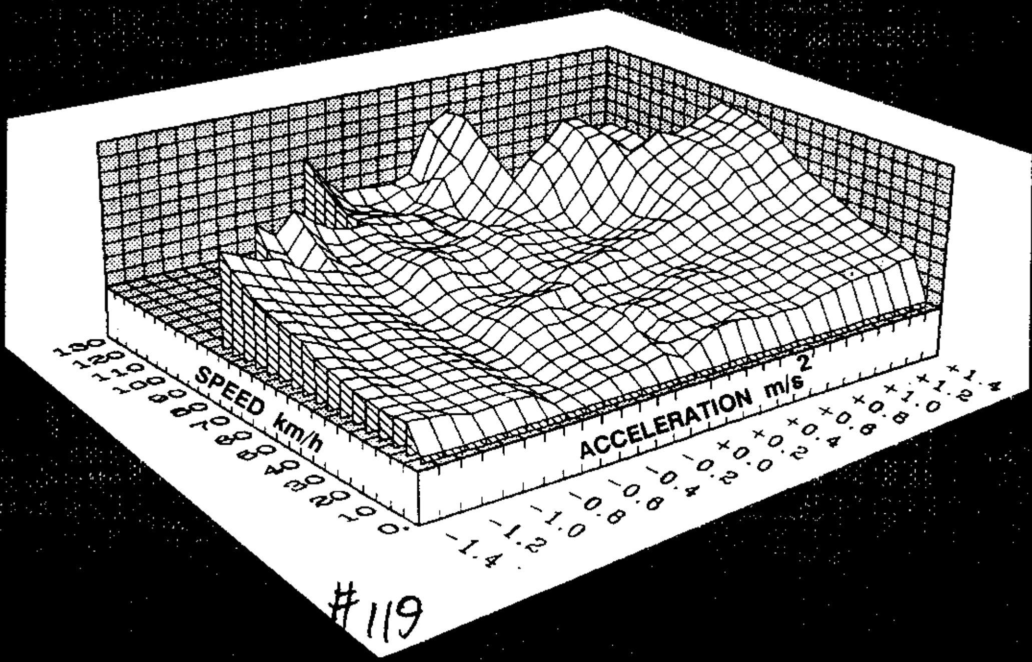
Computer Techniques in Environmental Studies III

Computer Techniques in Environmental Studies III



Computer Techniques in Environmental Studies III

Proceedings of the third International Conference on Development and Application of Computer Techniques to Environmental Studies, held in Montreal, September 11-13, 1990.

Editor: P. Zannetti

Computational Mechanics Publications
Southampton Boston

Co-published with

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo
Hong Kong Barcelona

Hydrologic Model Discretization Error

T.V. Hromadka II

Water Resources Engineering, Williamson and Schmid, Irvine, California 92714, USA and *Department of Mathematics, California State University, Fullerton, California 92634, USA*

ABSTRACT

In this paper, the premise that the uncertainty in the storm rainfall distribution over the catchment, R, is a dominant factor in the uncertainty in hydrologic model output, is used to develop useful equations regarding the distribution of hydrologic modeling uncertainty.

Also developed herein are estimates in the variance of the peak flow rates for single area unit hydrograph (UH) models and highly discretized models based on several subareas.

It is shown that the discretization of a catchment into subareas, without subarea rainfall-runoff data, introduces a significant modeling error which results in a departure from the natural variance in the correlation of the available rainfall-runoff data.

INTRODUCTION

Three major assumptions are used herein. 1) the watershed has negligible detention and backwater effects. Such a catchment would be relatively steep and with a fully developed drainage system. The subarea, R_j unit hydrograph (UH), $\phi_j^i(s)$, is variable between storms, and can approximate any other surface runoff submodel. 2) Translation routing (no peak attenuation) is used for channel flow routing. This assumption is augmented by the use of variable channel link travel time, for each storm i, which is reflected in the appropriate sum of travel times, τ_j^i, used for R_j in storm i. The kinematic wave channel routing technique also does not allow for peak attenuation (when the articificial numerical diffusion is eliminated), and hence the above two assumptions are somewhat allied to the KW technique. (Again, channel storage effects can be included by using the linear routing technique of Hromadka[1], but are ignored in order to simplify the mathematics.) 3) The effective rainfall distribution over the catchment R is linear with respect to the measured effective rainfall data; that is in each R_j,

$$e_j^i(t) = \sum_{k=1}^{n_j^i} \lambda_{jk}^i \, e_g^i(t - \theta_{jk}^i) \tag{1}$$

where λ_{jk}^i are coefficients for R_j and storm i; and θ_{jk}^i are constant timesteps for R_j and storm i. Equation (1) is used to approximate as closely as possible the effective rainfall distributions over R_j given only the $e_g^i(t)$. Convoluting Eq. (1) with the UH gives

$$Q_m^i(t) = \int_{s=0}^{t} e_g^i(t-s) \sum_{j=1}^{m} \sum_{k=1}^{n_j^i} \lambda_{jk}^i \, \phi_j^i(s-\tau_j^i-\theta_{jk}^i) \, ds \tag{2}$$

