Journal of Applied Statistics,
Vol. 31, No. 9, 1083–1094, November 2004

Using Markov Chains for Non-perennial Daily Streamflow Data Generation

HAFZULLAH AKSOY
Department of Civil Engineering, Istanbul Technical University, Turkey

ABSTRACT *The use of Markov chains to simulate non-perennial streamflow data is considered. A non-perennial stream may be thought as having three states, namely zero flow, increasing flow and decreasing flow, for which a three-state Markov chain can be constructed. Alternatively, two two-state Markov chains can be used, the first of which represents the existence and non-existence of flow, whereas the second deals with the increment and decrement in the flow for periods with flow. Probabilistic relationships between the two alternatives are derived. Their performances in simulating the state of the stream are compared on the basis of data from two different geographical regions in Turkey. It is concluded that both alternatives are capable of simulating the state of the stream.*

KEY WORDS: Daily streamflow, data generation, Markov chain, simulation

Introduction

The analysis, estimation, forecasting, extension, and simulation of streamflow data are important in hydrological applications. Of these, simulation – synthetic data generation – is of particular importance because of the short hydrological records that hydrologists face in many cases. This task becomes more important in the analysis of non-perennial streams where hydrological records are usually shorter than those of perennial streams. Non-perennial streams here refer to both intermittent and ephemeral streams. Synthetic data generation aims to obtain the population of streamflow data from which the streamflow record is extracted. In this technique, it is assumed that the recorded data set is a sample representing the population. Thus, the synthetic data sequence corresponds to different realizations with the same statistical behaviour as the recorded sample.

Synthetic data sets are of particular use in water resources studies. For example, the planning, development, management, and operation of water resources systems require large amounts of synthetic data. Many water resources-related studies – such as the development of river ecology, the design of small-scale rural water supply schemes, river regulation, flood control and routing, and the release

Correspondence Address: Hafzullah Aksoy, Hydraulics Division, Department of Civil Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey. Email: haksoy@itu.edu.tr

0266-4763 Print/ 1360-0532 Online/04/091083-12
DOI: 10.1080/0266476042000280418

of water for water quality control and fisheries during a low flow period – use streamflow data in the daily time interval.

Different techniques are available for streamflow data generation. Those used for the generation of daily streamflow data are in three main groups: (i) autoregressive models, (ii) shot noise models, and (iii) Markov chain models. Autoregressive models were first studied by Quimpo (1967) for the analysis of the structure of daily streamflow data. Shot noise models were first developed and applied by Weiss (1977). The advantage of shot noise models over autoregressive models is that shot noise models are capable of reproducing the short-term characteristics (ascent and descent curves) of the daily streamflow hydrograph in addition to its long-term features, e.g. mean, standard deviation, and higher order central moments.

The earliest use of probability in the streamflow goes back to the study by Savarenskiy (1940), who developed methods of calculating the flow coefficients of rivers for which long records were available. Later, Moran (1954, 1959) developed the stochastic theory of storage based on independent flow inputs leading to a homogeneous Markov chain for the storage. The assumption of independence of inputs is only an approximation, as many annual river flows exhibit a significant serial correlation. Because of this, Fiering (1967) explored Markov autoregressive schemes to take the dependent structure of annual river flow into account. The Markov chain models were later used in streamflow data generation by Treiber & Plate (1977) for perennial streams. Aksoy & Bayazit (2000a,b) and Aksoy (2003, 2004) extended the use of the Markov chain model to intermittent streams by using two two-state Markov chains or, alternatively, a three-state Markov chain.

This study investigates how these two-state Markov chains are related to the three-state Markov chain by deriving relationships between the transition probabilities of the chains. We apply both approaches to two data sets taken from two geographically different regions in Turkey. We conclude that both approaches simulate the daily non-perennial streamflow time series satisfactorily.

Markov Chain

A non-perennial streamflow time series may be divided into three states, namely zero flow, increasing flow, and decreasing flow (Figure 1). Two approaches are proposed. The first approach uses two two-state Markov chains whereas the second uses one three-state Markov chain only.

