

Wood River Water Collaborative Predictive Streamflow and Curtailment Date Model Details

Dr. Kendra Kaiser, Boise State University, Department of Geosciences

12/20/2020

Contents

1. Introduction	2
2. Methodology & Model Fits	2
Overview	2
Reproducibility	2
2.1 Data Downloading and Organization	2
USGS	3
Snotel	3
Agrimet	3
Snow Cover Extent	3
Diversion & Curtailment Data	3
2.2 Temperature Model	4
2.3.1 Streamflow Models	4
Big Wood at Stanton Crossing	6
Silver Creek	7
Camas Creek	7
Big Wood Diversions	8
Silver Creek Diversions	8
2.3.2 Streamflow Correlations	9
2.4 Streamflow Simulation	11
2.5 Curtailment Model and Simulation	13
3. Overview of modeling results	13
4. Recommendations	13

```
getwd()
```

```
## [1] "/Users/kendrakaizer/Desktop/Data/WRWC"
```

1. Introduction

The Wood River Collaborative is a grassroots effort to tackle water usage challenges among irrigators, municipalities, and protect minimum flows for fish and wildlife habitat. Its many, basin-wide participants include private citizens, representatives of water agencies, non-profit organizations, private interests and the public sector. The outcome of the collaboration is to bring all stakeholders together and develop strategies and tools for best use of water for consumptive use, while conserving water for groundwater and in-stream flows.

The following suite of modeling tools were developed in response to stakeholder interests in improving management of surface and groundwater resources for agriculture and conservation purposes. These tools include automated data retrieval and organization for use in predictive models of irrigation season streamflow volume and timing in the Big Wood River Basin at the Hailey and Stanton Crossing gages, Camas Creek, and Silver Creek at Sportsman's Access (Figure 1). Annual river diversions are also predicted to estimate curtailment dates for three water right priority dates.

2. Methodology & Model Fits

Overview

Individual multivariate linear regression models were developed for each of these locations using USGS streamflow data, Snotel SWE and temperature data, AgriMet temperature data, and diversion data from Water District 37 (Table 1). The Bayesian Information Criterion (define & cite) was used to select model parameters for each of the gage locations for both total irrigation season streamflow volume and timing, characterized by the center of mass. Center of mass definition here. . . .

Once the linear regression models were developed for total irrigation season volume and timing at each location, multivariate distributions were used to stochastically model hydrographs for each location. The residuals (standard error) from each of the regression models and correlations between gauge stations are used to create the multivariate distributions. This ensures that given a set of predictor variables (e.g. SWE, temp) the predicted volumes will be statistically consistent across gage locations (e.g. the models wont predict that Camas Creek will have really low runoff year while the Big Wood has a really high runoff year because they are statistically correlated). Repeated, random selection from these multivariate distributions produces a sample of predicted volumes and timing of streamflow. The samples of total volume and streamflow timing are then used to create simulations of the irrigation season hydrograph. Variability in final model outputs is quantified by percentiles of the resulting predictions.

Reproducibility

All model scripts have been developed using gitHub as the code repository. This enables tracking of all model changes, sharing of model code with WRWC members and a mechanism for users to post 'issues' to the code repository (<https://github.com/kendrakaizer/WRWC>). When the model is updated a versioning standard will be used to update

2.1 Data Downloading and Organization

Automation of data downloads and processing ensures that all data is formatted properly. Creating a local folder for each model run where all formatted data is saved will be valuable for reproducibility purposes.

Script	Purpose	Output
run_models.R	Run all scripts	Model_run.pdf
data_scraping.R	Data automation & harmonization	Formatted data
temperature_model.R	Mixed-effects temperature model	Predicted April - June average temperatures
streamflow_model.R	Multivariate streamflow models	Multivariate streamflow models
	Bootstrap volumes from multivariate distribution	Predicted volumes
	Bootstrap water year from center of mass multivariate distribution	Predicted analog water years
streamflow_simulation.R	Simulate hydrograph by normalizing analog WY with sampled volume	Simulated hydrograph & prediction intervals
curtailment_model.R	Curtailment model	Curtailment dates

USGS

Snotel

Agrimet

explain how to download specific RNRCS package `devtools::install_github(repo = "rhlee12/RNRCS", subdir = "/RNRCS/", force = TRUE)`

Snow Cover Extent

summary details of Google Earth Explorer outcomes and reference to the other summary report

Diversion & Curtailment Data

This data was compiled by WRWC by manually entering data from the irrigation district black books and should be updated annually for future model revisions. Currently the following diversions are included: BWB: SC:

Smaller diversions were not included in this model version as they were considered to be minor given time constraints.

