GENERATION OF SIMULATED MONTHLY RAINFALL ACCUMULATIONS AND ASSOCIATED MONTHLY EVAPORATION DEPTHS FOR USE IN THE MANAGEMENT OF SMALL RESERVOIRS AND WATER HARVESTING FOR CENTRAL TEXAS

GEORGE R. HERRMANN, P. E., P. H., CFM
Transportation Engineer,
Texas Department of Transportation
4502 Knickerbocker Road
San Angelo, Texas 76904
325-947-9321
gherrma@dot.state.tx.us

and

THEODORE G. CLEVELAND, Ph.D., P.E.
Associate Professor
Department of Civil and Environmental Engineering
Texas Tech University
10th and Akron, Room 203F
Lubbock, Texas 79409
806-742-2801
theodore.cleveland@ttu.edu

ABSTRACT

A stochastic simulator is presented that projects monthly rainfall based on observed monthly rainfall for selected cities in central Texas. The simulator is built in the R environment (R Development Core Team, 2008) to take advantage of the built-in statistical and graphical functions and existing external packages.

The simulator is demonstrated using (1) an assumption of serial independence in monthly totals, and (2) a model of month-to-month dependence for sequential months using selected copulas provided by Asquith, W.H. (2009). Using these two dependence structures, associations between monthly rainfall and concurrent monthly evaporation depths are examined. The evaporation values also are generated using selected copulas.

The realization trajectories (monthly moving sums) for repeated simulations display three distinct regimes: some trajectories are inherently “drought” trajectories, whereas a larger fraction are “ordinary climate,” and the remaining represent extraordinarily large accumulations of rainfall. These trajectories, ordered yearly totals, moving sums, and moving averages are discussed in the context of water resources management for the three regimes (drought, ordinary, and wet).

Key Words: Rainfall Harvesting, Stochastic Rainfall Simulation, Arid Hydrology, Evaporation

INTRODUCTION

In arid and semi-arid areas such as west Texas, water is a critical resource the supply of which is highly variable. The management of this resource is becoming increasingly critical as population in these areas increases. Climate tends to exhibit some serial persistence, which manifests itself as extended periods of rainfall extremes, both high and low. While high rainfall periods do not imply water supply problems, they may well imply flood potential and the potential for geotechnical problems due to unusually high soil moisture.

Low rainfall periods represent the most challenging situation for the management of water supply, demand, and usage. Traditional methods of hydrologic modeling are not necessarily well adapted to the representation of extended drought conditions, and the management of water supply may have to rely on real-time response to conditions. This response is usually influenced by past experience, public perception, and chance of rainfall.

The harvesting of storm runoff is the chief source of water in many arid areas. Scale may vary considerably; most often moderate to large reservoirs serve as principle supply. However, as demand for water increases, the unreliability of surface water in the event of extended drought is acknowledged, and the actual economic value of water as a commodity is recognized, the construction of smaller projects, often called "water harvesting", are increasing. The principle use of "harvested" water is often the irrigation of landscaping vegetation. The irrigation of landscaped areas constitutes a significant fraction of municipal water consumption, particularly in areas of low rainfall.

In most such areas, not only is rainfall relatively sparse, but evaporation potential may be several times rainfall potential (as depth). Water captured in surface impoundments cannot be stored indefinitely; it may well be consumed by evaporation before it is needed.

Both rainfall potential and evaporation potential are seasonal in nature, as is the demand for water for landscaping irrigation. Rainfall is often scarce enough that daily or weekly estimation of supply and demand is unnecessary, however the ability to estimate both accumulated rainfall and accumulated evaporation at the end of each month could be very beneficial from the standpoint of the design of small water supply impoundments, and for the development of management strategies to maximize the benefits of captured water.

Toward that end, the authors have developed a stochastic monthly rainfall simulator for the generation of time-series of monthly rainfall accumulation depths, based on historical data. Two variations on this theme were examined; one assuming that each individual monthly depth of rainfall was a random variate drawn from a distribution fitted to historical data for that month, and one assuming that there is an association or dependence from one month to the next. These simulators have been developed using real monthly rainfall data for San Angelo, Texas, although the method could be used for any city for which data is available to parameterize the model. Monthly rainfall data can be obtained from the National Climatic Data Center, commercial sources such as HydroSphere, and occasionally other, open sources. Evaporation data for all of Texas is available from the Texas Water Development Board website. Concurrent rainfall and evaporation data are necessary to parameterize the copula linking rainfall and evaporation. Such concurrent data was available for 49 years for San Angelo, while 166 years of data were used to develop the rainfall distribution parameters. Gaps in the published rainfall data left necessitated the truncation of that series to a contiguous series of 61 years for the parameterization of the rainfall copula.

Assuming Serial Independence

Monthly rainfall accumulation data for San Angelo was obtained and screened for suitability. These data contain many values shown as "trace", i.e. where there was rainfall, but not a measurable amount. Trace values were set to zero for this exercise. There were 116 complete years (not a contiguous stream- some years were missing or incomplete) of suitable data. Using the "lmomco" R package for L-moment statistics, the L-moments and L-moment ratios for each month were computed, and the ratios plotted on an L-moment diagram to aid in the selection of a suitable statistical distribution. These values indicated that the generalized Pareto distribution (GPA) was the strongest candidate. The GPA is a relatively flexible three-parameter distribution. This distribution has some philosophical attraction also; the Pareto distribution arises from the study of asymmetrical distribution of such things as wealth, i.e. that a large fraction of wealth resides in the possession of a small fraction of people. Rainfall for the area in question exhibits a similar quality in that statistics seem to be strongly influenced by a small fraction of very large values. Overall, rainfall is highly variable in this region, and extended periods of low values (drought) have been seen.

