Nordic Hydrology, 17, 1986, 371-382

Transformation of Input to a Stochastic Model using a Distributed Deterministic Model

Paper presented at the Nordic Hydrological Conference
(Reykjavik, Iceland, August – 1986)

Á. Snorrason and K. Einarsson
Orkustofnun, National Energy Authority,
Reykjavik, Iceland

In recent years, hydrologists have attempted, with encouraging results, to synthesize the accomplishments of time series analysis and deterministic systems methods into the more general framework of stochastic dynamic systems. The purpose of this paper is to attempt an improvement of a stochastic time series model by the use of inputs generated by a distributed, deterministic model.

The analysis is based on a simplified transfer function-noise model, where the system output is modeled as a linear combination of present and past inputs as well as past outputs.

The systems chosen for the analysis are two partially glaciated basins in Iceland.

The results were in general promising and showed that a careful selection of input series could overcome to a great extent the non-linearities as well as the non-stationarities in the relationship between discharge and meteorological variables. The results also showed that transformation of the meteorological series using distributed, deterministic techniques gave superiour results, compared with the standard degree-day approach.

Introduction

The purpose of this paper is to attempt to improve a simplified time series model of the relationship between discharge and meteorological variables. This is done by generating the inputs to the system by a distributed, deterministic model. This approach is applied to the analysis of river flow data from two partly glaciated basins in Iceland. Two sets of inputs are tried. Firstly, the temperature series is transformed into a degree-day series, and secondly, the inputs are generated by the use of a distributed deterministic model.

The transformation of the meteorological series is done with three things in mind; firstly, the non-stationarity of the transfer function is hopefully minimized. Secondly, the non-linearities between temperature and melted water will hopefully be reduced. Finally, to account for the fact that the drainage area is all at higher elevation than is the meteorological station.

The results of the analysis are compared by studying the resulting generated series for each set of input series, and when possible, in relation with measured flow.

Methodology

The discussion of the methodology is separated into two parts. First, the model is presented and described, then the generation of the inputs is discussed.

Model

A general class of discrete linear transfer function-noise models was introduced by Box and Jenkins (1970). It has been discussed, extended, and applied to hydrological systems by many investigators, e.g. Gudmundsson (1970), Chow (1973, 1978), Clark (1974), Anselmo and Ubertini (1979), Barcos et al. (1981), Hsu and Adamowski (1981), Snorrason (1983), Snorrason et al. (1984), Hipel (1985), and Snorrason (1985).

The transfer function-noise, TFN, model relates the output of a system to its inputs, adding a noise series. The schematic form of such a model, coupled with a system which transforms the inputs, is shown in Fig. 1. For a system with a single input as well as a single output the TFN model takes the form

$$(1-\delta_1 B-\ldots-\delta_r B^r)Y_t = (\omega_0+\omega_1 B+\ldots+\omega_s B^s)X_{t-b}+N_t \tag{1}$$

In Eq. (1), B is the backward shift operator defined as $B^kY_t=Y_{t-k}$, Y_t is the output series, X_t is the input series, b is the pure delay in the system, and N_t is the noise series, which can be represented as a moving average process of order q

$$N_t = \alpha_t = \theta_1\alpha_{t-1}-\ldots-\theta_q\alpha^q_{t-q} \tag{2}$$

in which α_t is a white noise process with mean equal to zero. By substituting Eq. (2) into Eq. (1), and writing the polynominal operators in B as δ (B), $\omega(B)$ and $\theta(B)$ gives

$$\delta(B)Y_t = \omega(B)X_{t-b}+\theta(B)\alpha_t \tag{3}$$

Extending Eq. (3) to a system with multiple inputs, we have

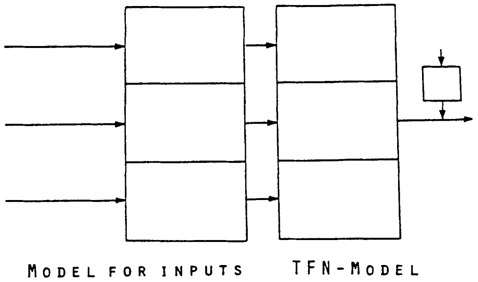

Fig. 1. Schematic representation of a system.

$$\delta(B)Y_t = \sum_{j=1}^{k} \omega_j(B)X_{j,t-b_j} + \theta(B)\alpha_t \tag{4}$$

For the purpose of this paper, Eq. (4) will prove to be sufficient, but for further extension and application of the TFN model, see Snorrason (1983).

