FORMALIZATION OF MASS-CONSERVATION RAINFALL-RUNOFF MODELS

T.V. Hromadka II[1] and R.J. Whitley[2]

Abstract

Debate continues over topics involving rainfall-runoff (R-R) model structures and the appropriateness of various model structures in resolving questions regarding storm runoff. Recently, a mathematical formalization was developed for rainfall-runoff models such as included in the computer program HEC-1 and related programs. With this formalization, a precise examination can be made of the algorithmic underpinnings of various R-R modeling structures. In this paper, the formalization is extended to R-R models whose algorithms and processes satisfy the mass conservation equation and also assume storage is a function of outflow such that the first derivative of storage with respect to outflow is positive. The formalization introduces two types of R-R computer model structures, called "Type 1" or "Type 2", that describe almost all R-R computer models in use today. For example, the classic unit hydrograph method is found to be a Type 1 model structure. A R-R model structure that satisfies mass conservation, but is not a Type I model is said, in this paper, to be a Type II model. The formalization is then applied to introduce a procedure useful in deciding which of the Type 1 or Type 2 model structures is "best" for a particular application given a R-R data set for model calibration purposes.

---

1 Professor, Mathematics and Environmental Studies, Department of Mathematics, California State University, Fullerton, California 92634.

2 Professor, Department of Mathematics, University of California, Irvine, California 92616

INTRODUCTION

In Hromadka and Whitley (1998), a mathematical formalization of the link-node modeling system used in the computer program HEC-1 and related computer programs was introduced that provided a mathematical description of these frequently used rainfall-runoff computer-model structures. It was shown that many rainfall-runoff (R-R) computer-model structures involve algorithms and processes that are described by an assemblage of Toeplitz matrices that describe the computer model's transformation of effective rainfall (rainfall less losses) into a runoff hydrograph. For example, the unit-hydrograph-method algorithm for generating subarea runoff and the hydrograph-routing algorithms for convex, Muskingum, translation, convolution and modified Puls (with storage being equal to the product of a constant parameter with outflow, i.e., a linear storage method) have all been shown to be described by Toeplitz matrices. Because the sum and product of Toeplitz matrices is a Toeplitz matrix, the entire modeling network methodology of subdividing a watershed into subareas, combining subarea hydrographs, and routing flood hydrographs through links, can be described by the Toeplitz matrix formalization for many frequently used computer-model structures such as contained in HEC-1. Thus, a classification is possible of R-R computer model structures regarding whether a particular R-R computer model can be mathematically described by a Toeplitz matrix system.

In this paper, an examination is made of R-R computer models that are not resolvable into Toeplitz matrix systems. By expanding upon the basic continuity equation of mass transport in a closed system, another R-R computer model structure class is identified. This new class of R-R model structure is shown to be mathematically described by the sum and product of lower triangular matrices that are not of the Toeplitz type. The new class of model structure provides further insight as to how R-R model algorithms and components integrate, and provides another link to standard mathematical optimization techniques and concepts. Additionally, a procedure is advanced as to how to select a particular type of R-R computer-model structure, between the two model structure types identified in this paper, for use in runoff prediction. It is noted that this paper develops a mathematical description of the R-R computer model itself rather than examining the conceptual model that the R-R computer model attempts to approximate. Therefore, the mathematical formalization provides a precise description of R-R computer model structures, and is not a prescription for formulating a new class of R-R model structure.

The Continuity Equation

The well-known continuity equation of mass transport in a closed system, as applied to storm runoff, is given by

$$I = O + \frac{dS}{dt} \tag{1}$$

where I = runoff inflow rate; O = runoff outflow rate; S = runoff storage; and t is time.

For a small time step Δt, Eq. (1) is typically approximated in R-R computer models by

$$\frac{I_i + I_{i+1}}{2} = \frac{O_i + O_{i+1}}{2} + \frac{S_{i+1} - S_i}{\Delta t} \tag{2}$$

where, for example, I_i is notation for runoff inflow at time t_i, and I_{i+1} is inflow at time $t_i + \Delta t$.

A key assumption typically used in hydrologic-modeling algorithms is that Eq. (2) applies to the particular algorithm (e.g., hydrograph routing) and also that runoff storage and outflow are functionally related by

$$\frac{dS}{dt} = \frac{dS}{dO}\frac{dO}{dt} \tag{3}$$

where S and $\frac{dS}{dO}$ are positive (and nonzero) functions of outflow, O. This assumption may be problematic in applications involving backwater or hysterisis effects, among other issues. For example, a lengthy routing link may involve a rising or falling limb of a hydrograph that is not well approximated by Eq. (3).

Assuming that the time step size, Δt, is sufficiently small, the term $\frac{dS}{dO}$ is typically assumed constant during the time step such that

$$\frac{dS}{dt} = k_i \frac{dO}{dt} \tag{4}$$

where Eq. (4) and the constant k_i applies to a particular time step, Δt_i.

Combining (2) and (4) gives, for a particular time step,

$$\frac{I_i + I_{i+1}}{2} = \frac{O_i + O_{i+1}}{2} + k_i \left(\frac{O_{i+1} - O_i}{\Delta t}\right) \tag{5}$$

Recombining terms we have

$$a_i I_i + a_i I_{i+1} + b_i O_i = O_{i+1} \tag{6}$$

where $a_i = 1/(1 + 2k_i/\Delta t)$; $b_i = (2k_i - \Delta t)/(2k_i + \Delta t)$; $k_i = \left(\frac{dS}{dO}\right)\Big|_{t_i}$, where t_i is model time $(i-1)\Delta t$; and $I_o = O_o = 0$.

Matrix Representations

The usual procedure applied in R-R computer models to represent storm rainfall or a runoff hydrograph is to discretize the total storm duration into unit-period time intervals of constant duration, Δt. Both the effective rainfall over a subarea, and an inflow hydrograph to a routing link, can be handled as an inflow vector $\underset{\sim}{I}$ where

$$\underset{\sim}{I} = \begin{Bmatrix} I_1 \\ I_2 \\ I_3 \\ \bullet \\ \bullet \\ \bullet \\ I_n \end{Bmatrix} \tag{7}$$

where the dimension n of $\underset{\sim}{I}$ is chosen to be consistent with the entire matrix system to be developed. Similarly, the subarea runoff hydrograph, or an outflow hydrograph from a routing link, can be handled as an outflow vector $\underset{\sim}{O}$ where

$$\underset{\sim}{O} = \begin{bmatrix} O_1 \\ O_2 \\ \bullet \\ \bullet \\ \bullet \\ O_n \end{bmatrix} \tag{8}$$

Note that in (7), for example, $\underset{\sim}{I}_3 = I(2\Delta t)$ where Δt is the constant time step.