A pre-processing routine which reads a data file and computes and stores parameters for the GPA distribution for each month, based on L-moments, was written. A series of functions and walking routines were written that step sequentially through a "year" by generating a uniformly distributed random number on the interval (0,1) and passing that number through the quantile function parameterized for the appropriate "month", accumulating those values into an annual total, plotting them as a year-long string of monthly values, and repeating the process for a specified number of years. Each individual month is stored in sequence, and the annual totals are stored, both for further statistical examination.

The result is a vector of values, each of which is independently drawn from a distribution parameterized for a given month, and a vector of sums of each 12 month cycle, representing a year.

Computing the summary statistics (mean, standard deviation, and interquartile range) for the sample, and the same summary statistics for the simulated data, the sample data show greater dispersion than is obtained by simulation. This is to be expected; real data are produced by more complex processes than are represented by simulation. However, even though the magnitude of variation is not simulated, the basic underlying structure of the data is, and from that there is still utility in the simulated data.

Simulating Serial Dependence

Recognizing that there exists some “persistence” to climate at any particular time, which expresses itself by serial dependence or association from month to month, it was decided to attempt to represent such dependence by the use of a copula to connect each month to the next. In this way, some influence is manifest between each month and the month preceding it, although the entire range of potential values is still open, should conditions favor it.

For the purposes of comparison, the data were restricted to contiguous years, such that there would be no “gaps” where there may not be a valid association. Values of Kendall’s tau and Spearman’s rho were computed between the data series for successive months. Values for the December to January association were computed by offsetting values such that each December was compared to the following January, rather than the January of the year in which it occurred. These taus and rhos indicated varying degrees of dependence between months, although none of them are extremely strong. Of the 12 values each of tau and rho, 11 are positive and one negative. While the magnitudes are small, if there were no association at all, it could be expected by sampling error that the algebraic sign of tau and rho values would be relatively evenly split. In this case, 11 out of 12 are positive, indicating that a weak positive association might exist.

It was decided to represent this association by the use of a Plackett copula. A Plackett was chosen rather than another copula because if it’s flexibility and ease of parameterization. The Plackett is a one-parameter copula. That parameter, usually called theta, can be estimated from Spearman’s rho.

In this case, the pre-processing routine was modified to read both the entire data series (for parameterizing the distributions) and a truncated series of contiguous data (for parameterizing the copulas) and to compute and store the theta values. In the simulation process, rather than individually drawing from the uniform distribution and passing to the quantile function, an entire vector of monthly probability values was generated. This was done by beginning with a single random value, then passing it to the copula simulator, which drew the next value based on the first, the association between the first and second based on the parameterized copula, and a uniform random variate. In this way, each monthly value but the first is weakly dependent on the previous monthly value, combined with a random component. An entire series of probability values, cycling through the copula parameters for each monthly serial association, was generated for the desired period.

These probability values were then sequentially passed through the quantile function for the GPA distribution in cycle for the appropriate month to obtain a vector of sequential values. This vector was broken into “years” for accumulation into “annual” sequentially summed series.

The series of annual values (Figure 2) was plotted and can be compared to those for the measured data. It can easily be seen that the simulated data still does not exhibit the variation evident in the measured data. The geographic area in question, west-central Texas, is subject to mixed populations of rainfall generating events, and this mixed behavior is one explanation of the larger variation in measured data than can be simulated. The real data exhibits periodic or episodic persistence on the scale of decades; the record length does not allow simulation of that effect. However, over reasonable periods, the simulation routine developed by the authors offers at least some ability to approach water management problems.

The series of values generated by accounting for serial association in this way exhibited summary statistics with more dispersion than from those with month-to-month independence. These statistics more nearly match those for the real data series.

Routines were written to compute moving sums and moving averages for any desired number of months, and to plot empirical distributions of those sums and averages. Moving sums intuitively should be a measure of “persistent state” of recent climate, i.e. drought or wet periods will show low or high moving sums, respectively.

The Spearman’s rho value for one month-to-month association (April to May) was a small negative number. A negative rho indicates a counter-association. While the magnitude of rho for that month is small and results in a small negative value of the copula parameter, the overall effect of having a negative copula parameter is to introduce additional variability, a yearly “kick” to the overall variation in rainfall values. This is very interesting, and suggests that manipulation of the negative parameter on a periodic basis, could simulate global periodic phenomena such as the El Nino/Southern Oscillation.

EVAPORATION

Evaporation from the surface of a reservoir in arid and semi-arid areas is often quite significant. In order to manage a small reservoir such as might be used in water harvesting, this loss must also be estimated; otherwise the operators might over-commit their harvest water, and face a dry reservoir when anticipate water need.

Monthly evaporation totals for San Angelo were obtained for a period of 55 years. Intuitive suggests that monthly evaporation total may be associated with monthly rainfall total, so the values of Kendall’s tau and Spearman’s rho were computed between rainfall and evaporation for the period of evaporation record in order to examine that conjecture. When compared to the taus and rhos for month-to-month rainfall associations, the values for rainfall to evaporation are considerably stronger, although consistently negative. Negative values of these statistics indicate inverse dependence, or counter-association. Large rainfall values associate with small evaporation values (relative to their overall distribution). This observation makes sense from a thermodynamic standpoint- rainfall is usually accompanied by cloud cover, cool temperatures, and high humidity, all of which reduce potential evaporation.

An examination of evaporation data by an L-moment ratio diagram indicates that the appropriate distribution for this data should be a generalized extreme value (GEV) rather than the GPA that was used for rainfall. In the same manner as with the monthly rainfalls, distribution parameters for monthly evaporation values were computed and stored. Once again choosing the Plackett copula for the same reasons as previously, the parameters for the copula for all months were computed. For evaporation, the association chosen was not month-to-month serial, but that between rainfall and evaporation. Each monthly rainfall probability was fed to the copula simulation routine and an evaporation probability computed based on rainfall probability, the copula parameter, and a uniform random variate.

