Approximating rainfall-runoff modelling uncertainty using the stochastic integral equation method

T. V. Hromadka II

Water Resources Engineering, Williamson and Schmidt, 17782 Sky Park Boulevard, Irvine, CA 92714, USA; and Applied Mathematics Department, California State University, Fullerton, CA 92634, USA

Although the uncertainty in rainfall-runoff model predictions has been frequently demonstrated, little attention has been paid towards including uncertainty in rainfall-runoff models, or providing a methodology that not only accommodates the uncertainty, but can be readily integrated into most rainfall-runoff modelling structures. In this paper, the uncertainty problem is approached by providing a methodology which can be incorporated into almost all rainfall-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. The stochastic integral formulation is used to represent the total variation between a record of measured runoff data and model estimates.

Because of the simplicity of the technique, the stochastic integral equation method can be integrated into most currently available rainfall-runoff models. The method provides the capability to develop probability distributions of almost all criterion variable estimates as produced from almost any rainfall-runoff modelling approach.

INTRODUCTION

The most recent literature regarding rainfall-runoff hydrologic models has demonstrated the considerable magnitude of uncertainty in modelling predictions (e.g., Loague and Freeze; Schilling and Fuchs; among other references). Hromadka et al. and Hromadka and McCuen provide an extensive review of the rainfall-runoff modelling literature which includes several papers and reports which examine the large errors in modelling predictions. Due to the nondeterministic nature of the rainfall-runoff processes occurring over the catchment, the mathematical descriptions of these processes result in stochastic equations. Additionally, the so-called deterministic surface runoff models used to describe the several physical processes contain parameters or coefficients which have well-defined physically-based meanings, but whose exact values are unknown. The boundary conditions of the problem itself are unknown (e.g., the temporal and spatial distribution of rainfall over the catchment for the storm event under study and also for all prior storm events) and often exhibit considerable variations with respect to the assumed boundary conditions. Thus the physically-based parameters and coefficients, and also the problem boundary conditions, are not given by the assumed values but are random variables and stochastic processes whose variations about the assumed values are governed by certain probability distributions. Consequently, due to the significant errors in rainfall-runoff modelling estimates reported in the literature, it is more realistic to formulate a stochastic model of the rainfall-runoff modelling error rather than to assume a totally deterministic model (of the rainfall-runoff phenomena) which entirely neglects the significant random error contribution.

Although the uncertainty problem has been frequently demonstrated, little attention has been paid towards including uncertainty in rainfall-runoff models, or providing a methodology to accommodate the uncertainty.

In this paper, the uncertainty problem is approached 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 life sciences and chemical engineering (e.g., Tsokos and Padgett, provide a thorough development). The stochastic integral formulation is used to represent the total variation between a record of measured runoff data and the corresponding model estimates.

The methodology developed in this paper focuses only upon the problem where a future estimate of a criterion variable (e.g., peak flow rate, or volume of runoff, or average flow velocity, or cost, etc.) corresponding to a hypothetical storm event is needed at a stream gauge, assuming only the data (for the subject storm) given at a single rain gauge (i.e., the single rain gauge/stream gauge problem). That is, the stochastic integral equation method answers the question: 'based upon the historic rainfall-runoff data record and the model's accuracy in estimating runoff, what is the distribution of values of the subject criterion variable given a hypothetical rainfall event?'. The generalization of the methodology to
accommodate problems where several rain gauges or several stream gauges are available, or where either the runoff data or rainfall data are unavailable, involves the concept of regionalizing the stochastic integral equation transfer function distribution of realizations, and is the subject of a companion paper (although the topic of regionalization is briefly alluded to in this paper).