A model represented by Eq. (4) can easily be applied to the modeling of relatively uncomplicated systems. It can be written as

$$Y_t = \sum_{j=1}^{k} \frac{\omega_j(B)}{\delta(B)} X_{j,t-b_j} + \frac{\theta(B)}{\delta(B)}\alpha_t \tag{5}$$

or as a lagged-regression equation where the present output is modeled as a linear combination of past and present inputs as well as past outputs, adding a noise series.

$$Y_t = \sum_{i=1}^{r} \delta_i Y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{s_k} \omega_{j,i} X_{j,t-b_j-i} + N_t \tag{6}$$

Eq. (6) is quite suitable for the analysis of simple systems and can be estimated by a least squares estimation. The structure of the model is identified by the use of a cross- and autocorrelation analysis. However, in general it is first necessary to remove from each series, the deterministic variations in the mean. It should be stressed, that both the identification and estimation of distributed lag models is in general difficult, especially if the inputs are highly autocorrelated and/or cross correlated (Granger and Newbold 1974, 1977)

Generation of Input Series

The input series to be compared are both based on temperature and precipitation measurements. The first set consists of degree-days series based on different threshold values and precipitation at nearby meteorological stations. The second set consists of series of glacial snow and icemelt as well as of precipitation and/or snowmelt from non-glaciated parts of the watershed in question, both are generated by the NAM2 semi-deterministic model.

The NAM model was originally described by Nielsen and Hansen (1973) and

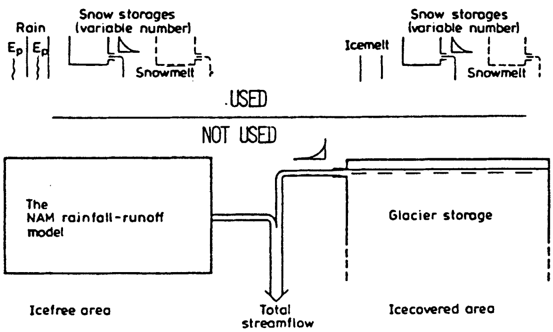

Fig. 2. NAM2 model structure. Modified from Gottlieb (1980).

modified to yield NAM2 by Gottlieb (1980). The additions made by Gottlieb centered around the melting of ice and snow and were mainly based on ideas from Laramie and Schaake (1972), or the so called MIT-model. In the last few years some minor modifications have been made in the NAM2 model in Iceland, Hólm (1984 personal communication), see Einarsson (1984).

The general model structure is shown in Fig. 2. In the icefree area input is collected from subareas, both as rain and (eventually) snowmelt, if so directed by the accounting process used, in conjunction with observed temperature and precipitation. The accounting process is aimed at describing the heat exchange between the atmosphere and the snow. Estimated evapotranspiration is subtracted. In the icecovered area input is collected from the subareas as snowmelt by the same accounting process, and as icemelt by solving the equation of heat conduction to determine the heat exchange between the surface layer and the underlying ice.

The complete NAM2 model operates further on the input, routing it with a partly serial, partly parallel linear reservoir structure, but this further structure is not within the scope of our study, as it isn't used.

Data

The hydrological systems chosen for analysis are the Jökulsá í Fljótsdal river basin in East-central Iceland (stream gauge vhm109), and the Blanda river basin in North-western Iceland (stream gauge vhm054). In the latter river, a hydropower plant is now under construction, but the former is under development for hydropower generation. The geographical location of the basins is shown in Fig. 3.

The river above the stream gauge vhm109, shown in Fig. 3, drains about 560 km^2, of which 133 km^2 are glaciated. These figures are quite reliable since water divides on the glacier, called Eyjabakkajökull, have been mapped by electromagnetic soundings (Björnsson 1982). The mean elevation of the watershed is 670 m a.s.l., maximum elevation 1,833 m and water gauge elevation 60 m a.s.l. The Eyjabakkajökull is a surging glacier with recorded surges in 1890, 1931 and 1972.