In this way, a statistically based dual series of probabilistically associated, sequential monthly values of rainfall and evaporation of arbitrary length can be generated. These series can then be used in a pond geometry/withdrawal management water balance model to investigate optimum pond geometry, withdrawal protocol, drought duration, drought management, water availability, and possibly contribute to impounded water quality estimation.

The overall distribution of both rainfall and evaporation values produced by the simulation process are surprisingly close to the distribution of measured values. A small number of anomalously high rainfall values (attributable to tropical cyclones that reach the area occasionally, but rarely) dramatically increase the maximum measured values. High outliers such as these are not heavily weighted in the L-moment distributional fitting process, and not reproduced in the simulation. This suppressed influence is the same phenomenon that results in a visible difference in dispersion between Figures 1 and 2.

DISCUSSION

After the generation of the simulated data series, it is useful to compare these series to the measured data in ways that have value in the context of water management. Figures 3 and 4 compare the measured and simulated empirical distribution of evaporation and rainfall (respectively). For the purpose of further comparison, routines were written to easily compute moving sums and moving averages of both rainfall depth and evaporation depth of any length (in months). Moving averages of rainfall for measured and simulated series are shown in Figures 4 and 5. Moving averages are a reasonable measure of water availability at any given time, and would tend to indicate prevailing conditions of wet climate or drought. Moving sums are an equivalent measure, simply indicating total rainfall for the desired period.

The distribution of moving averages for periods of 24 and 48 months for the measured data and simulated data are shown in Figures 7 and 8 respectively. It is curious to note that the variation in measured data is very high- there appears to be no real modal range. When compared to the measured data, the simulated data exhibits a flatter distribution through the middle of the data, although the ranges are similar. It is conjectured that periodic or persistent episodic behavior in real climate on the scale of decades might account for this difference. Such mid-range behavior cannot be imitated by the current method.

Nonetheless, for small reservoir design, where the storage of water for years is not a consideration, this variation on the order of years is of little consequence. Even for larger reservoirs where storage on the order of years is anticipated, droughts and wet periods are simulated with sufficient fidelity as to be useful.

USES

The graphs presented are based on 100 years of simulated data. For actual design purposes, it would be desirable to simulate several hundred, or even several thousand, years of data. Given a watershed contributing area, simple rational runoff model, basic reservoir geometry such as a stage-area relationship and a basic withdrawal protocol, the operation of a reservoir can be simulated on a monthly basis.

Monthly rainfall depth results in an input volume from the watershed to the pond. Evaporation from the pond combined with user withdrawals results in an end-of-the month volume and depth carried into the next month, where the process is repeated. The volumes of water leaving the outlet overflow works, lost to evaporation, withdrawn by users, and stored in the reservoir to be passed to the next month as an initial condition can all be tabulated. Adjustments to pond geometry and withdrawal protocol can be made and the computations re-run to compare and optimize these variables prior to construction. A much better idea of the actual return on investment can be made than results from traditional event-simulation models; more importantly, a risk-based performance of the harvest program becomes accessible to the operators.

Returning to Figures 1 and 2 of real and simulated annual series of monthly rainfalls, it can be seen that there are three zones of trajectories of series represented. The highest zone of sparse series represents years of abnormally high rainfall (wet periods, the middle, dense zone the ordinary rainfall condition, and the lower, also sparse zone represents years of abnormal drought. Because evaporation is high, often nearly three times precipitation depth, historical attempts to store water for long periods of time in surface reservoirs and shift availability to drought periods are chronically problematic. Years, or series of years, of high rainfall occur frequently enough to affect summary statistics and to impart a sense of well-being on the populace; but overall the prevalent and limiting condition is dry, such that the effective climate is more arid than represented by summary statistics.

CONCLUSIONS

This simulation model, while still in infancy, provides a framework for continued development in order to optimize water management facility design and protocols. No attempt has been made to simulate individual events; but monthly water balance is a more reasonable approach to water supply management. Further development of this process might involve the selection and parameterization of more ideal copula forms, if necessary, for each of the copula applications. The application of a trivariate copula for evaporation, which would exploit both the link between rainfall and evaporation and the monthly serial association seen in evaporation data, may prove to be practical and useful.

As stated earlier, the artificial manipulation of the difference between positive and negative copula parameters might allow for the simulation of weak periodicity.

The ability to easily generate synthetic series of monthly rainfall and evaporation values that are statistically based, linked by internal dependence, and of any arbitrarily great length should allow for virtually unlimited modeling of management methods, and should also provide considerable insight into the range of conditions that can reasonably be expected to be seen. Once parameterized with appropriate climatic data, such a model might prove useful in any area where water is of high value.

REFERENCES

Asquith, William H., R package lmomco
Austin, TX

Hosking, J. R. M. and J. R. Wallis, Regional Frequency Analysis: An Approach Based on L-Moments
Cambridge University Press, Austin, Texas, 1997

Nelsen, Roger B., An Introduction to Copulas, second edition
Springer Science & Business Media, LLC New York, NY, 2006

R Development Core Team, 2008, R—A language and environment for statistical computing
R Foundation for Statistical Computing, ISBN 3– 900051–07–0, version 2.9.0, Vienna, Austria ,
http://www.R-project.org

