USE OF A STOCHASTIC WEATHER GENERATOR IN A WATERSHED MODEL FOR STREAMFLOW SIMULATION

by

ADAM N. HOBSON

B.A., Wesleyan University, 1997

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Master of Science

Department of Civil, Environmental and Architectural Engineering

2005

This thesis entitled:
USE OF A STOCHASTIC WEATHER GENERATOR IN A WATERSHED MODEL FOR STREAMFLOW SIMULATION
written by Adam N. Hobson
has been approved for the Department of Civil, Environmental and Architectural Engineering

____
(Balaji Rajagopalan)

____
(Kenneth Strzepek)

____
(Edith Zagona)

Date__________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptance presentation standards of scholarly work in the above mentioned discipline.

ABSTRACT

Hobson, Adam N. (M.S. Civil, Environmental and Architectural Engineering)

Use of a Stochastic Weather Generator in a Watershed Model for Streamflow Simulation

Thesis directed by Assistant Professor Balaji Rajagopalan

Current methods of streamflow forecasting rely on historic climate sequences that are inadequate in length and have statistical relationships that are difficult to fit and condition. Coupling a stochastic weather generator with a deterministic watershed model can provide better streamflow forecasts. This study presents a technique to couple a *k*-nearest neighbor stochastic weather generator and the Precipitation-Runoff Modeling System (PRMS) watershed model to simulate historic streamflow statistics and provide a framework for forecasting flows. The weather generator uses weather data in the Upper Truckee River Basin on the California and Nevada border (USA) to produce a simulated dataset. The simulated dataset sufficiently preserves the statistics of the historic record of precipitation and maximum and minimum temperature. Simulated weather variables were used as input to PRMS, which adequately simulated modeled historic streamflows. A conditioned forecast based on wet/dry years was used to demonstrate the utility of the forecasting framework.

ACKNOWLEDGEMENTS

This research and thesis would not have been possible without the help of many people. First and foremost, I would like to thank my advisor, Dr. Balaji Rajagopalan for all his assistance in getting this project together, supporting me through the process and believing in me to the end. Secondly, I would like to thank Dr. George Leavesley and Steve Markstrom at the U.S. Geological Survey, Water Resources Division in Lakewood, CO for securing funding for this project under the STEP program and supporting this work from the beginning. I wish to thank Dr. Edie Zagona and the Center for Advanced Decision Support for Water and Environmental Systems (CADSWES) for their support and use of their computing facilities. I would also like to thank Dr. Ken Strzepek for encouragement along the way.

Personally, this has been a long adventure and many people have been influential. You all know who you are and thanks for the support and keeping me sane during this journey.

Lastly, I wish to thank the person, who, at the beginning of this adventure was merely a friend of a friend, and now at the end, will be my wife. Thanks, Anne, this one's for you.

CONTENTS

Chapter

TABLES

FIGURES

CHAPTER 1

1. INTRODUCTION

Realistic and accurate streamflow forecasts are an essential tool for water resources planning and management. In many regions, agricultural, municipal, and environmental water uses place increased demands on limited and variable freshwater resources. In the Upper Truckee River Basin, like many basins in the Western United States, planning and management of water resources is particularly challenging because snowmelt from the Sierra Nevada Mountains is essentially the only source of streamflow in the semi-arid desert of Western Nevada. As a result, inter-annual variability of streamflow in such basins tends to be high. Streamflow forecasts can be used as a tool to facilitate effective basin management by providing accurate forecasts for water quality, volume, timing, and flow rates.

1.1. BACKGROUND

Simulated streamflows are used in conjunction with hydrologic and hydraulic models to generate flow forecasts which can then be used to assist with water resources management in a basin. The ensemble or extended streamflow prediction (ESP) procedure is one method used to forecast streamflows in hydrologic models used by the U.S. National Weather Service (NWS) and others [*Day,* 1985; *Leavesley et al,* 2002*; Mastin and Vaccaro,* 2002]. The ESP method assumes that past meteorological conditions are representative of what may occur in the future and uses historic weather data as input variables to initialize a deterministic watershed model. The model is then run in a simulation mode to forecast for the given time period. For short term streamflow forecasts (i.e. one-week), the NWS uses seven-day weather forecasts to drive the simulation model. For longer forecast periods (i.e. two-weeks, monthly, seasonal), weather forecasts are unavailable. Therefore, historic time series weather data from the representative time period are run through the watershed model to generate an ensemble of flows that can be analyzed statistically to give a stochastic forecast [*Day*, 1995].

The ESP approach has many advantages including simplicity in implementation and assumption that historic weather conditions can be surrogates for future weather. The method is easy to apply in any basin over any time scale where historic weather data are available. By utilizing historic data records, the spatial and temporal distributions of weather variables are inherently preserved. However, the drawback to this method is that it relies on the length of the historic record to create the members of the ensemble which, in many cases, does not provide enough series to be statistically significant. For example, if only 19 years of basin-wide weather data are available, only 19 flow hydrographs are possible, reducing the variability of the simulated flows. Additionally, if a conditioned forecast is required, such as one based upon large-scale climate indices or wet/dry relative forecast, the number of possibilities is further reduced. Given the inherently short historic records in many water-stressed basins and changing climate patterns a better streamflow forecasting tool is required. Modifying the ESP approach by using a weather generator to create statistically similar, synthetic weather data based on the historic record would increase the number of records from which to generate ensembles and improve the ability to make conditioned streamflow forecasts.

1.2. STUDY AREA

The Upper Truckee River Basin (Figure 1.1) is a snow-melt dominated watershed that requires significant streamflow forecasting to meet competing demands from various water users. The basin encompasses an area of approximately 3,060 square miles on the border of California and Nevada. The Truckee River originates as the only drainage outflow from Lake Tahoe in California, runs northeastward approximately 105 miles, and terminates in Pyramid Lake in Nevada. The Truckee River has an average annual flow of 548,200 acre-feet (1973-1994 period of record) at the Farad gaging station on the California- Nevada border [*Horton*, 1995].

The Upper Truckee Basin is a steep, high alpine or forested environment with elevations reaching 9,000 to 10,000 feet [*Horton*, 1995]. The area receives 30 - 60 inches of precipitation annually, primarily in the form of snow [*Taylor*, 1998]. Consequently, most of the runoff results from snowmelt during the spring months. The Upper Truckee River Basin has seven major storage reservoirs: Lake Tahoe, Donner Lake, Independence Lake, Martis Creek Lake, Prosser Creek Reservoir, Stampede Reservoir, and Boca Reservoir. These reservoirs are used for both flood control and storage of water for downstream uses.

The Truckee River is used to produce hydropower at a plant upstream of the Truckee Meadows in Nevada, an area that encompass the cities of Reno and Sparks.

Reno and Sparks use Truckee River water for municipal and industrial (M&I) purposes. Downstream of the Truckee Meadows, the Truckee River flows to Derby Dam where an annual average of almost 187,000 acre-feet of water is diverted from the Truckee Basin through the Truckee Canal into Lahontan Reservoir for use in the Newlands Project irrigation district. The Newlands Project diversion comprises the most significant single withdrawal of the Truckee River's waters.

The portion of the river that is not diverted continues through desert and into Pyramid Lake within the Pyramid Lake Indian Reservation. Two fish, the endangered cui-ui and the threatened Lahontan cutthroat trout live in Pyramid Lake and are culturally and economically important to the Pyramid Lake Paiute Tribe. Low flows and shallow depths in Truckee River below Derby Dam, however, have inhibited spawning, egg incubation, and survival of these species [*Taylor*, 1998].

The Truckee River has been, and continues to be, crucial to the sustainment of life in western Nevada. The river has played a major role in the settlement and development of the area. Consequently, the policies and operations on the river extend back to before the turn of the century and continue to be negotiated to this day. Current negotiations seek to balance the demands of M&I for the cities of Reno and Sparks, irrigation for Truckee Meadows, power production, as well as protection of endangered and threatened species [*Horton*, 1995].

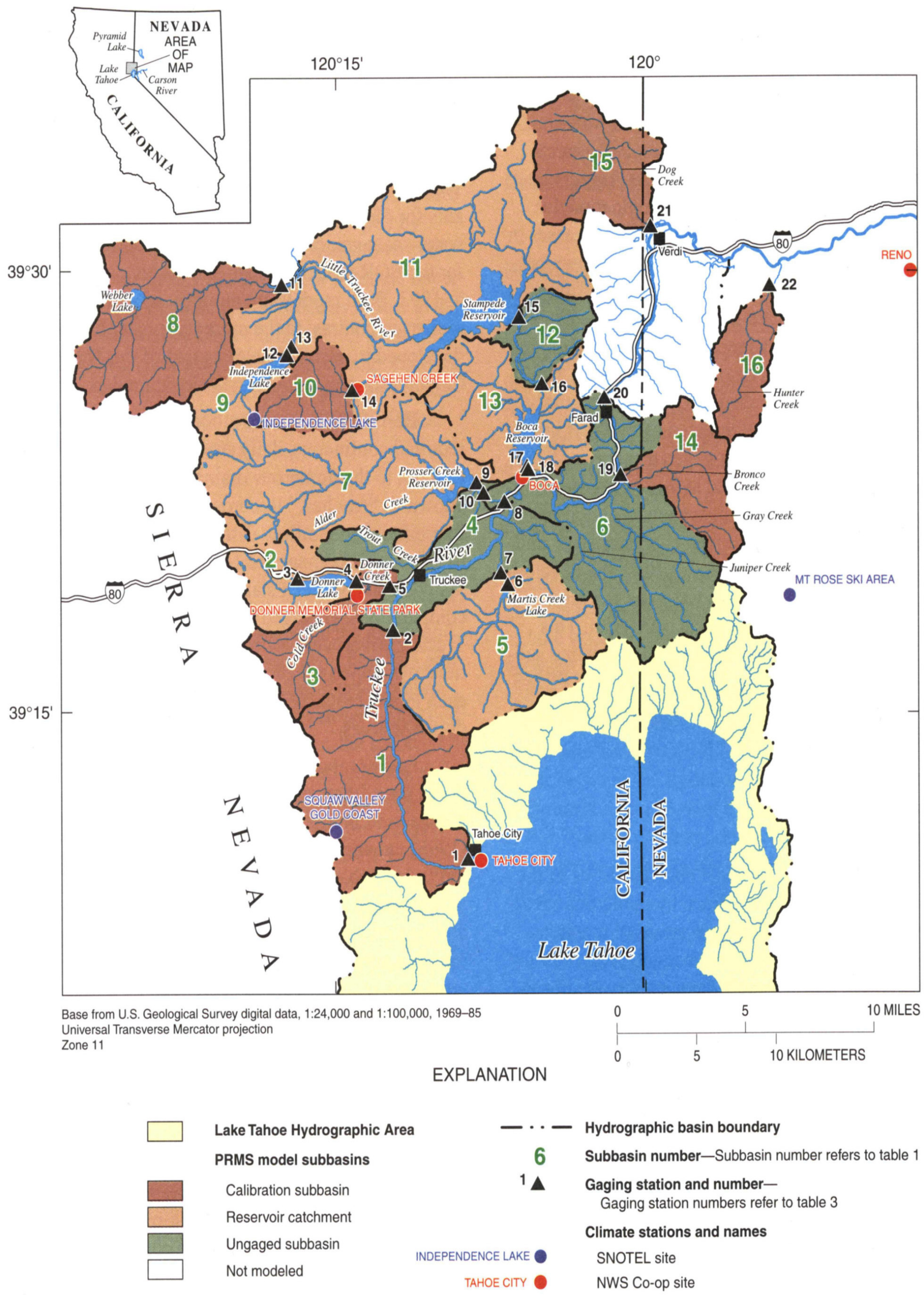

Figure 1.1. The Upper Truckee River Basin (*Jeton,* 1999).

Currently, the U.S.Bureau of Reclamation (USBR) techniques use linear regression analysis based on snow water equivalent (SWE) information to forecast streamflows in the Truckee. Forecasters typically regress streamflow data against monthly basin average snow water equivilent (SWE), percent of normal snowpack, total accumulated precipitation, and observed runoff data to develop regression equations for each month. Forecasters then use the monthly regression equations to predict the most probable streamflow value based on snow data. USBR forecasts also include information from the Natural Resource Conservation Service (NRCS) official forecasts, whenever they are available. The monthly distribution of the forecasted runoff is determined from similar years selected based on forecasted volume forecasts.

The main drawback to this forecasting system is the lack of historical data from which to draw from and the inability to condition on other factors such as large scale climate indices. Additionally, correlation between weather variables is not considered which can have a significant impact on the outcome. For example, with similar snowpack data between different years, differences in temperature can dramatically change the magnitude and timing of snowmelt. A modified ESP approach, specifically using a stochastic weather generator to create ensembles of weather based on the historic record, could significantly overcome these issues and improve streamflow forecasting of the Truckee River.

1.3. PROPOSED RESEARCH

This paper presents a framework to couple simulated weather data from a modified *K-nearest-neighbor (K-nn)* stochastic weather generator of *Yates, et al.* [2004] with the Precipitation-Runoff Modeling System (PRMS), an existing deterministic, physically-based watershed model to generate alternative streamflow scenarios (**Error! Reference source not found.**). The method, although universal in nature, is applied to three sub-basins in the Upper Truckee River Basin.

A stochastic weather generator simulates weather variables based on the historic record and creates ensembles of weather data for any period of time. A weather generator eliminates the historic-record limitation of the traditional ESP approach. Simulated weather variables can be incorporated into a deterministic watershed model to generate ensembles of simulated streamflows which can be statistically analyzed to assess possible streamflow scenarios. The goal of this technique is to: 1) preserve the statistics from the historic weather data; 2) preserve the temporal and spatial relationships between weather variables both at and between stations in the basin; and 3) allow for conditioned forecasting based on other relationships such as wet/dry years.

The overall objective of this paper is to present a framework that couples a stochastic weather generator with a watershed model. The specific objectives are three-fold: (1) demonstrate the ability of the weather generator to adequately reproduce the weather variables in the Upper Truckee River Basin; (2) couple the weather generator with an existing watershed model; and (3) generate alternative streamflow scenarios for one sub basin in the Upper Truckee River Basin.

In chapter 2, a brief overview of weather generators and a description of the *K-nn* approach are provided. Results from the Upper Truckee River Basin using the historic weather record to create synthetic data series are presented and discussed. Chapter 3 describes the watershed model and its application in the Upper Truckee River Basin using the historic record of weather. Chapter 4 presents the coupling of the synthetic weather data with the watershed model to create an ensemble of streamflow forecasts. In chapter 5 the forecasting capabilities of the technique are presented by using a simple wet/dry/normal forecast comparison. In Chapter 6, the results and finding are discussed along with a summary of the project and recommendations for future studies.

