A method for forecasting second year peak season streamflow

Cameron Bracken

CVEN6833 Fall, 2009

1. Introduction

The 24 month study is the BOR's primary midterm operational forecast model for the Colorado River Basin. The 24 month study takes inputs of forecasted inflow and outflow on a monthly time step and provides managers with projections of water availability at major reservoirs[1]. In the 24 Month Study streamflow after the first year is set to climatology for lack of skilful actual forecasts. In fact most current forecasting methods in the literature which use ocean and atmospheric variables only extend out a single water year (*Bracken et al.*, 2009; *Grantz et al.*, 2005; *Regonda et al.*, 2006). In large reservoirs, skilful forecasts for inflow are useful for operations, especially in extreme cases. For example, the knowledge that two years in a row will be extremely low could significantly affect operations.

Prairie et al. (2008) develop a method for synthetic sreamflow generation which incorporates state information from paleoreconstructed data. A nonhomogeneous markov chain is used to simulate the state of the system m and magnitudes are then generated based on a KNN resampling of the historical record. In this study we extend these ideas to a forecast setting in order generate estimates of peak season runoff 2 years in the future based on current system conditions.

2. Methodology

Consider a qth order markov chain model in which the states of the system $S_t, S_{t-1}, ..., S_{t-q+1}$, determine the state S_{t+1} but also comes with information about the state at $t+2, ..., t+r$ (which we will consider a forecast). Let $x(t)$ be the historical timeseries of interest. Let $\mathbf{S}^-(t) = [S_{t-q+1}, S_{t-q+2}, ..., S_t]$ and $\mathbf{S}^+(t) = [S_{t+1}, ..., S_{t+r}]$, where q is the number of previous states used to determine determine r future states. In the same way let the quantiles of the current and future flows be represented by $\mathbf{Q}^-(t) = [Q_{t-q+1}, Q_{t-q+2}, ..., Q_t]$, and $\mathbf{Q}^-(t) = [Q_{t+1}, ..., Q_{t+r}]$. We are interested in constructing a model for forecasting the future magnitudes of flow given $\mathbf{S}^-(t)$ and $\mathbf{Q}^-(t)$. Conceptually we wish to construct

$$\mathbf{f}(\mathbf{x}^+(t)|\mathbf{S}^-(t), \mathbf{Q}^-(t))$$

the probability distribution of the next r years of seasonal total flow. Lack of data for higher order markov chains such as this is a very real concern if we are interested in annual or seasonal total data. For example, with 100 years of data and we use a 5th order ($q = 5$) model to estimate transition probabilities, there are $2^5 = 32$ different transitions to estimate, which is just over 3 data points per transition. This lack of data motivates the search for other sources of data

[1] `http://www.usbr.gov/uc/water/crsp/studies/index.html`

to estimate transition probabilities. Paleo reconstructed streamflow data provides an attractive solution. Paleoreconstucted streamflow tends not to agree on magnitude but agrees well on the "wet" or "dry" state of of the system *Prairie et al.* (2008); *Woodhouse et al.* (2006). Their much longer period of record, provides ample information for higher order markov models.

In this study we additionally rely on the quantiles of the paleo data to map paleo values to historical values. This method of bias correction is commonly used to downscale global climate models (*Wood et al.*, 2004). This also allows use to use annual flow in the paleo record and seasonal total flow in the historical record side by side. One future state, $\mathbf{S}^+(t)$, determines many possible sets, of quantiles, $\mathbf{Q}^+(t)$. In this approach we sample from the set of possible future quantiles based on a KNN approach (*Lall and Sharma*, 1996; *Rajagopalan and Lall*, 1999). The measure of distance we use is the euclidean norm of the difference between the historical and paleo current quantiles

$$||_p\mathbf{Q}^-(t) - \mathbf{Q}^-(t)||.$$

We select from the $K = \sqrt{n}$ nearest neighbors (*Lall and Sharma*, 1996) based on the weight function

$$W(i) = \frac{1/i}{\sum_{i=1}^{K} 1/i}$$

Combining these ideas, a forecasting algorithm is presented:

1. Use the paleoreconstructed data to generate the transition probabilities from all 2^q starting states to all 2^r following states.
2. Use the current state of the system, $\mathbf{S}^-(t)$, to simulate a future state of the system, $_p\mathbf{S}^+(t)$, using the transition probabilities.
3. Sample a nearest neighbor future quantile from the marginal probability distribution $f(_p\mathbf{Q}^+(t) \mid {}_p\mathbf{S}^+(t))$ based on the distance measure $||_p\mathbf{Q}^-(t) - \mathbf{Q}^-(t)||$. Use the resampled quantile to obtain a future magnitude $\mathbf{x}^+(t)$.
4. Repeat 2-4 to obtain ensembles.