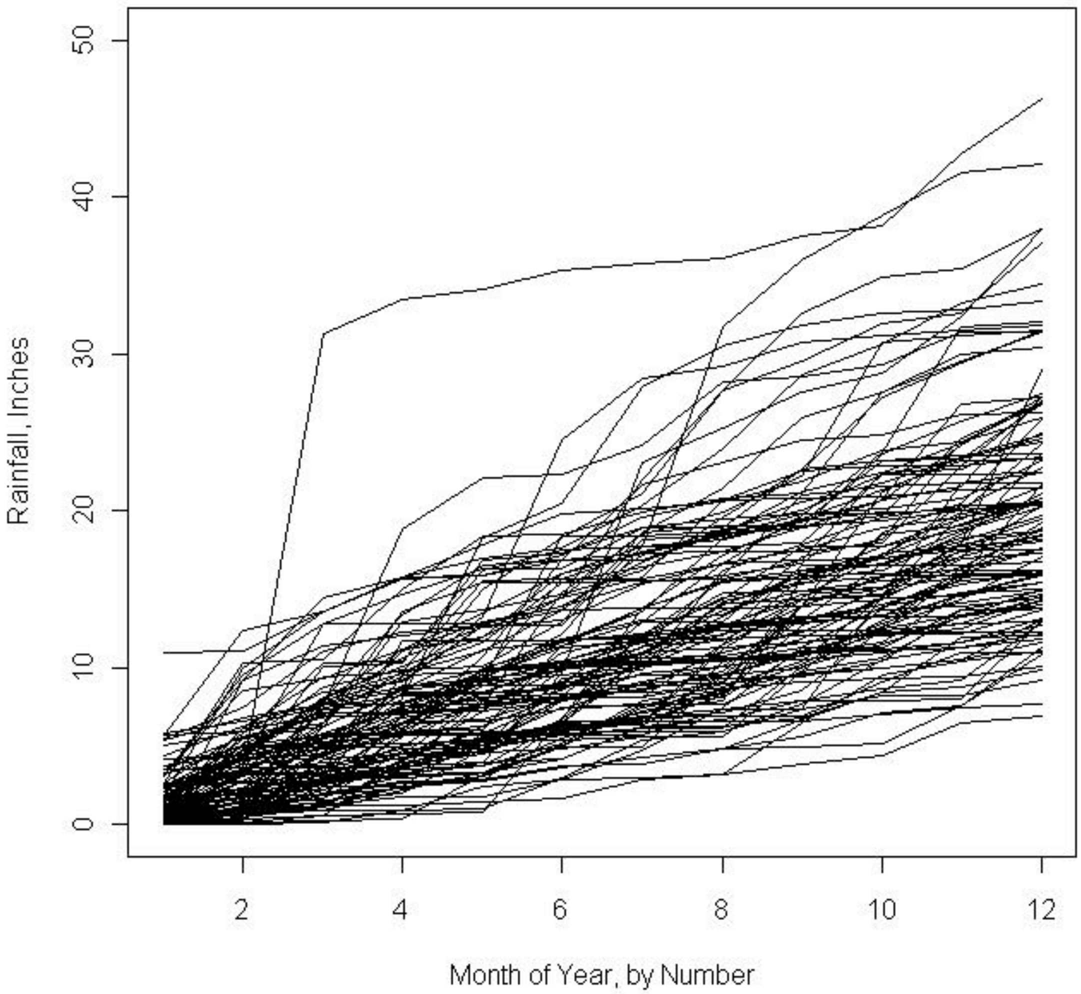

FIGURE 1
116 years (discontinuous) of actual rainfall data for San Angelo, Texas.
Shown as strings of summed monthly accumulations for each year.