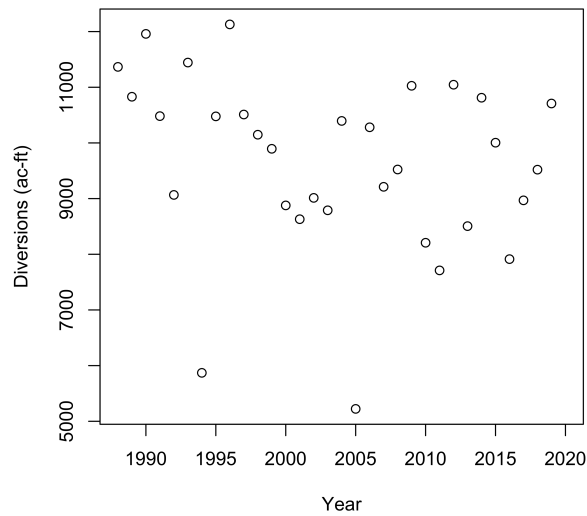


Figure 1: Diversions Above Hailey, will update to show total diversions

2.2 Temperature Model

A linear mixed effects model was developed to predict mean April - June temperatures in each sub basin. April - June temperatures are predictors in the center of mass regressions, so a bootstrapped sample of predicted temperatures are used.

```
#subset sites to those for each sub basin
tdata.bwb<- tdata[tdata$site %in% c("galena","galena summit", "lost-wood divide"),]
trend.reml<-lme(fixed=Apr.Jun.tempF ~ year, random=~1+year|site, correlation = corAR1(), data=tdata.bwb)
# predict this years temperature
pred<-predict(trend.reml,new.data,0:1)$predict.fixed[1]
fits<-fitted(trend.reml,0:1)[(1:nyrs),1]
# Bootstrap to estimate variance on new prediction, based on fixed-effects covariance matrix
mu<-trend.reml$coef$fixed
sig<-trend.reml$var
rand.coefs<-mvrnorm(nboots,mu,sig)
var.est<-var(rand.coefs*%c(1,last.yr+1))
var.site<-var(summary(trend.reml)$coeff$random$site[,1])/length(site.key)
se.pred<-sqrt(var.est+var.site)
aj.temps.bwh<-rnorm(nboot,mean=pred,sd=se.pred)
```

2.3.1 Streamflow Models

Initial model development was explored using the `streamflow_model_exploration.R` script. The full suite of predictor variables were subset for each gage and the final set of predictor variables were determined using the `regsubsets` package, which enables visualization of adjusted R2 and BIC of each parameter set (e.g. Figure X).

