2006

A hybrid, Markov chain-based model for daily streamflow generation at multiple catchment sites

Jozsef Szilagyi
Budapest University of Technology and Economics

Gabor Balint
National Hydrological Forecasting Service of Hungary

Andras Csik
National Hydrological Forecasting Service of Hungary

Hybrid, Markov Chain–Based Model for Daily Streamflow Generation at Multiple Catchment Sites

Article *in* Journal of Hydrologic Engineering · May 2006

DOI: 10.1061/(ASCE)1084-0699(2006)11:3(245)

3 authors, including:

Joe Szilagyi
Budapest University of Technology and Economics
104 PUBLICATIONS **1,845** CITATIONS

SEE PROFILE

Gábor Bálint
Vituki Environmental Protection And Water Management Research Institute
19 PUBLICATIONS **190** CITATIONS

SEE PROFILE

Hybrid, Markov Chain-Based Model for Daily Streamflow Generation at Multiple Catchment Sites

Jozsef Szilagyi[1]; Gabor Balint[2]; and Andras Csik[3]

Abstract: A hybrid, seasonal, Markov chain-based model is formulated for daily streamflow generation at multiple sites of a watershed. Diurnal increments of the rising limb of the main channel hydrograph were stochastically generated using fitted, seasonally varying distributions in combination with an additive noise term, the standard deviation of which depended linearly on the actual value of the generated increment. Increments of the ascension hydrograph values at the tributary sites were related by third- or second-order polynomials to the main channel ones, together with an additive noise term, the standard deviation of which depended nonlinearly on the main channel's actual increment value. The recession flow rates of the tributaries, as well as of the main channel, were allowed to decay deterministically in a nonlinear way. The model-generated daily values retain the short-term characteristics of the original measured time series (i.e., the general shape of the hydrograph) as well as the probability distributions and basic long-term statistics (mean, variance, skewness, autocorrelation structure, and zero-lag cross correlations) of the measured values. Probability distributions of the annual maxima, means, and minima of the measured daily values were also well replicated.

DOI: 10.1061/(ASCE)1084-0699(2006)11:3(245)

CE Database subject headings: Streamflow; Markov chains; Hydrographs; Hybrid methods; Catchments.

Introduction

Hydrologists involved with operational stream forecasting and flood control may be interested in hypothetical but quite possible scenarios of flood events. This may help them prepare for events that have not yet been observed in the past for which measurements are available but nonetheless can be expected in the future. While statistical analyses of, e.g., annual maxima, may offer information on the return period of floods with different magnitudes, they do not provide information on the possible time sequence of the expected flood event. Such information may encompass duration of different water levels during a flood, the speed at which stream levels may rise or the flood may recede, and all of which potentially influence how flood protection works ought to be planned and built as well as flood defense activities organized.

Traditional autoregressive models (Quimpo 1968; Payne et al. 1969; McGinnis and Sammons 1970) are generally inadequate at capturing the typically asymmetric shape of the hydrograph observable in daily streamflow series (Sharma et al. 1997). Shot noise models, originally developed in electrical engineering, were introduced to the hydrologic literature in the 1970s (Bernier 1970; Weis 1973, 1977; Cowpertwait and O'Connell 1992; Murrone et al. 1997) and were specifically formulated for working with daily flow values, having become capable of producing—besides basic long-term statistics such as mean, variance, and serial correlations—realistic-looking hydrographs at a single location. The same can be said about the types of models where input pulses are transformed into flow values using a transfer function approach (Treiber and Plate 1977; Kottegoda and Horder 1980) or where daily rainfall series are generated and converted to streamflow series using conceptual models (Kelman 1980; Koch 1985; Bierkens and Puente 1990). Because adequate information of precipitation over the watershed may often be lacking, and even when it is available, little may be known of the effective precipitation that actually forms the flood event; stochastic techniques that do not require information on precipitation may be practical to pursue.

Xu et al. (2001, 2003) made attempts to extend synthetic streamflow generation for a single site to multiple sites with possibly high cross correlations of the daily values among these sites. Such cases may be of importance when stochastically generated flow values at several upstream tributary sites are subsequently routed with a flow routing model that can account for anticipated changes in channel flow of the tributaries and/or of the main stem of the river due to a change in channel conditions downstream of the data generation sites. Such changes may result from reservoir construction, altered operation schedule of existing reservoirs, channelization effects, and flood protection works, just to name a few. One may be interested in how these disturbed channel conditions, both on the main channel and on the tributaries as well, may alter the behavior of the main channel flow downstream of the confluence with the tributaries.

[1]Associate Professor, Dept. of Hydraulic and Water Resources Engineering, Budapest Univ. of Technology and Economics, Muegyetem Rkpt. 1-4, H-1111, Budapest, Hungary (corresponding author). E-mail: jszilagyil@unl.edu; and, Conservation and Survey Division, Univ. of Nebraska-Lincoln, Lincoln, NE 68588-0517.

[2]Senior Research Associate, National Hydrological Forecasting Service of Hungary, VITUKI, H-1095 Kvassay J. ut 1, Budapest, Hungary.

[3]Junior Research Associate, National Hydrological Forecasting Service of Hungary, VITUKI, H-1095 Kvassay J. ut 1, Budapest, Hungary.

Note. Discussion open until October 1, 2006. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on May 27, 2004; approved on July 18, 2005. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 11, No. 3, May 1, 2006. ©ASCE, ISSN 1084-0699/2006/3-245–256/$25.00.

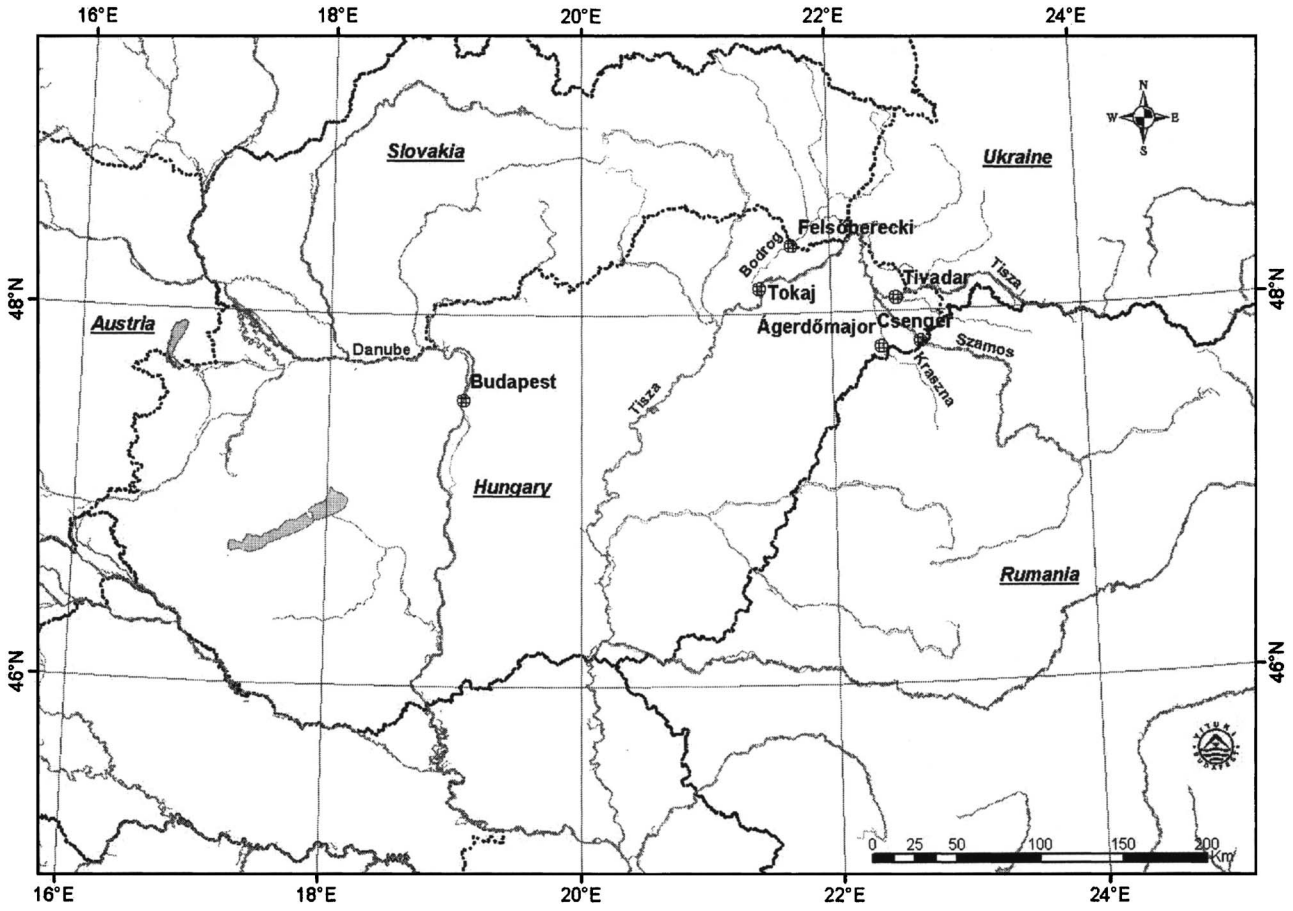

Fig. 1. Location of gauging stations and rivers in study

Our multivariate, seasonal streamflow-generation algorithm detailed below uses components of the shot noise models in a Markov chain based approach, together with a conceptual framework describing flow recession without the need for information on precipitation. It is built around the concept of conditional heteroscedasticity originally established in the ARCH models (Engle 1982) of time series analysis, when it is assumed that the noise term is not independent of the process to be modeled, nor it is identically distributed.