Two-state Markov chains

In the first approach we construct a two-state Markov chain for determining whether the stream has flow '1' or not '0' (Figure 2) for which the transition probability matrix can be written as:

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} \tag{1}$$

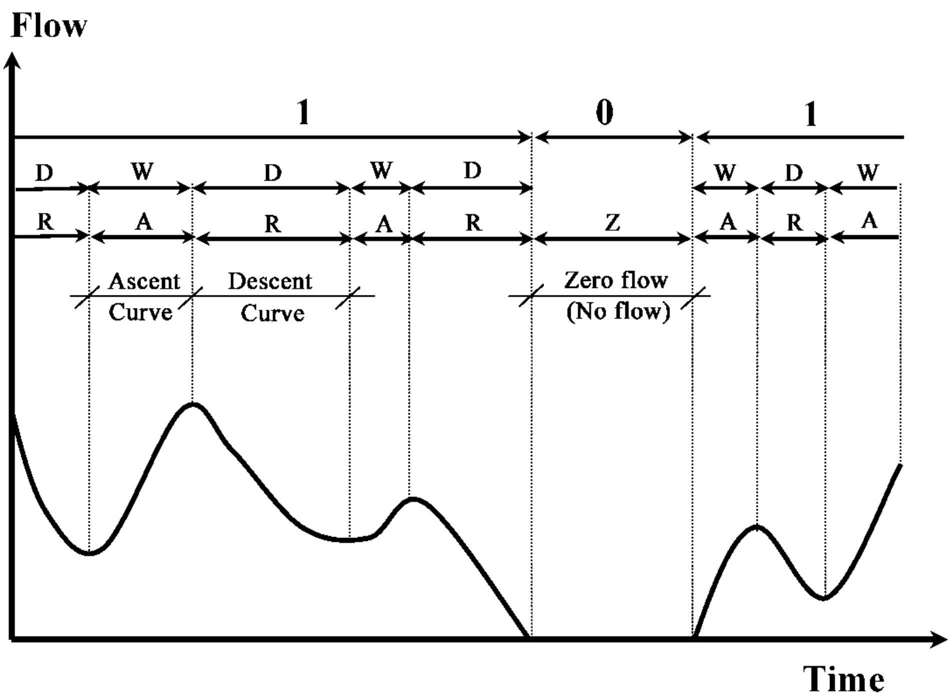

Figure 1. Schematic representation of the daily non-perennial streamflow hydrograph

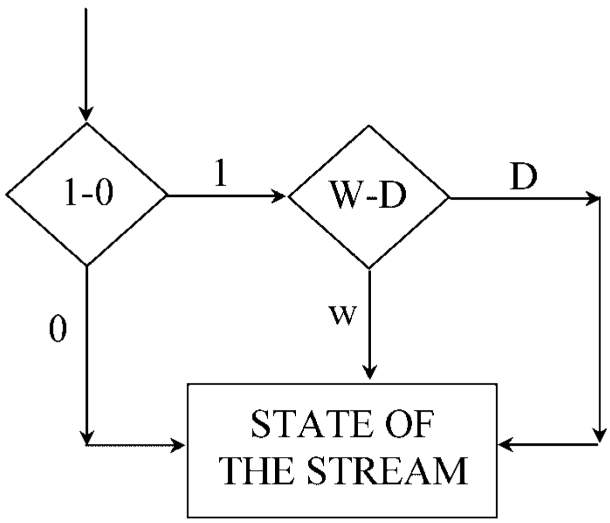

Figure 2. 1-0/*W-D* two-state Markov chain approach

Once days with and without flow are determined, another two-state Markov chain is used in determining the days on which flow increases '*W*' or decreases '*D*' (Figure 2). The matrix of this Markov chain is given by:

$$P = \begin{bmatrix} P_{WW} & P_{WD} \\ P_{DW} & P_{DD} \end{bmatrix} \tag{2}$$

The number of parameters required is two for each matrix as the sum of the probabilities equals one for each row of the matrices.

Three-state Markov chain

Instead of the 1-0 and *W-D* two-state Markov chains, a three-state Markov chain can be used for the determination of the state of the stream. The matrix of such a Markov chain can be written as:

$$P = \begin{bmatrix} P_{AA} & P_{AR} & P_{AZ} \\ P_{RA} & P_{RR} & P_{RZ} \\ P_{ZA} & P_{ZR} & P_{ZZ} \end{bmatrix} \tag{3}$$

where *A*, *R* and *Z* represent an increase in the flow (upward curve of the hydrograph), a decrease in the flow (downward curve of the hydrograph), and zero flow (day with no flow), respectively (Figure 3). The number of parameters is six. However, $P_{ZR} = 0$ as the flow cannot decrease further after a day with no flow, and $P_{AZ} \cong 0$, since a day with no flow is very rare after a day with an increase in the flow. Thus, the number of parameters is reduced from six to four.