Final streamflow models are defined and synthesized in `streamflow_models.R`, this script also creates predictions for each model for the user defined year. Data is imported, and data structures are set up to save model output. The `modOut` function returns relevant metrics and statistics from the modeled results for the year being predicted.

```

modOut<- function(mod, pred.dat, wq, vol, meanSWE, lastQ){
  ,
  mod:      input model
  pred.dat: data.frame of prediction variables
  wq:       array of historic winter flows (e.g. hist$cc.wq)
  vol:      array of historic april-sept volumes (hist$cc.vol)
  meanSWE:  mean(arrays of historic SWE from ws snotel sites) #mean(hist$ccd+hist$sr, na.rm=T)
  lastQ:    last years summer streamflow volume (ac-ft) #var$cc.vol[var$year == pred.yr-1]
  ,

  pred.params.vol<-array(NA,c(1,2))
  output.vol<-array(NA,c(1,8))

  sig<-summary(mod)$sigma
  pred.params.vol[1,2]<-sig
  #predict this years total volume at 95 % confidence
  predictions<-predict(mod,newdata=pred.dat,se.fit=T,interval="prediction",level=0.95)
  pred.params.vol[1,1]<-mean(predictions$fit, na.rm=T)
  #This years percent of mean winter flow
  output.vol[1,1]<-round(pred.dat[1,1]/mean(wq, na.rm=T),3)
  #percent of mean SWE
  output.vol[1,2]<-round(sum(pred.dat[1,2:3])/meanSWE,3)
  # back-transformation of log-transformed data to expected value in original units, with lognormal residuals
  output.vol[1,3]<-round(exp(predictions$fit[1]+sig^2/2)/(1.98*183),0)
  #Division by long-term mean to generate % of average volume, with lognormal residuals
  output.vol[1,4]<-round(exp(predictions$fit[1]+sig^2/2)/mean(vol, na.rm=T),3)

  #this years total volume at 80 % confidence
  predictions<-predict(mod,newdata=pred.dat,se.fit=T,interval="prediction",level=0.8)
  #bottom of 80% CI (statisticians) converted to cfs
  output.vol[1,5]<-round(exp(predictions$fit[2])/(1.98*183),0)
  # 90% exceedance flow as a percent of long-term mean
  output.vol[1,6]<-round(exp(predictions$fit[2])/mean(vol, na.rm=T),3)
  output.vol[1,7]<-round(lastQ/(1.98*183),0) # last years volume in cfs
  output.vol[1,8]<-round(lastQ/mean(vol, na.rm=T),3) # Last years percent of average historic volume
  return(list(output.vol, pred.params.vol))
}

```

After this function is defined, the same set of steps occurs for each linear model. 1) The model parameters are subset from the full dataset, 2) the linear model is created & summary metrics are saved, 3) Prediction data is subset, 4) Predictions are made and outputs (estimated volume and standard error) are saved.

```

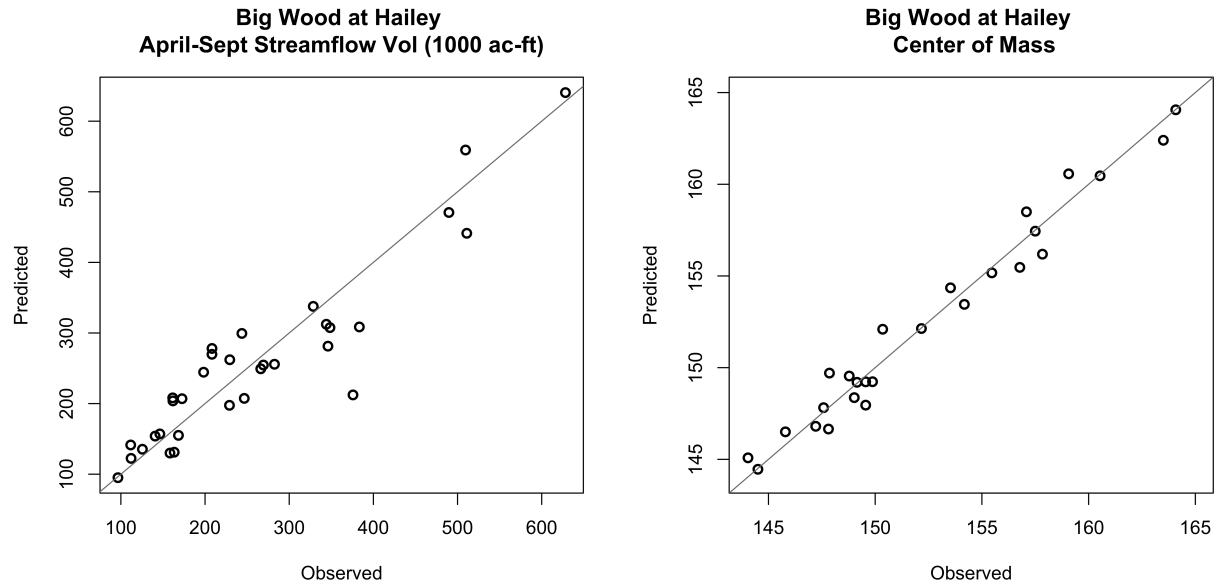
# 1. Subset Big Wood Winter flows, snotel from Galena & Galena Summit, Hyndman
hist <- var[var$year < pred.yr,] %>% select(bwb.vol.nat, g.swe, gs.swe, hc.swe)
# 2. Create Big Wood at Hailey linear model
bwb_mod<-lm(log(bwb.vol.nat)~ g.swe+ log(gs.swe)+ hc.swe, data=hist)
mod_sum[1,1]<-summary(bwb_mod)$adj.r.squared
# 3. Subset April 1 Prediction Data
pred.dat<-var[var$year == pred.yr,] %>% select(g.swe, gs.swe, hc.swe)
# 4. Big Wood at Hailey Model output
mod_out<- modOut(bwb_mod, pred.dat, hist$bwb.wq, hist$bwb.vol.nat, mean(hist$g.swe,
hist$gs.swe, hist$hc.swe, trim=0, na.rm=T), var$bwb.vol.nat[var$year == pred.yr-1])
output.vol[1,] <- mod_out[[1]]
pred.params.vol[1,] <- mod_out[[2]] # standard error, "sigma"

