APPENDIX 2

AMEC MEMORANDUM
SUBJECT: STOCHASTIC STREAMFLOW SIMULATIONS FOR THE OTOWI GAGE

MEMORANDUM

TO: Dr. Nabil Shafike, P.E., New Mexico Instream Commission

FROM: Subhrendu Gangopadhyay and Ben Harding
AMEC Earth & Environmental

SUBJECT: Stochastic Streamflow Simulations for the Otowi gage

DATE: June 24, 2008

In this memorandum we present the results of stochastic simulation of Otowi ensemble flows developed using two approaches. The first set of simulations is based on the non-homogeneous Markov Chain (NHMC) algorithm of *Prairie et al* [2008]. The second set is based on the homogeneous Markov Chain (HMC) algorithm [*Haan*, 1977] using transition probability matrix developed by *Ward* [2008] for the Otowi gage.

Non-homogeneous Markov Chain Simulations

The non-homogeneous Markov Chain algorithm with an application to the Lees Ferry gage (Colorado River, AZ) ensemble streamflow simulation is described in the paper by *Prairie et al.* [2008]. Conceptually, this approach provides a method to combine transient state information (wet/dry) from paleo reconstructions of hydrologic markers (e.g., streamflow, Palmer Drought Severity Index – PDSI) with observed streamflow records to generate streamflow ensembles. Please refer to *Prairie et al* [2008] paper for the detailed description of this algorithm.

The starting point to apply this algorithm for the Otowi simulations was to study the correlation between the gridded paleo PDSI [*Cook et al*, 2004] for grid points in the vicinity of the Otowi gage and the natural streamflow of the Otowi gage. Naturalized streamflow for the Otowi gage is presently available for the period 1940-2007. The gridded PDSI data is available from the NCDC (National Climate Data Center) website, http://www.ncdc.noaa.gov/paleo/newpdsi.html. Results of the correlation analysis are shown in Figure 1. The grid point (longitude, 110.0W; latitude, 37.5N; number, 103) was found to have the highest correlation with the Otowi naturalized flow. Paleo reconstructed PDSI [*Cook et al*, 2004] for this grid point (number, 103) was used to define the flow states of the Otowi gage – 604 years of PDSI values for the period 1400-2003. Two states (dry – 0, wet – 1) were assigned to each of the 604 years of the PDSI time series. The mean and median PDSI for the 604 year period was found to be, -0.5490 and -0.3150 respectively. Consequently, positive PDSI values (PDSI > 0.0) were assigned to state wet, and PDSI values less than and equal to zero were assigned to state dry. This resulted in a binary [0, 1] time series of length 604.

This binary time series was used to develop the transient transition probability matrix corresponding to four transitions, dd (dry-dry); dw (dry-wet); wd (wet-dry); and wet-wet (ww). The transient transition probability plot is given in Figure 2. The transient transitions were developed using a non-parametric kernel smoothing algorithm following *Prairie et al.* [2008]. It should be noted in Figure 2 that certain epochs have a higher (lower) probability of transitioning from dry (wet) to wet (dry) states. The NHMC algorithm then randomly selects an epoch of given length (100 in this case), randomly selects the initial state (dry or wet), and marches through the transitions [*Haan,* 1977] to get the simulated 100 year state time series. In addition, the historical Otowi naturalized flow time series is also classified into two states. Median flow over the 1940-2007 period, 833 kaf, was used to assign dry and wet states to each of the 68 years. If annual flow for a given year was greater than 833 kaf that year was identified to be wet, else the year was in dry state. Using the above simulated states from the paleo PDSI, the flow state from the observed period and flow magnitudes of the observed period, a conditional K-nearest neighbor (K-nn) bootstrap was performed [*Prairie et al.,* 2008] to generate an ensemble member of years and hence flow. The step of randomly selecting an epoch and conditional K-nn was repeated to generate a 1000 member ensemble of traces, each of length 100. The simulated years and flows are available in files *NM_paleo.txt* and *NM_paleo_flow.txt* respectively (the files are in matrix format, 100 rows x 1000 columns) in directory *NHMC.*

To test the simulated flows a suite of basic distributional statistics were computed from the flow ensemble including the ensemble member (i) mean, (ii) standard deviation, (iii) coefficient of skew, (iv) maximum, (v) minimum, and (vi) lag-1 autocorrelation. Drought statistics include, (i) average deficit, (ii) maximum deficit, (iii) average drought run intensity, and (iv) maximum drought run length. The drought statistics were derived using the respective median values as the threshold (drought – below median flows).

