Uncertainty in Model Calibration

R.J. Whitley
University of California, Irvine, California, USA
T.V. Hromadka II
Williamson and Schmid, 17782 Sky Park Blvd., Irvine, California USA and California State University, Fullerton, California, USA

ABSTRACT

In this paper, the unit hydrograph (UH) method is used for the analysis of three model structures for an arbitrary storm event i: (1) a single area UH model noted as $Q_1^i(t)$; (2) an m-subarea link-node model of the catchment, R, with linear unsteady flow routing, and with variable subarea UH's, and where the subarea effective rainfalls are linear in the measured effective rainfall, noted as $Q_m^i(t)$; and (3) an estimator, $\hat{Q}_m^i(t)$, which represents the $Q_m^i(t)$ model except that all model parameters are estimated, and the storm effective rainfall distribution over R is estimated. Available for model calibration purposes is a rain gauge site which is monitored such as to produce an effective rainfall distribution, and a stream gauge located at the downstream point of R.

Using the above three model structures for R, the modeling calibration can be analyzed as to parameter calibration efficiency, reliability, and rationality.

INTRODUCTION

The issue of parameter calibration for hydrologic models has been a topic of considerable concern since the inception of computer modeling techniques. Hromadka and Whitley[1] reviews several of the negative reports obtained from the open literature concerning the lack of success in obtaining "optimum" parameter sets for hydrologic models of all types. Inherent in the cited literature review is the growing weight of evidence that (1) simple models (e.g., a single area unit hydrograph (UH) model) generally can perform as good as or better than complex models (e.g., a highly discretized, link-node model, possibly with a soil-moisture accounting subalgorithm), and (2) the uncertainty in the effective rainfall distribution (rainfall less losses) over the catchment, R, generally is a dominant factor in the uncertainty in hydrologic model output.

In this paper, the unit hydrograph (UH) method is used for the analysis of three model structures for an arbitrary storm event i:

(1) a single area UH model, noted as $Q_1^i(t)$;
(2) an m-subarea link-node model of the catchment, R, with linear unsteady flow routing (Hromadka[2]), and with variable subareas UH's, and where the subarea effective rainfalls are linear with respect to measured effective rainfall monitored at the single available rain gauge site, noted as $Q_m^i(t)$; and
(3) an estimator, $\hat{Q}_m^i(t)$, which represents the $Q_m^i(t)$ model except that all model parameters are estimated, and the storm i effective rainfall distribution over R is estimated.

Available for model calibration purposes is a rain gauge site which is monitored such as to produce an effective rainfall distribution, and a stream gauge located at the downstream point of R.

Using the above three model structures for R, the modeling calibration process can be analyzed as to parameter calibration efficiency, reliability, and rationality.

HYDROLOGIC MODEL STRUCTURE

Each of the above three models and modeling assumptions are described in detail in Hromadka[2], and will not be fully redeveloped herein.

The catchment R is subdivided into m nearly homogeneous subareas, R_j, such that the effective rainfall distribution in R_j is given by $e_j^i(t)$ for storm i where

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

where λ_{jk}^i are coefficients, θ_{jk}^i are timing offsets, and $e_g^i(t)$ is the effective rainfall distribution measured at the rain gauge site for storm event i.

The runoff hydrograph from R_j for storm i is $Q_j^i(t)$ where

$$Q_j^i(t) = \int_{s=0}^{t} e_j^i(t - s) \, \phi_j^i(s) \, ds \qquad (2)$$

where $\phi_j^i(s)$ is the storm i UH for R_j. Because $\phi_j^i(s)$ is variable between storms, and correlates $Q_j^i(t)$ to $e_j^i(t)$, then whether a UH model or another runoff model is used has no major effect in this analysis. Obviously subarea rainfall-runoff data would be needed to evaluate $e_j^i(t)$ and $Q_j^i(t)$, and hence $\phi_j^i(s)$.

Assuming initially that channel flow routing storage effects are minor, the true runoff hydrograph model for the

response from R from storm i is given by $Q_m^i(t)$ where

$$Q_m^i(t) = \sum_{j=1}^{m} Q_j^i(t - \tau_j^i) \qquad (3)$$

where τ_j^i is the sum of link translation times from R_j to the stream gauge. The τ_j^i will vary between storms, and hence stream gauges would be required on the links in order to evaluate the respective travel times for each link, for each storm, i.