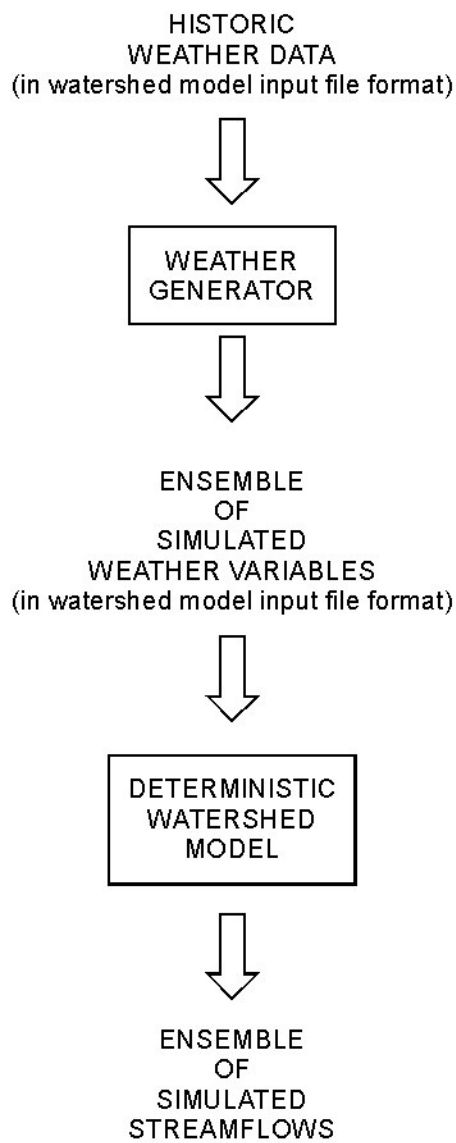

Figure 1.2. Flow chart of coupling the weather generator with a watershed model to simulate streamflows.

CHAPTER 2

2. WEATHER GENERATOR

As previously described, a group of methods referred to as "weather generators" can provide alternative weather sequences that compensate for the inadequate length of climate records. Stochastic weather generators simulate climate variables such as temperature and precipitation based upon the statistical distribution of historical data. Generally, stochastic weather generators need to consider the correlation and/or dependence of the weather variables with each other on the same day as well as over time (persistence) both at a station (temporal) and between stations (spatial). Many weather variables are correlated to precipitation. Consequently, many parametric weather generators simulate precipitation occurrence and amount independently and generate other variables based on the stochastically generated precipitation [*Rajagopalan and Lall,* 1999*; Yates, et al.,* 2003]. Other weather variables are generated at each station based upon a unique probability distribution function given the state-of-precipitation on the given day, the location of the station and the time of year. However, precipitation occurrence and amounts may depend on the weather variables being simulated and the dependence may vary throughout the year. Accordingly, the dependence of variables other than precipitation on the daily weather sequence should be considered as well as the unique relationship between them.

Non-parametric methods generate weather sequences by resampling from the historical record [e.g. *Rajagopalan and Lall,* 1999*; Buishand amd Brandsma,* 2001*;*

Yates, et al., 2003*, Clark, et al.,* 2004*; and Gangopadhyay, et al.,* 2005] and can improve upon the parametric models. Like all resampling methods, these methods assume past meteorological conditions are representative of what may occur in the future. These methods compare a vector of all weather variables for a given day, t, against a vector including the same variables from similar dates in the historic record. The k most similar days are taken as the k-nearest neighbors, where $k = \sqrt{n}$ and n is the number of similar dates used in the comparison. One of these neighbors is randomly selected, and the day following the selected neighbor is taken as the next simulated day (day $t + 1$). By resampling from the all data in the historic record unique spatial and temporal dependencies can be preserved and allow for simple and effective multivariate, multi-site weather generation.

2.1. METHOD

The weather generator used in this study is a modification of the *K-nn* algorithm method presented by *Yates et al.* [2003] which is based on the methods of *Lall and Sharma* [1996], *Rajagopalan and Lall* [1999], and *Buishand and Brandsma* [2001]. The method was modified to incorporate the data format used by the PRMS watershed model for input and output.

Yates, et al. [2003] describes in detail the process steps of the weather generator. In brief, the algorithm generates weather variables as follows:

1. The regional means of all weather variables: precipitation depth (PPT), maximum temperature (TMX), and minimum temperature (TMN) are computed for a given day t from all stations.

2. All days within a user-specified temporal window centered on day *t* are selected as potential candidates for the simulated day.
3. The regional means of each potential day are calculated
4. The Mahalnobis distances between the mean vectors of potential days and the current day are computed
5. The distances are sorted and the first *k*-nearest neighbors are retained. In this case, $k = \sqrt{n}$ where *n* is the number of days in the window.
6. A weighted probability function is assigned to each day with the nearest neighbors having the higher probability.
7. One of the days is randomly selected and the day following is taken as the next simulated day (day $t + 1$).

The weather generator uses daily precipitation (PPT) and maximum and minimum temperature (TMX and TMN) data from the National Weather Service (NWS) cooperative network (COOP) of climate observing stations throughout the Upper Truckee River Basin in the Eastern Sierra Nevada (Table 1). Data were formatted as input files to the watershed model to facilitate coupling between the weather generator and the watershed model. Currently, the weather generator requires a complete dataset (no missing values) to run and therefore only stations with a continuous record can be utilized with the weather generator. In this study, data from six climate stations provided data from 1 January 1981 to 31 December 1999 (19 years). A fixed window of three days on either side of the current day was used yielding a total of 132 potential days and $k = 11$. Output from the weather generator includes an index of days referencing days in the historic record selected to represent the simulated day's weather. Other statistically outputs are described below.

Meterological Station Name	Station Identification	Operating Agency	Weather Generator ID
Independence Lake	20K05S	NRCS	
Marlette Lake	19K04S	NRCS	
Mt. Rose Ski Area	19K07S	NRCS	
Hagan's Meadow	19L03S	NRCS	
Squaw Valley Gold Coast	20K30S	NRCS	
Rubicon #2	20L02S	NRCS	
Echo Peak	20L06S	NRCS	
Fallen Leaf	20L10S	NRCS	
Boca	CA0931	NWS	1
Donner Memorial State Park	CA2467	NWS	2
Sagehen Creek	CA7641	NWS	3
Tahoe City	CA8758	NWS	4
Glenbrook	NV3205	NWS	5
Reno Airport	NV6779	NWS	6
Tahoe Valley FAA AP	CA8762	NWS	

Table 2.1. Meteorological stations used in weather generator and watershed model for the Upper Truckee River Basin, California and Nevada, USA.

2.2. RESULTS AND DISCUSSION

The *K-nn* algorithm weather generator was used to simulate daily maximum and minimum temperature and precipitation at six stations in the Upper Truckee River Basin for a total of 100 independent years. A set of statistics were computed for each month to compare the simulated data with the historic record. The statistics include means, standard deviations, skewness coefficients, cross correlations between the variables at a station, and cross correlations between stations. Total depth and average number of days between events are used for precipitation data to provide a better description of the data and to capture elements of concern. The computed statistics are summarized with boxplots presented in FiguresFigure 2.1 toFigure 2.9. In these boxplot figures, the box indicates the interquartile range, the whiskers show the 5th and 95th percentile of the simulation, dots indicate values outside this range, and the horizontal line within each box indicates the median value. The values of the statistics from the observed data are represented by a solid line. In general, if the statistics of the observed data lie within the box of the simulated values, it suggests that the simulated values adequately reproduced the statistics of the historical data. Although, the weather simulations were made on a daily timescale, the statistics from the daily data have been aggregated to a monthly timescale.

2.2.1. PRECIPITATION

The upper left plots in FiguresFigure 2.1,Figure 2.2, andFigure 2.3 show total precipitation at three stations in the basin: Tahoe City, Sagehen Creek, and Donner Memorial. The simulations adequately reproduced the historical monthly precipitation suggesting that the annual precipitation was captured. There is a tendency to underestimate precipitation totals at all stations, however, the overall performance was good when compared to the patterns of the historical data.

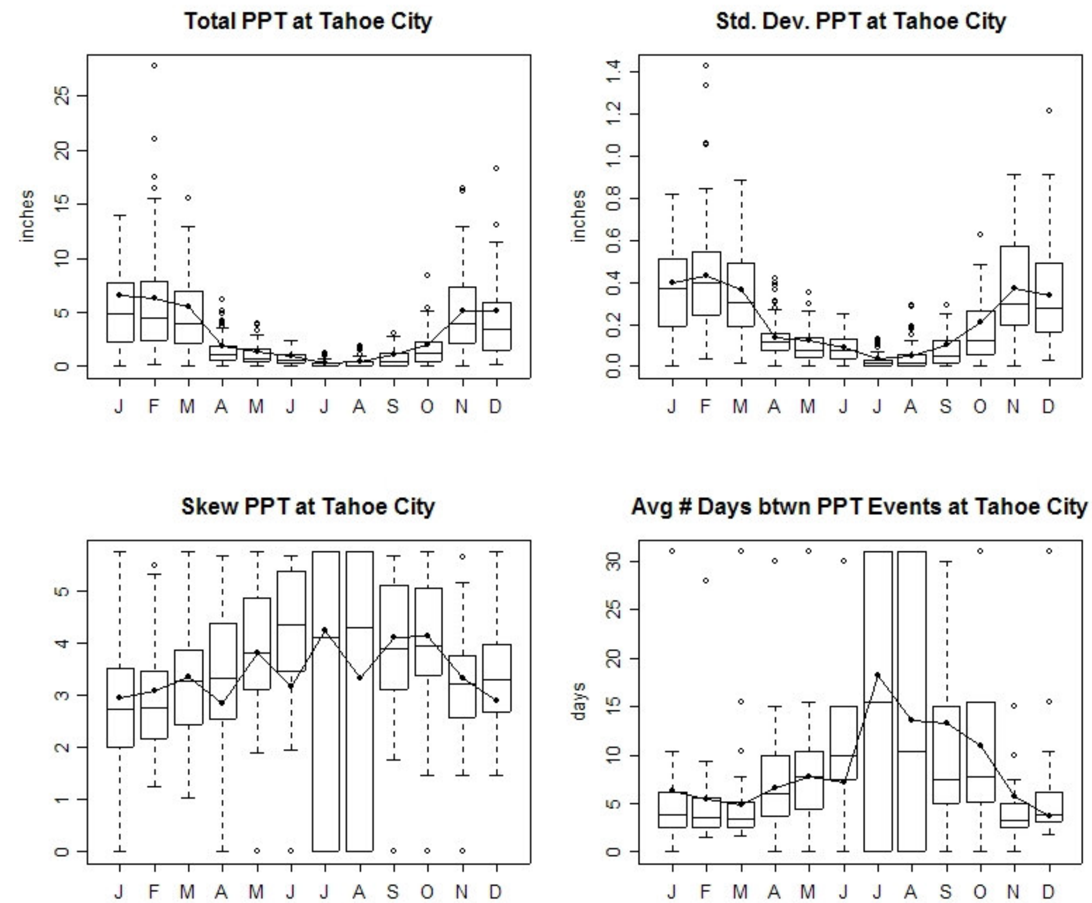

Figure 2.1. Box plots of precipitation statistics for Tahoe City. Top left graph is total monthly precipitation, top right graph is standard deviation of daily precipitation, bottom left graph is the skewness of daily precipitation, and the bottom right graph is the average number of days between precipitation events for the month. The solid line represents the same statistics from the historical record for the period 1981-1999.

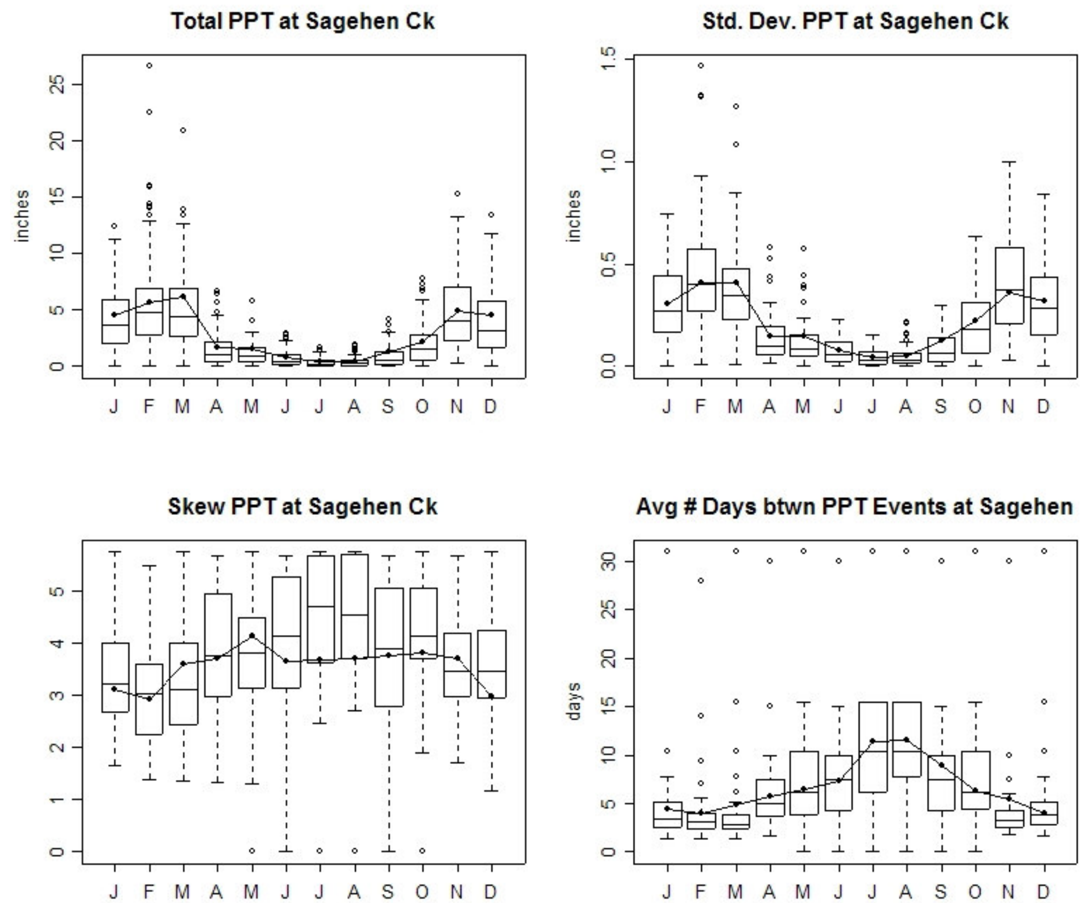

Figure 2.2. Same as Figure 2.1, but for Sagehen Creek.