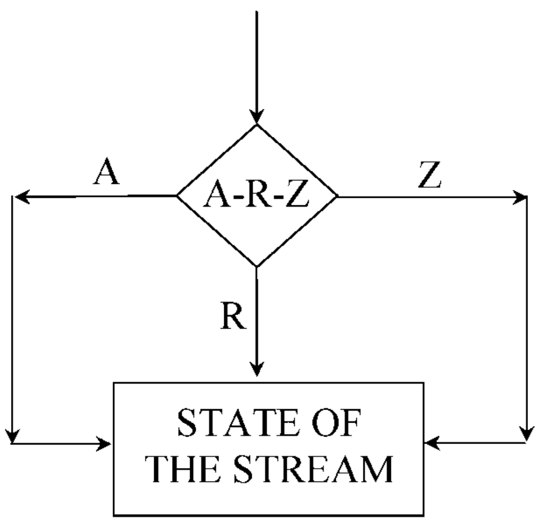

Figure 3. *A-R-Z* three-state Markov chain approach

Relation Between Two- and Three-state Markov Chains

Figure 1 depicts how the states 1-0, *W-D* and *A-R-Z* are related to each other. State '0' in the 1-0 Markov chain corresponds to state '*Z*' in the *A-R-Z* Markov chain, and state '1' to states '*A*' and '*R*'. Also, it is clear that states '*W*' and '*D*' in the *W-D* Markov chain correspond to states '*A*' and '*R*' of the *A-R-Z* three-state Markov chain. However, their transition probabilities were found not to be equal in months where the stream dries up. The reason for this is that the *W-D* Markov chain operates only in the period where the flow is not zero, whereas the *A-R-Z* Markov chain takes the streamflow series as a whole.

Relations between transition probabilities of the two- and three-state Markov chains are derived in the following sections. First relations between the 1-0 and *A-R-Z* Markov chains are given. Then relations between the *W-D* and *A-R-Z* Markov chains are derived.

Relation Between the 1-0 and A-R-Z Markov Chains

As stated before, state '0' of the 1-0 Markov chain corresponds to state '*Z*' in the *A-R-Z* Markov chain (Figure 1). Therefore it is obvious that:

$$P_{00} = P_{ZZ} \tag{4}$$

Knowing that the probability of transition from one state to another is a conditional probability and that state '1' in the 1-0 Markov chain corresponds to states '*A*' and '*R*' in the *A-R-Z* Markov chain, we see from the definition of conditional probability, that:

$$P_{01} = P(1|0) = P(A|Z) + P(R|Z) = P_{ZA} + P_{ZR} \tag{5}$$

Since $P_{ZR} = 0$, then:

$$P_{01} = P_{ZA} \tag{6}$$

Using the same procedure and knowing that $P_{AZ} \cong 0$, the following relations can be derived:

$$P_{10} \cong \frac{P_R}{P_A + P_R} P_{RZ}, P_{11} \cong \frac{P_A + (R_{RA} + P_{RR}) P_R}{P_A + P_R} \tag{7}$$

The first equation in (7) can be written alternatively as:

$$P_{10} \cong \frac{P_Z}{P_A + P_R} (P_{ZA} + P_{ZR}) \tag{8}$$

P_A and P_R in equations (7) and (8) are defined as the probability of days with states *A* and *R*, respectively. Equations (4), (6) and (7) (or alternatively (8)), express the relations between the transition probabilities of the 1-0 and *A-R-Z* Markov chains.

Relation Between the W-D and A-R-Z Markov Chains

The following relations between the probabilities of the *W-D* and *A-R-Z* Markov chains can similarly be obtained:

$$P_{WD} = \frac{P_{AR}}{P_{AA} + P_{AR}} \cong P_{AR}, P_{WW} = \frac{P_{AA}}{P_{AA} + P_{AR}} \cong P_{AA}, P_{DW} = \frac{P_{RA}}{P_{RA} + P_{RR}},$$
$$P_{DD} = \frac{P_{RR}}{P_{RA} + P_{RR}} \tag{9}$$

Application

Two different but similar approaches, both based on Markov chains, have been proposed for the simulation of the state of non-perennial daily streamflows. The first uses two Markov chains, 1-0 and *W-D*, whereas the second uses a single *A-R-Z* Markov chain only.