Combining Eqs. (1) and (2) gives the R_j runoff contribution from storm i,

$$Q_j^i(t) = \int_{s=0}^{t} \sum_{k=1}^{n_j^i} \lambda_{jk}^i e_g^i(t - \theta_{jk}^i - s) \, \phi_j^i(s) \, ds \qquad (4)$$

where the $\phi_j^i(s)$ is the same UH used in Eq. (2). Again, rainfall-runoff data in R_j would be needed to evaluate the several λ_{jk}^i and θ_{jk}^i, as well as the $\phi_j^i(s)$ from Eq. (2).

Combining Eqs. (3) and (4) gives the "true" runoff hydrograph model for storm i (for translation routing),

$$Q_m^i(t) = \sum_{j=0}^{m} \int_{s=0}^{t} \sum_{k=1}^{n_j^i} \lambda_{jk}^i \, e_g^i(t - \theta_{jk}^i - s) \, \phi_j^i(s - \tau_j^i) \, ds \qquad (5)$$

Rewriting the timing of the subarea effective rainfall distribution and the subarea UH simplifies Eq. (5) to be

$$Q_m^i(t) = \int_{s=0}^{t} e_g^i(t - s) \sum_{j=1}^{m} \sum_{k=1}^{n_j^i} \lambda_{jk}^i \, \phi_j^i(s - \tau_j^i - \theta_{jk}^i) \, ds \qquad (6)$$

Equation (6) is the output from a m-subarea link-node model of R, with the effective rainfall distribution properly defined (by Eq. (1)) in each individual R_j; and with variable subarea UH's, $\phi_j^i(s)$; and with variable translation flow routing travel times for each storm, as summed in the τ_j^i.

Given only the effective rainfall distribution at the rain gauge site, $e_g^i(t)$, and the stream gauge measured runoff hydrograph, $Q_g^i(t)$, and assuming the effective rainfall distribution on each subarea R_j is given by Eq. (1), and assuming that channel link flow routing is given by translation with a storm-dependent travel time, then for storm event i it is assumed that

$$Q_g^i(t) = Q_m^i(t) \qquad (7)$$

A single area UH model, $Q_1^i(t)$, correlates $\{e_g^i(t), Q_g^i(t)\}$ by the correlation distribution $\eta^i(s)$ where

$$Q_1^i(t) = \int_{s=0}^{t} e_g^i\ (t - s)\ \eta^i(s)\ ds \qquad (8)$$

Thus for the above assumptions, $Q_m^i(t) = Q_1^i(t)$ where

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

Hence the standard single area UH model, $Q_1^i(t)$, and the link node model $Q_m^i(t)$ with subarea rainfall-runoff data and link runoff data sufficient to define all submodel parameters, are equivalent, assuming the subarea effective rainfalls are related to the rain gauge measured $e_g^i(t)$, by Eq. (1).

Equation (6) can be extended to include linear unsteady flow routing (Hromadka[2]), but the results of Eqs. (7) and (8) would still be valid. Because of the reduction in mathematical notation, only translation flow routing effects will be considered hereafter.

MODEL PARAMETER CALIBRATION

From the above development it is seen that $Q_1^i(t) = Q_m^i(t)$, where $Q_m^i(t)$ is the "true" model.

In practice, an estimator of $Q_m^i(t)$, denoted by $\hat{Q}_m^i(t)$, is used where

$$Q_m^i(t) = \sum_{j=1}^{m} \int_{s=0}^{t} \hat{e}_j(t - s)\ \hat{\phi}_j^i(s - \hat{\tau}_j^i)\ ds \qquad (10)$$

where hats are notation for estimates. The subarea UH (or kinematic wave (KW) overland flow planes) are empirically defined without rainfall-runoff data. Additionally, the channel flow routing parameters, represented by the $\hat{\tau}_j^i$, are also estimated. It is noted that in Eq. (10), the $\hat{\phi}_j^i$ and $\hat{\tau}_j^i$ may still be variable between storms in order to approximate a specific modeling technique. It is also recalled that detention or backwater effects are assumed to be minor (Hromadka[2]).

