Stochastic Hydrol. Hydraul. 1 (1987) 303-318

Stochastic Hydrology and Hydraulics

Rainfall-runoff forecasting methods, old and new

M. Karlsson and **S. Yakowitz**

Systems and Industrial Engineering Dept., College of Engineering and Mines,
The University of Arizona, Tucson, AZ 85721, USA

Abstract: Here we review the main thrusts of rainfall-runoff modelling with an eye toward the advantageous use of the massive date sets being accumulated and the modern computers capable of dealing effectively with such sets.

More than a tutorial, this study is aimed at providing a unifying structure for analyzing available techniques. The closing section draws attention to the existence of an alternative methodology.

Key words: Rainfall-runoff models, forecasting, unit hydrograph.

1 Introduction

Relating precipitation information to runoff is a fundamental activity of hydrologic research. The "information age" has transformed the underpinnings of this activity. Whereas in previous times, it was feasible only to perform a few calculator-sized computations using a small set of laboriously-gathered rainfall-runoff numbers, now investigators may work with widely available data bases consisting of millions of measurements. Even these bases have scarcely tapped the potential of electronic gathering, telemetering, and remote-sensing technology.

The intention of the present study is to survey the major directions in rainfall-runoff (R-R) forecasting methodology, with attempts at deriving a unified framework for categorizing various modelling classes. We offer commentary on implications of modern computational and statistical theory regarding major forecasting schemes. Our attention is devoted to "empirical" (in the sense of Clarke 1973) or, synonymously, "systems-theoretic" (e.g. Sorooshian 1983) techniques. These constitute the major portion of the forecasting literature. This work is not a comprehensive review: however, we do believe that we have hit upon the major themes, and have sketched their logical foundations.

The organization of this study is the following: The best-known R-R forecasting tool, the instantaneous unit hydrograph (IUH), is reviewed and its characteristics are examined with regard to suitability in the case of large data bases. A conclusion is that when runoff (as well as precipitation) data are available, one should use them in the forecasting mode. The line of thinking leads us (still holding to "second order" theory) from the unit hydrograph (which is a linear regression on excess precipitation) to the ARMAX model. We remind the reader that "ARMAX" denotes ARMA (on the runoff series) with an exogenous variable, which, in this context, is precipitation. Our framework for viewing linear prediction allows fresh insight into the relations of IUH, and ARMAX forecasting. Curiously, we are aware of only a few studies in which the resulting forecast actually depends on both rainfall and runoff.

The section on the linear forecasters (i.e., IUH, ARMA, and ARMAX) is followed by a short synopsis of nonlinear predictors, with special attention to ideas based on i) Volterra expansions and ii) random IUH's.

In the closing section we draw the reader's attention to a new procedure based on a pattern recognition device, which has some advantages over all techniques we have reviewed. The essential feature is that under fairly general circumstances, convergence to the optimal forecaster is assured as the data set increases without bound. In fact, an aim of the present exposition is to whet the reader's appetite for our methodology, which is expounded elsewhere (Karlsson and Yakowitz 1987).

The present study is part of a three-pronged attack on R-R forecasting. Its purpose is to give perspective to the subject and motivation for steps made by the other "prongs". Karlsson and Yakowitz (1987) concentrates on experimentally exploring the nearest-neighbor approach on actual data (from the Cochocton Watershed), comparing our methodology with ARMAX and unit hydrograph techniques. Yakowitz (1987), the third prong, is a theoretical development which demonstrates the consistency and convergence of the method in a rigorous mathematical framework.

2 Unit hydrograph application to forecasting

The instantaneous unit hydrograph (IUH) is a fundamental pillar of surface hydrology theory and application. For purposes of the present study, there are two salient aspects of the IUH R-R modelling:

1. It is used for runoff forecasting.
2. There is reason to suspect that when large data bases are available, it might be needlessly inaccurate.

Toward substantiating the first point, Sorooshian (1983) has written: "It is perhaps disconcerting that the most widely applied technique for short interval (24 hours or less) on-line flow forecasting is still the unit hydrograph".

Linsley et al. (1982, Sec. 16.11) describes and offers a detailed IUH forecasting procedure and an example (the only R-R forecaster in their book). Freeze (1982) offers a chart categorizing usages of various R-R modelling devices. The IUH is classified as being a forecasting tool (as well as a design aid). The book by Bras and Rodriguez-Iturbe (1985, p. 164) states, "Possibly one of the most popular hydrologic concepts is the [IUH]". Chander and Shanker (1984), for example, give a detailed analysis of on-line prediction of base flow and excess precipitation index, so that the classical IUH model can be used in the forecasting mode.

Let us develop some notation and terminology. Assume that rainfall and runoff are stationary sampled data, which we denote by sequences $\{p(i)\}$ and $\{q(i)\}$, respectively. Presume that the $p(i)$'s and $q(i)$'s have already been "preprocessed" to remove baseflow. Then to say that R-R is a linear time-invariant, causal relation is equivalent to saying that there exist constants $\{a(k)\}_{k\geq 0}$ so that for any time i,

$$q(i) = \sum_{j=0}^{M} a(j)p(i-j). \tag{1}$$

In the general case, M can be infinity. A nonlinear (M-step) relationship implies that flows and precipitation are related, for all i, by an expression of the form

$$q(i) = f(p(i), p(i-1), p(i-2), \ldots, p(i-M)), \tag{2}$$

where $f(\ )$ is an arbitrary function of the indicated variables.

Linearity is a very restrictive assumption. On the other hand, it is to be granted that linear approximations have served engineers and statisticians well, especially in circumstances (thermodynamics or electromagnetic waves) in which it has been established that the measurables do satisfy linear differential equations, or in statistical analysis when large data sets are unavailable. In precomputer times, when IUH techniques earned a central place in hydrology, one could only deal with data sets within the computational limitations of human endurance and in the absence of the anything like the foundational theory of mathematical statistics of present day. Eagleson et al. (1966) articulate this viewpoint. However, a theme of the present study and its companion papers (Karlsson and Yakowitz 1987 and Yakowitz 1987) is that developments in computers and statistics have made it seem advisable to reexamine the utility of the IUH concept.