Model Formulation and Application

The model works with daily streamflow data from which a time series of diurnal increments can be obtained by differencing the original series. These increments define a two-state Markov chain for perennial streams. State one is observed when the increment is positive (termed as the "wet" state), and state two (termed as "dry"), otherwise. The two states result in four different state transitions: wet–wet (P_{ww}), wet–dry (P_{wd}), dry–wet (P_{dw}), and dry–dry (P_{dd}). The state transition probabilities can be estimated from the observed data as

$$P_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}, \quad i,j = w,d \tag{1}$$

where n_{ij}=number of observed transitions from state i to j. The state transition probabilities typically vary with seasons. Often these transitions are written on a monthly basis (e.g., Xu et al. 2001, 2003), which result in very similar values among neighboring months (thus raising the question whether they are statistically different or not), plus in a large number of model parameters. From the viewpoint of parameter parsimony, a seasonal resolution should suffice in most cases, as was adapted here.

The Tisza River is the major tributary of the Danube within Hungary (Fig. 1). Besides the gaging station of Tivadar on the Tisza River, three additional sites on tributaries of the Tisza were included in the study. Table 1 lists the corresponding drainage areas and daily mean flow rates measured at the gaging stations. Flow data were provided by the Hungarian Hydrological Forecasting Service of the Institute of Water Resources Research (VITUKI).

50 years (from the period of 1951 to 2000) of daily instantaneous flow-rate values were employed for all four gaging stations for statistical inference. Table 2 displays the estimated state transition probabilities at Tivadar on a seasonal basis. It shows that a wet-to-wet transition has the highest likelihood in spring, which

Table 1. Drainage Area and Daily Mean (Base Period of 1985–1994) Flow Rate of Gaging Stations Included in Study

	Drainage area (km^2)	Daily mean flow rate ($m^3 s^{-1}$)
Tivadar (Tisza River)	12,540	233
Csenger (Szamos River)	15,283	102
Felsoberecki (Bodrog River)	12,886	104
Agerdomajor (Kraszna River)	1,974	4.81

Table 2. Estimated State Transition Probabilities (%) at Tivadar

	P_{dd}	P_{dw} $(=1-P_{dd})$	P_{wd}	P_{ww} $(=1-P_{wd})$
Winter	80.44	19.56	37.9	62.1
Spring	79.71	20.29	37.55	62.45
Summer	76.15	23.85	51.34	48.66
Fall	80.51	19.49	48.62	51.38

comes from two sources: (1) it is the season of most abundant precipitation in Hungary; and (2) it is the time of year when melting snow in the Carpathian Mountains feeds the streams, occasionally (especially when combined with rain) causing major flooding in the region.

Ascension Limb of Hydrograph

Main Channel

Positive diurnal increments (or wet states) designate the ascension limb of the hydrograph. Sargent (1979) and Aksoy (2003) recommended a two-parameter gamma distribution for these increments. For the Tisza River at Tivadar (Fig. 1), we found that the Weibull distribution better fits the observed data taken from the period of 1951–2000. Figs. 2 and 3 display the seasonal *Q-Q* plots of the observed, as well as the hypothetical, two-parameter gamma and Weibull distributions, respectively, both parameterized with the help of the observed data using the Matlab functions "gamfit" and "weibfit". In the Weibull distribution case, the points align closer to the perfect fit straight line, suggesting that diurnal increments can be probably better described by the Weibull, rather than by a two-parameter gamma distribution, although none of the eight cases displayed pass the Kolmogorov–Smirnov test at the typical 5% level.

During Monte Carlo simulation of these increments, dQ $[L^3\,T^{-1}]$, for wet states the computer uses the fitted Weibull distributions (on a seasonal basis) for its random number generation. The so-obtained values for the main channel, which is the Tisza River now, are subsequently disturbed with an additive noise term, $W[L^3\,T^{-1}]$, taken from a normal distribution of zero mean (m). The noise, however, is not identically distributed, because its standard deviation (σ) is conditioned on the Weibull-distributed random number, dQ_{gen}, to be disturbed

$$W(m,\sigma) = W(0, a \cdot dQ_{gen}^{b}) \tag{2}$$

where a $[L^{3(1-b)}\,T^{(b-1)}]$=scale-coefficient; and b [-]=exponent. From the generated W values, those that are negative and have larger magnitudes than the corresponding dQ_{gen} values are discarded and replaced by zero. This results in noise values that follow a positively skewed distribution (whose mean is no longer zero). Fig. 4 displays the *Q-Q* plots of the model-obtained positive diurnal increments versus the observed increments on a seasonal basis. The scale coefficient, a, and the exponent, b, are model parameters that were calibrated using trial and error upon visually inspecting the resulting *Q-Q* plots of each prescribed (a,b) combinations. Calibration leads to values of 1.1, 1.2, 1, 0.7 for the four seasons, starting with winter, for a, and to 1 for b. This so-called conditional heteroscedasticity assures that the Monte Carlo generated diurnal increments this way will better approximate the observed distribution than a simple Weibull-distributed number generator (compare Figs. 3 and 4).

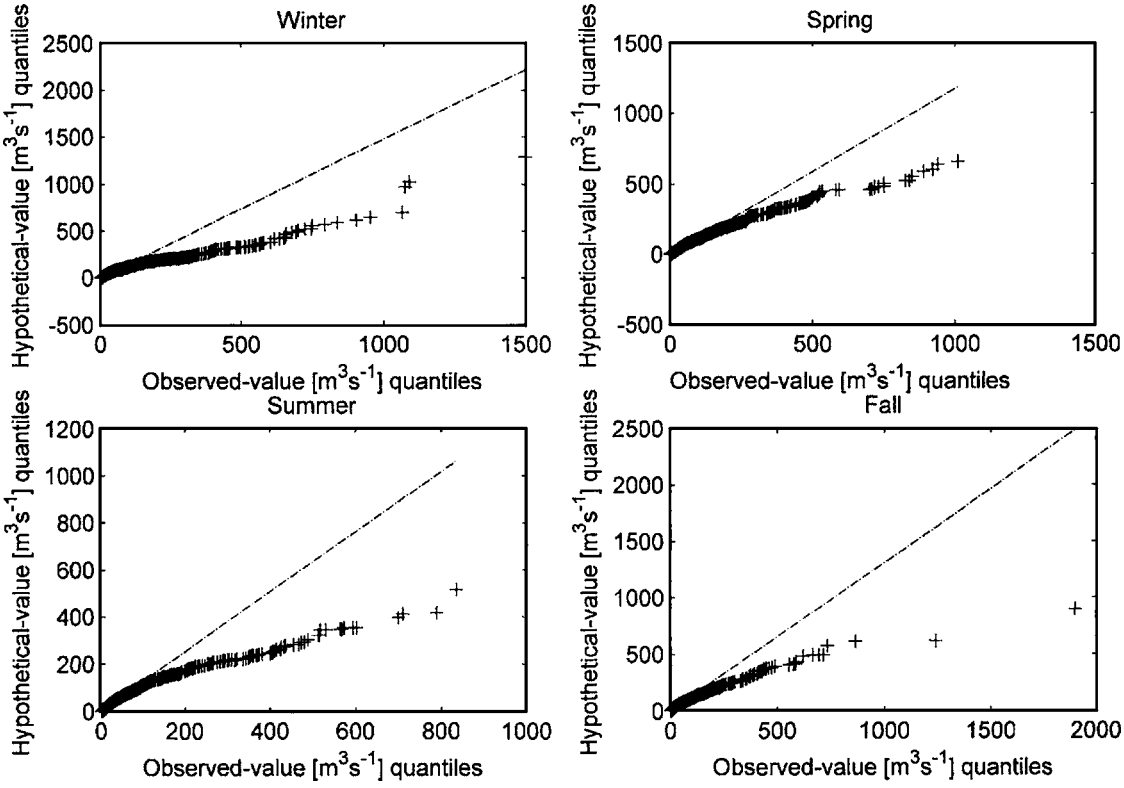

Fig. 2. Seasonal *Q-Q* plots of observed and two-parameter gamma-distributed positive diurnal increment values of Tisza River at Tivadar. Theoretical distribution was parameterized using observed values.