STOCHASTIC INTEGRAL EQUATION METHOD (SIEM) FOR RAINFALL-RUNOFF MODELS

Let \( \omega \) be an element of the probability space, \( \Omega \), with probability \( P(\omega) \). In our problem setting, \( \omega \) is a storm event occurrence which has associated precipitation, runoff, and other effects. Let \( R \) be the study catchment which freely drains to a single stream gauge (\( R \) is assumed to be reasonably homogeneous in loss rates over the catchment (such that a lumping of the loss rates into a single loss rate parameter set is appropriate), and have a free-draining collector channel system which exhibits negligible backwater effects, such as is assumed in almost all rainfall-runoff models in common use today). The key hydrologic assumptions employed in this analysis focuses upon three points as follows: (1) there are no detention effects – i.e., the catchment does not contain any storage elements such as dams, basins, or significant channel infiltration effects; (2) channel routing effects are such that unit water flow routing models in common use (e.g., kinematic wave, convex Muskingum) are employed; and (3) the basin is homogeneous such that a basin-averaged loss rate (i.e., 'lumped' parameter) loss function is appropriate. These assumptions, although restrictive, occur frequently in practice. Additionally, the error analysis techniques discussed in the following can be extended by subdividing the catchment into homogeneous regions, if needed, should data be available. However in this paper, the above hydrologic assumptions are utilized in order to make a clearer presentation of the mathematical underpinnings employed. Should the above hydrologic assumptions not be satisfied, the presented error analysis procedures may be inappropriate. A single rain gauge is available for data analysis. For a given \( \omega \), the measured rainfall and runoff are the realizations \( P(t; \omega) \) and \( Q(t; \omega) \), respectively, where \( t = 0 \) is time.

Let \( M \) be a surface runoff model which operates on the rain gauge data to produce an estimate of runoff at the stream gauge for event \( \omega \) by

\[
M = P(t; \omega) \rightarrow M(t; \omega)
\]  

Then for event \( \omega \), the measured runoff, \( Q(t; \omega) \), and the model estimate, \( M(t; \omega) \), are related by the stochastic integral equation:

\[
Q(t; \omega) = M(t; \omega) + \int_0^t k_1(t-s; \omega)h(s, M(s; \omega)) \, ds
\]

where \( k_1(\cdot; \omega) \) and \( k_2(\cdot; \omega) \) are functions of time for event \( \omega \), and \( h(\cdot; k_1(\cdot; \omega)) \) is a correlation between the \( k_1(\cdot; \omega) \) and the total model error, \( Q(\cdot; \omega) - M(\cdot; \omega) \).

From equation (2), \( h(\cdot; k_2(\cdot; \omega)) \) depends on \( k_2(\cdot; \omega) \). In this study, the correlation distribution, \( h(\cdot; k_2(\cdot; \omega)) \), is assumed to be highly dependent upon the model estimate, \( M(\cdot; \omega) \), for event \( \omega \). Other possible choices for \( k_2(\cdot; \omega) \) exist, but only \( M(\cdot; \omega) \) is considered in this paper. Thus

\[
h^*(\cdot; k_2^*(\cdot; \omega)) = h(\cdot; M(\cdot; \omega))
\]

Additionally, the integral of equation (2) is assumed to depend only upon prior data in the kernel in that given equation (3),

\[
Q(t; \omega) = M(t; \omega) + \int_0^t k_1(t-s; \omega)h(s, M(s; \omega)) \, ds
\]

Suitable choices of \( k_1(\cdot; \omega) \) are \( P(\cdot; \omega) \), \( M(\cdot; \omega) \), or even the mean effective rainfall (rainfall less losses, rainfall excess) over the catchment, \( F(\cdot; \omega) \). Selection of the \( k_1(\cdot; \omega) \) is the subject of current research, but is similar to the concept of regression analysis using various transformations of the independent variable.

For a reasonably homogeneous catchment, the mean effective rainfall over \( R \) for event \( \omega \) is given by the operator \( F \) where \( F: P(\cdot; \omega) \rightarrow F(\cdot; \omega) \). Using \( F(\cdot; \omega) \) in equation (4),

\[
Q(t; \omega) = M(t; \omega) + \int_0^t F(t-s; \omega)h(s, M(s; \omega)) \, ds
\]

The choice of \( F(\cdot; \omega) \) in equation (5) is made for the current study due to convenience only.

Equation (5) provides for a method of including uncertainty in runoff predictions. For \( m \) events, \( m \) sets of realizations \( \{P(\cdot; \omega), Q(\cdot; \omega)\} \) are available. Given the model structure \( M \) and the mean effective rainfall estimator \( F \) used in \( M \), then unique \( M(\cdot; \omega) \) and \( F(\cdot; \omega) \) realizations are available for each event, \( \omega \). Consequently, the \( h(\cdot; M(\cdot; \omega)) \) can be categorized into equivalence classes depending on some prescribed characteristics of \( M(\cdot; \omega) \), say \( C(M(\cdot; \omega)) = Z \), where \( Z \) is a vector of characteristic values of \( M(\cdot; \omega) \); for example, peak flow rate, peak 1-hour average flow rate, etc. That is, for class \( M_2 \),

\[
M_2 = \{M(\cdot; \omega); C(M(\cdot; \omega)) = Z\}
\]

where \( C \) is the characteristic definition operator.