Probably most important in Eq. (10), the R_j effective rainfalls are estimates as only $e_g^i(t)$ data is available. Indeed, the usual practice is to assign the same storm rainfall to each R_j which is measured at the rain gauge, $P_g^i(t)$. Rewriting the estimator, $\hat{Q}_m^i(t)$, in a form comparable to Eq. (6), and recalling Eq. (1),

$$\hat{Q}_m^i(t) = \sum_{j=1}^{m} \int_{s=0}^{t} \hat{\lambda}_j\ e_g^i(t - s)\ \hat{\phi}_j^i(s - \hat{\tau}_j^i)\ ds \qquad (11)$$

where the $\hat{\lambda}_j$ are constants, independent of the storm event i, and reflect only the assumed nonhomogeneity in loss rates between subareas. Obviously, the variability in $\hat{\lambda}_j$ due to storm magnitude, and also variations in storm timing (e.g., $_{jk}^i$), are all unknown.

Rewriting Eq. (11) again,

$$\hat{Q}_m^i(t) = \int_{s=0}^{t} e_g^i(t - s) \sum_{j=1}^{m} \hat{\lambda}_j\ \hat{\phi}_j^i(s - \hat{\tau}_j^i)\ ds \qquad (12)$$

which is similar in form to the standard single area UH model. Writing Eq. (12) in the form of Eq. (8) gives

$$\hat{Q}_m^i(t) = \int_{s=0}^{t} e_g^i(t - s)\ \hat{\eta}^i(s)\ ds \qquad (13)$$

where $\hat{\eta}^i(s)$ is the estimated correlation distribution between $e_g^i(t)$ and $Q_g^i(t)$.

In comparing the "true" model $Q_m^i(t)$ to the estimator $\hat{Q}_m^i(t)$ of Eqs. (6) and (10), respectively, it is seen that the differences occur in the correlation distributions of $\eta^i(s)$ and $\hat{\eta}^i(s)$, respectively.

In Hromadka and Yen[3], all storms were assumed categorized into storm classes $\langle\xi_q\rangle$ such that any two elements (storms) in a specific class $\langle\xi_o\rangle$ would result in nearly identical effective rainfall distributions at the rain gauge site such that one would expect nearly identical runoff hydrographs from the catchment, R. Consequently in a predictive mode, where a design storm effective rainfall $e_g^D(t)$ is assumed to occur at the rain gauge site, the storm class $\langle\xi_o\rangle$ is identified such that $e_g^D(t)$ is in $\langle\xi_o\rangle$ and the resulting model prediction is the random variable

$$[Q_1^D(t)] = \int_{s=0}^{t} e_g^D(t - s)\ [\eta_o(s)]\ ds \qquad (14)$$

where $[\eta_o(s)]$ is the distribution of correlation distributions (see Eq. (8)) in correlating rainfall-runoff information for storms in class $\langle\xi_o\rangle$. It is recalled that all elements in $\langle\xi_o\rangle$ result in nearly an identical effective rainfall distribution; consequently, $e_g^D(t)$ being in $\langle\xi_o\rangle$ implies that each storm in $\langle\xi_o\rangle$ results in an effective rainfall distribution, $e_g^o(t)$, which nearly duplicates $e_g^D(t)$.

Now compare the models $Q_m^i(t)$ and $Q_m^i(t)$ for elements $e_g^o(t) \in \langle\xi_o\rangle$. From Eqs. (8), (9), and (14),

$$[Q_m^o(t)] = \int_{s=0}^{t} e_g^o(t - s)\ [\eta_o(s)]\ ds \qquad (15a)$$

whereas from Eqs. (12) and (13),

$$\hat{Q}_m^o(t) = \int e_g^o(t - s)\,[\hat{\eta}_o(s)]\,ds \qquad (15b)$$

But the $\hat{\eta}_o(s)$ must be nearly identical distributions for any element of $<\xi_o>$ due to the imprecise (constant) definition of effective rainfalls in each R_j. This "rigidity" in $\hat{\eta}_o(s)$ becomes apparent when studying severe storms of flood control interest, where nonlinear effects approach linearity (Hromadka[2]). Hence for any element $e_g^o(t)$ in $<\xi_o>$, Eq. (15b) is

$$\hat{Q}_m^o(t) = \int_{s=0}^{t} e_g^o(t - s)\,\eta_o(s)\,ds \qquad (16)$$

That is, the estimator $\hat{Q}_m^o(t)$ would predict nearly identical responses for runoff given nearly identical effective rainfall distributions occurring at the rain gauge site.

