Journal of Hydrology, 114 (1990) 17–45
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

[1]

ON IMPROVED HYDROLOGIC FORECASTING — RESULTS FROM A WMO REAL-TIME FORECASTING EXPERIMENT

KONSTANTINE P. GEORGAKAKOS[1] and GEORGE F. SMITH[2]

[1]*Department of Civil and Environmental Engineering and Iowa Institute of Hydraulic Research, The University of Iowa, Iowa City, IA 52242 (U.S.A.)*
[2]*Hydrologic Research Laboratory, W/OH3, National Weather Service, NOAA, Silver Spring, MD 20910 (U.S.A.)*

(Received April 10, 1989; accepted for publication June 9, 1989)

ABSTRACT

Georgakakos, K.P. and Smith, G.F., 1990. On improved hydrologic forecasting — results from a WMO real-time forecasting experiment. J. Hydrol., 114: 17–45.

A methodology for improved hydrologic forecasting of streamflows in headwater basins has been developed. The methodology is based on the mathematical modeling of the hydrologic catchment processes via conceptual hydrologic models. Real-time updating procedures have been designed for the on-line incorporation of past observed discharges in the forecasts. Updating is accomplished via a state estimator that exploits a priori knowledge on the statistical characteristics of model input and parameter errors. Hydrologic expertise is the only essential requirement for routine application. The value of the methodology in real-time hydrologic forecasting is illustrated for two hydrologically different headwater basins: the Bird Creek basin in Oklahoma, U.S.A., and the Orgeval basin in France. Verification of the methodology with data from the aforementioned basins was performed as part of a World Meteorological Organization (WMO) workshop held in Vancouver, British Columbia, Canada, during the period: 30 July–8 August, 1987. Real-time conditions were simulated with verification data made known after forecasts were issued. The verification results appear encouraging especially in light of the fact that all the forecasts were obtained in an automatic fashion (without manual intervention). The methodology developed has been implemented on the National Weather Service River Forecast System forecast component at the Hydrologic Research Laboratory, National Weather Service (NWS), and is currently undergoing testing in a true real-time environment.

OPERATIONAL FORECASTING OF STREAMFLOWS

Operational streamflow forecasting has been the subject of extensive research in recent years. The use of mathematical models that simulate the dynamics of the terrestrial runoff process and the utilization of improved monitoring systems have increased our capability to reliably predict floods and flash floods. Along these lines of research, conceptual mathematical models for soil moisture accounting and flood routing have been devised and used (Peck, 1976) in real-time flood forecasting. State estimators that process the observations of discharge and automatically update the model states (model variables

0022-1694/90/$03.50

that summarize past inputs) have been designed based on concepts borrowed from modern estimation theory (Kitanidis and Bras, 1980a,b; Georgakakos, 1986a,b). The estimators use information on the magnitude of the uncertainty that enters the mathematical model (because of inaccurate specification of input, model parameters and model structure) and produce forecasts of the expected flood discharges and associated estimation-error variances. They also utilize observations of discharge and associated observation error statistics, and they produce improved initial conditions for the state variables in a sequential manner (i.e. real-time updating) for improved forecasts.

Most of the past work has been based on recursive state estimation algorithms of the type of the linear Kalman Filter (Gelb, 1974). Since the nonlinearity of the natural process is represented in the conceptual mathematical models, and the Kalman Filter algorithm requires linear models, linearization techniques of various types have been used to generate linear approximations to nonlinear models (Georgakakos and Bras, 1982).

Denote by $\mathbf{f}(\mathbf{x},\mathbf{u};\boldsymbol{\theta})$ the nonlinear function that signifies the derivative of the state vector $\mathbf{x}$ with respect to time t. It is assumed that dependence of $\mathbf{f}$ on t is through the state vector $\mathbf{x}$, the vector of input variables $\mathbf{u}$, and the vector of parameters $\boldsymbol{\theta}$. The mathematical expression for the derivative vector function $\mathbf{f}$ will be referred to as the conceptual model in the following. If the model errors (in input, model parameters and structure) can be represented as an additive sequence of random vector disturbances $\mathbf{w}(t)$, then

$$\mathrm{d}\mathbf{x}(t)/\mathrm{d}t = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\theta}) + \mathbf{w}(t) \tag{1}$$

Consider an observation $z(t_k)$ of the natural process of streamflow at time instant t_k, and an expression $h(\mathbf{x}, \mathbf{u}; \boldsymbol{\theta})$ that models the relationship between the model state vector $\mathbf{x}(t_k)$ and the input vector $\mathbf{u}(t_k)$ at time t_k, and the observation at the same time. If $v(t_k)$ signifies the observation error at time t_k, then

$$z(t_k) = h(\mathbf{x}(t_k), \mathbf{u}(t_k); \boldsymbol{\theta}) + v(t_k) \tag{2}$$

where it has been assumed that the drainage basin of interest is a headwater one with available observations of streamflow at the outlet of the basin (a realistic assumption in the operational environment).

The set of eqns. (1) and (2) represents a state-space mathematical formulation of the terrestrial streamflow process. Conceptualization of the component processes, such as overland flow, groundwater flow and flood-wave propagation, gives expressions for the functions $\mathbf{f}$ and h.

The Kalman Filter applied to the nonlinear system (1)–(2) (the algorithm is called Extended Kalman Filter in this case) results in the following set of recursive differential and algebraic equations for step $[t_{k-1}, t_k]$:

$$(\mathrm{d}\hat{\mathbf{x}}(t)^-)/(\mathrm{d}t) = \mathbf{f}(\hat{\mathbf{x}}(t)^-, \mathbf{u}(t); \boldsymbol{\theta});\ t_{k-1} \leqslant t \leqslant t_k \tag{3}$$

$$(\mathrm{d}P(t)^-)/(\mathrm{d}t) = F(t)P(t)^- + P(t)^- F(t)^{\mathrm{T}} + Q(t);\ t_{k-1} \leqslant t \leqslant t_k \tag{4}$$

$$\mathbf{K}(t_k) = P(t_k)^- H(t_k)^{\mathrm{T}}[H(t_k)P(t_k)^- H(t_k)^{\mathrm{T}} + R(t_k)]^{-1} \tag{5}$$

$$\hat{\mathbf{x}}(t_k)^+ = \hat{\mathbf{x}}(t_k)^- + \mathbf{K}(t_k)[z(t_k) - h(\hat{\mathbf{x}}(t_k)^-, \mathbf{u}(t_k); \boldsymbol{\theta})] \tag{6}$$

$$P(t_k)^+ = [1 - \mathbf{K}(t_k)H(t_k)]P(t_k)^- [1 - \mathbf{K}(t_k)H(t_k)]^{\mathrm{T}} + \mathbf{K}(t_k)R(t_k)\mathbf{K}(t_k)^{\mathrm{T}} \tag{7}$$

where

$$[F(t)]_{i,j} = [\partial f_i(\hat{\mathbf{x}}(t)^-, \mathbf{u}(t);\boldsymbol{\theta})]/[\partial x_j(t)] \tag{8}$$

$$[H(t)]_{1,j} = [\partial h(\hat{\mathbf{x}}(t)^-, \mathbf{u}(t);\boldsymbol{\theta})]/[\partial x_j(t)] \tag{9}$$

and with the following symbol definitions

f_i the ith element of vector function $\mathbf{f}$
x_j the jth element of the state vector $\mathbf{x}$
$\hat{\mathbf{x}}(t)^-$ the unbiased minimum variance estimate of $\mathbf{x}(t)$ given all observations before time t
$\hat{\mathbf{x}}(t)^+$ the unbiased minimum variance estimate of $\mathbf{x}(t)$ given all observations up to and including time t
$P(t)^-$ the covariance matrix of the state estimation error vector given all observations before time t
$P(t)^+$ the covariance matrix of the state estimation error vector given all observations up to and including time t
$Q(t)$ the covariance parameter matrix of the model error process $\mathbf{w}(t)$
$R(t_k)$ the variance function of the random sequence $v(t_k)$.

It is noted that the derivatives in eqns. (8) and (9) are functions of the state vectors evaluated at the best estimate of the state vector before the observation at time t, $z(t)$, becomes available.

The set of eqns. (3)–(7) can be utilized recursively to (a) process the observations $z(t_k)(k = 1,2, \ldots, M)$ as they become available in real time, and (b) produce forecasts of the state vector $\mathbf{x}(t_k)(\mathbf{x}(t)^-)$ and associated estimation-error covariance matrix $P(t_k)(P(t_k)^-)$. (Initial conditions $\mathbf{x}_0$ and P_0 are required for the initiation of the algorithm.) Given these forecasts, forecasts of streamflow $\hat{z}(t_k)$ and associated estimation-error variance $P_z(t_k)^-$ can be produced by

$$\hat{z}(t_k) = h(\hat{\mathbf{x}}(t)^-, \mathbf{u}(t_k);\boldsymbol{\theta}) \tag{10}$$

$$P_z(t_k)^- = H(t_k)P(t_k)^- H(t_k)^{\mathrm{T}} \tag{11}$$

If desired, use of a Gaussian assumption can be made for the sequence of streamflow forecast errors resulting in capability for probabilistic flood occurrence predictions. Thus, if ϕ denotes the discharge corresponding to a prespecified flood stage at the outlet of a headwater basin, and the Gaussian probability density function of the streamflow $\zeta(t_k)$ at time t_k is denoted by $\not{p}(\zeta(t_k); \hat{z}(t_k), P_z(t_k)^-)$, then the probability $\mathscr{PR}(\phi, t_k)$ that the threshold ϕ will be exceeded at time t_k is given by

$$\mathscr{PR}(\phi, t_k) = \left(1 - \int_{-\infty}^{\phi} \not{p}(\zeta(t_k); \hat{z}(t_k), P_z(t_k)^-)\mathrm{d}\zeta(t_k)\right) \tag{12}$$

The key assumptions that lead to the numerically convenient form of eqns. (3)–(7) are that the continuous-time error process $\mathbf{w}(t)$ is a Guassian white-noise stochastic process with mean vector equal to the zero vector and covariance parameter matrix equal to $Q(t)$, and that the discrete-time error sequence $v(t_k)$ is a Gaussian independent sequence with mean equal to zero and variance equal to $R(t_k)$. The initial vector $\mathbf{x}(t_0)$ is also assumed Gaussian with mean vector $\mathbf{x}_0$ and covariance matrix P_0.