Each equivalence class, \( M_2 \), is composed of several events such that \( C(M(\cdot; \omega)) = Z \), and it is assumed that the associated correlations, \( h(\cdot; M_2(\cdot; \omega)) \), are all equally likely to occur for a future event, \( \omega_0 \), such that \( C(M(\cdot; \omega_0)) = Z \). In order to develop statistical significance, the equivalence classes, \( M_2 \), may be required to include a wide range of events, so that a reasonable number of realizations are elements of the class.

The realizations, \( h(\cdot; M_2(\cdot; \omega)) \), are all readily determined analogous to standard convolution methods (e.g., unit hydrograph techniques) by noting

\[
\int_0^t F(t-s; \omega)h(s, M(s; \omega)) \, ds = Q(t; \omega) - M(t; \omega)
\]

In the following development, the correlation \( h(\cdot; M(\cdot; \omega)) \) will be simply written as \( h(\cdot; Z) \) where it is understood that \( Z \) indicates that \( C(M(\cdot; \omega)) = Z \), and \( h(\cdot; Z) \) is a realization of the stochastic process associated to equivalence class \( M_2 \). Thus equation (5) is written as the stochastic integral equation. (again, see Tsokos and...
Padgett\textsuperscript{18}, for details regarding uniqueness, existence, and other analysis,

\[ Q(t; \omega) = M(t; \omega) + \int_0^t F(t-s; \omega) h(s, Z_0) \, ds; \]
\[ C(M(\cdot; \omega)) = \mathcal{Z} \]  

(8)

From equation (8), it is seen that the integral term represents the total modelling error. The formulation implies that the total modelling error is linear on an equivalence class basis. And in prediction, where the model is used to predict the runoff at the stream gauge for a future event, \( \omega_0 \), then equation (8) develops a distribution of outcomes (realizations) of runoff distributed as \([Q(\cdot; \omega_0)]\) where

\[ [Q(t; \omega_0)] = M(t; \omega_0) + \int_0^t F(t-s; \omega_0) [h(s, Z_0)] \, ds \]  

(9)

where \([h(\cdot, Z_0)]\) is the distribution of realizations of correlations derived from all prior events, \( \omega \), such that \( C(M(\cdot; \omega)) = \mathcal{Z} \).

In practice, the runoff model is calibrated such that

\[ E[Q(t; \omega_0)] \equiv M(t; \omega_0) \]  

(10)

where \( E \) is the usual expected value operation. Then the variance of the total modelling error, \( \sigma_g(t) \), in prediction for future event \( \omega_0 \) is estimated by

\[ \sigma_g^2(t) \approx \frac{1}{m-1} \sum_{i=1}^{m} \left( \int_0^t F(t-s; \omega_0) h(s, Z_0) \, ds \right)^2, \quad t > 0 \]  

(11)

where \( m \) realizations are available in the equivalence class \( M_{Z_0} \) and \( h(s, Z_0) \) is a sample realization from the associated set of correlations. If the variance \( \sigma_g(t) \) is correlated to catchment characteristics, then a regionalization of the variance can be prepared (analogous to unit hydrograph methods) in order to transfer the information to other catchments. Obviously, equations (10) and (11) are subject to the usual sampling error considerations that are associated with all statistical estimators.

From equations (9)–(11), it is seen that all predictions depend strongly upon the selected model structure, \( M \), and the associated mean loss function, \( F \). However, should \( k_M(\cdot; \omega) = M(\cdot; \omega) \) in equation (4), then predictions depend only on the model estimator.

