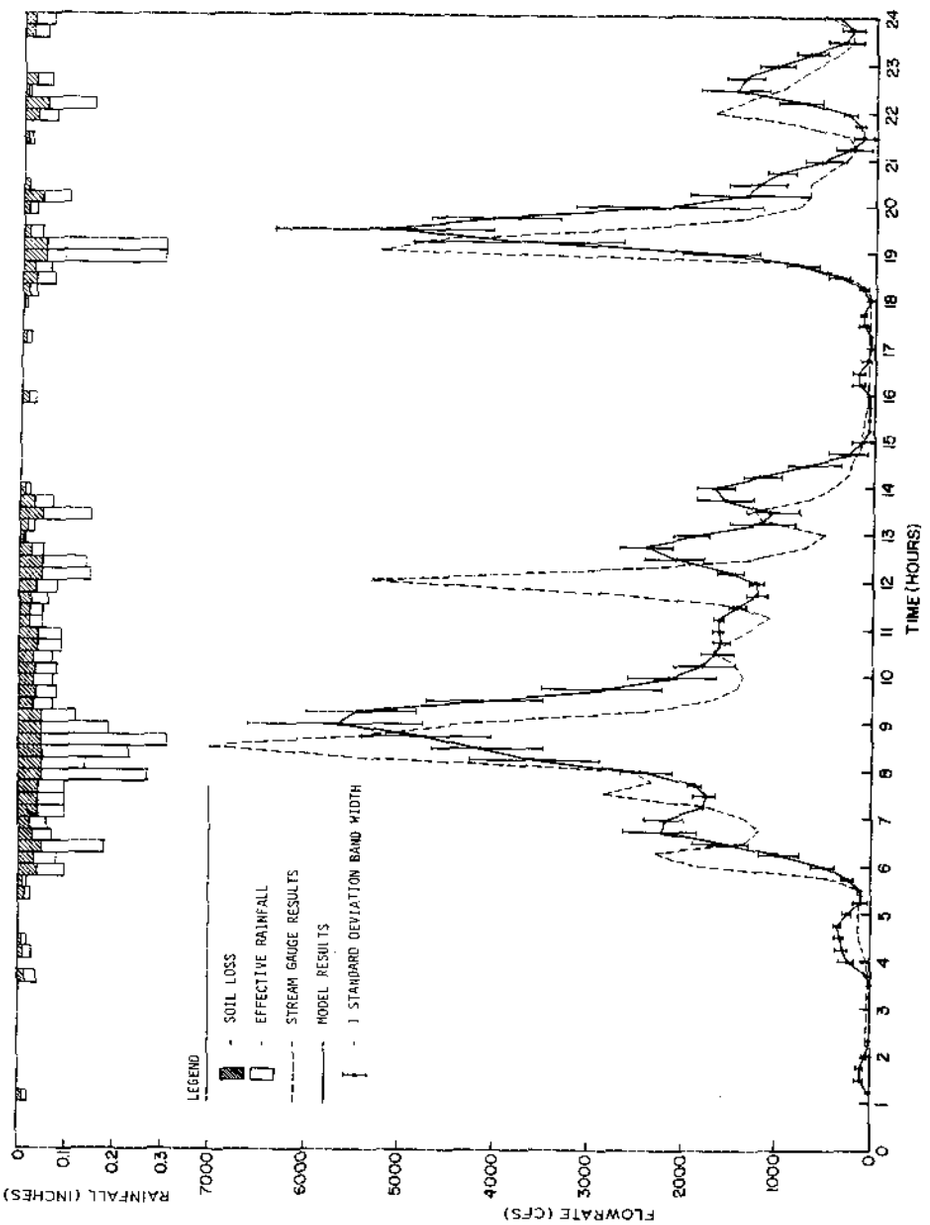


Fig. 9. Monte Carlo analysis of the March 1, 1983 storm for the Alhambra Wash watershed.

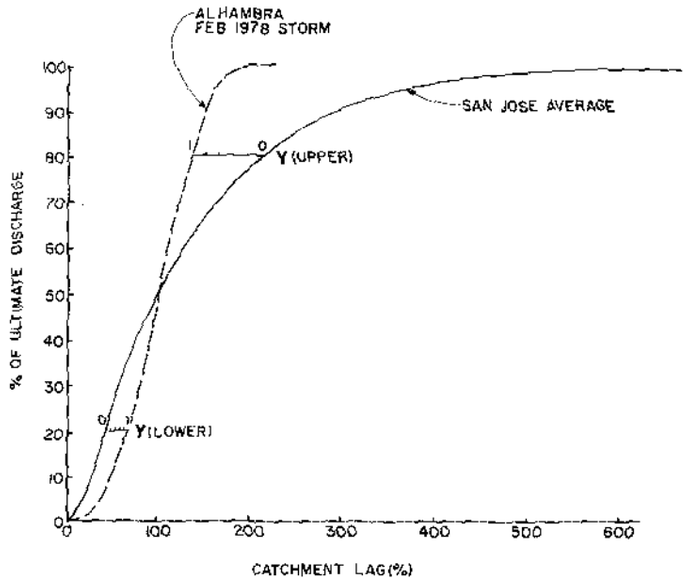


Fig. 7. S-graph scaling for $S(Y)$.

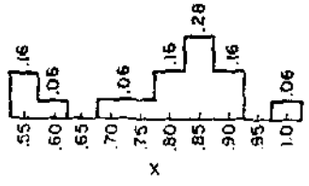


Fig. 8. Distribution-frequency for S-graph Y parameter.

parameters of Y, lag, and mass. Consequently, the randomness in the unit hydrographs can be represented by these mutually dependent random variables. With the normalization process, the variations in total mass and the lag parameter must be also accounted, and were assumed to be normally distributed. Finally, the random variables of $\{Y, \text{lag, mass}\}$ were assumed to be independent for the considered storm class to develop $[\approx_M(\cdot)]$. It is noted that the random variables $\{Y, \text{lag, mass}\}$ are actually mutually dependent and that in order to preserve this dependency, a large sample size of the $\approx_M(\cdot)$ would be needed.