But due to the unknown variations in rainfall over R, among other factors, the runoff hydrographs measured at the stream gauge do vary for nearly identical rain gauge measurements. As a result, the unknown variations in effective rainfall are propogated differently between the single area model, $Q_1^o(t)$, and the link-node model estimator $\hat{Q}_m^o(t)$, during the parameter calibration process.

For the $Q_1^o(t)$ model, the correlation of $Q_g^o(t)$ to $e_g^o(t)$ gives a distribution which is a sample from the random variable, $[\eta_o(s)]$.

For the $\hat{Q}_m^o(t)$ model, however, the calibration effort results in a distribution for the model's effective rainfall estimator, now denoted by $\hat{e}_g^o(t)$ for storm class $<\xi_o>$.

Hence due to the rigidity in $\hat{\eta}^o(s)$ in neglecting the variations in effective rainfall over R, these variations must be transferred to the parameters used to model $\hat{e}_g^o(t)$ in $\hat{Q}_m^o(t)$. Hence, $[\hat{Q}_m^o(t)]$ essentially becomes for storm class $<\xi_o>$,

$$[\hat{Q}_m^o(t)] = \int_{s=0}^{t} [\hat{e}_g^o(t - s)]\,\hat{\eta}_o(s)\,ds \qquad (17)$$

where the variations in $[\hat{e}_g^o(t)]$ produce a minor variation in $\hat{\eta}_o(s)$.

THE CALIBRATION PROCESS

In calibrating $Q_1^o(t)$, the data $Q_g^o(t)$ and $e_g^o(t)$ is used to determine $[\eta_o(s)]$.

In calibrating $\hat{Q}_m^o(t)$, the data $Q_g^o(t)$ and the rigid $\hat{\eta}_o(s)$ is used to determine a best fit $\hat{e}_g^o(t)$. But the $\hat{Q}_m^o(t)$ submodel structure used to compute the effective rainfall typically cannot fit the derived $\hat{e}_g^o(t)$. As a result, the $\hat{e}_g^o(t)$ becomes a best fit to the integrated effects of the set $\{e_g^o(t), \lambda_{jk}^i, \theta_{jk}^i, n_j^i\}$ as well as errors in the parameters $\{\phi_j^i(s), \tau^i\}$. Consequently, the calibrated $\hat{Q}_m^o(t)$ must have more error than the $Q_1^o(t)$ model, and cannot achieve the correct uncertainty distribution of output due to the structure of the $\hat{e}_g^o(t)$ function.

To demonstrate the above discussion, a 25-subarea link-node model of an idealized catchment is used which satisfies the several assumptions leading to $Q_m^i(t)$, (see Fig. 1). The single "available" rain gauge is shown as a triangle in Fig. 1. Not shown in Fig. 1 are subarea-centered rain gauges and downstream stream gauges which are used for $Q_m^i(t)$, but are "unavailable" to the estimator, $\hat{Q}_m^i(t)$. The catchment, R, is 500 acres in size, with each R_j being 20 acres. All channel links are rectangular channels with dimensions of depth = 20-feet, width = 5-feet, slope = 0.01 ft/ft, and a Mannings friction factor of 0.015.

Each subarea has its own UH which is assumed to be a function of its time of concentration, Tc, as shown in Fig. 2. Each subarea is assumed to have a uniform loss rate. The rain gauge site is monitored to determine the effective rainfall, $e_g^o(t)$, (Fig. 1).

To evaluate the calibration process, a series of identical effective rainfall distributions (i.e., storms $e_g^o(t)$) are defined at the rain gauge site, which satisfy that each storm is in the same storm class, $<\xi_o>$. For $\hat{Q}_m^o(t)$, the subarea effective rainfalls are assumed related to the $e_g^o(t)$ by the factors $\hat{\lambda}_j$ listed in Table 1. Other parameter data is also listed in this table. The "true" distributions of $e_j^o(t)$ are random variables distributed according to Fig. 3 for λ_{jk}^o, and Fig. 4 for timing offsets, θ_{jk}^o, where mean values are listed in Table 1. The "true" runoff hydrographs are developed for each storm using $Q_m^i(t)$.