From (6),

$$\begin{aligned} O_1 &= a_1 I_1 \\ O_2 &= a_2 I_1 + a_2 I_2 + b_2 O_1 \\ &= a_2 I_1 + a_2 I_2 + b_2 a_1 I_1 \\ O_3 &= a_3 I_2 + a_3 I_3 + b_3 O_2 \\ &= a_3 I_2 + a_3 I_3 + b_3 a_2 I_1 + b_3 a_2 I_2 + b_3 b_2 a_1 I_1 \end{aligned} \tag{9}$$

In matrix form, a particular R-R computer-model algorithm given by (9) is mathematically described by

$$\begin{bmatrix} O_1 \\ O_2 \\ O_3 \\ \bullet \\ \bullet \\ \bullet \\ O_n \end{bmatrix} = \begin{bmatrix} a_1 & 0 & 0 & \bullet\bullet\bullet & 0 \\ (b_2 a_1 + a_2) & a_2 & 0 & \bullet\bullet\bullet & 0 \\ b_3(b_2 a_1 + a_2) & (b_3 a_2 + a_3) & a_3 & \bullet\bullet\bullet & 0 \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & \bullet\bullet\bullet & a_n \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \bullet \\ \bullet \\ \bullet \\ I_n \end{bmatrix} \tag{10}$$

or in simpler notation,

$$\underset{\sim}{O} = \mathbb{H} \underset{\sim}{I} \tag{11}$$

where $\underset{\sim}{O}$ and $\underset{\sim}{I}$ are nx1 column vectors; and $\mathbb{H}$ is a nxn lower triangular matrix. The matrix system (10) applies to any algorithm or component used in a R-R computer model that satisfies assumptions (1) and (3), and is based upon a discretization of time such as (2), (4) and (5). Although the $\mathbb{H}$ matrix of (10) is a function of the particular R-R model algorithm's properties, the notation "$\mathbb{H}$" is used for any such $\mathbb{H}$ matrix without further descriptive notation specifying that particular algorithm's attributes. In order to maintain consistent dimensions, the vectors $\underset{\sim}{O}$ and $\underset{\sim}{I}$ often require additional zero value entries in order to extend the vector's dimension.

The particular $\mathbb{H}$ matrix of (10) is generated by noting that the entries of row i+1 are equal to the entries of row i multiplied by the value b_{i+1}, and adding the value a_{i+1} to $\mathbb{H}(i+1,i)$ and $\mathbb{H}(i+1,i+1)$. Hereafter in this paper, a R-R computer model, whose component algorithms and processes all satisfy Eqs. (1) to (6), is called a "Type 2" model.

An example of a Type 2 R-R model is a HEC-1 model network, with significant detention basin effects modeled by use of the modified-Puls method such that basin outflow and basin storage are not linearly related.

A computer model network of a catchment is composed of several linkages and sources of runoff. Consequently, the corresponding subarea runoff estimators and hydrograph routing algorithms all combine in a complex way. The nonlinearity of a routing algorithm may be dampened to insignificance by another routing algorithm or by the contribution from a subarea runoff estimation method, resulting in a global model that is essentially a Type I model structure. Similarly, the nonlinearity effects evident at a particular process in a catchment may preclude the use of a Type I model structure due to the improvement in accuracy afforded by a Type II model structure.

Subarea Runoff Hydrographs

Let the study watershed, denoted as Ω, be discretized into subareas Ω_j, $j = 1,2,\ldots,N$. Each subarea has an associated effective rainfall vector $\underset{\sim}{e_j}$ developed by some prescribed algorithm, where each $\underset{\sim}{e_j}$ is a nx1 column vector composed of sequential Δt unit-period effective-rainfall depths (rainfall less losses), analogous to the vector of (7). If the subarea j runoff generation algorithm is from a Type 2 model, then the runoff hydrograph in vector form (analogous to Eq. (8)), denoted as $\underset{\sim}{q}_j$, is given by

$$\underset{\sim}{q}_j = \mathbb{T}_j \underset{\sim}{e}_j \tag{12}$$

where $\mathbb{T}_j$ is the appropriate $\mathbb{H}$ matrix of Eq. (11) for this particular process and subarea j.

Flood Hydrograph Routing

In a model link-node network, subareas are linked together by hydrograph-routing links that include a model algorithm for transporting a hydrograph along the length of the routing link. A link is described as a connection between nodal points (nodes).

Given an inflow hydrograph at node i, and a link that connects node i to downstream node i+1, then the outflow hydrograph, at node i+1, is developed by routing the inflow hydrograph along the length of the link, $L_{i,i+1}$. If the routing algorithm is from a Type 2 model, then the following equation in the form of (11) is obtained

$$\underset{\sim}{Q}_{i+1} = \mathbb{R}_{i,i+1} \underset{\sim}{I}_i \tag{13}$$

where $\underset{\sim}{I}_i$ is the inflow hydrograph at node i in Δt unit-period vector form; $\underset{\sim}{Q}_{i+1}$ is the outflow hydrograph at node i+1; and $\mathbb{R}_{i,i+1}$ is the appropriate $\mathbb{H}$ matrix for the hydrograph-routing algorithm selected for link $L_{i,i+1}$.

Link-Node Models

Link-node model applications of Type 2 R-R computer models are readily developed using the above formulations.