```

After the streamflow volume model section of code, the same procedure is done for creating multivariate

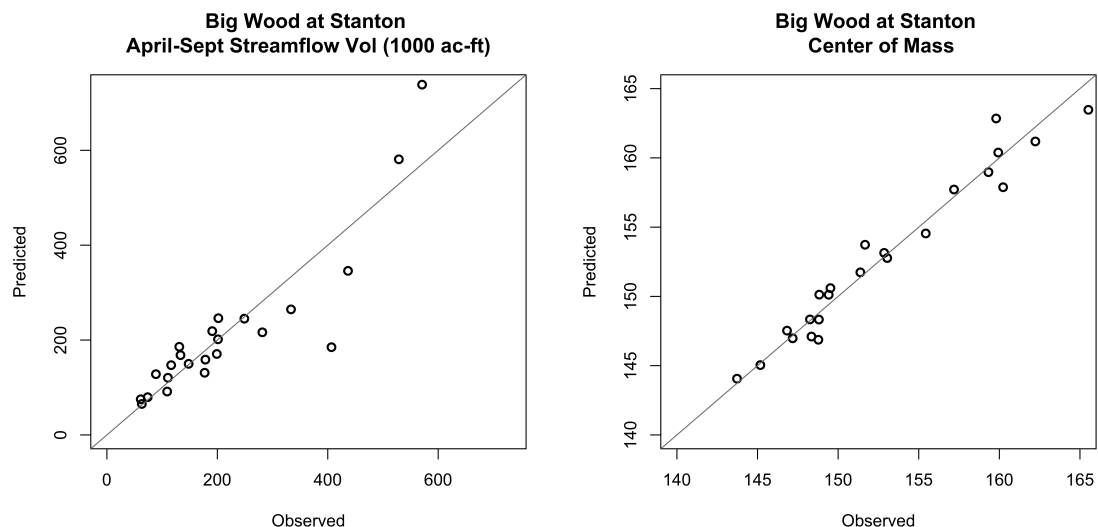
linear regressions for predicting center of mass. Here, model fits for irrigation season volume and center of mass are shown together for each gage.

```
# Big Wood at Hailey Natural Flow Volume model
bwb_mod<-lm(log(bwb.vol.nat)~ g.swe+ log(gs.swe)+ hc.swe, data=hist)
# BW Hailey Natural Flow Center of Mass model
bwb_mod.cm <-lm(bwb.cm.nat ~ log(bwb.wq) + g.swe+ hc.swe+ t.g +t.gs+t.lw+
log(cg.swe)+log(gs.swe), data=hist)
```



