VOL. 21, NO. 4

WATER RESOURCES BULLETIN
AMERICAN WATER RESOURCES ASSOCIATION

AUGUST 1985

THRESHOLD TIME SERIES MODELING OF TWO ICELANDIC RIVERFLOW SYSTEMS[1]

H. Tong, B. Thanoon, and G. Gudmundsson[2]

ABSTRACT: This paper reports our experience in building time series models which connect the flows in two Icelandic rivers with the meteorological variables of precipitation and temperature. Two rivers with different hydrological characteristics were studied. In areas where precipitation may be either in the form of rain or snow linear models are inadequate to describe the relationship between the river and the meteorological variables. The methodology of threshold models recently developed seems to be well suited for taking into account the sharp difference in the relationship according to whether it is freezing or not. The possibility of identifying an alternative threshold variable is also explored.
(KEY TERMS: threshold models; nonlinear autoregression; riverflow; precipitation; temperature; Icelandic rivers; glacier; snow melting.)

INTRODUCTION

The observable riverflow, at a point in a river and at a particular time instant, may be thought of as the observable output of a system the input of which is the past effective principitation. The dynamics and the memory length of the system are dictated by the geography, geology and topography of the river region. Commonly mentioned factors are catchment retention, losses through evaporation, transpiration from plants, infiltration into the ground, underground sources, catchment storage and melting snow. Deterministic models describing the relationship between riverflow and meteorological variables must, therefore, inevitably be elaborate and require extensive measurements. The possibility of constructing simpler stochastic models, based on a few meteorological variables, may therefore be worth exploring. Such models can be useful for simulation and prediction and may provide some quantitative information about the relationship between the riverflow and some of the more important meteorological variables. A substantial literature is available on the stochastic modeling of the riverflow alone without incorporating any meteorological variables. An excellent review in this respect is given by Lawrance and Kottegoda (1977). An extension of the transfer function-noise model approach of Box and Jenkins (1976) to analyze the relationship between riverflow and three input series was described by Snorrason, *et al.* (1984).

The statistical analysis of time series data is greatly facilitated if the mean and covariances do not change with time. However, this is not a realistic assumption in hydrology. Seasonal variations and nonlinear relationships between meteorological variables and the river imply that riverflow is neither Gaussian nor stationary. Seasonal variations in second order properties of riverflow data were described by Gudmundsson (1975), and Kavvas and Delleur (1984). Gudmundsson (1970) examined seasonal variations in the relationship between riverflow and meteorological variables.

For many rivers, the associated precipitation may alternate between rainfall and snowfall. In addition, there may be glaciers on the drainage area. In such cases, the meteorological variable of temperature plays a naturally important role. The *threshold* at the freezing point has here a readily identified hydrological-meteorological meaning and it seems reasonable to expect that it will explain some of the nonlinearity (see Figure 1). A class of time series models which gives a special role to the notion of a threshold was developed by Tong in a series of papers. (See Tong, 1983, for an up-to-date detailed account or Tong and Lim, 1980, for a shorter introduction, including hydrological application.)

A THRESHOLD SYSTEM

Suppose that $\{Y_t\}$ and $\{Z_t\}$ denote the inputs and $\{X_t\}$ the output of a composite system. The composite system is supposed to be made up of two subsystems denoted by $L^{(1)}$ and $L^{(2)}$, each of which is linear. The selection of a subsystem is indicated by a command signal, say J_t, which takes the value 1 or 2. That is, $L^{(i)}$ is selected if and only if $J_t = i$, $(i = 1,2)$. Typically, the random variable J_t is defined by reference to the crossing of a certain *threshold* (e.g., freezing point) of an appropriate observable random variable (e.g., temperature). As usual, we may allow past outputs $X_{t-1}, X_{t-2}, \ldots$, to be incorporated in the linear subsystems. They may, in theory, be eliminated if we wish although, in practice, this is not a trivial matter. Mathematically, we may express the set-up as follows:

[1]Paper No. 85091 of the *Water Resources Bulletin.* **Discussions are open until April 1, 1986.**

[2]Respectively, Professor, Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong; Lecturer, Department of Mathematics and Statistics, University of Masul, Iraq; and Statistician, The Central Bank of Iceland, Reykjavik, Iceland.