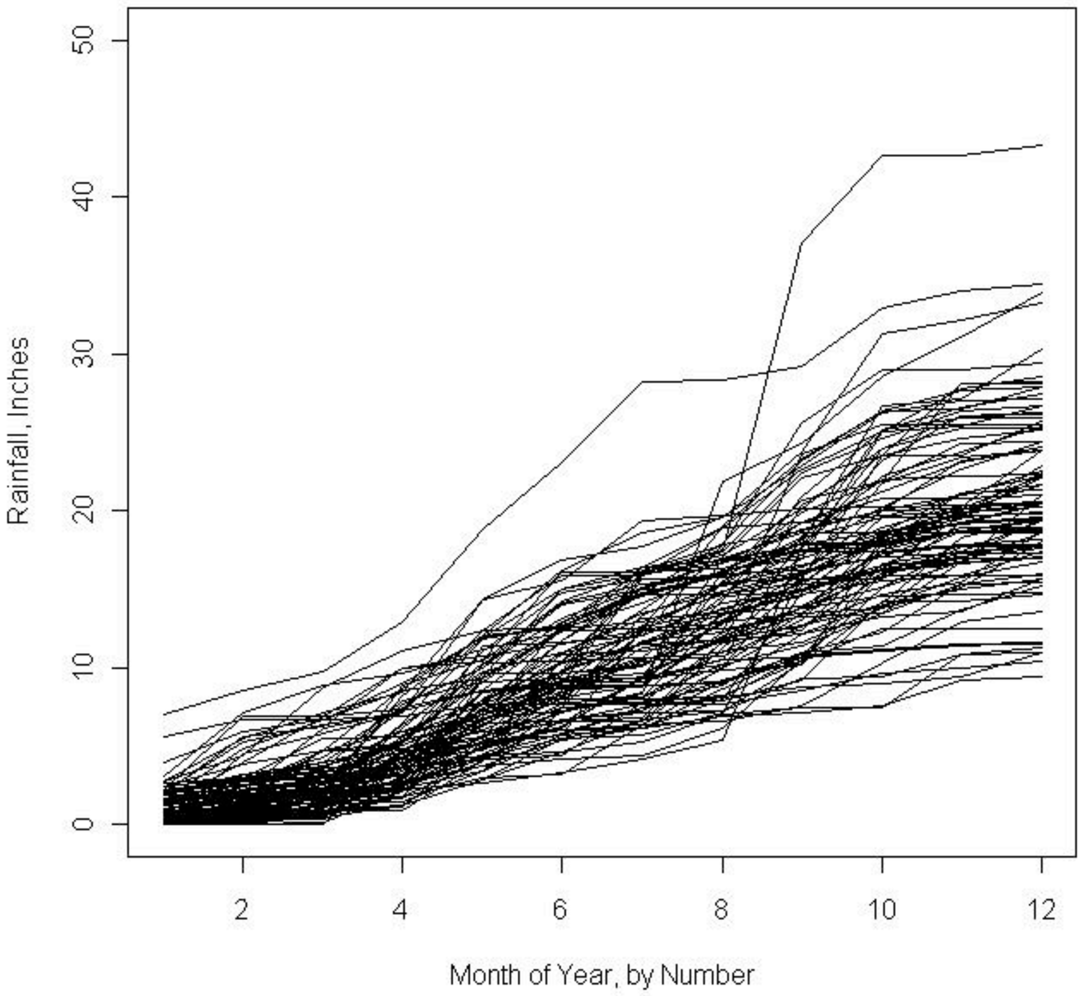

FIGURE 2
100 years of simulated rainfall depths, as annual strings of monthly accumulations.

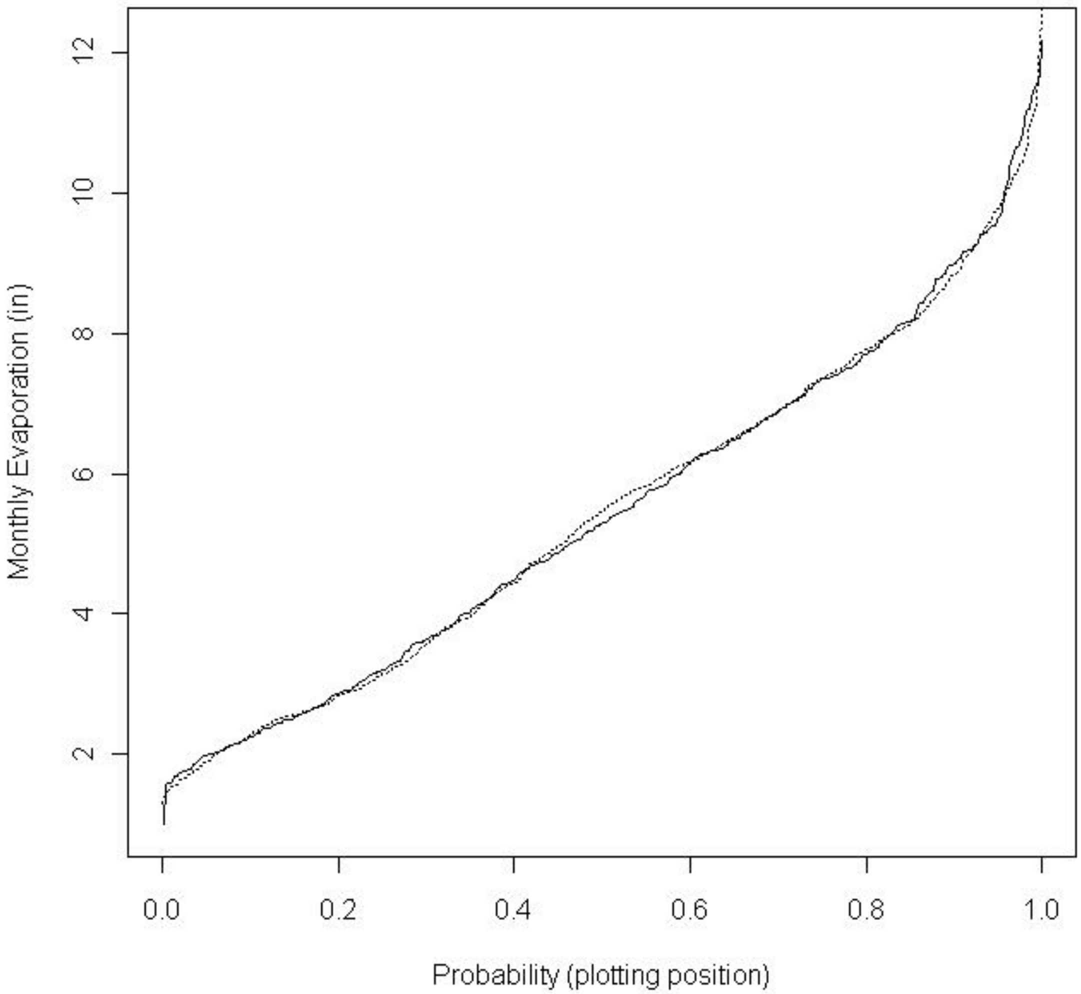

FIGURE 3
Distribution of measured (solid line) and
simulated (dotted line) evaporation values.

