Seasonal to Inter-Annual Streamflow Simulation and Forecasting on the Upper Colorado River Basin and Implications for Water Resources Management

Cameron Bracken
M.S. Candidate

Advised by Prof. Balaji Rajagopalan and Prof. Edith Zagona

September 21, 2011
OUTLINE

- Intro and Background
- Seasonal Forecasts
- Hidden Markov Models
  - Annual Simulations
  - Two Year Forecasts
- Stochastic Reservoir Ops Model
Study Area

- Parts of five states with an area of 279,300 mi².
- Elevations ranging from 4000 to 14,200 ft.
- Nearly 80% of the streamflow in the UCRB is due to snowmelt during the Apr-July period.
- 19 important gages in the Upper Colorado River Basin.
  - Lees Ferry and three main tributaries
**Motivation**

- The recent dry period (2000-2010) in the Upper Colorado River Basin (UCRB).
- Lowest inflows in Lake Powell since filling.
**Motivation**

- The recent dry period (2000-2010) in the Upper Colorado River Basin (UCRB).
- Lowest inflows in Lake Powell since filling.
Why is skillful long lead forecasting important?

- Flood control
- Drought mitigation
- Municipal and agricultural water supply
- Recreation
- Trans-basin diversions
- Hydropower
- E-flows
- ...
**SECOND YEAR**

- Decisions for Powell are made annually
- Annual Operating Plan, August
- Project Equalization

<table>
<thead>
<tr>
<th>Elevation (feet)</th>
<th>Operation According to the Interim Guidelines</th>
<th>Live Storage (maf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,700</td>
<td>Equalization Tier, Equalize, avoid spills or release 8.23 maf</td>
<td>24.3</td>
</tr>
<tr>
<td>3,646.26</td>
<td></td>
<td>16.75</td>
</tr>
<tr>
<td>3,575</td>
<td>Upper Elevation Balancing Tier² Release 8.23 maf; if Lake Mead &lt; 1,075 feet, balance contents with a min/max release of 7.0 and 9.0 maf</td>
<td>9.5</td>
</tr>
<tr>
<td>3,525</td>
<td>Mid-Elevation Release Tier Release 7.48 maf; if Lake Mead &lt; 1,025 feet, release 8.23 maf</td>
<td>5.9</td>
</tr>
<tr>
<td>3,490</td>
<td>Lower Elevation Balancing Tier Balance contents with a min/max release of 7.0 and 9.5 maf</td>
<td>4.0</td>
</tr>
<tr>
<td>3,370</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

1 January 1, 2012, projections are based on the August 2011 24-Month Study.
Why is Skillful Long Lead Forecasting Difficult?

- Highly variable
  - Low autocorrelation
- Highly seasonal / Very dependent on snowpack
  - Nearly 80% of flow occurs in April-July.
- Very dependent on snowpack
CURRENT FORECASTING

- Current forecasts are made using snowpack and soil conditions (and some climate information).
  - ESP + SWS + expertise = Coordinated Forecast
  - First peak season forecasts made starting in January
  - No second year forecast
**Forecasting Goals**

- Skillful Seasonal Flow forecasts
- Skillful Second year forecasts
- Earlier lead times than current operational forecast
- Ensemble Forecasts
LARGE-SCALE CLIMATE INFLUENCE ON BASIN-SCALE HYDROLOGY

How can skillful predictions be made in the earlier in the winter/spring season when snowpack data is unavailable or incomplete?
LARGE-SCALE CLIMATE INFLUENCE ON BASIN-SCALE HYDROLOGY

How can skillful predictions be made in the earlier in the winter/spring season when snowpack data is unavailable or incomplete? Large-scale climate variables can be used as predictors of peak season streamflow [Grantz et. al. 2005] [Regonda et. al. 2006].
Bracken et al. 2010 demonstrated the feasibility of simultaneously forecast many spatial locations while preserving spatial dependencies.