The results are displayed as boxplots in Figures 3 and 4, where the box represents the interquartile range (IQR) and whiskers approximate the 5th and 95th percentile of the simulations and outliers are shown as open circles beyond the whiskers. The statistics of the observed record are represented as a filled red triangle. Performance on a given statistic is considered good when the observed statistic of interest falls within the box (i.e., IQR) of the boxplots, while increased variability is indicated by a wider boxplot.

Generally, in all cases the observed record (red triangle) falls within the IQR of the boxplots except the drought statistic, maximum run length. The maximum drought run length in the observed flow record is 4 years (1953-1956), the maximum drought length from the PDSI record is 14 years (1870-1883). The NHMC is able to generate flows that span extended drought periods not experienced in the observed period. This we believe is valuable to the ISC in using these simulated flows in the URGWOM model.

In the next section we present the results from the homogeneous Markov Chain simulations using transition probability matrix from *Ward* [2008].

Homogeneous Markov Chain Simulations

Markov chain simulation using a single (hence the notion of homogeneity) transition probability matrix is given in *Haan* [1977]. The transition probability matrix (four state - very dry, dry, wet, very wet) used in simulating the Otowi flows is based on the matrix derived using 2000-year paleo record (Table 4, *Ward*, 2008). The transition probability matrix is given in Table 1. The flows for the observed, 1940-2007 period were assigned the four states based on thresholds defined in Table 6 of *Ward* [2008] (2000-year record upper bounding flows – very dry (444.5 kaf); dry (843.1 kaf); wet (1463.2 kaf); very wet (maximum)).

The simulated years and flows (1000 simulations each of length 100) are available in files *Ward_SimYear.out* and *Ward_SimFlow.out* respectively (the files are in matrix format, 100 rows x 1000 columns) in directory *Ward*. Similar to the NHMC simulations, basic distributional statistics and drought statistics were calculated from the simulated flows and the results are presented as boxplots in Figures 5 and 6. Generally, the observed statistic (red triangle) is within the IQR of the boxplots, except for the statistic maximum drought run length. Using the homogeneous Markov Chain model we were able to generate flow traces with longer drought lengths (median length of 7 to 8 years) than the observed maximum of 4 years (1953-1956). This again, we believe is of value to ISC in using these simulated flows in the URGWOM model.

Transition Probability Comparison

As part of the above analyses we also compared the transition probability matrix (four state – very dry, dry, wet, very wet) of *Ward* [2008] derived using 2000 year of paleo record with the one we developed (four state) using the 604 years (1400-2003) PDSI values for grid 103 (longitude, 110.0W; latitude, 37.5N; *Cook et al.,* 2004). This may not be the case for the Otowi gage, but from our experience on working with tree-ring chronologies on the Colorado River for the Lees Ferry gage (e.g., *Gangopadhyay et al.,* 2008), limited tree-ring chronologies are available prior to c. 1400, and we believe any hydrologic reconstruction prior to this time will likely have large uncertainties. This was the motivation for using the PDSI time series starting in 1400.