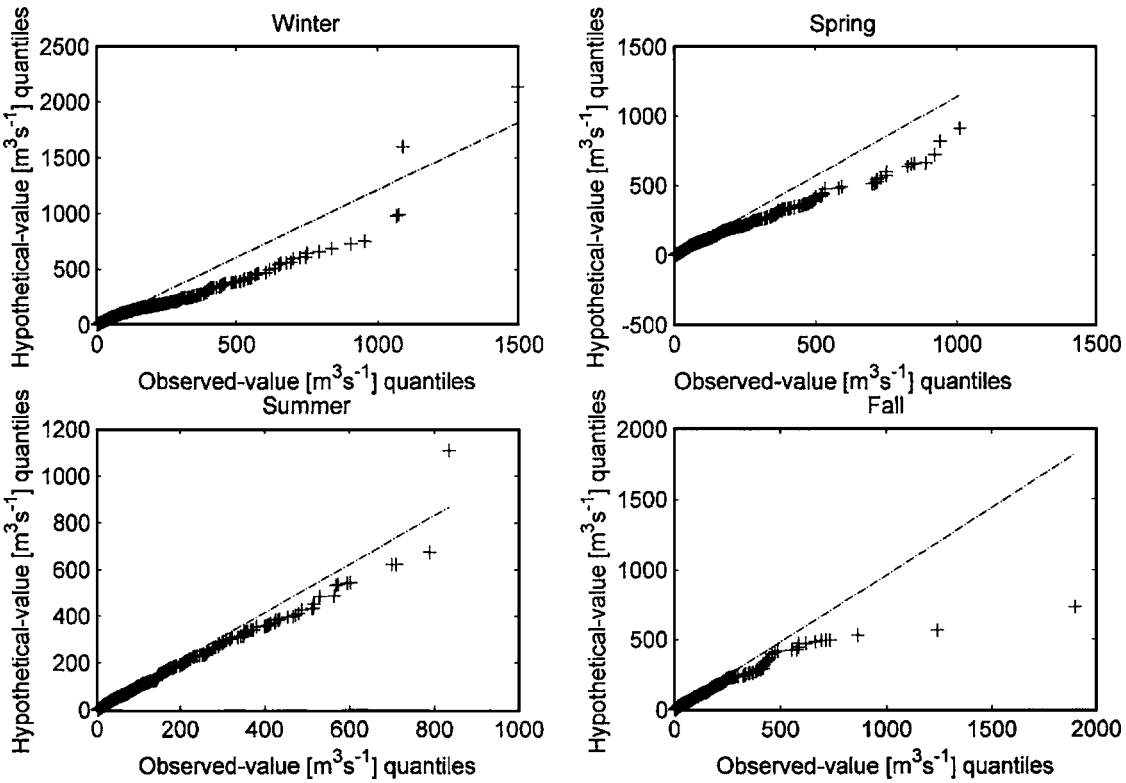

Fig. 3. Seasonal *Q-Q* plots of observed and Weibull-distributed positive diurnal increment values of Tisza River at Tivadar. Theoretical distribution was parameterized using observed values.

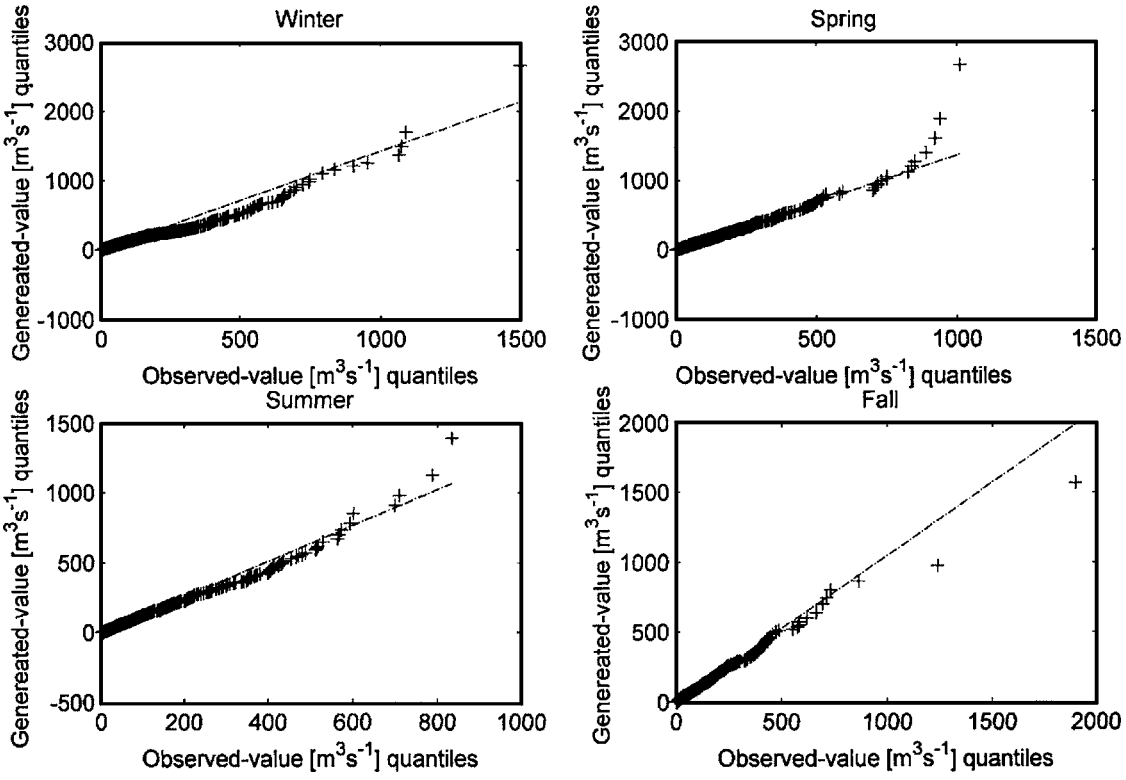

Fig. 4. Seasonal *Q-Q* plots of observed and model-generated positive diurnal increment values of Tisza River at Tivadar

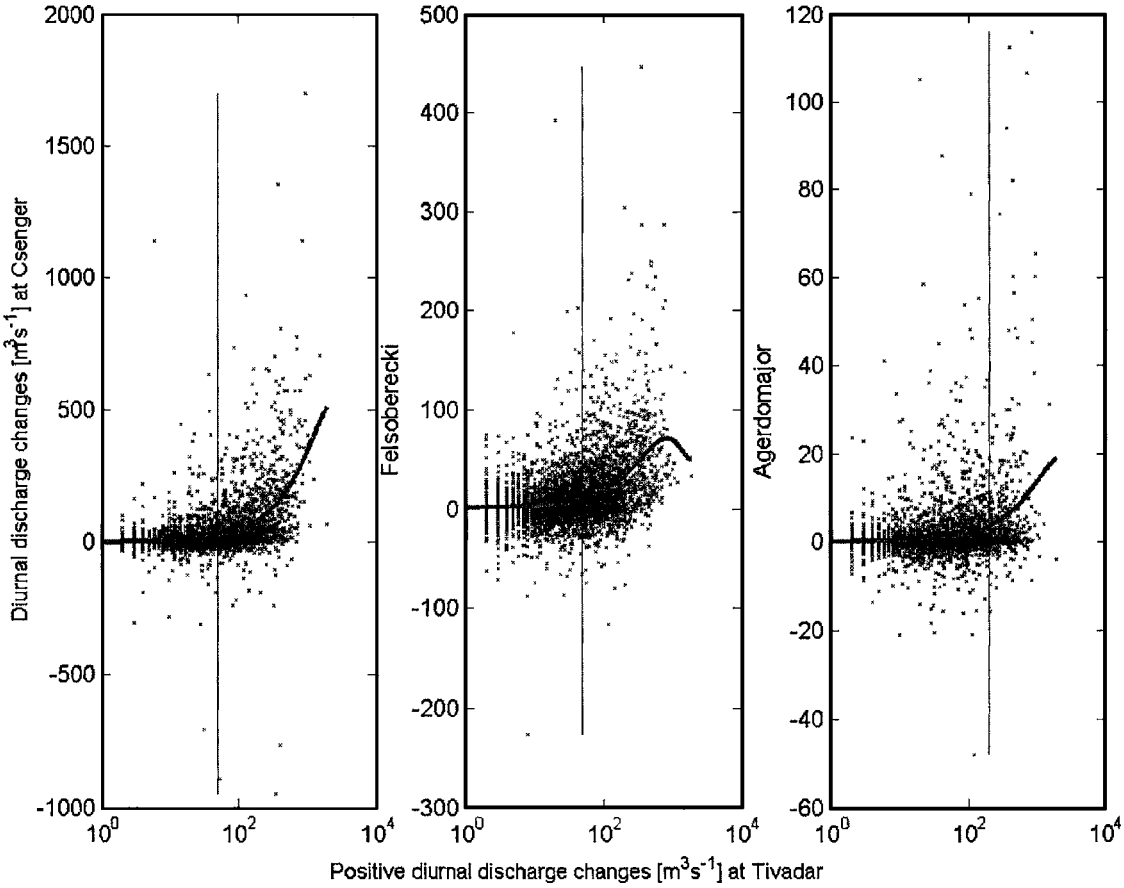

Fig. 5. Diurnal tributary versus main channel discharge changes for wet periods of main channel. Best fit second- or third-order (Felsoberecki) polynomials are also shown together with their minimum arguments [50, 50, 200 ($m^3 s^{-1}$)] applied for tributary data generation.