Example 1. Figure 1 depicts a model schematic where a single subarea runoff hydrograph concentrates at node #1 and is then routed to node #2 via model link $L_{1,2}$. For a Type 2 model, the runoff hydrograph from subarea #1 is given by

$$\underset{\sim}{q}_1 = \mathbb{T}_1 \underset{\sim}{e}_1$$

and the runoff hydrograph at node #1 is given by $\utilde{Q}_1 = \utilde{q}_1$. The hydrograph at node #2 is

$$\utilde{Q}_2 = \mathbb{R}_{1,2}\,\utilde{Q}_1 = \mathbb{R}_{1,2}\,\mathbb{T}_1\,\utilde{e}_1$$

Example 2. Figure 2 depicts a model schematic involving four subareas and two links. From the figure, the various runoff hydrographs developed from a Type 2 computer model are given by

$$\utilde{Q}_1 = \utilde{q}_1 + \utilde{q}_2 = \mathbb{T}_1\,\utilde{e}_1 + \mathbb{T}_2\,\utilde{e}_2$$

$$\utilde{Q}_2 = \mathbb{R}_{1,2}\,\utilde{Q}_1 + \utilde{q}_3 = \mathbb{R}_{1,2}(\mathbb{T}_1\,\utilde{e}_1 + \mathbb{T}_2\,\utilde{e}_2) + \mathbb{T}_3\,\utilde{e}_3$$

$$\utilde{Q}_3 = \mathbb{R}_{2,3}\,\utilde{Q}_2 + \utilde{q}_4 = \mathbb{R}_{2,3}\,(\mathbb{R}_{1,2}(\mathbb{T}_1\,\utilde{e}_1 + \mathbb{T}_2\,\utilde{e}_2) + \mathbb{T}_3\,\utilde{e}_3) + \mathbb{T}_4\,\utilde{e}_4$$

where again $\utilde{Q}_3$ is the runoff hydrograph at node #3 developed from a particular Type 2 computer model, and $\utilde{q}_3$ is the runoff hydrograph from subarea #3 developed from the particular Type 2 computer model.

$\mathbb{H}$ Matrix Principles

From the above, a Type 2 R-R computer model can be resolved into sums and products of $\mathbb{H}$ matrices such as developed in Eq. (10). In order to use this result effectively, some of the properties associated with $\mathbb{H}$ matrices must be considered.

Let $\mathbb{H}(n)$ be the set of all nxn lower triangular matrices, and any element of $\mathbb{H}(n)$ is said to be a $\mathbb{H}$ matrix. Key properties of $\mathbb{H}(n)$ are as follows:

Property 1: Let $\mathbb{A}$ and $\mathbb{B} \in \mathbb{H}(n)$. Then $\mathbb{A} + \mathbb{B} \in \mathbb{H}(n)$.
Property 2: Let $\mathbb{A}$ and $\mathbb{B} \in \mathbb{H}(n)$. Then $\mathbb{A}\,\mathbb{B} \in \mathbb{H}(n)$.
Property 3: Let λ be a real constant and $\mathbb{A} \in \mathbb{H}(n)$. Then $\lambda\,\mathbb{A} \in \mathbb{H}(n)$.
Property 4: $\mathbb{A} + \mathbb{B} = \mathbb{B} + \mathbb{A}$.
Property 5: $(\mathbb{A} + \mathbb{B}) + \mathbb{C} = \mathbb{A} + (\mathbb{B} + \mathbb{C})$.
Property 6: Let $\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$ all be $\in \mathbb{H}(n)$. Then $\mathbb{A}(\mathbb{B} + \mathbb{C}) = \mathbb{A}\mathbb{B} + \mathbb{A}\mathbb{C}$.

With the above properties, the manipulation of the $\mathbb{H}$ matrices is straightforward. It is noted, however, that the product of $\mathbb{H}$ matrices is not necessarily commutative. This mathematical result is consistent with nonlinear R-R computer model results that indicate sensitivity to the ordering and arrangement of links and other processes. Using the above properties, the example 2 problem results for node #3 can be expanded into a series, as

$$\begin{aligned} \underset{\sim}{Q_3} &= \mathbb{R}_{2,3}\,\mathbb{R}_{1,2}\,\mathbb{T}_1\,\underset{\sim}{e_1} + \mathbb{R}_{2,3}\,\mathbb{R}_{1,2}\,\mathbb{T}_2\,\underset{\sim}{e_2} + \mathbb{R}_{2,3}\,\mathbb{T}_3\,\underset{\sim}{e_3} + \mathbb{T}_4\,\underset{\sim}{e_4} \\ &= \mathbb{H}_1\,\underset{\sim}{e_1} + \mathbb{H}_2\,\underset{\sim}{e_2} + \mathbb{H}_3\,\underset{\sim}{e_3} + \mathbb{H}_4\,\underset{\sim}{e_4} = \sum_{j=1}^{4} \mathbb{H}_j\,\underset{\sim}{e_j} \end{aligned} \tag{14}$$

where each $\mathbb{H}_j$ is a $\mathbb{H}$ matrix and, for example, $\mathbb{H}_1 = \mathbb{R}_{2,3}\,\mathbb{R}_{1,2}\,\mathbb{T}_1$.

Comparison to a Toeplitz Matrix Structure

In Hromadka and Whitley (1998), a mathematical formalization of the link-node modeling system used in the computer program HEC-1 and related programs produced the result that many frequently used R-R computer-model structures were resolvable into Toeplitz matrix systems. Toeplitz matrices of dimension n, denoted by T(n), are also $\mathbb{H}$ matrices except that a particular circulant matrix structure occurs; for example, for $\mathbb{U} \in T(n)$,

$$\mathbb{U} = \begin{bmatrix} u_1 & 0 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots & 0 \\ u_3 & u_2 & u_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ u_n & u_{n-1} & u_{n-2} & \cdots & u_1 \end{bmatrix} \tag{15}$$

Note that in (15), the diagonal and off-diagonal terms are constant. Many R-R computer models utilize processes and algorithms that are shown, in Hromadka and Whitley (1998), to be Toeplitz matrices (see the Introduction of this paper). Hereafter in this paper, computer models whose algorithms and processes resolve into Toeplitz matrices will be called "Type 1" models.

Because Type 1 models are also Type 2 models, they satisfy the above properties 1 to 6. It is noted that for Toeplitz matrices, matrix multiplication is commutative, which is a different result than for $\mathbb{H}$ matrices in a Type 2 model.