We wish to make certain observations:

2.1 Among linear predictors using precipitation only, the IUH is inclusive:

To say that a forecaster $q(i)$ depends linearly on the data $\{p(j)\}_{j \leq i}$ is mathematically equivalent to demanding that the forecaster be of the Eq. (1). If the coefficient sequence $\{a(i)\}$ is absolutely summable (i.e., $\sum |a(i)| < \infty$), one can approximate a predictor with infinite M to arbitrary accuracy by a predictor with M finite, but sufficiently large. The statistical problem of choosing the IUH parameters $\{a(j)\}$ is not trivial.

2.2 Regression would seem advisable if M, in Eq. (1) is "small":

A standard hydrological practice for choosing parameters is to select the IUH to "fit" the R-R data from one or perhaps a few rainstorms. Thus the coefficients are determined by solving the system

$$q(i) = \sum_{k=0}^{M} a(k) p(i-k), \quad M \leq i \leq N-M \tag{3}$$

for the $a(i)$'s, the $p(i)$'s and $q(i)$'s being observed areal precipitation and runoff of a single event yielding N data pairs. If $M = N+1$, then barring some unlucky event of probability 0, the $a(i)$'s are uniquely determined. If $N > M+1$, then one customarily seeks a solution minimizing the squared residual error $J(\underline{a})$, given by

$$J(a(0), \ldots a(M)) = \sum_{i=M}^{N-M} (q(i) - \sum_{k=0}^{M} a(k) p(i-k))^2. \tag{4}$$

Linsley et al. (1982, Sec. 7.8) recommend repeating the process over several rainfall events and then in a tricky way averaging the coefficient vectors determined by each event. Eagleson et al. (1966) and Hino (1970) describe ways for using Wiener filter theory to infer the optimum unit hydrograph. However, it follows from Priestly (1981, Chap. 10), for example, that least-squares methods must also yield the same optimum linear filter.

In this computer age, all data analysis options are open: Statistician Efron encourages us to "think the unthinkable". We believe that a feasible initial plan of action is to use the entire data base for inference of the IUH parameter set $\underline{a} = (a(0), \ldots, a(M))$. Some hydrologists have argued that an automated procedure is foolhardy: One should restrict attention to particularly informative or representative events. There is pragmatic merit in this view. But for purposes of scholarship, such investigators should state clearly and algorithmically what it is that constitutes good data for R-R calibration. Some hydrologists argue further that hydrologic intuition and experience cannot be encoded. We agree that the working hydrologist should incorporate all the information and intuition he can into analysis of specific basins when undertaking actual applications. But to the authors of this manuscript, science rests upon two pillars: (1) the reproducible experiment and (2) mathematical deduction. Remove either pillar and stochastic hydrology is not a science. For this reason, we urge theoreticians in their research expositions, to be fully algorithmic about their procedures, explicitly incorporating into their modelling efforts and resulting formulas those guiding principles which they think lead to good hydrographs.

Presume for now that the number of R-R data points ranges into the thousands, as it does in the data sets we deal with in Karlsson and Yakowitz (1987). If the number of regression coefficients M in Eq. (1) is relatively modest (say $M = 10$ or 12) and if the IUH model is approximately correct, or even stationary, the estimates $\underline{\hat{a}}$ of $\underline{a}$ obtained by regression should be accurate. One need not make any further modelling assumptions, by way of specialized "kernel" classes.

On the other hand, for larger IUH parameter sets, the perils of least-squares methods are legion. Perhaps weighted regression, ridge regression, and so forth, have roles here, and in any case, numerical techniques such as related in Lawson and Hanson (1974) should be employed to overcome the usual ill-conditioning difficulties of high-dimensional least-squares problems. (Such difficulties have been illustrated in Yakowitz and Szidarovszky 1986, Sec. 6.2.4).

3 General framework for linear predictors

The most general framework for linear forecasting is afforded by the Gauss-Markov (GM) discrete-time model (e.g. the model in Anderson and Moore 1979, pp. 44-45). The model is also known variously as the state-space model (Ljung and Soderstrom, 1983, pp. 15), and the Kalman filter model. In the stationary case, this model assumes the form

$$x(k+1) = A\,x(k) + B\,p(k) + D\,w(k)\,, \qquad (5)$$

$$q(k) = C\,x(k) + v(k)\,,$$

where $\{p(k)\}$ is a given input sequence, $\{w(k)\}$ and $\{v(k)\}$ are uncorrelated noise processes, and A, B, C, and D are matrices of appropriate orders. In the IUH context, we will take $q(k)$ to be the runoff. We will view $p(k)$ as precipitation, which is presumed to be observable. The $v(k)$ terms can be regarded as measurement noise, and are unobservable. The statistical assumptions in IUH modelling are seldom made explicit, but it is consistent with this literature to take the $w(k)$'s to be zero, and the $v(k)$'s to be white noise.

It is instructive to see how standard IUH models can be viewed as various special cases of the GM model. Having investigated that issue, it is appropriate to ask if there is anything useful left over.

The general IUH model (Eq. (1) with M finite) can be modelled by Eq. (5), with the $\{w(k)\}$ sequence set to zero. The intention is to make the state $x(k)$ store the past M precipitation observations. Then it is clear that C should be taken to be the horizontal vector of coefficients $\{a(j)\}$ in Eq. (1). Specifically

$$A = \begin{bmatrix} 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & & & & & \cdot \\ 0 & \cdot & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 & 0 \end{bmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad C = (a(0),\dots,a(M)). \tag{6}$$

We mention that the discrete-time version of the Dooge (1959, pp. 248) triangular IUH takes the form Eq. (1), with M finite since the kernel is zero except on a bounded time interval. Thus the preceding discussion covers this case. In fact, for the Dooge kernel composed of two lines, the coefficients can be stated explicitly as

$$a(j) = \begin{cases} C(j/T_1), & 0 \leq j \leq T_1, \\ C(1-j/T_1), & T_1 \leq j \leq T_2. \end{cases} \tag{7}$$