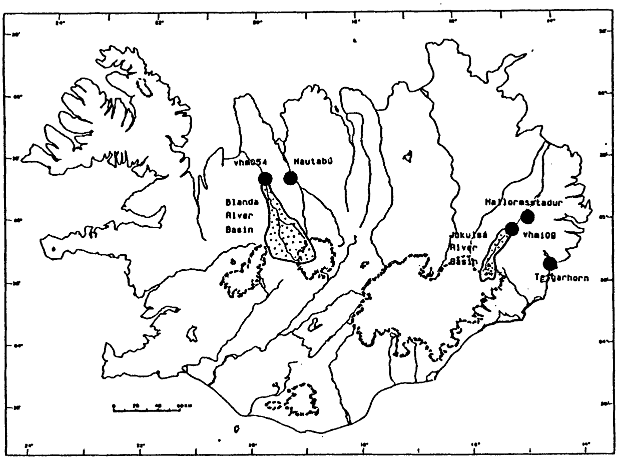

Fig. 3.
Geographic location of the basins under study.

The meteorological stations used are situated at Hallormsstadur, and Teigarhorn (see Fig. 3), at 60 and 18 m a.s.l. respectively. The NAM2 model is calibrated on a daily basis for the period Sep. 1 1962-Dec. 31 1983. Explained variance is 0.79, linear correlation coefficient r=0.89 and water balance +1.2% (Vatnaskil Consulting Engineers 1985). The intermediate input series used are made available by kind permission of Sigurdur Lárus Hólm at VCE.

The river above the stream gauge vhm054, shown in Fig. 3, drains 1,609 km^2, of which 198 km^2 are glaciated. These figures are reasonably reliable since water divides on the glacier, called Blöndujökull, have been mapped by electromagnetic soundings, except for areas where the glacier was heavily fractured (Björnsson personal communication). The mean elevation of the watershed is about 650 m a.s.l., maximum elevation about 1,700 m and water gauge elevation about 100 m a.s.l.

The meteorological station used is situated at Nautabú (see Fig. 3) at 115 m a.s.l. The NAM2 model for the gauge vhm054 is calibrated on a daily basis for the period Jan. 1 1975-Dec. 31 1983. Explained variance is 0.70, linear correlation coefficient r=0.83 and water balance -1.22% (Vatnaskil Consulting Engineers 1986). As in the case of Jökulsá, the intermediate input series are made available by kind permission of Sigurdur Lárus Hólm at VCE.

The basic data used for the analysis of the first case is composed of the discharge series, measured at stream gauge vhm109, and meterological series of air temperature and precipitation, measured at Hallormsstadur. The discharge series is composed of accumulated discharge of river Jökulsá for every two weeks, for the years 1974-1982 (water years, starting September 1st.). This volume is in units of 10^{6}m^3, so the units on the discharge series, P'_t, are 10^{6}m^3/2-weeks (10^{6}m^3/2-weeks = 0.83 m^3/s). The discharge series measured over the period 1962-1982, will be used for comparison with generated discharge values. The input series were the derived series of accumulated degree-days per two weeks, above 0, 2, 4, 6, and 8 degrees Celcius, identified as O'_t, T'_t, F'_t, X'_t, and Y'_t, respectively, since it has been shown by a earlier study (Snorrason 1985) that mean air temperature is inferior to the degree-days series as input. The intermediate input series derived from the NAM2-model were the lumped precipitation and/or snowmelt (less estimated evapotranspiration) series in million m^3 per two weeks calculated for each non-glaciated subarea, identified as B'_t, and the lumped precipitation and/or snowmelt/icemelt series in million m^3 per two weeks calculated for each glaciated subarea, identified as J'_t. Eventual negative values of the B'_t series were allowed to pass.

The basic data used for the analysis of the second case is composed of the discharge series, measured at stream gauge vhm054, and meteorological series of air temperature and precipitation, measured at Nautabú. The discharge series is constructed as in the first case. The input series are also formulated as in the first case, except adding precipitation series, N_t. The description of the input series derived from the NAM2 model is consequently also the same as above for Jökulsá.

Analysis

The objective of the analysis was to identify and estimate a model, relating the discharge of the two rivers to the most suitable of the input series and to compare the performance of the two sets of input data on the basis of the generated discharge records.

The analysis was initiated by estimating and removing from each series, seasonal variation in the mean. No attempt was made to account for, or to access the seasonal variation in higher moments of the processes involved, (see Gudmundsson 1975). The seasonal variation was estimated using a Fourier series of three periods (1, 1/2, and 1/3 years) as well as an overall mean.

The analysis of the data from the two river basins will now follow.

The Blanda Basin

The results of the modeling of the seasonal variations are shown in Table 1 for the Blanda basin. These consist of the mean, the original variance, the variance of the residual series and its percentage of the original variance, for each series.