The application of the first approach considers the consecutive runs of the 1-0 and *W-D* Markov chains (Figure 2). First, the 1-0 Markov chain is run so that time periods, with and without flow, of the stream are determined. Then, as the second step, the *W-D* Markov chain is run in order to determine if the streamflow increases or decreases during the time period with flow, excluding the time period during which the stream is dry. This also means that the *W-D* Markov chain divides the time period with flow into two parts, ascent and descent (Figure 1). The second approach using the *A-R-Z* Markov chain is such that the three states of the non-perennial streamflow time series are obtained at once (Figure 3).

Data Description

Comparisons of the two alternative Markov chain-based approaches are made by using daily streamflow data of Seytan Deresi at Babaeski in Thrace (European part of Turkey), and Culap Suyu within the Euphrates River basin in South-eastern Anatolia, both being intermittent streams. The data sets, of which information is given in Table 1, are provided by the Electrical Power Resources,

Table 1. Data sets used in the study

River	Seytan Deresi	Culap Suyu
Location	27°06′04″E 41°25′39″N	39°02′02″E 37°09′50″N
Elevation (m)	50	467
Drainage Area (km^2)	478.4	525.2
Observation Period	1958–1992 (35 years)	1964–1992 (28 years, 1969 is missed)
Long-Term Average Flow (m^3/s)	2.44	0.70
No Flow Period (%)	10.6	19.6

Survey and Development Administration (EIE) of Turkey. Parameters (transition probabilities, P_{11}, P_{00}, P_{WW}, P_{DD}, P_{AA}, P_{RA}, P_{RR}, P_{ZA}) of the 1-0, *W-D*, and *A-R-Z* Markov chains are calculated for each month of the year using daily streamflow data and are listed in Tables 2 and 3 for Seytan Deresi and Culap Suyu, respectively. As seen in these tables parameters are listed starting from October rather than January, the first month of the calendar year. The reason, as is clear to hydrologists and water resources planners, is that the term 'year' in this study refers to 'water year' starting with the first day of October of the previous year and ending with the last day of September of the current year (e.g. from 1 October 1992 to 30 September 1993, in the year 1993).

From Table 1, we see that Culap Suyu is a drier stream than Seytan Deresi and hence has a wider dry period from April to December. The dry period for Seytan Deresi is limited to June–October (Tables 2–3).

Table 2. Transition probabilities of the 1-0, W-D, and A-R-Z Markov chains calculated from the observed daily streamflow data of Seytan Deresi

Month	P_{11}	P_{00}	P_{WW}	P_{DD}	P_{AA}	P_{RA}	P_{RR}	P_{ZZ}
October	0.998	0.869	0.778	0.855	0.776	0.145	0.853	0.869
November	0.999	0.000	0.719	0.845	0.717	0.155	0.845	0.000
December	1.000	0.000	0.645	0.824	0.645	0.176	0.824	0.000
January	1.000	0.000	0.568	0.832	0.568	0.168	0.832	0.000
February	1.000	0.000	0.588	0.877	0.588	0.123	0.877	0.000
March	1.000	0.000	0.535	0.897	0.535	0.103	0.897	0.000
April	1.000	0.000	0.486	0.893	0.486	0.107	0.893	0.000
May	1.000	0.000	0.486	0.890	0.486	0.110	0.890	0.000
June	0.998	1.000	0.513	0.892	0.513	0.108	0.890	1.000
July	0.985	0.971	0.458	0.898	0.449	0.100	0.886	0.971
August	0.964	0.981	0.657	0.930	0.657	0.067	0.887	0.981
September	0.981	0.967	0.755	0.853	0.744	0.144	0.834	0.967

Table 3. Transition probabilities of the 1-0, W-D, and A-R-Z Markov chains calculated from the observed daily streamflow data of Culap Suyu

Month	P_{11}	P_{00}	P_{WW}	P_{DD}	P_{AA}	P_{RA}	P_{RR}	P_{ZZ}
October	0.998	0.979	0.823	0.788	0.822	0.212	0.784	0.979
November	1.000	0.933	0.833	0.847	0.833	0.153	0.847	0.933
December	1.000	0.500	0.854	0.889	0.854	0.111	0.889	0.500
January	1.000	0.000	0.852	0.878	0.852	0.122	0.878	0.000
February	1.000	0.000	0.836	0.891	0.836	0.109	0.891	0.000
March	1.000	0.000	0.703	0.878	0.703	0.122	0.878	0.000
April	0.994	0.935	0.643	0.907	0.639	0.092	0.902	0.935
May	0.999	0.990	0.758	0.871	0.758	0.129	0.869	0.990
June	0.993	0.980	0.626	0.904	0.622	0.096	0.897	0.980
July	0.972	0.992	0.597	0.929	0.571	0.070	0.905	0.992
August	0.989	0.992	0.729	0.848	0.718	0.150	0.841	0.992
September	0.997	0.993	0.841	0.783	0.841	0.216	0.778	0.993

