A Multi-Site Stochastic Weather Generator for Improved Streamflow Forecast

Nina Caraway & Balaji Rajagopalan

University of Colorado, Boulder

Seasonal to Two-Year Colorado River Streamflow Prediction Workshop
CBRFC, Salt Lake City

March 21, 2011

Multi-Model Ensemble Combination and Conditional Stochastic Weather Generation Tool for Improved Streamflow Forecasts

- National Oceanic and Atmospheric Administration (NOAA) funded project
- NOAA program element
  - Climate Prediction Program for the Americas (CPPA)
  - Hydrologic and water resources applications
- Principal Investigator: Balaji Rajagopalan
- Collaborators from Colorado Basin River Forecasting Center (CBRFC)
  - Dr. Kevin Werner
  - Dr. Michelle Schmidt

Outline

- Background
- Motivation
  - Current Ensemble Streamflow Prediction (ESP) method
  - Need for weather generator
- Stochastic Weather Generators
  - K-Nearest Neighbor re-sampling approach
- Application
- Results
- Future plans

Background

- Stress on water resources
  - Drought
  - Socio economic growth
- Need for efficient water resources management
  - Requires skillful streamflow predictions
    - For short term (weeks) time scales
    - And long term (seasonal to inter-annual)
  - Also incorporate seasonal climate forecasts based on large-scale climate forcings

Motivation

- CBRFC & Natural Resource Conservation Service (NRCS) work together to predict streamflows at two time scales
- At seasonal time scale, CBRFC uses the following models:
  - Statistical Water Supply (SWS), a regression based method that relates observed data with future streamflow
  - **Ensemble Streamflow Prediction (ESP)**
    - **Based on historical daily weather sequence and a physically-based watershed model (such as the Sacramento Soil Moisture model, SAC-SMA)**
- NRCS uses a principle components regression technique
- Forecasts from these models are qualitatively combined to issue a single 'coordinated' forecast for U.S. Bureau of Reclamation (USBR)

Motivation: ESP

- The ESP forecasts start at time 't', proceeding as follows:
  - Calibrated watershed model is run with historical data until time 't'
  - Thus obtaining the initial condition for the hydrologic state of the basin
  - Historical weather sequence for the period 't+1' through 't+k' (desired length of forecast) is used to force the hydrologic model
  - Which creates an ensemble of streamflow sequences
    - Where the number of ensembles is equal to number of historical years

Motivation: ESP

- Limited historical data means limited ensembles
- Incorporating seasonal forecasts further reduces the number of ensembles
  - For instance, forecasting based on warm ENSO phase
- Need arises for a simple and efficient approach to generate a 'rich variety' of streamflow ensembles
  - Will allow for robust estimation of probability density functions (PDFs)
- Hence bringing in stochastic weather generators

Stochastic Weather Generators

- Traditional weather generators are parametric
  - Generate ensembles of weather sequences
  - Employ Markov chain for precipitation occurences
  - PDFs (Gamma, Log-Normal, etc) for precipitation amounts
  - AR-1 for maximum and minimum temperatures (e.g., Richardson 1981)
- Extensions to multi-site were not trivial
- Extensions to 'conditional' generation (i.e., conditioned on seasonal climate forecasts) also proved difficult

Stochastic Weather Generators

- Nonparametric weather generators offer attractive alternative
  - Data-driven
    - Thus can 'capture' deviations from theoretical probability distributions
    - And also nonlinearities between variables
  - Can be based on kernel density estimators (Rajagopalan et al., 1996)
  - Or use resampling (Lall and Sharma, 1996; Rajagopalan and Lall, 1999)

K-Nearest Neighbor Weather Generator

- Semi-parametric weather generator (Apipattanavis et al., 2007)
  - First precipitation state is generated by a Markov chain fitted to the historical data (wet or dry)
  - Then precicipation time series is created using Markov chain
  - A K-Nearest Neighbor (KNN) method is applied to the time series, which can be expressed as simulating from the conditional PDF:

$$f(x_t \mid x_{t-1}, S_t, S_{t-1})$$

  - where x_t and x_{t-1} are the weather states and S_t and S_{t-1} are the precipitation state on day t and $t-1$