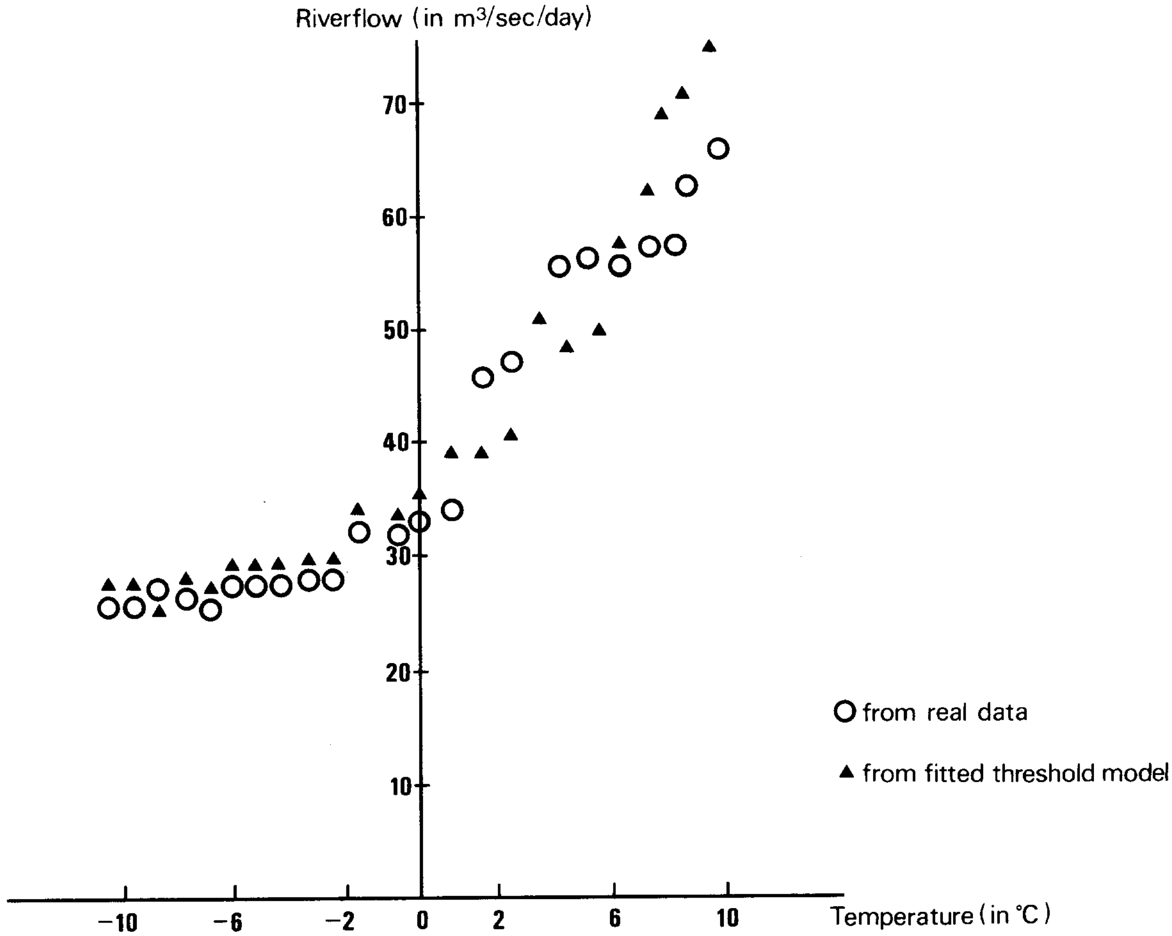

Figure 1. Nonparametric Regression, $E[X|Z = z]$, of riverflow (X) on temperature (Z) for the river Jökulsá eystri, Iceland. The method does not assume an a priori functional form but is based on a (kernel) smoothing on the data historgrams. Specifically, let $\{\delta_N(z)\}$ denote a sequence of symmetric and nonnegative functions of z, of area 1, with the property that $\delta_N(z) \to$ Dirac Delta functions as $N \to \infty$. Then $\widehat{E[X|Z = z]} = \sum_{j=1}^{N} x_j \delta_N(z - z_j) / \sum_{j=1}^{N} \delta_N(z - z_j)$, where $(x_1, z_1), \ldots, (x_N, z_n)$ denote the N data points.

$$X_t = L^{(J_t)}(1; X_{t-1}, X_{t-2}, \ldots; Y_t, Y_{t-1}, \ldots; Z_t, Z_{t-1}, \ldots), \tag{1}$$

the arguments of each linear function $L^{(\cdot)}$ are arranged into obvious groups for convenience, the first one corresponding to a constant.

To simplify the notation, the convention typified by the following example, with fictitious numbers, will be adopted:

$$X_t = \begin{cases} L^{(1)}(2.5 \parallel 1.7 \parallel 3.7, 7.7 \parallel 2.4, -5.6) \text{ if } J_t = 1 \\ L^{(2)}(3.1 \parallel 2.1, -3.2 \parallel -2.6, 1.2 \parallel 3.4) \text{ if } J_t = 2 \end{cases} \tag{2}$$

is to mean that

$$X_t = \begin{cases} 2.5 + 1.7X_{t-1} + 3.7Y_t + 7.7Y_{t-1} + 2.4Z_t \\ \quad - 5.6Z_{t-1} + \epsilon_t^{(1)} \text{ if } J_t = 1 \\ 3.1 + 2.1X_{t-1} - 3.2X_{t-2} - 2.6Y_t + 1.2Y_{t-1} \\ \quad + 3.4Z_t + \epsilon_t^{(2)} \text{ if } J_t = 2 \end{cases} \tag{2'}$$

where $\{\epsilon_t^{(1)}\}$ and $\{\epsilon_t^{(2)}\}$ are two white noise sequences, independent of each other.

The above composite system may be generalized in an obvious manner to one with multi-input, multi-output and multiple subsystems $L^{(1)}, L^{(2)}, \ldots, L^{(k)}$ say, all of which are linear. However, we restrict our discussion to the case of 2-input, 1-output and $k \leqslant 4$ in this paper.

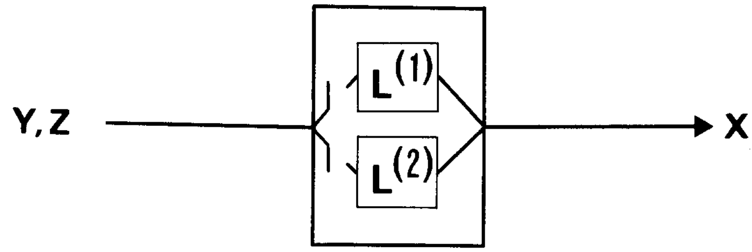

Figure 2. A Composite System.

Given a realization $(X_1, \ldots, X_n; Y_1, \ldots, Y_n; Z_1, \ldots, Z_n; J_1, \ldots, J_n)$, parameters defining $L^{(i)}$, $(i = 1, 2)$, may be estimated by the least-squares method in an obvious way. The first few observations are reserved as initial values and the estimation is applied to the remaining observations which start with the suffix m + 1, say. The observations $(J_{m+1}, \ldots, J_n)$ will sort the data $(X_{m+1}, \ldots, X_n)$ into two groups, one following $L^{(1)}$ and the other $L^{(2)}$. The parameters of each linear subsystem may be obtained by ordinary least squares estimation. Let RSS_1 and RSS_2 denote the residual sum of squared errors corresponding to $L^{(1)}$ and $L^{(2)}$, respectively. Suppose that the sorting assigns n_1 and n_2 observations to $L^{(1)}$ and $L^{(2)}$, respectively, where $n_1 + n_2 = n - m$. We may then calculate

$$\{n_1 \ell n(RSS_1/n_1) + 2k_1\} + \{n_2 \ell n(RSS_2/n_2) + 2k_2\}, \quad (3)$$

where k_1 and k_2 denote the number of independent parameters for $L^{(1)}$ and $L^{(2)}$, respectively. This quantity normalized by the total number of effective observations, i.e., n − m, may be interpreted as a measure per observation of the divergence of the fitted system from an underlying 'optimal' system. Let this normalized quantity be denoted by NAIC(k_1, k_2) ("Normalized Akaike Information Criterion"). After comparing the NAIC values for various combinations of (k_1, k_2), we may adopt that fitted system which has the minimum NAIC. Of course, other decision rules for system identification are available; it is certainly not our intention here to advocate the NAIC approach to the exclusion of others. Suffice it to say that as far as the Icelandic riverflow systems are concerned, the results seem to be encouraging, as we shall see. For the most recent development of AIC, reference may be made to Akaike (1983).

The basic idea of using different linear dynamics over different range of inputs is, of course, not new. References may be made, e.g., to Andronov, *et al.* (1959), within the general context, and to Whittle (1954) and Box and Jenkins (1976, p. 369) within the time series context. More examples are given in Tong (1983).

HYDROLOGICAL AND METEOROLOGICAL CONDITIONS

In order to examine empirically the possibilities of threshold models in dealing with the nonlinear effects associated with the melting of snow and ice we obtained observations of two rivers in Northwest Iceland, Vatnsdalsa and Jökulsá eystri. Geographical and meteorological conditions are, in many aspects, rather similar on the drainage areas of both rivers. The bedrock consists mainly of basalts of low permeability. These are partly covered by sediments. There are no woods and vegetation is negligible. No direct observations of evaporation are available, but it might be of the order of 20 percent with a substantial seasonal variation. A detailed description of the hydrological conditions in this area was given by Richter and Schunke (1981).