DEVELOPING DISTRIBUTIONS OF CRITERION VARIABLE PREDICTIONS USING THE SIEM

Equation (9) can now be used to develop predictions of a criterion variable for a future storm event, \( \omega_0 \). Let \( A \) be a criterion variable of interest such as peak flow rate at the stream gauge, or mean flow velocity for the peak 1-hour of runoff, etc. Then for event \( \omega_0 \),

\[ A(\omega_0) = A(Q(\cdot; \omega_0)) \]  

(12)

and in prediction for some event \( \omega_0 \), \( A(\omega_0) \) is a random variable distributed as \([A(\omega_0)]\) where from equations (9), the SIEM gives the estimate

\[ [A(\omega_0)] = [A(Q(\cdot; \omega_0))] \]  

(13)

where in equation (13), \([A(Q(\cdot; \omega_0))]\) is notation for operating on each sampled runoff realization to derive a sampling of the criterion variable value, \( A(\omega_0) \).

For example, should \( A(\omega_0) = \) be the peak flow rate from event \( \omega_0 \), then

\[ A(\omega_0) \equiv \max \{Q(t; \omega_0)\} \]  

(14)

And in prediction for future event \( \omega_0 \), the peak flow rate is a random variable distributed as

\[ [A(\omega_0)] = \max \{[Q(t; \omega_0)]\} \]  

(15)

Given \( m \) realizations in equivalence class \( M_{Z_0} \), then in equation (15), \([A(\omega_0)]\) is the frequency-distributed of peak flow rate values, \( \max \{Q(t; \omega_0)\} \), for \( i = 1, 2, \ldots, m \), where from equations (9) and (11),

\[ Q(t; \omega_0) = M(t; \omega_0) + \int_0^t F(t-s; \omega_0) h(s, Z_0) \, ds \]  

(16)

Example problem

To demonstrate the use of the SIEM, a fully urbanized watershed in Los Angeles, California is considered. The subject watershed has three continuous recording rain gauges available, and also a stream gauge located in a large concrete flood control channel. The 13 square-mile catchment contains densely developed residential and commercial lots over 98-percent of the land area, and is served by a fully improved storm drain and collector channel system. Of interest are hydrologic predictions of storm runoff occurring at the stream gauge.

The rainfall-runoff model estimator used, \( M \), is composed of 93 subareas, each of near-equal size, and linked to respective storm drain or channel hydraulic elements. Because of the relatively steep gradients involved, all unsteady flow routing processes are assumed adequately modelled by the diffusion method (i.e., zero-inertia) for flow routing\textsuperscript{8}.

Subarea runoffs are approximated by use of the standard SCS unit hydrograph\textsuperscript{11} where effective rainfall (i.e., rainfall less losses, rainfall excess) is estimated in each subarea \( R_j \) by use of the loss function \( f_j(t) \) where

\[ f_j(t) = f_{0,j} + (f_{0,j} - f_{\infty,j}) e^{-k_j t} \]  

(17)

where \( f_{0,j} \) and \( f_{\infty,j} \) are initial and ultimate loss rates, respectively; and \( k_j \) is a timing parameter chosen such that \( f_j(t) \) is within 5-percent of \( f_{\infty,j} \) within 30-minutes of storm time. The values of \( f_{0,j} \) and \( f_{\infty,j} \) are estimated based upon the subarea percent impervious area, and soil-water percolation tests conducted by the local flood control agency. A uniform initial abstraction of 0.10 inches of rainfall is assumed for the entire catchment, which is also assumed to be fully recovered after one day of no rainfall. Subarea unit hydrograph timing estimates of time-to-peak, \( T_j \), are developed by summing normal depth travel times for each appropriate channel link used to estimate subarea time of concentration\textsuperscript{11}. Thiessen

polygons are used to partition the catchment into subregions where rainfall data are assumed to apply. For storm event \( \omega \), the model estimate of runoff at the stream gauge is \( M(\cdot; \omega) \), the measured runoff is \( Q(\cdot; \omega) \), and the model error is \( E(\cdot; \omega) \).