K-Nearest Neighbor Method

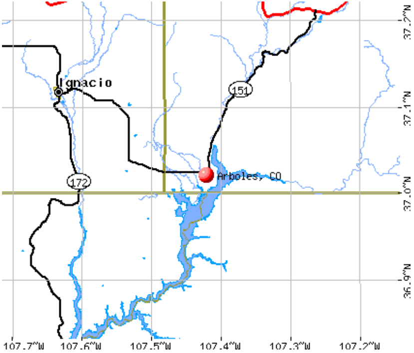
- Suppose January 1st is the simulated day of interest
- A 7-day (can be user defined) window is placed on January 1st (i.e. December 29th to January 4th)
  - This window around a given point is defined as a 'neighborhood'
- Calculates weighted Euclidean distance between weather variables of current day and neighbors
- Nearest neighbor receives a higher weight and the kth neighbor gets the least
- One of the historical days within the 7-day window is selected based on the previously calculated weights
- For example, a simulated January 1st could be January 3rd, 1985 from historical data

K-NN: A Physical Example

- Unconditional resampling
  - Drawing cards from a well-shuffled deck
  - Corresponds to selecting a (single or a set of) historical years from the record, with equal chance
- Conditional resampling
  - Drawing cards from a *biased* deck
  - Corresponds to selecting a (single or a set of) historical years with *unequal* chance

Application

Single Site Application



- Arboles, Colorado in San Juan Basin
- Three elevation zones available, lower was chosen
- Daily weather variables 1961-2004
  - Precipitation
  - Maximum temperature
  - Minimum temperature
- 50 simulations each 44 years long
- Statistics of simulated and historical data were compared

Multi Site Application

- The three stations (upper, middle, and lower) were spatially averaged to produce a 'synthetic' single site time series
  - Daily weather is generated at all locations simultaneously
    - Captures spatial dependency automatically
- Future modifications include elevation weighting or Principal Component Analysis (Yates et al., 2003)

Validation

- A suite of statistics were computed from the simulations and compared with historical observations
  - Displayed as boxplots
- Distributional Statistics
  - Mean
  - Standard Deviation
  - IQR
  - Skew
  - Probability density functions (PDFs)
- Threshold exceedances and extremes
  - Average number of wet and dry days
  - Total rainfall exceeding a threshold (e.g, 75th percentile)