If M in Eq. (1) is infinite, only certain $a(i)$ sequences admit GM representation. But the IUH model given by Nash (1959, pp. 115) falls into this representable class. In fact, the modelling discussion in that paper essentially tells us how to construct the representation. In Nash's notation, the number of reservoirs is n, and they are identical, and are determined by the relation

$$\begin{aligned} &q_v(j+1) = [(1-1/k)q_v(j) + q_{v-1}(j)]\Delta t, \quad \text{if } v = 1,\dots,n, \\ &q_v(j) = p(j), \quad \text{if } v = 0, \end{aligned} \tag{8}$$

which we obtain by substituting the finite difference approximation $(q_v(j+1) - q_v(j))/\Delta t$ for $d/dt\ q_v(t)$ in the continuous reservoir model

$$\frac{dq_v}{dt} = -1/k\, q_v + q_{v-1}.$$

In Eq. (8), $q_v(j)$ denotes the outflow of the vth reservoir at time instant $j\Delta t$.

A way to put this model into the mold of Eq. (5) is to take as state vector

$$x(j) = (q_1(j), q_2(j), \dots, q_n(j)).$$

Then we get a system consistent with Eq. (8) by defining

$$A = \begin{bmatrix} 1-\Delta t/k, & 0, & \cdot\ \cdot\ \cdot, & 0 \\ \Delta t & & & \cdot \\ 0 & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & \cdot\ \cdot\ \cdot & \Delta t, & 1-\Delta t/k \end{bmatrix}, \quad B = \begin{pmatrix} \Delta t \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad C = (0, 0, \dots, 0, 1). \tag{9}$$

From linear systems theory, one can conclude that the sequence $\{a(j)\}$ is GM-representable only if it can be written in the form of a finite sum of geometric terms

$$a(j) = \sum_{q=o}^{Q} \sum_{m=0}^{N(q)} C_{q,m} (\lambda_q)^j j^m, \tag{10}$$

where the λ_q's are real or complex numbers with modulus less than one, and the $N(q)$'s are non-negative integers. In essence, the constraint that the IUH be GM-representable is equivalent to requiring that the outflow can be viewed as precipitation routed through a finite number of Nash-like reservoirs. But we observe that in contrast to Nash (1959), there is no mathematical reason to take the reservoirs as identical, and they can be interconnected any which way, and furthermore, each could have its own tributary, in addition to inflows from other reservoirs. Dooge (1959) has some generalizations in these directions. Klemes and Boruvka (1975) have derived this hydrograph representation in detail.

4 Connections with ARMA streamflow modelling

The assertion that runoff is ARMA implies that $\{q(i)\}$ satisfies the linear inhomogeneous difference equation,

$$\sum_{j=0}^{N1} f(j)q(i-j) = \sum_{v=0}^{N2} e(v)w(i-v). \tag{11}$$

One may place the ARMA model Eq. (11) into the GM model class by setting the $p(k)$ sequence in Eq. (5) to zero and using the state to store past values of w and q. This can be accomplished by first assuming (without loss of generality) that $f(0) = 1$, and then rewriting Eq. (11) as

$$q(i) = -\sum_{j=1}^{N1} f(j)q(i-j) + \sum_{v=0}^{N2} e(v)w(i-v). \tag{12}$$

Now take the state vector to be

$x(n)^T = (q(n), q(n-1), \ldots, q(n-N1), w(n), \ldots, w(n-N2))$.

Then Eq. (5) satisfies Eq. (11) if A and D are given respectively by

$$\mathbf{A} = \begin{array}{c} \left[\begin{array}{ccccccccc} -f(1), & -f(2), & \ldots & -f(N1), & e(0), & e(1), & \ldots & \ldots & e(N2) \\ 1, & 0, & \ldots & 0, & 0, & 0, & .. & 0,0, & 0 \\ 0, & 1, & \ldots & 0, & 0, & 0, & .. & 0,0, & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0, & 0, & 0,1,0, & 0, & 0, & 0, & .. & 0,0, & 0 \\ 0, & 0, & 0,0,0, & 0, & 0, & 0, & .. & 0,0, & 0 \\ 0, & 0, & 0,0,0, & 0, & 1, & 0, & .. & 0,0, & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0, & 0, & 0,0,0, & 0, & 0, & 0, & .. & 0,1, & 0 \end{array}\right] \end{array} \begin{array}{l} \\ \\ \\ \\ \\ \leftarrow \text{row } N1+1 \\ \\ \\ \\ \end{array} \tag{13}$$

$D = (-e(0), 0, \ldots, 0, -1, 0, \ldots 0)^T$ (The "-1" is in row $N1 + 2$). The observable output is flow $q(n)$, and one obtains this by setting $C = (1, 0, \ldots, 0)$.

The z-transform of any sequence of numbers or vectors such as $\{b_j\}$ is denoted by $B(z)$ and defined by

$$B(z) = \sum b_j z^{-j} \tag{14}$$

with j in Eq. (14) ranging over the domain of the sequence index.

The z-transform transfer function effected by the ARMA model Eq. (11) is

$$Q(z) = \frac{E(z^{-1})}{F(z^{-1})} W(z) \tag{15}$$

where $E(.)$ and $F(.)$ are polynomials of the variable z^{-1}, $Q(z)$ and $W(z)$ being the z-transforms of the sequences $\{q(i)\}$ and $\{w(i)\}$, respectively.

Suppose now that the ARMA runoff is the result of some sort of linear response to precipitation. The model of Eq. (11) forces us to conclude that the $w(i)$'s, which are assumed to be white noise in GM ARMA modelling, must somehow be related to the precipitation values $\{p(i)\}$. This is to say, the z-transformation $P(z)$ of the precipitation sequence must satisfy

$$P(z) = G(z^{-1})W(z)$$

for some polynomial $G(\cdot)$. Also, the presumption (12) that runoff depends on a moving average of a forcing function implies that the forcing function must be a moving average of precipitation, if it is to be believed that runoff depends linearly on precipitation, and nothing else. Writing this observation in the form of an equation, we conclude that for some polynomial $H(\cdot)$,

$$F(z^{-1})Q(z) = H(z^{-1})G(z^{-1})W(z) = E(z^{-1})W(z).$$

Thus $G(z^{-1})$ must factor the polynomial $E(z^{-1})$ in Eq. (15). Perhaps from this information and with the help of conceptual modelling, one can discover from the roots of $E(\cdot)$ the factor that is attributable to the precipitation moving average (i.e., the polynomial $G(\cdot)$), and the factor that is attributable the hydrogeologic process converting rainfall into runoff (the $H(\cdot)$ part of the filter). The interesting possibilities stem from the fact that $E(\cdot)$ can be found, up to a scale factor, without using precipitation data at all, but solely from the runoff data, because this is enough to calibrate ARMA models.