Prior to implementation of the estimation algorithm for a particular conceptual model it is required that values for the model parameters, $\boldsymbol{\theta}$, and for the estimator parameters, $Q(t)$ and $R(t_k)$ (for all t and k), be identified from available morphometric and hydrologic historical data from the site of interest. Thus, before the application of algorithm in real-time streamflow forecasting, a parameter identification study is required.

A detailed survey of parameter estimation techniques and applications in the area of real-time flow forecasting has recently been presented by Rajaram and Georgakakos (1987) and will not be attempted here. Some conclusions of that review relevant to the present study will be discussed below.

Even though estimation techniques have been extensively studied in the model parameter estimation context, very few methodologies have been proposed and tested for the estimation of the estimator parameters $Q(t)$ and $R(t_k)$. Of particular importance is the identification of the elements of $Q(t)$ since very little a priori information is available for it. The variance $R(t_k)$ of the streamflow observation error can be estimated in most cases reliably (especially in headwater basins with no looped rating curves) since it depends on the characteristics of the measuring device. The stochastic approximations technique has been used (Kitanidis and Bras, 1978; Georgakakos, 1984b) as the basis of a methodology for the determination of selected elements of matrix $Q(t)$ assumed independent of time t. The idea of the methodology is to enforce consistency at all times between the actual statistics of the one-step-ahead predicted residuals (residual = observed − predicted streamflow) and those statistics as they are predicted by the estimator in real time. This methodology is well suited to low dimensional Q matrices. It fails to perform well in cases of high dimensional systems and in cases when the model error is highly nonstationary (the matrix Q depends on time t). Recognizing the deficiency of existing methods to provide a general solution to the problem of determining values for the estimator parameters, Rajaram and Georgakakos (1987, 1989) proposed a conceptual framework that leads to an identification methodology that does not suffer from the aforementioned deficiencies.

Rajaram and Georgakakos (1987, 1989) proposed the decomposition of the model error process into two component processes: the model-parameter error process and the input error process. The conceptual model equations and a priori estimates of the model-parameter and input estimation errors were used to parameterize the statistics of the component processes. For model-parameter calibration by automated techniques, the final parameter error covariance matrix can be used as an a priori estimate of the model-parameter estimation error. For manual calibration of the model parameters by hydrologists, the hydrologists' degree-of-belief estimates of the parameter error variances can be used instead.

The proposed methodology requires the identification of only two free parameters from historical data. Because of the small number of free parameters, almost any statistically sound parameter estimation technique can be used for their identification. Previous application to large-sclae hydrologic–chemical models that simulate watershed response to acidic deposition has given encouraging results.

The objectives of this study were two-fold: (1) to apply the Rajaram and Georgakakos methodology to the real-time forecasting of streamflows using the National Weather Service River Forecast System (NWSRFS) soil moisture accounting scheme and a nonlinear storage routing model; (2) to verify the predictions of the calibrated model in a World Meteorological Organization (WMO) experiment that simulated real-time conditions and used data from two hydrologically dissimilar drainage basins: the Bird Creek basin in Oklahoma, U.S.A., and the Orgeval basin in France.

The next section develops the pertinent mathematical formulation that leads to the parameterization of the second moment of the model error process. It also indicates various approaches for the identification of the two free parameters of the formulation. Sections on pages 9, 16 and 22 present applications of the methodology to the two drainage basins of interest, and the verification results from the WMO experiment. Conclusions and recommendations for future research work are included in the section on p.25. For the details of the derivations and a complete mathematical formulation the reader is referred to Georgakakos et al. (1988).

PHYSICALLY BASED STATISTICAL CHARACTERIZATION OF HYDROLOGIC MODEL ERRORS

Model errors result from inaccurate specification of the model structure (e.g. nonlinear processes are represented as linear and spatially distributed processes are represented as spatially lumped), because of estimation errors in the values of model parameters, and because of measurement or estimation errors in the model input (e.g. estimation of mean areal precipitation and evapotranspiration). Use of the powerful results of linear estimation theory presented in the previous section requires the characterization of the model errors as an additive zero-mean white-noise random process (i.e. the process $\mathbf{w}(t)$ in eqn. (1)). The purpose of this section is to parameterize the covariance parameter matrix $Q(t)$ of this random process.

Errors in model structure are difficult to quantify because of the absence of "ground truth" models for the large-scale hydrologic process of runoff. It is expected, however, that structural errors would lead to a non-zero-mean random process $\mathbf{w}(t)$. The latter is because of the fact that structural errors would become important during the periods of time when the natural component process which is inaccurately represented in the model formulation becomes significant. During those periods of time, model structure errors are expected to be biases rather than zero-mean random errors. For the aforementioned reasons no attempt will be made herein to characterize errors in model structure. For the following development it is tacitly assumed that model errors are dominated by errors in the estimation of model parameters from historical data and by input estimation (or observation) errors. It is noted, however, that the proposed methodology can be used in a recursive parameter estimation framework to detect structural errors (Rajaram and Georgakakos, 1987, 1989).

The strategy for the derivation of a formula for the model error covariance parameter matrix $Q(t)$ is based on the following steps:

Step 1: Hypothesize models for the model errors because of parameter estimation and input estimation.

Step 2: Linearize the nonlinear vector function $\mathbf{f}(\mathbf{x},\mathbf{u};\,\boldsymbol{\theta})$ by expanding in a Taylor series expansion about the best estimates of $\mathbf{x}$, $\mathbf{u}$ and $\boldsymbol{\theta}$.

Step 3: Derive the differential equation that governs the propagation of the state covariance matrix using the linearized form of eqn. (1) obtained in Step 2 and the models of uncertainty hypothesized in Step 1.

Step 4: Obtain an expression for $Q(t)$ by comparing the equation derived in Step 3 with eqn. (4) and estimate free parameters (if any) from historical data.

It is noted that the methodology is based on a linearization of the process dynamics, eqn. (1), about the best estimates of the model state, input variables and model parameters. Thus it is expected to be most successful in cases of small deviations from these best estimates. Various applications to date however indicate that it is still satisfactory in cases of large estimation errors in input variables, state variables and parameters (Rajaram and Georgakakos, 1987, 1989). Also, note that the procedure outlined can use the state estimator-produced best estimates of the state variables in the linearization (Step 3). The mathematical development corresponding to each step of the procedure is given below.

Step 1

Consider the "true" values of the state, input, and parameter vectors as functions of their best estimates and residual errors. It holds that

$$\mathbf{x}(t) = \hat{\mathbf{x}}(t) + \mathbf{e}(t) \tag{13}$$

$$\mathbf{u}(t) = \hat{\mathbf{u}}(t) + \mathbf{e}_{\mathrm{u}}(t) \tag{14}$$

$$\boldsymbol{\theta}(t) = \hat{\boldsymbol{\theta}} + \mathbf{e}_{\theta}(t) \tag{15}$$

where the left-hand sides of eqns. (13)–(15) denote the true values, the first term of the right-hand side denotes the best estimates and the last term denotes the residual errors of the state, input and parameter vectors, respectively.

A time invariant estimate of the parameter vector has been assumed. This is consistent with the limited data bases existing for the calibration of large-scale conceptual hydrologic models. It is noted that a time varying model parameter vector has been assumed as the true parameter vector. This is in an effort to account for model structure errors that are manifested by requiring a different parameter set for various regimes of physical system response. The interested reader is referred to Rajaram and Georgakakos (1987) for an implementation of the procedure under the hypothesis of a constant true parameter vector. Estimates of the model state vector can be obtained recursively by the linear state estimator of eqns. (3)–(7). In particular the predicted estimate $\hat{\mathbf{x}}(t)^-$ should be used. Estimates of the input vector can be obtained from observations or estimation procedures external to the model dynamics (e.g. the estimation of the mean areal precipitation input).

It is hypothesized that the residual error terms in eqns. (14) and (15) are independent white-noise random processes with zero means and covariance parameters given by

$$E\{\mathbf{e}_{\mathrm{u}}(t)\mathbf{e}_{\mathrm{u}}(t')^{\mathrm{T}}\} = U(t)\,\delta(t - t') \tag{16}$$

$$E\{\mathbf{e}_{\theta}(t)\mathbf{e}_{\theta}(t')^{\mathrm{T}}\} = W(t)\,\delta(t - t') \tag{17}$$

where $E\{\cdot\}$ denotes expectation of a random variable or process, and $\delta(t - t')$ denotes the Kronecker delta function which is zero everywhere else but at the origin where it becomes infinity. Superscript T denotes transpose of a matrix or vector quantity.