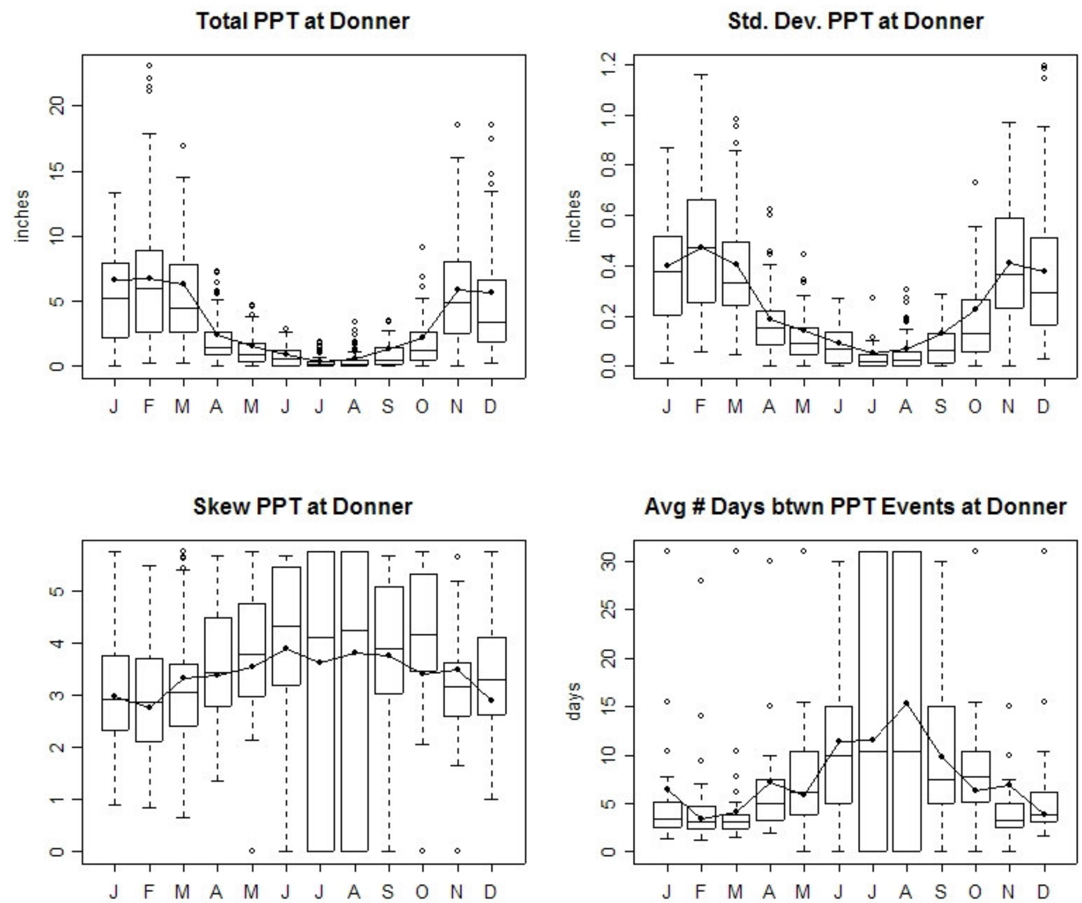

Figure 2.3. Same as Figure 2.1, but for Donner Memorial

The upper right plots of FiguresFigure 2.1,Figure 2.2, andFigure 2.3 represent the standard deviations of daily precipitation for each month quantifying the spread of the data. The simulations adequately reproduced the historical statistics. Skewness, a description of rainfall distribution symmetry, is shown in the lower left graphs of FiguresFigure 2.1,Figure 2.2, andFigure 2.3. Typically, precipitation data are positively skewed, placing the mean in the upper half of the data. The degree of positive skewness illustrates that rainfall typically occurs as many small events with a few large events that elevate the mean. The simulations capture the historical statistics well. The months of June and July typically have almost zero precipitation with occasional small events and one large event which create a wide range of skew values.

The lower right graphs in FiguresFigure 2.1,Figure 2.2, andFigure 2.3 show the average number of days between precipitation events as a measure of persistence. This statistic gives insight to the ability of the technique to reproduce dry and wet spells. The simulated statistics do a fairly good job at capturing the historic statistic.

2.2.2. TEMPERATURE

FiguresFigure 2.4,Figure 2.5, andFigure 2.6 show the minimum and maximum temperature statistics for Tahoe City, Sagehen Creek, and Donner Memorial, respectively. The upper panels show the mean and standard deviation of the minimum daily temperature while the bottom panels shows the same statistics for the maximum temperature. The simulations reproduce the historic statistics very well.

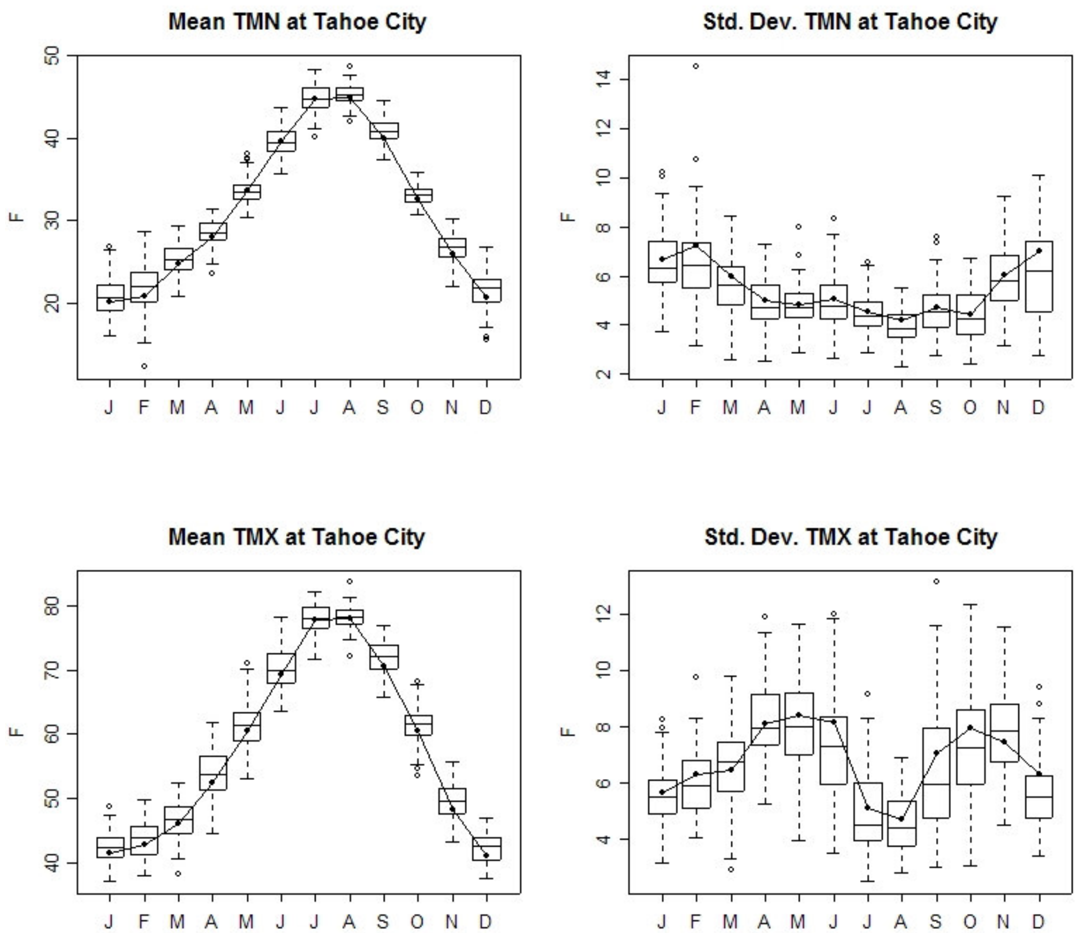

Figure 2.4. Box plots of daily mean and standard deviation minimum and maximum temperature for Tahoe City. The solid lines represent the same statistics derived from the historical record for the period 1981 to 1999.

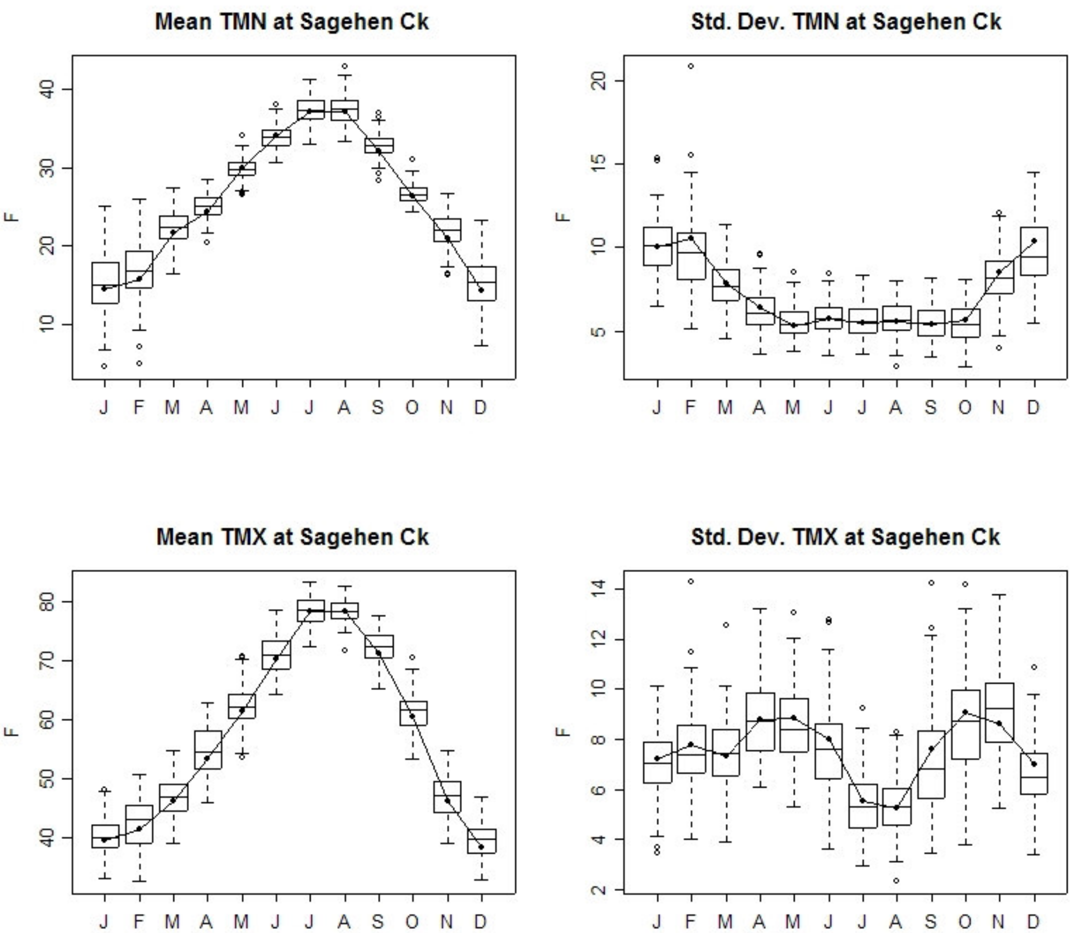

Figure 2.5. Same as Figure 2.4, but for Sagehen Creek.

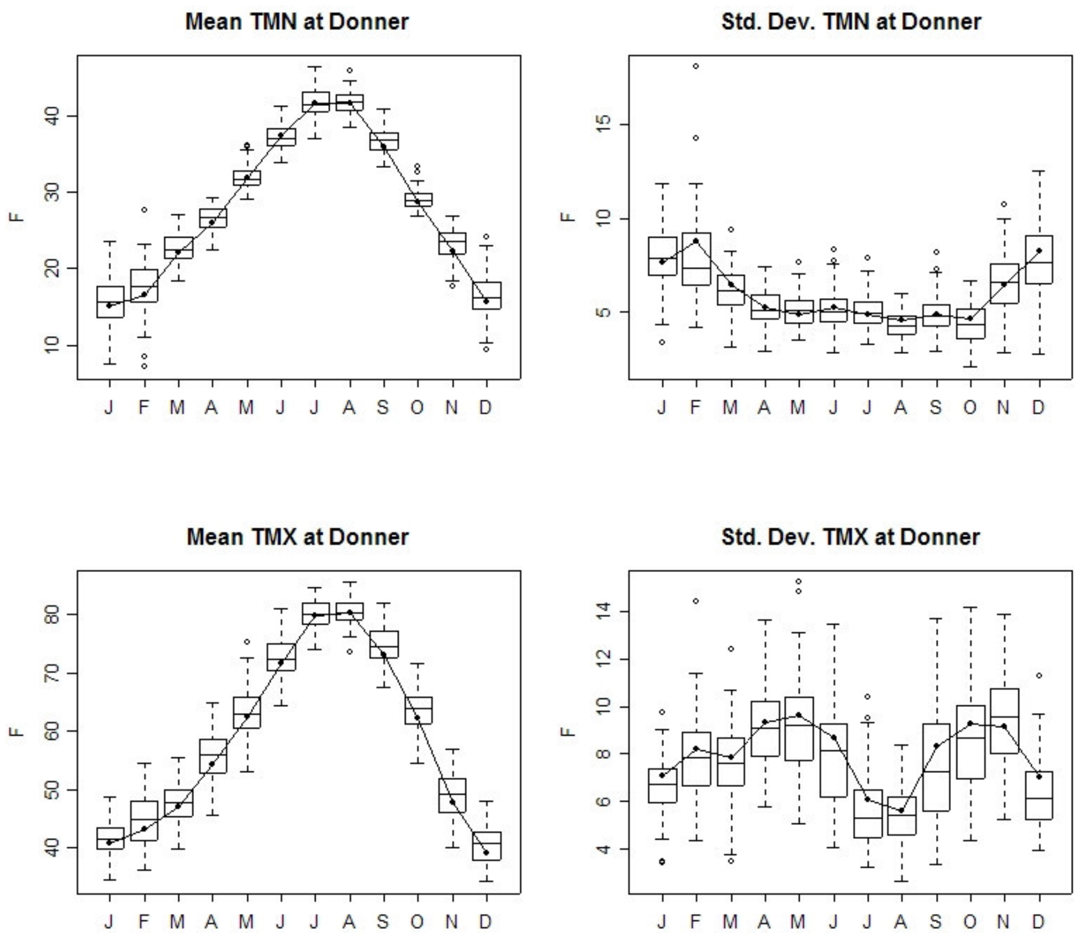

Figure 2.6. Same as Figure 2.4, but for Donner Memorial.

2.2.3. CORRELATION BETWEEN VARIABLES AT A STATION

Lag-0 correlation between variables at a station is the relationship between variables at a station on a given day (e.g. PPT-TMX, PPT-TMN, and TMX-TMN). Strong lag-0 correlations between variables at a station would be expected since the algorithm selects all weather variables from a station for a given day. The weather generator reproduces the historical statistics well. Figure 2.7 shows the lag-0 correlations between variables at stations 1 (Boca), 4 (Tahoe City), and 6 (Reno) for the months of January, June and October.

The weather generator simulates each day's weather based on the previous day. Lag-1 correlation of a variable at a station is the relationship of the variable's value on a given day with its value on the next day. Lag-1 correlations illustrate the ability of the weather generator to simulate the temporal variability at a station. The simulated values reproduce the observed statistics well, indicating that the weather generator maintains the historic temporal variability. Figure 2.8 shows the lag-1 correlations of each variable at stations 1, 4, and 6.