The main characteristic of Vatnsdalsa is direct runoff. There are, however, some highly permeable post-glacial lavas on the southern part of the drainage area which contribute a component of groundwater that is not sensitive to short-term variations in the weather. The drainage area is 450 km^2. The recording station is at a height of 70 m but most of the drainage area is at an altitude of 400-800 m.

The most important difference of the drainage area of Jökulsá eystri as compared with Vatnsdalsa is that it includes a glacier covering 155 km^2 out of a total area of 1200 km^2. The presence of the glacier has the effect that temperatures above zero at its altitude, 1000-1800 m, always produce meltwater, whereas from July and into the autumn there is negligible snow on other parts of the drainage areas.

The meteorological station at Hveravellir lies between the two drainage areas on a level with their southern borders at an altitude of 641 m. The average temperature is about -1^{o}C and the amplitude of the annual variation 6-7^{o}C. The temperature measurements should provide a fairly good indicator of the temperatures on the drainage areas, but it must be kept in mind that, as a result of different altitudes and different distance from the sea, the temperature (at any time) will differ within each area. A difference in altitude of 100 m may correspond to a difference in temperature of 0.5-1.0^{o}C. The diurnal variation implies that on a day when the average temperature is at the freezing point the actual temperature is somewhat higher for a couple of hours.

The precipitation is less well represented by the observations at Hveravellir. It is difficult to measure accurately, especially when wind is high as is common in these areas. Precipitation is subject to much more local variation than temperature, so that no single station will provide an accurate indicator of precipitation within a large area.

THRESHOLD MODELING OF VATNSDALSA RIVER, ICELAND (1972-1974)

The data consist of the daily flow of the river Vatnsdalsa (X_t) in m^3/sec., the daily precipitation (Y_t) in mm, and the mean daily temperature (X_t) in $^{\circ}$C at the meteorological station at Hveravellir. The data span the period of 1972, 1973 and 1974. The precipitation record is actually several hours late as the recorded value is the accumulated rain at 9 a.m. from the same time the day before. (We have adjusted for this in our modeling by a forward translation by one day.) Some of the data are illustrated in Figures 3 and 4.

Before examining threshold models, it is informative to look at ordinary linear models with precipitation and temperature as inputs and the riverflow as output. For the data in 1972 alone, the model is

$$X_t = 9.40 - 0.07Y_t + 0.17Z_t + 0.11Z_{t-1} + \epsilon_t , \qquad (4)$$

where var ϵ_t = 215.64 and NAIC = 3.09. The inadequacy of these models is indicated by the negative coefficient of precipitation and the magnitude of var ϵ_t which is larger than the squared average value of X_t. An explanation of this lies in the model's inability to cope with the obviously highly nonlinear relationship between the riverflow and meteorological variables; an increase in temperature from -15°C to -5°C has very different effects from an increase from 0°C to 10°C.

Within the linear framework, the fit is greatly improved by including past values of X:

$$\begin{aligned} X_t = {} & 0.73 + 1.12X_{t-1} - 0.23X_{t-2} + 0.12X_{t-3} \\ & - 0.09X_{t-4} + 0.01Y_t + 0.07Y_{t-1} \\ & - 0.06Y_{t-2} + 0.02Y_{t-3} + 0.09Z_t \\ & - 0.03Z_{t-1} - 0.04Z_{t-2} + \epsilon_t \qquad (5) \end{aligned}$$

where var ϵ_t = 2.85 and NAIC = 1.11. A great deal of water that is released on the drainage area by rain or melting snow reaches the point of observation on the same day, but part of it arrives later because of long distances, low slopes and delays through the groundwater system. Past and present flow is therefore a useful indicator of tomorrow's flow. In fact, the magnitudes of the parameters show that the role of the meteorological variables in (5) is limited to modifying the dynamics described by the autoregressive part rather than providing a description of the actual relationship between X and (Y, Z).

The nonlinear effects of Z on X, described earlier, suggest that the 'autoregressive' dynamics may well be state-dependent; e.g., (1) it may depend on whether there has been a prolonged period of frost or a prolonged period of warm weather; and (2) it may depend on the unobserved state of the groundwater.

Now, there are three obvious methods of selecting J_t:

$$(1) \quad J_t = \begin{cases} 1 & \text{if } X_{t-1} \leq r_x, \\ 2 & \text{if } X_{t-1} > r_x, \end{cases} \qquad (6a)$$

$$(2) \quad J_t = \begin{cases} 1 & \text{if } Y_{t-1} \leq r_y, \\ 2 & \text{if } Y_{t-1} > r_y, \end{cases} \qquad (6b)$$

$$(3) \quad J_t = \begin{cases} 1 & \text{if } Z_t \leq r_z, \\ 2 & \text{if } Z_t > r_z . \end{cases} \qquad (6c)$$

Note that Y_{t-1} consists of 15-hour accumulation of rain on day (t−1) and a nine-hour accumulation of rain on day t. Thus, it is not realistic to use Y_t for J_t.

Experimentations with the identification package documented in Tong (1983) suggest that method (2) is the least efficient and that X_{t-1} is a parsimonious choice for J_t (see Table 1). A hydrological explanation of this is that X_{t-1} gives an indiciation of whether and how much water is being released on the drainage area.

TABLE 1. Choice of Indicator Variables (for 1972 data set).

Method	Threshold Estimate	Minimum Normalized AIC
(1)	$\hat{r}_x = 12$	0.2482
(2)	$\hat{r}_y = 9$	0.8336
(3)	$\hat{r}_z = 0$	0.5405

We now use method (1) and the pooled data from 1972, 1973 and 1974. The pooling is essential because the spring floods which account for a large proportion of the variation only last for a few days. The following threshold model is identified:

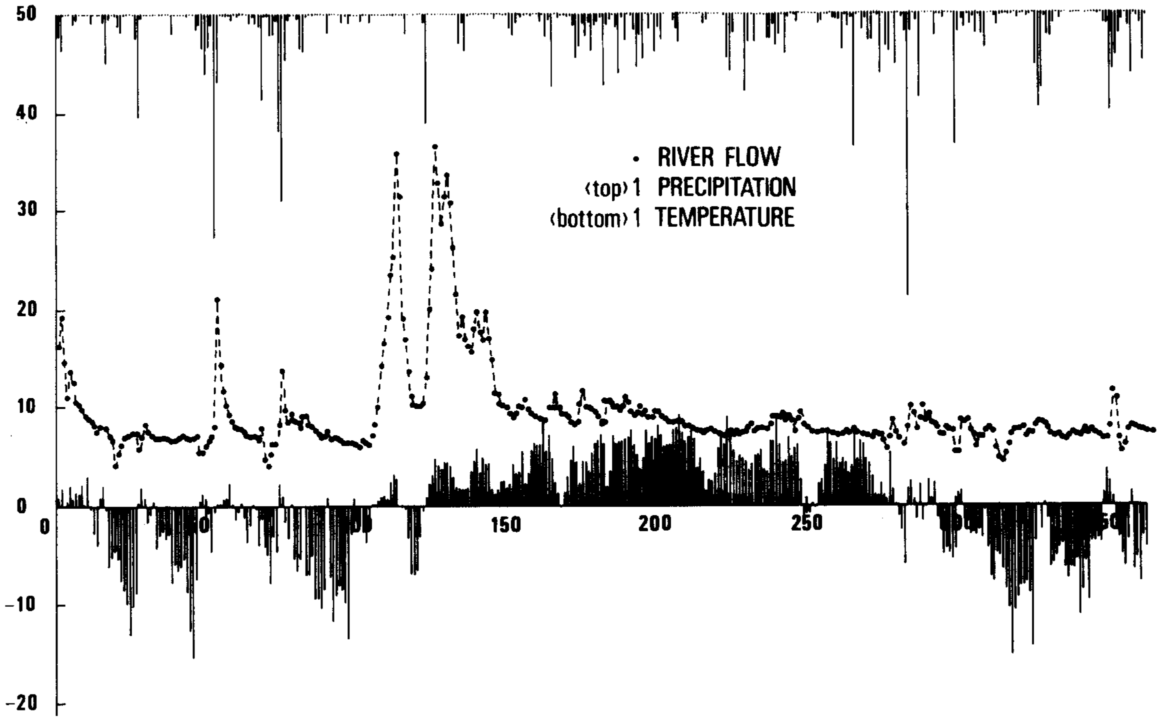

Figure 3. Vatnsdalsa River Data (1972).

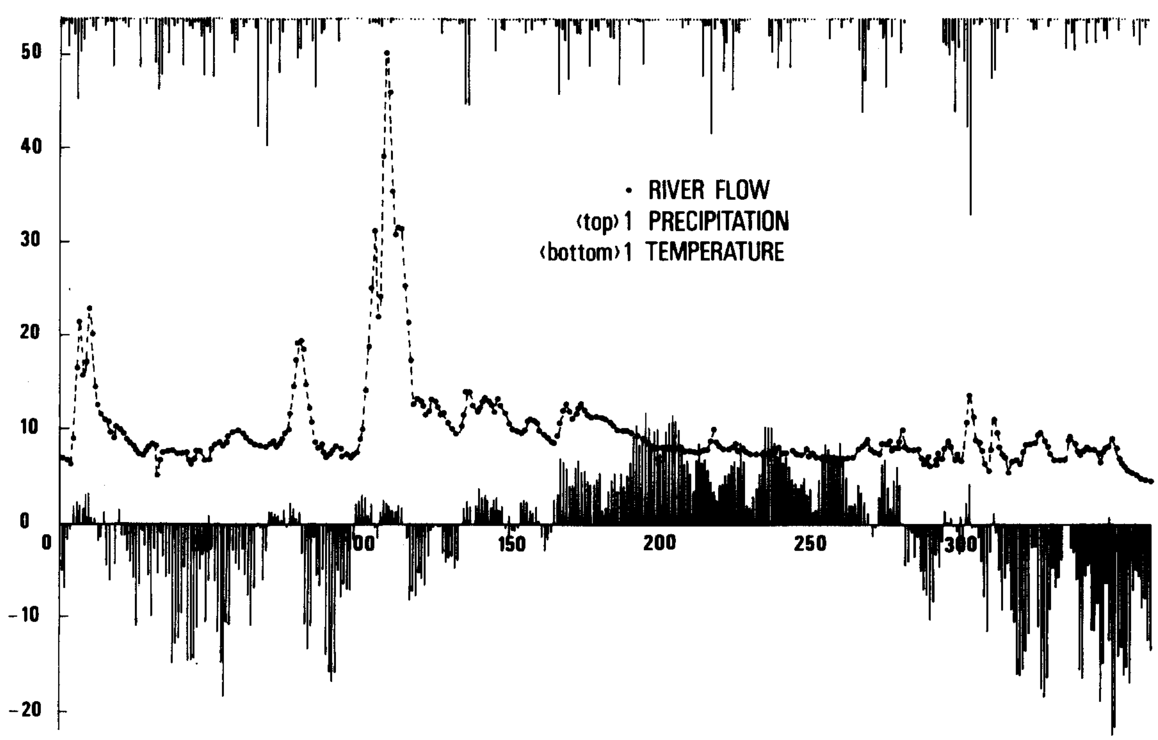

Figure 4. Vatnsdalsa River Data (1973).

$$X_t = \begin{cases} L^{(1)} \; \underset{(0.13)}{(0.14} \,\|\, \underset{(0.05)}{1.31}, \underset{(0.06)}{-0.57}, \underset{(0.04)}{0.22} \,\|\, \underset{(0.00)}{0.02}, \underset{(0.01)}{0.09}, \underset{(0.01)}{-0.02} \,\|\, \underset{(0.01)}{0.03}, \underset{(0.02)}{-0.04}, \underset{(0.01)}{0.01)} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{with var } \epsilon_t^{(1)} = 1.03 \\ L^{(2)} \; \underset{(1.49)}{(7.49} \,\|\, \underset{(0.08)}{0.98}, \underset{(0.13)}{-0.24}, \underset{(0.13)}{-0.05}, \underset{(0.13)}{0.24}, \underset{(0.13)}{-0.13}, \underset{(0.13)}{-0.13}, \underset{(0.12)}{0.17}, \underset{(0.07)}{-0.11} \,\|\, \underset{(0.20)}{0.47}, \underset{(0.11)}{-0.19}, \underset{(0.06)}{-0.02}, \underset{(0.05)}{-0.06}, \underset{(0.06)}{-0.07}, \\ \underset{(0.13)}{-0.50}, \underset{(0.07)}{0.16}, \underset{(0.06)}{0.11} \,\|\, \underset{(0.26)}{0.41}, \underset{(0.45)}{1.20}, \underset{(0.51)}{-2.27}, \underset{(0.43)}{0.59}, \underset{(0.46)}{-0.87}, \underset{(0.42)}{0.29}, \underset{(0.45)}{0.46}, \underset{(0.33)}{-0.69)} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{with var } \epsilon_t^{(2)} = 15.47 \text{ (pooled var} = 2.40) \end{cases} \quad (7)$$

$\hat{r}_X = 13$ and NAIC = 0.3528. Estimated standard deviations are presented in parentheses below respective parameter. These estimates are based on the assumption of independent Gaussian residuals. The actual residual distribution is somewhat skewed and long-tailed so that the parameter estimates are less accurate than suggested by these values.