The distinction between Type 1 and Type 2 model algorithms and processes occurs at Eq. (3). If $\frac{dS}{dO}$ is a constant for a particular algorithm, then in Eqs. (6), (9), (10), the terms a_i and b_i are the constants a_o and b_o, respectively, in which case (10) simplifies to

$$\begin{Bmatrix} O_1 \\ O_2 \\ O_3 \\ \bullet \\ \bullet \\ \bullet \\ O_n \end{Bmatrix} = \begin{bmatrix} a_o & 0 & 0 & \bullet\bullet\bullet & 0 \\ a_o(b_o+1) & a_o & 0 & \bullet\bullet\bullet & 0 \\ b_o a_o(b_o+1) & a_o(b_o+1) & a_o & \bullet\bullet\bullet & 0 \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & \bullet\bullet\bullet & a_o \end{bmatrix} \begin{Bmatrix} I_1 \\ I_2 \\ I_3 \\ \bullet \\ \bullet \\ \bullet \\ I_n \end{Bmatrix} \tag{16}$$

which is a Toeplitz matrix system. Several algorithms used in HEC-1 and related programs are resolvable into Toeplitz matrices; for example the hydrologic routing procedures of translation, convex, Muskingum, convolution, modified-Puls (with $S = kO$; i.e., a linear routing method), and the unit-hydrograph method for generating runoff are all Toeplitz matrix systems (Hromadka and Whitley, 1998). Similarly, modified-Puls routing where storage and outflow are nonlinear, or in a nonlinear model of hydrograph routing, the corresponding matrix system is not of the Toeplitz form, yet the continuity equations (1) to (6) are satisfied; these algorithms are found in Type II models.

Matrix System Representation Series Expansions

Suppose a Type 1 computer model is used to study a watershed with m "sources" of effective-rainfall information vectors. For example, several subareas of the link-node model may have identical effective-rainfall vectors or have subarea effective rainfalls that are linear combinations of source effective-rainfall data. The concept introduced here of "sources" of effective rainfall is analogous to the principles involving mutually independent vectors where, in this case, a subarea's effective-rainfall vector is a linear combination of other subarea effective-rainfall vectors. The minimum number of linearly independent effective-rainfall vectors is called a basis, where each subarea's effective-rainfall vector is a linear combination of the basis or "source" effective-rainfall vectors. The number of linearly independent vectors in the basis is called the dimension. Hromadka (1993) provides more details. Then, by generalizing the series expansion of Eq. (14), the computed runoff hydrograph is the vector $\underset{\sim}{Q}$ where

$$\underset{\sim}{Q} = \sum_{k=1}^{m} \mathbb{A}_k \underset{\sim}{e_k} \tag{17}$$

where $\underset{\sim}{Q}$ and each $\underset{\sim}{e_k}$ are nx1 column vectors; each $\mathbb{A}_k$ is an nxn Toeplitz matrix; and m is the dimension of the effective-rainfall vector basis.

Similarly, if a Type 2 computer model is used, then from the series expansion of Eq. (14),

$$\underset{\sim}{Q} = \sum_{k=1}^{m} \mathbb{H}_k \underset{\sim}{e_k} \tag{18}$$

where each $\mathbb{H}_k$ is an nxn $\mathbb{H}$ matrix.

If only a single source of effective-rainfall information (i.e., the effective-rainfall vector basis has a dimension = 1) is used, the vector $\underset{\sim}{e_0}$, then Eqs. (17) and (18) simplify to

$$\underset{\sim}{Q} = \begin{cases} \mathbb{A}_o \underset{\sim}{e_0} \text{ ; Type 1 model} & (19) \\ \mathbb{H}_o \underset{\sim}{e_0} \text{ ; Type 2 model} & (20) \end{cases}$$

where $\mathbb{A}_o$ is a nxn Toeplitz matrix; and $\mathbb{H}_o$ is a nxn $\mathbb{H}$ matrix. The matrices $\mathbb{A}_o$ and $\mathbb{H}_o$ are consistent with the dimensions of the $\mathbb{A}_k$ and $\mathbb{H}_k$ matrices used in (17) and (18), respectively. The zero subscript notation in $\mathbb{A}_o$ and $\mathbb{H}_o$ is used to indicate that, in the case of (19) and (20), both $\mathbb{A}_o$ and $\mathbb{H}_o$ are equivalent to the entire assemblage of $\mathbb{A}_k$ and $\mathbb{H}_k$ matrices used in (17) and (18), respectively. The use of Eqs. (19) and (20) is further investigated for deciding whether a Type 1 or Type 2 R-R computer model is "best".

R-R Model Calibration and Model Structure Selection

In order to simplify the analysis, a single storm and a single source of effective-rainfall information are considered for R-R model calibration purposes. Equations (17) to (20) are important because they show that the application of a Type 1 or Type 2 R-R computer model to a highly discretized watershed involving numerous subareas linked together by numerous hydrograph-routing links results in a nxn matrix series expansion (Eqs. 17, 18) that, in the case of a single source of effective rainfall information, simply sums to a single matrix-vector product (Eqs. 19, 20) and this resulting matrix-vector product involves a nxn matrix that is either a nxn Toeplitz matrix, $\mathbb{A}_o$, (Eq. 19) or an nxn $\mathbb{H}$ matrix, $\mathbb{H}_o$ (Eq. 20). That is, if the effective rainfall vector set has a basis of dimension = 1, then regardless of the number of links and subareas, the matrix systems of (19) or (20) still result. Of course, different arrangements or different choices of hydrograph-routing algorithms (e.g. Muskingum versus convex, etc.) or use of a different subarea hydrograph generator algorithm will result in different $\mathbb{A}_k$ and $\mathbb{H}_k$ matrices in (17) and (18), and also different $\mathbb{A}_o$ and $\mathbb{H}_o$ matrices in (19) and (20), respectively. Consequently, a wide range of matrices must be considered due to the variety of R-R model schematics and algorithms. The key issue, then, is determining the "best" matrices to be used in Eqs. (17) to (20). One approach is to utilize standard optimization techniques, which are summarized below.

Let $\mathbb{A}_o$ be in T(n). Then,

$$\mathbb{A}_o = \begin{bmatrix} a_1 & 0 & 0 & \bullet\bullet\bullet & 0 \\ a_2 & a_1 & 0 & \bullet\bullet\bullet & 0 \\ a_3 & a_2 & a_1 & \bullet\bullet\bullet & 0 \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ a_n & a_{n-1} & a_{n-2} & \bullet\bullet\bullet & a_1 \end{bmatrix} \qquad (21)$$

Then $\mathbb{A}_o$ is seen to a function of n variables (or degrees of freedom)

$$\mathbb{A}_o = \mathbb{A}_o(a_1, a_2, \ldots, a_n) \qquad (22)$$

Similarly, the $\mathbb{H}_0$ matrix of (20) is in $\mathbb{H}(n)$ where

$$\mathbb{H}_0 = \begin{bmatrix} h_1 & 0 & 0 & \cdots & 0 \\ h_2 & h_3 & 0 & \cdots & 0 \\ h_4 & h_5 & h_6 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \cdot & \cdot & \cdot & \cdots & h_{(n+1)n/2} \end{bmatrix} \tag{23}$$

Thus, $\mathbb{H}_0$ is a function of (n+1)n/2 degrees of freedom:

$$\mathbb{H}_0 = \mathbb{H}_0(h_1, h_2, \ldots, h_{(n+1)n/2}) \tag{24}$$

When rainfall-runoff data are available to calibrate the R-R model at a specific point of study, the question arises whether to calibrate the matrix systems of Eqs. (19) or (20), or whether to attempt to calibrate each of the numerous algorithms and processes leading to Eqs. (17) and (18).