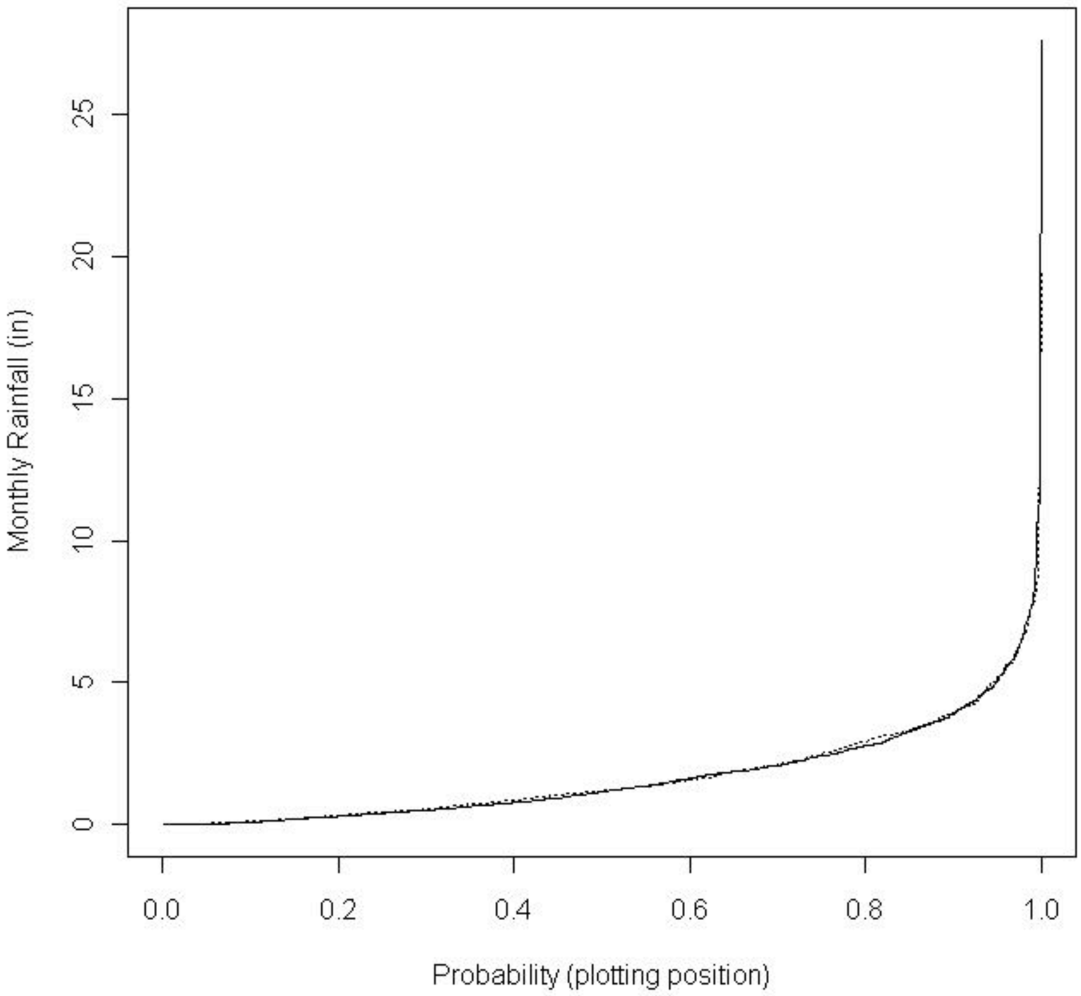

FIGURE 4
Distribution of measured (solid line) and
simulated (dotted line) rainfall values

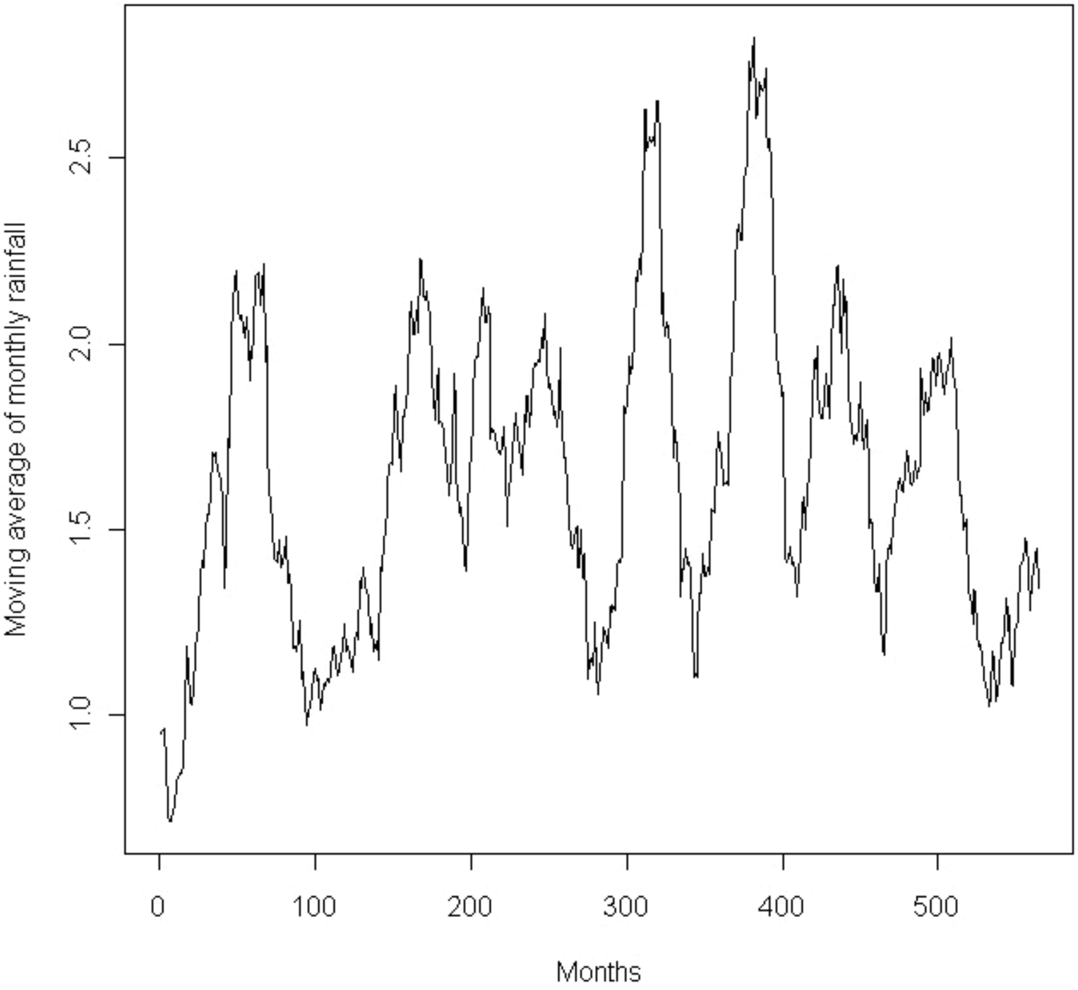

FIGURE 5
24 month moving average of measured rainfall
(49 years of contiguous record)