Having made the connection between IUH kernels and ARMA runoff models, we observe that while there are mountains of publications on ARMA models for runoff prediction with no measurements on precipitation, and mountains of publications on runoff prediction based on the IUH, which depends solely on rainfall, it is curious that there seems to be few who have noticed the potential of using the GM model of Eq. (5) to incorporate both rainfall and runoff data into the runoff prediction. In this context, the ARMA formula is called the ARMAX model (e.g., Ljung and Soderstrom 1983), the "X" standing for the exogenous variable (the $p(i)$'s in Eq. (5)). It is not clear to us why a ground rule of forecasting seems to be that past runoff observations do not appear in the prediction formula. Usually, one has rainfall-runoff pairs available for calibration of the IUH, according to the standard formula (4). Presumably the same measuring devices are still available at forecasting time.

The essential distinction between the ARMA and ARMAX models is that in the latter, part of the forcing term is observable. That is, one can suppose $q(i)$ in Eq. (12) can, in an ARMAX regime, be written as

$$q(i) = -\sum_{j=1}^{N1} f(j)q(i-j) - \sum_{j=1}^{N2} e(j)w(i-j) - \sum_{j=1}^{N3} b(j)p(i-j). \tag{16}$$

Here the $w(k)$'s are still presumed to be white noise, but $\{p(j)\}$ can be any sequence.

In comparing the ARMA and ARMAX viewpoints, we may say that under the latter regime we can partially observe the forcing function and allow the decision to use this extra information. This agrees with our earlier statement that in ARMA modelling, part of the forcing term is a moving average of white noise, and part of it is a linear combination of precipitations. The latter component is taken as observable in ARMAX but not in ARMA. This information should certainly reduce the mean squared prediction error.

The ARMAX model is accommodated into the GM structure of Eq. (5) by breaking out the (observable) precipitation from the ARMA representation of Eq. (12). Thus the state vector now stores precipitation as well as white noise:

$$x(n)^T = (q(n),...,q(n-N1),w(n),...,w(n-N2),p(n),...,p(n-N3)).$$

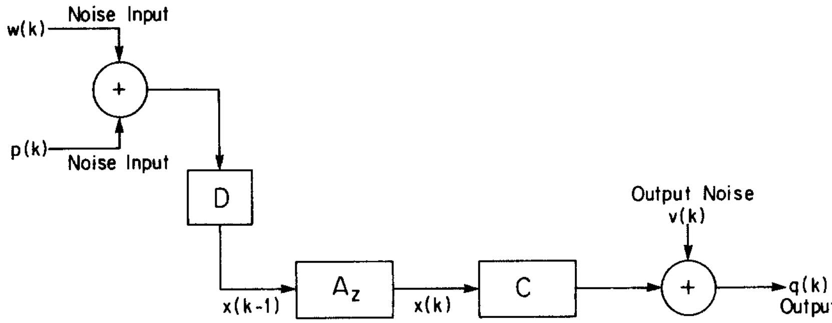

Figure 1. ARMA state space model

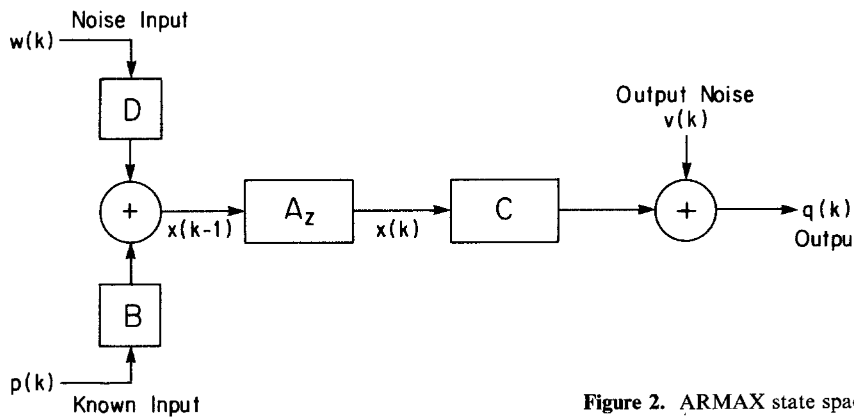

Figure 2. ARMAX state space model

Let us compare the ARMA and ARMAX R-R models with an eye toward evaluating their mean-square prediction errors. In Figs. 1 and 2, we have given the schematic block diagrams of the two processes. An important insight comes from noting that because of superposition, the effect of the measured precipitation inputs $\{p(i)\}$ can be eliminated entirely from the prediction error in the ARMAX model. To see this, evaluate the trajectory $\{\tilde{x}(j)\}$ of the linear system determined by A and D and the forcing sequence $\{p(i)\}$, presuming the initial state to be zero. Then subtract this trajectory $\{\tilde{x}(j)\}$ from the full ARMAX trajectory, and what is left (call it $\bar{x}(k) = x(k) - \tilde{x}(k)$, $k = 1,2,...$) is exactly the ARMA process generated by the input sequence of $w(k)$'s alone. Thus the effective noise of the ARMAX process is the noise of the ARMA process in Fig. 1, but with the $p(k)$ sequence removed. Intuitively, it should be clear that the optimal ARMA predictor ought to be more accurate with the input noise component reduced (by removal of the $p(k)$ sequence). But we can demonstrate this point by use of equations from Anderson and Moore (1979, Sec. 3.1). In keeping with their developments, let Q_w and Q_p be the variances of the sequences $w(k)$ and $p(k)$ respectively. Then at each iteration k, the squared error $\Sigma_{k+1\,|\,k}$ of one-step-ahead prediction under an optimal (Kalman) filter is recursively calculated