An autocovariance analysis was performed on the residual series. All cases had small but significant autocorrelation structure. The autocorrelation function of the P'_t series of river Blanda suggests, that the residual discharge series is an autoregressive process of order one, ar(1). The analysis showed, furthermore, that the glacial runoff was a moving average process of order three, ma(3), the surface runoff was an ar(1) process, as well as all the degree days series. Theoretically, it is neccessary to remove this structure from the input series, i.e. pre-whiten them, to ensure a successful identification and estimation of the models, (see Granger and Newbold 1977), but its magnitude is in this case so small, that it will be ignored.

Table 1 – Blanda River Basin; Summary of Deterministic Components

Series	Mean	Var(.)	Residual	Var(.) %
P'_t	52.5	4189	674	16.1
O'_t	56.3	5509	439	8.0
T'_t	39.0	3099	313	10.1
F'_t	24.9	1533	212	13.8
X'_t	14.2	626	126	20.0
Y'_t	6.97	195	58.5	30.0
J'_t	12.1	533	126	23.6
B'_t	38.3	5852	2688	45.9

A covariance analysis was used to identify the relationship between the series. A significant correlation existed between the residual series of discharge and the residual input series. No cross correlation was, however, observed between the NAM2-generated inputs. The covariance structure suggests the following model for both cases

$$P_t = \frac{\omega_0}{1-\delta_1 B} X_{1t} + \frac{\omega_1}{1-\delta_2 B} X_{2t} + \frac{1}{1-\theta_1 B} N_t \qquad (7)$$

in which N_t is a white noise series and X_{it}, $i = 1,2$ are any of the input series. Assuming that the coefficients δ_1, δ_2 and θ_1 are equal, Eq. (7) can be written as a lagged regression equation

$$P_t \equiv \beta+\delta_1 P_{t-1}+\omega_0 X_{1t}+\omega_1 X_{2t}+ N_t \qquad (8)$$

This equation was then estimated by LSE routine for both sets of input series. The results are shown in Table 2.

These results show a slight advantage for the model with the NAM2-generated inputs over the model with the most suitable degree series as input. Note also that the autoregressive term is not significant for the model with the NAM2-generated inputs. The discharge series was extended using both sets of input series. The generating equation had the form

$$P_t \equiv \beta+\delta_1 P_{t-1}+\omega_0 X_{1t}+\omega_1 X_{2t} \qquad (9)$$

where X_{it}, $i = 1,2$ are any of the residual input series. The initial values were set equal to zero. The generated series were compared to the measured one, both quantitatively as well as qualitatively. The average yearly volume of flows, both measured and generated on the basis of the two sets of input series are compared in Fig. 4. It is apparent that both models follow the basic trends in measured discharge, but fail slightly for extreme years. The generated discharges outside of the calibration period are quite similar for the period 1950-1973.

In Fig. 5, four years of measured and generated values from the calibration

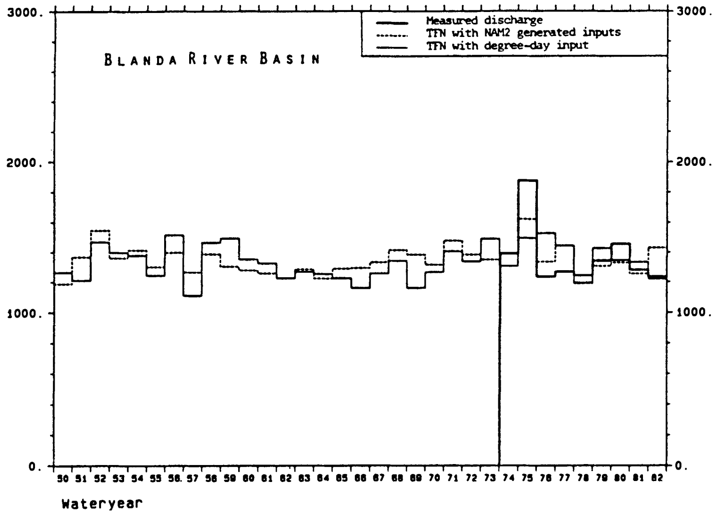

Fig. 4.
Yearly volumes of measured and generated discharge, 10^6 m^3/year.