The equation for yesterday's flow below 13 m^3/sec. applies to 982 days out of 1085. The coefficients of the meteorological variables are on the whole negligible, so that either their effect is small or the model is inadequate to describe it. In the equation for yesterday's flow above the threshold, which applies to only 103 days, meteorological variables have a substantial influence. The coefficient of yesterday's flow is 0.98 and the sum of the coefficients of past flows is 0.73. As a result of the large autoregressive effects, the coefficients of the meteorological variables, say, Y_{t-j}, Z_{t-j}, cannot be interpreted directly as a measure of the effects of these variables j days ago on the present flow. Nor is it appropriate to eliminate the past values of the flow and describe the flow as a function of past values of the meteorological variables alone; the model thus obtained differs substantially from what is obtained by a direct estimation of this form. The past flow contains information which cannot be expressed by the meteorological variables within the present model.

In interpreting the above model, it is useful to keep in mind the fact that the equation for yesterday's flow above 13 m^3/sec. will not be applicable unless yesterday's temperature is about or above the freezing point as otherwise yesterday's flow will be below the threshold value. In this connection, therefore, it is interesting to reanalyze the data using a temperature threshold. The following model is identified:

$$X_t = \begin{cases} L^{(1)}(1.79 \,\|\, 0.76, -0.05) \text{ with var } \epsilon_t^{(1)} = 0.69, \\ L^{(2)}(0.87 \,\|\, 1.30, -0.71, 0.34) \text{ with var } \epsilon_t^{(2)} = 7.18, \end{cases} \quad (8)$$
$$\text{(pooled var} = 4.50)$$

$\hat{r}_Z = -1$, and NAIC = 1.01. We have essentially two different autoregressive models, one for frost and the other for thaw. In the lower temperature range, the model mainly describes a convergence towards a constant flow of about 6 m^3/sec. reached in a few days after frost sets in. This agrees well with the fact that stable flows in this range are often observed for days or weeks. There is a great difference in residual variation between the two models. A few large floods, caused by the melting of a large proportion of the snow on the drainage area, are responsible for much of the variation in the model above the threshold temperature. This applies also to $L^{(2)}$ in Equation (7); the peaks of the floods are preceded and followed by days of large flows. The temperature is $\leq -1^{\circ}$C for 448 days. Thus, in spite of its relatively inferior NAIC value, model (8) may be used to complement model (7).

Prompted by model (8) and guided by Figure 5, the following more elaborate model (9) may be entertained:

$$X_t = L^{(1)} \; \underset{(.13)}{(0.75} \,\|\, \underset{(.05)}{1.06}, \underset{(.06)}{-0.26}, \underset{(.00)}{0.09}, \underset{(.04)}{-0.11}, \underset{(.02)}{0.08} \,\|\, \underset{(.01)}{-0.02}, \underset{(.01)}{-0.01}, \underset{(.01)}{-0.00}, \underset{(.01)}{0.01} \,\|\, \underset{(.01)}{0.02}, \underset{(.01)}{-0.03}, \underset{(.01)}{0.01}, \underset{(.01)}{-0.02)} + \epsilon_t^{(1)} \quad \text{if } Z_t \leq -2$$

$$X_t = L^{(2)} \; \underset{(.29)}{(1.21} \,\|\, \underset{(.05)}{0.97}, \underset{(.07)}{-0.29}, \underset{(.07)}{0.04}, \underset{(.05)}{0.11} \,\|\, \underset{(.02)}{0.03}, \underset{(.02)}{0.12}, \underset{(.02)}{-0.04}, \underset{(.01)}{-0.02} \,\|\, \underset{(.14)}{0.53}, \underset{(.08)}{0.02}, \underset{(.05)}{-0.02)} + \epsilon_t^{(2)} \quad \text{if } -2 < Z_t \leq 2$$

X_t

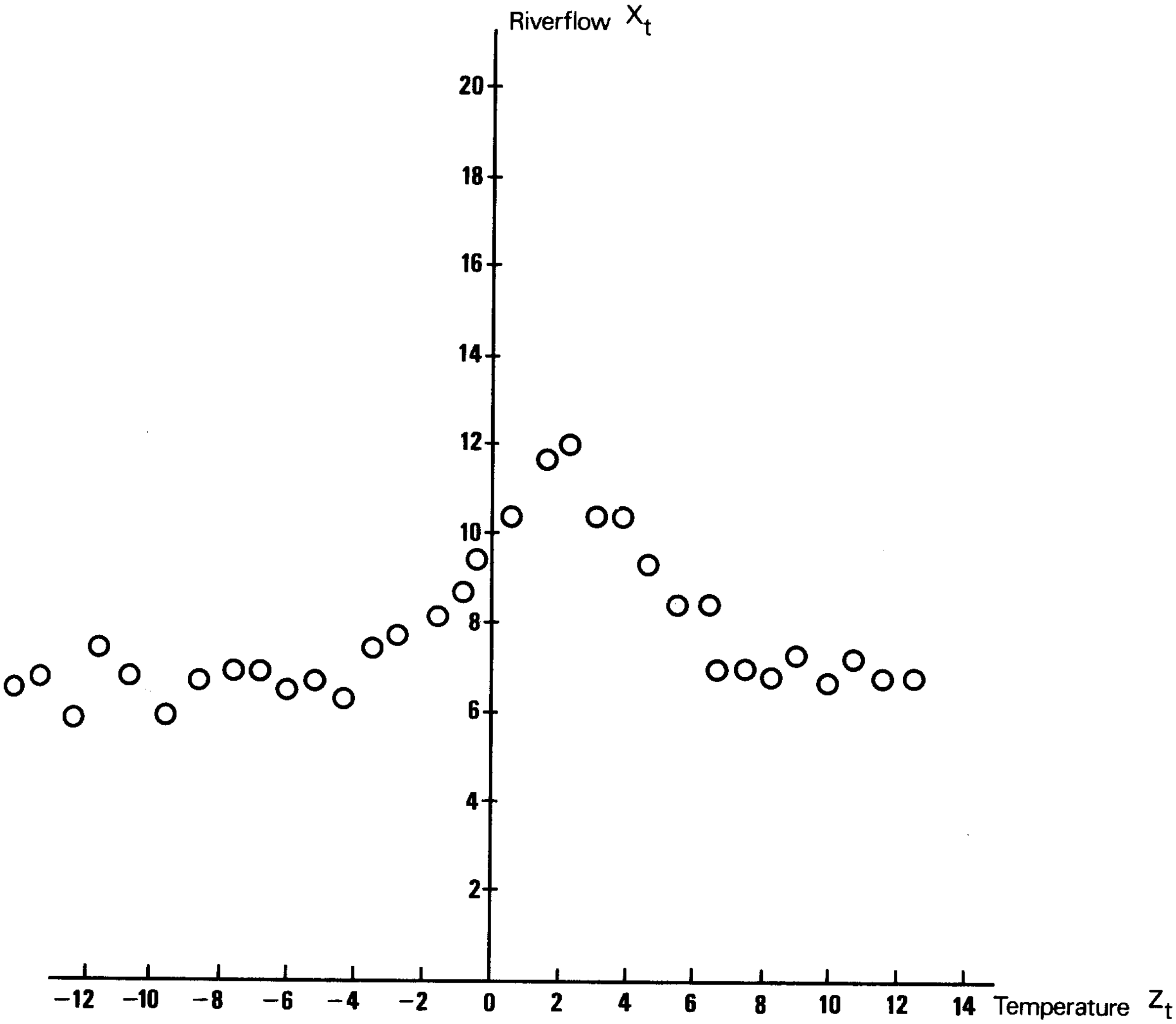

Figure 5. Nonparametric Regression of X_t on Z_t.