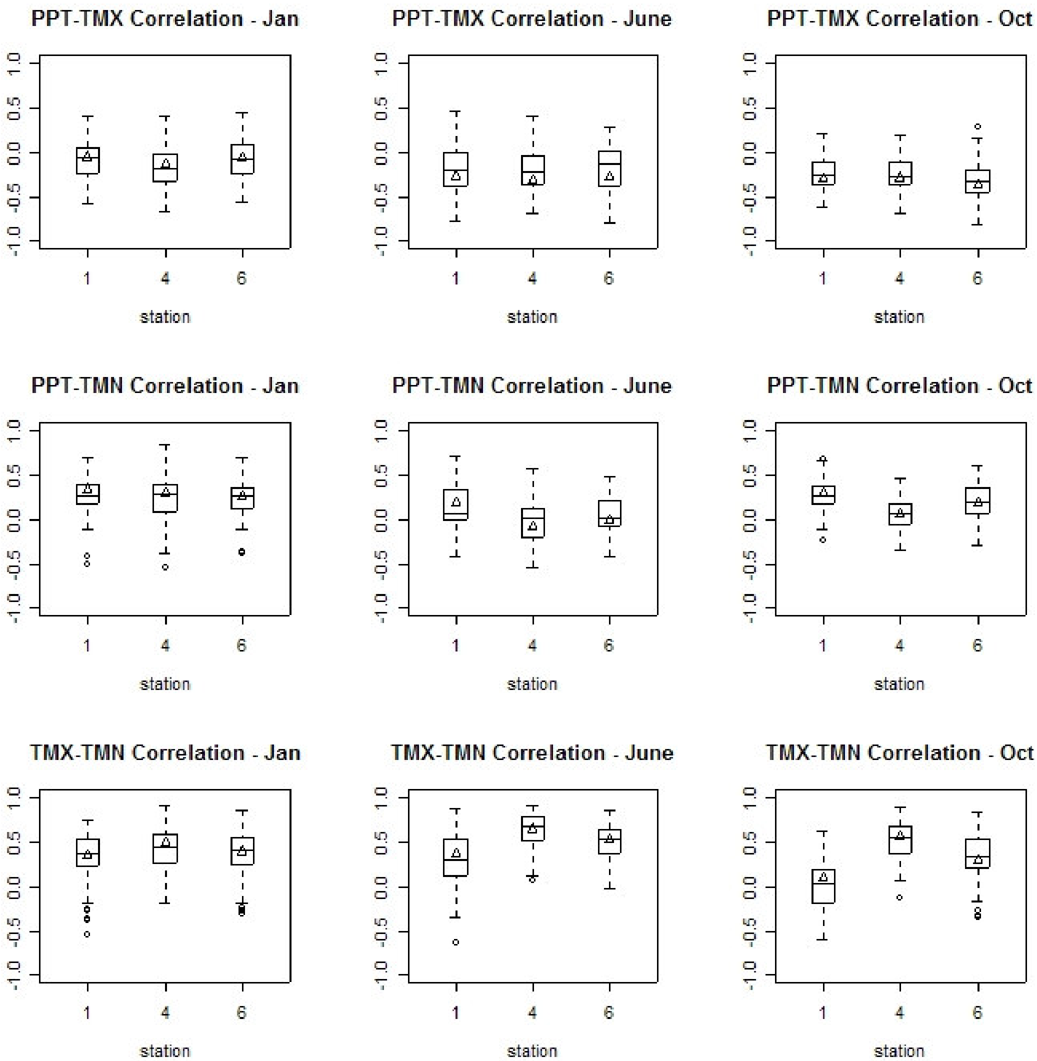

Figure 2.7. Box plots of lag-0 correlation between variables at stations 1, 4, and 6 for January, June, and October. The triangles represent the mean of the historic correlations for the period 1981 to 1999. Stations numbers correspond to Weather Generator ID's indicated in Table 2.1.

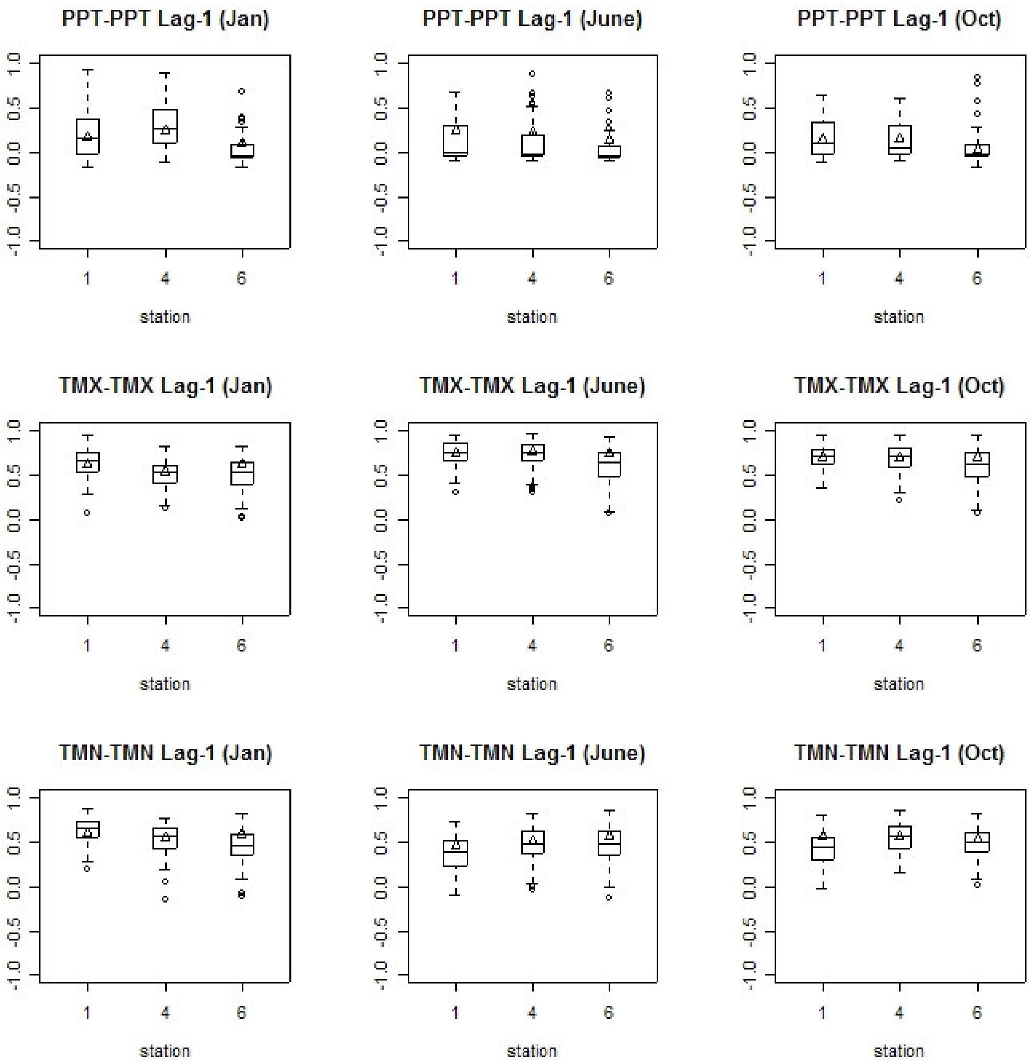

Figure 2.8. Same as Figure 2.7, but lag-1 correlation between same variables.

2.2.4. CORRELATION BETWEEN STATIONS

Accurate spatial distribution of variables between stations is vital for use in a watershed model to represent the distribution of weather patterns over a basin. Lag-0 correlations of variables between stations are shown in Figure 2.9. The plots indicate that the simulated values adequately reproduce the historical statistics and illustrate the effectiveness of the weather generator to distribute variables across the basin.

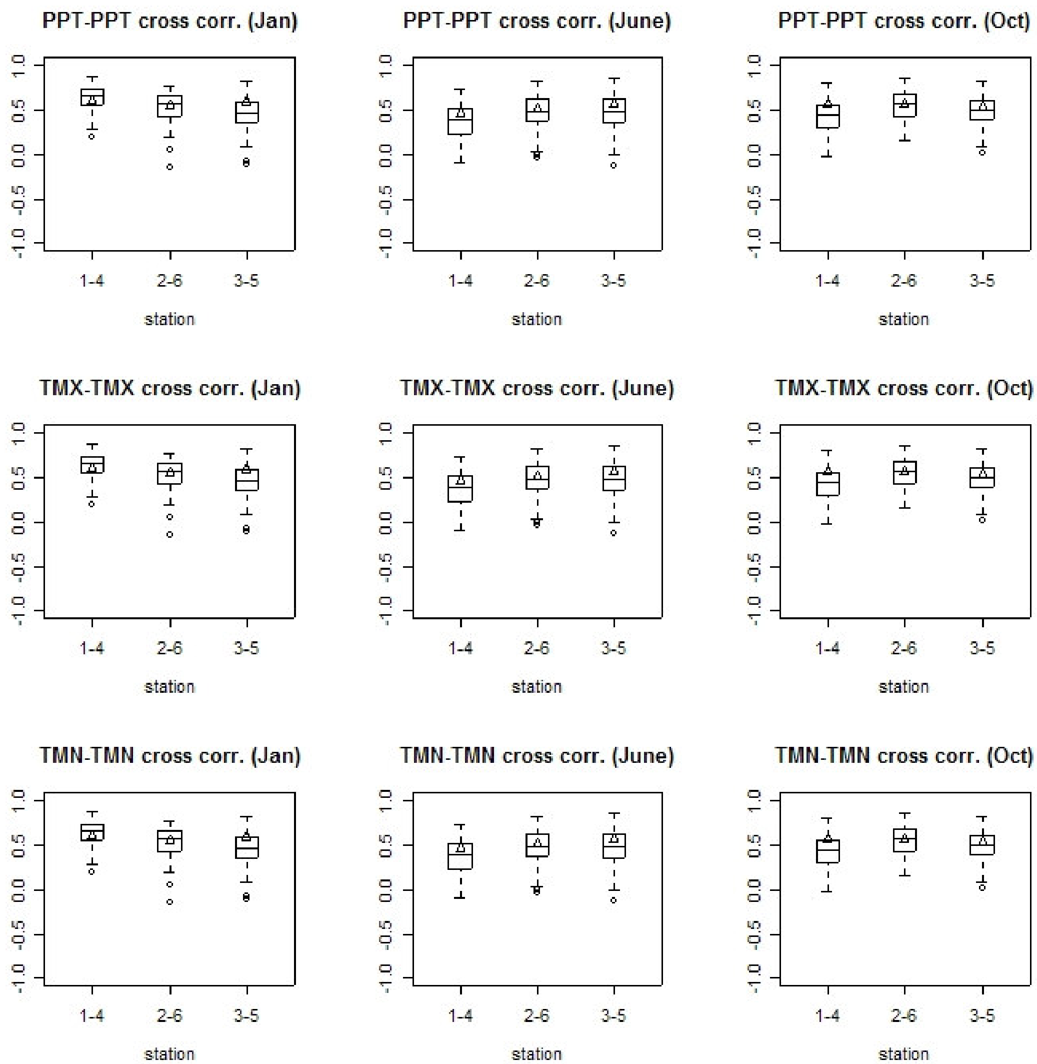

Figure 2.9. Box plots of lag-0 correlation of variables between stations 1 and 4, 2 and 6 and 3 and 5 for January, June, and October. The triangles represent the mean of the historic correlations for the period 1981 to 1999. Stations numbers correspond to Weather Generator ID's indicated in Table 2.1.

2.3. SUMMARY

In summary, the K-nn weather generator methodology proved to be effective at creating multi-station realizations of weather data in the Upper Truckee River Basin. Mean and standard deviation statistics were reproduced well for daily maximum and minimum temperature. Precipitation statistics of total, standard deviation, skew and average-number–of-days-between-events were adequately reproduced, however, total amounts were slightly underestimated by the weather generator. The simulated weather variables largely reproduced the temporal and spatial statistics of the historic record as indicated by the lag-0 and lag-1 correlations. The results indicate that the weather generator reproduced the historic distributions of the weather variables and adequately distributed them throughout the basin.

CHAPTER 3

3. WATERSHED MODEL

Watershed models, like all models, are approximations of physical processes and, as such, are not perfect. Specifically, they are assemblages of sub-models describing different components of the hydrologic cycle intended simulate basin responses to various inputs. Watershed models can be classified according to the spatial distribution and description of hydrologic processes used to represent these components.

In general, watershed models have five components: 1) system (watershed) geometry, 2) input data, 3) governing laws, 4) initial and boundary conditions, and 5) modeled output. These components are variously combined in different types of watershed models. The processes within the model include the hydrologic processes that contribute to the system output. Based on the description of these processes, in conjunction with the system characteristics, models can be described as lumped or distributed, deterministic or stochastic.

Lumped models, in general, are expressed by simplified ordinary differential equations and empirical algorithms that do not account for spatial variability of the model components and processes. Alternatively, distributed models explicitly account for spatial variability of processes, input, boundary conditions, and system characteristics. In practice, lack of data prevents such a general formulation of distributed models. In many instances only the processes linked directly to output are distributed [*Singh,* 1995].

Processes, and therefore models, can be described as either deterministic or stochastic. A deterministic process is one whose time evolution can be predicted exactly, while a stochastic process is one that cannot be predicted exactly but the probability of certain behaviors is known. If all components are deterministic, the watershed model is deterministic and, likewise, if all components are stochastic, the model is stochastic. For this study, a stochastic input (weather) is applied to a distributed, deterministic watershed model to yield discrete values of streamflow that are combined to create an ensemble that can be statistically analyzed to yield a stochastic estimate of streamflow.

3.1. MODULAR MODELING SYSTEM AND PRECIPITATION-RUNOFF MODELING SYSTEM

The Modular Modeling System (MMS) is a modeling framework developed by the U.S. Geological Survey (USGS) that provides a flexible modular approach to model design [*Leavesley, et al.,* 1996, 2002]. MMS use a module library containing algorithms for simulating a variety of water, energy, and biogeochemical processes. If appropriate algorithms are not provided in a module, new modules can be developed and incorporated into MMS. The modeling component used for this study was the USGS Precipitation-Runoff Modeling System [PRMS; *Leavesley, et al.,* 1983].

PRMS is a physically-based (deterministic), distributed-parameter model designed to simulate runoff from precipitation and snowmelt as well as alpine snowpack accumulation and snowmelt processes. The model analyzes the effects of temperature, precipitation, and land use on general basin hydrology. Spatially variable land characteristics affecting runoff, such as altitude, slope, aspect, vegetation, soil, geology, and climate, are accounted for by dividing the modeled area into Hydrologic Response Units (HRU). Within each HRU, it is assumed that the hydrologic response to uniformly distributed precipitation and simulated snowmelt is homogeneous.

PRMS computes a daily water-energy balance within each HRU. The total basin response is computed by taking the area-weighted sum of the daily hydrologic fluxes from all HRU's. The model is driven by daily values of precipitation and maximum and minimum temperature. Solar radiation can be supplied as a driving variable or can be computed from the other variables and adjusted for slope and aspect. Each HRU is indexed to climate stations not necessarily in that HRU. Spatial and altitude differences between the weather station and the HRU are accounted for by a lapse rate computation and a precipitation correction factor. Within each HRU, the form of precipitation (rain, snow, or mixed) is dependent on a user-defined relationship between a specified snow-rain threshold temperature and estimated minimum and maximum temperature.

PRMS is composed of several "reservoirs" that serve as water-budget terms: interception, snowpack, impervious zone, recharge and lower zone, subsurface, and groundwater. Algorithms within each water-budget term determine the fluxes such as evapotranspiration, snowmelt, surface runoff, subsurface and groundwater flow.