In this example, the model error, \( E(\cdot; \omega) \), is correlated to an area-averaged effective rainfall, \( F(\cdot; \omega) \), defined by the sum of effective rainfall:

\[
F(t; \omega) = \sum \left( P_j(t; \omega) - f_j(t; \omega) \right) A_j / \sum A_j \tag{18}
\]

where \( A_j \) is the area of subarea \( R_j \); \( f_j(\cdot; \omega) \) is the subarea loss rate for event \( \omega \); and \( P_j(\cdot; \omega) \) is the assigned subarea rainfall for event \( \omega \). \( F(\cdot; \omega) \) also includes the recoverable initial abstraction of 0.10 inches of rainfall. The transfer function, \( h(\cdot; M(\cdot; \omega)) \), used in the correlation \( \langle E(\cdot; \omega) \rangle \), \( F(\cdot; \omega) \) is developed from the stochastic integral (see equation (5))

\[
E(t; \omega) = \int_0^t F(t - s; \omega) h(s, M(s; \omega)) ds \tag{19}
\]

Thirty seven storm events are considered in this example problem, which represent the most productive runoffs which have occurred over the last twenty years, the catchment has been in a relatively stable development condition during this time period. Consequently, these considered storms form a storm class to be used in the SIEM where \( C(M(\cdot; \omega)) = Z_0 \).

The rainfall-runoff model error for outcome \( \omega \), given in equation (19) as \( E(\cdot; \omega) \), follows directly from equation (16). In equation (16), each realization of measured runoff is compared to the modelled runoff, and a total modelling error is obtained by subtraction. Assuming the total model error realization, \( E(\cdot; \omega) \), to be highly correlated to the basin-averaged effective rainfall (in this example application) realization, \( F(\cdot; \omega) \), implies that a unique realization of the transfer function, \( h(\cdot; \omega) \), exists for each event, \( \omega \). Because we are basing this example error analysis upon effective rainfall, the total space of effective rainfall realizations is partitioned into storm classes of effective rainfalls. The assumption that all of the considered 37 events are elements of the same storm class, \( Z_0 \), implies that each realization of the transfer function, \( h(\cdot; \omega) \), obtained by solving equation (19) for each event \( \omega \), is an element of the distribution of realizations, \( h(\cdot; Z_0) \), for storm class \( Z_0 \). As a result, 37 realizations are defined to be elements of \( h(\cdot; Z_0) \). In this application, each element \( h(\cdot; \omega) \) was computed by solving equation (19) by a least-squares numerical solution.

The realizations of \( h(\cdot; Z_0) \) are approximately shown in Fig. 1 in point density graph form. From the figure, it is seen that the model estimate appears to generally underestimate runoff quantities. By decreasing the subarea loss rates uniformly by 20 percent, the associated error realizations \( E(\cdot; \omega) \) are found to be near equally distributed about zero. The model, now with adjusted subarea loss rates (and hence an adjusted \( F(\cdot; \omega) \)), is to be used for predicting runoff estimates at the stream gauge.

Of concern is the possible runoff at the stream gauge should a hypothetical storm event, \( \omega_0 \), occur outside the subject catchment. It is assumed that \( \omega_0 \) is an element of the storm class used to develop the above \( h(\cdot; Z_0) \). The storm event \( \omega_0 \) considered is the 24-hour design storm pattern described in HEC Training Document 1519.

From the SIEM, the runoff prediction at the stream gauge is the distribution defined by equations (9) and (13).

For the criterion variable \( A \) of peak flow rate resulting from \( \omega_0 \), as measured at the stream gauge, the distribution \( [A] \) is shown in Fig. 2.

Should the stream gauge be the site of a hypothetical dam, with a prescribed outlet rating curve and storage curve, the probability distribution of peak volume demand on the dam for event \( \omega_0 \) is shown in Fig. 3.

Other criterion variable distributions are developed by means of the SIEM as applied in equations (13) and (16). From the various distributions, confidence interval estimates can be readily obtained for decision-making purposes.

Discussion

The above example problem utilized a certain rainfall-runoff model to develop estimates of runoff at the stream gauge, and then the SIEM is used to couple to the model.

---

CONCLUSIONS

The stochastic integral equation method (SIEM) is used to include the uncertainty in rainfall-runoff modelling estimates in the prediction of runoff criterion variable values. The technique is based upon the well-established theory of stochastic integral equations, and can be readily integrated into most currently available rainfall-runoff models. With the magnitude of surface runoff modelling errors demonstrated in the literature, it is important to attempt to include a measure of the uncertainty in the modelling estimates; the SIEM provides a convenient procedure for including modelling uncertainty in runoff criterion variable predictions.

REFERENCES