The definition of the four states for the above PDSI data as used in this comparison is given in Table 2. We believe these state definitions to be objective as it is based on the standard PDSI state classification system [e.g., http://weather.nmsu.edu/drought/053102/drought6.pdf]. Using the PDSI classification from Table 2, each of the 604 years (1400-2003) were assigned a state – very dry, dry, wet and very wet. The four state transition probability matrix is given in Table 3. Overall the transitions in Table 3 compare well with the transitions in Table 1 [*Ward,* 2008] except that we found no transition in the 604 year record and based on our state definition (Table 2) from very dry to very wet state. It is intuitive that for the semi-arid Rio Grande valley to move from a very dry state (very low soil moisture) to a very wet state (very high soil moisture) within a year is mostly unlikely. From a steady state transition probability estimation analysis (alternatively, Table 2 column labeled – *Count Pct)* we found that the system will be in very dry, dry, wet and very wet states 17%, 39%, 36%, and 8% of the time respectively. *Ward* [2008, Table 5] estimated these states to be 17% (very dry), 35% (dry), 33% (wet), and 15% (very wet). Again these state percentages compare well, except for the very wet state, where we get nearly half (8%) as compared to 15% by *Ward* [2008]. No flow simulations were carried out using our transition probability matrix (Table 3) at this time.

Future Work

The following additional work is proposed to be carried out on this project.

(1) Use the transition probability matrix (Table 3) to generate flow ensembles using the homogeneous Markov Chain (HMC) algorithm.

(2) Derive two state homogeneous transition probability matrix (e.g., average of the NHMC transient transition probability matrix) and generate flow ensembles using the *Prairie et al.*[2008] conditional bootstrap algorithm.

(3) Repeat (2) above using alternative PDSI thresholds corresponding to mean and median flows from the observed period. The current transition probability matrix is derived using PDSI value of zero as the dry/wet transition threshold.

(4) Perform NHMC simulations using the additional two state transition probability matrix derived in (3) above.

(5) Perform HMC simulations using a three state (based on observed Otowi flow terciles) transition probability matrix with unconditional and possibly conditional resampling of flows. Current conditional resampling codes are based on a two state system, will need additional development for extending it to more states.

(6) Evaluate cross-validation statistics for selected Markov Chain simulations.

References

Cook, E.R., C.A. Woodhouse, C.M. Eakin, D.M. Meko, and D.W. Stahle, 2004. Long-Term Aridity Changes in the Western United States. *Science*, Vol. 306, No. 5698, pp. 1015-1018, 5 November 2004.

Gangopadhyay, S., B. Harding, B. Rajagopalan, and T. Fulp, 2008. A Non-Parametric Approach for Paleo Reconstruction of Annual Streamflow Ensembles. *Water Resources Research*, submitted.

Haan, C.T., 1977. *Statistical Methods in Hydrology*, Iowa State University Press.

Prairie, J, K. Novak, B. Rajagopalan, U Lall, and T. Fulp, 2008. A Stochastic Nonparametric Approach for Streamflow Generation Combining Observational and Paleo Reconstructed Data. *Water Resources Research*, in press.

Ward, T.J., 2008. Generation of Synthetic Flow Sequences for Ecological Modeling in the Middle Rio Grande, NM. *Internal discussion draft report, PHVA / Hydrology Ad Hoc Work Group.*

Table 1. Transition probability matrix used in homogeneous Markov Chain simulations.

	Very Dry	*Dry*	*Wet*	*Very Wet*	**Steady State**
Very Dry	0.22	0.42	0.29	0.07	17%
Dry	0.20	0.41	0.29	0.10	35%
Wet	0.16	0.31	0.35	0.19	33%
Very Wet	0.07	0.25	0.40	0.29	15%

Source: Ward [2008].

Table 2. Definition of four states for the Cook et al. [2004] PDSI data.

State	*Description*	*Definition*	*Count*	*Count Pct*
1	Very Dry	PDSI ≤ -3.00	105	17%
2	Dry	-2.99 ≤ PDSI ≤ 0.00	233	39%
3	Wet	0.01 ≤ PDSI ≤ 2.99	220	36%
4	Very Wet	PDSI ≥ 3.00	46	8%

Table 3. Four state transition probability matrix using state definition from Table 2 and Cook et al. [2004] PDSI time series for grid 103 for the period 1400-2003.