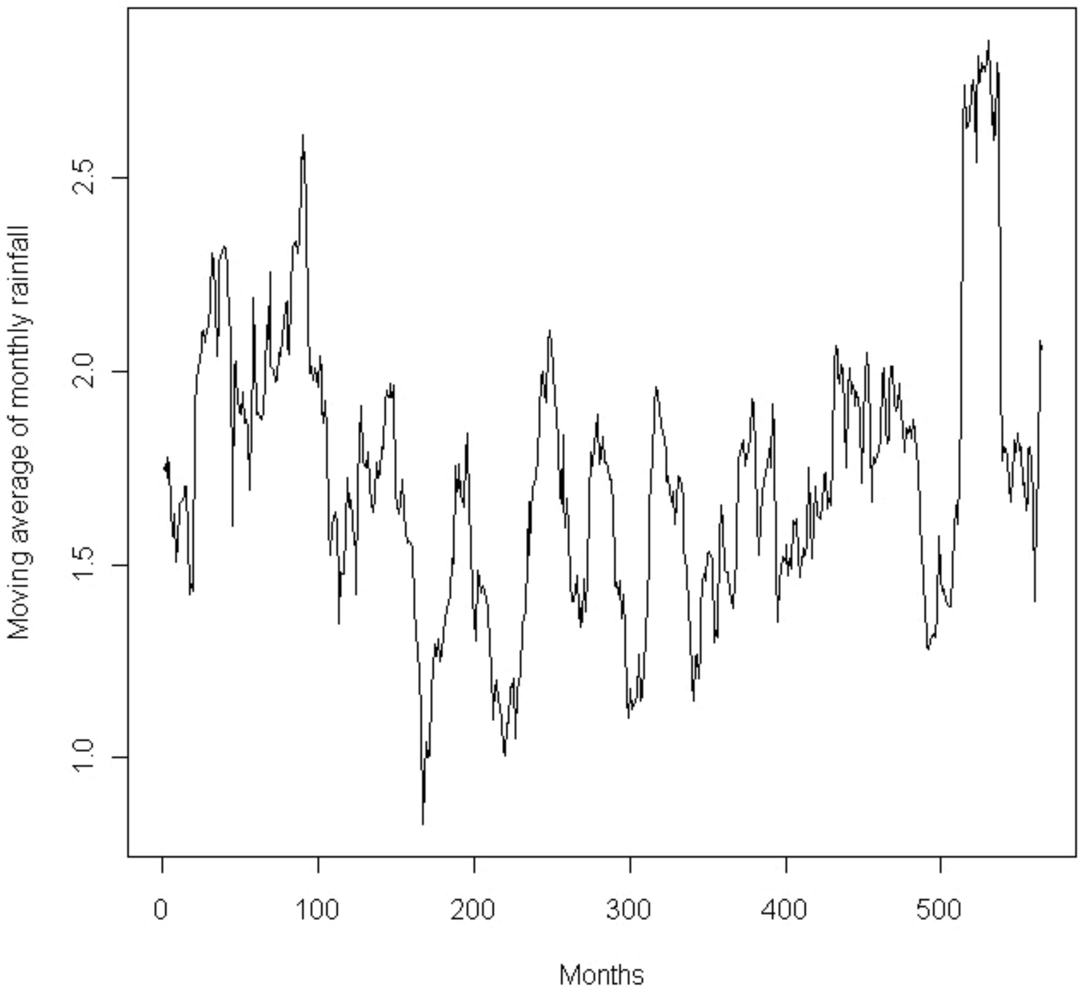

FIGURE 6
24 month moving average of simulated rainfall
(49 simulated years)