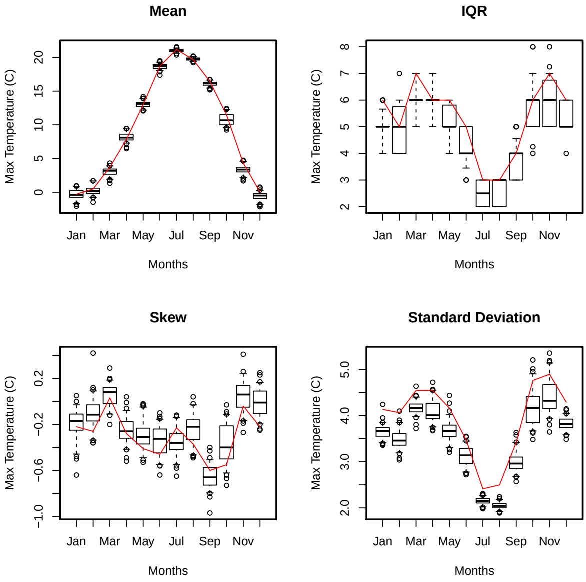


Figure: Single Site Max. Temperature

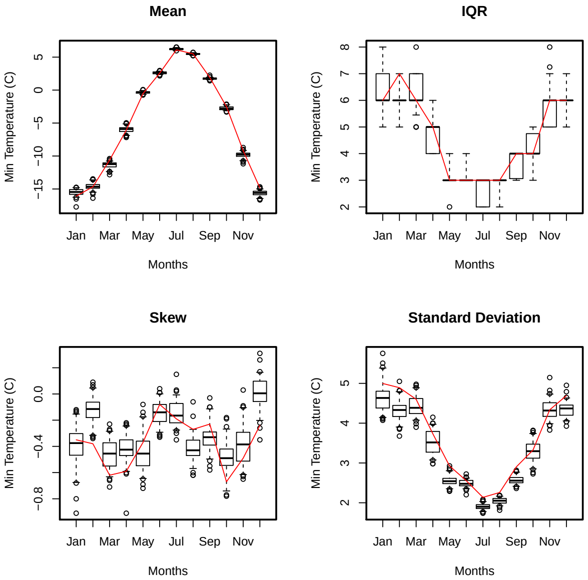


Figure: Single Site Min. Temperature

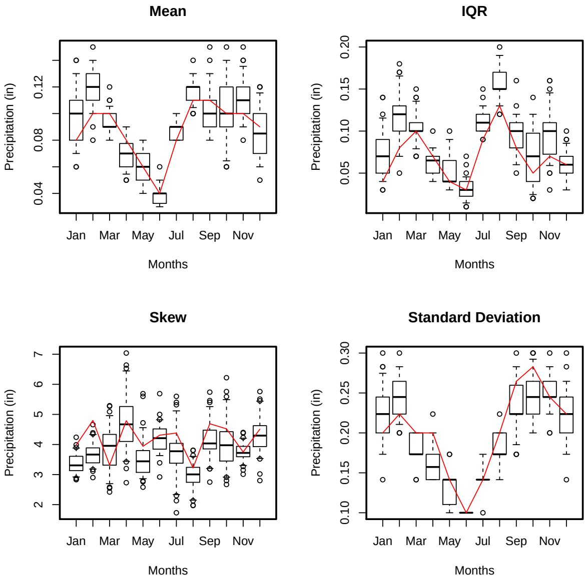


Figure: Single Site Precipitation

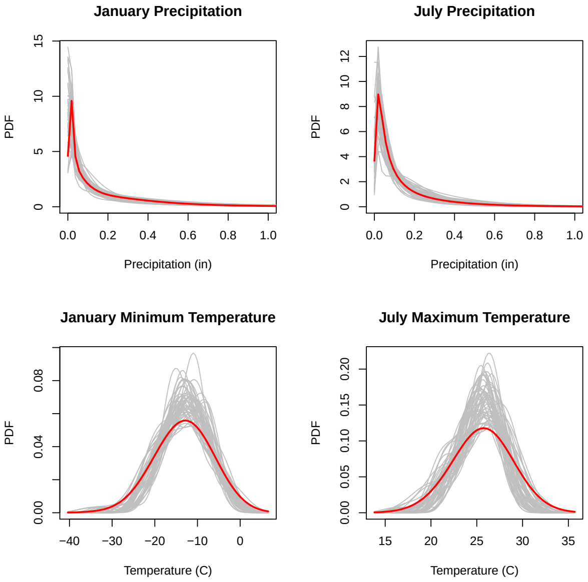


Figure: Single Site PDFs

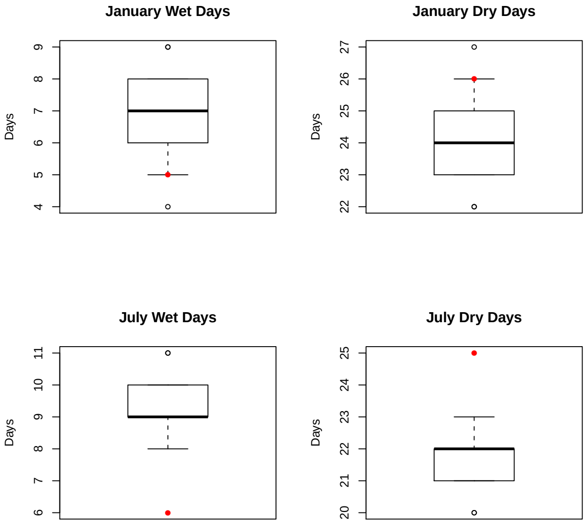


Figure: Average Wet and Dry Days in Selected Months

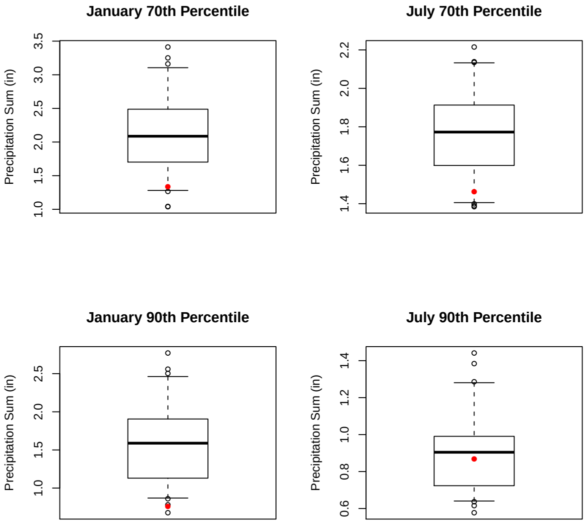


Figure: Precipitation Sums Above Thresholds

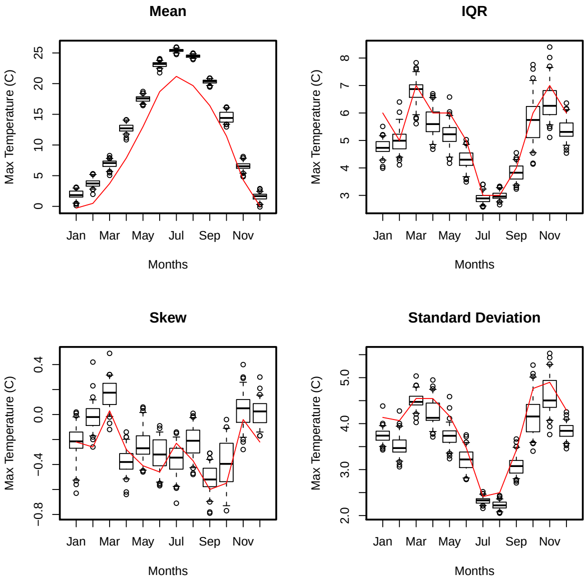


Figure: Single Site Max. Temperature from Averaged Multi Site

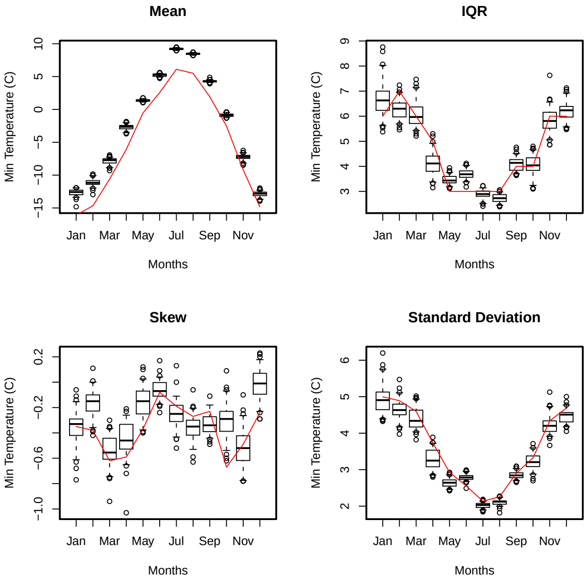


Figure: Single Site Min. Temperature from Averaged Multi Site

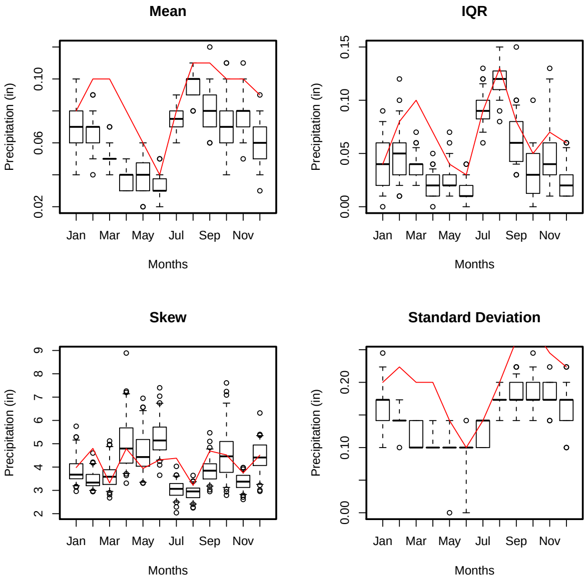


Figure: Single Site Precipitation from Averaged Multi Site

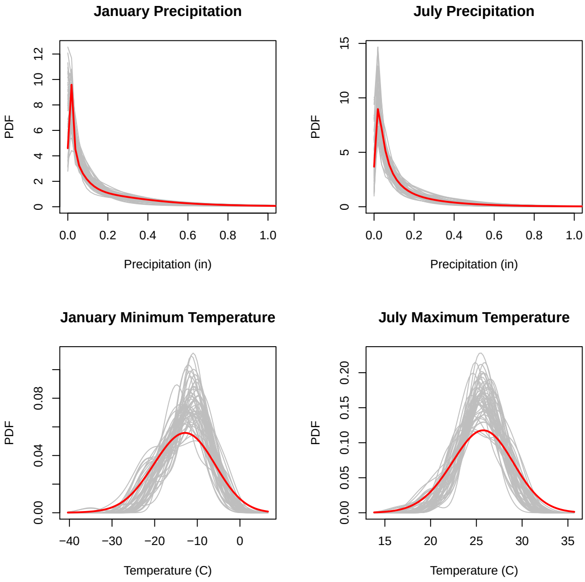
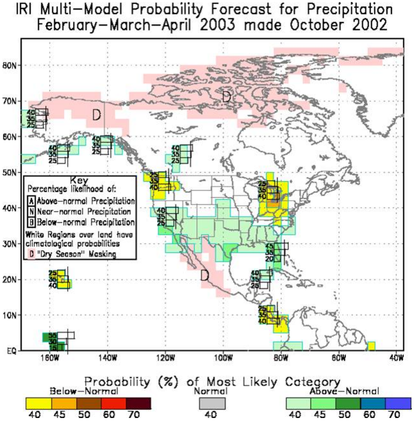


Figure: Single Site PDFs from Averaged Multi Site