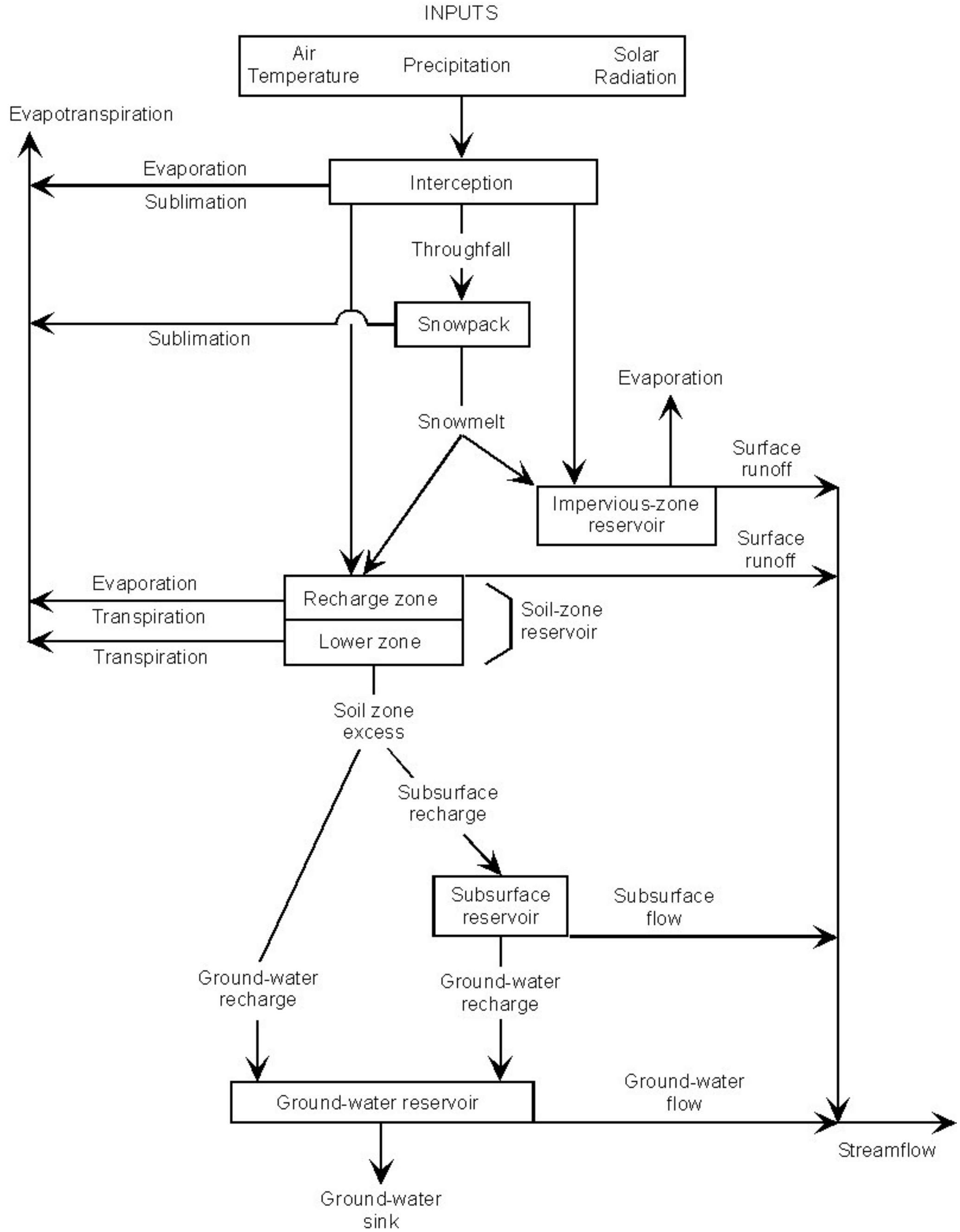

Figure 3.1. Schematic diagram of the conceptual system used in PRMS

3.2. APPLICATION OF MMS-PRMS IN THE TRUCKEE BASIN

Jeton [1999] developed MMS-PRMS models as part of a river/reservoir model for the Upper Truckee River Basin for the Truckee-Carson Program of the USGS to assist the U.S. Department of the Interior in implementing the Truckee River Operations Agreement. The model was intended to 1) simulate streamflow at ungaged sites, 2) extend streamflow records at gaged sites, and 3) forecast inflows.

The model is comprised of 16 sub-basin models that are run independently and subsequently routed to generate a total basin outflow (Table 3.1). Each sub-basin model is run using precipitation and temperature data from 15 NWS-COOP and National Resources Conservation Service Snow Telemetry (SNOTEL) climate stations throughout the basin (Table 2.1).

Jeton [1999] provides a thorough description of the water-budget terms and fluxes in the Upper Truckee River Basin model. For this study, snowmelt and streamflow fluxes, warrant attention.

Because the Upper Truckee Basin is a snowmelt dominated basin accurate simulation of snowmelt is critical to determining the basin response to melt of winter snowpack. In the model, snowmelt is determined by a simplified energy balance approach based on shortwave and longwave radiation and a term representing latent and sensible heat. At the watershed scale, mean area values of snow accumulation and melt can be estimated by using measured components of the snow-energy budget and parameterizing or neglecting components where data aren't available [*Leavesley,* 1989]. For the Upper Truckee River Basin, the dominant energy contributor to snowmelt is radiation fluxes, rather than turbulent transfers from the atmosphere.

This would suggest that snow accumulation, melt, and timing of runoff are highly sensitive to temperature. The Sierra Nevada commonly experience rain-on-snow events which are challenging to models designed to simulate colder, higher alpine snowpacks. Melting of snow in PRMS is triggered when the cold content of the snowpack is low (warm) enough to ripen and melt the snowpack, therefore timing is indirectly linked to the annual temperature cycle [*Jeton,* 1999].

Sub-Basin Number (Fig.1)	Sub-Basin Name	Drainage Area (square miles)	Mean HRU altitude range (feet above sea level)
1	Tahoe City-to-Truckee Reach	47.0	5,882 - 8,609
2	Donner Lake	13.2	5,932 - 8,612
3	Cold Creek	14.7	5,879 - 8,356
4	Ungaged Area 1 (Truckee to below Prosser Creek Reservoir)	25.7	5,610 - 7,224
5	Martis Creek Lake	39.6	5,794 - 8,320
6	Ungaged Area 2 (Juniper and Gray Creeks)	46.6	5,200 - 10,184
7	Prosser Creek Reservoir	49.0	5,777 - 8,487
8	Webber Lake	35.4	6,470 - 9,062
9	Independence Lake	6.5	6,955 - 8,806
10	Sagehen Creek	10.7	6,430 - 8,320
11	Stampede Reservoir	76.7	5,948 - 8,419
12	Ungaged Area 3 (between Stampede and Boca Reservoirs)	9.8	5,669 - 8,215
13	Boca Reservoir	24.9	5,630 - 7,254
14	Bronco Creek	15.4	5,731 - 10,358
15	Dog Creek	21.7	4,908 - 8,323
16	Hunter Creek	11.3	5,292 - 9,528

Table 3.1. Modeled sub-basins in the Upper Truckee River Basin, California and Nevada, USA.

Simulated streamflow from PRMS is a summation of three flow components: 1) surface flow; 2) subsurface flow; and 3) groundwater flow. Surface, or overland, flow is generated from saturated soils and runoff from impervious surfaces. Subsurface flow is shallow subsurface flow that is originates from soil water in excess of available water-holding capacity of the soil. Groundwater flow, or baseflow, is sourced from both the soil zone and subsurface reservoir.

MMS-PRMS must be specifically calibrated to a basin. In the Upper Truckee Basin models, calibrating parameters focused on timing and volume of spring runoff using a few years of data and validating with other years. During calibration, it was determined that the model is highly sensitive to non-distributed, temperature dependent parameters and subsurface and groundwater flow routing coefficients. These parameters and coefficients have an impact on timing of spring runoff, the shape of the base flow recession, and the overall seasonal distribution of flows. Only a subset of the sub-basin models was calibrated due to the unavailability and uncertainty of data. Uncalibrated basins used parameters distributed from calibrated basins [*Jeton,* 1999].

3.2.1. SIMULATION OF AVERAGE CONDITIONS

For this study, the model developed by *Jeton* [1999] and subsequently modified by others [*Markstrom, personal communication*] was used to generate average historic streamflows from three sub-basins in the Upper Truckee River Basin using historic weather data.

FiguresFigure 3.2,Figure 3.3, andFigure 3.4 illustrate the modeled and observed runoff from the historic record (January 1981 to December 1999) for the

Tahoe City-to-Truckee Reach, the Sagehen Creek Reach, and the Prosser Creek Reach. It can be seen that the modeled historic flows do not accurately simulate the observed flows, however, limitations of the model have been well documented [*Jeton,* 1999]. The models were calibrated using a few years of data and validating with other years. As such, the models were not specifically calibrated for the average conditions, yet it is reasonable to assume that the models should adequately represent the historic flow statistics since the parameters were fit to specific years in the historic record.

The Tahoe-to-Truckee Reach is a calibrated basin that defines the most upstream reach of the Truckee River. It drains approximately 47 square miles of forested, mountainous terrain adjacent to the Truckee River. Development in the basin is primarily limited to ski resorts. Figure 3.2 indicates that the model accurately captures the timing of the peak (May) but appears to under represent the volume of flow during the runoff season (AMJ) and beyond. The model slightly over represents flow in February and March.

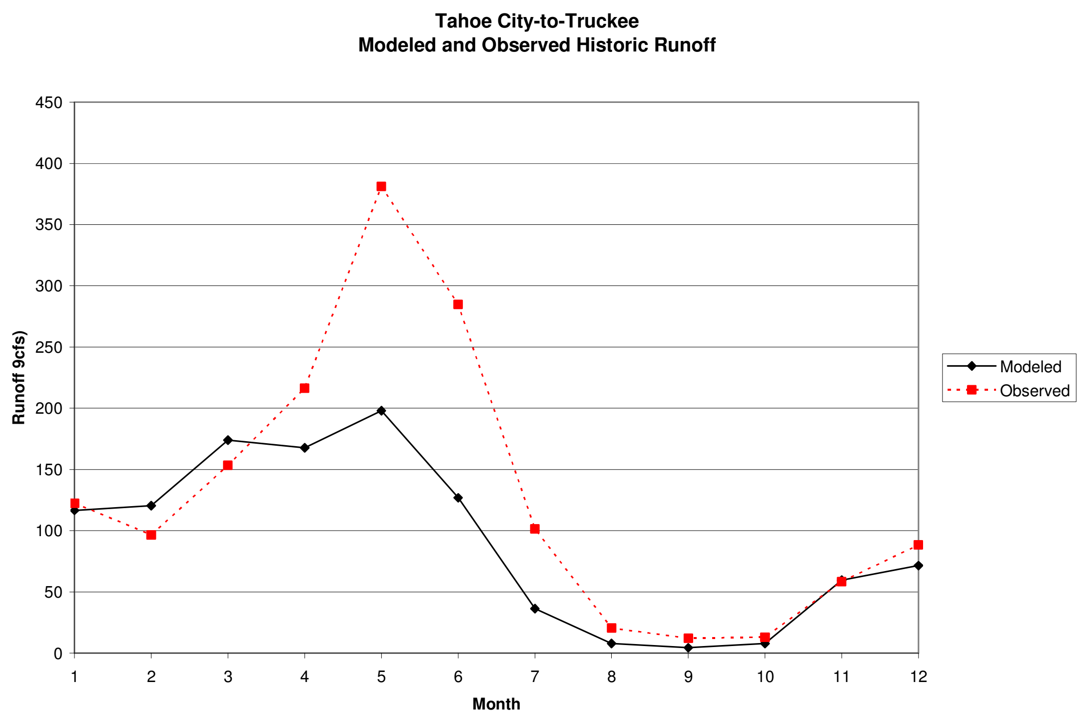

Figure 3.2. Mean runoff in Tahoe City-to-Truckee Reach. The dashed line is the mean observed runoff for the period 1981-1999. The solid line is the PRMS-modeled runoff for the period 1981-1999. Months represent the calendar year (1 = January, 12 = December).

The Sagehen Creek Reach is a calibrated basin that drains an area of approximately 10.7 square miles. The basin has no regulation of streamflow or urban development and contains the longest continuous flow record in the study area. Figure 3.3 shows that on average the modeled runoff underestimates the AMJ and annual volume and the timing of spring runoff. Modeled flows for January, February, and March appear to overestimate the observed mean.

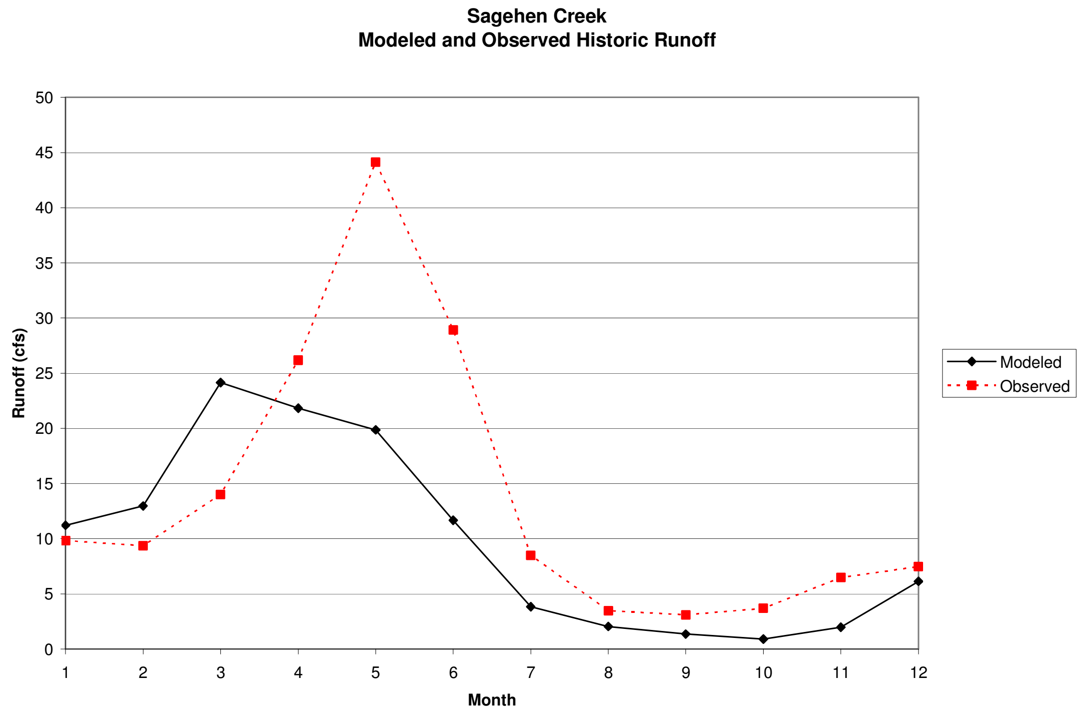

Figure 3.3. Same as Figure 3.2 except for Sagehen Creek Reach.

Prosser Creek Reservoir is an uncalibrated basin draining an area of approximately 49 square miles. A water balance was simulated to compute reservoir inflows. Due to the lack of total inflow data and the uncertainty of precipitation and evaporation no calibration was performed. Figure 3.4 indicates that the modeled flows represent the annual volume of flow well but place the average timing of the peak runoff prior to the observed value.

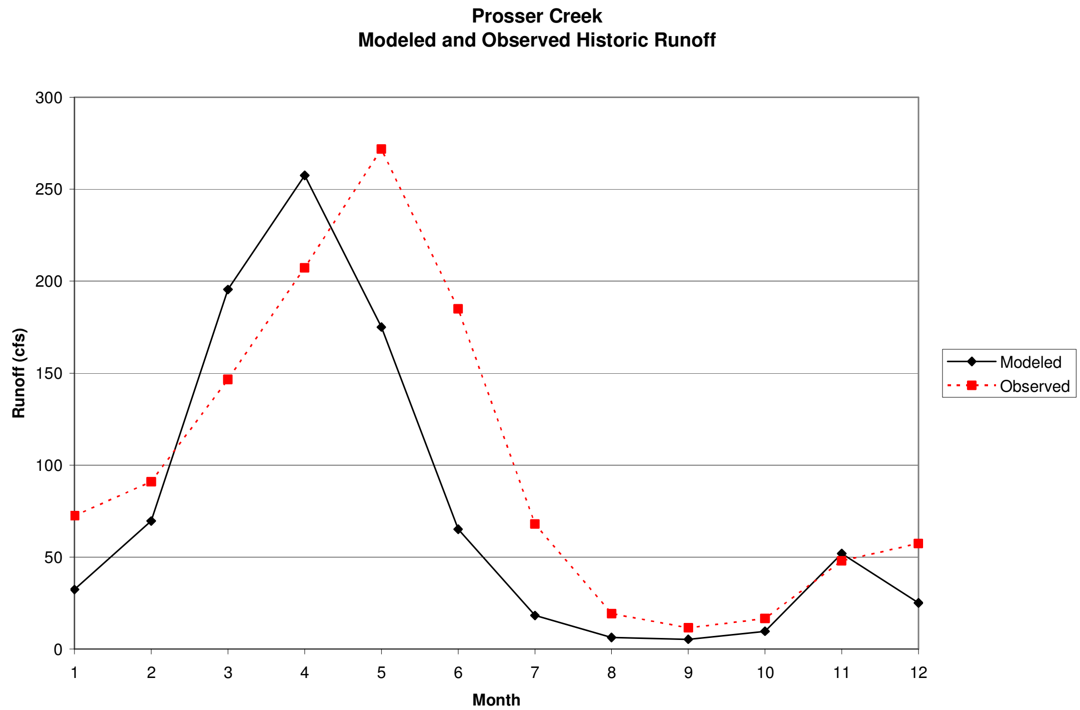

Figure 3.4. Same as Figure 3.2 except for Prosser Creek Reach.

