Stochastic Nonparametric Framework for Basin Wide Streamflow and Salinity Modeling

Application to Colorado River basin

Study Progress Meeting
James R. Prairie
May 2, 2006
Progress Update

• **Completed Flow Disaggregation**
  – Revising journal article for Water Resources Research
    • Describes algorithm and applies to 4 key gauges
• **Completed Basin Wide Salinity Model**
  – Reviewing draft journal article
    • Describes algorithm and applies to 4 key gauges
• **Generating streamflow conditioned on Paleo Climate data**
  – Using block bootstrap resampling
  – Using transition probability matrices
    • Determine transition probability with reconstructed streamflows
    • Generate flows with Markov chains
Masters Research
Single site
Modified K-NN streamflow generator
Climate Analysis
Nonparametric Natural Salt Model

Stochastic Nonparametric Technique for Space-Time Disaggregation

Basin Wide Natural Salt Model

Incorporate Paleoclimate Information
Transition probability matrices (TPM)
Generate streamflow conditioned with TPM

Policy Analysis
Impacts of drought
Hydrology
Water quality
Proposed Methods

- Block bootstrap resampling
- Markov model

Generate system state \((S_t)\)

Generate flow conditionally

\(f(x_t | S_t, S_{t-1}, x_{t-1})\)
Datasets

• Paleo reconstruction from Woodhouse et al. 2006
  – Water years 1490-1997
• Observed natural flow from Reclamation
  – Water years 1906-2003
Generating paleo conditioned flows

- Combines
  - Observed streamflow record: models magnitude
  - Transition probability matrices: models frequency/state
- Conditioned on flow state (e.g., wet or dry)

- Methods to generate flow state series
  1. Block bootstrap resampling
  2. Markov chain model (Transition Probability Matrix)
Steps to generate flows with resampling

1. Block bootstrap reconstructed (paleo) flow
   - 30 year block
   - 98 year simulation length \( N \)
   - 500 simulations

2. Compute states for each simulation
   - 0011011110001010010000100

\[ S(t) \quad t = 1, \ldots, N \]

3. Determine flow by resampling observed flow \((x_t)\) with a K-NN algorithm conditioned on current and previous state and previous flow

\[ f(x_t | S_t, S_{t-1}, x_{t-1}) \]
Transition Probability Matrix (TPM)

- **Markov Chain defined by** \( P(t) \)
  \[
  \begin{bmatrix}
  p_{00} & p_{01} \\
  p_{10} & p_{11}
  \end{bmatrix}
  \]
  \[
  \begin{bmatrix}
  1 & 0 \\
  0 & 1
  \end{bmatrix}
  = 1
  \\
  \begin{bmatrix}
  0 & 1 \\
  1 & 0
  \end{bmatrix}
  = 1
  
  - **Multiple** \( P(t) \) **computed with moving window**
  - **From** \( P(t) \) **we simulate the state (wet or dry)**
Steps to generate flows with TPM

1. Compute TPM for required states
   - 2 states (wet and dry)
   - 30 year moving window

2. Generate Markov chains with TPM
   - 00110111100010100100001000100 $S(t) \quad t = 1, \ldots, N$
   - 98 years in length $= N$
   - 500 simulations

3. Determine flow by resampling observed flow $(x_t)$ with a K-NN algorithm conditioned on previous flow and previous state
   $$f(x_t | S_t, S_{t-1}, x_{t-1})$$
Observed

- 30 year Block bootstrap
- 500 simulations
- 98 year length
Observed

- 30 year Block bootstrap
- 500 simulations
- 98 year length
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulations
- 98 year length
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulations
- 98 year length
Paleo Conditioned

- 30 block bootstrap
- 30 year Markov chains
- 500 simulations
- 98 year length
Sequent Peak Algorithm

- Determine required Storage Capacity ($S_c$) at various demand levels given specified inflows.
- Evaluate risk of not meeting the required $S_c$

\[
S'_i = \begin{cases} 
S'_{i-1} + d - y_i & \text{if positive} \\
0 & \text{otherwise}
\end{cases}
\]

\[y = \text{inflow time series}\]
\[d = \text{demand level}\]
\[S = \text{storage}\]
\[S'_0 = 0\]

\[S_c = \max \left[ S'_1, \ldots, S'_N \right]\]
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulations
- 98 year length
Paleo Conditioned