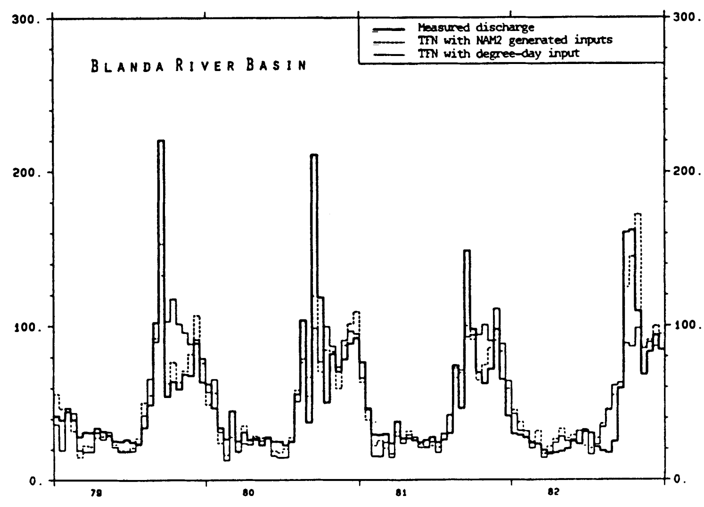

Fig. 5.
A comparison of measured and generated discharge within the calibration period, 10^6 m^3/2-weeks.

Table 2 – Blanda River Basin; Summary of residual models.

Model 1:	Model 2:
$P_t = 0.16 * P_{t-1} + 0.72 * T_t - 1.$	$P_t = 0.28 * B_t + 0.73 * J_t - 1.$
(0.06) (0.08) (1.)	(0.02) (0.10) (1.)
Explained variance 27% of the variance of P_t	Explained variance 48% of the variance of P_t

period are shown. The generated discharge seems to follow the measured one quite well, and the series generated with the NAM2-generated inputs is clearly superiour to the one generated with the degree days series as input. The difference is especially clear for the period between the peak spring melting and peak glacial melting in the years 1979 and 1981. Also, there is considerable difference in performance in the beginning of the melting season in 1982. Both models fail to generate extreme springfloods.

The Jökulsá í Fljótsdal Basin

The analysis of the data from the Jökulsá basin followed the same course as the analysis for the Blanda basin. The results of the modeling of the seasonal variations are shown in Table 3 for the Jökulsá basin. These consist of the mean, the original variance, the variance of the residual series and its percentage of the original variance, for each series.

An autocovariance analysis was performed on the residual series. The autocorrelation function of the P_t series of river Jökulsá suggests, that the residual discharge series is an autoregressive process of order one, ar(1). The analysis showed, furthermore, that the glacial runoff was a moving average process of order one, ma(1), the surface runoff was an ma(1) process. The residual series of degree days and precipitation did not show a significant autocorrelation structure.

A covariance analysis was used to identify the relationship between the series. A significant correlation existed between the residual series of discharge and the residual series. The covariance structure suggests similar models as in the Blanda case, and here, also, no cross correlations were observed between the inputs to each model. The results are shown in Table 4.

These results show a slight advantage for the model with the NAM2-generated inputs over the model with the most suitable degree series as input. All relations are more significant for the Jökulsá basin, than for the Blanda basin. Also, the autocorrelation term here is more significant. The relationship between precipitation and discharge is rather small, but still significant.

The discharge series was extended using both sets of input series in the same

Table 3 – Jökulsá River Basin; Summary of Deterministic Components

Series	Mean	Var(.)	Residual	Var(.) %
P'_t	34.4	2554	347	13.6
N'_t	26.8	1541	710	46.1
O'_t	65.9	7002	529	7.5
T'_t	47.0	4112	381	9.2
F'_t	31.3	2163	261	12.1
X'_t	18.8	970	169	17.4
Y'_t	9.76	347	96.7	27.9
J'_t	14.7	788	169	21.5
B'_t	18.5	1307	550	42.1