Because $e_g^o(t)$ is fixed, the $\hat{Q}_m^o(t)$ model must have a fixed output. Therefore, because $\hat{\eta}_o(s)$ is fixed, a least-squares best fit for $\hat{e}_g^o(t)$ can be developed for each storm in $<\xi_o>$. Some of the resulting plots of effective rainfall distributions are shown in Fig. 5.

For $Q_1^o(t)$, the variations in $e_j^o(t)$ are reflected in the $\eta_o(s)$ variations. Some of the elements of the set $\{\eta_o(s)\}$ are shown in summation (distribution) graph form in Fig. 6.

From Fig. 5, the set of $\hat{e}_g^o(t)$ plots needed to correlate the $Q_g^o(t)$ to the single $\hat{\eta}_o(s)$ cannot be duplicated by a fixed loss rate model structure because the storm precipitation is identical for each event and, therefore, a loss in accuracy must occur during parameter calibration. Additionally, the

TABLE 1. APPLICATION PROBLEM DATA

Subarea R_j	Tc^1	$\lambda_j{}^2$	$\lambda_{jk}{}^{o3}$	$\theta_{jk}{}^{o4}$
1	30	1	1	0
2	30	1	1	0
3	45	1	1	0
4	45	1.1	1.1	5
5	30	1.1	1.1	5
6	30	.9	.9	5
7	45	.8	.8	5
8	30	.8	.8	5
9	30	.7	.7	5
10	30	.7	.7	5
11	45	.8	.8	10
12	45	1.	1.	10
13	45	1.	1.	10
14	45	1.3	1.3	10
15	30	1.3	1.3	10
16	30	1.2	1.2	10
17	45	1.2	1.2	10
18	30	1.1	1.1	10
19	30	1.1	1.1	10
20	45	1.	1.	10
21	30	1.	1.	10
22	30	1.	1.	10
23	30	.9	.9	10
24	45	.9	.9	10
25	45	.8	.8	10

Notes:

1. Tc = time of concentration in minutes
2. $\hat{\lambda}$ = assumed ratio of effective rainfall at subarea to rain gauge site
3. $\hat{\lambda}_{jk}$ = mean value for $\lambda_{jk}{}^i$. Note that $\bar{\lambda}_{jk} = \hat{\lambda}_j$
4. $\bar{\theta}_{jk}$ = mean value for $\theta_{jk}{}^i$, in minutes

final calibrated parameters lose some of the physical meaning for what they were intended, in that they reflect variations in effects other than the loss rate.

In Fig. 6, however, the resulting $\eta_o(s)$ plots (summation graph form) are used to populate a frequency distribution for $[\eta_o(s)]$, to develop the uncertainty distribution for $[Q_1{}^o(t)]$ using $e_g{}^o(t)$ as the model input.

It is noted that in this application, the estimated $\hat{\lambda}_j$ are assumed "correctly" in that the $\hat{\lambda}_j$ equal the mean value of $\lambda_{jk}{}^i$ (see Table 1). Hence in actual applications, the discrepancies between $\hat{e}_g{}^o(t)$ and $e_g{}^o(t)$ shown in Fig. 5 could be augmented.

DISCUSSION

The application demonstrates how the unknown effective rainfall distribution manifests itself in the single area UH, $Q_1{}^i(t)$, model, and in a discretized link node model, $\hat{Q}_m{}^i(t)$, when using storms of a similar class to calibrate model parameters. For the $Q_1{}^i(t)$ model, the uncertainties are incorporated into the UH correlation distribution, $\eta^i(s)$. In the estimator, $\hat{Q}_m{}^i(t)$, however, the uncertainties are transferred to the effective rainfall submodel parameters used in $\hat{e}_g{}^i(t)$.

Because the $\eta^i(s)$ are allowed to freely vary, the frequency distribution $[\eta(s)]$ of the $\eta^i(s)$ reflect the several modeling uncertainties as well as the important uncertainty in the effective rainfall distribution over R.