Step 2

Expanding the nonlinear vector function $\mathbf{f}$ in a Taylor series about the best estimates of the state, input and parameter vectors yields to a first order approximation:

$$\mathbf{f}(\mathbf{x}, \mathbf{u}; \hat{\boldsymbol{\theta}}) = \mathbf{f}(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}}) + F(\hat{\mathbf{x}},\hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})\mathbf{e} + M(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})\mathbf{e}_{\mathrm{u}} + N(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})\mathbf{e}_{\theta} \tag{18}$$

where the residual error vectors have been defined in eqns. (13)–(15), $F(\cdot)$ has been defined in eqn. (8), and the (i,j)th elements of the matrices $M(\cdot)$ and $N(\cdot)$ are defined as follows:

$$[M(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})]_{i,j} = [\partial f_i(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})]/(\partial u_j) \tag{19}$$

$$[N(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})]_{i,j} = [\partial f_i(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})]/(\partial \theta_j) \tag{20}$$

The derivatives in eqns. (19) and (20) are evaluated at the best estimate values of the vectors $\mathbf{x}$, $\mathbf{u}$, and $\boldsymbol{\theta}$; and u_j and θ_j represent the jth elements of vectors $\mathbf{u}$ and $\boldsymbol{\theta}$, respectively. Also, dependence on time has been omitted for notational convenience.

Step 3

Given the assumptions in eqns. (16) and (17) regarding the covariance properties of the vector random processes $\mathbf{e}_{\mathrm{u}}$ and $\mathbf{e}_{\theta}$, the last two terms in the right-hand side of eqn. (18) can be replaced by the vector random process $\mathbf{w}$ with covariance parameter matrix Q given by

$$Q = M(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})UM(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})^{\mathrm{T}} + N(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})WN(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})^{\mathrm{T}} \tag{21}$$

The matrices U and W have been defined in eqns. (16) and (17), respectively.

Upon the introduction of the process $\mathbf{w}$, the propagation equation for the state estimator remains as in eqn. (14), with $Q(t)$ replaced by the expression in the right-hand side of eqn. (21).

Step 4

Implementation of the algorithm requires specification of the matrix covariance parameters U and W. Statistical input and model parameter estimation procedures provide estimates of U and W directly. In lieu of statistical estimation procedures for estimating $\mathbf{u}$ and $\boldsymbol{\theta}$, degree-of-belief estimates of the errors in input and parameter values can be used to specify the diagonal elements of these covariance parameter vectors with the off-diagonal elements set to zero. In order to account for erroneously specified U and W, two free parameters α_{u} and α_{p} are introduced so that the expression for Q in eqn. (21) is modified to

$$Q = \alpha_{\mathrm{u}}M(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})UM(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})^{\mathrm{T}} + \alpha_{\mathrm{p}}N(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})WN(\hat{\mathbf{x}}, \hat{\mathbf{u}}; \hat{\boldsymbol{\theta}})^{\mathrm{T}} \tag{22}$$

Estimation of the two free parameters from historical data is necessary before the proposed state estimator is used in real-time forecasting of streamflows. A criterion should be set and best estimates of the free parameters should be obtained by minimizing the estimation criterion. Since the free parameters under study are estimator parameters rather than model parameters, it is proposed that they be estimated by forcing consistency between the estimator-predicted streamflow residual variance and the actual streamflow residual variance. The stochastic approximations technique can be utilized for this purpose. Rajaram and Georgakakos (1987) present the relevant formulation. However, since there are only two free model parameters, one can also construct contours of various objective functions in the $(\alpha_{\mathrm{u}}, \alpha_{\mathrm{p}})$ parameter space and, then, find the values of the two free parameters that optimize the objective functions. A suitable objective function in this case would be the variance of the normalized streamflow prediction residuals. An optimal value of 1 enforces consistency between the predicted and actual streamflow residuals.

It is noted that the procedure outlined in Steps 1 to 4 reflects the nonstationary nature of model errors in a physically consistent manner. Even though the matrices U and W may remain constant at all times, the derivative matrices $M(\cdot)$ and $N(\cdot)$ are time dependent and modify the strength of the uncertainty in various input and model-parameter variables in accordance with the dynamics represented in the conceptual model. This is perhaps the most important contribution of this novel approach. In addition and for operational applications of the methodology, hydrologic expertise is utilized for the specification of degree of belief estimation of U and W. It is also noted that a plot of the normalized residuals as a function of time (after the two free estimator parameters have been identified) can indicate periods of potential errors in the model structure. Persistent excursions of the normalized residuals beyond the interval: $[-5, +5]$ should be associated with such errors.

HYDROLOGIC MODELS AND DRAINAGE BASINS USED IN THE APPLICATIONS

This section presents the components of a conceptual hydrologic rainfall–runoff model that was used in the applications of the forecasting methodology previously described. The hydrologic model is composed of the modified Sacramento soil-moisture accounting model and a nonlinear reservoir channel routing model. Data from two different drainage basins, the Bird Creek basin in Oklahoma, U.S.A., and the Orgeval basin in France, are used in the test cases. Following is a short description of the hydrologic-model components and of the drainage basins. For a more detailed description the reader is referred to Georgakakos et al. (1988).

The time-continuous hydrologic rainfall–runoff model

The rainfall–runoff model used in this work is of the conceptual type. That is, even though the model equations are not expressions of the differential laws of conservation of heat, mass and momentum for the water substance, they do describe large spatial and temporal scale conservation and response laws that are in accordance with the observed large scale behavior of water in hydrologic drainage basins. The model is a spatially lumped one in that it treats the hydrologic basin as one unit with no regard for the spatial distribution of the modeled physical processes. One of the model features that make it particularly attractive for real-time prediction applications is its modest requirements in input data for adequate performance. At the current stage of development, the hydrologic model has no component for the computation of snowmelt runoff.

The modified Sacramento model

The soil-moisture accounting component of the rainfall–runoff model used in the tests is the modified Sacramento model, recently documented in Georgakakos (1986a). It is a conceptual spatially-lumped model suitable for application to headwater drainage basins. It accepts mean areal precipitation and mean areal evapotranspiration as input, and produces total channel inflow as output. The flow components that contribute to the total channel inflow are: direct runoff from impervious areas, surface runoff in cases of excessive rainfall rates, interflow through the upper soil layers, and groundwater flow. The model subdivides the drainage basin into two zones, an upper and a lower zone. The upper zone simulates water stored in the upper soil layers which is available for evapotranspiration, percolation, surface runoff and interflow. The lower zone simulates groundwater storage. Both zones have tension-water elements and free-water elements. The former simulate water which can be extracted only via evapotranspiration while the latter simulate water that is "free" to move under the action of gravity. The upper soil zone contains one tension and one free-water element, while the lower zone contains one tension and two free-water elements. The two free-water elements of the lower zone were introduced by the originators of the model in order to better approximate the recession limbs of hydrographs using the linear outflow from the lower zone free-water elements. An additional impervious area storage element was introduced in an effort to account for the temporal changes of the saturated soil area during flooding events and, therefore, to partially account for the spatial distribution of that area as a function of time.

Several versions of the Sacramento soil-moisture accounting model have been published in the literature. The original version (Burnash et al., 1973) was a discrete-time version of the model. Kitanidis and Bras (1978) presented a simplified state-space form of the discrete-time version. Georgakakos (1986a) presented a continuous-time state-space form of the model that closely approximates the original discrete-time model. It is the Georgakakos (1986a) version of the model that we have used in the applications. We will refer to that version as the modified Sacramento model. For the details of model formulation the interested reader is referred to Georgakakos (1986a) and Georgakakos et al. (1988).

The channel routing model

When it is impossible to collect cross-sectional data and to identify the spatial distribution of channel inflow along the channel length, conceptual models of channel routing can be used. A class of such models is based on a generic element that is a linear or nonlinear reservoir. Thus, cascades or tree-like patterns of conceptual reservoirs can be used to simulate the water flow of a natural stream network.

Several channel routing models based on a cascade of reservoirs have appeared in the literature (Unny and Karmeshu, 1984). For the purposes of this work and in accordance with the data available, regime of application (headwater basins with "kinematic" water flow), and soil-moisture accounting model used (see previous section), a series of n nonlinear reservoirs is used as the channel routing model. This model is based on the formulation originally proposed by Mein et al. (1974) and brought to a state-space form by Georgakakos and Bras (1980, 1982).

The channel routing model contains two parameters that define the relationship between the storage and the outflow of each of the conceptual reservoirs. It is noted that the relationship implies a one-to-one correspondence between discharge and stage which is a characteristic of kinematic channel routing methods. Thus, use of this model should be restricted to headwater areas with relatively steep slopes.

The conservation of mass (or volume) equation applied to each of the conceptual reservoirs with the assumption that the total channel inflow enters the cascade at the upstream end (a modeling assumption) provides an expression for the dynamical equation of the routing component. Estimation of the two free routing-model parameters can be achieved with trial and error procedures or recursive identification methods based on modern estimation theory (Georgakakos and Bras, 1980). The number n of reservoirs in the series can be obtained from the response time of the drainage network (Georgakakos, 1987).

State updating

The state updating procedure described in the introductory section (see eqns. (3)–(9)) uses the differential equations of the hydrologic rainfall–runoff model together with descriptors of the input and parameter uncertainty and produces updated estimates of the model states given observations of discharge at the drainage basin outlet. Fundamental in the state-update equations is a linearization of the hydrologic model equations about the current "best" estimates of the states. Such linearization provides a means by which uncertainty is propagated in time and is updated when observations become available. It is noted that the linearization does not affect the predictions of the mean value of the state vector between observations and, thus, those predictions are identical to the predictions that the deterministic model would compute if it were using the same input, parameters and initial conditions.

Integration of model differential equations

A variable step-size, fourth-order predictor-corrector method was used to integrate the coupled state mean and covariance eqns. (3) and (4). The state mean equations are $(6 + n)$ first-order nonlinear differential equations, where n is the number of conceptual reservoirs in the channel routing model and 6 is the number of the states of the modified Sacramento model. The state covariance equations are $(6 + n)^2$ first-order linear time-varying differential equations that are coupled with the state mean equations through the elements of $F(t)$ (see eqn. (4)). Since the state covariance matrix is a symmetric matrix, only $(6 + n)(6 + n + 1)/2$ differential equations were integrated during each time step.