The Tahoe-to-Truckee, Sagehen Creek, and Prosser Creek Reservoir models appear to inadequately simulate the historic observed mean runoff for each respective basin. This can partially be due to deficiencies in the watershed model. Deterministic hydrologic models are only an approximation of the physical system and cannot be perfectly calibrated. Model uncertainties in PRMS stem from simplifications made in representing hydrologic processes. Discrepancies in matching modeled to observed streamflow appear to be primarily due to difficulties in modeling the temporal and spatial distribution of precipitation and temperature, the form of precipitation and the snowpack melt rates [*Jeton,* 1999]. These issues are being addressed by others such as Hwang [2005] and are out of the scope of this study.

CHAPTER 4

4. COUPLING THE WEATHER GENERATOR WITH MMS

Despite the differences between modeled and observed streamflow, the MMS-PRMS models of the Upper Truckee River provide a baseline from which to evaluate the coupling of the weather generator with a watershed model. By coupling ensembles of weather data with the watershed model, ensembles of streamflows can be produced and statistically analyzed to generate more robust stochastic forecasts of streamflows.

4.1. METHODS

Simulated daily weather values selected by the weather generator were used as input to the Upper Truckee River Basin MMS watershed model to simulate modeled historic streamflow statistics. As described in previous chapters, the weather generator was used to simulate an ensemble of weather data. The index days from each series in the ensemble were used to select days from the original MMS-PRMS input data file. Six stations were used in the weather generator to select representative days, however, the MMS-PRMS model requires data from 15 stations to operate. Therefore, for a chosen day, weather data from all 15 stations in the historic recod were selected using the index day from the weather generator to represent the weather on a simulated day. Specifically, the six stations used in the weather generator are the NWS COOP stations within the basin that tend to be located in more populated areas. These stations were selected based on the length of record. The NRCS SNOTEL stations, with an inconsistent record, were completely left out of the weather generator process which could bias the data towards lower elevations and therefore less snow accumulation. The daily values of weather were incorporated into an MMS data input file along with the historic values to provide initialization of the model. The model was run to asses the ability of the weather generator to adequately reproduce the statistics of the modeled historic streamflows.

4.2. RESULTS AND DISCUSSION

FiguresFigure 4.1 toFigure 4.9 show boxplots of selected streamflow statistics from the simulated weather variables for each of the watershed models. The solid lines in the plots show the mean modeled historic flow statistics for each month from previous MMS-PRMS model runs. The dashed line shows the mean observed flow statistics.

FiguresFigure 4.1,Figure 4.2, andFigure 4.3 show the monthly mean values for steamflow. The weather generator values reproduce the modeled statistics fairly well with the exception of May and June values that consistently underestimate the modeled value. May values are within the 75th and 95th percentile, while the June values appear to be just outside the 95th percentile. The values for July, August, September and October are discounted due to the extremely low flows during that period. The timing of peak spring runoff appears to be consistent with the modeled values with the exception of the Tahoe-to-Truckee Reach.

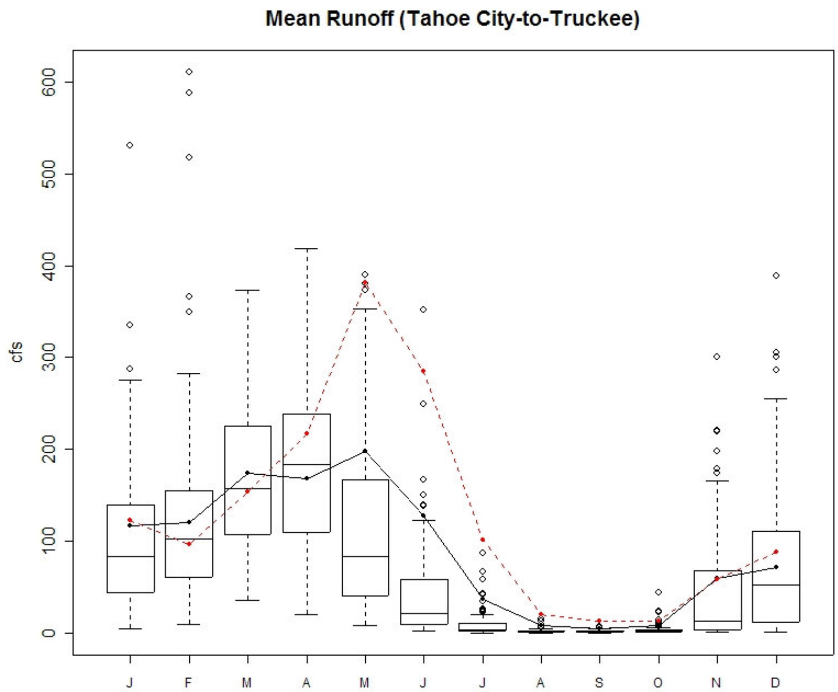

Figure 4.1. Monthly mean runoff for Tahoe City-to-Truckee Reach. Box plots represent the ensembles of output from the watershed model using the parameters simulated by the weather generator. The solid line indicates the watershed model output using the historic data (1981-1999). The dashed line represents the observed values (1981-1999).

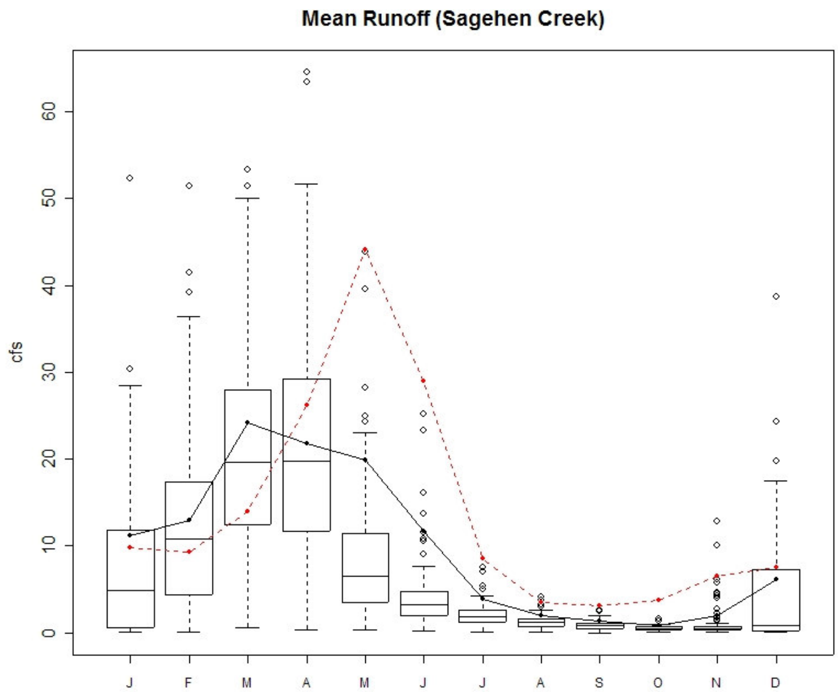

Figure 4.2. Same as Figure 4.1 but for Sagehen Creek Reach.

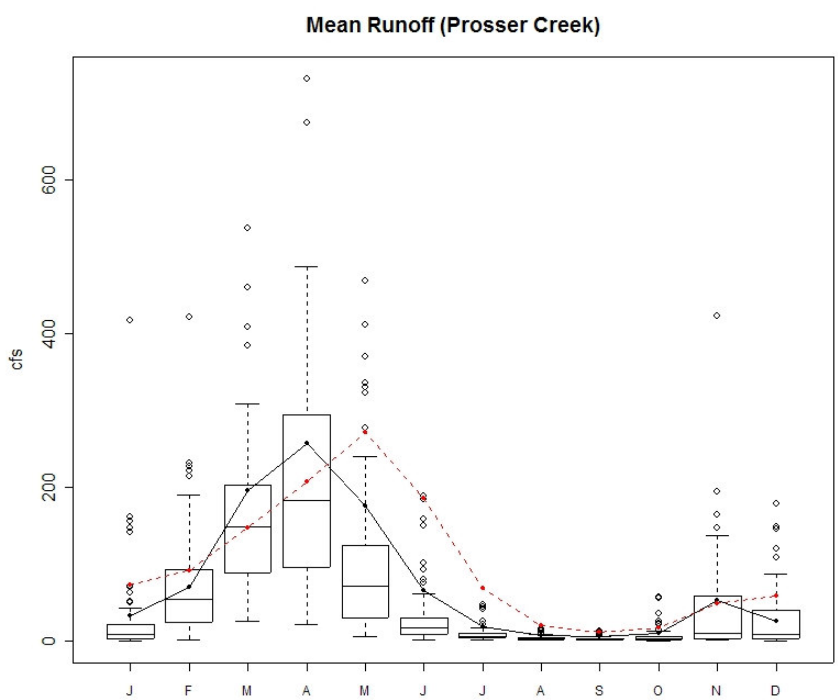

Figure 4.3. Same as Figure 4.1 but for Prosser Creek Reach.

FiguresFigure 4.4,Figure 4.5, andFigure 4.6 show the standard deviations of runoff describing the spread of data. The results are similar to the comparison of the means, indicating a fairly good approximation of the modeled statistics.

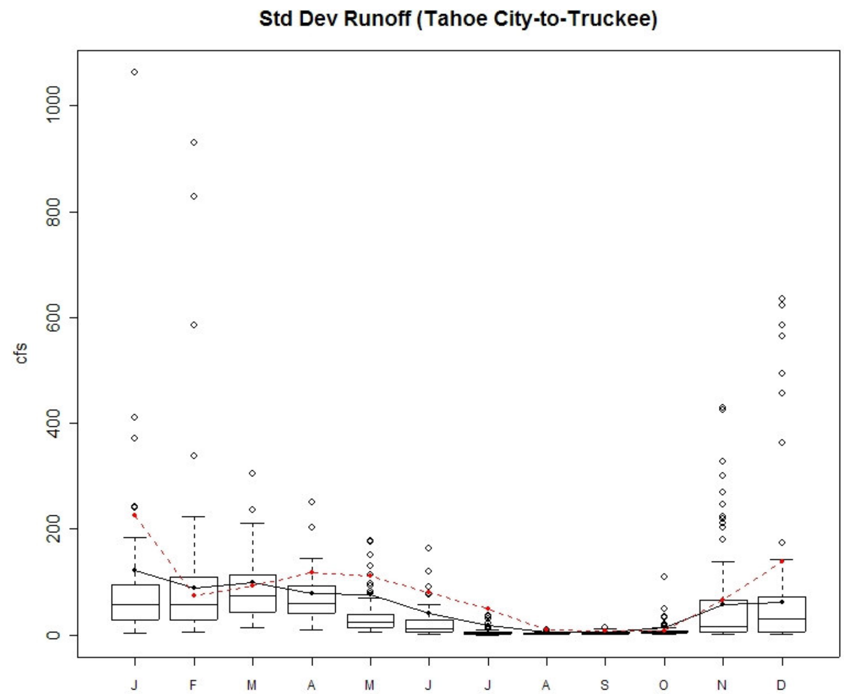

Figure 4.4. Monthly standard deviation of runoff for Tahoe City-to-Truckee Reach. Box plots represent the ensembles of output from the watershed model using the parameters simulated by the weather generator. The solid line indicates the watershed model output using the historic data (1981-1999). The dashed line represents the observed values (1981-1999).

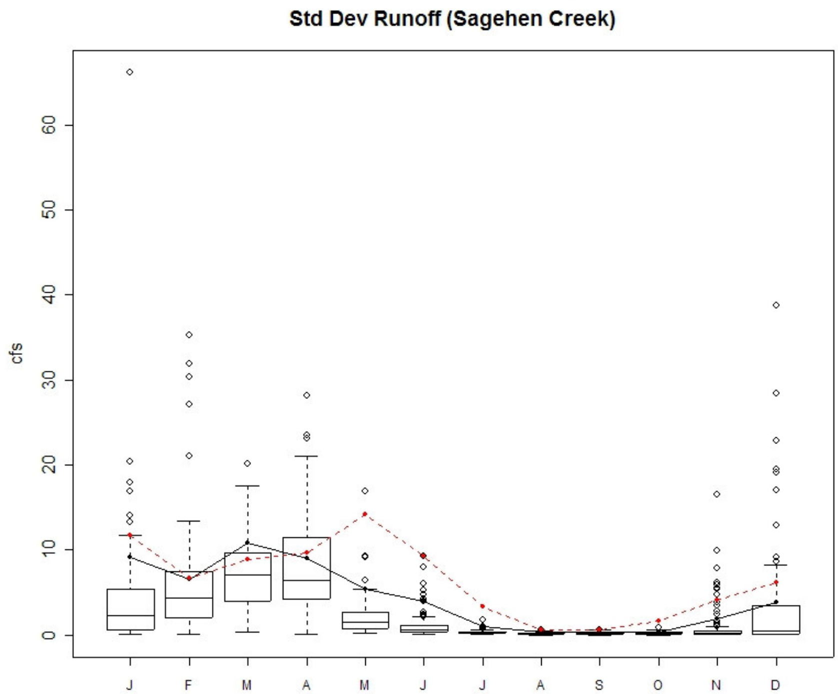

Figure 4.5. Same as Figure 4.4 but for Sagehen Creek Reach.

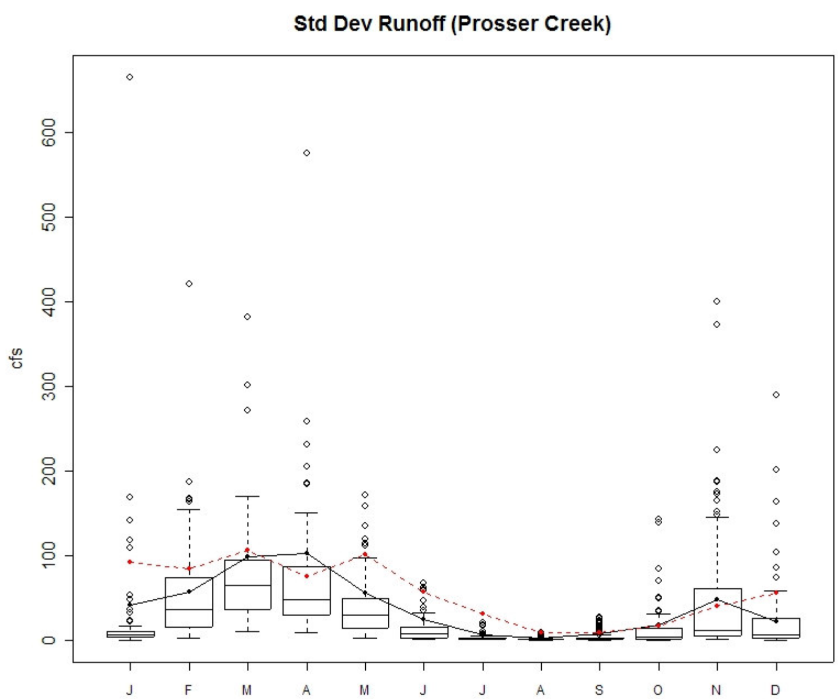

Figure 4.6. Same as Figure 4.4 but for Prosser Creek Reach.