In order to differentiate between these two types of calibration strategies, let $\mathbb{H}_0^{LN}$ be the $\mathbb{H}_0$ matrix resulting from the calibration of each link-node model algorithm and process leading to the development of (20), and let $\mathbb{H}_0^*$ be the $\mathbb{H}_0$ matrix resulting from the direct calibration of Eq. (23) in Eq. (20). Similarly, let $\mathbb{A}_0^{LN}$ and $\mathbb{A}_0^*$ be the resulting Toeplitz matrices from calibration of each of the various link-node processes and algorithms leading to (19) and the calibration of Eq. (21), respectively (Hromadka and Whitley (1998) focus on this particular topic).

As shown in Hromadka and Whitley (1998), for a Type 1 computer model, the best calibration is achieved using the $\mathbb{A}_0^*$ matrix, because the $\mathbb{A}_0^{LN}$ matrix cannot do any better than the $\mathbb{A}_0^*$ matrix in reducing the residual error between computed runoff and the runoff data. It is recalled that setting $\mathbb{A}_0 = \mathbb{A}_0^*$ in Eq. (19) results in a matrix system representation of the classic single area unit-hydrograph method. Similarly, the $\mathbb{H}_0^{LN}$ matrix cannot outperform the $\mathbb{H}_0^*$ matrix in reducing modeling residual error in the Type 2 computer-model calibration effort.

Thus, the best calibration possible from the Type 1 or Type 2 computer model, to a specific storm, given a single source of effective-rainfall data, is achieved by developing the $\mathbb{A}_0^*$ or $\mathbb{H}_0^*$ matrices, respectively, as used in Eqs. (19) and (20). This conclusion is used in developing a decision-making procedure for deciding whether to use a Type 1 or Type 2 R-R computer model. In other words, provision of an answer to the question, "Are the R-R data sufficient to conclude that use of a Type 2 model will necessarily provide better results than use of a Type 1 model?" will be attempted. This question is posed more precisely in the following Type 1 R-R Model Structure Selection Test.

TYPE 1 R-R MODEL STRUCTURE SELECTION TEST:

Let T_1 and T_2 be the sets of all Type 1 and Type 2 R-R computer models, respectively. Let $T_3 = T_2 - T_1$, because T_1 is a subset of T_2, it is necessary to distinguish models in T_2 that are not simply Toeplitz systems and, therefore, also in T_1.

Let $\mathfrak{D}$ be a set of M (M > 1) mutually independent rainfall-runoff events to be used for R-R model calibration and verification purposes. Choose a split-sample size, $0 < M_s < M$ where M_s is the number of storms to be used in model calibration, and $(M-M_s)$ storms are used in model verification tests. Then there are $\eta = \binom{M}{M_s}$ possible split-sample combinations of the M R-R events in $\mathfrak{D}$. Let S_η be the set of all such split-sample combinations. The elements of S_η are denoted by $s_i \in S_\eta$; $i = 1,2,\ldots,\eta$.

Let the minimum least-squares residual error in calibration of a Type 1 and a Type 2 R-R computer model to split-sample data set $s_i \in S_\eta$ be achieved by elements $M_{1i}^* \in T_1$ and $M_{2i}^* \in T_2$, respectively; $i = 1,2,\ldots,\eta$. M_{1i}^* (and also M_{2i}^*) is optimized with respect to all M_s storms in s_i, simultaneously.

Three cases are identified:

(a) If $M_{2i}^* \in T_1$, then $M_{2i}^* = M_{1i}^*$ and a Type 1 model structure provides the minimum calibration residual error.

(b) If $M^*_{2i} \in T_3$ and the residual error for the verification events (of s_i) produced by M^*_{1i} is less than that produced by M^*_{2i}, then a Type 1 model structure provides the minimum verification test residual error.

(c) Neither case (a) or (b) applies.

If either case (a) or (b) applies for data set s_i, then the above described test is considered "passed". Let P be the total number of passed tests; for $i = 1,2,...,\eta$. Define the Type 1 Model Structure performance ratio, r, by $r = P/\eta$.

In this paper, if r > 50-percent, the Type 1 R-R model structure is selected as the "best" model structure, i.e., more than 50-percent of the time the use of Toeplitz matrices, as opposed to the use of the more general lower triangular matrices, provides either the smallest calibration error or the smallest verification error.

Choosing Between a Type 1 or Type 2 Model Structure

To apply the above concepts, the optional $\mathbb{A}^*_0$ and $\mathbb{H}^*_0$ matrices must be determined with respect to several storms simultaneously. Specifically, suppose M storms with similar effective rainfalls (similar in timing and magnitude) are available. Then, regardless of the R-R model structure, similar computed runoff hydrographs are expected. Additionally, similar matrix systems should be developed for each of these storms (although, of course, the Type 1 matrices would differ from the Type 2 matrices).

In order to choose a Type 1 or Type 2 R-R computer model for future use, the split-sample test procedure, provided below, can be used:

R-R Model Split Sample Test

<u>Step 1.</u> Divide the M similar storm data sets of effective rainfall and runoff into two sets, calibration and verification.

<u>Step 2.</u> Develop optimized matrices $\mathbb{A}^*_0$ and $\mathbb{H}^*_0$, each optimized for the entire set of calibration storms, using a least squares residual minimization and Eqs. (19) and (20). Constraints may be considered such as imposing a requirement that all matrix entries are nonnegative; or that particular matrix entry values are bounded by proportions of other matrix entry values; among other constraints. (The following application case study demonstrates the optimization process for 4 calibration storms, simultaneously.)

<u>Step 3.</u> For the verification storms, use $\mathbb{A}^*_0$ and $\mathbb{H}^*_0$ in Eqs. (19) and (20) to "predict" runoff quantities for each verification event.