The integration method is based on Hamming's modified predictor–corrector method (Ralston and Wilf, 1960; Ralston, 1962) and it allows for a tolerance error input and a time-step initial subdivision indicator. If during integration the error bound is not satisfied by the local truncation error, the time interval is halved and integration is attempted again for the new discretization. If, on the other hand, during integration the errors are estimated to be significantly smaller than the input error bound, the time interval is doubled and integration continues. A maximum of 10 halvings is allowed.

Integration errors are another type of model errors. Since the nonlinearities of the hydrologic model differential equations (in particular those of the upper zone elements of the modified Sacramento model) affect the model predictions the most when the elements are reaching saturation, integration errors depend significantly on the magnitude of the rainfall input and on the degree of saturation of the model elements (conceptual reservoirs). This, of course, is particularly true when the duration of the time step is long. Georgakakos et al. (1988) present results from a sensitivity analysis with respect to the parameters of the integration procedure. It was evident that for the same accuracy, the saturated-soil case requires many more initial subdivisions of the 6-h time step than the unsaturated case. However, the more subdivisions of the time step one enforces the longer the central processing unit (CPU) time becomes. Since the case of saturated soil and excessive rainfall is a rare one in the calibration period and in the interest of completing the calibration and the verification runs in reasonable time, it was decided to set the number of initial subdivisions equal to 4. This might lead to inaccuracies in cases of rare events in a saturated soil. An input and soil-saturation degree dependent integration method for at least the state mean equations might improve the accuracy without increasing CPU time.

The Bird Creek hydrologic basin: characteristics and data

The Bird Creek drainage basin is in Oklahoma, located near the northern state border with Kansas. The outlet of the basin is near Sperry, OK, and it is located ~10 km north of Tulsa, OK (Fig. 1). The area of the drainage basin is equal to 2344 km^2. The terrain is gently rolling hills, and there are no mountains or large water surfaces that influence its climatic conditions. The elevation ranges from 175 m to 390 m above mean sea level. The basin vegetative cover is mainly grassland with 20% forested area. The soils have large storing capacities and the entire basin is underlain by groundwater aquifers (Smith, 1986).

The climate of the Bird Creek basin area is classified as humid because of significant rainfall during most years, but the basin also experiences extended periods with very light rainfall. Spring and summer are the wettest seasons with rain in the form of showers and thundershowers of convective origin. Hail accompanying thunderstorms is not uncommon. Snowfall amounts are light and remain on the ground for a very short time. Sunshine is abundant. Significant evapotranspiration losses occur in the basin because of the high air temperatures (100°F is common) from the latter part of July to September, accompanied by low relative humidity and a good southerly breeze.

The calibration and verification data for Bird Creek consisted of 6-h values of mean areal precipitation, daily values of mean areal potential evaporation, and 6-h and daily values of outflow discharge. The mean areal precipitation values were computed based on data from stations located both within and outside the basin boundaries (Fig. 1) and on U.S. National Weather Service procedures (Larson, 1975; Larson and VanDemark, 1979; Smith, 1986). A total of 12 precipitation stations were used of which five were hourly recording stations. The daily potential evaporation values were computed from meteorological data recorded at the Wichita, Kansas, Air Force base, located ~112 km northwest of the Bird Creek basin. Computation was based on operational U.S. National Weather Service procedures (Hydrologic Research

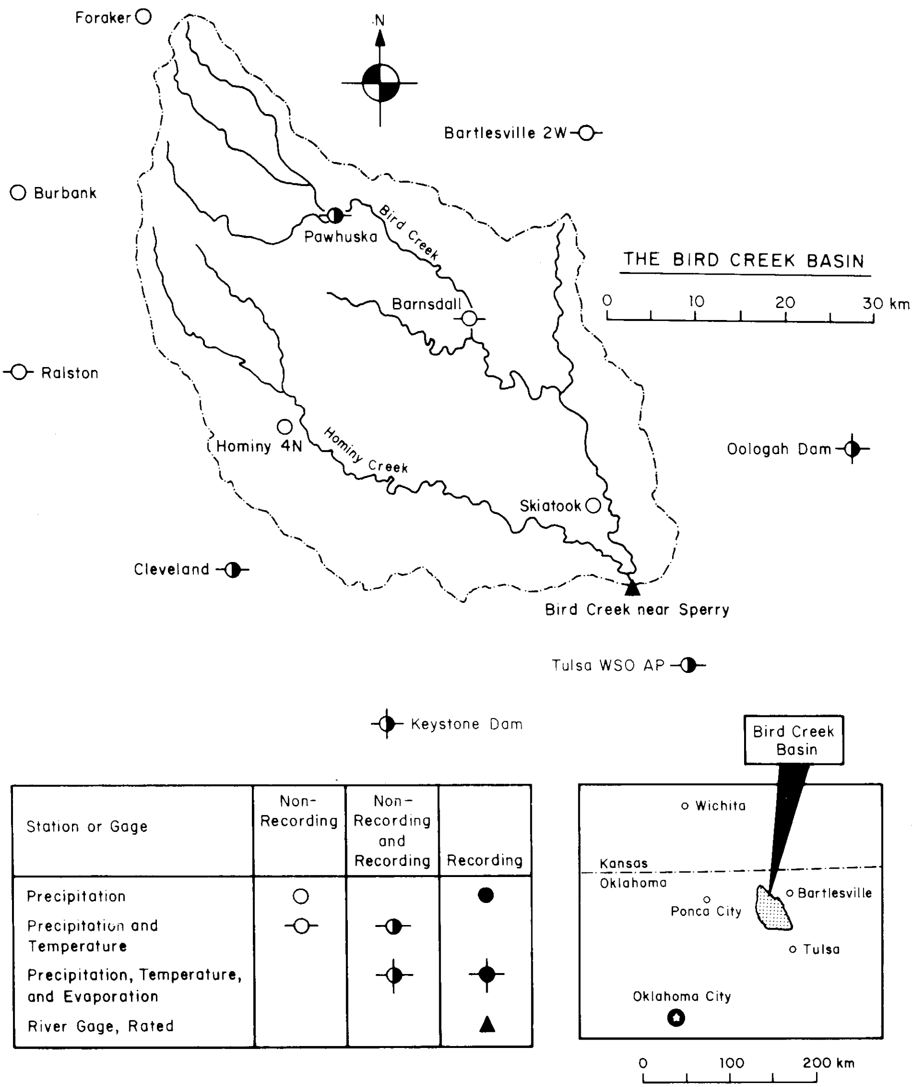


Fig. 1. The Bird Creek drainage basin in Oklahoma, U.S.A., and its observation stations.

Laboratory, 1972; Day and Farnsworth, 1982). The daily mean areal potential evaporation values were multiplied by an adjustment factor, determined by staff of the Hydrologic Research Laboratory, NWS, National Oceanic and Atmospheric Administration (NOAA), during the calibration phase, to account for the local basin conditions and transpiration. Six-hourly values of mean areal potential evapotranspiration were computed by dividing the adjusted daily mean areal potential evapotranspiration into four equal parts. Six-hourly discharge data were used in this work and they are available for high flow periods of record. The discharge values were computed from continuous stage recorder data (U.S. Geol. Surv. St. No. 07177500) and rating tables provided by the U.S. Geological Survey. Consistency tests were performed on the three data time series for the period October 1955–December 1974 (Smith, 1986). All three data sets were found to be consistent for the entire period.

The eight-year calibration period for Bird Creek was from October 1955–September 1963, while the verification events were in the period from November 1972–November 1974. It is noted that 6-h discharge values were only available for high flow periods while daily discharge values were available for the whole period. The discharge at the basin outlet ranged from $0\,m^3\,s^{-1}$ to $2540\,m^3\,s^{-1}$ during the calibration period. The highest discharge of the verification period was $1506\,m^3\,s^{-1}$.

The Orgeval hydrologic basin: characteristics and data

The Orgeval basin is located at a distance of ~ 40 km east of Paris. The Orgeval is a secondary tributary of the Marne River joining the right bank of the Grand Morin River (Fig. 2). The area of the catchment is approximately equal to $104\,km^2$. Its average elevation is 148 m with a highest elevation of ~ 182 m and a lowest elevation of ~ 70 m. A prominent feature of the basin is a sharp decrease in elevation below the elevation of 130 m (Fig. 2). The catchment is situated entirely in rural areas, with only 1% of the total surface area devoted to urban areas and roads. Most of the plant cover consists of crops — cereals and beetroot, or fenced rangeland. Forests occupy 18% of the total catchment area. Approximately 50% of the basin's arable land has been drained. The groundwater table is very close to the ground surface all over the basin during humid periods and its form follows the topography to a considerable extent. This is a result of low permeability formations causing a build-up of shallow water during humid periods (Askew, 1986).