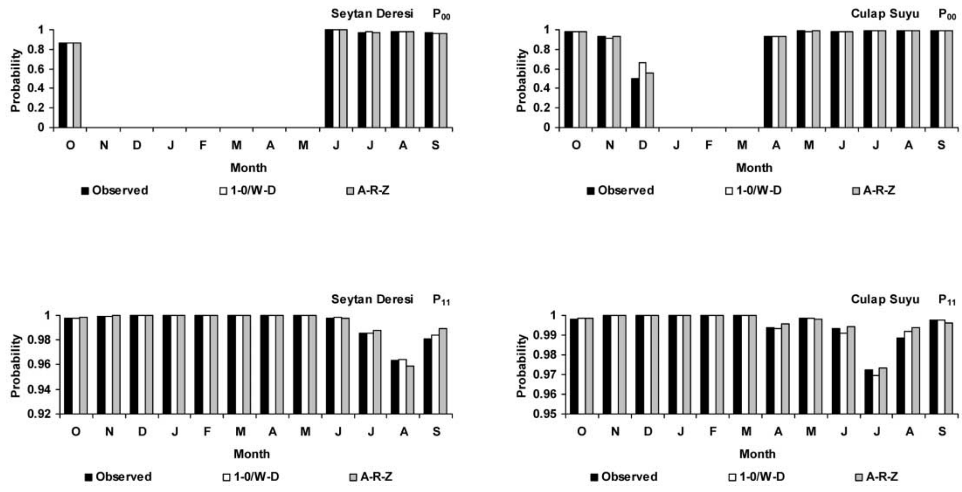

Figure 4. Transition probabilities of the 1-0 Markov chain for observed and simulated series

Results

Transition probabilities of the 1-0 Markov chain calculated from the observed daily streamflow time series are given in Figure 4 together with their counterparts calculated as the average of ten simulations. Almost the same transition probabilities were obtained for P_{11} and P_{00} by both approaches for the two streams. Relative error, the ratio of the difference between the observed and simulated transition probabilities to the observed transition probability, was found to be less than 1%, for both approaches. This means that the 1-0 and *W-D* Markov chains (together) are as suitable as the *A-R-Z* Markov chain in simulating the time periods during which the stream is with and without flow.

The probability P_{00} is close to one throughout the year during the period with flow for both streams, except for P_{00} of Culap Suyu in December. Only two days with the previous state '0' of the 1-0 Markov chain were observed in December 1973. Following the dry day of 30 November 1973, 1 December was recorded as a dry day with no flow followed by a wet day, 2 December, with flow. This resulted in $P_{00} = 0.5$ for December. The same situation holds for the probability P_{ZZ} (Table 3). Both the 1-0/*W-D* and *A-R-Z* Markov chains could not simulate P_{00} properly in this particular month. This is due to the inadequate number of simulations. Increasing the number of simulations from ten to a large enough number is likely to solve the problem and result in simulated probabilities P_{00} that are in agreement with the observed one.

Figure 5 shows the transition probabilities of the *W-D* Markov chain. It is again seen that both approaches are as suitable as each other in simulating the state of the stream for the ascent and descent parts of the hydrograph. As stated before, the *W-D* Markov chain covers the period during which the streamflow is not zero.

Not only the first-order but also the second-order transition probabilities of the *W-D* Markov chain were compared in Figure 6 to check how good the