<u>Step 4.</u> Compare the computed runoff hydrographs from Step 3 to the runoff data for the verification storm set and compute the total residual error.

<u>Step 5.</u> Select the computer model type based on the least total residual error.

<u>Step 6.</u> Repeat steps 1 to 5 by trying all possible split-sampling combinations to evaluate sensitivity to selection of model type.

From the previous section, the Type 1 model that solves Eq. (19), using $\mathbb{A}^*_0$, is the best Type 1 model in calibrating to the R-R data at a specific watershed location. Similarly, the best Type 2 model performance in calibration to the R-R data is achieved using Eq. (20) with the $\mathbb{H}^*_0$ matrix. Because there are more degrees of freedom in the $\mathbb{H}^*_0$ matrix, Eq. (20) will typically provide a lower calibration residual error than Eq. (19), and could appear to be the better model structure; however, a better model in calibration is not necessarily a better model in prediction; hence, the split-sample test procedure.

If the $\mathbb{A}_0^*$ matrix Type 1 model provides the best split-sample testing verification results in comparison to the $\mathbb{H}_0^*$ matrix Type 2 model, then no other Type 2 model can perform any better than the $\mathbb{A}_0^*$ matrix Type 1 model for the given test calibration specifications; i.e., the classic single-area unit hydrograph technique is the best approximation in the specific case study and for the available R-R data.

It is noted that the above calibration procedure results in an $\mathbb{A}_0^*$ matrix that can be used for a variety of storms just as a calibrated unit hydrograph is commonly used for a variety of storms. In contrast, the above $\mathbb{H}_0^*$ matrix is only appropriate for use with storms whose effective rainfalls are similar in both magnitude and timing with the set of storm effective rainfalls used in the calibration and development of the $\mathbb{H}_0^*$ matrix. This is because the $\mathbb{H}_0^*$ matrix was constructed from a set of $\frac{dS}{dO}$ values (see Eqs. (1) to (6)) that are all mutually dependent on storm timing and magnitude, which dictates the condition of runoff storage throughout the link-node model for each model time step, Δt.

If the split-sample test results indicates strong evidence that a Type I model structure does not achieve the success, in verification, that a Type II model structure achieves, then there may be several issues that are involved. For example, there may be highly nonlinear R-R or routing processes involved, there may be issues regarding R-R data, among other topics.

Case Study: Choosing Between a Type 1 or Type 2 R-R Computer Model Structures for a Catchment in Los Angeles, California

Seven significant storms $\{\underset{\sim}{e}i \ ; i=1,2,\ldots,7\}$ were selected from a single rain gauge in Los Angeles, California. The rain gauge is located near the centroid of a 7.4 square mile watershed. The watershed's condition of urbanization and storm drainage is essentially constant for all seven storms. The time of concentration is approximately 50 minutes as computed from a sum of flow velocity travel times along the main watercourse. Each of the selected storms had similar prior rainfall histories (i.e., antecedent moisture), and the rainfall pattern timing and magnitudes were such that all rainfall intensities were within ten percent of each other, for any storm time, t. Only the initial 2-hours of each event was used in this analysis. The storms were of such similarity that one would expect similar runoff responses for all seven storms, $\{\underset{\sim}{Q}^i ; i=1,2,...,7\}$. A five-minute unit time period was used in the analysis.

<u>Step 1.</u> For calibration purposes, the seven storms were split into a set of four calibration and three verification storms. All possible combinations of four calibration and three verification storms were determined (a total of 35) for eventual decision-making sensitivity analysis, in step 6.

<u>Step 2a: Calibration of $\mathbb{A}_o^*$ matrix</u>. The entries of the $\mathbb{A}_o^*$ matrix, where

$$\mathbb{A}_o^* = \begin{bmatrix} a_1^* & 0 & 0 & \cdots & 0 \\ a_2^* & a_1^* & 0 & \cdots & 0 \\ a_3^* & a_2^* & a_1^* & \cdots & 0 \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ \bullet & \bullet & \bullet & & \bullet \\ a_n^* & a_{n-1}^* & a_{n-2}^* & \cdots & a_1^* \end{bmatrix} \quad (25)$$

must be determined such that the least-squares residual between the computed runoff hydrographs and the measured runoff hydrographs is a minimum.

For a single storm with effective rainfall $\underset{\sim}{e}^1$ (for example), the usual least-squares difference between the vectors $\underset{\sim}{Q}^1$ and $\underset{\sim}{Q}^{*1}$ must be minimized, where $\underset{\sim}{Q}^1$ is the measured runoff hydrograph and $\underset{\sim}{Q}^{*1}$ is the computed runoff hydrograph given by

$$\underset{\sim}{Q}^{*1} = \mathbb{A}_o^* \underset{\sim}{e}^1 \quad (26)$$

and the dimensions of $\underset{\sim}{Q}^{*1}$ and $\underset{\sim}{e}^1$ are nx1. The vector superscript notation refers to the storm number. This is accomplished by noting, for $e_1 \neq 0$,

$$\mathbb{A}^*_o \underset{\sim}{e}^1 = a^*_1 \begin{Bmatrix} e_1 \\ e_2 \\ e_3 \\ \bullet \\ \bullet \\ \bullet \\ e_n \end{Bmatrix}^1 + a^*_2 \begin{Bmatrix} 0 \\ e_1 \\ e_2 \\ \bullet \\ \bullet \\ \bullet \\ e_{n-1} \end{Bmatrix}^1 + a^*_3 \begin{Bmatrix} 0 \\ 0 \\ e_1 \\ e_2 \\ \bullet \\ \bullet \\ e_{n-2} \end{Bmatrix}^1 + \cdots + a^*_n \begin{Bmatrix} 0 \\ 0 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ e_1 \end{Bmatrix}^1 \tag{27}$$

Then, from Eq. (27), the values of $\{a^*_i;\ i=1,2,\ldots,n\}$ are uniquely determined by the well-known Gramm-Schmidt procedure in minimizing the least squares difference between the recorded runoff $\underset{\sim}{Q}^1$, and the product $\mathbb{A}^*_o \underset{\sim}{e}^1$ (see Hromadka and Whitley, 1989, Chapter 5).