The calibration and verification data for the Orgeval drainage basin consisted of hourly values of mean areal precipitation, daily values of mean areal potential evaporation, and values of hourly averaged outflow discharge. The mean areal precipitation was computed by the staff of Centre National du Machinisme Agricole, du Genie Rural, des Eaux et des Forets, Hydrologic-Hydraulic Division, based on the data from three rainfall recording stations located at elevations of 130 m, 146 m, and 174 m (Fig. 2). The daily mean areal potential evaporation data was based on the readings of a C-type evaporation pan located at the Biossy-le-Chatel climate station near the catchment boundary (Fig. 2). Monthly adjustment factors for Orgeval were computed by staff of the Hydrologic Research Laboratory, NWS, NOAA, during the calibration phase. The procedure followed for the determination of 6-h mean areal potential evapotranspiration values in the case of Bird Creek was also followed in the case of Orgeval to determine the hourly mean areal potential evapotrans-

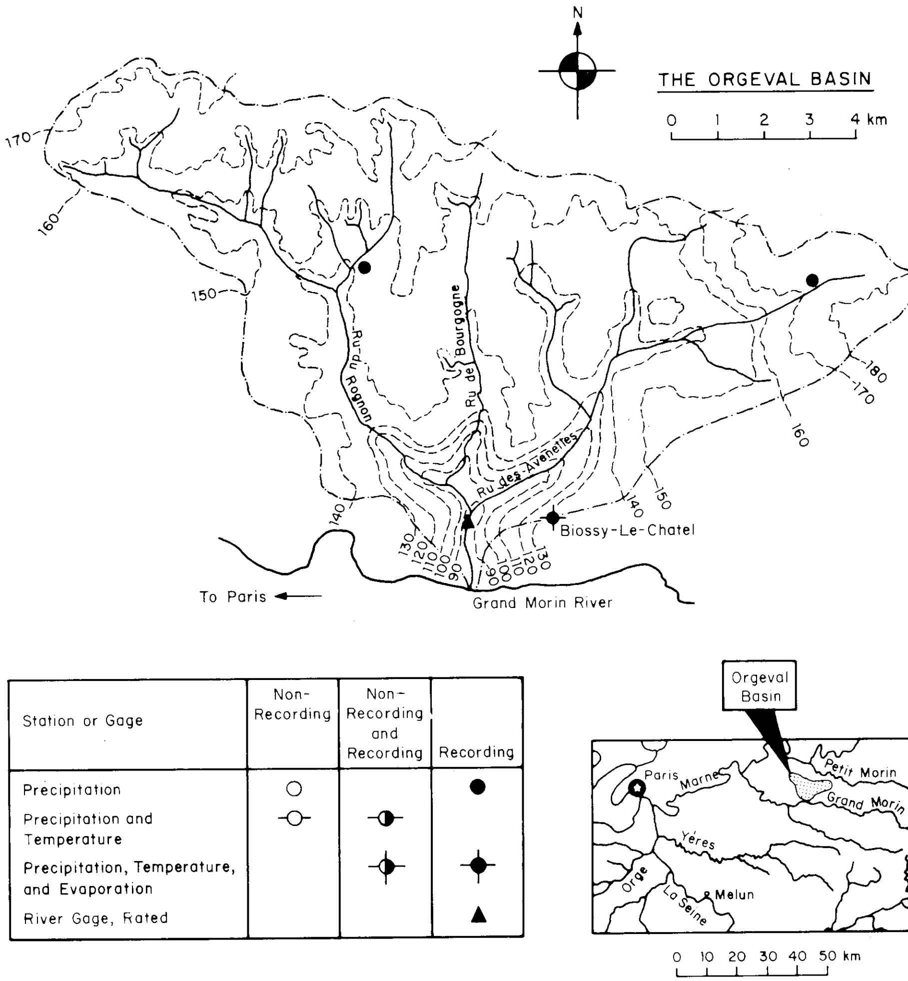


Fig. 2. The Orgeval drainage basin in France and its observation stations.

piration values. In the Orgeval case, the daily values of mean areal potential evapotranspiration were divided by 24 in order to generate hourly values.

The six-year calibration period for Orgeval spanned the interval from October 1972 to September 1978, while the verification events were in the period from December 1978 to July 1980. During the calibration period of record the hourly averaged discharge at the basin outlet ranged from 0 $m^3 s^{-1}$ to about 21 $m^3 s^{-1}$. The highest hourly averaged discharge of the verification period was about 29 $m^3 s^{-1}$.

ESTIMATION OF THE PARAMETERS OF THE STOCHASTIC-DYNAMICAL HYDROLOGIC MODEL

This section presents a summary of the procedures used for the calibration of the parameters of the hydrologic model and for the estimation of the two free parameters of the state estimator (namely, α_u and α_p).

The modified Sacramento model and the channel routing model described previously contain several free-model parameters whose values should be identified based on historical time-series and drainage basin data. The procedure followed for the determination of estimates for the model parameters that are consistent with the observed data will be referred to as calibration in the following. Actual observations of the model input variables were used as input data for the calibration and verification periods of record, instead of forecasted values of input variables as the case would be in true real-time operation. This "perfect foresight" scenario was used in an effort to examine the performance of the hydrologic model free from the dominating inconsistencies between model forecasts and observations of streamflow caused by forecast-input errors (espcially in forecasting precipitation in real time with a forecast lead time of several hours). In that respect the case studies presented in this work indicate upper bounds of performance rather than typical performance of the model under real-time conditions. Georgakakos and Hudlow (1984) present operationally available quantitative precipitation forecast techniques suitable for use in hydrologic forecasting. Georgakakos (1986b) presents an analysis of the hydrologic model forecast errors as a result of erroneous input forecasts for the Bird Creek drainage basin.

The calibration procedures and results for both the modified Sacramento and the channel routing models are presented below.

Estimation of the modified Sacramento model parameters

Recently, Georgakakos and Brazil (1987) described the calibration strategy followed routinely by the Hydrologic Research Laboratory, NWS, NOAA, for the calibration of the modified Sacramento model. It is a combination of manual (trial and error guided by hydrologic experience) and automatic optimization procedures. Such procedures were used to determine values for the parameters of the modified Sacramento model for the two drainage basins. Calibration of the model for the Bird Creek basin was performed in the early 1970s by R.J.C. Burnash, an experienced hydrologist at the Sacramento River Forecast Center, for the first World Meteorological Organization rainfall-runoff modeling intercomparison project. The discrete-time version of the model was used with a layered coefficient routing model. Degree-of-belief estimates for the standard deviation of the estimation errors were determined by the staff of the Hydrologic Research Laboratory, NWS, NOAA. E.A. Anderson, an experienced senior research hydrologist at the Hydrologic Research Laboratory, calibrated the discrete-time modified Sacramento model coupled to a Lag and K channel routing model for the Orgeval basin following a combination of manual and automatic optimization methods. A total of seven manual calibration runs were performed to establish good initial estimates for the parameters. An automatic gradient search procedure was used to improve on the estimation of a few parameters such as the interflow recession coefficient. Several comments (by E.A. Anderson) regarding the Orgeval calibration procedure and results follow.

Based on an air-temperature time series for the climatic station near Orgeval it was determined that during the calibration period there were a few cases of snowmelt runoff. Thus, runoff was delayed during these periods because of melting of presumably solid precipitation. An appropriate time shift was applied to the discharge data during these periods so that calibration runs for the entire calibration period could be made without the use of a snowmelt model. Since these periods were very few, no significant effects of the adjustment procedure on the quality of the parameter estimation results are expected.

The upper zone free-water element never filled during the October 1972–September 1978 period. The interflow recession coefficient was found to be dependent on the contents of the upper zone free-water element and it is questionable if a surface runoff threshold exists for this element. Since the modified Sacramento model assumes a linear recession of the element, a value for the recession coefficient was selected that was representative of the values it took for the range of values of the upper zone free-water element obtained during the calibration runs. This is a potential structural error of the model and it may have significant effects on the model predictions during periods of flows dominated by interflow (e.g. winter months for the Orgeval basin). A nonlinear interflow recession is needed. Because of these considerations an inflated standard deviation was assigned to the estimate of the saturation value of the upper zone free-water element.

Estimation of the channel routing model parameters

Two methods are possible for the estimation of the routing model parameters depending on the data available for calibration. The first method uses hydromorphological data collected from the particular basin of interest. The second method uses historical time series data of channel inflow and channel outflow. Mein et al. (1974) and Georgakakos and Bras (1980) describe the first method of parameter estimation. It is based on the subdivision of the channel network into reaches of homogeneous hydrological properties, and use of the equations for kinematic routing in wide rectangular channels for the flow computation in each reach. The main shortcoming of the method is that it is usually difficult to identify a few such homogeneous reaches in natural headwater-basin channel networks. The second method uses (a) the channel network response time to identify the number of conceptual reservoirs in the series that defines the routing model, and (b) parameter estimation methods based on modern estimation theory applied to historical time series input and output data, in order to find estimates for the parameters of the storage–discharge relationship. Because of the available data for the two catchments of interest we used the second method of parameter estimation for the test cases examined.

Determination of the parameters of the conceptual reservoir routing model for the two hydrologic drainage basins was based on a systematic search procedure. The calibrated deterministic modified Sacramento model was used with data from the calibration period to produce a time series of total channel inflow values for the two basins under study. Using the total channel inflow as an input time series, several runs of the dterministic channel routing model were performed with various combinations of parameter values and the set of parameters that reproduced the "best" performance in terms of predicted discharge hydrographs was selected as the "optimal" set. Hydrologic performance criteria were used to quantify performance. In particluar, hydrograph peak time and peak-magnitude errors were minimized over the set of parameter values considered during the search.

Georgakakos et al. (1988) presents the values of the best estimates of the hydrologic model parameters. Examples of predictions of the calibrated hydrologic model for the two basins are given in Figs. 3 and 4 for the Bird Creek basin, and in Figs. 5 and 6 for the Orgeval basin. Those figures present predictions of discharge by the calibrated deterministic model (dashed lines) together with the corresponding observations for selected historical flood events (solid lines). The corresponding values of the mean μ and standard deviation σ of the one-step-ahead predicted residuals are shown on the figures. Also shown are the coefficient C_f of efficiency and the coefficient C_p of persistence. The former coefficient is a measure of the variance explained by the model and it takes the value of one (1) for perfect performance. The latter coefficient compares the least-squares performance of the model with the performance of a naive model that predicts the current observation at each time step. Positive values of the coefficient of persistence are desirable. Negative coefficient values indicate that the performance of the hydrologic model is

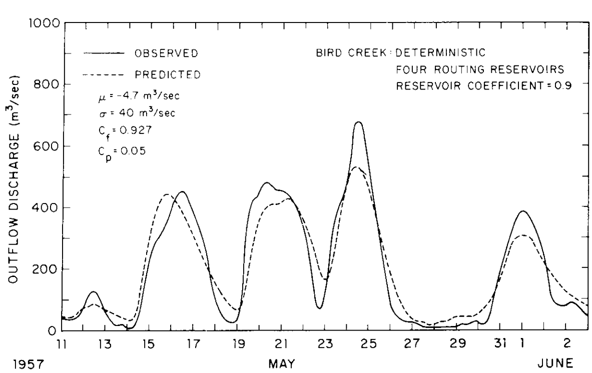


Fig. 3. Six-hourly predictions of discharge (dashed line) in $m^3 s^{-1}$ by the calibrated deterministic hydrologic model for the Bird Creek basin, May 1957. Observations are shown as a solid line.