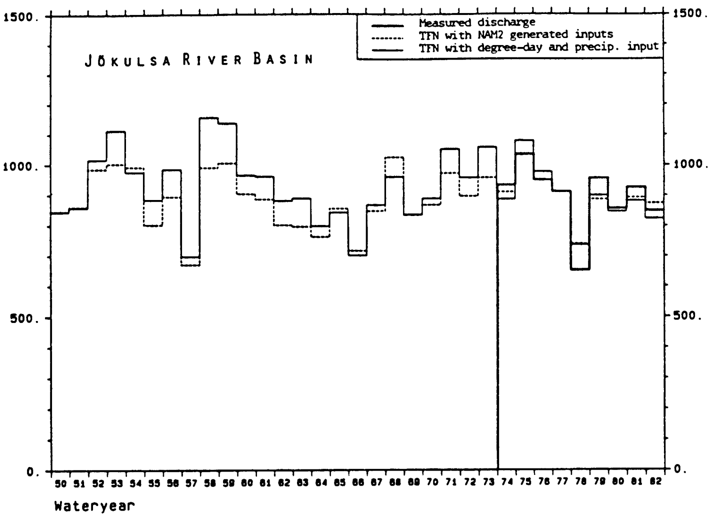

Fig. 6.
Yearly volumes of measured and generated discharge, 10^6 m^3/year.

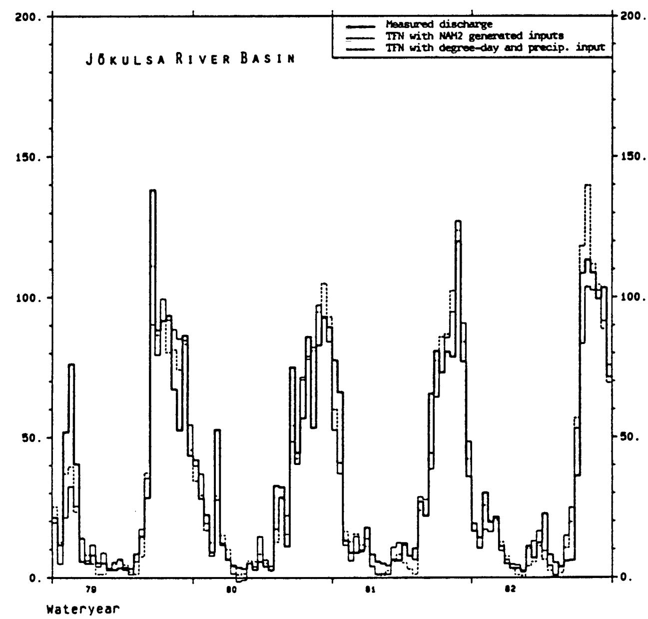

Fig. 7.
A comparison of measured and generated discharge within the calibration period, 10^6 m^3/2-weeks.

Table 4 – Jökulsá River Basin; Summary of residual models.

Model 1:	Model 2:
$P_t = 0.36 * P_{t-1} + 0.64 * F_t + 0.07 * N_t + 0.6$	$P_t = 0.27 * P_{t-1} + 0.30 * B_t + 0.70 * J_r - 0.3$
(0.05) (0.06) (0.03) (0.8)	(0.04) (0.03) (0.06) (0.7)
Explained variance 59% of the variance of P_t	Explained variance 66% of the variance of P_t

manner as before. The generated series were compared to the measured one, both quantitatively as well as qualitatively. Values were compared both for the calibration period, i.e. the model fit, as well as outside of it. The average yearly volume of flows, both measured and generated on the basis of the two sets of input series are compared in Fig. 6. It is apparent that both models follow the measured discharge quite well. The figure also shows that there is systematic difference in the generated series for the period 1950 to 1963, but they are comparable for the period 1964 to 1973.

In Fig. 7, four years of measured and generated values from the calibration period are shown. The measured and generated discharge compares in general well, and it is difficult to distinguish the performance of the two models.

Conclusions

The relationship of meteorological variables and river flow from two glaciated basins in Iceland was investigated by the use of simplified linear transfer function-noise models. Two sets of input series were compared, one consisted of degree days and precipitation series, and the second set consisted of series generated by a distributed, deterministic model, the NAM2 model. The analysis showed that by transforming the meteorological series to a suitable input series, it was possible to overcome to a great extent, the non-linearities and non-stationarities in the relationship between temperature, precipitation and discharge. The results showed that the generated series seemed to contain both the general trends, as well as extreme values in good agreement with measured discharge, with the series generated on the basis of the NAM2-generated inputs considerably better for one case and marginally better for the second case.

References

Anselmo, V., and Ubertini, L. (1979) Transfer function-noise model applied to flow forecasting, *Hydrological Science Bulletin, Vol. 24,* No. 3, p. 353-360.

Baracos, P. C., Hiepel, K. W. and McLeod, A. I. (1981) Modeling hydrologic time series from the artic, *Water Resources Bulletin, American Water Resources Association, Vol. 17,* No. 3, p. 414-422.