With the estimator, $\hat{Q}_m{}^i(t)$, however, the effective rainfall estimator, $\hat{e}_g{}^i(t)$, is usually a fixed model structure which cannot fit the irregular effective rainfall distributions needed to correlate measured runoff data to the $Q_m{}^i(t)$ model UH correlation distribution, $\hat{\eta}^i(s)$. As a result, the calibration of $\hat{e}_g{}^i(t)$ must be imprecise and, therefore, the $\hat{Q}_m{}^i(t)$ must be a more uncertain model in the predictive mode than the $Q_1{}^i(t)$ model.

CONCLUSIONS

The unit hydrograph method is used to evaluate how modeling uncertainty is propogated through a single area UH model and a discretized link node model estimator. It is shown that the modeling uncertainties and the important uncertainty in the effective rainfall distribution over the catchment are manifested in the single area model unit hydrograph, whereas they are manifested in the link node model estimator's effective loss rate function. Because the effective loss rate submodel is of a prescribed structure, the calibration of the loss rate submodel of the link node model will result in imperfect fits of the effective rainfall distributions needed to correlate the measured runoff data to the hydraulic response function.

In comparison, calibration of the single area UH model results in the several uncertainties (including hydraulic response uncertainties and the unknown distribution of effective rainfall over the catchment) being integrated into the UH, with the loss rate submodel being calibrated to loss rate information.

REFERENCES

1. Hromadka II, T.V. and Whitley, R.J. (1988), The Design Storm Concept in Flood Control Design and Planning, Stochastic Hydrology and Hydraulics, in-press.

2. Hromadka II, T.V. (1988), Back to the Unit Hydrograph Method, Proceedings: Envirosoft 88 Conference, Computational Mechanics, Porto Carras, Greece.

3. Hromadka II, T.V. and Yen, C.C. (1988), Unit Hydrograph Models and Uncertainty Distributions, Proceedings: Environsoft 88 Conference, Computational Mechanics, Porto Carras, Greece.

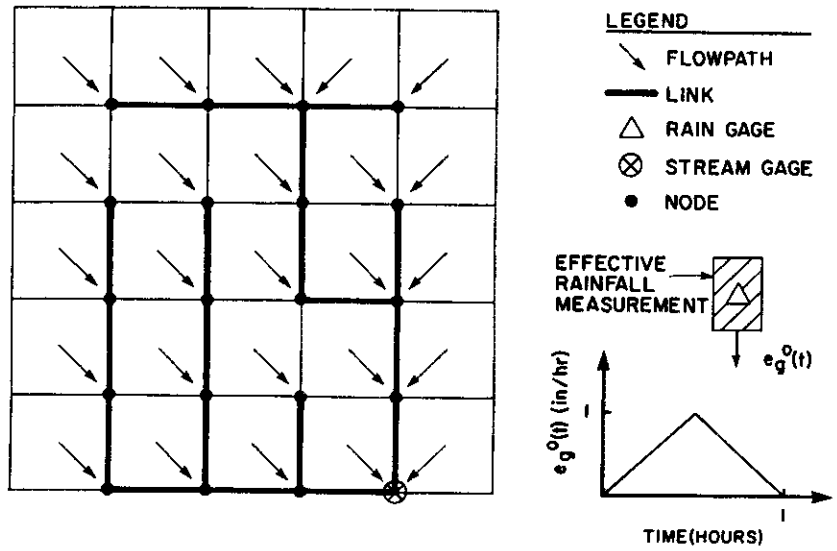

Fig. 1. Application Problem Schematic.

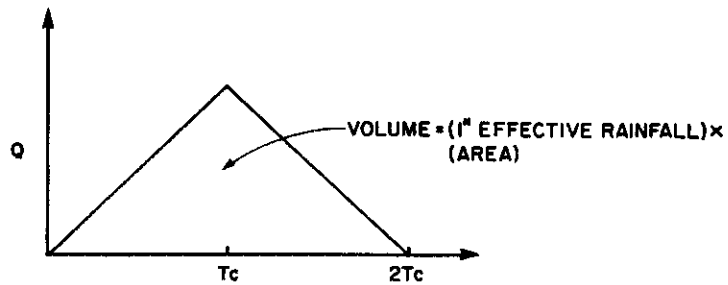

Fig. 2. Subarea Unit Hydrograph.