Big Wood at Stanton Crossing

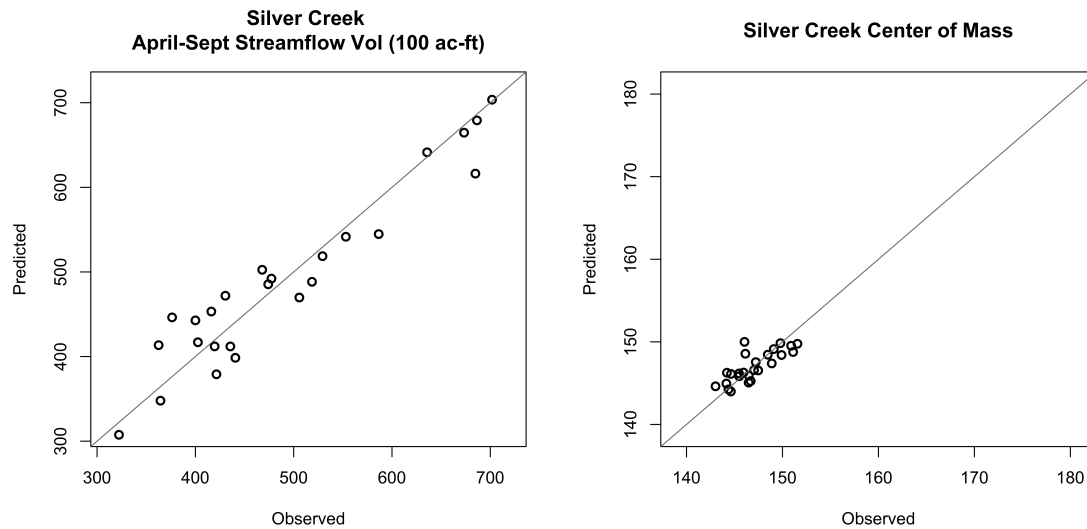
```
# Big Wood at Stanton Natural Flow Volume model
bws_mod<-lm(log(bws.vol.nat)~bws.wq+ log(g.swe) + log(gs.swe)+ log(hc.swe), data=hist)
# Big Wood at Stanton Natural Flow Center of Mass model
bws_mod.cm <-lm(bws.cm.nat ~ lwd.swe +log(cg.swe)+log(hc.swe) + t.cg + t.g + t.hc + t.lw, data=hist)
```



Silver Creek

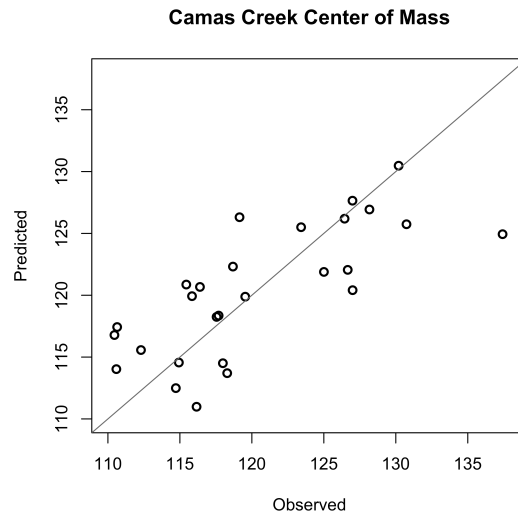
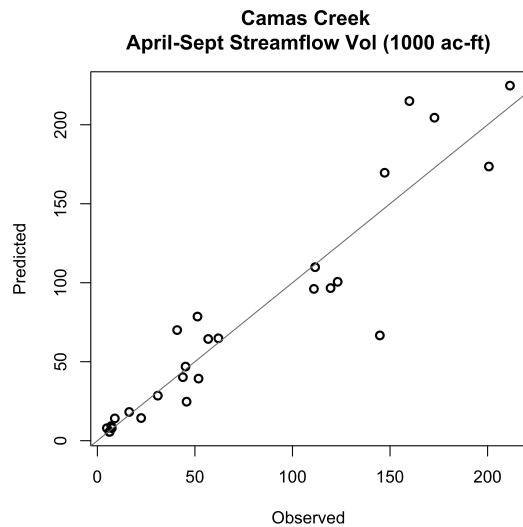
The Silver Creek Model is unique in that it uses a mixture of SWE data from both the Big Wood and Little Wood Basins. While it is not traditional to use SWE from outside of a delineated HUC basin, none of the SNOTEL sites are adequate representations of the snow that is contributing to the Silver Creek Watershed. Given this data limitation and the groundwater interactions between Big Wood and Silver Creek, the model includes SWE data from Galena, Chocolate Gulch, and Swede Peak. This is one utility of using a statistical model, namely if SWE from these locations are correlated to flows in Silver Creek historically, they can do a sufficient job of predicting flows in the basin. Further discussion on viable next steps for the Silver Creek Model are discussed in the Recommendations section.

```
# Silver Creek Flow model, note mixture of SWE from Big Wood and Little Wood basins
sc_mod<-lm(log(sc.vol.nat)~ sc.wq+ga.swe + g.swe + log(hc.swe) + log(bwb.wq), data=hist)
# Silver Creek Natural Flow Center of Mass model
```