Björnsson, H. (1982) Nidurstödur íssjármælinga á Eyjabakkajökli. Mimeographed report in Icelandic, Rafmagnsveitur ríkisins, The State Electric Power Works, Raunvísindastofnun háskólans, Science Institute, University of Iceland, No. Rarik 82504/RH-82-02.

Box, G. E. P., and Jenkins, G. M. (1970) *Time Series Analysis: Forecasting and Control,* Holden-Day, San Fransisco, USA.

Chow, V. T. (1973) Analysis of Multiple-Input Stochastic Hydrological Systems, Water Resources Center, Univ. of Illinois, USA, Research Report No. 67, 66 p.

Chow, V. T. (1978) Stochastic modeling of watershed systems. In: *Advances in Hydrosciences.* Ed. V. T. Chow, Vol. 11, Academic Press, New York, USA, p. 1-93.

Clark, R. T. (1974) The representation of a short period of experimental catchment data by a linear stochastic difference equation, Proceedings of the Warsaw Symposium on Mathematical Models in Hydrology, July 1971, International Association of Hydrological Sciences, Publication No. 100, IAHS-UNESCO-WMO, Vol. 1, p. 3-16.

Einarsson, K. (1984) Nedbør/afstrømningsmodeller i Island. In: Killingtveit, Å., (Ed.), Anvendelse af hydrologiske modeller i vannbruksplanlegging, NHP-rapport nr. 6, p. 35-41, KOHYNO, Oslo 1984, In Danish.

Gottlieb, L. (1980) Development and Applications of a Runoff Model for Snowcovered and Glacierized Basins, *Nordic Hydrology, Vol. 11,* p. 255-272.

Granger, C. W. J., and Newbold, P. (1974) Spurious regressions in econometrics, *Journal of Econometrics, Vol. 2,* p. 111-120.

Granger, C. W. J., and Newbold P. (1977) *Forecasting Economic Time Series,* Academic Press, New York, USA, 1977, 333 p.

Gudmundsson, G. (1970) Short term variations of glacier-fed river, *Tellus XXII, No. 3,* p. 341-353.

Gudmundsson, G. (1975) Seasonal variations and stationarity, *Nordic Hydrology, Vol. 6,* p. 137-144.

Hipel, K. (1985) Stochastic research in multivariate analysis. In: Multivariate Analysis of Hydrologic Processes, Proc. Fourth International Hydrology Symposium, Fort Collins, Colo., USA, p. 2-50.

Hsu, C. T., and Adamowski, K. (1981) Multivariate hydrologic time series analysis, *Advances in Water Resources, Vol. 4,* No. 2, p. 83-95.

Laramie, R. L., and Schaake, J. C., Jr. (1972) Simulation of the continuous snowmelt process, MIT, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Report No. 143, Cambridge, Ma., USA.

Nielsen, S. A., and Hansen, E. (1973) Numerical simulation of the rainfall-runoff process on a daily basis, *Nordic Hydrology, Vol. 4,* p. 171-190.

Snorrason, A. (1983) Analysis of Multivariate Stochastic Hydrological Systems using Transfer Function-Noise Models, Ph. D. Thesis, University of Illinois, Urbana, Illinois, unpublished.

Snorrason, A. (1985) Analysis of river flow data from glaciated basin in Iceland. In: Multivariate Analysis of Hydrologic Processes, Proc. Fourth International Hydrology Symposium, Fort Collins, Colo., USA, p. 651-663.

Snorrason, A., Newbold, P. and Maxwell, W. H. C. (1984) Multiple input transfer function-noise modeling of river flow, In: *Frontiers in Hydrology.* Ed. W. H. C. Maxwell and L. R. Beard, Water Resources Publications, Littleton, Colorado, p. 111-126.

Vatnaskil Consulting Engineers (1985) Reiknad rennsli Jökulsár í Fljótsdal við Hól 1941-1983. Mimeographed report in Icelandic. Orkustofnun, National Energy Authority, Hydro Power Division, No. OS-85009/VOD-04 B.

Vatnaskil Consulting Engineers (1986) Blönduvirkjun. Rennslislíkan. Mimeographed report in Icelandic. Landsvirkjun, National Power Company, No. 86.01.

Received: 1 October, 1986

Address: Orkustofnun, National Energy Authority,
Grensásv. 9,
108 Reykjavik, Iceland.