Another alternative could be the use of an additive noise term that follows a Weibull rather than a normal distribution. In that case, the prescribed mean and standard deviation of this noise distribution (the letter providing one with a rough idea of the spread of the distribution) could be converted to the parameters of the Weibull distribution. However, the Weibull distribution is a monotonic function (i.e., no peak in the distribution function's shape) for a wide range of the parameters, while the distribution of the diurnal increments is not. Consequently, it is more convenient to employ a normal distribution instead, and make it skewed by specifying the lower limit with each value of the increment to be disturbed.

Once the positive increment values have been generated for a wet spell, they are ranked in an increasing order to make sure that the larger increments are closer to the peak of the hydrograph. This recreates the general shape and ensures preservation of the correlation structure of the ascension limb of the hydrograph (Aksoy 2003).

Table 3. Coefficients of Third- and Second-Order Polynomials between Main Channel (Tisza River) and Tributary Increments during Wet Spells of Tisza

Tivadar	Third-order term	Second-order term	First-order term	Constant
Csenger	—	$-8.60 \cdot 10^{-5}$	0.4323	−0.806
Felsoberecki	$4.29 \cdot 10^{-8}$	$-1.72 \cdot 10^{-4}$	0.1918	0.5555
Agerdomajor	—	$-4.14 \cdot 10^{-6}$	0.0177	0.1363

Tributaries

For small to medium-sized watersheds [i.e., the drainage area is less than 30,000 km^2 as defined by the U.S. Geological Survey's Hydrologic Unit Code system for subbasins (DeBarry 2004)] tributary flow values are typically correlated with the main channel values (especially when the corresponding drainage areas are of the same order as in our case); therefore, one may want to avoid generation of positive increment values for the tributaries independently of the main channel ones. One way of linking tributary increments to the main channel state could be achieved by conditioning the state transition probabilities of the tributaries to that of the main channel, since for correlated flow series, the probability of a wet-to-wet transition is higher for the tributary when the main channel is in a wet state too. Unfortunately, such conditioning of the state transition probabilities did not meet expectations in our study; the cross-correlation value between the (measured) main channel and simulated tributary flow rates remained much lower than observed. As an alternative, the following was performed.

Table 4. Calibrated Parameter Values Applied in Model

	a ($m^{3(1-b)} s^{(b-1)}$)	b (-)	d ($m^3 s^{-1}$)	f (%)	g (-)	h (-)	k'_{max} (-)	k'_{min} (-)
Csenger	0.5	0.95	50	20	0.2	0.1	0.62	0.01
Felsoberecki	0.35	0.9	50	20	0.2	0.05	0.33	0.045
Agerdomajor	0.04	1	200	20	0.3	0.05	0.4	0.1

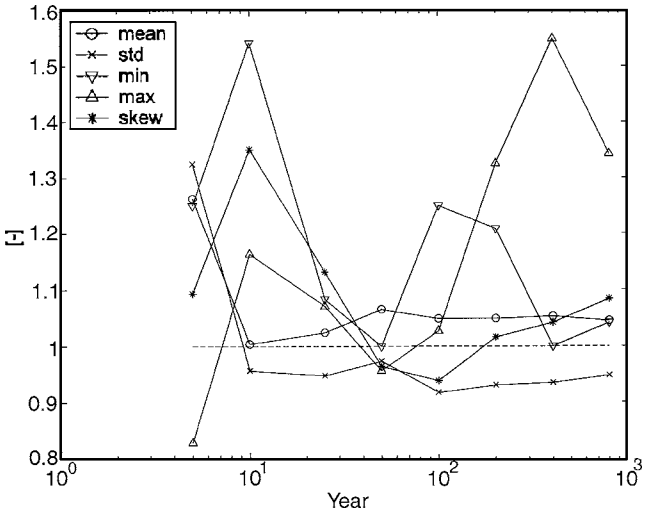

Fig. 6. Ratio of Monte Carlo generated and 50 years of observed daily flow statistics (mean, standard deviation, minimum, maximum, and skewness) as function of simulation length

Diurnal increments of the ascension hydrograph at the main channel were related by third- or second-order polynomials to corresponding increments at the tributary sites. The polynomial regression-derived tributary increment values were again disturbed by an additive noise term in the form of Eq. (2). As before, Eq. (2) includes dQ_{gen} of the main channel and not of the tributary. Alternatively, one may choose to apply the regression-derived tributary value in place of dQ_{gen}. In either case, the coefficients, a and b, must be calibrated anew for each tributary site.

Two new parameters must be introduced into the model at this point. The first, d ($m^3 s^{-1}$), is a minimum discharge value the positive diurnal discharge change in the main channel must exceed before the above described tributary discharge generation starts. This is necessary to avoid tributary flooding for every tiny increase of discharge in the main channel. The other parameter, $f(\%)$, ensures that the tributary-generated value on any given day would not fall below a certain percentage (given by f) of the previous day's value, since the speed of recession certainly has a natural limit. Whenever the tributary value would be lower than this percentage of the previous day's discharge (as a result of the additive noise term), the value of the actual noise term is changed to zero. See Fig. 5 for the polynomial fit with the d value marked. When employing the polynomials for arbitrary data generation, it may be prudent to not apply polynomials with arguments larger than the observed maximum diurnal change of the main channel. Instead, it is recommended that the polynomial value of the largest observed argument be maintained with such possible values.

Table 3 lists the coefficients of the third- and second-order polynomials applied in the study, while Table 4 displays the calibrated a, b, d, and f parameter values for each tributary site. Note that while the second- and third-order coefficients in Table 3 have small values, they are not negligible because the $(dQ_{gen})^3$ and $(dQ_{gen})^2$ terms are typically on the order of 10^6–10^4, respectively. Calibration of the tributary a, b, d, and f values is explained below.

Recession Curve

Observed flow recession in the channel is generally of a nonlinear nature (Aksoy et al. 2001). This is so because the upper part of the recession limb of the hydrograph is influenced by channel

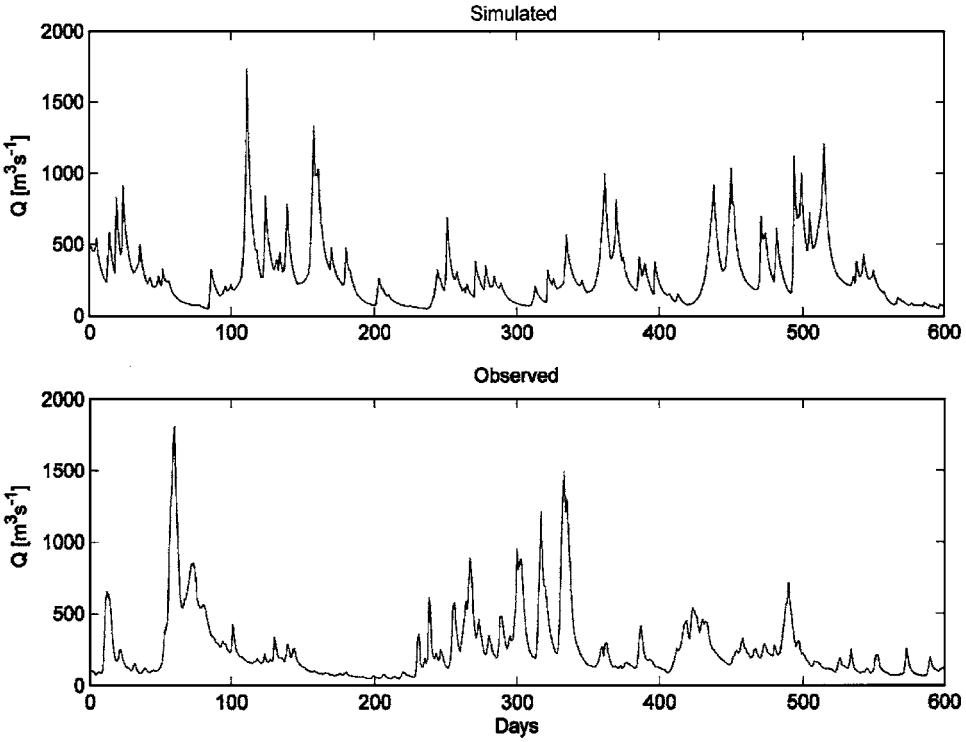

Fig. 7. Sample observed and Monte Carlo generated time series of daily flow rates of Tisza River at Tivadar

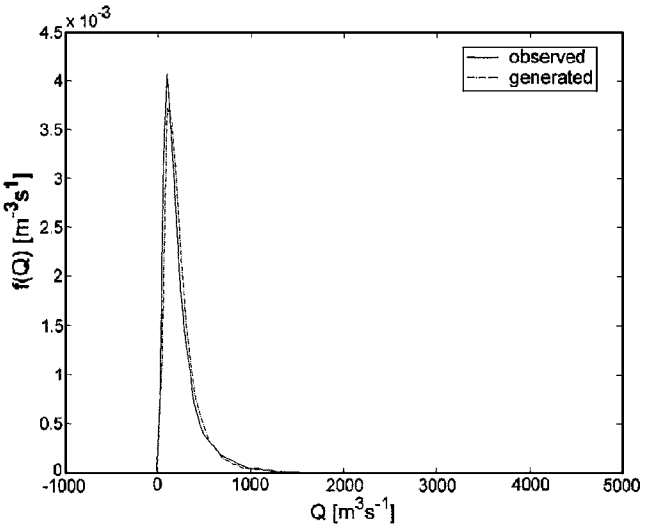

Fig. 8. Estimated distribution functions of 50 years of observed and simulated daily flow rates of Tisza River at Tivadar