Comparing $Q_m^i(t)$ to $Q_1^i(t)$ it is seen that for the above assumptions, $Q_m^i(t) = Q_1^i(t)$ where necesarily in Equation (3),

$$\eta^i(s) = \sum_{j=1}^{m} \sum_{k=1}^{n_j^i} \lambda_{jk}^i \, \phi_j^i(s-\tau_j^i-\theta_{jk}^i) \tag{3}$$

Hence by the accumulation of $\eta^i(s)$ distributions, the output from $Q_1^i(t)$ (in the predictive mode) is the distribution

$$[Q_1(t)] = \int_{s=0}^{t} e_g(t-s) \, [\eta(s)] \, ds \tag{4}$$

where $[\eta(s)]$ corresponds to a storm class, $\langle \xi_q \rangle$, which represents the model input, $e_g(t)$. The variations in the random variable $[Q_1(t)]$ represents the natural variance between the effective rainfall data measured at the rain gauge site, and the stream gauge data.

In practice, Equation (4) is not used in hydrologic predictive studies, but instead the expected value of the $\eta^i(s)$ is used giving the expected output,

$$E[Q_1(t)] = \int_{s=0}^{t} e_g(t-s) \, E[\eta(s)] \, ds \tag{5}$$

where $E[\eta(s)]$ is the simple averaging of ordinates of the correlation distributions, $\eta^i(s)$. Use of Equation (5) is but an estimate in the distribution of outcomes, and usually has a 50-percent confidence limit (one-sided) associated with its estimates of hydrologic quantities such as peak flow rate, q_p.

The use of the m-subarea link node model estimate, $\hat{Q}_m{}^i(t)$, results in

$$\hat{Q}_m{}^i(t) = \int_{s=0}^{t} e_g{}^i(t-s) \sum_{j=1}^{m} \sum_{k=1}^{n_j{}^i} \hat{\lambda}_{jk}^i \hat{\phi}_j{}^i(s-\hat{\tau}_j{}^i-\hat{\theta}_{jk}^i) \, ds \qquad (6)$$

where all hats are estimates without subarea rainfall-runoff data. Without subarea data for a correlation effort, the $\hat{\lambda}_{jk}{}^i$, which include the effects of loss rate nonhomogeneity and the effects of the rainfall variations over R with respect to the rain gauge site, can only be specified to be $\hat{\lambda}_j$, which represents only the loss rate nonhomogeneity. Similarly, the storm timing offsets, $\hat{\theta}_{jk}{}^i$, are all unknown and are set to zero, $\hat{\theta}_{jk}{}^i \equiv 0$.

$$\hat{Q}_m{}^i(t) = \int_{s=0}^{t} e_g{}^i(t-s) \sum_{j=1}^{m} \hat{\lambda}_j \hat{\phi}_j{}^i(s-\hat{\tau}_j{}^i) \, ds \qquad (7)$$

DISCRETIZATION ERROR

Assuming that the effects of rainfall variations over R is a dominating factor in hydrologic model accuracy, then the $\lambda_{jk}{}^i$ and $\theta_{jk}{}^i$ are random variables $[\lambda_{jk}]$, $[\theta_{jk}]$, respectively. Equation (2) is now written as

$$[Q_m(t)] = \int_{s=0}^{t} e_g(t-s) \sum_{j=1}^{m} \sum_{k=1}^{n_j{}^i} [\lambda_{jk}] \, \phi_j{}^i(s-\tau_j{}^i - [\theta_{jk}]) \, ds \qquad (8)$$

In comparing Equations (4) and (8) it is seen that

$$[\eta(s)] = \sum_{j=1}^{m} \sum_{k=1}^{n_j{}^i} [\lambda_{jk}] \, \phi_j(s - \tau_j{}^i - [\theta_{jk}]) \, ds \qquad (9)$$

where the $\phi_j{}^i$ and $\tau_j{}^i$ are known precisely from the rainfall-runoff data available only to $\hat{Q}_m{}^i(t)$, but the $[\eta(s)]$ can be developed from the available effective rainfall and runoff data.