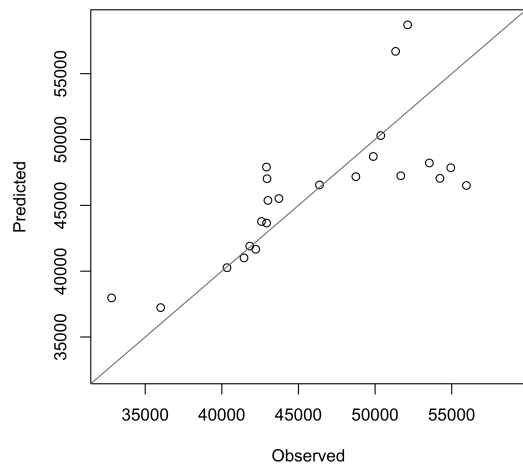
Camas Creek

```
# Camas Creek Flow Volume model
cc_mod<-lm(log(cc.vol)~log(cc.wq)+sr.swe+ccd.swe, data=hist)
# Camas Creek Center of Mass model
cc_mod.cm<-lm(cc.cm~ccd.swe + sr.swe+ t.f, data=hist)
```



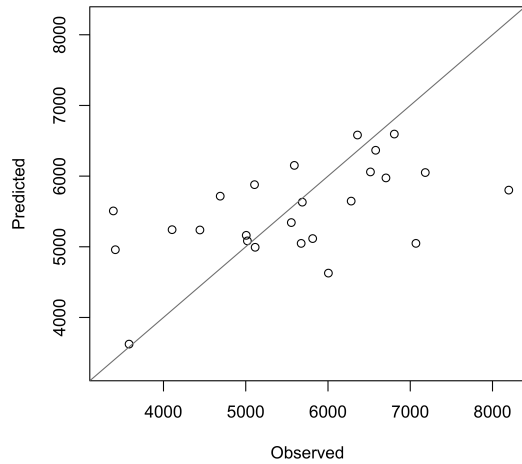
Big Wood Diversions

```
# Total Big Wood Diversion Volume linear model
div_mod<-lm(log(var$div[var$year >=1997 & var$year < pred.yr]) ~ log(cg.swe)+log(hc.swe)
+log(bws.wq), data=hist)
# Data is subset to after 1997 when the Big Wood at Stanton gage came on board,
# revisit this model using bub.wq so we can use the full dataset
```



Silver Creek Diversions

```
sc.div_mod<-lm(log(var$sc.div[var$year>1993 & var$year < pred.yr]) ~ g.swe+ temps+log(cg.swe)+log(lwd.s
```

2.3.2 Streamflow Correlations

Given the proximity of the three basins, correlations between the basins' total annual irrigation season streamflow, diversions, and center of mass allow us to ensure that the predicted flows at each gage are representative of how regional climatic patterns will be effecting all locations. For example, we would not expect Camas Creek to have an exceptionally dry year in a year when the Big Wood is experiencing an exceptionally high streamflow year. The correlation between sites is combined with the standard error from each linear model to create a covariance matrix which is use to bootstrap model predictions.

```
# Correlation matrix between streamflow volumes, diversions and centers of mass
cor.mat<-cor(cbind(flow.data[c(1,3,5,7,9,10)],flow.data[c(2,4,6,8)]),use="pairwise.complete")
# Create covariance matrix by multiplying by each models standard error
# pred.pars[1,]: fitted values; pred.pars[2,]: sigma (standard error)
pred.pars<-rbind(pred.params.vol, pred.params.div, pred.params.cm)
outer.prod<-as.matrix(pred.pars[,2])%*%t(as.matrix(pred.pars[,2]))
cov.mat<-cor.mat*outer.prod
```