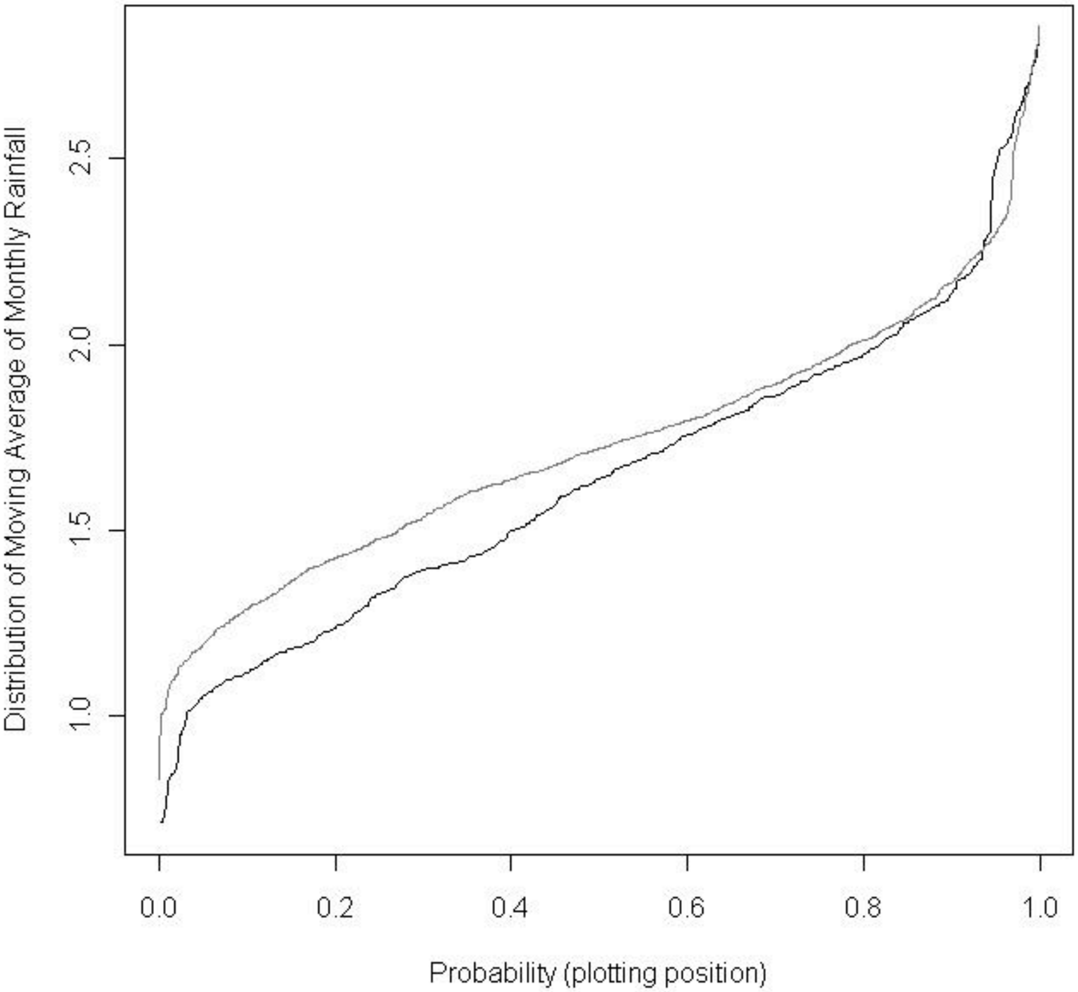

FIGURE 7
Distribution of 24 month moving average rainfall depths.
Black line is measured data, grey line is simulated data.

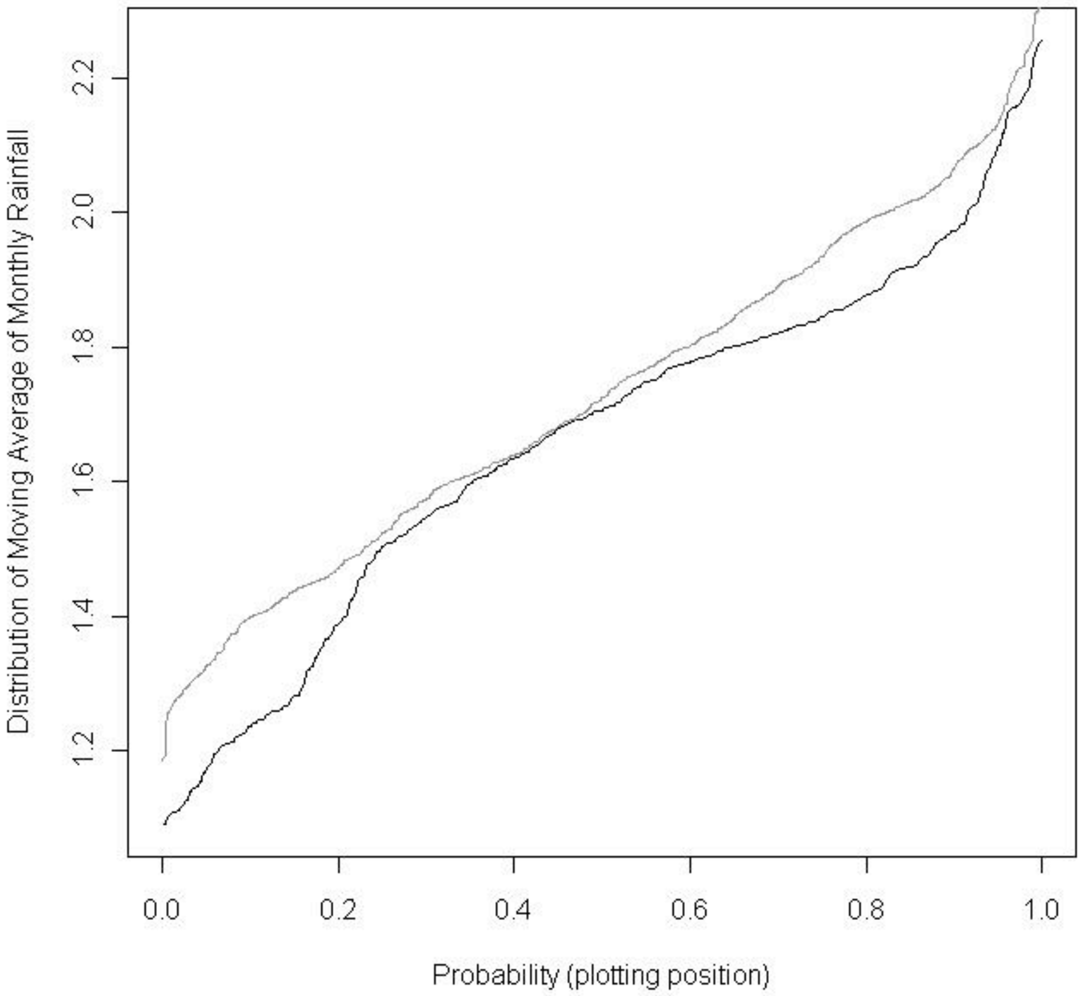

FIGURE 8
Distribution of 48 month moving average rainfall depths.
Black line is measured data, grey line is simulated data.