- Predictors are: PDSI, SST, Zonal/Meridional Winds, Geopotential Height, SWE
DISAGGREGATION

Useful for input to other models
**Seasonal Forecast Results: Cross-Validation**

<table>
<thead>
<tr>
<th>Validation mode</th>
<th>apr1</th>
<th>feb1</th>
<th>jan1</th>
<th>nov1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leave-one</td>
<td>0.85</td>
<td>0.74</td>
<td>0.49</td>
<td>0.30</td>
</tr>
<tr>
<td>Retroactive</td>
<td>0.62</td>
<td>0.58</td>
<td>0.55</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Table: Lees Ferry total flow forecast skills

(a) RPSS = 0.90 MC = 0.86
Two Year Forecasts

1. Poor/missing climate/snowpack information
2. A logical step is to use time series methods, ARMA, KNN, MC, frequency Domain Methods
3. Goal to make predictions of the following (second) year seasonal flow.

Year 0
- Nov1
- Feb1
- Nov-Mar

Year 1
- Apr-Jul

Year 2
- Jan-Dec
TWO YEAR FORECASTS

“Lees Ferry (Natural Flow) series is a time series modeler’s nightmare”

- Balaji Rajagopalan, 2011
“Lees Ferry (Natural Flow) series is a time series modeler’s nightmare”

- Balaji Rajagopalan, 2011

Regime switching behavior.
**Hidden Markov Models**
**Hidden Markov Models**

\[
\Pr(C_t | C^{(t-1)}) = \Pr(C_t | C_{t-1}), t = 2, 3, \ldots
\]

\[
\Pr(X_t | X^{(t-1)}, C^{(t)}) = \Pr(X_t | C_t), t \in \mathbb{N}
\]

1. General time series model
2. Markov process determines ‘hidden’ state, state dictates component distribution
3. A model that includes discrete states makes intuitive sense given the concept of climate regimes (such as El Niño)
**Hidden Markov Models**

The process generating the observations is demonstrated again in Figure 2.3, for state-dependent distributions $p_1(x)$ and $p_2(x)$, stationary distribution $\delta = (0.75, 0.25)$, and t.p.m. $\Gamma = (0.9, 0.1, 0.3, 0.7)$. In contrast to the case of an independent mixture, here the distribution of $C_t$, the state at time $t$, does depend on $C_{t-1}$. As is also true of independent mixtures, there is for each state a different distribution, discrete or continuous.

We now introduce some notation which will cover both discrete- and continuous-valued observations. In the case of discrete observations we define, for $i = 1, 2, \ldots, m$, $p_i(x) = \Pr(X_t = x | C_t = i)$. That is, $p_i$ is the probability mass function of $X_t$ if the Markov chain is in state $i$ at time $t$. The continuous case is treated similarly: there we define $p_i$ to be the probability density function of $X_t$ if the Markov chain is in state $i$ at time $t$.

Flexibility over explicit MC states.
PARAMETERS

Model Order

\[ m \]

Distribution parameters

Gamma- \( \beta_i, k_i \)
Normal - \( \mu_i, \sigma_i \)

Transition Probabilities

\[ \Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{bmatrix} \]

Stationary Distribution

\[ \delta \]
**LEES FERRY HMM**

\[ \delta = [0.8, 0.2] \]

\[ \delta = [0.79, 0.21] \]

<table>
<thead>
<tr>
<th>Model</th>
<th>BIC</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM2G</td>
<td>621.34</td>
<td>605.42</td>
</tr>
<tr>
<td>HM2N</td>
<td>620.91</td>
<td>604.98</td>
</tr>
<tr>
<td>HM3G</td>
<td>647.26</td>
<td>615.42</td>
</tr>
<tr>
<td>HM3N</td>
<td>643.73</td>
<td>611.89</td>
</tr>
</tbody>
</table>
GLOBAL DECODING

Flow Volume [MAF]