3. Application

The proposed framework is applied to observed naturalized seasonal total (April-July) volumes at Lees Ferry for the period (1906-2006) (*Prairie and Callejo*, 2005). We use the recent paleoreconstructed annual flows at Lees Ferry for the period (1490-1997) from *Woodhouse et al.* (2006). We use 11 year running means for each data set *Gangopadhyay et al.* (2009), though is is not certain if this has an impact on the forecasts. Since we are using seasonal totals, it is not appropriate to use the median of the historical to compare

The parameters q and r are available to change in this framework. We fix r at 2 since we are only interested in the first and second year forecasts. Presumably increasing r would degrade the forecast but this has yet to be investigated. A few values of q =2, 3 and 4 were tested.

4. Results

Results are shown for two sets of forecasts using 11 year running means. Figure 1 shows first and second year forecasts withe $q = 2$ and $r = 2$. Qualitatively, the first year forecasts show better agreement than the second year forecasts which is to be expected. Figure 2 shows first and second year forecasts with $q = 3$ and $r = 2$. Figure 3 shows first and second year forecasts withe $q = 4$ and $r = 2$. These results are similar to those with $q = 3$.

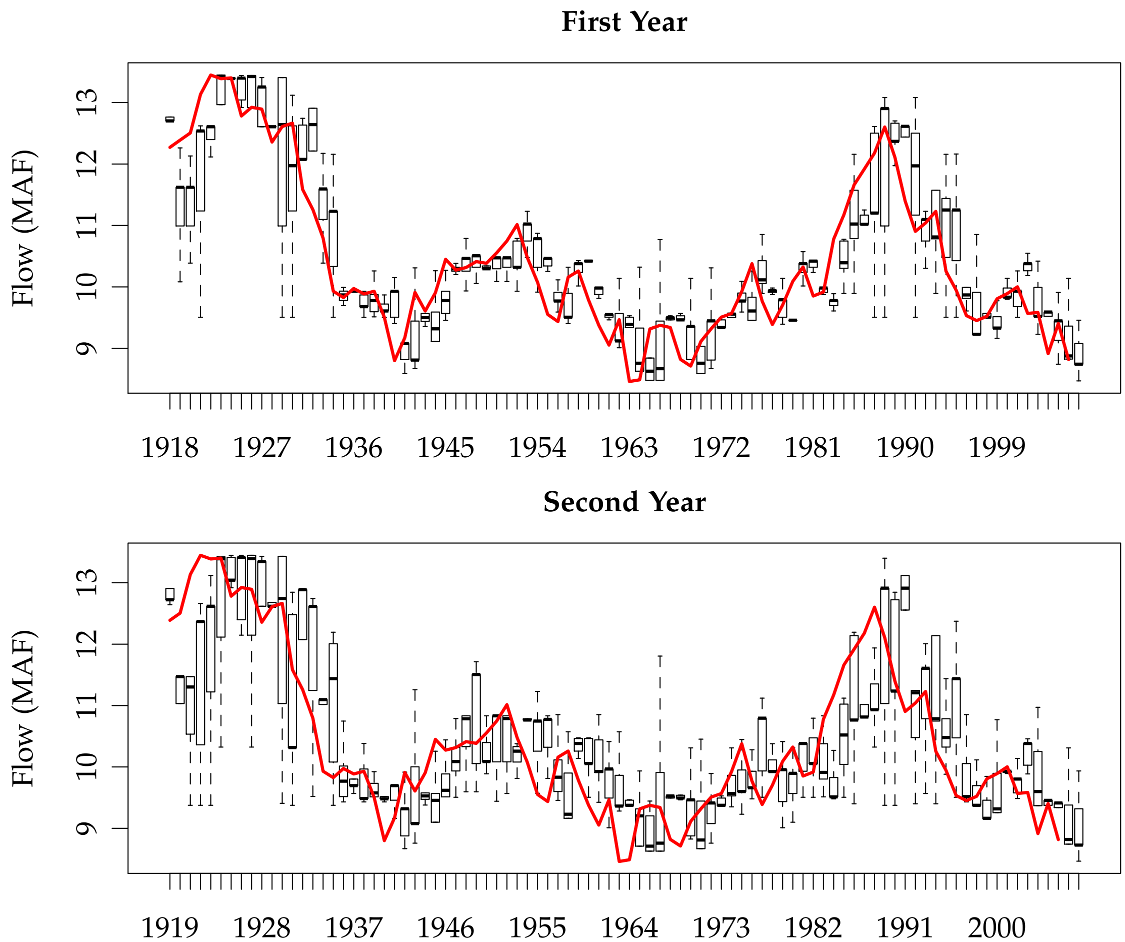

Figure 1: First and second year forecasts with $q = 2$ and $r = 2$

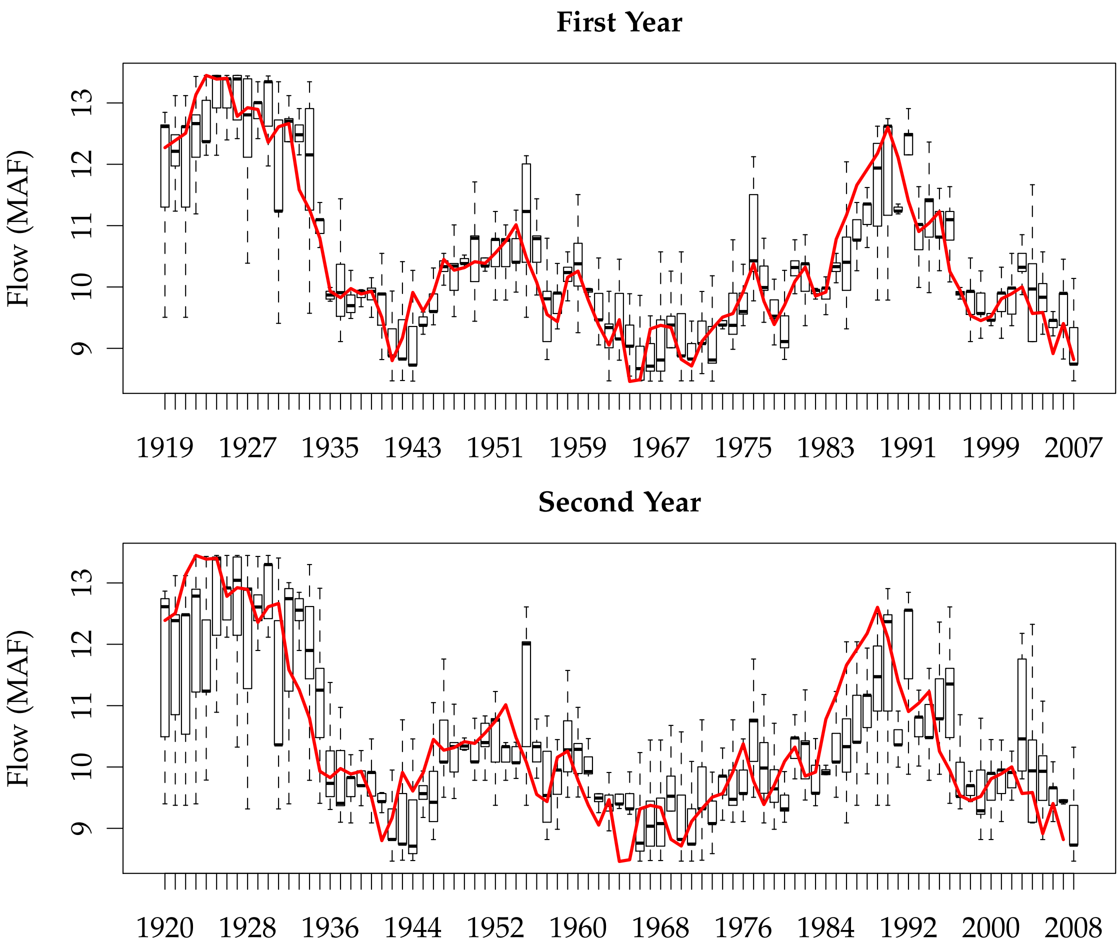

Figure 2: First and second year forecasts with $q = 3$ and $r = 2$

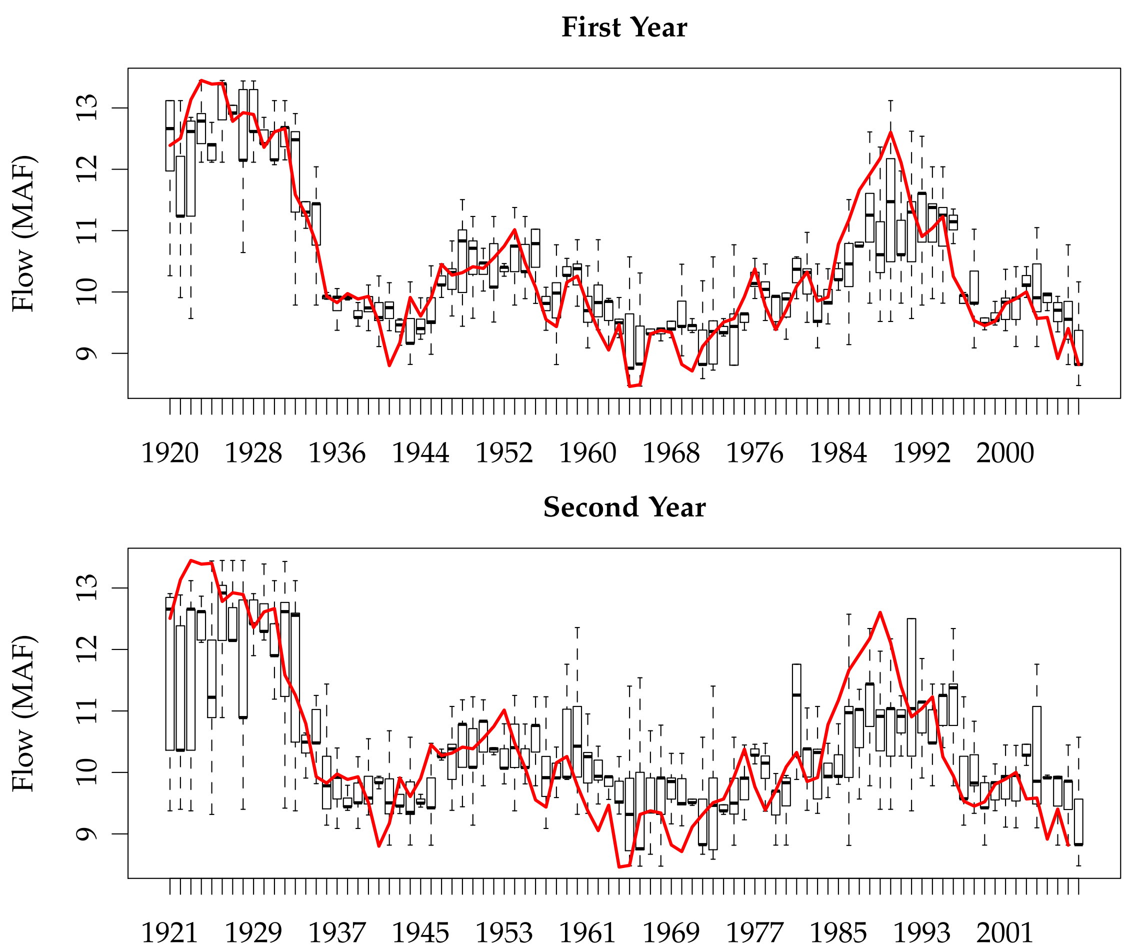

Figure 3: First and second year forecasts with $q = 4$

5. Conclusions

A framework for obtaining ensemble forecasts of second year peak season streamflow has been presented. This method uses paleoreconstructed streamflow data to obtain information about the state and quantile of future years. This forecast could be issued on September 1st for the subsequent 2 years. These forecasts could be disaggregated to obtain diverse spatial and temporal information. The forecasts show good agreement with historical data, though a much more thorough investigation of sensitivity and forecast errors is necessary.

Many things still need to be investigated:

1. The use of different running mean lengths
2. Many more combinations of q and r
3. The use of different values of K, the heuristic criteria of $K = \sqrt{n}$ was developed for use with magnitudes and the criteria for quantiles may need to be refined.
4. Quantify forecast skill with RPSS.
5. Quantify the benefit from second year forecasts.

Possible extensions include:

1. The incorporation of a multivariate state from other paleoreconstructed variables.
2. The use of nonstationary transition probabilities.

References

Bracken, C., B. Rajagopalan, and J. Prairie (2009), A multi-site seasonal ensemble streamflow forecasting technique, *Water Resour. Res*, pp. 1–27.

Gangopadhyay, S., B. Harding, B. Rajagopalan, J. Lukas, and T. Fulp (2009), A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles, *Water Resour. Res.*, *45*(6), 1–14, doi:10.1029/2008WR007201.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona (2005), A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, *Water Resour. Res.*

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resour. Res.*, *32*(3), 679–693.

Prairie, J., and R. Callejo (2005), Natural flow and salt computation methods, pp. 1–31.