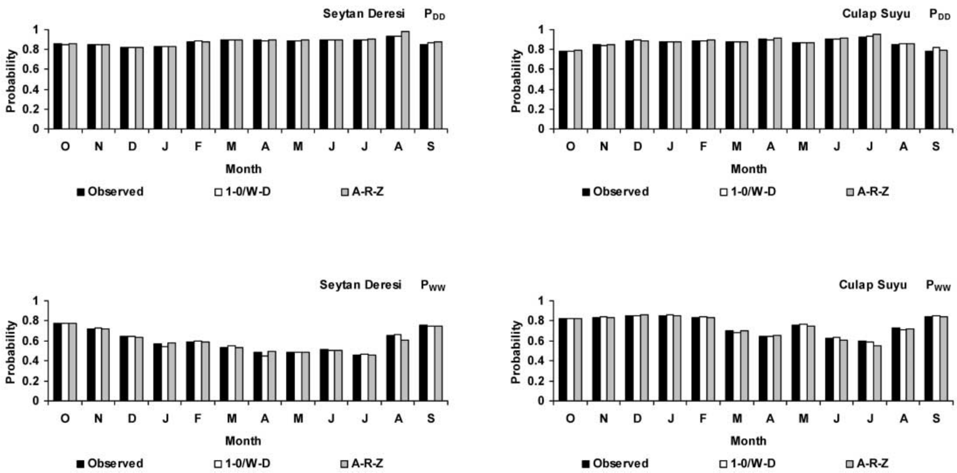

Figure 5. Transition probabilities of the *W-D* Markov chain for observed and simulated series

proposed approaches are in simulating the state of the stream. P_{WWW} shows the probability of transition from a day on the ascent curve, with a previous day on the ascent curve, to a day on the ascent curve again. Similarly, for example, P_{WDW} shows the transition probability of a day on the descent curve following a day on the ascent curve and followed by a day on the ascent curve. Figure 6 shows how the proposed approaches can generate the second-order transition probabilities of the *W-D* Markov chain, although only the first-order Markov chain was used in the generation algorithm. Based upon results in Figure 6, we may conclude that there is no need to extend the proposed first-order Markov chain approach into the second-order.

In Figure 7, transition probabilities of the *A-R-Z* Markov chain obtained as the average of ten simulations were compared to those of observed series. Again, it is seen that the transition probabilities are very close to their observed counterparts.

Summary, Discussion and Conclusion

The streamflow '1' or not '0' was simulated by the 1-0 Markov chain. Flow on any day of the non-dry period is either higher '*W*' or lower '*D*' than the previous day's flow. Based upon the configuration in Figure 2, a Markov chain-based algorithm with two steps was proposed. The 1-0 and *W-D* Markov chains are run consecutively to generate the state of the streamflow. Alternatively, a more practical algorithm, the three-state *A-R-Z* Markov chain can be used. Such a Markov chain consists of states *A*, *R* and *Z*, referring respectively to the ascent curve, descent curve, and zero flow (no flow) period of the daily non-perennial streamflow hydrograph.

A two-state Markov chain has two parameters (transition probabilities) to be determined from the observed time series. This means the 1-0/*W-D* approach