However in comparing Equations (7) and (8), the estimator $\hat{Q}_m{}^i(t)$ specifies that $[\lambda_{jk}] = \hat{\lambda}_j$, and $[\theta_{jk}] = 0$. Hence, a considerable error is entered into the discretized model due to the discretization process itself. Assuming that the subarea effective rainfalls are linear in $e_g{}^i(t)$, then in order to avoid the above discretization error, a stream gauge is needed in each subarea in order to correlate the subarea runoff, $Q_j{}^i(t)$, to the available effective rainfall data at the rain gauge site, $e_g{}^i(t)$, for storm event i. Should the subarea stream gauge be supplied, then not only would the errors in defining $[\lambda_{jk}]$ and $[\theta_{jk}]$ be eliminated, but also errors in the $\hat{\phi}_j{}^i(s)$ estimates, and the discretized model would then achieve the accuracy of the simpler $Q_1{}^i(t)$ model. Should the $\hat{e}_j{}^i(t)$ be nonlinear in $e_g{}^i(t)$, then a rain gauge is also needed in R_j.

ERRORS IN EVALUATING MODELING UNCERTAINTY

Should the estimator, $\hat{Q}_m{}^i(t)$, be used to evaluate the uncertainty in model estimates, the usual procedure is to let the subarea hydrologic parameters vary independently in a Monte Carlo sense, and analyze the modeling output from $\hat{Q}_m{}^i(t)$. For example, Garen and Burges[2] assume mutual independence among model parameters in evaluating the Stanford Watershed Model for modeling uncertainty. Should the catchment, R, be discretized into m subareas, R_j, each subarea having p- parameters for the subarea model, then the estimator, $\hat{Q}_m{}^i(t)$, has mp parameters to vary independently (ignoring routing parameters).

Each subarea R_j runoff model, $\hat{Q}_j{}^i(t)$, is based on parameters developed from a single regionalized or local data base. Assuming for simplicity that R is nearly homogeneous, then the R_j parameters are selected from the same probability distribution function (pdf) for each parameter, X_i, used in the model.

Runoff hydrograph models develop an output which is a sequence of unit interval flowrates (usually a 5-minute interval). Each unit flowrate can then be thought of as a random variable. Should the peak flow rate, q_p, be the hydrologic quantity of interest, then the several unit intervals where q_p occurs (for different parameter sets) become the focal point of the analysis.

Subarea Peak Flow Rates in Alignment

Should each subarea $\hat{Q}_j{}^i(t)$ peak flow rate ($\hat{Q}_j$) occur in the same unit interval for all parameter combinations, then

$$\text{Var}\,[\hat{Q}_m] = \text{Var}\left[\sum_{j=1}^{m} \hat{Q}_j\right] \tag{10}$$

where Var $[\hat{Q}_m]$ is the variance of the peak flow estimate from the estimator, $\hat{Q}_m{}^i(t)$. Similarly, $[Q_m]$ is the distribution of peak flow values Q_m from the 'true' $Q_m{}^i(t)$ model.

From the $Q_1{}^i(t)$ model, let $[Y]$ be a random variable defined by

$$[Y] = [Q_1]/E[Q_1] = [Q_m]/E[Q_m] \tag{11}$$

where $[Q_1]$ is the distribution of peak Q values from $Q_1(t)$ for a given storm. Thus, $[Y]$ is supported by the available rainfall-runoff data for the entire catchment. (It is noted here that in practice, the expected value of the distribution, $E[Q_1]$, has been historically used as "the model output" for peak flow. Hence, hydrologic models are developed and calibrated, but only the expected value of the output distribution is used for study purposes.)

To proceed with Equation (10), some assumption is needed regarding the subarea distributions of each $[\hat{Y}_j] = [\hat{Q}]/E[\hat{Q}_j]$. Without subarea rainfall-runoff data to evaluate $[\hat{Y}_j]$ directly, one can define constants, λ_j, such that

$$\mathrm{Var}\,[\hat{Y}_j] = \lambda_j \,\mathrm{Var}\,[Y] \tag{12}$$

From the above,

$$\mathrm{Var}\,[\hat{Y}_j] = \mathrm{Var}\,([\hat{Q}_j]/E[\hat{Q}_j]) \tag{13}$$

or from Equations (14) and (15),

$$\mathrm{Var}\,[\hat{Q}_j] = \lambda_j \,E[\hat{Q}_j]^2 \,\mathrm{Var}\,[Y]$$

$$= \lambda_j \;\frac{E[\hat{Q}_j]^2 \,\mathrm{Var}\,[Q_1]}{E[Q_1]^2} \tag{14}$$

For the subarea peak flow rates occurring in the same unit interval, and assuming independence between subarea runoffs (due to the mutual independence of subarea model parameters), Equations (10) and (14) give

$$\mathrm{Var}\,[\hat{Q}_m] = \mathrm{Var}\,[Q_1] \left(\sum_{j=1}^{m} \lambda_j \,\frac{E[\hat{Q}_j]^2}{E[Q_1]^2} \right) \tag{15}$$