1920 1940 1960 1980 2000
5 10 15 20 25

State
1 2
1920 1940 1960 1980 2000
Hidden Markov Models for Simulation

- HMMs are also useful for simulation (used in risk analysis).
- Alternative to AR simulations
- Can we capture longer period variability?
SIMULATION STATISTICS

**Mean**

**SD**

**Skew**

**Min**

**Max**

**Lag 1 AC**
**Simulated Spell Lengths**

HMMs can capture longer spells than AR models.
HMM Forecasting

<table>
<thead>
<tr>
<th>Model</th>
<th>MC</th>
<th>RPSS</th>
<th>Dry RPSS</th>
<th>Ave RPSS</th>
<th>Wet RPSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM2G</td>
<td>0.31</td>
<td>0.21</td>
<td>0.26</td>
<td>−0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>HM2N</td>
<td>0.24</td>
<td>0.17</td>
<td>0.20</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>AR1</td>
<td>0.07</td>
<td>−0.03</td>
<td>0.07</td>
<td>−0.03</td>
<td>−0.25</td>
</tr>
</tbody>
</table>
DECISION-SUPPORT MODEL
DEcision-Support Model

- A forecast is only as good as the decisions it is used to make
- Decision support models aid in the decision making process
Current Operational Modeling

9 Upper Basin Reservoirs

The 24 Month Study
- Single trace
- Outflows are input
- Deterministic output
- Run monthly since late 90’s
**Current Operational Modeling**

9 Upper Basin Reservoirs

The 24 Month Study
- Single trace
- Outflows are input
- Deterministic output

- Run monthly since late 90’s

Midterm Operations Model
- Arbitrary number of traces
- Outflows computed on-the-fly
- Probabilistic output

- In development since Early 2009
EXPERIMENTAL SETUP

- Combine seasonal and two year forecasts
- Must run at a recent time because operations change
- Three lead times

- Nov 07
- Jan 08
- Apr 08
- Aug 08
- Jan 09
- Nov 09
- Jan 10
- Aug 10
- Dec 10

- CBRFC Forecast
- Seasonal Forecast
- Historical Resample
- HMM Year 1
- HMM Year 2
EXPERIMENTAL SETUP

![Graph showing seasonal forecasting and hidden Markov model results.]

- **Y-axis:** Lake Powell Inflow [KAF]
- **X-axis:** Time (Mar 08 to Nov 10)
- **Legend:**
  - CBRFC Forecast
  - Seasonal Forecast
  - Historical Resample
  - HMM Year 1
  - HMM Year 2

The graph illustrates the inflow data for Lake Powell, with different forecasting models represented by distinct colored bars and box plots. The data is segmented by months and years, with specific annotations for Nov 07 to Dec 10.
**Sample Results - Navajo Pool Elevation**

Blue - Observed  
Green - 24MS  
Purple - Control Run  
Boxplots - Midterm Model
**Sample Results - Flaming Gorge Outflow**

Blue - Observed  
Green - 24MS  
Purple - Control Run  
Boxplots - Midterm Model
**Sample Results - Powell Pool Elevation**

![Graph showing pool elevation over time with different color lines representing observed, 24MS, and control run predictions.]

- **Blue** - Observed
- **Green** - 24MS
- **Purple** - Control Run
- **Boxplots** - Midterm Model
CONCLUSIONS

- Long lead seasonal forecast model using large scale climate info
- Hidden Markov model for simulation and second year forecasts
- Persistence is not the same as autocorrelation
- Forecasts combined as input to a new probabilistic operations model for the Colorado River Basin
CONCLUSIONS

- Long lead seasonal forecast model using large scale climate info
CONCLUSIONS

- Long lead seasonal forecast model using large scale climate info
- Hidden Markov model for simulation and second year forecasts
  - Persistence is not the same as autocorrelation
CONCLUSIONS

- Long lead seasonal forecast model using large scale climate info
- Hidden Markov model for simulation and second year forecasts
  - Persistence is not the same as autocorrelation
- Forecasts combined as input to a new probabilistic operations model for the Colorado River Basin