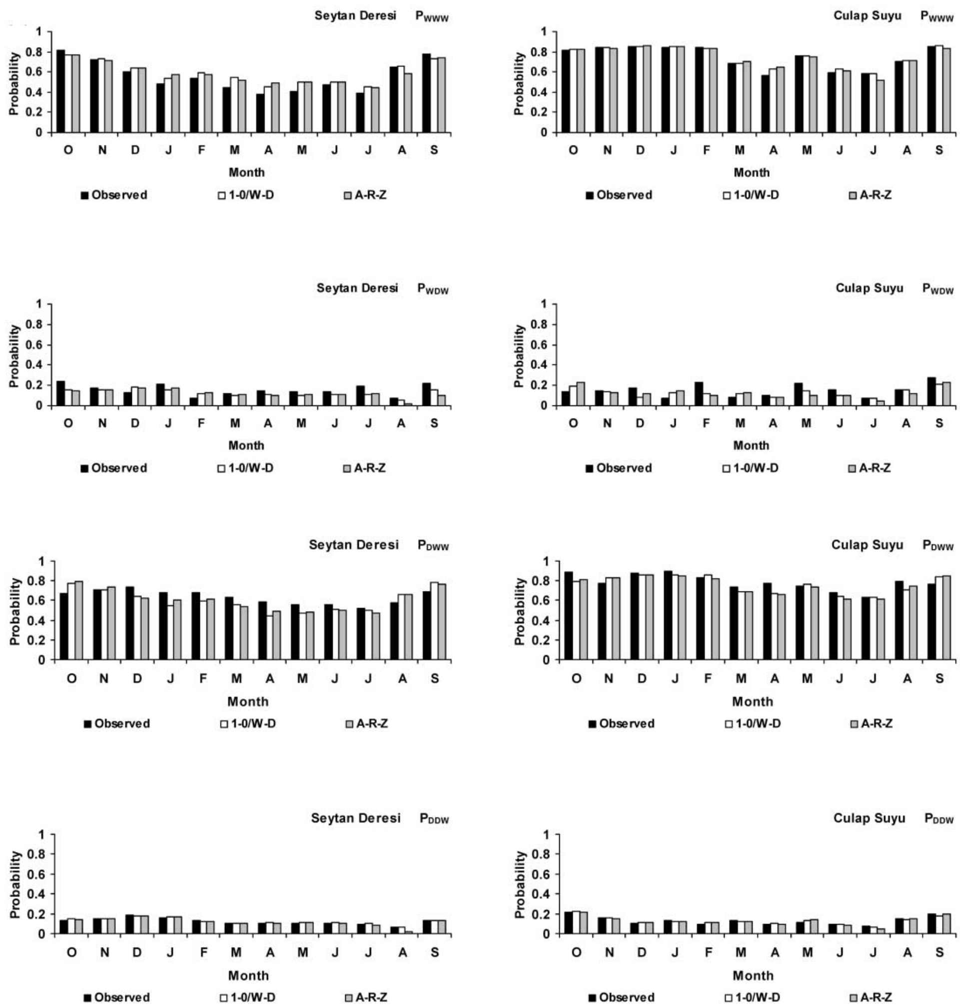

Figure 6. Second-order transition probabilities of the *W-D* Markov chain for observed and simulated series

requires four transition probabilities to be calculated for each month of the year making 48 parameters in total. Meanwhile, a three-state Markov chain is defined by six transition probabilities. However, the number of parameters required for the *A-R-Z* Markov chain in this study was reduced from six to four, which again resulted in 48 parameters in total. One of the parameters was dropped ($P_{ZR}=0$) owing to the nature of the problem, as the streamflow cannot decline further after a day with no flow. The second parameter was cancelled by making the assumption $P_{AZ}\cong 0$, which states that a day with no flow after a day with an increasing flow is very rare. This is the only assumption made in this study for reducing the number of parameters. However, it may be expected that P_{AZ} takes higher values for ephemeral streams from arid zones of the world than for