Based upon the model M defined by eqns. (56)–(59), a severe storm of March 1, 1983 (which was not used in the development of $[\approx_M(\cdot)]$) is analyzed for Alhambra Wash and Compton Creek. The distributions $[Q_R^D(\cdot)]$ are plotted along with the recorded stream gage in Figs. 9 and 10. From the figures, the uncertainty in model predictions of $[Q_R^{sD}(\cdot)]$ is significant and should be included when analyzing an operator $\mathscr{A}$ on the runoff predictions.

SURFACE RUNOFF MODEL OPERATOR FREQUENCY DISTRIBUTIONS

In this section, the distribution of the model estimates is considered in order to evaluate the annual T-year values.

From eqn. (52), the correlation distribution for storm event i, $\eta^i(s)$, includes all the uncertainty in the effective rainfall distribution over R, as well as the uncertainty in the several flow routing processes, for the given assumptions about the catchment's runoff response. That is, $\eta^i(\cdot)$ is a realization of the stochastic process, $[\eta(\cdot)]$, where:

$$\eta^i(s) = \sum_{j=1}^{m} \sum_{\langle k \rangle_j} a^i_{\langle k \rangle_j} \sum \lambda_{jk}(1 + X^i_{jk})\phi^i_j (s - \theta^i_{jk} - \alpha^i_{\langle k \rangle_j}) \tag{60}$$

for storm event i, which is an element of some storm class $[\![\xi_z]\!]$, see eqn.(52).

Although use of the $[\eta(\cdot)]$ realizations combines the uncertainties of both the effective rainfalls and also the channel routing and other processes, eqn.(60) is useful in motivating the use of the probabilistic distribution concept in design and planning studies for all hydrologic models, based on just the magnitude of the uncertainties in the effective rainfall distribution over R. That is, although one may argue that a particular model is "physically based" and represents the "true" hydraulic response distributed throughout the catchment, the uncertainty in model input still remains and is not reduced by increasing hydraulic routing modeling complexity. Rather, the uncertainty in input is reduced only by the use of additional rainfall–runoff data, (and input error is different from model error). In eqn.(60), the use of mean value parameters for the routing effects implicitly assumes that the variances of the random variables distributed as $[X_{jk}]$ and $[\theta_{jk}]$ are such that a single set of linear routing parameters can be used on a channel link-by-link basis, for a given storm class.

The distribution of the criterion variable

As before, let R be a free-draining urban catchment without significant detention effects (e.g., dams), nor baseflow, with a single stream gage and rain gage for data analysis purposes. The goal is to develop estimates of rare occurrence values of a runoff criterion variable (or operator), $\mathscr{A}$, evaluated at the stream gage site. Examples of $\mathscr{A}$ are the peak flow rate, or a detention basin peak volume for a given outlet structure located at the stream gage. Thus, $\mathscr{A}$ is the peak demand value of a hydrologic variable from a given runoff hydrograph, evaluated at the stream gage site.

For simplicity, let all the effects of one year's precipitation be identified with an annual storm event $P_i(\cdot)$; the underlying probability space is then the space of all such annual storms. Event $P_i(\cdot)$ may have a duration of a few hours or a few weeks in order to include all the precipitation assumed to be of importance in relating the event to the stream gage measured runoff, $Q_i(\cdot)$.