- 30 year block bootstrap
- 30 year Markov chains
- 500 simulations
- 98 year length
Drought and Surplus Statistics

- Threshold (e.g., mean)
- Drought Length
- Surplus Volume
- Drought Deficit

flow vs. time
Observed

- 30 year block bootstrap
- 500 simulations
- 98 year length
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulation
- 98 year length
Paleo Conditioned

- 30 year block bootstrap
- 30 year Markov chains
- 500 simulation
- 98 year length
Transition Probability Matrices

- **Wet**
  
  \[
  \begin{bmatrix}
  0 & 1 \\
  0 & 0.2 & 0.8 \\
  1 & 0.125 & 0.875
  \end{bmatrix}
  \]

- **Dry**
  
  \[
  \begin{bmatrix}
  0 & 1 \\
  0 & 0.765 & 0.235 \\
  1 & 0.417 & 0.583
  \end{bmatrix}
  \]
Wet TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0 & 1 \\
0.2 & 0.8 \\
0.125 & 0.875
\end{bmatrix}
\]
Wet TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0 & 1 \\
0 & 0.2 & 0.8 \\
1 & 0.125 & 0.875
\end{bmatrix}
\]
Wet TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0 & 1 \\
0 & 0.2 & 0.8 \\
1 & 0.125 & 0.875
\end{bmatrix}
\]
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulations
- 98 year length
Wet TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0 & 1 \\
0.2 & 0.8 \\
0.125 & 0.875
\end{bmatrix}
\]
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulation
- 98 year length
Dry TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0 & 1 \\
0 & 0.765 & 0.235 \\
1 & 0.417 & 0.583 
\end{bmatrix}
\]

![Annual Flow vs Probability Density Graph]

- Observed record (WY 1906-2003)
Dry TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0 & 1 \\
0 & 0.765 & 0.235 \\
1 & 0.417 & 0.583
\end{bmatrix}
\]
Dry TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{bmatrix}
0.765 & 0.235 \\
0.417 & 0.583 \\
\end{bmatrix}
\]
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulations
- 98 year length
Dry TPM

- 30 year Markov chains
- 500 simulations
- 98 year length

\[
\begin{pmatrix}
0.765 & 0.235 \\
0.417 & 0.583
\end{pmatrix}
\]
Paleo Conditioned

- Random TPM
- 30 year Markov chains
- 500 simulation
- 98 year length
Conclusions

• Combines strength of
  – Reconstructed paleo streamflows: system state
  – Observed streamflows: flows magnitude

• Develops a rich variety of streamflow sequences
  – Generates sequences not in the observed record
  – More variety: block bootstrap reconstructed streamflows
  – Most variety: develop Markov chain from TPM

• TPM provide flexibility
  – Use TPM to mimic climate signal (e.g., PDO)
  – Generate drier or wetter than average flows
Next steps

• Determine proper Markov order
  – Maximum Likelihood

• Develop for a 3 state system
  – Wet, Dry, Normal

• Generate paleo conditioned streamflows for network
  – Disaggregate streamflows to appropriate temporal and spatial scale
  – Drive decision support system
  – Perform policy analysis
    • Compare results from two hydrologies
      – Our conditioned streamflows
      – Index Sequential Method (current Reclamation technique)
Additional Research Information

http://cadswes.colorado.edu/~prairie/ResearchHomePage.html
Incorporate paleo state information

• Magnitudes of Paleo data in question?
  – Address issue, use observed data to represent magnitude and paleo reconstructed streamflows to represent system state
  – Generate streamflows from the observed record conditioned on paleo streamflow state information

• Outline technique
  – State conditioned KNN streamflow generation
Block Bootstrap Data (30 year blocks)

Determine TPM (30 year moving window)

Compute state information

Markov model

Use KNN technique to resample natural flow data consistent with paleo state information

Paleo Reconstructed Streamflow Data

Natural Streamflow Data

Choose one path

Categorize natural flow data
Acknowledgements

• To my committee and advisor. Thank you for your guidance and commitment.
  – Balaji Rajagopalan, Edith Zagona, Kenneth Strzepek, Subhrendu Gangopadhyay, and Terrance Fulp
• Funding support provided by Reclamation’s Lower Colorado Regional Office
• Logistical support provided by CADSWES