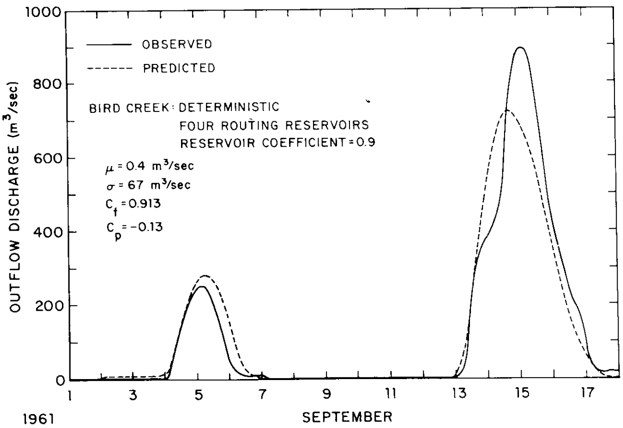


Fig. 4. Six-hourly predictions of discharge (dashed line) in $m^3\,s^{-1}$ by the calibrated deterministic hydrologic model for the Bird Creek basin, September 1961. Observations are shown as a solid line.

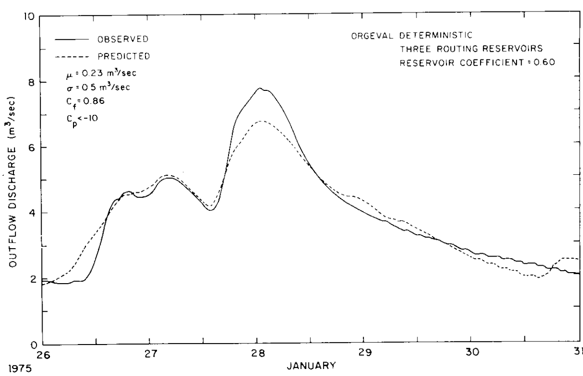


Fig. 5. Hourly predictions of discharge (dashed line) in $m^3\,s^{-1}$ by the calibrated deterministic hydrologic model for the Orgeval basin, January 1975. Observations are shown as a solid line.

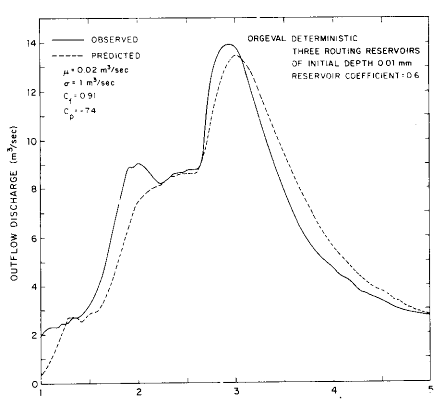


Fig. 6. Hourly predictions of discharge (dashed line) in $m^3\,s^{-1}$ by the calibrated deterministic hydrologic model for the Orgeval basin, beginning of February 1978. Observations are shown as a solid line.

worse (in a least-squares sense) than the performance of the naive model. Both coefficients are defined below in terms of model predictions and observations of discharge. C_f is defined by

$$C_f = 1 - S^2/S_0^2 \tag{23}$$

with

$$S^2 = \sum_{i=1}^{N} [z_Q(t_i) - z_{\text{PRED}}(t_i)]^2 \tag{24}$$

and

$$S_0^2 = \sum_{i=1}^{N} [z_Q(t_i) - z_A]^2 \tag{25}$$

where N is the number of observations in the period under study, $z_Q(t_i)$ is the observed discharge at time t_i, $z_{\text{PRED}}(t_i)$ is the model predicted discharge at time t_i and z_A is the time average of the observed discharges during the period from t_1 to t_N. Using the same notation, C_p is defined by:

$$C_p = 1 - S^2/S_p^2 \tag{26}$$

with

$$S_p^2 = \sum_{i=2}^{N} [z_Q(t_i) - z_Q(t_{i-1})]^2 \tag{27}$$

Estimation of model error parameters

The methodology proposed in the section on p.5 was followed for the determination of the model error covariance parameter Q. The two free parameters α_u and α_p of the proposed formulation (see eqn. (22)) were estimated via a systematic search in the two-dimensional parameter space for the optimal value of several objective functions that reflected both least-squares and hydrologic criteria. Only one-step-ahead predictions were utilized in the computation of performance criteria.

The least-squares criteria used were the absolute value of the time-average of the normalized residuals E_1, and the difference from 1 of the time-standard-deviation of the normalized residuals E_2. The hydrologic criteria used were the average errors in the hydrograph peak timing and magnitude for the significant floods of the calibration record. The first two criteria represent time-averaged statistics for the calibration period without special consideration for flooding periods or low-flow periods. The last two quantify errors in hydrologically significant features of the flood hydrographs.

Optimal values of criteria E_1 and E_2 ensure that the state estimator performance is as predicted by the optimal linear theory results. E_1 takes a value of zero (0) for optimal performance. E_2 also has an optimal value of zero (0).

Optimal regions in the space of the two state-estimator parameters were determined rather than optimal values of parameters. Use of an optimal region rather than an optimal set of values of α_u and α_p is necessitated because of: (a) the fact that only a finite calibration period (4 months duration) was used in the determination of parameter values, and (b) expected numerical errors in the integration of model equations and in the filter computations that prevent very precise computation of performance criteria values.

Degree-of-belief estimates of the standard deviation of the errors in the model input variables: mean areal precipitation and mean areal potential evapotranspiration, were used. The standard deviation of the errors at observation time t_i, $\sigma_u(t_i)$, was given by

$$\sigma_u(t_i) = c_u z_u(t_i) + c'_u \tag{28}$$

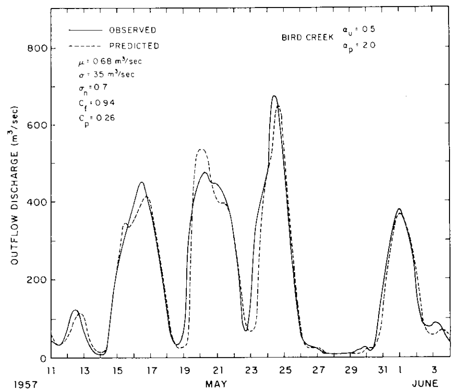
with u representing any of the two input variables, $z_u(t_i)$ representing the observed value of u at time t_i and c_u, c'_u are positive constants which were defined for each case of input variable and drainage basin (see Georgakakos et al., 1988). It is noted that the assumed input errors are modeled as nonstationary random sequences and that the errors increase as the magnitude of the observations increases. The cross-correlation between the errors in mean areal precipitation and in mean areal potential evapotranspiration was assumed equal to zero. It is important to note that given the perfect foresight scenario adapted in this work, the errors whose statistics are given by eqn. (28) are only because of the on-site observation errors of the instruments and the spatial interpolation procedure used to obtain mean areal values from point estimates. No input forecast error exists.

A similar model was assumed for the standard deviation of the errors in the observations of discharge at the outlet of the drainage basins studied. That is:

$$\sigma_Q(t_i) = c_Q z_Q(t_i) + c'_Q \tag{29}$$

where $\sigma_Q(t_i)$ is the standard deviation of the error in the observation of discharge $z_Q(t_i)$ at time t_i, and c_Q and c'_Q are constants defined for each drainage basin (see Georgakakos et al., 1988). It is noted that this too is a nonstationary model with the magnitude of errors increasing as the discharge observation increases. The values indicate that the observed discharge time series for both basins was of high quality. The variance of the observation errors $R(t_i)$ (see eqn. (5)) is given by the square of the right-hand side of eqn. (29). The Fortran Program TSFP (Georgakakos, 1984a) was used for the parameter search to generate contours of the performance criteria in the two dimensional parameter space of α_u and α_p.

A general comment on the 2-D parameter search runs is that they are CPU time intensive. This is especially true for the run corresponding to the Orgeval basin with hourly observations. Thus, the selection of the time period of the search run and the preliminary runs that define the parameter space to be searched become important factors in the successful identification of the optimal regions.



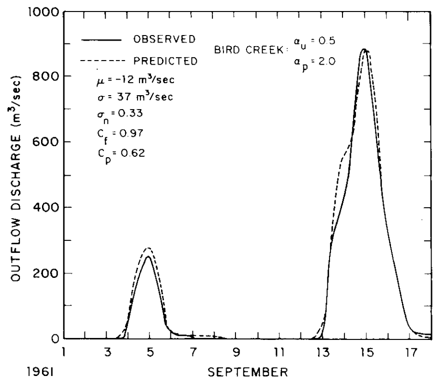


Fig. 7. Bird Creek 6-h predictions of discharge (dashed line) in $m^3 s^{-1}$ by the stochastic–dynamic hydrologic model with $\alpha_u = 0.5$ and $\alpha_p = 2.0$, May 1957. Observations are shown as a solid line.