For four storms, however, the Gramm-Schmidt procedure is extended to minimizing the residual error for all four storms simultaneously. This is accomplished by "stacking" all four calibration storms $\underset{\sim}{Q}^i$ and $\underset{\sim}{e}^i$ vectors, i=1,2,3,4 such that Eq. (27) is extended to choosing the a^*_i values of $\mathbb{A}^*_o$ to minimize the least squares residual in the vector equation:

$$
\begin{bmatrix}
\begin{Bmatrix} q_1 \\ q_2 \\ \bullet \\ \bullet \\ \bullet \\ q_n \end{Bmatrix}^1 \\
\begin{Bmatrix} q_1 \\ q_2 \\ \bullet \\ \bullet \\ \bullet \\ q_n \end{Bmatrix}^2 \\
\bullet \\ \bullet \\ \bullet \\
\begin{Bmatrix} q_1 \\ q_2 \\ \bullet \\ \bullet \\ \bullet \\ q_n \end{Bmatrix}^4
\end{bmatrix}
\equiv a_1^*
\begin{bmatrix}
\begin{Bmatrix} e_1 \\ e_2 \\ \bullet \\ \bullet \\ \bullet \\ e_n \end{Bmatrix}^1 \\
\begin{Bmatrix} e_1 \\ e_2 \\ \bullet \\ \bullet \\ \bullet \\ e_n \end{Bmatrix}^2 \\
\bullet \\ \bullet \\ \bullet \\
\begin{Bmatrix} e_1 \\ e_2 \\ \bullet \\ \bullet \\ \bullet \\ e_n \end{Bmatrix}^4
\end{bmatrix}
+ a_2^*
\begin{bmatrix}
\begin{Bmatrix} 0 \\ e_1 \\ e_2 \\ \bullet \\ \bullet \\ e_{n-1} \end{Bmatrix}^1 \\
\begin{Bmatrix} 0 \\ e_1 \\ e_2 \\ \bullet \\ \bullet \\ e_{n-1} \end{Bmatrix}^2 \\
\bullet \\ \bullet \\ \bullet \\
\begin{Bmatrix} 0 \\ e_1 \\ e_2 \\ \bullet \\ \bullet \\ e_{n-1} \end{Bmatrix}^4
\end{bmatrix}
+ \ldots + a_n^*
\begin{bmatrix}
\begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ 0 \\ e_1 \end{Bmatrix}^1 \\
\begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ 0 \\ e_1 \end{Bmatrix}^2 \\
\bullet \\ \bullet \\ \bullet \\
\begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ 0 \\ e_1 \end{Bmatrix}^4
\end{bmatrix}
\tag{28}
$$

The resulting values a_i^*; i=1,2,...,n, are the components of the $\mathbb{A}_o^*$ matrix, for the available data and split-sample.

Step 2b. Calibration of $\mathbb{H}_o^*$ matrix. Analogous to step 2a, the entries of the $\mathbb{H}_o^*$ matrix, where

$$
\mathbb{H}_o^* =
\begin{bmatrix}
h_1^* & 0 & 0 & \bullet\bullet\bullet & 0 \\
h_2^* & h_3^* & 0 & \bullet\bullet\bullet & 0 \\
h_4^* & h_5^* & h_6^* & \bullet\bullet\bullet & 0 \\
\bullet & \bullet & \bullet & & \bullet \\
\bullet & \bullet & \bullet & & \bullet \\
\bullet & \bullet & \bullet & & \bullet \\
\bullet & \bullet & \bullet & \bullet\bullet\bullet & h_{(n+1)n/2}^*
\end{bmatrix}
\tag{29}
$$

must be determined such that the least-squares residual between all of the recorded hydrographs and the computed hydrographs using Eqs. (29) and (20), is minimized.

Again, for storm 1 with effective rainfall $\underset{\sim}{e1}$, this Type 2 model computed hydrograph is

$$\underset{\sim}{Q}^{*1} = \mathbb{H}^{*}_{o} \underset{\sim}{e1} \tag{30}$$

The corresponding expansion of (30) is

$$\mathbb{H}^{*}_{o} \underset{\sim}{e1} = h^{*}_{1} \begin{Bmatrix} e_1 \\ 0 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^{1} + h^{*}_{2} \begin{Bmatrix} 0 \\ e_1 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^{1} + h^{*}_{3} \begin{Bmatrix} 0 \\ e_2 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^{1} + \ldots + h^{*}_{(n+1)n/2} \begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ \\ e_n \end{Bmatrix}^{1} \tag{31}$$

For all four storms, the Gramm-Schmidt procedure would be applied to the vector equation:

$$
\begin{bmatrix} \begin{Bmatrix} q_1 \\ q_2 \\ \bullet \\ \bullet \\ \bullet \\ q_n \end{Bmatrix}^1 \\ \begin{Bmatrix} q_1 \\ q_2 \\ \bullet \\ \bullet \\ \bullet \\ q_n \end{Bmatrix}^2 \\ \bullet \\ \bullet \\ \bullet \\ \begin{Bmatrix} q_1 \\ q_2 \\ \bullet \\ \bullet \\ \bullet \\ q_n \end{Bmatrix}^4 \end{bmatrix}
\cong h^*_1
\begin{bmatrix} \begin{Bmatrix} e_1 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^1 \\ \begin{Bmatrix} e_1 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^2 \\ \bullet \\ \bullet \\ \bullet \\ \begin{Bmatrix} e_1 \\ 0 \\ \bullet \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^4 \end{bmatrix}
+ h^*_2
\begin{bmatrix} \begin{Bmatrix} 0 \\ e_1 \\ 0 \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^1 \\ \begin{Bmatrix} 0 \\ e_1 \\ 0 \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^2 \\ \bullet \\ \bullet \\ \bullet \\ \begin{Bmatrix} 0 \\ e_1 \\ 0 \\ \bullet \\ \bullet \\ 0 \end{Bmatrix}^4 \end{bmatrix}
+ \ldots + h^*_{(n+1)n/2}
\begin{bmatrix} \begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ 0 \\ e_n \end{Bmatrix}^1 \\ \begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ 0 \\ e_n \end{Bmatrix}^2 \\ \bullet \\ \bullet \\ \bullet \\ \begin{Bmatrix} 0 \\ 0 \\ \bullet \\ \bullet \\ 0 \\ e_n \end{Bmatrix}^4 \end{bmatrix}
\tag{32}
$$

The resulting values of h^*_i are the matrix components of $\mathbb{H}^*_0$.