Conditional Simulation



- Conditioned on International Research Institute (IRI) seasonal forecast
- So if prediction is A:N:B = 40:35:25
- Divide historical (seasonal) total into 3 tercile categories
- Bootstrap 40, 35, and 25 samples of historical years from wet, normal, and dry categories
- Then apply weather generator

Summary and Plans for Near-Term

- K-NN weather generator implemented and tested on small region in CO river basin
  - Historical statistics were well reproduced
- Further testing will include
  - Additional statistics, such as extreme precipitation, hot and cold spells, etc
  - Better weighting approaches to generate the 'synthetic' single site
- Also will perform testing on conditional simulation based on seasonal climate forecasts
- Then testing will be performed on other sites in Upper Colorado River Basin
- Ultimately, multi-site weather sequences will be driven through SAC-SMA
  - Performance of the streamflow ensembles will be evaluated
  - Has been done with Precipitation-Runoff Modeling System (PRMS) before (Apipattanavis et al., 2007)

How will USBR and CBRFC benefit?

- Project has 2 key things to develop
  - A conditional stochastic weather generator to provide daily weather ensembles based on NWS short term and NOAA seasonal outlooks
  - An optimal multi-model ensemble combination, which will provide a combined ensemble forecast from physical and statistical models
- Project will also build on a multi-model statistical streamflow forecast tool
  - Demonstrated on the Gunnison River Basin (Regona et al., 2006) and Upper Colorado River Basin (Bracken et al., 2010)
- These new and improved forecasts will be used for efficient operation and management of major reservoirs
  - Thus impacting water resources, agriculture, hydropower, and aquatic environments in the southwest and inter mountain regions of western U.S.

Questions?