	bwb.vol.nat	bws.vol.nat	cc.vol	sc.vol	div	bwb.cm.nat	bws.cm.nat	cc.cm	sc.cm
bwb.vol.nat	1	0.98	0.92	0.96	0.7	0.23	0.13	0.08	-0.59
bws.vol.nat	0.98	1	0.92	0.96	0.72	0.24	0.14	0.11	-0.59
cc.vol	0.92	0.92	1	0.92	0.7	0.32	0.23	0.05	-0.58
sc.vol	0.96	0.96	0.92	1	0.68	0.11	0.01	-0.03	-0.66
div	0.7	0.72	0.7	0.68	1	0.16	0.11	0.19	-0.49
bwb.cm.nat	0.23	0.24	0.32	0.11	0.16	1	0.99	0.63	0.26
bws.cm.nat	0.13	0.14	0.23	0.01	0.11	0.99	1	0.64	0.33
cc.cm	0.08	0.11	0.05	-0.03	0.19	0.63	0.64	1	0.5
sc.cm	-0.59	-0.59	-0.58	-0.66	-0.49	0.26	0.33	0.5	1

Figure 2: Correlation matrix between gages

Flow volumes are then sampled from the multivariate distribution.

```
vol.pars<-rbind(pred.params.vol, pred.params.div) # only use predictions from volume models
vol.sample<-mvrnorm(n=5000,mu=(vol.pars[,1]),Sigma=cov.mat[1:5,1:5]) # historical covariance of volumes
```

This results in a distribution of potential volumes for each gage, given the input predictor variables. – put in a figure here showing the distribution of results as an example of the output