$$
\begin{aligned}
X_t = L^{(3)} \Big(& \underset{(.92)}{1.97} \;\|\; \underset{(.08)}{1.38}, \underset{(.15)}{-0.70}, \underset{(.15)}{0.47}, \underset{(.14)}{0.02}, \underset{(.14)}{-0.19}, \\
& \underset{(.12)}{-0.02}, \underset{(.09)}{0.34}, \underset{(.07)}{-0.23} \;\|\; \underset{(.04)}{0.03}, \underset{(.03)}{0.03}, \\
& \underset{(.03)}{-0.01}, \underset{(.03)}{0.04}, \underset{(.03)}{-0.03}, \underset{(.03)}{-0.04}, \\
& \underset{(.03)}{0.13}, \underset{(.02)}{0.01} \;\|\; \underset{(.25)}{-0.59}, \underset{(.14)}{0.07}, \\
& \underset{(.14)}{-0.11}, \underset{(.15)}{-0.05}, \underset{(.17)}{0.07}, \underset{(.16)}{0.13}, \\
& \underset{(.12)}{-0.25} \Big) + \epsilon_t^{(3)} \qquad \text{if } 2 < Z_t \leqslant 5
\end{aligned}
$$

$$
\begin{aligned}
X_t = L^{(4)} \Big(& \underset{(.21)}{0.59} \;\|\; \underset{(.06)}{1.22}, \underset{(.09)}{-0.49}, \underset{(.08)}{0.30}, \underset{(.08)}{-0.17}, \\
& \underset{(.08)}{0.27}, \underset{(.07)}{-0.26}, \underset{(.04)}{0.11} \;\|\; \underset{(.00)}{0.01}, \underset{(.00)}{0.01}, \\
& \underset{(.01)}{-0.01}, \underset{(.08)}{-0.01}, \underset{(.01)}{0.01}, \underset{(.01)}{-0.02} \;\|\; \\
& \underset{(.02)}{-0.02}, \underset{(.02)}{-0.02}, \underset{(.03)}{0.01}, \underset{(.02)}{-0.01}, \underset{(.02)}{0.02}, \\
& \underset{(.02)}{-0.04} \Big) + \epsilon_t^{(4)} \quad \text{if } Z_t > 5,
\end{aligned}
\tag{9}
$$

where

$$\mathrm{Var}\,(\epsilon_t^{(1)}) = 0.3730, \qquad \mathrm{Var}\,(\epsilon_t^{(2)}) = 5.7714$$

$$\mathrm{Var}\,(\epsilon_t^{(3)}) = 6.0659, \qquad \mathrm{Var}\,(\epsilon_t^{(4)}) = 0.1642$$

$$\text{Pooled Variance} = 2.8432, \qquad \text{NAIC} = 0.2466$$

$$n_1 = 387, \; n_2 = 295, \; n_3 = 199, \; n_4 = 205 \,.$$

Brief comments on model (9) will be made in the final section.

THRESHOLD MODELING OF JÖKULSA EYSTRI

Some of the data are illustrated in Figures 6 and 7. Note the much bigger dynamic range of the data.

An important hydrological feature of the river Jökulsa eystri is that there is a glacier on the drainage area. This effects a substantial difference in the response of the two rivers to changes in temperature. In Vatnsdalsa, high temperature only enhances the flow when there is snow to melt, whereas the glacier is still there after all the winter snow has vanished at Jökulsa. It turns out that for 1972 the minimum normalized AIC using method (1) (i.e., with X_{t-1} as the indicator), is 1.9106 and that using method (3) (i.e., with Z_t as the indicator), is 2.0401. For the other two years, the results are quite similar with the minor difference that, for 1973, the result based on method (1) has a slightly bigger minimum normalized AIC value than that based on method (3). Method (2) gives a consistently larger value. We may draw the conclusion that for Jökulsa Z_t is just as good an indicator of hydrological conditions as X_{t-1}.

The following threshold model has been fitted to the Jökulsa data of 1972-1974:

$$X_t = \begin{cases} L^{(1)}\,(\underset{(0.48)}{6.15} \,\|\, \underset{(0.04)}{0.70}, \underset{(0.05)}{0.04}, \underset{(0.03)}{0.03}, \underset{(0.02)}{-0.05}, \underset{(0.02)}{-0.01}, \underset{(0.01)}{0.06} \,\|\, \underset{(0.01)}{-0.02}, \underset{(0.02)}{-0.01}, \underset{(0.01)}{-0.03}, \underset{(0.02)}{0.02}, \underset{(0.01)}{0.00}, \underset{(0.01)}{0.01}, \,\|\, \underset{(0.01)}{0.04}, \\ \qquad \underset{(0.02)}{0.00}, \underset{(0.02)}{-0.01}, \underset{(0.02)}{-0.04}, \underset{(0.02)}{-0.00}, \underset{(0.01)}{0.01}) \text{ with } \mathrm{var}\,\epsilon_t^{(1)} = 0.67 \\ L^{(2)}\,(\underset{(0.87)}{1.11} \,\|\, \underset{(0.04)}{1.18}, \underset{(0.06)}{-0.47}, \underset{(0.06)}{0.32}, \underset{(0.06)}{-0.20}, \underset{(0.06)}{0.15}, \underset{(0.06)}{-0.11}, \underset{(0.06)}{0.01}, \underset{(0.03)}{0.05} \,\|\, \underset{(0.04)}{0.01}, \underset{(0.04)}{0.37}, \underset{(0.05)}{-0.21}, \underset{(0.04)}{-0.05}, \underset{(0.05)}{0.05} \\ \qquad \underset{(0.05)}{-0.01}, \underset{(0.05)}{0.06}, \underset{(0.04)}{0.11} \,\|\, \underset{(0.16)}{0.72}, \underset{(0.18)}{0.56}, \underset{(0.17)}{-0.10}, \underset{(0.18)}{-0.21}, \underset{(0.19)}{0.01}, \underset{(0.18)}{-0.13}, \underset{(0.18)}{-0.02}, \underset{(0.14)}{-0.21}) \end{cases}$$

$$\text{with var } \epsilon_t^{(2)} = 48.99 \text{ (pooled variance} = 31.77) \qquad (10)$$

$\hat{r}_Z = -2^{\circ}$C. A nonparametric estimate of $E[X_t|Z_t]$ suggests that the latter is piecewise linear consisting of a horizontal line cutting the vertical axis at 26 m^2/sec. and a line of positive slope, the knot being at $Z_t = -2^{\circ}$C. Thus, the estimate $\hat{r}_Z = -2^{\circ}$C seems reasonable (see Figure 1).