$$\Sigma_{k+1\,|\,k} = \Pi(\Sigma_{k\,|\,k-1}) + BQB^t \,, \tag{17}$$

where the exact form (given precisely by Eq. (1.9) of Anderson and Moore) of the function $\Pi(\Sigma)$ is of no particular significance, except that it increases monotonically with Σ. Now we may see inductively that for $\tilde{\Sigma}_{k+1\,|\,k}$ the error variance of ARMA and $\overline{\Sigma}_{k+1\,|\,k}$ the effective error variance of ARMAX,

$$\tilde{\Sigma}_{k+1|k} > \overline{\Sigma}_{k+1|k} \,. \tag{18}$$

For (by Eq. (1.10) of Anderson and Moore),

$$\tilde{\Sigma}_{1|0} = Q_p + Q_w > Q_w = \overline{\Sigma}_{1|0} \,.$$

Now use monotonically $\Pi(\cdot)$ in Eq. (17), and again the fact that $Q_p + Q_w > Q$ to conclude Eq. (18), by finite induction.

The few works known to us which use the ARMAX estimator and actually make the estimate depend on both rainfall and runoff, are Todini (1978), Party and Moreno (1982), Copper and Wood (1982a,b), and Kitanidis and Bras (1980). The earliest references we know of are Hino (1970), a conference proceeding report which Bras and Rodriguez-Iturbe (1985, Sec. 8.4) review, and a very nice and complete exposition by Kashyap and Rao (1973), based on an IFIPS conference in 1972.

In our view, ARMAX modelling is the most elegant method and final word in empirical linear forecasting for the runoff response to rainfall. The ARMAX model could, in principle, also be made to depend on other hydrologic variables such as snowpack, soil moisture, temperature, etc., however an increase in dimensionality makes the identification problem more complicated. Perhaps a conceptual approach to modelling, which incorporates the hydrogeology of the watershed, will eventually lead to more parsimonious and effective linear modelling, but the authors are not optimistic, since nonlinearities show up in the most elementary physical considerations of the process.

The identification problem has been adequately studied in the systems literature (e.g. Ljung and Soderstrom (1983)). The IMSL library has packages for ARMAX (as well as ARMA) inference. A conclusion is that estimates of ARMA and ARMAX are consistent and the error variance of standard estimators goes to 0 as $1/n$, n being here the length of the data set.

Admirable surveys of linear R-R estimation methods include Clarke (1973), O'Connell and Clarke (1981), Freeze (1982), Bras and Rodriguez-Iturbe (1986), and Sorooshian (1983).

There are two obvious and fundamental weaknesses to IUH and linear systems forecasting:

1. There is no reason to consider that the R-R relationship is linear. Indeed standard models for channel flow and infiltration would suggest the contrary.
2. Statistical methods for IUH, Kalman, ARMA, and ARMAX forecasters are second order methods. Second order methods are suited only to least-squares estimation (unless the Guassian assumption is made). In particular, second order methods are of no theoretical use in estimating the probability

   $$p\,[q_{n+1} \geq T \mid q_n, q_{n-1}, \ldots; p_n, p_{n-1}, \ldots]\,, \tag{19}$$

   of flood, conditioned on observations in the recent past. However, it is clear that stochastic hydrologists would like tools for flood warning problems.

Whereas Linsley et al. (1982) tell us, "the unit hydrograph has been the mainstay of the flood hydrologist...," the case against second-order methodology (e.g. IUH) has been aptly stated by Freeze (1982) as follows:

> In order to arrive at a theoretically defensible design flood, it is necessary that one consider the measured frequency distribution of the peaks as an estimate of the probability density function for the population of peaks from which the measured peaks have been drawn. Unlike time series analysis, frequency analysis requires knowledge of the full distribution, not just the first two moments.

5 Parametric nonlinear forecasting

There exists a multitude of the oft-cited works which describe nonlinear stretchings of the IUH mold to make it work better. Some years after Sherman (1932) had proposed the unit hydrograph concept, Paynter (1952) argued from purely physical grounds that linear models in this context were untenable. Later, Minsall (1960) offered experimental measurements which he regarded as contradicting the unit hydrograph paradigm. The literature questioning the linearity of the R-R relationship is enormous. For example, Linsley et al. (1982, pp. 382) have written:

> The question of adjustment of the unit hydrograph is basically an assumption that the unit hydrograph is nonlinear .... empirical data on which to judge the unit hydrograph linearity .... concept are not conclusive.

O'Connell (1982) states flatly:

> It is well-known that the response of a catchment to rain is highly nonlinear.

One can accept the fundamental elaborations of IUH constructs such as baseflow and excess runoff at face value as modifications dictated by physical considerations, or more skeptically view them as not totally implausible fudge factors to twiddle for better fit than linearity will allow. Having established that the invention of a physically-motivated but unproven nonlinear fudge-factor elaboration of the basic IUH is a respectable and publishable activity, there is no apparent end to this line of inquiry. We mention a few of the more interesting nonlinear modelling efforts to come to our attention.

Early in the age of IUH, Amorocho (1967) proposed using a nonlinear modelling device (Volterra series) as a generalization of the IUH. We dwell for a moment on this approach because it has been influential, and because it has the potential of being more inclusive than many of the other nonlinear models. The theory, as sketched in Chaps. 2 and 3 of Wiener (1958), shows that a wide class of nonlinear functionals mapping the space of squared integrable functions onto itself (the mapping taken by Amorocho to model the transformation of rainfalls into runoffs) can be represented as an infinite series of compositions of convolutions. The richness of this series gives confidence that any conceivable R-R transfer function effected by "nature" must be in this class. The motivation of the series is that in principle one can obtain an adequate approximation, in the spirit of trigonometric polynomials of Fourier theory, by truncation. Standard least-squares methods can then be used to obtain the coefficients of the truncated expansion. Of all the nonlinear R-R modelling efforts, this Volterra series approach has attracted the greatest attention by far. We mention Diskin and Boneh (1973), Patry and Marino (1983a,b), Jacoby (1966), Napiorkowski and Strupczewski (1981). Amorocho and Brandstetter (1971) describe the Volterra scheme in an algorithmic (computer-ready) form, and apply it to some actual R-R data. They then compare their technique with a standard IUH forecaster; the nonlinear forecaster seems much more accurate. Of course, in any algorithmic version, one must truncate the Volterra series (most authors restrict attention to two terms). This cuts off the nonparametric potential of the Volterra approach. A plan (untried to our knowledge) which we think may have some promise is to let statistical methodology be a guide as to how many terms to take. In other contexts, Shibata (1984) has explored the idea of using Akaike-like criteria to test when the data warrant additional terms in a series expansion.