Prairie, J., K. Nowak, B. Rajagopalan, U. Lall, and T. Fulp (2008), A stochastic nonparametric approach for streamflow generation combining observational and paleoreconstructed data, *Water Resour. Res.*, *44*(6), 1–11, doi:10.1029/2007WR006684.

Rajagopalan, B., and U. Lall (1999), A k–nearest-neighbor simulator for daily precipitation and other weather variables, *Water Resour. Res.*, pp. 1–14.

Regonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona (2006), A multimodel ensemble forecast framework: Application to spring seasonal flows in the gunnison river basin, *Water Resour. Res.*, *42*.

Wood, A., L. Leung, V. Sridhar, and D. Lettenmeir (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs, *Climatic Change*, *15(62)*, 189–216.

Woodhouse, C., S. Gray, and D. Meko (2006), Updated streamflow reconstructions for the upper colorado river basin, *Water Resour. Res.*

Appendix - Source code

```r
#!/usr/bin/env Rscript

#load packages, functions and data
source('setup.R')

q <- 2 #years back
r <- 2 #years ahead, 2 or greater

nsims <- 250
paleo <- wood
paleo.b <- wood.b
hist <- lees
```

```
hist.b <- lees.b

n <- length(paleo)
nh <- length(hist)

    #binary combinations
b <- expand.grid( rep( list(c(T,F)), r ) )

        #all state tansitions of paleo
paleo.bqr <- laglead(paleo.b,q,r)

        #block transition probabilities
tr.paleo <- trprob(paleo.bqr$lag,paleo.bqr$lead)

        #Quantiles of paleo and historical
Qp <- ecdf(paleo)(paleo)
Qp.qr <- laglead(Qp,q,r)

sim.y1 <- sim.y2 <- matrix(NA,nsims,nh-q+1)

        #y is the year to forecast FROM
for(y in q:nh){
        cat('Forecasting From Year',time(hist)[y],'\n')
        Qh <- ecdf(hist[-((y-q+1):y)])(hist)
        Qh.qr <- laglead(Qh,q,r)

                #current binary and quantile state
        state <- hist.b[(y-q+1):y]
        Qstate <- Qh[(y-q+1):y]
        Qsum <- sum(Qstate)
        this.p <- tr.paleo$p[rows.equal(tr.paleo$from,state)]

        cat('\tThe Current State is', state,'\n')
        cat('\tThe Quantile State is', Qstate,'\n')

        this.sim <- matrix(NA,nsims,r)

        for(i in 1:nsims){

                        #simulate transition
                rand <- runif(1)
                which.to <- rank(c(rand,cumsum(this.p)))[1]
                state.to <- tr.paleo$to[which.to,]

                        #conditional pool, quantiles given state.to
```

```
                pool.from <- Qp.qr$lag[rows.equal(paleo.bqr$lead, state.to),]
                pool.to <- Qp.qr$lead[rows.equal(paleo.bqr$lead, state.to),]

                        #current and pool's quantile sum
                pool.Qsum <- apply(pool.from,1,sum)

                        #number of neighbors
                k <- round(sqrt(nrow(pool.to)))

                        #weight function
                w <- numeric(k)
                for(j in 1:k)
                        w[j] <- (1/j)/(sum(1/(1:j)))
                w <- cumsum(w/sum(w))

                        #find nearest quantile neighbors
                d <- sqrt(apply((pool.from - Qstate)^2,1,sum))
                neighbors <- order(d)[2:(k+1)]

                rand <- runif(1)
                this.neighbor <- rank(c(rand,w))[1]
                this.quant <- pool.to[neighbors,][this.neighbor,]
                #this.quant <- pool.to[sample(1:nrow(pool.to),1),]

                this.sim[i,] <- this.quant

        }

        sim.y1[,y-q+1] <- quantile(hist,this.sim[,1])
        sim.y2[,y-q+1] <- quantile(hist,this.sim[,2])
}
colnames(sim.y1) <- (start(hist)[1]+q-1):(end(hist)[1])+1
colnames(sim.y2) <- (start(hist)[1]+q-1):(end(hist)[1])+2
hist.y1 <- window(hist,start(hist)[1]+q)
hist.y2 <- window(hist,start(hist)[1]+q+1)

 layout(rbind(1,2))
 boxplot(sim.y1,outline=F)
 lines(1:length(hist.y1),hist.y1,col='red',lwd=2)
 boxplot(sim.y2,outline=F)
 lines(1:length(hist.y2),hist.y2,col='red',lwd=2)

save(sim.y1,sim.y2, hist.y1, hist.y2 ,file='data/fc.Rdata')
```