	Very Dry	*Dry*	*Wet*	*Very Wet*
Very Dry	0.27	0.52	0.21	0.00
Dry	0.20	0.42	0.33	0.05
Wet	0.13	0.30	0.46	0.10
Very Wet	0.02	0.33	0.39	0.26

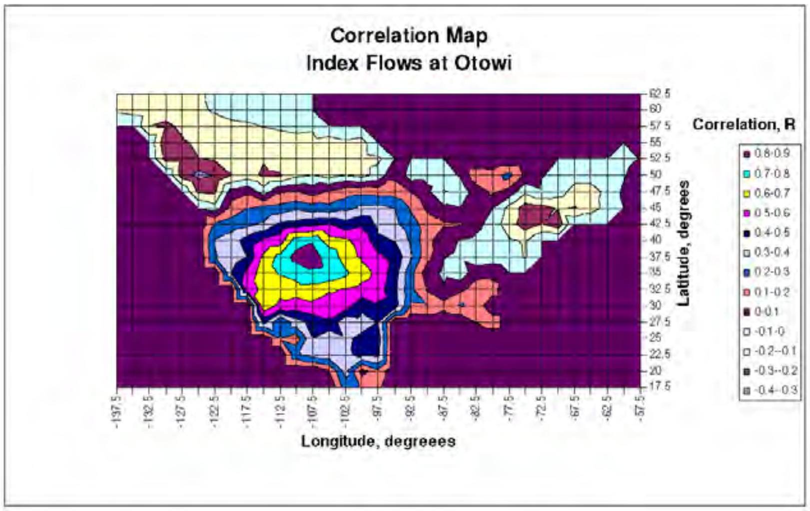

Figure 1. Correlation of gridded PDSI to Otowi index flows.

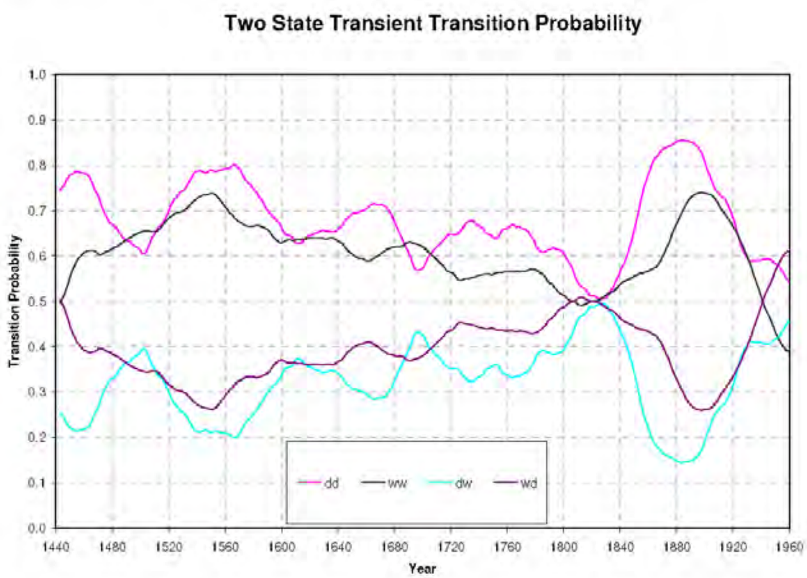

Figure 2. Two state – four transitions (dd, dw, wd, ww) transient transition probability based on the binary dry/wet time series of length 604 (1400-2003).