FiguresFigure 4.7,Figure 4.8, andFigure 4.9 illustrate the skew of the runoff. It appears that the weather generator values reproduce the modeled statistic well with the exception of July and August. This would suggest that the distribution of stream flows is similar in both datasets.

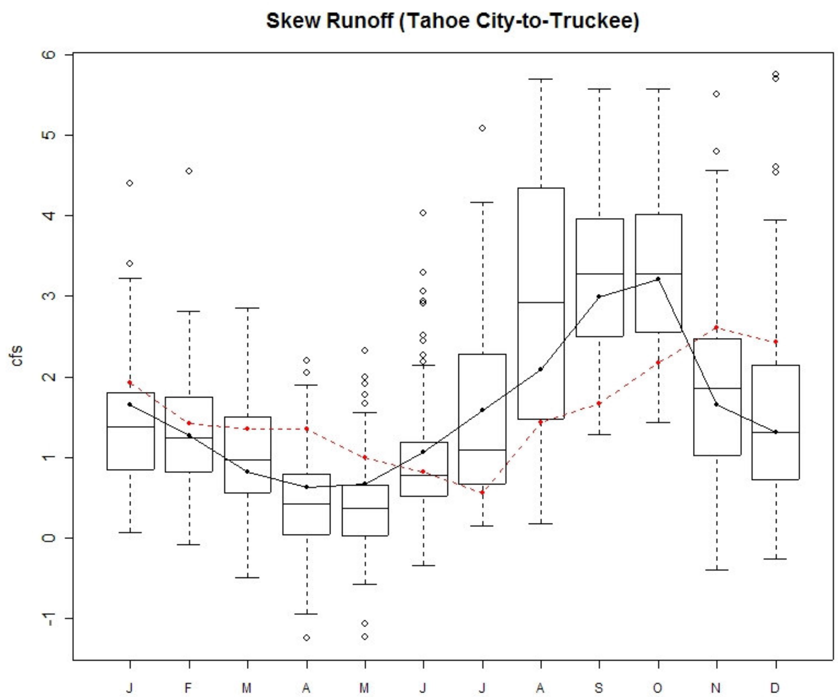

Figure 4.7. Monthly skew of runoff for Tahoe City-to-Truckee Reach. Box plots represent the ensembles of output from the watershed model using the parameters simulated by the weather generator. The solid line indicates the watershed model output using the historic data (1981-1999). The dashed line represents the observed values (1981-1999).Simulated, modeled historic and observed skews for Tahoe City-to-Truckee Reach

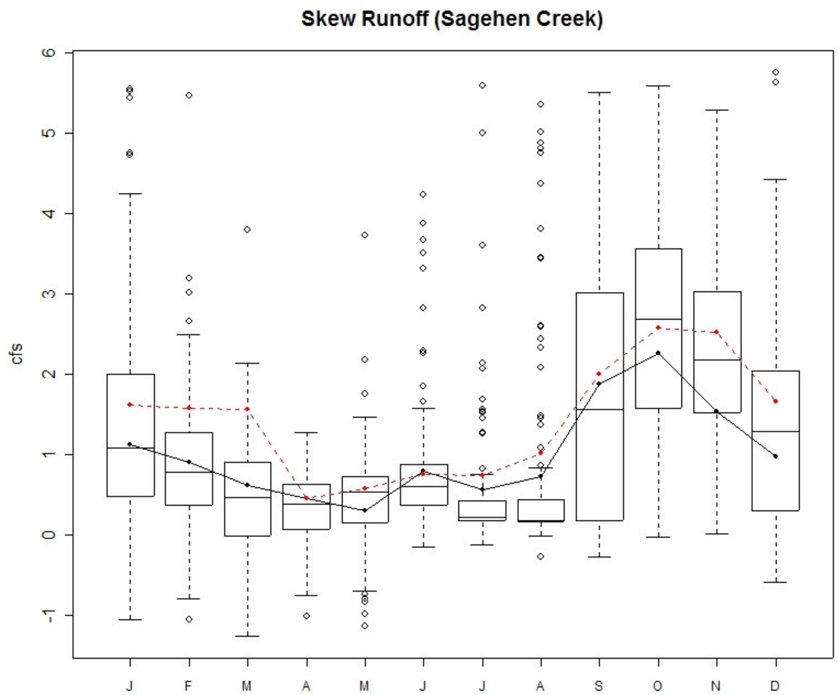

Figure 4.8. Same as Figure 4.7 but for Sagehen Creek Reach.

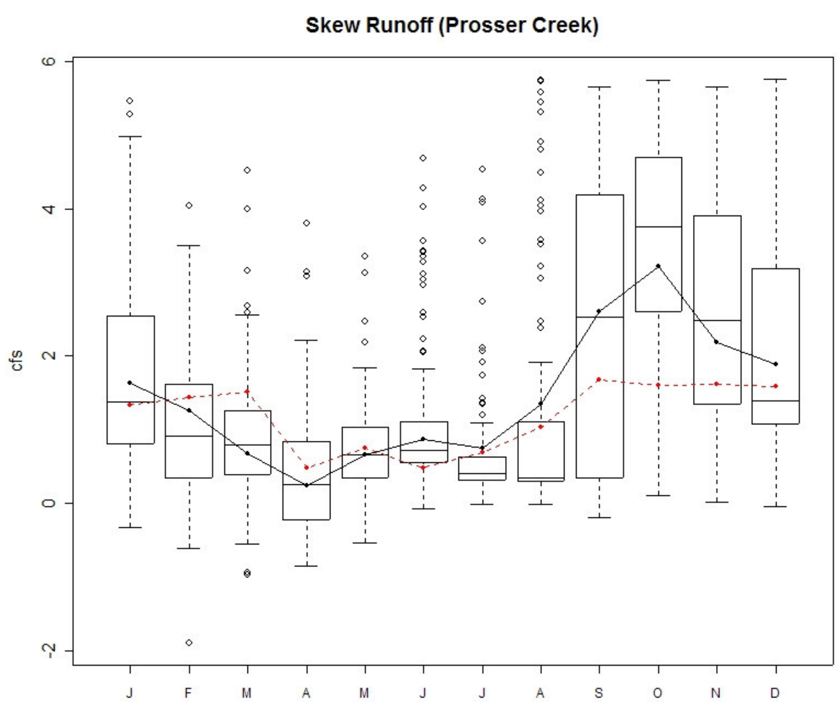

Figure 4.9. Same as Figure 4.7 but for Prosser Creek Reach.

CHAPTER 5

5. CONDITIONED FORECASTING

Adequately simulating the historic streamflow record provides a base from which to develop forecasting models. Many forecasting methods, such as ESP, rely simply on historic records to generate forecasts making the assumption that historic conditions represent future conditions. Given the lack of historic data discussed previously, this simple assumption cannot account for deviations from the static historic climate. Due to the increasing and complex demands on water resources within basins, planning and forecasting techniques must be able to account for changes in the static system. Depending upon the forecast period, different conditioners can be used. For example, Grantz [2003] demonstrated the link between streamflows in the Truckee-Carson system with Pacific Decadal Oscillation which provides an index on which to condition forecasts. In other basins, other large-scale climate indices can be used to condition forecasts. A simple technique is to condition the simulated values based on some prediction of the relative wetness of the future period of concern. If a prediction is made for a wet time period, the forecasting technique resamples from only wet years to give an ensemble of possible wet scenarios. This method can be applied within the ESP approach; however, the lack of data in the historic record can make the sample size too small to be statistically significant. Using the modified *K-nn* approach presented here the same technique can be applied to a much larger data set yielding a statistically-significant conditioned forecast.

5.1. METHODS

As a simple demonstration of this utility, the historic record of weather data was divided in to two sections, wet and dry, and run through the hydrologic models. The monthly observed historic values were used to calculate precipitation averages for each water year (October-September). Water years with means greater than the overall mean were classified as wet years while those with means less than the overall mean were classified as dry. Daily data from each conditioned water year along with all water years were processed through the weather generator to simulate wet, dry, and normal forecasts respectively. Output data were used as input to each of the hydrologic models in this study to illustrate the changes in streamflow volume and timing with the different forecasts.

5.2. RESULTS AND DISCUSSION

The monthly mean hydrographs of the conditioned and unconditioned forecasts for each sub-basin model are shown in Figures Figure 5.1, Figure 5.2, and Figure 5.3. All three basins show a dramatic change in volume between the three forecasted periods. As expected, the dry-conditioned dataset yields a forecast with a lower volume than using all of the data as is the practice with many ensemble techniques. Similarly, the volume of runoff from the wet-conditioned dataset is greater than the unconditioned case. Timing of the peak runoff between the different forecasts is approximately the same for Sagehen and Prosser Creek reaches yet interestingly, the Tahoe City-to-Truckee reach shows a shift in runoff timing with the different forecasts. Forecasted peak runoff occurs later in the season with increased runoff volumes. This can be attributed to the increased snowpack from above average precipitation. Typically, during wet periods effective incoming solar radiation and air temperatures are decreased due to increased cloud cover and increased snow-covered areas which reduce the amount of available energy to melt the snowpack.

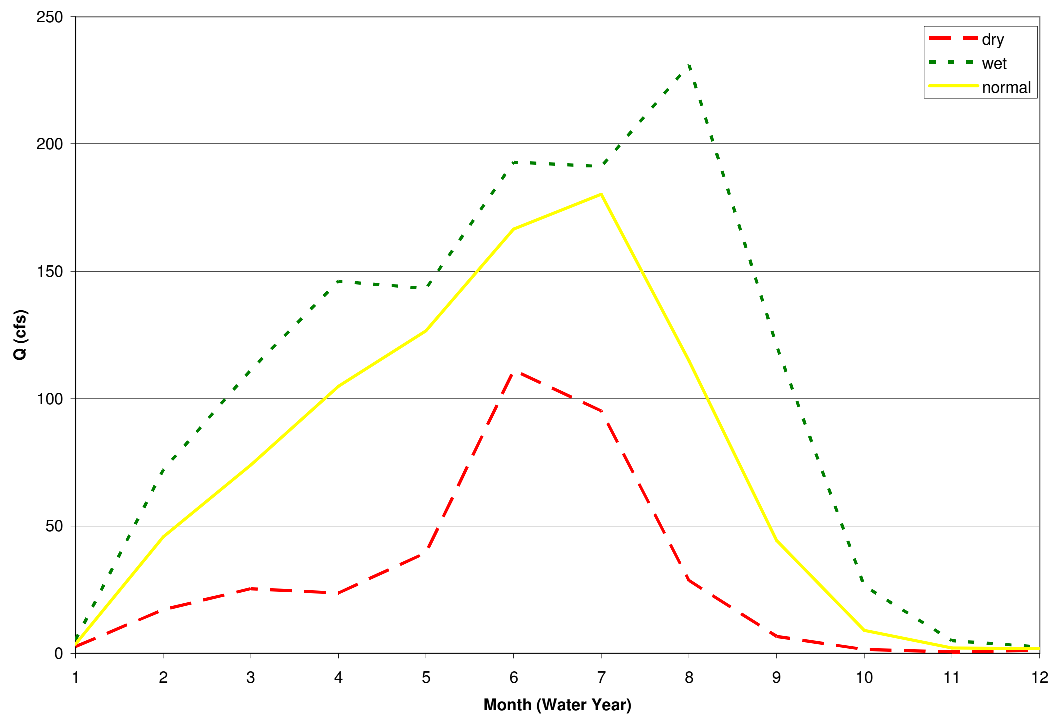

Figure 5.1. Tahoe City to Truckee Reach conditioned simulated streamflows

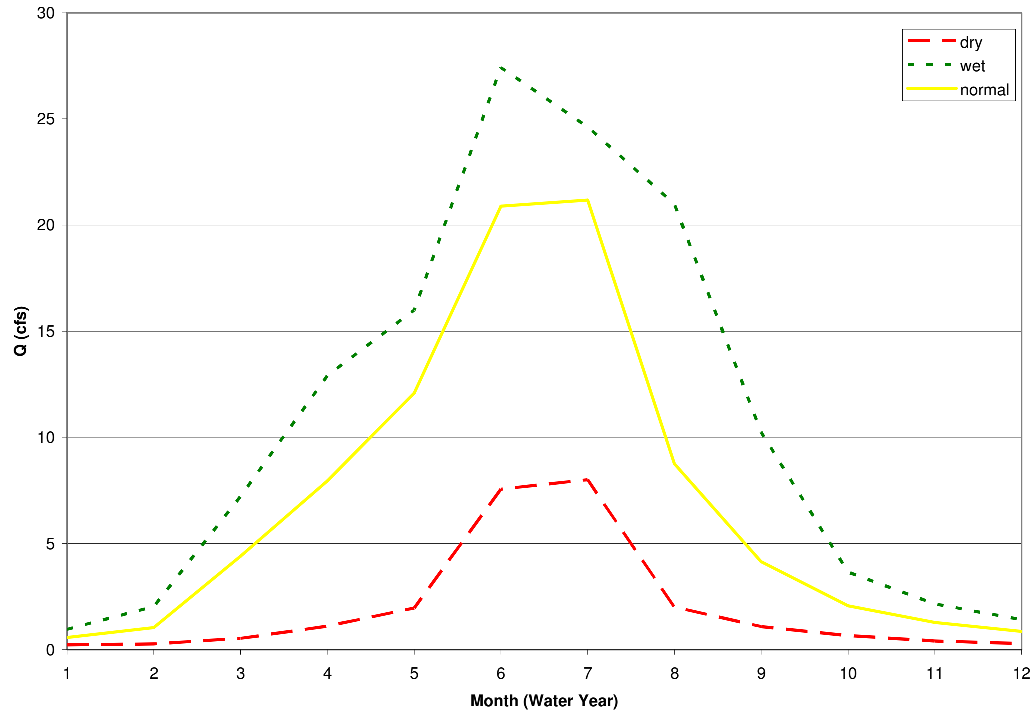

Figure 5.2. Sagehen Creek conditioned simulated streamflow

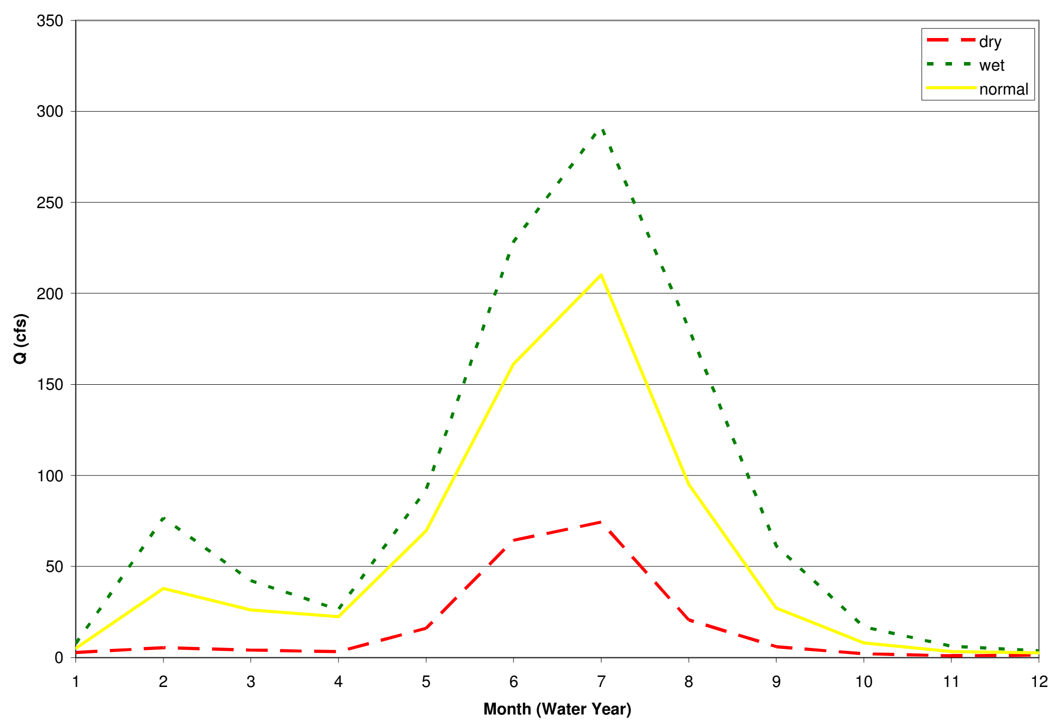

Figure 5.3. Prosser Creek conditioned simulated streamflow

Figures Figure 5.4, Figure 5.5, and Figure 5.6 show the probability distribution functions (PDF) of each forecast in each sub-basin. The PDFs demonstrate the shift in probabilistic forecasts using the conditioned datasets. Unconditioned forecasts, such as those used in traditional ESP methods and represented as the normal forecast, do not adequately capture the probabilities of flows that could be encountered. The conditioned forecasts shift the flow probabilities to provide a more accurate forecast of flows. To illustrate this point, if a forecast was given using the unconditioned case, the probability of low flow would be dramatically underestimated as compared to the dry-conditioned forecast.