Step 3. Using the $\mathbb{A}^*_0$ and $\mathbb{H}^*_0$ matrices developed above, computed runoffs are developed for each verification storm $\mathbb{A}^*_0 \underset{\sim}{e}i$; i=5,6,7, and $\mathbb{H}^*_0 \underset{\sim}{e}i$; i=5,6,7. These are the Type 1 and Type 2 model verification estimates to be compared to the recorded runoffs from storms 5, 6, and 7.

Step 4. A least-squares weighted residual is developed for both the Type 1 and Type 2 model structures, for storms 5, 6, and 7, and summed, respectively.

Step 5. Based on the total residual error from the verification set of storms, the smaller-residual error model type is selected as "best".

Step 6. The above steps 1 to 5 are repeated for each combination of the calibration/verification storms.

In this application, the Type 1 model structure was selected as "best" 33 of the 35 split-sample test combinations. The resulting $\mathbb{A}_0^*$ Toeplitz matrices all differed, but the variations observed in the matrix components was only on the order of 4-percent. Additionally, the calibrated $\mathbb{H}_0^*$ matrices were all quite similar to a Toeplitz matrix structure. For example, the main diagonal terms of the $\mathbb{H}_0^*$ matrices had a standard deviation of less than 5-percent about the mean, for each $\mathbb{H}_0^*$ matrix.

CONCLUSIONS

A mathematical formalization is introduced for describing rainfall-runoff (R-R) computer models and their component algorithms and processes. With this formalization, a precise examination can be made of the algorithmic underpinnings of various R-R modeling structures. In this paper, the formalization is extended to R-R models whose algorithms and processes all satisfy the mass conservation equation and also assume storage is a function of outflow such that the first derivative of storage with respect to outflow is positive. The formalization introduces two types of R-R computer model structures, called "Type 1" or "Type 2", that describe almost all R-R computer models in use today; for example, the classic unit hydrograph method is found to be a Type 1 model structure. The formalization is then applied to develop a procedure useful in evaluating whether a Type 1 or Type 2 model structure may be best for a particular application given a R-R data set for model calibration purposes.

A computer model network of a catchment is composed of several linkages and sources of runoff. Consequently, the corresponding subarea runoff estimators and hydrograph routing algorithms all combine in a complex way. The nonlinearity of a routing algorithm may be dampened to insignificance by another routing algorithm or by the contribution from a subarea runoff estimation method, resulting in a global model that is essentially a Type I model structure. Similarly, the nonlinearity effects evident at a particular process in a catchment may preclude the use of a Type I model structure due to the improvement in accuracy afforded by a Type II model structure.

REFERENCES

1. Hromadka II, T.V. and Whitley, R.J., 1998, On Formalization of Unit-Hydrograph and Link-Node Hydrograph-Routing Systems, Journal of Hydrology, in-review.

2. Hromadka II, T.V., and Whitley, R.J., 1989, Stochastic Integral Equations in Rainfall-Runoff Modeling, Springer-Verlag, 350 pgs.

3. Hromadka II, T.V., 1993, The Best Approximation Method in Computational Mechanics, Springer-Verlag, 250 pages.

Figure 1. Model Schematic for a Single Subarea and a Single Hydrograph Routing Link.

Figure 2. Model Schematic for Four Subareas and Two Hydrograph Routing Links.

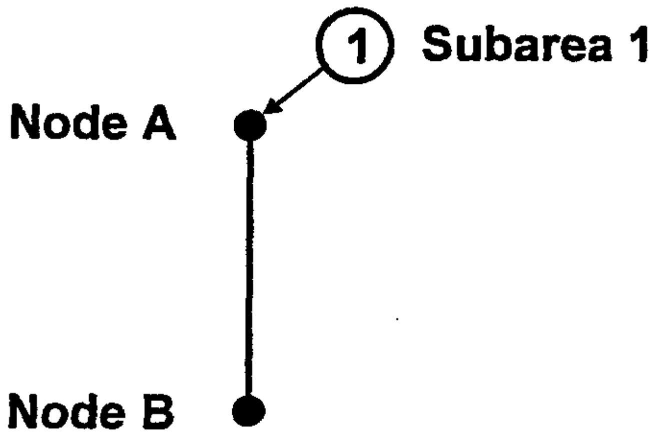

Figure 1. Model Schematic for a Single Subarea and a Single Hydrograph Routing Link.

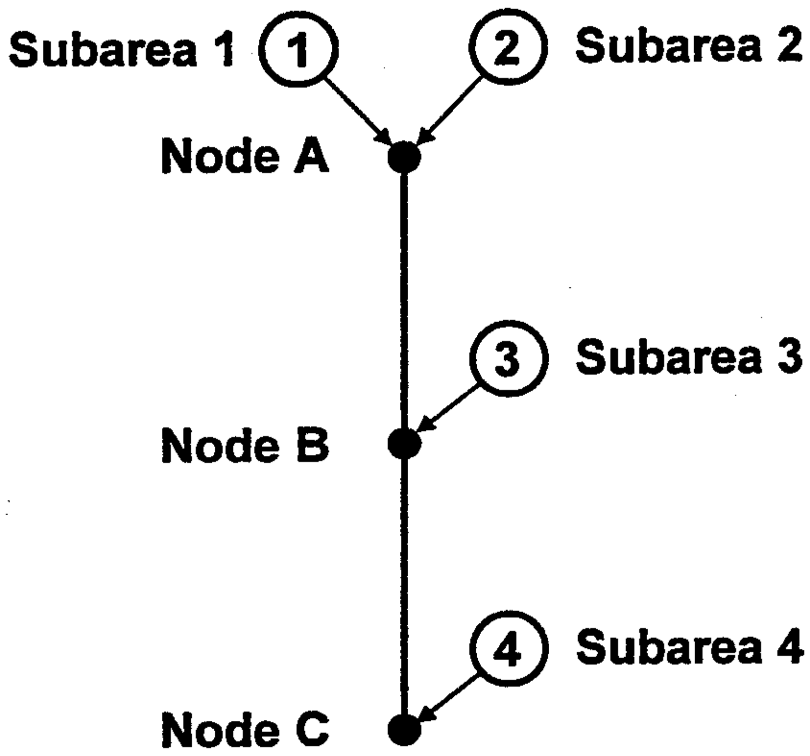

Figure 2. Model Schematic for Four Subareas and Two Hydrograph Routing Links.