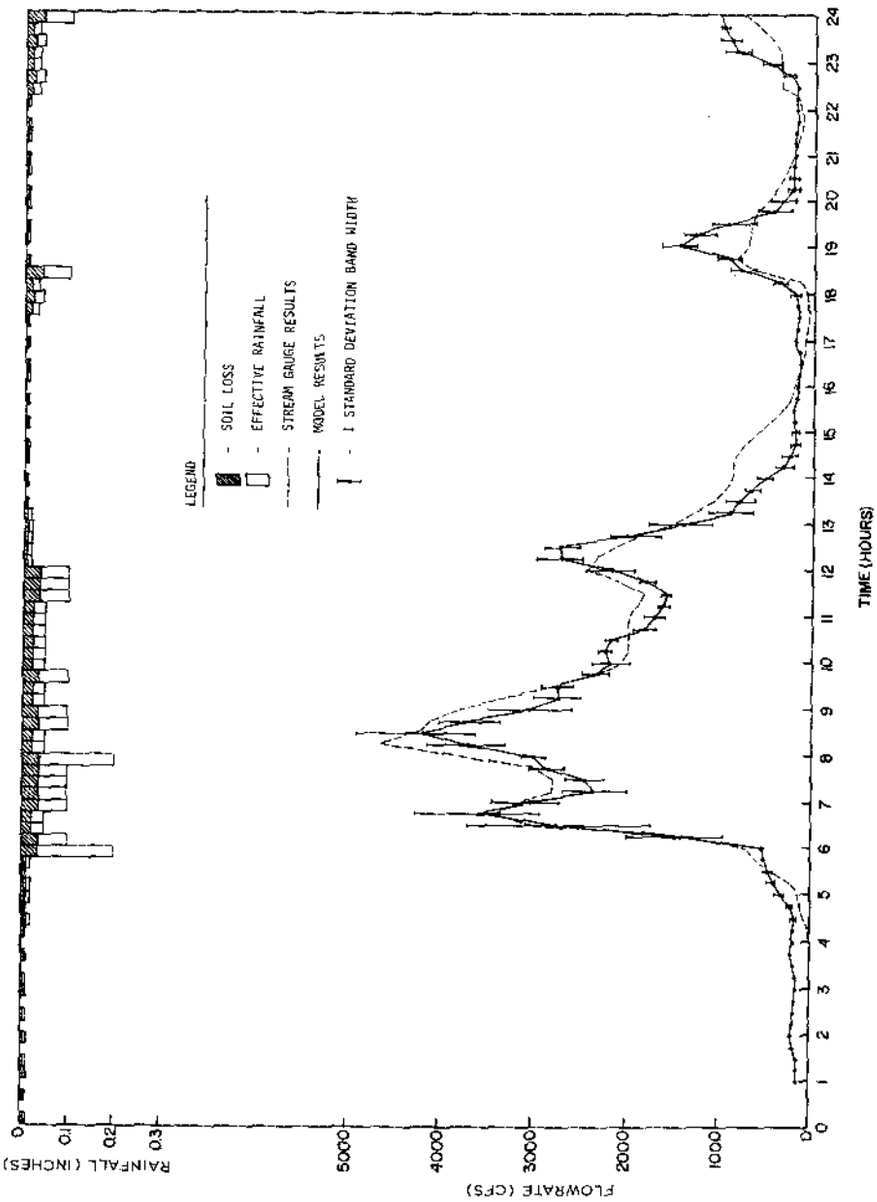


Fig. 10. Monte Carlo analysis of the March 1, 1983 storm for the Compton Creek watershed.

Base input peak duration analysis

Given the base input, $F_i(\cdot)$, let I_δ be the operation of locating the δ-time interval of peak area in $F_i(\cdot)$. Then (see Fig. 11):

$$I_\delta : F_i(\cdot) \rightarrow F_i^\delta(\cdot) \tag{67}$$

where $F_i^\delta(t) \equiv 0$ for all $t \notin I_\delta$; $F_i^\delta(t) = F_i(t)$ for $t \in I_\delta$; and where $\delta > 0$. That is, $F_i^\delta(\cdot)$ is the peak δ-time portion of $F_i(\cdot)$, and $F_i^\delta(\cdot)$ is zero outside of I_δ. It is noted that I_δ is also used as the notation for the peak interval itself. (It is also noted that hereafter I indicates the base input information, and not hydrograph inflow.)

The contribution to $Q_i(\cdot)$ from $F_i^\delta(\cdot)$ is determined by:

$$Q_i^\delta(t) = \int_{s=0}^{t} F_i^\delta(t - s)\, \eta_i(s)\, ds \tag{68}$$

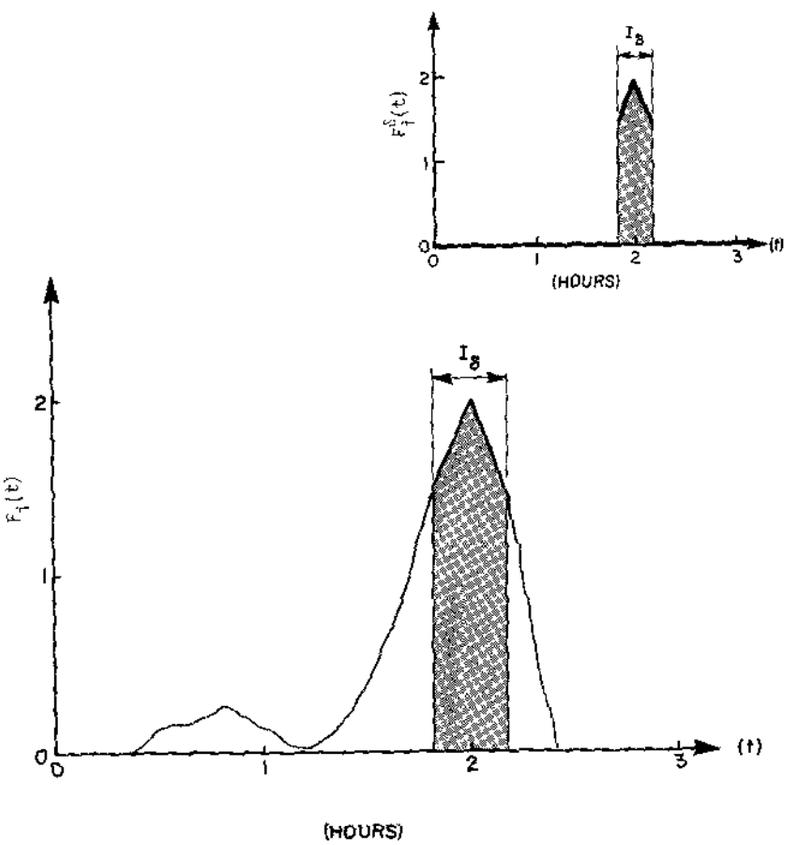


Fig. 11. Locating the peak I_δ interval of $F_i(\cdot)$.

The criterion variable of interest is noted by A_i for annual event i where

$$A_i = \mathscr{A}[Q_i(\cdot)] \tag{61}$$

where $\mathscr{A}[Q_i(\cdot)]$ is notation for finding the maximum value of the criterion variable resulting from the hydrograph, $Q_i(\cdot)$, and where each A is evaluated at the stream gage site; for example, peak discharge is max $[Q_i(t)$: t real] and volume of discharge is $\int Q_i(t)dt$.

The distribution $[A]$ can be estimated from a finite sample A_1, A_2, A_3, and this empirical distribution can be used to obtain the desired T-year return frequency estimates, A_T, of the criterion variable where by definition of exceedence probability:

$$P(A_i \geqslant A_T) = \frac{1}{T}, \text{ for } T > 1 \tag{62}$$

It is noted that A_i is the maximum value of the criterion variable, A, for year i; and A^i is the peak demand of A from arbitrary storm event i.