Other nonlinear modelling efforts to catch our fancy include a random IUH model (Rodrigues-Iturbe et al. 1982). Each storm generates a different hydrograph by virtue of randomly chosen parameters. For any particular height and time to peak, they use a triangular hydrograph, but presume that the times and heights are governed by some known random variable. Thus in a prediction mode, presumably one should let the IUH parameters "a" in Eq. (1) depend on the conditional distribution of times and heights, given precipitation measurements. The precipitation dependence of a would make the resulting predictor nonlinear. Patry and Marino (1982), Freeze (1982), and others have explored R-R relations of the (differential equation) form

$$q = -\frac{k_1}{k_2} N Q^{N-1} q - \frac{1}{k} Q + \frac{1}{k_2} P$$

which have a conceptual foundation in the Chezy relation of storage to outflow.

Other studies in replacing the IUH convolution kernel with nonlinear kernels include Patry and Marino (1983b), Rao and Rao (1984), and Shengjia (1984). Yet other approaches include a conference presentation by Todini and Wallis, which is outlined in O'Connell and Clarke (1981), who make the convolution kernel depend on the level of precipitation.

It seems plain from the above citations and the many related efforts of documented here, that whatever be the virtues of the IUH and its nonlinear variations, a sizable portion of influential stochastic hydrologists are dissatisfied with them. We urge the reader to note that in the next section, when we turn to (nonparametric) nearest-neighbor methods for the R-R process, we do depart from convention in that for a wide range of processes, our methods are consistent, in the strict statistical theory sense of the word: As the sample size increases, our model is guaranteed to converge to the correct R-R relation. No similar statement can be said for the type of nonlinear models cited above. If, for example, one uses a third order Volterra expansion model, and in fact, the true relation requires an infinite, or even a fourth order, relation, then beyond a certain limit, the model cannot get nearer the correct relationship no matter how large the calibration data base. Mathematicians would state that no parametric model is dense in the class of all possible R-R models.

Finally, we repeat an observation from Clarke (1973) and elsewhere that few modelers take the obvious step of using statistical theory to attempt to "validate" their selection. At present, statistical theory, computational means, and data bases of sufficient size are available; there is no excuse for not testing model hypotheses.

6 The nearest-neighbor forecaster

The goal of this brief section is to alert the reader to the existence of applicable methodologies fundamentally different from the general linear model. Here we describe in particular one such technique, the nearest neighbor (NN) method. In this section we give only the bare essentials of its construct and highlight its attractive properties. Our companion paper Karlsson and Yakowitz (1987) is dedicated to an investigation of the details and implementation of this technique. It gives an application to forecasting using data from the Cochocton watershed, and Yakowitz (1987), Yakowitz and Karlsson (1987) give some results on Bird Creek data. In the first reference, comparisons were made with the ARMAX and IUH models, and in the last two with the "Sacramento" model.

We do not intend that the present section should constitute an endorsement of nonparametric methods, but only that it should make the reader aware that they exist and in principle, have some promise.

With respect to very large R-R data sets, the R-R models hitherto considered have two fundamental drawbacks, in our opinion:

1: Being parametric, any of these model classes may not contain the true R-R relationship; as a result (and, mathematically speaking, since no parameter class is topologically dense in the space of all R-R models), the forecaster derived from such models may not be asymptotically optimal: It may never converge to the optimal forecaster, no matter how large the data base becomes.

2: Essentially none of them are probabilistic models. That is, the model specifications do not explicitly include provisions for modelling the random error of the forecast, nor do they possess the structure to serve as a foundation for a statistical theory for parameter calibration. We will admit that if the IUH approach is placed into a regression setting, as outlined at the end of Section 5, then one can achieve an asymptotically optimal linear estimator under the minimum squared-error criterion. But even under these restrictions, one can usually anticipate that there will be an even better nonlinear forecaster.

The authors have turned to nonparametric estimation theory in an attempt to overcome these drawbacks. Whereas results analogous to those to be outlined here can be achieved by the nonparametric kernel method (Yakowitz 1985 a,b), recently our attention has focussed on the nearest neighbor (NN) technique because it seems more intuitive, but nevertheless possesses statistical properties which are just as powerful as those of the kernel method.

To begin with, let us describe the NN method in the context of runoff forecasting. For each time epoch n, let $x(n)$ be some feature vector of past rainfall and runoff recordings. By feature vector, in the spirit of pattern recognition usage, we mean a vector that summarizes history as far as prediction is concerned. Mathematically, we hope that the conditional random variable $q(n+1) \mid (q(n), q(n-1), ...; p(n), p(n-1), ...)$ conditioned on the entire past, has the same probability distribution as the random variable $q(n+1) \mid x(n)$, which is conditioned on the nth feature vector, $x(n)$.

If x does not satisfy the above "history summarization" property, and the NN technique is nevertheless applied, the resulting forecaster will be asymptotically optimal among all forecasters defined on the feature vector $x(n)$. For example, if $x(n)$ is a vector of the past M precipitations, then the nearest neighbor will asymptotically (with increasing R-R record) be at least as good as any Mth order unit hydrograph.

We adopt some notation to help explain the NN algorithm. Let $m(x)$ denote the expectation of the next flow, conditioned on the current feature vector being x. That is,

$$m(x) = E[q(n+1) \mid x(n) = x]. \tag{20}$$

The nonlinear regression function $m(x)$ is the optimal one-step-ahead forecaster. The NN method is a statistical technique for approximating $m(x)$.