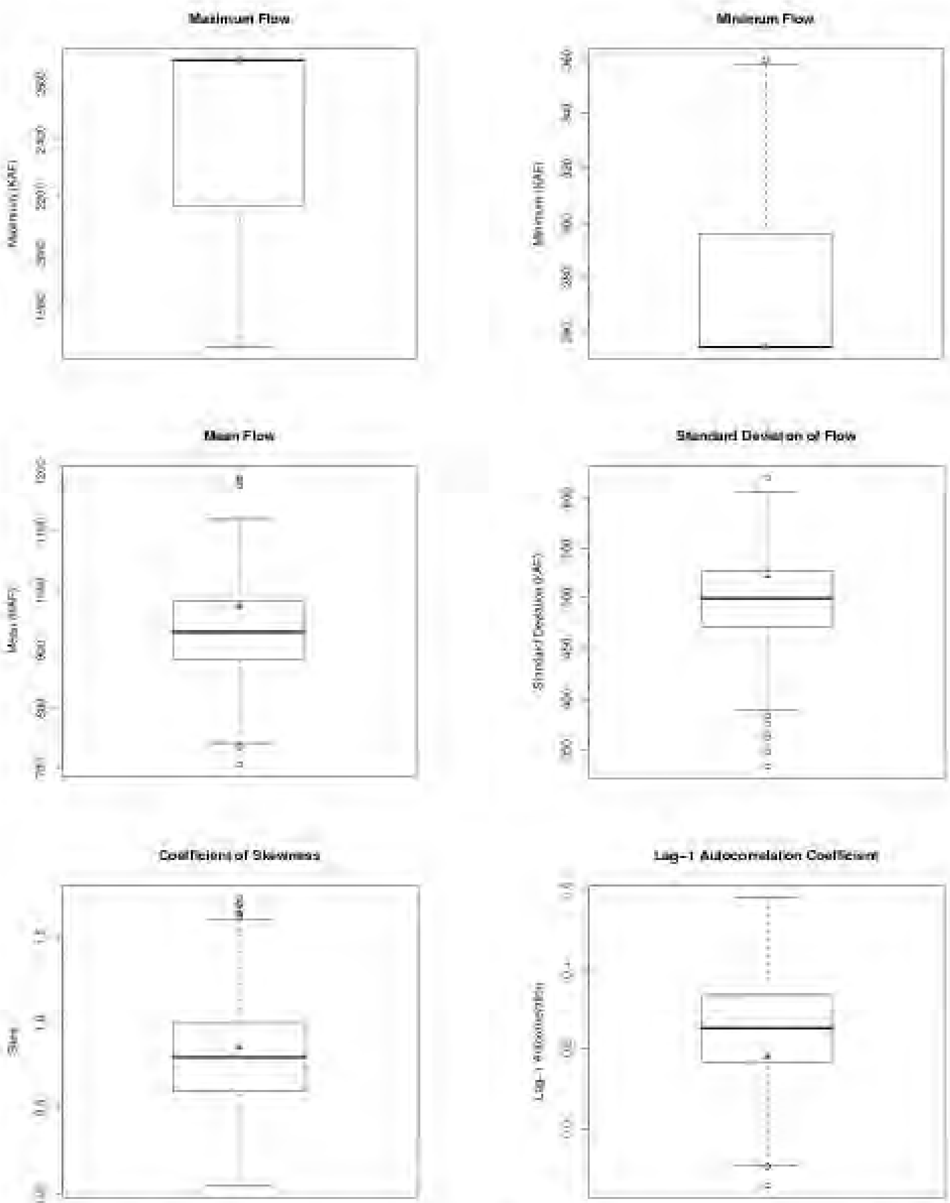

Figure 3. Distributions of statistics of annual flows for the traces generated by the NHMC.