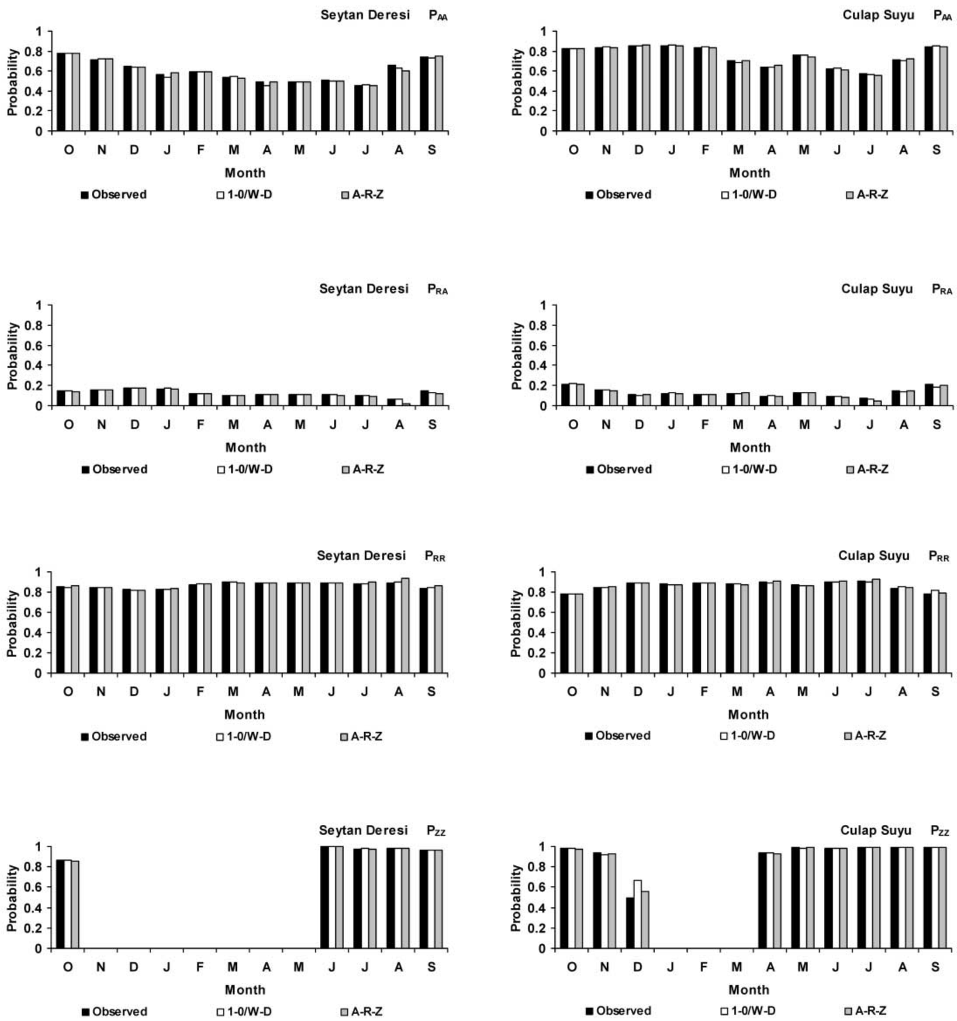

Figure 7. Transition probabilities of the *A-R-Z* Markov chain for observed and simulated series

intermittent streams of the semi-arid zones. Long and slowly decaying descent curves are typical of intermittent streams. Shorter and steeper descent curves of ephemeral streams can result in higher probabilities of P_{AZ} that may need to be considered in the analysis.

In conclusion, either the 1-0/*W-D* or the *A-R-Z* approach can be used for the simulation of the daily non-perennial streamflow time series. The *A-R-Z* approach is considered more practical than the 1-0/*W-D* approach, as it performs the simulation in one step. Following the simulation of the state of the streamflow time series, another analysis should be performed in order to determine the magnitude of the streamflow for the time period where the stream is not dry. A probability distribution function, the 2-parameter gamma distribution for instance, can be fitted to the ascent curve (Aksoy, 2000) and an exponential decay function (Aksoy, 1998) or a Markov chain-based approach (Aksoy *et al.*, 2001) can be used for the descent curve in order to determine the flow magnitude, thus completing the streamflow hydrograph.

Acknowledgements

This study was partially supported by the Research Fund of Istanbul Technical University and is based on the author's PhD thesis supervised by Professor M. Bayazit, whom the author sincerely thanks. The author wishes to extend his thanks to the anonymous referees whose review improved this paper substantially, and to his dear friend Dr Pieter H. A. J. M. van Gelder of TU Delft, the Netherlands who helped in providing essential references without which the paper would be incomplete.

References

Aksoy, H. (1998) Determination of recession curve parameters, *Low Flows Expert Meeting*, The University of Belgrade, 10–12 June 1998, Belgrade, pp. 89–96.

Aksoy, H. (2000) Use of gamma distribution in hydrological analysis, *Turkish Journal of Engineering and Environmental Sciences*, 24(6), pp. 419–428.

Aksoy, H. (2003) Markov chain-based modeling techniques for stochastic generation of daily intermittent streamflows, *Advances in Water Resources*, 26, pp. 663–671.

Aksoy, H. (2004) Pulse generation models for daily intermittent streamflows, *Hydrological Sciences Journal*, 49(3), pp. 399–411.

Aksoy, H. & Bayazit, M. (2000a) A model for daily flows of intermittent streams, *Hydrological Processes*, 14, pp. 1725–1744.

Aksoy, H. & Bayazit, M. (2000b) A daily intermittent streamflow simulator, *Turkish Journal of Engineering and Environmental Sciences*, 24, pp. 265–276.

Aksoy, H., Bayazit, M. & Wittenberg, H. (2001) Probabilistic approach to modelling of recession curve, *Hydrological Sciences Journal*, 46(2), pp. 269–285.

Fiering, M. B. (1967) *Streamflow Synthesis* (London: Macmillan).

Moran, P. A. P. (1954) A probability theory of dams and storage systems, *Australian Journal of Applied Science*, 5, pp. 116–124.

Moran, P. A. P. (1959) *The Theory of Storage* (London: Methuen).

Quimpo, R. G. (1967) Stochastic model of daily river flow sequences, *Hydrological Paper 18*, Colorado State University, Fort Collins.

Savarenskiy, A. D. (1940) *Gidrotekh. Stroit.*, 2, pp. 24–28.

Treiber, B. & Plate, E. J. (1977) A stochastic model for the simulation of daily flows, *Hydrological Sciences Bulletin*, 22(1), pp. 175–192.

Weiss, G. (1977) Shot noise models for the generation of synthetic streamflow data, *Water Resources Research*, 13(1), pp. 101–108.