Fig. 8. Bird Creek 6-h predictions of discharge (dashed line) in $m^3 s^{-1}$ by the stochastic–dynamic hydrologic model with $\alpha_u = 0.5$ and $\alpha_p = 2.0$, September 1961. Observations are shown as a solid line.

Several runs simulating real-time conditions were made with α_u and α_p values that were selected from within the optimal regions and for various time intervals within the calibration record of each basin (see Georgakakos et al., 1988). These runs were made in an effort to also take into account performance with respect to hydrologic criteria (timing and magnitude of peak flows) in the final selection of estimates for the two parameters.

Comparison of the predictions of the calibrated stochastic runs of Figs. 7 and 8 with the corresponding predictions of the deterministic runs of Figs. 3 and 4 for the Bird Creek basin shows significant improvement of the one-step-ahead predictions as a result of the use of the state estimator for real time state updating. A similar comment can be made after a comparison of Figs. 9 and 10 with Figs. 5 and 6 corresponding to the Orgeval basin. In Figs. 7, 8, 9 and 10, σ_n represents the standard deviation of the one-step-ahead predicted residuals which have been normalized by the filter-predicted innovations variance (Kitanidis and Bras, 1978). Both statistical indices and predictions of hydrograph characteristics are better in the stochastic-dynamic runs.

WMO TESTS SIMULATING REAL-TIME CONDITIONS

The World Meteorological Organization has sponsored several model intercomparison projects and most recently began in 1984 a project on the intercomparison of hydrologic models under simulated real-time conditions. The major event of this project was a workshop held in Vancouver, at the University of British Columbia in July and August of 1987. Modelers from throughout the world participated and compared their models and updating techniques for 10 days at the workshop.

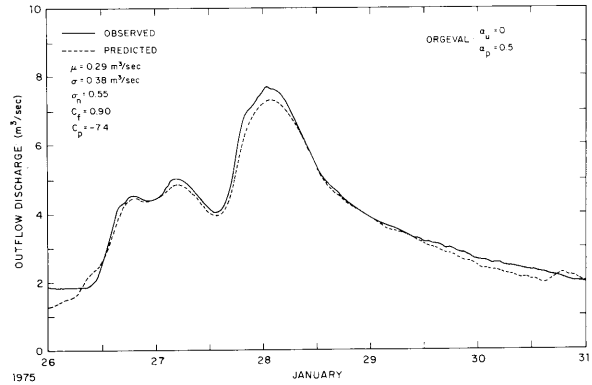


Fig. 9. Orgeval hourly predictions of discharge (dashed line) in $m^3 s^{-1}$ by the stochastic–dynamic hydrologic model with $\alpha_u = 0$ and $\alpha_p = 0.5$, January 1975. Observations are shown as a solid line.

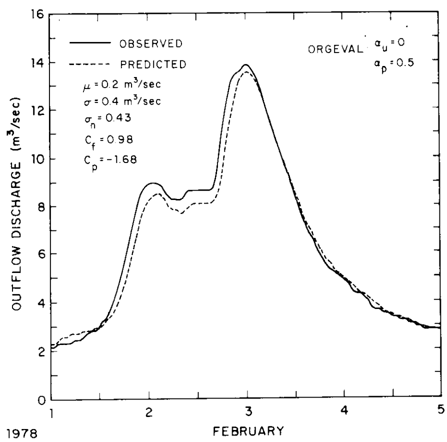


Fig. 10. Orgeval hourly predictions of discharge (dashed line) in $m^3 s^{-1}$ by the stochastic–dynamic hydrologic model with $\alpha_u = 0$ and $\alpha_p = 0.5$, beginning of February 1978. Observations are shown as a solid line.

The stochastic dynamic model and updating procedure described above participated in this model intercomparison workshop as the Hydrometorological Forecasting System (HFS). It will subsequently be referred to as the HFS model. Data for selected events on several river basins were available to the modelers. Each modeler was free to choose which basin or basins they wanted to model. The HFS model was run on the Bird Creek and Orgeval basins.

Under the simulated real-time conditions, data for each of six events were made available to the workshop participants as it would be for real-time forecasting. Seven forecasts were made for each event. The real-time nature of the forecasts was simulated by providing precipitation and observed discharge data up to the start of a forecast period, and then providing only precipitation data for the period to be forecast. When the forecast was turned in to the workshop moderator, new observed discharge and precipitation data were given to make the next forecast. In this way the WMO moderators were able to simulate the sequence in which data become available in real-time forecasting but greatly reduce the time needed to run the experiments.

Results for the HFS model for one of the six events for the Orgeval Basin are presented in Fig. 11. The solid line shows observed discharge data and the line marked with solid diamonds represents the deterministic simulation (i.e. no updating based on observed discharge). Each of the shorter lines indicates the results for a forecast with the HFS model at a specific time during the event. For example, the line with the open squares is the fourth forecast made by the HFS model for this event. It is a forecast for hours 11–20. The last observed discharge value for this forecast was for hour 10 and precipitation data were available to hour 20. Notice that as more observed discharge data become available the HFS model better simulates the peak of the event. See Georgakakos et al. (1988) for a similar presentation of results for each of the six events on the Bird Creek and Orgeval basins as simulated by the HFS model.

Figure 12 shows one set of summary data over all events for the Orgeval basin. Each of the vertical columns of points represents data for one event. Notice that the no-update results (open squares) lie farthest from the 45 degree line for every peak discharge forecast. For subsequent predictions (i.e. as more observed discharge data become available) the peak discharge prediction improves in each case.

Figure 13 presents the prediction error for various forecast lead times averaged over all forecasts for all of the six events used with the Bird Creek

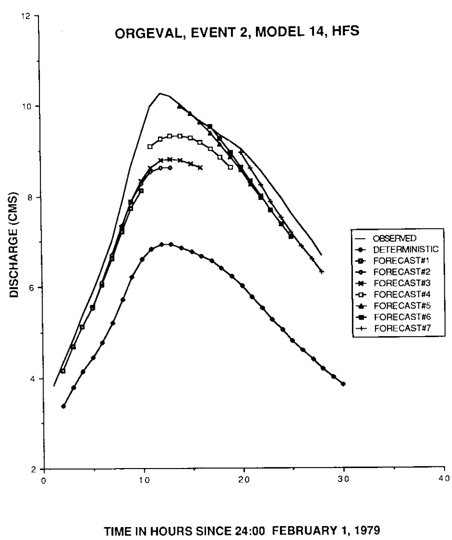
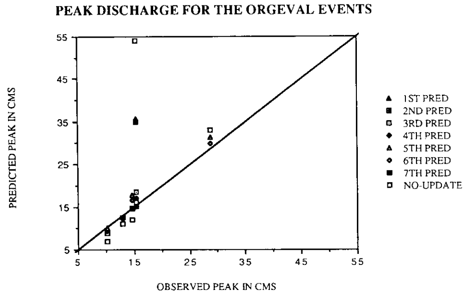


Fig. 11. Observed and forecast discharges for Orgeval, Event No. 2.



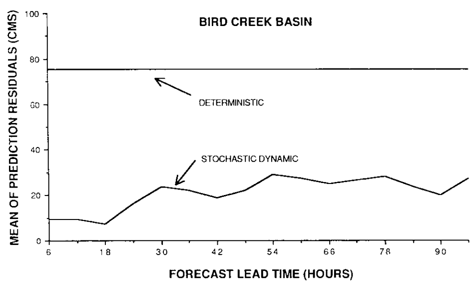


Fig. 12. Predicted vs. observed peak discharge in $m^3 s^{-1}$ for the Orgeval verification events. Predictions 1–7 are included if they cover periods with at least one hydrograph peak.

Fig. 13. Mean of prediction errors in $m^3 s^{-1}$ for various forecast lead times. Data from all six events of the Bird Creek basin were used.

basin. The deterministic prediction error is also averaged over the six events. The HFS model consistently produces a smaller prediction error than the deterministic model. The trend of the mean prediction error increases with the forecast lead time for the HFS model.

Additional information on the WMO Intercomparison Project including descriptions of all the participating models and their update procedures, as well as a compilation of comparative graphical and numerical verification criteria are given in the Technical Report to the Commission for Hydrology No. 23 (WMO, 1988). The complete set of HFS results corresponding to the simulated real-time tests is presented in Georgakakos et al. (1988).

CONCLUSIONS AND RECOMMENDATIONS

Improvements in real-time streamflow forecasting were realized when conceptual hydrologic models were coupled with state estimators (to form stochastic–dynamic hydrologic models) for state updating from observed streamflow. A physically based methodology for the estimation of the state estimator parameters was implemented for the case of the modified Sacramento model complemented by a nonlinear reservoir routing model. The estimation methodology utilizes hydrologists' degree-of-belief estimates of expected errors in model input variables and in model parameter estimates. It involves two free parameters and it is suitable for use with conceptual models containing large state vectors. It was exemplified for the Bird Creek basin in Oklahoma, U.S.A., and for the Orgeval basin in France. Verification consisted of running the stochastic–dynamic hydrologic model for several historical events, unavailable during the calibration phase, simulating real-time forecast conditions. Mean areal rainfall estimates obtained from actual observed raingauge rainfall were used as input during the verification tests in a "perfect foresight" scenario. Both statistical least-squares criteria and hydrologic criteria were used in both calibration and verification phases to quantify model performance.

Several conclusions follow:

(1) Use of stochastic–dynamic models in an operational environment for real-time flood prediction is feasible and requires minimal knowledge of estimation theory for realization of the advantages offered by a state estimator. The present study simplified the problem of state estimator calibration, so that the estimation framework is based on hydrologists' degree-of-belief estimates.