Sequence of annual base inputs

With only a single rain gage available, all rainfall–runoff models must operate on the annual precipitation events $P_i(\cdot)$. The notation of "effective rainfall" will be generated in the following.

Let F be a function on the precipitation measured at the rain gage:

$$F{:}P_i(\cdot) \rightarrow F_i(\cdot) \tag{63}$$

such that $F_i(\cdot)$ is a nonnegative, bounded, piecewise continuous function of time t. For example, two possible candidates are:

$$F{:}P_i(t) \rightarrow P_i(t); \qquad F{:}P_i(t) \rightarrow \int_{s=0}^{t} P_i(s)\, ds. \tag{64}$$

The rainfall–runoff model, M, is used to relate the synthetic "effective rainfall" $F_i(\cdot)$ to the measured runoff, $Q_i(\cdot)$. Note that $F_i(\cdot)$ depends very strongly on the mapping F chosen.

Thus for the multilinear surface runoff model, M, the base input, $F_i(\cdot)$, and the correlation distribution, $\eta_i(\cdot)$, are used to equate with $Q_i(t)$ by:

$$M{:} < F_i(\cdot), \eta_i(\cdot) > \rightarrow Q_i(\cdot) \tag{65}$$

where $F_i(\cdot)$ must not be strictly zero where $Q_i(\cdot)$ is not strictly zero. Equation (65) shows that M operates upon $F_i(\cdot)$ and $\eta_i(\cdot)$ realizations, for the annual event, i, to produce the annual runoff hydrograph event, $Q_i(\cdot)$.

Letting $\{P_i(\cdot);\ i = 1, 2, \ldots\}$ be the sequence of annual rainfall events measured at the rain gage, then the functional F transforms the rainfall data into the sequence of annual base inputs:

$$F{:}\ \{P_i(\cdot);\ i = 1, 2, \ldots\} \rightarrow \{F_i(\cdot);\ i = 1, 2, \ldots\} \tag{66}$$

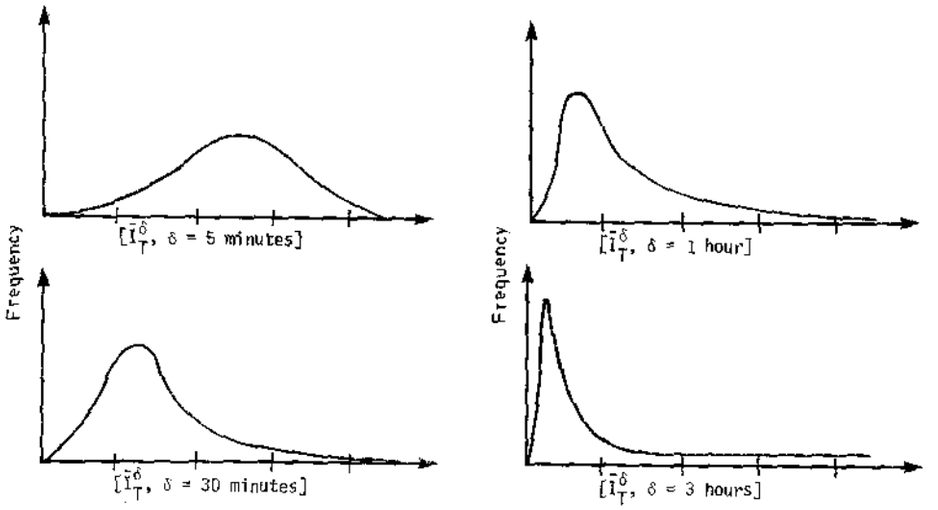


Fig. 13. Frequency distributions of $\bar{I}_T^\delta$ for δ = 5 min, 30 min, 1 h, 3 h.

From eqn.(72), T-year frequency values of $\bar{I}[F_i^\delta(\cdot)]$ are denoted by $\bar{I}_T^\delta$, where (see Fig. 13):

$$P\{\bar{I}[F_i(\cdot)] \geq \bar{I}_T^\delta\} = \frac{1}{T} \tag{75}$$

where $\bar{I}_T^\delta$ is the T-year return frequency value of the base input intensity, for peak duration δ (this is analogous to depth-duration frequency data for rainfall).

We are interested in the "average shape" of the base inputs which have, for a given I_δ, the same total mass of input. To proceed, the entire collection of realizations $F_i^\delta(\cdot)$ is translated in time to begin at a reference time $t = 0$. Thus each $F_i^\delta(\cdot)$ is zero except (possibly) for time $0 \leq t \leq \delta$. Given peak duration time increment δ, the $F_i^\delta(\cdot)$ are further categorized according to similar total mass. Thus, although several $F_i^\delta(\cdot)$ realization have similar total mass, they differ in their time distributions of mass. We want the expected shape of the $\Delta F_i^\delta(\cdot)$ in each grouping of similar mass and define:

$$E_i^\delta(\cdot) = E\{\Delta F_j^\delta(\cdot) | \bar{I}[F_j^\delta(\cdot)] \cong \bar{I}[F_i^\delta(\cdot)]\} \tag{76}$$

where i is the considered event, and j is independent of i.

It is recalled that in eqn. (76), each $\Delta F_j^\delta(\cdot)$ has been translated appropriately in time (Fig. 14), and the expectation is taken with respect to time, t.

Define $\varepsilon_i^\delta(t)$ by:

$$\varepsilon_i^\delta(t) = \Delta F_i^\delta(t) - E_i^\delta(t) \tag{77}$$

where $\varepsilon_i^\delta(t)$ is the variation in base input shape about the expected base input shape of all base inputs with the same mass (approximately) as $F_i^\delta(\cdot)$, (Fig.14).