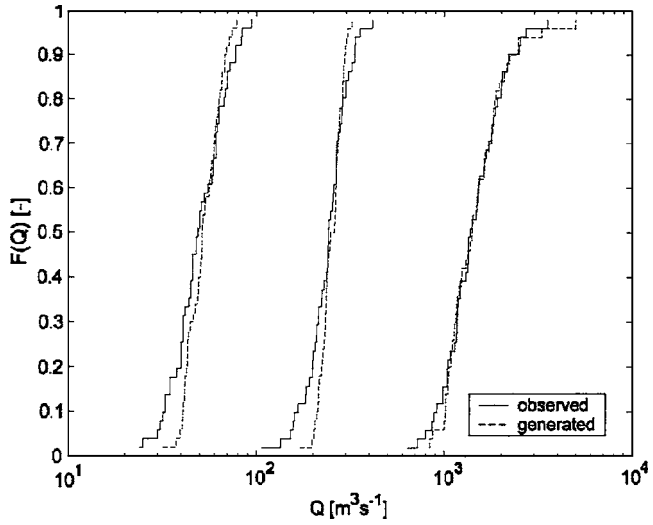

Fig. 9. Empirical cumulative distribution functions of 50 years of observed and simulated annual maxima, means, and minima of daily flow rates of Tisza River at Tivadar

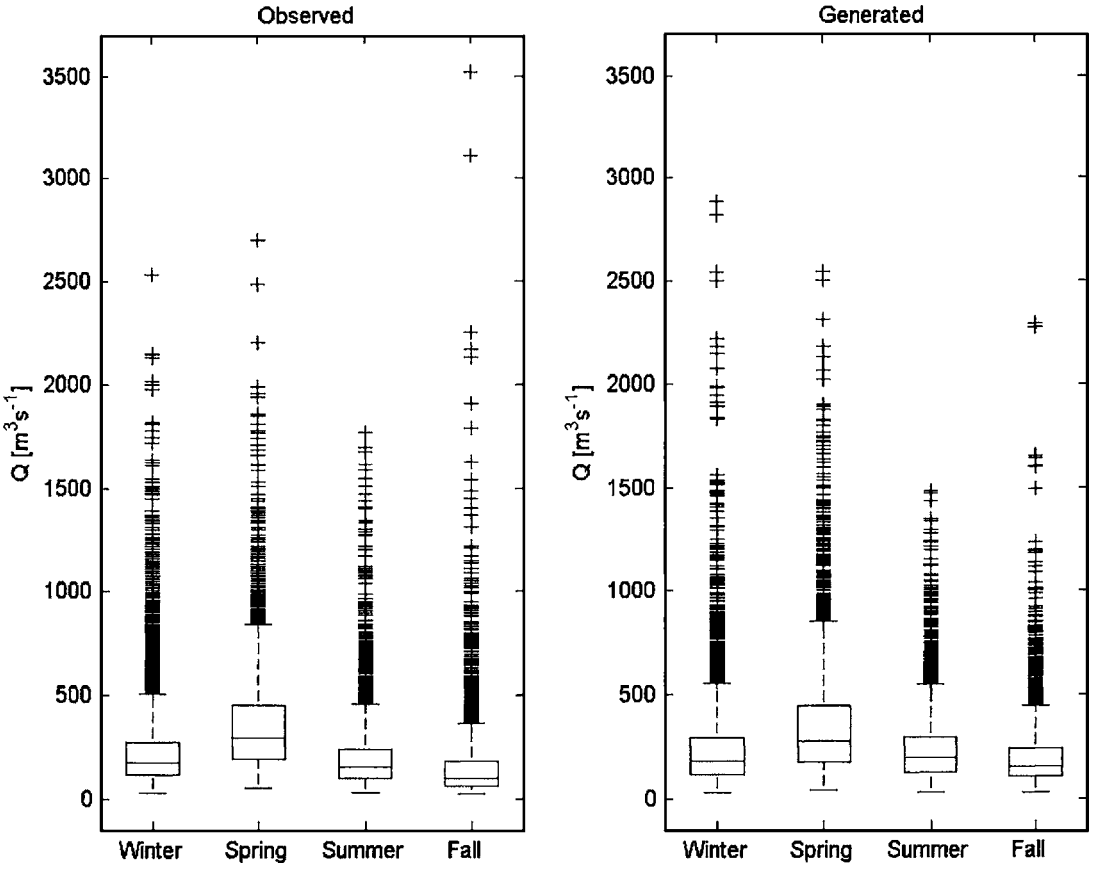

Fig. 10. Seasonal box plots of 50 years of observed and simulated daily flow values of Tisza River at Tivadar (see text for explanation)

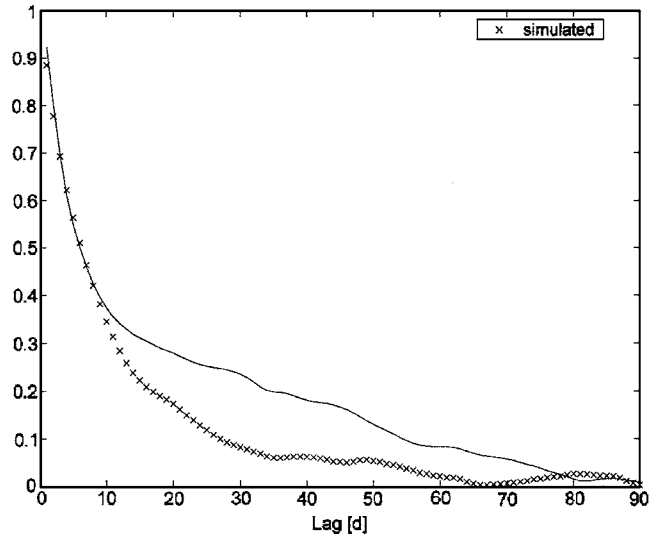

Fig. 11. Autocorrelation functions of 50 years of observed and simulated daily flow values of Tisza River at Tivadar

storage, while the lower part is mainly the result of baseflow recession (in the case of typical groundwater-fed streams), which has been demonstrated as a characteristically nonlinear process (e.g., Szilagyi 1999). Often a nonlinear reservoir approach is used to describe the behavior of the full range of the recession limb of the hydrograph as

$$Q = kS^n \tag{3}$$

where Q [$L^3 T^{-1}$]=observed streamflow; k [$L^{3(1-n)} T^{-1}$]=storage coefficient; S [L^3]=stored water volume; and n [-]=exponent. To accommodate for the different sources of water during recession flow, the value of the exponent may be changed with time (Kavvas and Delleur 1984). As an alternative, rather than changing n through time, the value of k may be changed (Aksoy et al. 2001; Aksoy 2003) with the $n=1$ choice. When n is unity, Eq. (3) can be written in a differentiated form as

$$\frac{dQ}{dt} = -kQ \tag{4}$$

which has a solution

$$Q(t) = Q_0 e^{-kt} \tag{5}$$

that can be written for $t=1$ day and with the $Q_0=Q(t-1)$ choice as

$$Q(t) = e^{-k'} Q(t-1) = c_1 Q(t-1) \tag{6}$$

where $k'(=1 \cdot k)$ (-). Employing a finite difference approximation of Eq. (4) with $t=1$ day yields

$$Q(t) = (1-k')Q(t-1) = c_2 Q(t-1) \tag{7}$$

which shows that, by the proper choice of c_2 in the finite difference scheme, one can obtain the analytical solution of Eq. (6). By letting the value of k' in Eq. (7) change through time, one can simulate the outflow of a nonlinear reservoir having a variable exponent through time. The following expression permits the