In days of frost the model describes a gradual decrease, very similar to the model for Vatnsdalsa, with negligible contribution from the meteorological variables. According to the model, the flow approaches 26 m^3/sec. in prolonged periods of frost, which is in reasonable agreement with the observations.

In $L^{(2)}$, the sum of the coefficients of past flows is 0.93 and yesterday's flow has the coefficient 1.18. Coefficients of the meteorological coefficients are not negligible. Considering the numerical values of the coefficients of the meteorological variables and the fact that the flow is usually greater than 30 m^3/sec. with a standard deviation of 7 m^3/sec. it is, however, clear that the role of these variables is to modify the dynamics described by the autoregressive part rather than to provide a description of the actual relationship between X and (Y, Z). The coefficient of Y_t is practically zero, which implies that the effect of today's precipitation on today's flow is adequately accounted for by multiplying yesterday's flow by 1.18. Yesterday's precipitation is, on the other hand, underrepresented by $1.18X_{t-1} - 0.47X_{t-2}$ and this is compensated by the term $0.37Y_{t-1}$, which is probably also related to the fact that it takes the water about a half-day to travel from the glacier to the point of observation. The coefficients of today's and yesterday's temperature are positive, but those of days t–2, t–3, . . . , t–7 are mainly negative. To some extent this does take into account the effects of snow melting. A present temperature of 8°C say in March-April, when there is snow to be melted, will usually be preceded by days of lower temperatures than a day of 8°C in August.

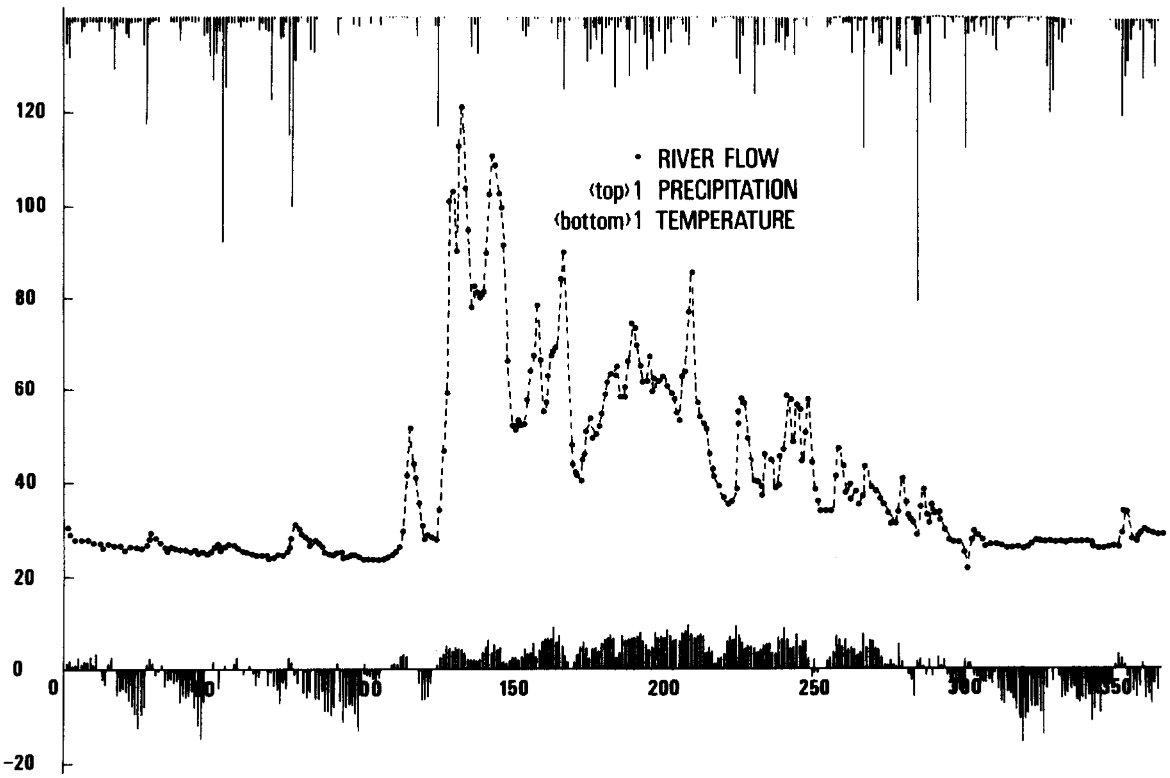

Figure 6. Jökulsa Data (1972).

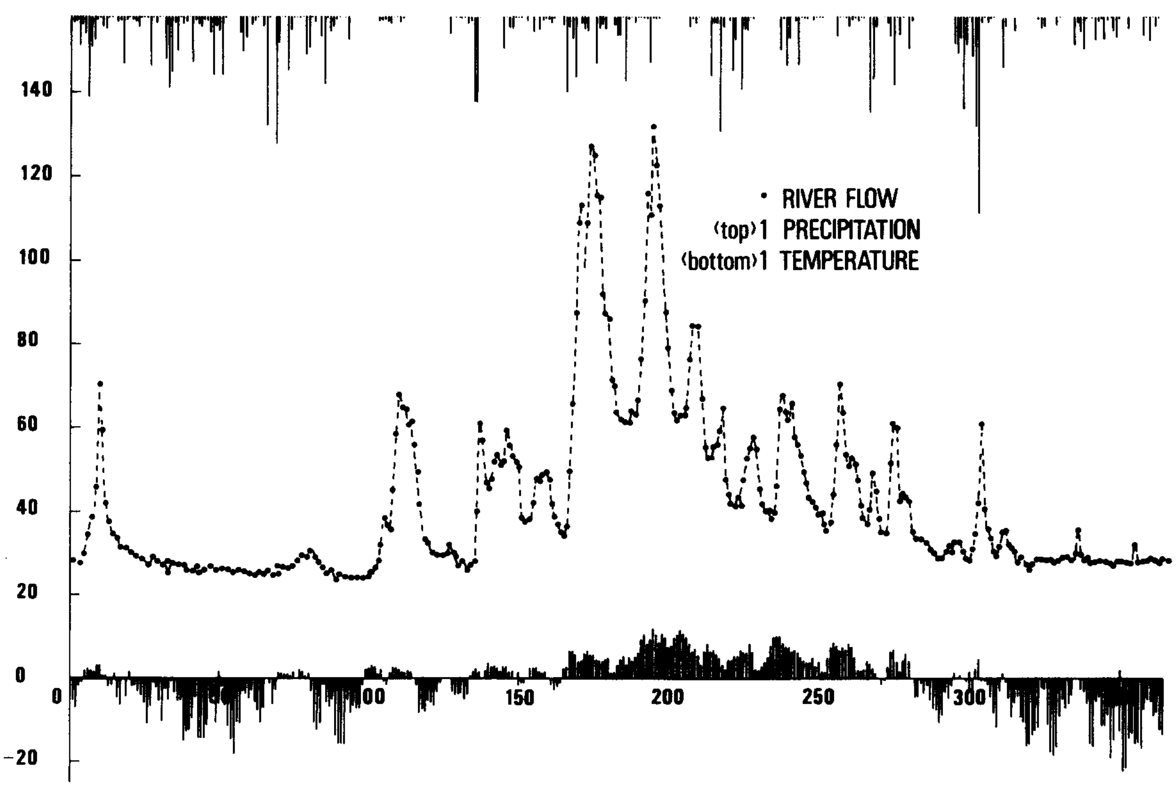

Figure 7. Jökulsa Data (1973).