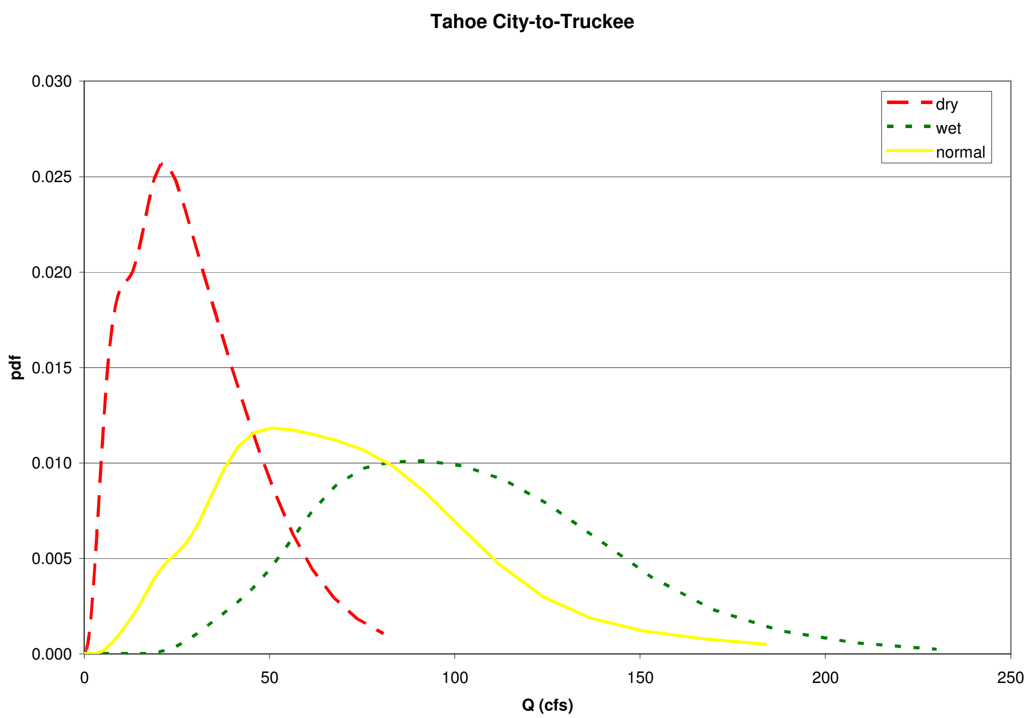

Figure 5.4. Probability distribution function of conditioned and unconditioned forecasts for Tahoe City-to-Truckee Reach. Dry-conditioned flows are represented by the large dashed line; wet-conditioned flows are represented by the small dashed line; unconditioned flows (normal) are represented by the solid line.

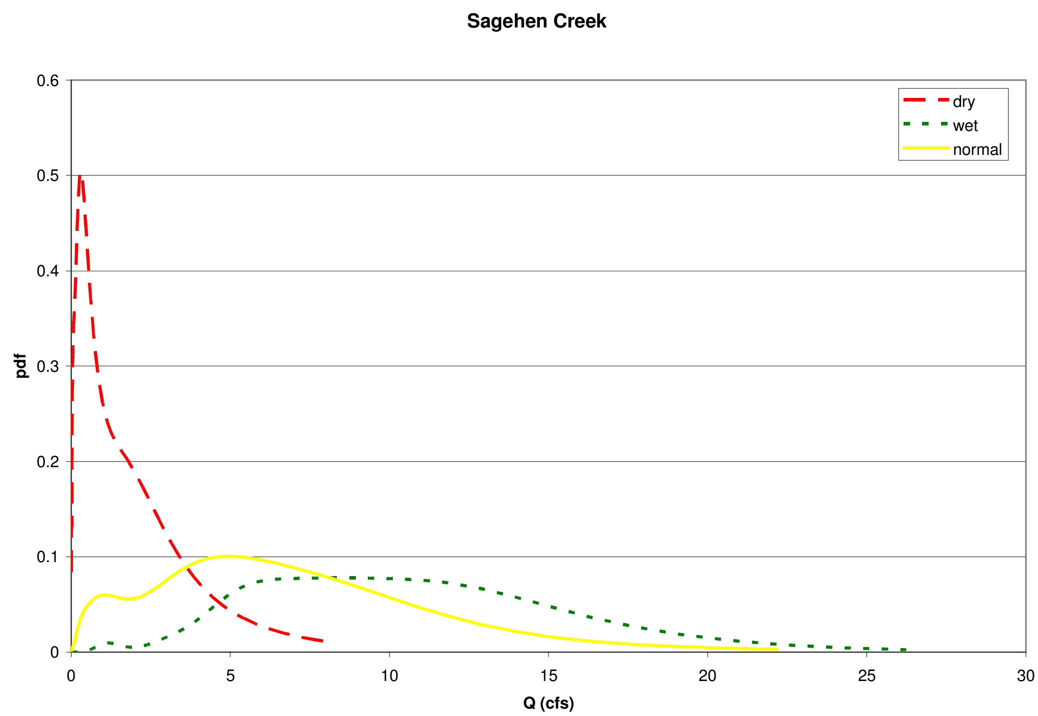

Figure 5.5. Same as Figure 5.4 except for Sagehen Creek Reach.

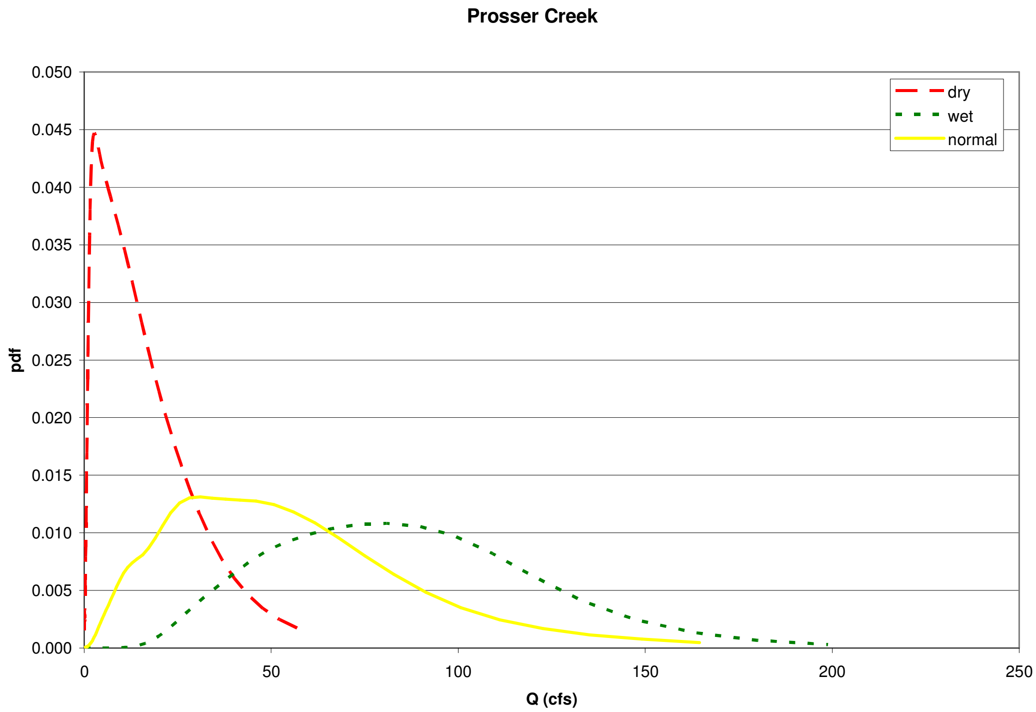

Figure 5.6. Same as Figure 5.5 except for Prosser Creek Reach

5.3. SUMMARY

Conditioned forecasts can provide better estimates of streamflow by providing additional information regarding the relationships of variables. Weather data from wet, dry, and normal years are used in the weather generator and input to the watershed model to demonstrate the utility of the weather generator in creating conditioned forecasts. Other variables such as large-scale climate indices can be used to condition data to develop a forecast [*Grantz,* 2003]. By conditioning the input to the weather generator a smarter, more robust forecast for short term and long term periods can be created.

CHAPTER 6

6. SUMMARY AND CONCLUSIONS

In an effort to improve upon current ESP techniques for streamflow forecasting, this research demonstrated the utility of coupling a stochastic weather generator with a deterministic watershed model and provided a robust framework to create streamflow forecasts. Regional daily weather variables were generated with a *K-nn* weather generator that preserved the spatial and temporal dependencies and adequately reproduced statistics of the historic weather variables in the Upper Truckee River Basin. The drawback to this technique is the requirement to utilize a complete record without missing values. Although techniques to fill in data are available, they provide additional uncertainty to the simulated dataset and may not prove to be an adequate representation of the historic record. In the case of the Upper Truckee River Basin, the only records that were acceptable were the NWS COOP stations. The NRCS SNOTEL stations had discontinuous data and could not be used with the weather generator. Neglecting the data from the high elevations would most likely bias the weather data towards the lower elevations where the snowpack tends to be smaller and melt sooner in the season.

The PRMS model within MMS was used to simulate streamflows. Weather data from the historic record were used to create an average distribution of streamflows which served as a baseline from which to evaluate simulated values. The average modeled historic streamflows did not represent the average historic observed streamflows due to the limited number of years used in calibration of the watershed model. Regardless, the calibrated model provided an adequate measure of the utility of coupling the weather generator with the watershed model and demonstrated a framework for streamflow forecasting.

Ensembles of simulated weather were used as input to the watershed model to generate an ensemble of simulated streamflows. The simulated flows reproduced the modeled historic streamflows fairly well, except in the case of the volume of modeled runoff. This can be attributed to the lack of accurate high-altitude snow data in the weather variable selection.

By demonstrating the ability of the technique to adequately simulate historic statistics, a framework is provided that illustrates how to proceed to conditioned forecasting. A conditioned forecast based on relative amounts of precipitation (wet/dry) is used to show the shift in streamflow volume. The framework provides a robust alternative to the ESP approach and allows for "what-if" scenarios for policy planning.

The next phase of this work should address the issue of missing values in the historic record of weather data. Incorporating non-continuous weather data records will reduce biases towards lower elevation data and will provide a better estimation of basin weather variables.

The weather generator should be applied to other basins in different hydrologic regimes as well as other models that require weather data as input. Appling the technique to rainfall runoff basins with different runoff timing and other process models will provide better validation of the technique for simulating weather variables.

REFERENCES

Buishand, T.A., and T. Brandsma, Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. *Water Resources Research, 37*(11), 2761-2776, 2001.

Clark, M.P., S. Gangopadhyay, D. Brandon, K. Werner, L. Hay, B. Rajagopalan, and D. Yates, A resampling procedure for generating daily weather sequences. Water *Resources Research ,40,* 2004.

Clark, M.P., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, The Schaake Shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology* 5, 243-262, 2004.

Day, G.N., Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management, 111,* 157-170, 1985.

Gangopadhyay, S., M.P. Clark, and B. Rajagopalan, Statistical Downscaling Using K-Nearest Neighbors. *Water Resources Research, 41*, 1-23, 2005.

Grantz, K.A., Using large-scale climate information to forecast seasonal streamflow in the Truckee and Carson Rivers, Civil, Environmental, and Architectural Engineering M.S. Thesis, University of Colorado, Boulder, CO, 2003.

Horton, G. A. *Truckee River Chronology,* Division of Water Planning, Department of Conservation and Natural Resources, Carson City, Nevada, 1995.

Horton, G. A., Carson River Chronology: A Chronological History of the Carson River and Related Water Issues, Nevada Division of Water Planning, Carson City, Nevada, 1996.

Hwang, Y., Impact of input uncertainty in ensemble stream flow generation, Civil, Environmental, and Architectural Engineering Ph.D. Thesis, University of Colorado, Boulder, CO, 2005.

Jeton, A.E., Precipitation-Runoff Simulations for the Upper Part of the Truckee River Basin, California and Nevada. *U.S. Geological Survey Water-Resources Investigations Report 99-4282*, Carson City, NV, 1999.

Leavesley, G.H., R.W. Lichty, B.M. Troutman, and L.G. Saindon, Precipitation-Runoff Modeling System - Users Manual. *U.S. Geological Survey Water Resources Investigations Report 83-4238*, Denver, Colorado, 1983.

Leavesley, G.H., Problems in snowmelt runoff modeling for a variety of physiographic and climatic conditions. *Hydrologic Sciences, 34,* 317-634, 1989.

Leavesley, G.H., P.J. Restrepo, S.L. Markstrom, M. Dixon, and L.G. Stannard, The Modular Modeling System (MMS) - User's Manual. *U.S. Geological Survey Open-File Report* 96-151, Denver, Colorado, 1996.

Leavesley, G.H., S.L. Markstrom, P.J. Restrepo, and R.J. Viger, A Modular Approach to Addressing Model Design, Scale, and Parameter Estimation Issues in Distributed hydrologic Modelling. *Hydrologic Processes, 16,* 173-187, 2002.

Markstrom, S., *personal communication*, 2005.

Mastin, M. C. and J. J. Vaccaro, Watershed Models for Decision Support in the Yakima River Basin, Washingtom. *U.S. Geological Survey Open-File Report 02-404*, Tacoma, Washington, 2002.

Rajagopalan, B., *personal communication*, 2005.

Rajagopalan, B., U. Lall, and D. G. Tarboton, Nonhomogeneous Markov model for daily precipitation. *Journal of Hydrologic Engineering*, *1*(1), 33-40, 1996.

Rajagopalan, B., and U. Lall, A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research, 35,* 3089-3101, 1999.

Singh, V.P., Watershed Modeling, *in* Computer Models of Watershed Hydrology, V.P Singh, eds. Water Resources Publications, Highlands Ranch, CO, 1995.

Taylor, R. L., Simulation of hourly stream temperature and daily dissolved solids for the Truckee River, California and Nevada, *Water-Resources Investigations*, USGS WRI 98-4064, 1998.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek, A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research, 39*(7)*,* 1199-1213, 2003.