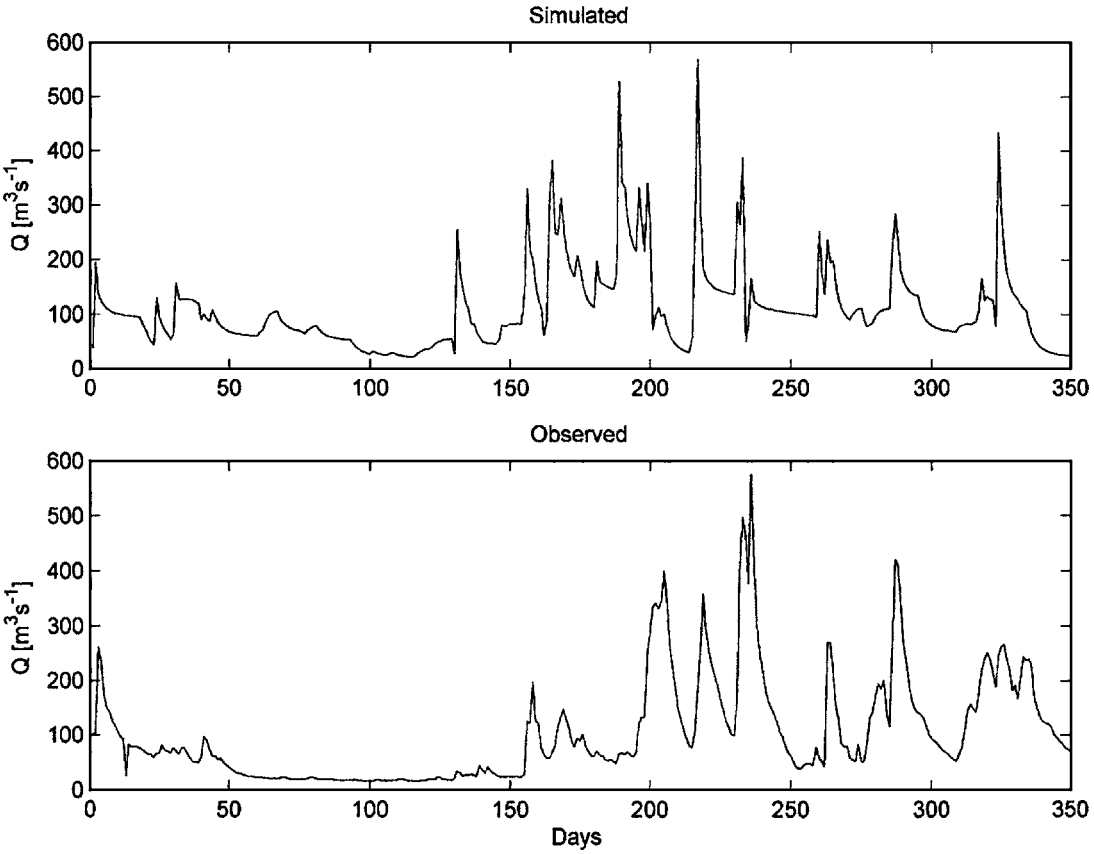

Fig. 12. Sample observed and Monte Carlo generated time series of daily flow rates of Szamos River at Csenger

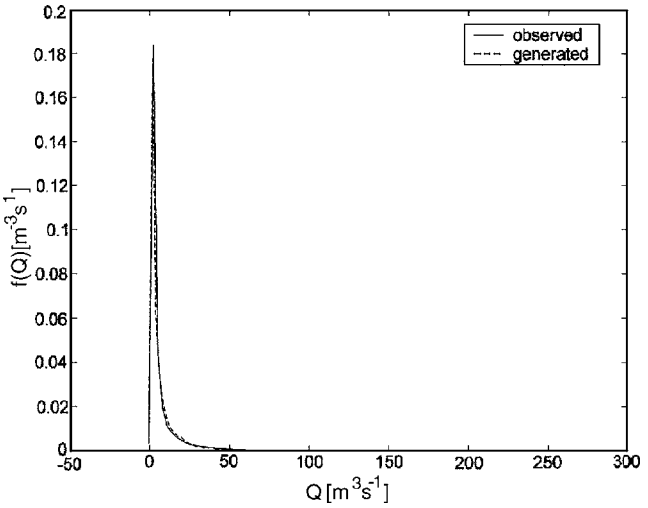

Fig. 13. Estimated distribution functions of 50 years of observed and simulated daily flow rates of Kraszna River at Agerdomajor

value of c_2 to increase in a logarithmic fashion from a minimum value at the time of the peak of the hydrograph to close to unity if $k'_{\min}$ is chosen sufficiently small

$$Q(t) = Q(t-1)\left[1 - k'_{\min} - \frac{k'_{\max} - k'_{\min}}{\ln\left(\frac{Q_{\max}}{Q_{\min}}\right)} \ln\left(\frac{Q(t-1)}{Q_{\min}}\right)\right] \quad (8)$$

Note that when $Q(t-1)=Q_{\max}$, $c_2=1-k'_{\max}$; and when $Q(t-1)=Q_{\min}$, $c_2=1-k'_{\min}$. Eq. (8) assures that the recession is steeper than a negative exponential function, and so fits observations (Kavvas and Delleur 1984).

The above description of recession flow cannot account for year-to-year or season-to-season variations in the volume of groundwater stored in the catchment. During wet years/seasons, this additional source of water will prevent very low flow rates in the channel for perennial streams. The model can account for this variability by adding a stochastic groundwater component to the recession flow model of Eq. (8) in the form

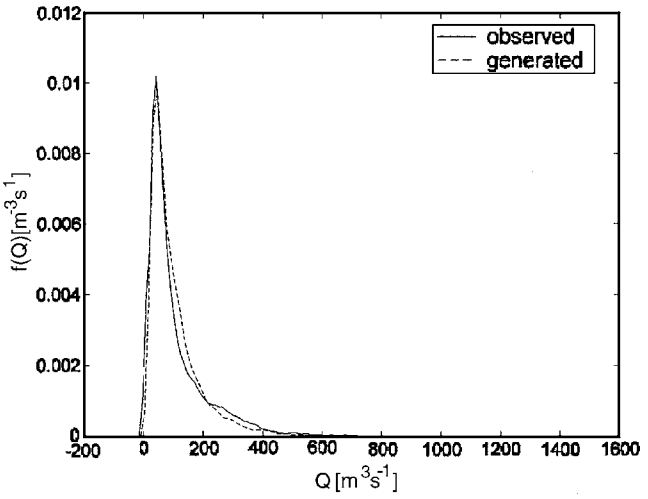

Fig. 14. Estimated distribution functions of 50 years of observed and simulated daily flow rates of Bodrog River at Felsoberecki

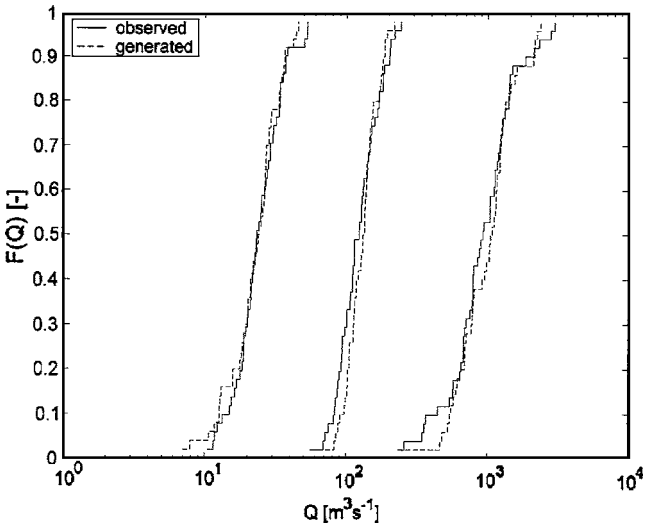

Fig. 15. Empirical cumulative distribution functions of 50 years of observed and simulated annual maxima, means, and minima of daily flow rates of Szamos River at Csenger

$$Q_{\text{gw}}(t) = (1 - k'_{\min})Q_{\text{gw}}(t-1) \quad (9)$$

where Q_{gw} designates the groundwater contribution to the channel flow, which thus becomes the sum of Eqs. (8) and (9) during recession flow periods. The starting value of Q_{gw} with a wet-to-dry transition at time $t=0$ is obtained as

$$Q_{\text{gw}}(0) = |W[g \cdot Q_{\text{gen}}(t), h \cdot Q_{\text{gen}}(t)]| \quad (10)$$

where, again, g (-) and h (-)=parameters to be calibrated; and W, as before, = normally distributed variable. The straight brackets are for taking the absolute value. In theory, the multiplier of Q_{gw} in Eq. (9) could change with time as in the channel flow case (Brutsaert and Nieber 1977; Szilagyi 1999, 2004), but that would further complicate the model, which is intended to be as simple as necessary.

The model has altogether ten parameters (a, b, d, f, g, h, $k_{\max}$, $k_{\min}$, $Q_{\max}$, $Q_{\min}$), to be specified for each gaging station. Two of them, $Q_{\max}$ and $Q_{\min}$, are the observed extrema and can be specified during Monte Carlo simulation to be somewhat larger and smaller, respectively, than their historical values, in order to accommodate for possibly larger or smaller generated values than what were observed. From the remaining eight parameters, only a varied with season in our study area and even that only for the main channel. The rest of the parameters were constant over the year.

