# Appendix H

# Historical Analog Streamflow Forecasting Model

The Historical Analog model (*Yao and Georgakakos, 2001*) produces forecasts based on information derived from the historical record. The underlying idea is that similar climatic and hydrologic conditions are followed by similar conditions for some time in the future. More specifically, streamflows and other hydrologic quantities materialize as a result of a nonlinear hydro-climatic process orbiting around an unknown attractor set. Although this set is not easily definable, this premise leads to the following conjecture: If the process is presently at a certain point in its orbit, its position in the near future can potentially be inferred by observing the movement it experienced on similar occasions in the past.

For example, streamflows are the result of the rainfall-runoff process, and the values they assume over a certain time period depend on various hydro-climatic factors including watershed rainfall, temperature, and soil moisture conditions. Thus, if the climate-watershed system tends to revisit the neighborhood of certain conditions (states), it should also tend to generate similar streamflow patterns.

In keeping with the above, the historical analog model "searches" the historical record and selects several inflow traces which, at some time in the past, have started from conditions similar to those of the current inflow sequence. Each one of these traces is a possible future realization of the inflow process and all together constitutes a set on which to base probabilistic, multi-lead forecasts.

Thus, suppose that the present time is April 1st, and the previous days' inflows were $W_1$, $W_2$,...,$W_n$ , where subscript "1" represents the last day in March, "2" the day before that, etc., and n is a parameter related to the process memory. Namely, [$W_1$, $W_2$,...,$W_n$] is the most recent inflow sequence to April 1st. The next step in the implementation of the historical analog model is to retrieve all inflow traces of the same month and date as the $W_1$, $W_2$,...,$W_n$ from the historical record and compute their Euclidian distance, $E_j$ , from the current sequence:

$$E_j= [(W_1 - W_{1,j})+ (W_2 - W_{2,j} )\ldots+ (W_n - W_{n,j} )]^{0.5} , \; j=1, 2, \ldots, m,$$

where $W_{i,j}$ is the historical inflow of year j at the same calendar date as $W_i$; m is the number of years in the historical record; and $E_j$ measures the proximity of [$W_{i,j}$ , i=1, 2, ..., n] to the most recent inflows [$W_i$ , i=1, 2, ..., n]. A small value of $E_j$ implies that the $W_{i,j}$ sequence is in the neighborhood of $W_i$. The inflows following $W_{i,j}$ are known (as part of the historical record) and can be used as the forecasts of the inflows following $W_i$. To generate multiple forecast traces, one can rank the $E_j$'s from smallest to largest, select the top portion of the ranked list, and use the corresponding historical inflows following $W_{i,j}$ as possible future inflow realizations. It is noted that several other ways may be used to measure the proximity of the historical to the most recent streamflow sequences. The reasons for using this particular scheme are that it is easy to implement while providing a benchmark for comparing other forecast schemes.