The model obtained with yesterday's flow as threshold parameter is very similar to the model (10) when the flow is larger than 30 m^3/sec. and the temperatures larger than -2^oC. When yesterday's flow is less than 30 m^3/sec., the former model gives "today's flow is practically the same as yesterday's."

DISCUSSION

To a certain extent the results support the view that threshold models may be suitable for analyzing these kinds of data. These models are more accurate than linear models and the temperature thresholds identified by the AIC are in accordance with the main characteristics of the rivers. Further details are available in the unpublished doctoral thesis of Thanoon (1984, University of Manchester, U.K.).

Apart from long-term changes in the geography the flow is completely determined by the weather; there is no feedback detectable on the level of accuracy attainable by actual observations. The estimated models hardly accord with this, however, for the autoregressive coefficients contribute much more to the description of the variations of the rivers than the meteorological variables.

It is not surprising, however, that past observations of river flow contain a great deal of information about its present level. The weather affects the river with "distributed lags." Thus some of the runoff after a burst of rain reaches the point of observation within a few hours as direct runoff along steep slopes over a short distance whereas the rain from other parts of the area, falling at the same time, may have to pass through layers of soil or travel longer distances over lower slopes. Observations of the river itself thus may be expected to contain information about its future.

If the models provide a satisfactory description of the relationship between the respective variables (included), elimination of the autoregressive part should lead to expressions,

$$X_t = L^{(J_t)}(1; Y_t, Y_{t-1}, \ldots; Z_t, Z_{t-1}, \ldots) + \eta_t^{(J_t)}, \qquad (11)$$

where the coefficients of the meteorological variables provide realistic estimates of the actual time lags. The model (9) of Vatnsdalsa for temperatures between -2^oC and 2^oC might be an example of this, and when temperatures are below -2^oC variations in temperature and precipitation have little effect upon the river so that the low coefficients of these variables are also realistic.

The residual term $\eta_t^{(J_t)}$ is a moving average of $\epsilon_t^{(J_t)}$. The new residuals would possess high positive autocorrelations. This is, however, not unrealistic. The river responds differently to the weather depending upon the temperature and the threshold models take account of this. However, the river also responds differently to a given sequence of meteorological variables depending upon the state of the drainage area. Here the amount of snow is most important, but groundwater and frost in the ground can also have large effects (Rist, 1983). The thresholds cannot fully cope with both the nonlinear character of the relationship in given hydrological conditions and the effect of the weather for altering the hydrological conditions. They are better suited to deal with the first, but the residual errors associated with variations in the state of the drainage area are obviously positively correlated.

Some of the models with large and apparently well determined autoregressive parts do not lead to sensible expressions when rewritten in the form of (11). Past values of the rivers therefore contain useful information which cannot be expressed by the meteorological variables within the present models.

The distributions of rainfall and meltwater production, which may be regarded as the inputs to the hydrological system, are very skewed. Therefore, residuals will not be Gaussian unless a model presents a fairly accurate description of the actual relationships. Presumably realistic and accurate hydrological models can be produced with nonGaussian residuals. Provided the residuals are approximately uncorrelated, the nonGaussianness does not seem to have much effect on the AIC or the large sample standard errors of the parameter estimates. However, available finite sample results are too scanty. (Some preliminary investigations are reported in Tong (1983) and Petruccelli and Woolford (1984).)

The threshold models described in this paper can only be considered partially successful from the point of view of reducing the data to "Gaussian white noise." There are various possibilities of improving upon the modeling. Among these are the introduction of more meteorological variables, such as wind and radiation or precipitation from more than one station. Direct observations of the state of the hydrological system, however, might be more to the point. With such information appropriate modeling would obviously differ a great deal from the models presented in this paper. However, sharp changes in relationships according to temperature and the presence of snow would still be present. Thresholds in some form could, therefore, also be convenient for dealing with the nonlinearity in more elaborate models.

ACKNOWLEDGMENTS

We are grateful to the Hydrological Survey of the National Energy Authority of Iceland for supplying us with the data analyzed in this paper.

REFERENCES

Akaike, H., 1983. Information Measures and Model Selection. Proc. 44th Session ISI, Madrid, Spain, Vol. 1:277-290.

Andronov, A. A., A. A. Vitt, and S. E. Khaikin, 1959. Theory of Oscillations (in Russian). English Translation published by Pergamon Press, 1966.

Box, G. E. P. and G. M. Jenkins, 1976. Time Series Analysis, Forecasting and Control (Revised Ed.). Holden-Day, San Francisco, California.

Gudmundsson, G., 1970. Short Term Variations of a Glacier-Fed River. Tellus 22:341-353.

Gudmundsson, G., 1975. Seasonal Variations and Stationarity. Nordic Hydrology 6:137-144.

Kavvas, M. L. and J. W. Delleur, 1984. A Statistical Analysis of the Daily Streamflow Hydrograph. J. of Hydrology 71:253-276.

Lawrance, A. J. and N. T. Kottegoda, 1977. Stochastic Modelling of Riverflow Time Series (with discussion). J. Roy. Statist. Soc. A(140):1-47.

Petruccelli, J. D. and S. W. Woolford, 1984. A Threshold AR(1) Model. J. Appl. Probability 21 (to appear).

Richter, K. and E. Schunke, 1981. Runoff and Water Budget of the Blanda and Vatnsdalsa Periglacial River Basins, Central Iceland. Bull. No. 34, Res. Inst. Nedri As, Hveragerdi, Iceland.

Rist, S., 1983. Floods and Flood Danger in Iceland. Jökull. 33:119-132.

Snorrason, A., P. Newbold, and W. H. C. Maxwell, 1984. Multiple Input Transfer Function-Noise Modelling of River Flow. *In:* Frontiers in Hydrology, W. H. C. Maxwell and L. R. Beard (Editors). Water Resources Publications, Littleton, Colorado, pp. 111-126.

Tong, H., 1983. Threshold Models in Non-Linear Time Series Analysis. Lecture Notes in Statistics No. 21, Springer Verlag, New York, New York.

Tong, H. and K. S. Lim, 1980. Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion). J. Roy. Statist. Soc. B(42): 245-292.

Whittle, P., 1954. The Statistical Analysis of a Seiche Record. Sears Foundation J. of Marine Resource 13:76-100.