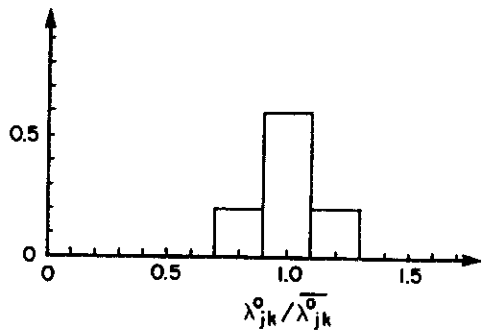

Fig. 3. Frequency Distribution for $\lambda^o_{jk}/\overline{\lambda^o_{jk}}$ (see table 1 for subarea $\overline{\lambda^o_{jk}}$).

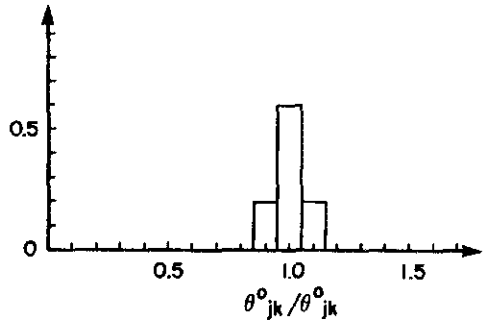

Fig. 4. Frequency Distribution for Θ^o_{jk} (see table 1 for subarea $\overline{\Theta^o_{jk}}$).

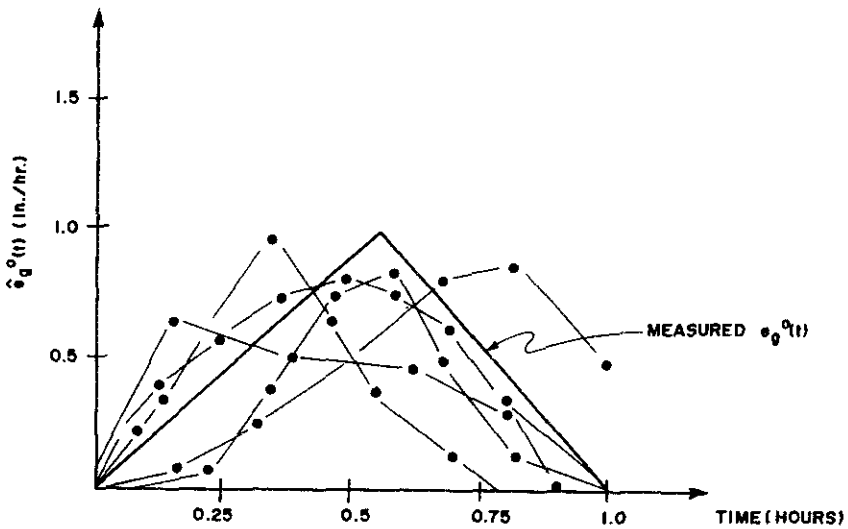

Fig. 5. Some Least-Squares Best Fit Effective Rainfall Distributions, $\hat{e}_g^o(t)$ which Correlate Measured Runoff, $Q_g^o(t)$, to the $\hat{n}_o(s)$ used in the Estimator, $\hat{Q}_m^o(t)$, for the Applicaton Problem.

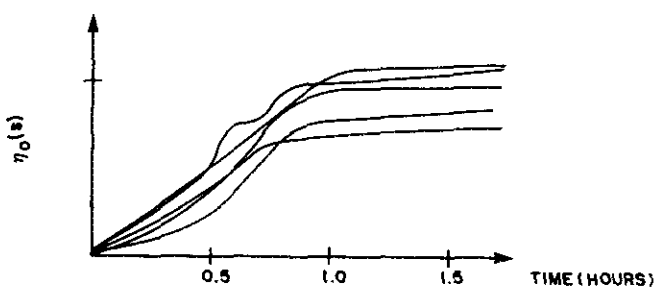

Fig. 6. Some Correlation Distributions $\eta_o(s)$, in Correlating $Q_g^o(t)$ and $e_g^o(t)$ for the Application Problem, Plotted in Summation (Distribution) Graph Form.