Definition: Let $\{k(n): n > 2\}$ be a nondecreasing sequence of positive integers such that $k(n) < n$, and let $\{(q(i+1), x(i)): 1 \le i \le n\}$ be the entire historical sequence of runoff/feature-vector pairs. With respect to these objects and for any feature vector x, the NN estimator of $m_n(x)$ is

$$m_n(x) = (1/k(n)) \sum_{j \in S(x,n)} q(j+1), \tag{21}$$

where $S(x,n)$ is the set of indices i of the $k(n)$ nearest neighbors of x amongst the observed vectors $x(i)$, $1 \le i \le n$. That is,

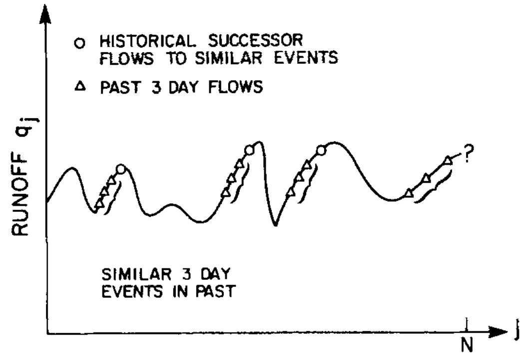

Figure 3. Illustration of nearest neighbor rule, estimate of $q_{N+1} \approx$ sample average of successors of similar 3 tuples to (q_{N-2}, q_{N-1}, q_N)

$$\text{if } j \in S(x,n) \quad \text{and } k \notin S(x,n)\,, \tag{22}$$

then with respect to Euclidean distance $||\cdot||$,

$$||x(j) - x|| \leq ||x(k) - x||\,.$$

In words, the NN-predictor of the next flow is the sample average of the successors to the $k(n)$ closest observed feature vectors to the current feature vector x. In Fig. 3, we have tried to illustrate the NN-method in the case that the feature vector depends only on the last three runoff observations. In this figure, it is presumed that $k(n) = 3$.

Rainfall-runoff analysis motivated the second author to undertake theoretical studies dedicated to justifying NN techniques that seem appropriate. We refer the interested reader to the companion statistical publication (Yakowitz 1987) for a complete and precise derivation of the asymptotic properties of the NN algorithm. There also the reader will find a summary of and references to foundational NN developments in the pattern recognition and statistics literature.

We mention here only facts and results most cogent to R-R analysis.

1 The NN method is nonparametric.

The assumptions of the convergence theory do not require that the jointly-distributed rainfall and runoff sequences $\{(p(i), q(i)\}$ belong to any parametric family such as normal, log Pearson, etc., or even ARMA (ARMA, of given AR and MA orders, are parametric inasmuch as they are continuously parameterized by the ARMA coefficient vectors). For the NN estimator to be consistent and converge at an optimal rate, it is enough that the sum

$$\sum_{i>0} (\,|E(q(0),q(i))| + |E(q(0),p(i))|\,) \tag{23}$$

of absolute cross correlations be finite; this is assured for the ARMA process, where cross correlations must vanish exponentially.

2 The NN algorithm is not difficult to program.

Our complete FORTRAN code contains about 500 statements. Typical Box-Jenkins inference packages are several times this size. Data sets involving several thousand points do require some CPU time; NN estimation is not something one does on a IBM-PC (yet), but a VAX is sufficient. The main expense in time is simply ordering the data to find the nearest neighbors. The NN estimation is effective for sample sizes in the order of 1000. We illustrate this point by way of example in the companion paper (Karlsson and Yakowitz 1987).

3 The asymptotic rate of squared error convergence is optimal.

The rate of convergence of a sequence of estimators is r, if for some constant C and all $n > 1$,

$$E[(m_n(x) - m(x)^2] < Cn^{-r} \tag{24}$$

and if no relation of the form Eq. (24) is satisfied for a value $r_1 > r$. An implication of a fundamental paper by Stone (1980) is that if the real dimension of feature vector $X(n)$ is d, and no parametric assumptions are made, the best possible rate r of convergence in squared error

$$E[m_n(x) - m(x))^2] < Cn^{-r} \tag{25}$$

is $r = 4/(4+d)$. Yakowitz (1987) has established that the NN achieves this optimal rate. For small d, this is almost as fast as one can hope for in the parametric (e.g. ARMA) case; this is because even in estimating the parameter of an iid sequence of Bernoulli observations, r cannot be made greater than 1.

4 The NN approach (unlike second-order methods) is effective for general decision problems.

Let $L(q,a)$ be the loss associated with action a and runoff q. Then define the regression function to be the expected risk under this action, conditional on the feature vector x:

$$m(x,a) = E[L(q(n+1),a) \mid x(n) = x]. \tag{26}$$

The obvious way to extend the NN algorithm of Eq. (21) to this setting is to define

$$m_n(x,a) = (1/k(n)) \sum_{j \in S(x,n)} L(q(j+1),a) \tag{27}$$

where $k(n)$ and $S(x,n)$ are as explained in connection with Eq. (21). Fundamental to flood warning problems (as explained at length in Yakowitz (1985b)) is the conditional probability of the next flow exceeding some threshold T:

$$m(x) = P[q(n+1) \geq T \mid X(n) = x]. \tag{28}$$

By now the reader may anticipate that the NN estimator is

$$m_n(x) = (1/k(n)) \sum_{j \in S(x,n)} 1_T(q(i)) \tag{29}$$

where $1_T(\cdot)$ is the indicator function for the event "$q > T$"; i.e., $1_T(q) = 1$, if $q > T$; $1_T(q) = 0$, otherwise. The assertions about convergence and asymptotic theory hold for the estimators we have just mentioned.

Acknowledgements

Both authors benefited greatly from advice and insights offered by Professor S. Sorooshian. Professor S. Burges helped by sending a packet of cogent reading material. The research was supported, in part, by NSF grants CEE-81-10788 and DPP-82-19439.