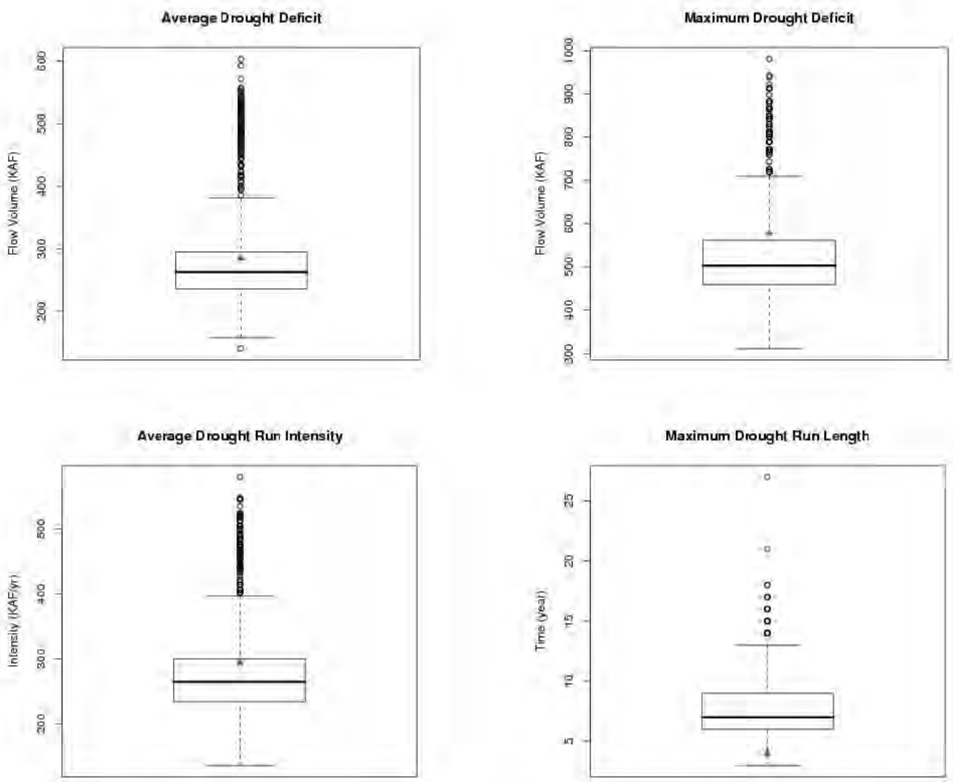

Figure 4. Distributions of statistics of drought for the traces generated by the NHMC