And the contribution to A_i from $F_i^\delta(\cdot)$ is:

$$A_i^\delta = A[Q_i^\delta(\cdot)] \tag{69}$$

Criterion variable distribution analysis

From the above:

$$A_i = \mathscr{A}\left[\int_{s=0}^{t} F_i(t - s)\,\eta_i(s)\,\mathrm{d}s\right] \tag{70}$$

and:

$$A_i^\delta = \mathscr{A}\left[\int_{s=0}^{t} F_i^\delta(t - s)\,\eta_i(s)\,\mathrm{d}s\right] \tag{71}$$

where $F_i^\delta(\cdot) \to F_i(\cdot)$ as $\delta\uparrow$ (i.e., as δ increases from zero). Then $A_i^\delta \to A_i$ as $\delta\uparrow$ where reasonable assumptions of continuity of $\mathscr{A}$ are assumed. The fact that $A_i^\delta \to A_i$ as $F_i^\delta(\cdot) \to F_i(\cdot)$ will be used in the following to identify the properties of the operator, F, which are involved in the estimates of T-year values of the distribution of annual outcomes, [A].

The base input $F_i^\delta(\cdot)$ is written as the sum of components $\bar{F}_i^\delta(\cdot)$ and $\Delta F_i^\delta(\cdot)$ where (Fig. 12):

$$\bar{I}[F_i^\delta(\cdot)] = \frac{1}{\delta}\int_{s=0}^{\infty} F_i^\delta(s)\,\mathrm{d}s = \frac{1}{\delta}\int_{I_\delta} F_i(s)\,\mathrm{d}s \tag{72}$$

$$\bar{F}_i^\delta(t) = \begin{cases} \bar{I}[F_i^\delta(\cdot)], & \text{for } t \in I_\delta \\ 0, & \text{otherwise} \end{cases} \tag{73}$$

$$\Delta F_i^\delta(t) = \begin{cases} F_i^\delta(t) - \bar{F}_i^\delta(t), & t \in I_\delta \\ 0, & \text{otherwise} \end{cases} \tag{74}$$

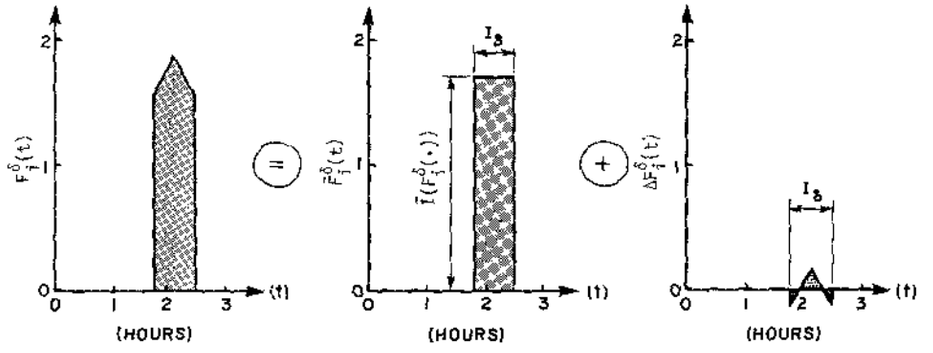


Fig. 12. Writing $F_i^\delta(t)$ of Fig. 11 as the sum $\bar{F}_i^\delta(t) + \Delta F_i^\delta(t)$.

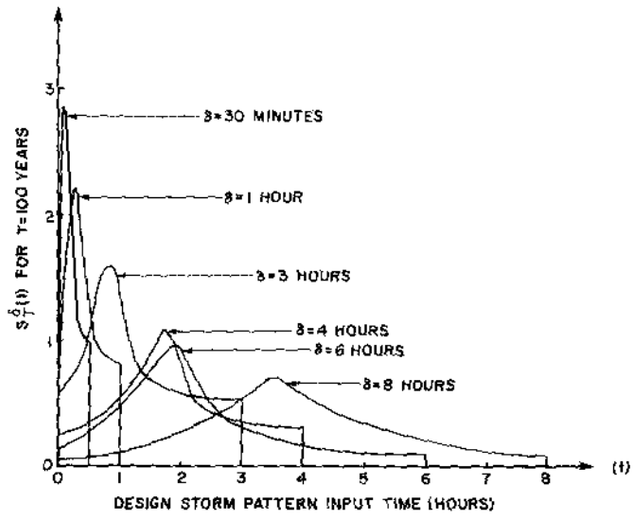


Fig. 15. Plots of $S_T^\delta(t)$ for various δ and $T = 100$ yr.

where in eqn. (81), return frequency, T, is allowed to vary as a real valued positive (nonzero) random variable; and $[\eta_z(\cdot)]$ is the distribution of realizations, $\eta_i(\cdot)$, when the parent $F_i(\cdot) \in [\![\xi_z]\!]$, [that is, there may be several $\eta(\cdot)$ realizations associated to the single realization of $F_i(\cdot)$].

Combining eqns. (80) and (81):

$$[Q^\delta(t)] = \int_{s=0}^{t} \{S_T^\delta(s) + [\varepsilon_T^\delta(s)]\} [\eta_z(t - s)]\, ds \tag{82}$$

and, for operator $\mathscr{A}$, eqn. (82) is used to provide the frequency distribution

$$[A^\delta] = \mathscr{A}[Q^\delta(\cdot)] \tag{83}$$

Figure 16 shows a flow-chart which implements the procedures leading to eqn. (83). Because $A_i^\delta \to A_i$ as $\delta\uparrow$, then necessarily $[A^\delta] \to [A]$ as $\delta\uparrow$.

T-year estimate model simplifications

The earlier sections dealt with uncertainty in predictions of the operator $\mathscr{A}$, which necessitated the inclusion of the distributions $[\eta_z(\cdot)]$ in the final model formulation.