Model Results and Conclusions

Before calibration, involving Monte Carlo simulations, one has to decide the number of values to be generated. A natural choice is the number of available observations. However, when the number of observations is small, then the statistics of the so-chosen equally small number of generated values may significantly differ between simulation runs, making any calibration based on such short runs questionable. To demonstrate this point, in Fig. 6 we plotted some basic statistical values (mean, standard deviation, skewness coefficient, minimum, and maximum) of individual simulation runs of our proposed model with differing length (i.e., 5, 10, 25, 50, 100, 200, 400, 800 years of daily data) for the Tivadar gaging station. The model parameter values were derived

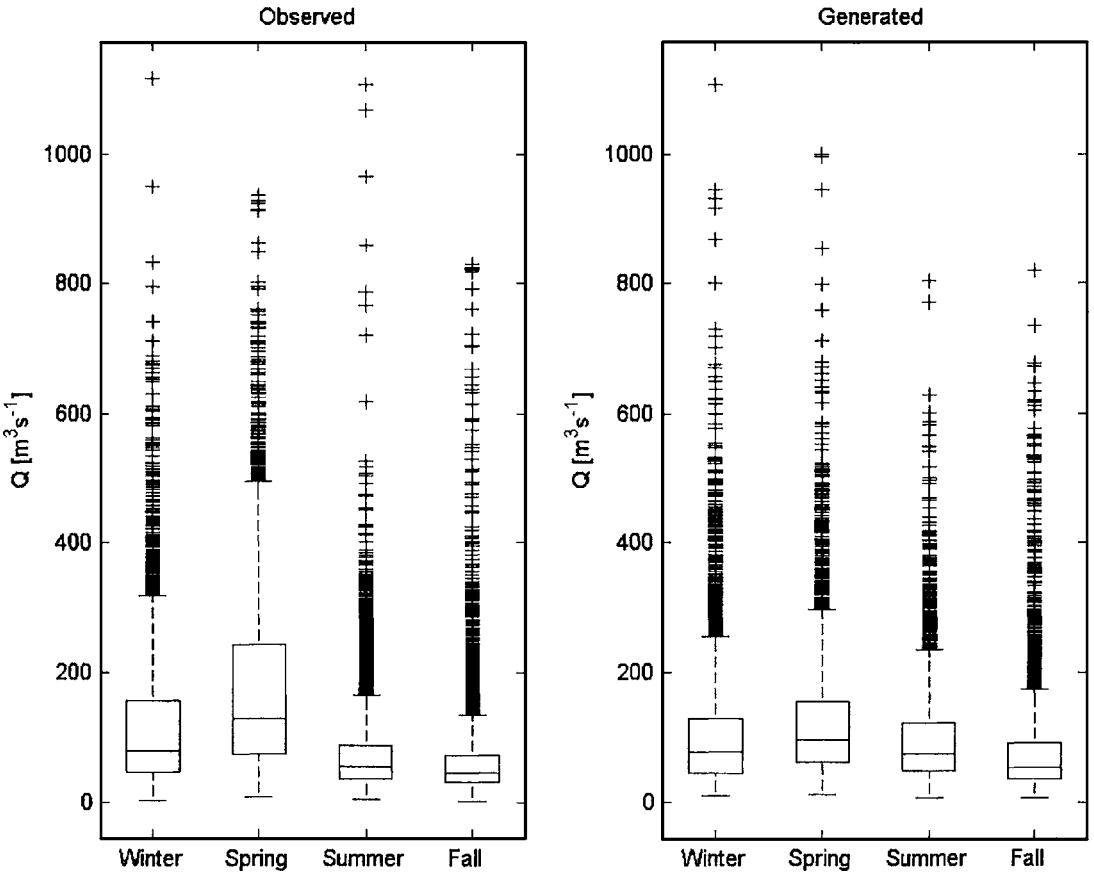

Fig. 16. Seasonal box plots of 50 years of observed and simulated daily flow values of Bodrog River at Felsoberecki

by calibration based on 50 year runs and remained unchanged for all cases shown in Fig. 6. It can be seen that all the statistics may change significantly between model runs shorter than 25 years. Fig. 6 thus suggests that calibration based on 50 or 100 years of generated daily data (involving altogether 18,262 or 36,524 data points, respectively) can most probably be considered as stable. Model calibration based on 50 year runs of simulated data was therefore employed in this study. With every new set of parameter values this way, the Monte Carlo simulation produces a new time series of 50 years which is then compared to the observed time series. Calibration again was performed by trial and error based on visually comparing: (1) the estimated distribution function of daily values; (2) the empirical cumulative distribution functions of the annual maxima, means, as well as minima of the generated values; and (3) the generated time series, with those of the observed ones.

For the main channel, recessions were modeled with $k'_{\max}=0.33$, $k'_{\min}=0.015$, $g=0.04$, and $h=0.02$. Sample observed and generated time series of daily discharges at Tivadar are displayed in Fig. 7. The asymmetric shape of the observed hydrographs is well conserved in the generated data. The distribution function estimates (using the Matlab function "ksdensity") of 50 years of observed and simulated daily flow rates are compared in Fig. 8. The empirical cumulative distribution functions of the annual maxima, means, and minima are displayed in Fig. 9, employing Weibull plotting positions. They all pass the Kolmogorov–Smirnov test at the 5% level for the null hypothesis that the observed and generated annual minimum, mean, and maximum values are from the same hypothetical distributions.

Fig. 10 exhibits box plots of observed and 50 years of simulated daily flow rates of the Tisza River at Tivadar for each season. The bottom and top of each box corresponds to the lower and upper quartiles of the data, respectively; the middle line represents the median, and the whiskers extend to the most extreme data within 1.5 times the interquartile range (i.e., the height of the box). The crosses are outlier values. The annual change in the median values (i.e., elevated water levels in spring, low flows in

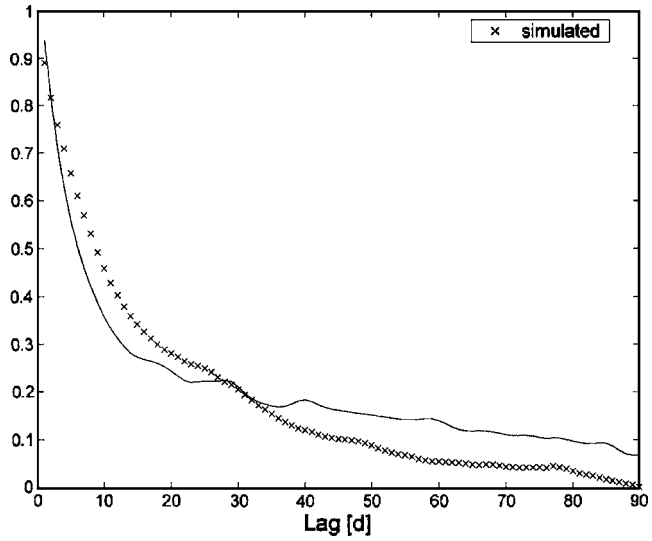

Fig. 17. Autocorrelation functions of 50 years of observed and simulated daily flow values of Kraszna River at Agerdomajor

Table 5. One-Step Serial and Cross Correlation (between Actual Station (.) and Tivadar [T]) Coefficients of Observed and Model-Generated Daily Flow Values

	One-step autocorrelation (r_1)		Cross correlation ($r_{0,T}$)	
Gaging station (stream)	Observed	Generated	Observed	Generated
Tivadar (Tisza)	0.92	0.89	—	—
Csenger (Szamos)	0.90	0.80	0.78	0.68
Felsoberecki (Bodrog)	0.98	0.93	0.69	0.66
Agerdomajor (Kraszna)	0.94	0.89	0.51	0.59

Table 6. Observed Values (Q_p) of Selected Percentiles of Empirical Cumulative Distribution Function of Daily Flow Values at Tivadar, Tisza

Percentile (%)	Q_p ($m^3 s^{-1}$)	d (%)	v (%)	n
99	1170	1 (0.87)	6.1 (5.27)	69 (91)
95	689	5 (4.27)	20.17 (16.74)	274 (314)
90	504	10 (9.43)	32.23 (28.54)	484 (593)
10	66	10 (4.31)	2.1 (0.93)	—
5	53	5 (1.51)	0.8 (0.27)	—
1	38	1 (0.19)	0.1 (0.03)	—

Note: Number of peaks (n), as well as relative duration (d) and volume (v), of observed and generated (in parentheses) flow rates larger (first three) or smaller (last three) than Q_p.

Table 7. Observed Values (Q_p) of Selected Percentiles of Empirical Cumulative Distribution Function of Daily Flow Values at Csenger, Szamos River

Percentile (%)	Q_p ($m^3 s^{-1}$)	d (%)	v (%)	n
99	747	1 (0.81)	8.68 (6.58)	74 (83)
95	394	5 (3.88)	24.95 (18.48)	286 (286)
90	275	10 (8.62)	37.56 (30.13)	462 (461)
10	28.8	10 (3.35)	1.74 (0.59)	—
5	23.4	5 (1.49)	0.75 (0.22)	—
1	16.2	1 (0.15)	0.1 (0.02)	—

Note: Number of peaks (n), as well as relative duration (d) and volume (v), of observed and generated (in parentheses) flow rates larger (first three) or smaller (last three) than Q_p.

Table 8. Observed Values (Q_p) of Selected Percentiles of Empirical Cumulative Distribution Function of Daily Flow Values at Felsoberecki, Bodrog River

Percentile (%)	Q_p ($m^3 s^{-1}$)	d (%)	v (%)	n
99	556	1 (0.51)	6.2 (3.37)	33 (37)
95	341	5 (3.03)	21.4 (13.65)	112 (160)
90	260	10 (6.08)	34.89 (22.48)	228 (265)
10	25.6	10 (7.22)	1.23 (1.38)	—
5	10.1	5 (0.18)	0.37 (0.01)	—
1	8.1	1 (0.05)	0.057 (0.003)	—

Note: Number of peaks (n), as well as relative duration (d) and volume (v), of observed and generated (in parentheses) flow rates larger (first three) or smaller (last three) than Q_p.