References

Amorocho, J. 1967: The nonlinear prediction problem in the study of the runoff cycle. Water Resour. Res. 3, 861-874

Amorocho, J.; Brandstetter, A. 1971: Determination of nonlinear functional in rainfall-runoff process. Water Resour. Res. 7, 1087-1101

Anderson, B.D.; Moore, J.B. 1979: Optimal filtering. New Jersey: Prentice-Hall

Bras, R.; Rodriguez-Iturbe, I. 1985: Random functions and hydrology. Reading, Mass.: Addison Wesley

Chander, S.; Shanker, H. 1984: Unit hydrograph based forecast model. J. Hydrological Sciences 29, 279-291

Clark, R.T. 1973: A review of some mathematical models used in hydrology, with observations on their calibration and use. J. Hydrology 19, 1-20

Cooper, D.M.; Wood, E.W. 1982a: Identification on multivariate time series and multivariate input-output models. Water Resour. Res. 18, 937-946

Cooper, D.M.; Wood, E.W. 1982b: Parameter estimation of multiple input-output time series models. Water Resour. Res. 18, 1352-1364

Diskin, M.H.; Boneh, A. 1973: Determination of optimal Kernels for second-order stationary surface runoff systems. Water Resour. Res. 9, 111-126

Dooge, J. 1959: A general theory of the unit hydrograph. J. Geophysical Res. 64, 241-256

Eagleson, P.S.; Mejia-R, R.; March, R. 1966: Computation of optimal realizable unit hydrograph. Water Resour. Res. 2, 755-765

Freeze, R.A. 1982: Hydrogeological concepts in stochastic and deterministic rainfall-runoff predictions. Special paper 189, Geological Society of America

Hino, M. 1970: Runoff forecasts by linear predictive filter. J. Hydraulics Division. Proc. Am. Soc. of Civil Engineers, HY3, 681-707

Jacoby, S.L.S. 1966: A mathematical model for nonlinear hydrologic systems. J. Geophysical Res. 71, 4811-4824

Karlsson, M.; Yakowitz, S. 1987: Nearest-neighbor methods for non-parametric rainfall-runoff forecasting. Water Resour. Res. 27, 1300-1308

Kashyap, R.L.; Rao, A.R. 1973: Real time recursive prediction of river flows. Automatica 9, 175-183

Kitanidis, P.K.; Bras, R.L. 1980: Real time forecasting with a conceptual hydrologic model. Water Resour. Res. 16, 1025-1033

Klemes, V.; Boruvka, L. 1975: Output from a cascade of discrete linear reservoirs. J. Hydrology 27, 1-13

Lawson, C.; Hanson, R. 1974: Solving least squares problems. New Jersey: Prentice-Hall

Linsley, R.K.; Kohler, M.A.; Paulhus, J.L.H. 1982: Hydrology for engineers. New York: McGraw-Hill (2nd edition)

Ljung, L.; Soderstrom, T. 1983: Theory and practice of recursive identification. Cambridge, Mass.: The MIT Press

Minsall, N.E. 1960: Predicting storm runoff on small experimental watersheds. ASCE J. Hydraulics Division 86, 12-58

Napiorkowski, J.J.; Strupczewski, W.G. 1981: The properties of the Kernels of the Volterra series describing flow deviations from a steady state in an open channel. J. Hydrology 51, 185-199

Nash, J.E. 1959: Systematic determination of unit hydrograph parameters. J. Geophysical Res. 64, 111-115

O'Connell, P.E.; Clarke, R.T. 1981: Adaptive hydrological forecasting -- A review. Hydrological Sciences Bulletin 26, 179-205

Patry, G.G.; Marino, M. 1982: Parameter identification of time varying noisy difference equations for real-time urban runoff forecasting. J. Hydrology 62, 25-55

Patry, G.G.; Marino, M. 1983a: Time-variant nonlinear functional runoff model for real-time forecasting. J. Hydrology 66, 227-244

Patry, G.G.; Marino, M. 1983b: Nonlinear runoff modeling: parameter identification. J. Hydraulic Engineering 109, 865-880

Paynter, H.H. 1952: Methods and results from MIT studies in unsteady flow. J. Boston Society of Civil Engineers 39, 120-165

Priestly, M.B. 1981: Spectral analysis and time series. Reading, Mass.: Academic Press

Rao, S.; Rao, R. 1984: Nonlinear stochastic model of rainfall runoff process. Water Resour. Res. 20, 297-309

Rodriquez-Iturbe, I.; Sanabria, M.G.; Caamano, G. 1982: On the climatic dependence of the IUH: A rainfall-runoff analysis of the Nash model and the geomorphoclimatic theory. Water Resour. Res. 18, 887-903

Sherman, L.K. 1932: Streamflow from rainfall by unit-graph method. Engineering New Record, 501-505

Shibata, R. 1984: Approximate efficiency of a selection procedure for the number of regression variables. Biometrica 71, 43-49

Sorooshian, S. 1983: Surface water hydrology: On-line estimation. Reviews of Geophysics and Space Physics 21, 706-721

Stone, C. 1980: Optimal rates of convergence for nonparametric estimators. Ann. Stat. 8, 1348-1360

Todini, E. 1978: Using a desk-top computer for an on-line flood warning system. IBM J. Res. Develop. 22, 464-471

Wiener, N. 1958: Nonlinear problems in random theory. Cambridge: The MIT press

Yakowitz, S. 1985a: Nonparametric density estimation, prediction and regression for Markov sequences. J. Amer. Stat. Assoc. 80, 215-221

Yakowitz, S. 1985b: Markov flow models and the flood warning problem. Water Resour. Res. 21, 81-88

Yakowitz, S. 1987: Nearest neighbor for time series analysis. J. Time Series Analysis 8, 235-247

Yakowitz, S.; Szidarovszky, F. 1986: An introduction to numerical computations. New York: Macmillan Publ. Co.

Yakowitz, S.; Karlsson, M. 1987: Nearest neighbor methods for time series, with application to rainfall-runoff prediction. In: MacNeil, I.B.; Umphries, G.J. (eds.) Stochastic Hydrology, pp. 149-160. New York: D. Reidel

Accepted July 23, 1987.