Equation (82) can be considerably simplified if it is assumed that eqn. (55) applies, and that:

$$[A^\delta] \simeq \mathscr{A}[E(Q^\delta(\cdot))] \tag{84}$$

in which case $E[\varepsilon_T^\delta(\cdot)] = 0$ and $E[\eta_z(\cdot)] = \eta_z(\cdot)$, and eqns. (82) and (84) can be

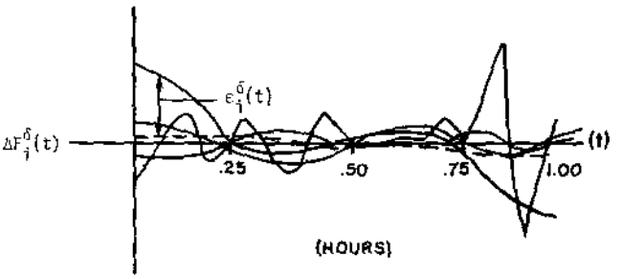


Fig. 14. Several plots of $\Delta F_i^\delta(t)$ for $\delta = 1$ h, and $\bar{I}[F_i^\delta(\cdot)] = I_\delta^\delta$. Dashed line is the expected value. (Note that all base inputs are translated in time to initiate at time 0.)

Then in summary, with all components appropriately translated in time:

$$Q_i^\delta(t) = \int_0^t [\bar{F}_i^\delta(s) + E_i^\delta(s) + \varepsilon_i^\delta(s)]\, \eta_i(t - s)\, ds \tag{78}$$

where $\bar{F}_i^\delta(\cdot)$ is the mean intensity of the base input, $F_i^\delta(\cdot)$, over the time interval $0 \leqslant t \leqslant \delta$ [where $F_i^\delta(\cdot)$ has been translated to begin at time $t = 0$]; $E_i^\delta(\cdot)$ is the expected shape of all possible δ-interval peak durations of base inputs with the same total mass of $F_i^\delta(\cdot)$; $\varepsilon_i^\delta(\cdot)$ is the variation of $\Delta F_i^\delta(\cdot)$ about the expected shape, $E_i^\delta(\cdot)$; and $\eta_i(\cdot)$ is the necessary multilinear model correlation distribution for the parent annual event $F_i(\cdot)$, in some storm class $[\![\xi_z]\!]$.

Estimation of T-year values of the criterion variable

For each peak duration, I_δ, the samples of $F_i^\delta(\cdot)$, see eqns. (73) and (74), are now analyzed to determine the underlying distribution of the annual outcomes of the values, $\bar{I}[F_i^\delta(\cdot)]$. From these distributions of mean intensity of I_δ base inputs, T-year values, I_T^δ of the $\bar{I}[F^\delta(\cdot)]$ can be derived (Fig. 13) and the unique T-year $\bar{F}_T^\delta(\cdot)$ defined by:

$$\bar{I}[\bar{F}_T^\delta(\cdot)] = I_T^\delta \tag{79}$$

Given I_T^δ, $\bar{F}_T^\delta(\cdot)$ is defined, and consequently $E_T^\delta(\cdot)$ and the distribution $[\varepsilon_T^\delta(\cdot)]$ are known. The "T-year I_δ base input", $S_T^\delta(\cdot)$ is defined as:

$$S_T^\delta(\cdot) = \bar{F}_T^\delta(\cdot) + E_T^\delta(\cdot) \tag{80}$$

Figure 15 shows a set of $S_T^\delta(\cdot)$ for $T = 100$ years, and various δ, using the data of application 3, and the model structure of eqns. (58) and (59). The T-year I_δ base input, $S_T^\delta(\cdot)$, varies in both shape and mass as either T or δ varies. The distribution $[Q^\delta(\cdot)]$ of realizations of $Q_i^\delta(\cdot)$ is now written from eqns. (68), (78), and (80) as:

$$[Q^\delta(t)] = \int_{s=0}^t \{\bar{F}_T^\delta(s) + E_T^\delta(s) + [\varepsilon_T^\delta(s)]\}\, [\eta_z(t - s)]\, ds \tag{81}$$