Without subarea rainfall-runoff data, the λ_j may be implicitly assigned as $\lambda_j = 1$, and Equation (15) becomes the estimator,

$$\mathrm{Var}\,[\hat{Q}_m] = \mathrm{Var}\,[Q_1] \sum_{j=1}^{m} \left(\frac{E[\hat{Q}_j]}{E[Q_1]} \right)^2 \tag{16}$$

Application 1

A catchment R is subdivided into five equal subareas, each with identical runoff responses. The 'true' model expected peak flow rate for R is $Q_1 = 5{,}000$ cfs, whreas the subarea expected peak flow rates are all $R_j = 1200$ cfs. The expected peak $\hat{Q}_j$ all are in alignment for all parameter sets. Calculate the Var $[\hat{Q}_m]$ given the true Var $[Q_m]$:

Solution. Using Equation (16), $(E[\hat{Q}_j]/E[Q_m])^2 = (1200/5000)^2 = 0.0576$. Thus, Var $[\hat{Q}_m]$ = (Var $[Q_m]$)(5)(0.0576) = 0.288 Var $[Q_m]$. That is, the estimator $\hat{Q}_m(t)$ will predict a significant reduction the true variance about the peak rate for R, due to the use of 5-subareas in the link-node model estimator.

Application 2.
A catchment R is subdivided into m subareas, R_j, each of identical runoff response. All expected peak flow rates, $\hat{Q}_j$, are in alignment so that Equation (15) applies. Calculate the variance in the peak flow rate estimate, Var $[\hat{Q}_m]$, using the estimator, $\hat{Q}_m(t)$, given the 'true' value, Var $[Q_m]$.

Solution. From Equation (15), Var $[\hat{Q}_m]$= Var $[Q_m]/m$. Thus as m gets large, Var $[\hat{Q}_m]$ reduces.

From the above applications it is seen that the more discretized a model estimate, $\hat{Q}_m(t)$, becomes, the less reliable are its estimates of the uncertainty distribution. The true variance in the prediction of peak flow rate, given the available data, is given by Var $[Q_m]$ which equals Var $[Q_1]$.

Should the subarea peak flow rates not be in alignment, then Equation (15) (or (16)) is still used except now the less useful inequality reqults:

$$\text{Var }[\hat{Q}_m] \leq n \text{ Var }[Q_m] \left(\sum_{j=1}^{m} \lambda_j \frac{E[\hat{Q}_j]^2}{E[Q_m]^2} \right) \tag{17}$$

where n is the number of unit intervals where the peak Q of $\hat{Q}_m$ may occur, and again Var $[Q_m]$ = Var $[Q_1]$.

CONCLUSIONS

The Var$[Q_1]$ = Var $[Q_m]$ is the 'true' variance and represents the natural variance in the correlation of the available rainfall-runoff data. The introduction of subareas, linked together by channel flow routing algorithms, given only a single rain gauge and stream gauge for data, results in a departure of the model from the true data correlation.

In comparison, the single area UH model preserves the natural variation, and can be used to develop the true modeling uncertainty distribution for the assumptions leading to the $Q_m{}^i(t)$ model (with linear flow routing).

The arbitrary use of discretization without subarea rainfall-runoff data results in a modeling error (discretization error) which is caused by the incorrect definition of the effective rainfall distribution over the catchment.

Care must be taken, therefore, to use subareas only when the discretization error is considered less than the error in not using subareas; for example, when significant detention effects are present such as swamps, lakes, or equivalent. Channel storage effects are included in the single area model UH (linear flow routing).

REFERENCES

1. Hromadka II, T.V., Back to the Unit Hydrograph Method, Proceedings: ENVIROSOFT '88 Conference, Porto Carras, Greece, Computational Mechanics Press, 1988.

2. Garen, D. and Burges, S., Approximate Error Bounds for Simulated Hydrographs, Journal of the Hydraulics Division, Proceedings of The American Society of Civil Engineers, Vol. 107, No. HY11, 1981.