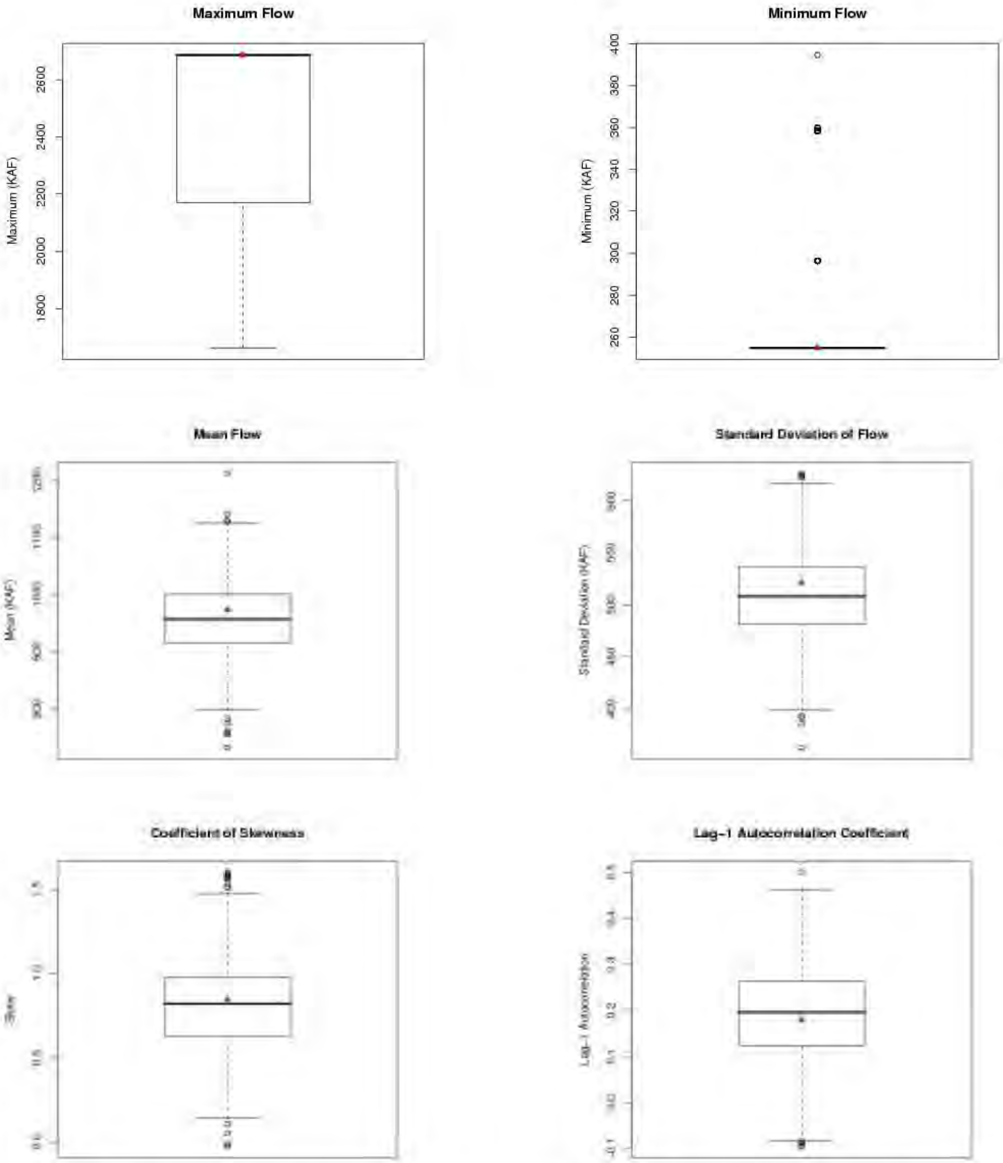

Figure 5. Distributions of statistics of annual flows for the traces generated by the HMC.

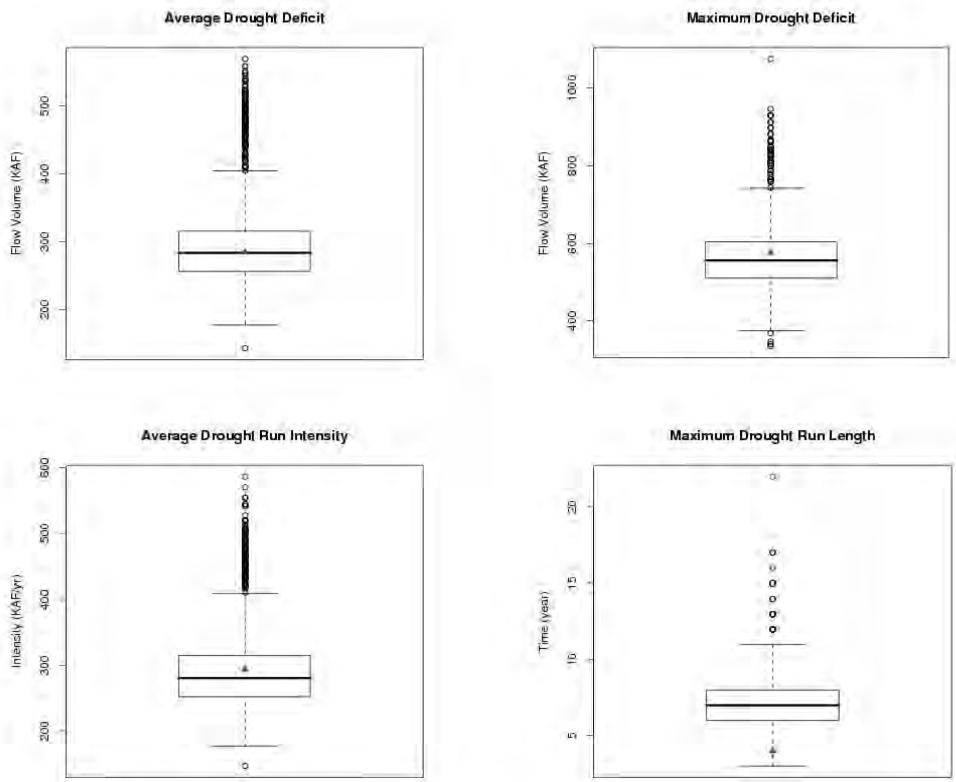

Figure 6. Distributions of statistics of drought for the traces generated by the HMC