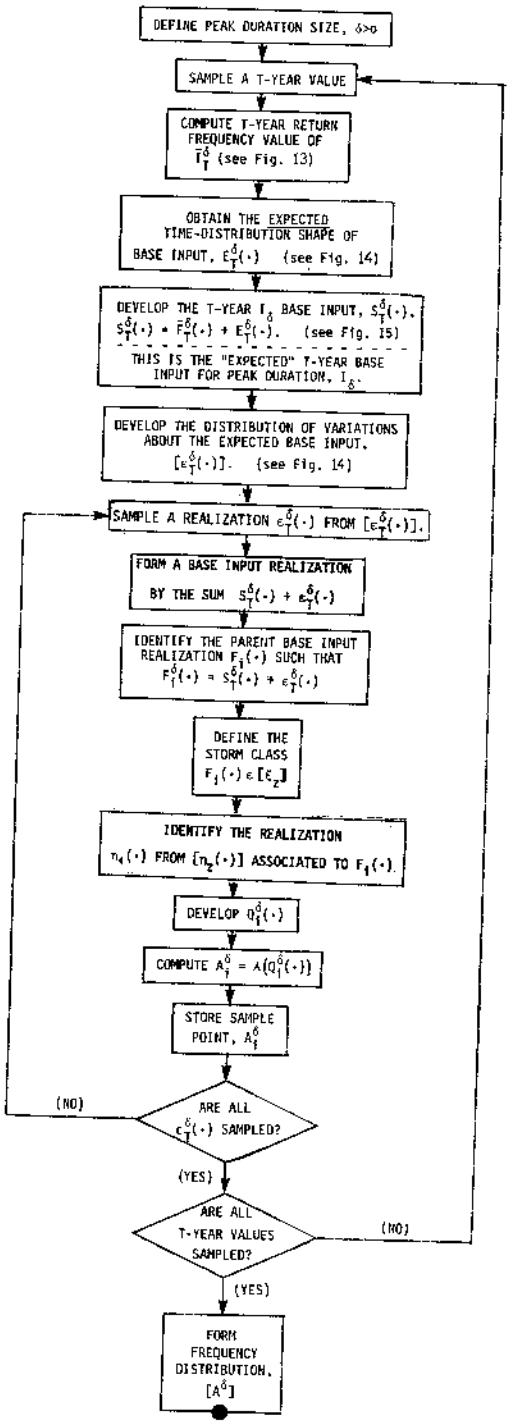


Fig. 16. Flow chart for estimation of T-year values of $[A^\delta]$ (preserving mutual dependency of random variables).

combined as:

$$[A^\delta] \simeq \mathscr{A}\left[\int_{s=0}^{t} S_T^\delta(t - s)\, \eta_z(s)\, ds\right] \tag{85}$$

where T is the annual series random variable. If furthermore it is assumed that the storm classes of base input, $[\![\xi_z]\!]$, are highly correlated to T-year values of base input mean intensity, then storm classes of T-year base input can be defined, $[\![\xi_T]\!]$, [perhaps on a duration basis such as 1-h, 3-h, etc.; see Pierrehumbert, 1974, for the case of eqn. (58)], and eqn. (85) becomes:

$$[A^\delta] \simeq \mathscr{A}\left[\int_{s=0}^{t} S_T^\delta(t - s)\, \eta_T(s)\, ds\right] \tag{86}$$

where T varies as an independent random variable. Finally, if it is assumed that the T-year value of $[A^\delta]$ monotonically increases as T-increases in eqn. (86), then the T_0 return frequency value of A is:

$$A_{T_0} \simeq \max_{\delta} \mathscr{A}\left[\int_{s=0}^{t} S_{T_0}^\delta(t - s)\, \eta_{T_0}(s)\, ds\right], \text{ as } \delta\uparrow \tag{87}$$

where $\eta_{T_0}(\cdot)$ is the expected realization of a multilinear surface runoff model response $[\eta(\cdot)]$ corresponding to storm class $[\![\xi_T]\!]$. Equation (87) is a form of the well-known design storm single area unit hydrograph procedure (e.g., Hromadka et al., 1987).

The choice of the base input estimator, F

In the above development it is shown that, generally speaking, T-year return frequency estimates of the criterion variable, A, estimated from a rainfall–runoff model, involves two major components: (1) a base input functional, F, which maps the rainfall data into a synthetic data set of effective rainfall; and (2) a catchment response model which can be approximated by a multilinear response model based on storm classes of base input.

In the eqns.(84)–(87), the most common practice in flood control hydrology studies is to assume that (1) the variation of base input shape (in time) about the expected base input mass and shape is negligible; and (2) the distribution of the criterion variable $[A]$ can be evaluated using the expected multilinear surface runoff response model, $E[\eta_T(\cdot)]$, for storm class $[\![\xi_T]\!]$. At present, there is insufficient rainfall–runoff data to conclude whether such simplifications are invalid. However, a basic question needs to be addressed: is there an optimum choice of functional, F, which best develops the distribution for the criterion variable $[A]$, given that the multilinear model structure of eqn. (52) applies?

The distributions of $[\eta_z(\cdot)]$ all depend on the choice of operator, F. Additionally, the various distributions involved with $S_T^\delta(\cdot)$ and $\varepsilon_T^\delta(\cdot)$ are completely defined by the sequence of annual base inputs produced by a particular F operating on the given sequence of annual rainfall, $\{P_i(\cdot)\}$.

Example

In order to illustrate the above concepts, the analysis is applied to a sequence of annual rainfalls where the effective rainfall estimator, F, produces particularly simple outcomes.

Let F:

$$P_i(\cdot) \to \begin{cases} 1 - \dfrac{1}{T}, & 0 \leqslant t \leqslant 1 \\ 0, \text{ otherwise} \end{cases}$$

where T is the annual outcome random variable. Then each T-year base input event is uniquely defined by:

$$F_T^\delta(\cdot) = 1 - \frac{1}{T}; \quad 0 \leqslant \delta \leqslant 1, 0 < t < 1$$

$$\bar{F}_T^\delta(\cdot) = F_T^\delta(\cdot); \quad 0 \leqslant \delta \leqslant 1, 0 \leqslant t \leqslant 1$$

$$E_T^\delta(\cdot) = 0$$

$$[\varepsilon_T^\delta(\cdot)] = 0$$

From the above, $S_T^\delta(\cdot) = 1 - 1/T$ for $0 \leqslant \delta \leqslant 1$ and $0 \leqslant t \leqslant 1$. (It is noted that even though this example for F appears highly restrictive, eqn.(52) still applies for the given model structure.) The above example demonstrates some of the possible ramifications in the choice of F. With the noted variability of effective rainfall over the catchment, and the high sensitivity of surface runoff model results to the said variability, is there a rational motivation to use a particular operator, F', over another operator, F?

One approach in choosing F is to incorporate the regionalized statistical tendencies of the rainfall data into the runoff predictions of the distributions for A. That is, given an operator F, should F be applied at several neighbouring catchments, the statistical tendencies of the base inputs $F_i(\cdot)$ could be developed at each catchment, and these statistics could then be regionalized together to form a broader pool of data for base-input/runoff correlation. Generally, rainfall data are locally regionalized and oftentimes represent considerable station-year data. Should the operator F be selected such that the statistical tendencies of the base input equate directly to the rainfall data, then F could be more readily regionalized to local conditions.