(2) Under the conditions of the present study, state estimation generally improves the deterministic model predictions not only from a least-squares point of view (coefficient of efficiency at the 0.9 level) but also with respect to hydrologic criteria such as errors in the timing and magnitude of the hydrograph peak. In correcting for timing errors, however, underestimation of excessively high flows can result in cases of fast rising hydrographs with a few observed discharges on the rising limb of the hydrograph. Such underestimation is attributed to lack of mass conservation during the state updating step of the predict-update sequence for nonlinear models.

(3) The stochastic–dynamic model predictions were quite sensitive to the parameters of the numerical integration scheme used to solve the state mean-covariance differential equations. Underestimation of the flows can result in cases of excessive rainfall and saturated model compartments for relatively long time discretization intervals.

(4) Both statistical performance criteria and hydrologic performance criteria should be used in assessing performance during calibration and verification of stochastic–dynamic hydrologic models. Optimality regions corresponding to each type criteria may not overlap.

(5) Application of the conceptual hydrologic model to the Orgeval basin surfaced the possibility of a structural error in simulating interflow via a linear function for basins with very little surface runoff.

Recommendations for future research

Perhaps the most important next step along the lines of research reported herein is the implementation of the stochastic–dynamic model developed in a true operational environment and its use by field hydrologists. Experience gained by such an effort can be utilized for the improvement of the stochastic–dynamic model formulation so that it is more robust to low quality real time data, and it can be easily used by field hydrologists with a minimal training in modern estimation theory. HFS has been implemented on the National Weather Service River Forecast System forecast component and it is currently undergoing testing in a true real-time environment.

Regarding the state estimator, research is needed to design constrained estimators that conserve water volume during state updating. Study of the effects of state updating to the long forecast-lead-time errors of nonlinear hydrologic models should be undertaken. In addition, efficient but accurate numerical schemes should be designed and tested, sutiable for use with large nonlinear stochastic–dynamic models. Input and state dependent time discretization of the coupled state mean and covariance differential equations should be utilized.

The formulation of the interflow production component of the modified Sacramento model should be re-examined in light of the Orgeval results. A nonlinear interflow function may be more suitable for basins with little or no surface runoff. The recursive parameter estimation methodology in Rajaram and Georgakakos (1987, 1989) and the formulation in this report can be utilized to examine the structural errors for the Orgeval basin.

The WMO Vancouver Workshop was a learning experience for modelers in that it simulated real-time conditions not available during the times devoted to the research and development of models. It is imperative however that a follow-up Workshop be planned during which forecasts of rainfall are utilized instead of actual future rainfall observations. It is the rainfall input component of the input uncertainty that contributes the most to prediction uncertainty (Georgakakos, 1986b).

ACKNOWLEDGMENTS

The research work was sponsored by the Hydrologic Research Laboratory (HRL), National Weather Service, National Oceanic and Atmospheric Administration, under the HRL–University of Iowa Cooperative Agreement No. NA86AA-H-HY126. The support work of the staff of HRL in determining values for the modified Sacramento model parameters and associated estimation-error variances for the two basins of study (Eric A. Anderson and Larry E. Brazil) is gratefully acknowledged. The staff of HRL and the staff of Centre National du Machinisme Agricole, du Genie Rural, des Eaux et des Forets, Hydrologic-Hydraulic Division, prepared the calibration and verification datasets for Bird Creek and Orgeval, respectively. Also acknowledged are the efforts of the WMO staff (particularly Arthur J. Askew) in organizing and executing the verification exercise. The computational work during the calibration phase was performed on an Apollo Domain 3000 workstation at the College of Engineering, The University of Iowa. H. Rajaram and S.G. Li assisted with the computational work of the calibration phase. During the verification phase, a similar computer workstation was used at the University of British Columbia (UBC), Vancouver, B.C., Canada. The efforts of Michael C. Quick and Allan Spears of UBC were essential to the success of the verification exercise and are gratefully acknowledged.

REFERENCES

Askew, A.J., 1986. WMO Project on simulated real-time intercomparison of hydrological models — Description of the Orgeval hydrological basin. Technical Correspondence, World Meteorological Organization, Geneva.

Burnash, R.J.C., Ferral, R.L. and McGuire, R.A., 1973. A generalized streamflow simulation system — Conceptual modeling for digital computers. National Weather Service, NOAA, CA Dep. Water Resourc. Tech. Rep., Joint Federal–State River Forecast Center, Sacramento, CA.

Day, G.N. and Farnsworth, R.K., 1982. Mean areal potential evaporation. National Weather Service River Forecast System (NWSRFS) User's Manual. Hydrologic Research Laboratory, NWS, NOAA, Silver Spring, MD.

Gelb, A. (Editor), 1974. Applied Optimal Estimation. MIT Press, Cambridge, MA.

Georgakakos, K.P., 1984a. Time series analysis-state space modeling — forecasting-parameter search (TSFP) user's manual, version 2. Hydro-Tech. Note No. 3. Hydrol. Res. Lab., National Weather Service, NOAA, Silver Spring, MD.

Georgakakos, K.P., 1984b. Model-error adaptive parameter determination of a conceptual rainfall prediction model. Proc. 16th IEEE Southeastern Symposium on System Theory. pp. 111–115.

Georgakakos, K.P., 1986a. A generalized stochastic hydrometeorological model for flood and flash-flood forecasting, 1, Formulation. Water Resourc. Res., 22(13): 2083–2095.

Georgakakos, K.P., 1986b. A generalized stochastic hydrometeorological model for flood and flash-flood forecasting, 2, Case studies. Water Resourc. Res., 22(13): 2096–2106.

Georgakakos, K.P., 1987. Flash Flood Prediction. J. Geophys. Res., 92(D8): 9615–9629.

Georgakakos, K.P. and Bras, R.L., 1980. A statistical linearization approach to real time nonlinear flood routing. Tech. Rep. No. 256, Dep. Civil Eng., MIT, Cambridge, MA.

Georgakakos, K.P. and Bras, R.L., 1982. Real time, statistically linearized adaptive flood routing. Water Resour. Res., 18(3): 513–524.

Georgakakos, K.P. and Brazil, L., 1987. Real-time flood and flash-flood forecasting using hydrometeorological models and modern estimation theory techniques. In: L. Berga and J. Dolz (Editors), Flood Warning Systems. Imprime Cooperativa de Publicaciones del Colegio de Ingenieros de Caminos, Canales y Puertos, Almagro, Madrid, pp. 184–227.

Georgakakos, K.P. and Hudlow, M.D., 1984. Quantitative precipitation forecast techniques for use in hydrologic forecasting. Bull. Am. Meteorol. Soc., 65 (11): 1186–1200.

Georgakakos, K.P., Rajaram, H. and Li, S.G., 1988. On improved operational hydrologic forecasting of streamflows. IIHR Tech. Rep. No. 325, Dep. Civil Environ. Eng., Iowa Inst. Hydraul. Res. University of Iowa, Iowa City, IA, 162 pp.

Hydrologic Research Laboratory, 1972. National Weather Service River Forecast System Forecast procedures. NOAA Tech. Memo. NWS HYDRO-14. HRL, National Weather Service, NOAA, Silver Spring, MD.

Kitanidis, P.K. and Bras, R.L., 1978. Real time forecasting of river flows. Tech. Rep. No. 235. Dep. Civil Eng., MIT, Cambridge, MA, 324 pp.

Kitanidis, P.K. and Bras, R.L., 1980a. Real time forecasting with a conceptual hydrologic model, 1, Analysis of uncertainty. Water Resourc. Res., 16(6): 1025–1033.

Kitanidis, P.K. and Bras, R.L., 1980b. Real time forecasting with a conceptual hydrologic model, 2, Applications and results. Water Resourc. Res., 16(6): 1034–1044.

Larson, L.W., 1975. Precipitation model. National Weather Service River Forecast System (NWSRFS) User's Manual. HRL, National Weather Service, NOAA, Silver Spring, MD.

Larson, L.W. and VanDemark, S., 1979. Mean areal precipitation program (MAP), 1979. National Weather Service River Forecast System (NWSRFS) User's Manual. HRL, National Weather Service, NOAA, Silver Spring, MD.

Mein, R.G., Laurenson, E.M. and McMahon, T.A., 1974. Simple nonlinear model for flood estimation. J. Hydraul. Div. Am. Soc. Civil Eng., 100(HY11): 1507–1518.

Peck, E.L., 1976. Catchment modeling and initial parameter estimation for the National Weather Service River Forecast System. NOAA Tech. Memo. NWS HYDRO-31, HRL, National Weather Service, NOAA, Silver Spring, MD.

Rajaram, H. and Georgakakos, K.P., 1987. Recursive Parameter Estimation of Conceptual Watershed Response Models. IIHR Rep. No. 319, Dep. Civil Environ. Eng., Iowa Inst. Hydraul. Res., University of Iowa, Iowa City, IA, 169 pp.

Rajaram, H. and Georgakakos, K.P., 1989. Recursive Parameter Estimation of Hydrologic Models. Water Resourc. Res., 25(2): 281–294.

Ralston, A., 1962. Runge-Kutta methods with minimum error bounds. MTAC, 16(80): 431–437.

Ralston, A. and Wilf, H.S., 1960. Mathematical Methods for Digital Computers. John Wiley, New York, pp. 95–109.

Smith, G.F., 1986. WMO Project on Simulated real-time intercomparison of hydrological models — Calibration data sets. Technical Correspondence, HRL, National Weather Service, NOAA, Silver Spring, MD.

Unny, T.E. and Karmeshu, 1984. Stochastic nature of outputs from conceptual reservoir model cascades. J. Hydrol., 68: 161–180.

WMO, 1988. Real-time intercomparison of hydrological models. Tech. Rep. to the Commission for Hydrology No. 23, WMO/TD No. 255, Geneva.