Table 9. Observed Values (Q_p) of Selected Percentiles of Empirical Cumulative Distribution Function of Daily Flow Values at Agerdomajor, Kraszna River

Percentile (%)	Q_p ($m^3 s^{-1}$)	d (%)	v (%)	n
99	62.7	1 (0.77)	15.01 (11.74)	53 (45)
95	27.5	5 (4.18)	39.43 (33.5)	208 (150)
90	15.8	10 (9.62)	55.31 (51.86)	360 (274)
10	0.6	10 (16.4)	0.467 (0.99)	—
5	0.34	5 (6.49)	0.137 (0.23)	—
1	0.115	1 (0.91)	0.014 (0.013)	—

Note: Number of peaks (n), as well as relative duration (d) and volume (v), of observed and generated (in parentheses) flow rates larger (first three) or smaller (last three) than Q_p.

autumn), as well as the skewness of the distributions are clearly maintained in the Monte Carlo generated daily flow rates. Fig. 11 displays the corresponding autocorrelation functions of 50 years of observed and simulated daily flow values.

Sample observed and simulated daily flow rates are displayed in Fig. 12 for the Szamos River at Csenger. Distribution function estimates are shown in Figs. 13 and 14 for the Kraszna at Agerdomajor and the Bodrog at Felsoberecki, respectively. Fig. 15 demonstrates the empirical cumulative distribution functions of observed and simulated annual maxima, means, and minima for the Szamos River at Csenger, all passing the Kolmogorov–Smirnov test at the 5% level. Seasonal box plots are exhibited for the Bodrog River at Felsoberecki in Fig. 16. It generally shows that model performance for tributaries is somewhat poorer than for the main channel (see Fig. 10). Finally, the autocorrelation functions (observed and simulated) of the daily flow values are displayed in Fig. 17 for the Kraszna River at Agerdomajor. Table 5 lists the one-step autocorrelation and zero-lag cross-correlation values of the observed and generated time series. Tables 6–9 document how the generated values reproduce observed frequency, duration, and volume characteristics of floods and low flows. It can be concluded that flood attributes are better retained than low flow features. This is in spite of the stochastic baseflow component in the model meant to account for a dynamic groundwater storage in the watershed which has been lacking in the approach of Aksoy (2003), and our results only emphasize the importance of groundwater–surface water interactions in daily synthetic flow generation.

Table 10. Sensitivity Analysis of Model Parameters at Csenger, Szamos River, Based on Coefficient of Variation Ratio (rC_v) of Generated and Original Time Series. Parameters Are Listed from Most Sensitive to Least Sensitive Order. rC_v=0.83 with Calibrated Parameters

	rC_v when parameter value is doubled	rC_v when parameter value is halved
b	>10	0.57
k_{max}	1.16[a]	0.53
g	0.35	1.12
a	1.03	0.69
h	0.78	0.79
d	0.86	0.79
k_{min}	0.84	0.82
f	0.82	0.83

[a]Parameter value is multiplied by 1.5 instead of two due to natural constraint ($k<1$).

Finally, Table 10 displays the outcome of a parameter sensitivity test performed at Csenger. The model is most sensitive to the value of the noise term exponent (and also to its scale parameter, a) for diurnal discharge changes and to the recession constant, k_{max}. The model is also highly sensitive to the starting value of baseflow recession through the parameter g, which corroborates our findings above, emphasizing the importance of the groundwater dynamics in daily flow-rate time series generation.

In conclusion, it can be stated that by applying the proposed hybrid, seasonal Markov chain-based approach of daily flow simulation at multiple catchment sites it is possible to generate arbitrarily long time series of daily flow rates that at least moderately well preserve basic long-term (mean, variance, skewness, autocorrelation structure, cross-correlations) statistics, as well as short-term behavior (asymmetric hydrograph) of the original time series. In a seasonal comparison, the model better works for the main channel than for the tributaries. This is so because tributary state transitions could not be linked to the main channel in a probabilistic way (i.e., the tributary state transitions ought to be conditioned by the corresponding state of the main channel); rather, they were linked through a deterministic polynomial expression of diurnal increases. The general modeling approach (main channel and tributaries as well), however, is centered around the concept of conditional heteroscedasticity, which means that the noise term of the stochastic model applied is not independent of the process to be modeled and neither is it identically distributed. The model has altogether nine parameters (in a seasonal formulation) for the main channel site to be calibrated, and eight additional parameters for each tributary gaging station. While the described approach is simple, calibration of the parameters may require some effort from the modeler, especially because no simple target function of calibration could be found since the generated time series must simultaneously satisfy both short- and long-term behavior of the observed time series.

Acknowledgments

This work has been supported by the Hungarian Research and Development Project: "Flood Risk Analysis," Grant No. NKFP 3/067/2001 by the Hungarian American Joint Research Fund (MAKA). The writers are grateful to Charles Flowerday for his editorial help and to Margit Horosz-Gulyas for her help with the figures. The views, conclusions, and opinions expressed in this paper are solely those of the writers and not the University of Nebraska, State of Nebraska, or any political subdivision thereof. This paper is a contribution of the University of Nebraska Agricultural Research Division, Lincoln, NE 68583, Journal Series No. 14518.

References

Aksoy, H. (2003). "Markov chain-based modeling techniques for stochastic generation of daily intermittent streamflows." *Adv. Water Resour.*, 26, 663–671.

Aksoy, H., Bayazit, M., and Wittenberg, H. (2001). "Probabilistic approach to modeling of recession curves." *Hydrol. Sci. J.*, 46(2), 269–285.

Bernier, J. (1970). "Inventaire des modeles de processus stochastiques applicable a la description des debits journaliers des riveres." *Int. Statist. Rev.*, 38(1), 49–61.

Bierkens, M. F. P., and Puente, C. E. (1990). "Analytically derived runoff models based on rainfall point processes." *Water Resour. Res.* 26(11), 2653–2659.

Brutsaert, W., and Nieber, J. L. (1977). "Regionalized drought flow hydrograph from a mature glaciated plateau." *Water Resour. Res.*, 13(3), 637–643.

Cowpertwait, P. S. P., and O'Connell, P. E. (1992). "A Neymann–Scott shot noise model for the generation of daily streamflow time series." *Advances in theoretical hydrology—A tribute to James Dooge*, J. P. O'Kane, ed., Elsevier, New York.

DeBarry, P. A. (2004). *Watersheds: Processes, assessment, and management*, Wiley, Hoboken, N.J.

Engle, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation." *Econometrica*, 50, 987–1007.

Kavvas, M. L., and Delleur, J. W. (1984). "A statistical analysis of the daily streamflow hydrograph." *J. Hydrol.*, 71, 253–275.

Kelman, J. (1980). "A stochastic model for daily streamflow." *J. Hydrol.*, 47, 235–249.

Koch, R. W. (1985). "A stochastic streamflow model based on physical principles." *Water Resour. Res.*, 21(4), 545–553.

Kottegoda, N. T., and Horder, M. A. (1980). "Daily flow model based on rainfall occurrences using pulses and a transfer function." *J. Hydrol.*, 47, 215–234.

McGinnis, D. F., and Sammons, W. H. (1970). "Discussion of 'Daily streamflow simulations.'" *J. Hydraul. Div., Am. Soc. Civ. Eng.*, 96(5), 1201–1206.

Murrone, F., Rossi, F., and Claps, P. (1997). "Conceptually-based shot noise modelling of streamflows at short time interval." *Stochastic Hydrol. Hydr.*, 11(6), 483–510.

Payne, K., Neumann, W. R., and Kerri, K. D. (1969). "Daily streamflow simulation." *J. Hydraul. Div., Am. Soc. Civ. Eng.*, 95(4), 1163–1180.

Quimpo, R. G. (1968). "Stochastic analysis of daily river flows." *J. Hydraul. Div., Am. Soc. Civ. Eng.*, 94(1), 43–57.

Sargent, D. M. (1979). "A simplified model for the generation of daily streamflows." *Hydrol. Sci. Bull.*, 24(4), 509–527.

Sharma, A., Tarboton, D. G., and Lall, U. (1997). "Streamflow simulation: a nonparametric approach." *Water Resour. Res.*, 33(2), 291–308.

Szilagyi, J. (1999). "On the use of semi-logarithmic plots for baseflow separation." *Ground Water*, 37(5), 660–662.

Szilagyi, J. (2004). "Heuristic continuous baseflow separation." *J. Hydrologic Eng.*, 9(4), 1–8.

Treiber, B., and Plate, E. J. (1977). "A stochastic model for the simulation of daily flows." *Hydrol. Sci. Bull.*, 22(1), 175–192.

Weis, G. (1973). "Shot noise models for synthetic generation of multisite daily streamflow data." *IAHS Publ.*, 108, 457–467.

Weis, G. (1977). "Shot noise models for the generation of synthetic streamflow data." *Water Resour. Res.*, 13(1), 101–108.

Xu, Z. X., Schumann, A., and Brass, C. (2001). "Markov autocorrelation pulse model for two sites daily streamflow." *J. Hydrologic Eng.*, 6(3), 189–195.

Xu, Z. X., Schumann, A., and Li, J. (2003). "Markov cross-correlation pulse model for daily streamflow generation at multiple sites." *Adv. Water Resour.*, 26, 325–335.