Example

Let F be defined by $F: P_i(\cdot) \to kP_i(\cdot)$, for k a positive nonzero constant. Then the depth–duration statistics of $F^\delta(\cdot)$ equate directly to the rainfall data: $\bar{I}[F^\delta(\cdot)] = k\bar{I}[P^\delta(\cdot)]$, and the temporal distributions of base input follow directly from the rainfall data (Pierrehumbert, 1974). Note that in this case, F is the simple runoff coefficient loss rate function (see Scully and Bender, 1969; Schilling and Fuchs, 1986).

Example

Let F be defined by:

$$F{:}P_i^{\delta}(\cdot) \rightarrow \begin{cases} P_i(\cdot) - \phi, & \text{when positive} \\ 0, \text{ otherwise} \end{cases}$$

where ϕ is a positive constant. This operator also preserves the rainfall data statistics (for durations of nonzero base inputs) and is widely used as the phi-index approach.

Example

$F{:}P_i(\cdot) \rightarrow \alpha P_i^{\beta}(\cdot)$, where α and β are the positive constants, preserves the rainfall data statistics.

The constant fraction loss function and the phi-index loss function are both operators on rainfall data which have been widely used, and continue to be used by practitioners in flood control planning and design. There is a wide range of other F operators reported in the literature, oftentimes involving considerable complexity. The lack of convincing evidence to use operators (F) more complex than the constant fraction or phi-index approach, and the advantage afforded in using these simple and well-known operators due to the ability of incorporating the local rainfall statistical tendencies, when available, leads to the motivation of recommending the use of these operators in surface runoff models.

CONCLUSIONS

A detailed mathematical analysis of approximate methods to include uncertainty in rainfall–runoff model predictions is presented. New notation is introduced in order to clarify and unify many of the concepts and techniques commonly employed in rainfall–runoff models, and also in order to analyze the mathematical underpinnings of these techniques. Two key problems are considered, namely: (1) the prediction of runoff quantities given an assumed rainfall event (and other conditions); and (2) the estimation of T-year return frequency values of a prescribed criterion variable (e.g., peak flow rate, detention basin volume).

Because rainfall–runoff models produce a single runoff hydrograph as the estimate of runoff for a future event, and because there is generally considerable error in criterion variable estimates developed from such runoff hydrographs, it is important to approximate the probabilistic distribution of values of the criterion variable in order to develop confidence interval estimates. As rainfall–runoff data increase in quantity, the uncertainty in runoff criterion variable estimates can be continually evaluated based on the best available data and the rainfall–runoff model being used.

REFERENCES

Akan, A.O. and Yen, B.C., 1981. Diffusion-wave flood routing in channel networks. ASCE J. Hydraul. Div., HY6: 719–731.

Beard, L. and Chang, S., 1979. Urbanization impact on stream flow. ASCE J. Hydraul. Div., 105 (6): 647–659.

Becker, A. and Kundzewicz, Z.W., 1987. Nonlinear flood routing with multi-linear models. Water Resour. Res., 23 (6): 1043–1048.

Doyle, W.H., Shearman, J.O., Stiltner, G.J. and Krug, W.R., 1983. A digital model for streamflow routing by convolution methods. U.S.G.S. Water Res. Invest. Rep., 83–4160.

Garen, D. and Burges, S., 1981. Approximate error bounds for simulated hydrographs. ASCE J. Hydraul. Div., 107 (11): 1519–1534.

Hromadka II, T.V. and Yen, C.C., 1986. A diffusion hydrodynamic model (DHM). Adv. Water Resour., 9 (3): 118–170.

Hromadka II, T.V., McCuen, R.H. and Yen, C.C., 1987. Computational Hydrology in Flood Control Design and Planning. Lighthouse Publications, Mission Viejo, Calif., 520 pp.

Huff, F.A., 1970. Spatial distribution of rainfall rates. Water Resour. Res., 6 (1): 254–260.

Loague, K. and Freeze, R., 1985. A comparison of rainfall runoff modeling techniques on small upland catchments, Water Resour. Res., 21 (2): 229–248.

Mockus, V.J., 1972. National Engineering Handbook, Section 4, Hydrology. U.S. Dep. Agric., Soil Conserv. Serv., Washington, D.C.

Nash, J. and Sutcliffe, J., 1970. River flow forecasting through conceptual models, Part 1 -- A discussion of principles. J. Hydrol., 10: 282–290.

Pierrehumbert, C.L., 1974. Point rainfall intensity–frequency–duration data capital cities. Dep. Sci., Bur. Meteorol., Bull. 49, Aust. Gov. Publ. Serv., Canberra, A.C.T.

Sarikelle, S., Chien, J.S. and French, G.L., 1978. Linearized subhydrographs urban runoff model. ASCE Proc. 26th Hydraul. Div. Spec. Conf., University of Maryland.

Schilling, W. and Fuchs, L., 1986. Errors in stormwater modeling — A quantitative assessment. ASCE J. Hydraul. Eng., 112 (2): 111–123.

Scully, D.R. and Bender, D.L., 1969. Separation of rainfall excess from total rainfall. Water Resour. Res., 5 (4): 887–883.

Troutman, B., 1982. An analysis of input in precipitation runoff models using regression with errors in the independent variables. Water Resour. Res., 18 (4): 947–964.

Tsokos, C.P. and Padgett, W.J., 1974. Random integral equations with applications to life, sciences and engineering. Academic Press, Math. Sci. Eng., Vol. 108.

U.S. Army Corps of Engineers, 1982. HEC, Hydrologic analysis of ungaged watersheds using HEC-1, Training Doc. No. 15.