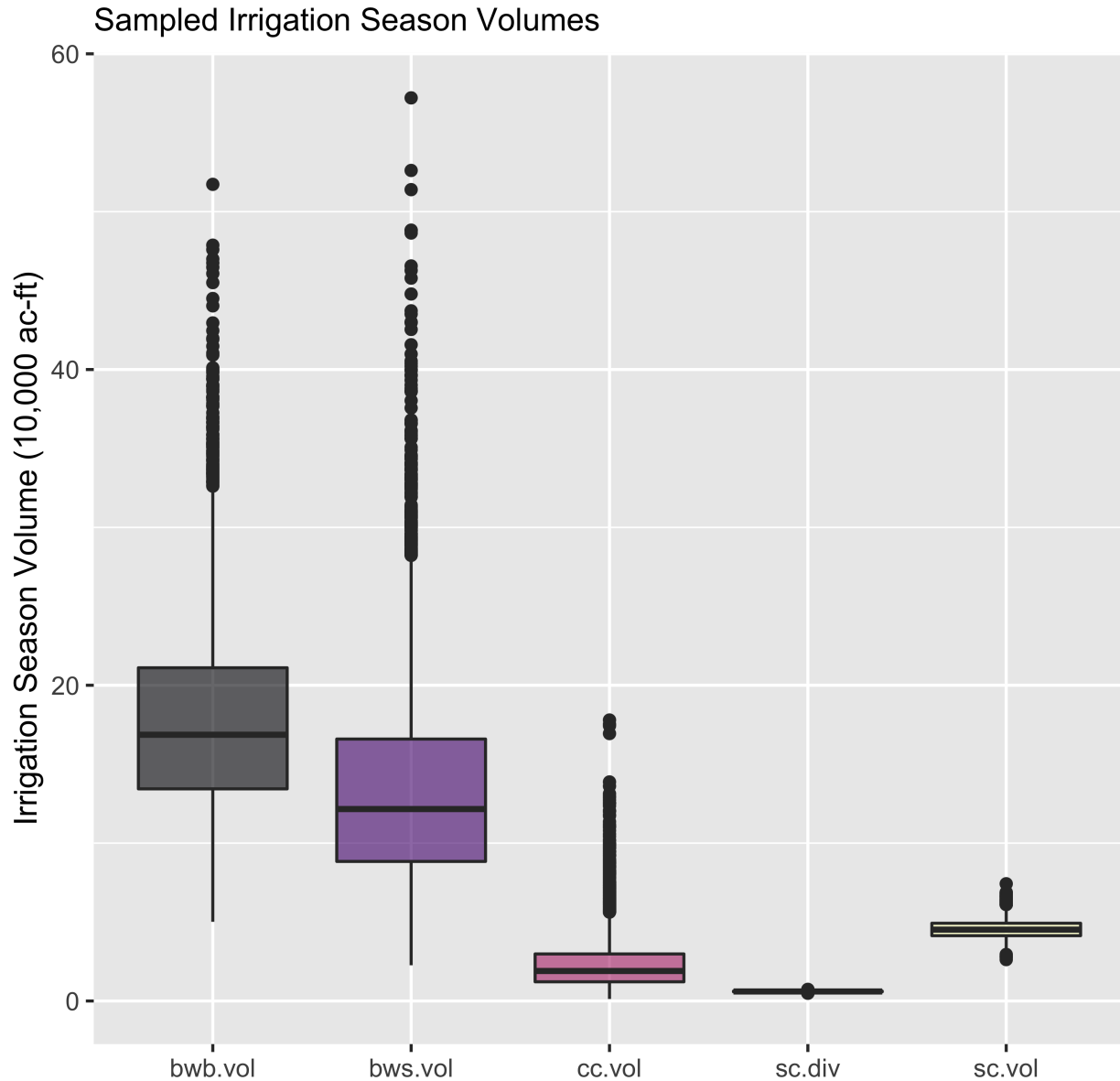


Figure 3: Distrubution of sampled volumes at each gage

A similar process is used for estimating the timing of runoff.

```
cm.data = var[var$year >= 1997 & var$year < pred.yr,] # only use complete dataset
cm.data = cm.data %>% select(year, bwb.cm.nat, bws.cm.nat,cc.cm, sc.cm)
```

```

cm.data$prob<-NA

# pmvnorm calculates the distribution function of the multivariate normal distribution
for(i in 1:dim(cm.data)[1]){
  vec<-cm.data[i,2:5]
  cm.data$prob[i]<-pmvnorm(lower=as.numeric(vec)-0.5,
                          upper=as.numeric(vec)+0.5,mean=pred.params.cm[,1],sigma=cov.mat[6:9,6:9])[1]
}
cm.data$prob<-cm.data$prob/sum(cm.data$prob) # turn into percentage
# create array of years based on their similarity to prediction year
CMyear.sample<-sample(cm.data$year,5000,replace=TRUE, prob=cm.data$prob)

```

	% of sample
2004	0.0066
2006	0.4524
2013	0.4386
2014	0.1006
2017	0.0018

Figure 4: Summary of center of mass sample

The resulting matrices are then saved as .csv to be used in the final simulation model.

2.4 Streamflow Simulation

The final irrigation season streamflow simulations are modeled in the `streamflow_simulation.R` script.

The original streamflow data, sampled volumes and centers of mass are imported and the irrigation season hydrographs are simulated. This is done by selecting the timeseries of natural flow that corresponds with a given year from the center of mass sample and normalizing it by a volume from the multivariate distribution sample. This ‘analog water year’ approach effectively uses the linear models to estimate the most similar year in runoff timing, and normalizes (another way to say this?) that hydrograph based on the predicted volume estimates.

```

for(k in 1:ns){ # ns = number of simulations, in our case 5000
  # Simulate natural flow supply at the four gages and total diversions
  year<-cm.year[k,1] # year sample
  vol<-volumes[k,] # volume sample

  # select the streamflow timeseries that corresponds with the center of mass sample

```

```

bwb<- bwb.wy[bwb.wy$wy == year, "bwb.nat.q"][183:365] # irrigation season
# normalize the sampled hydrograph by the sampled volume
bwb.flow.s[,k]<- bwb * vol/(sum(bwb)*1.98)
# 1.98 is the conversion from cfs to ac-ft, (cfs) * (ac-ft/ac-ft)

```

Prediction intervals are calculated from the relevant quantiles from the simulation results

```

pred.int<-function(location){
  lo<-apply(location,1,quantile,0.05, na.rm=TRUE)
  hi<-apply(location,1,quantile,0.95, na.rm=TRUE)
  meanQ<-apply(location,1,mean, na.rm=TRUE)

  return(cbind(lo, hi, meanQ))
}

```

The following figure is an example model output figure for each basin, the average simulated hydrograph (blue), the prediction interval (shaded grey), and the actual hydrograph (green) for 2019.

```

getwd()

## [1] "/Users/kendrakaizer/Desktop/Data/WRWC"
file.path('~/Desktop/Data/WRWC', "figures/BigWood_Hailey_Simulation.png")

## [1] "~/Desktop/Data/WRWC/figures/BigWood_Hailey_Simulation.png"

```

Big Wood Streamflow at Stanton Crossing

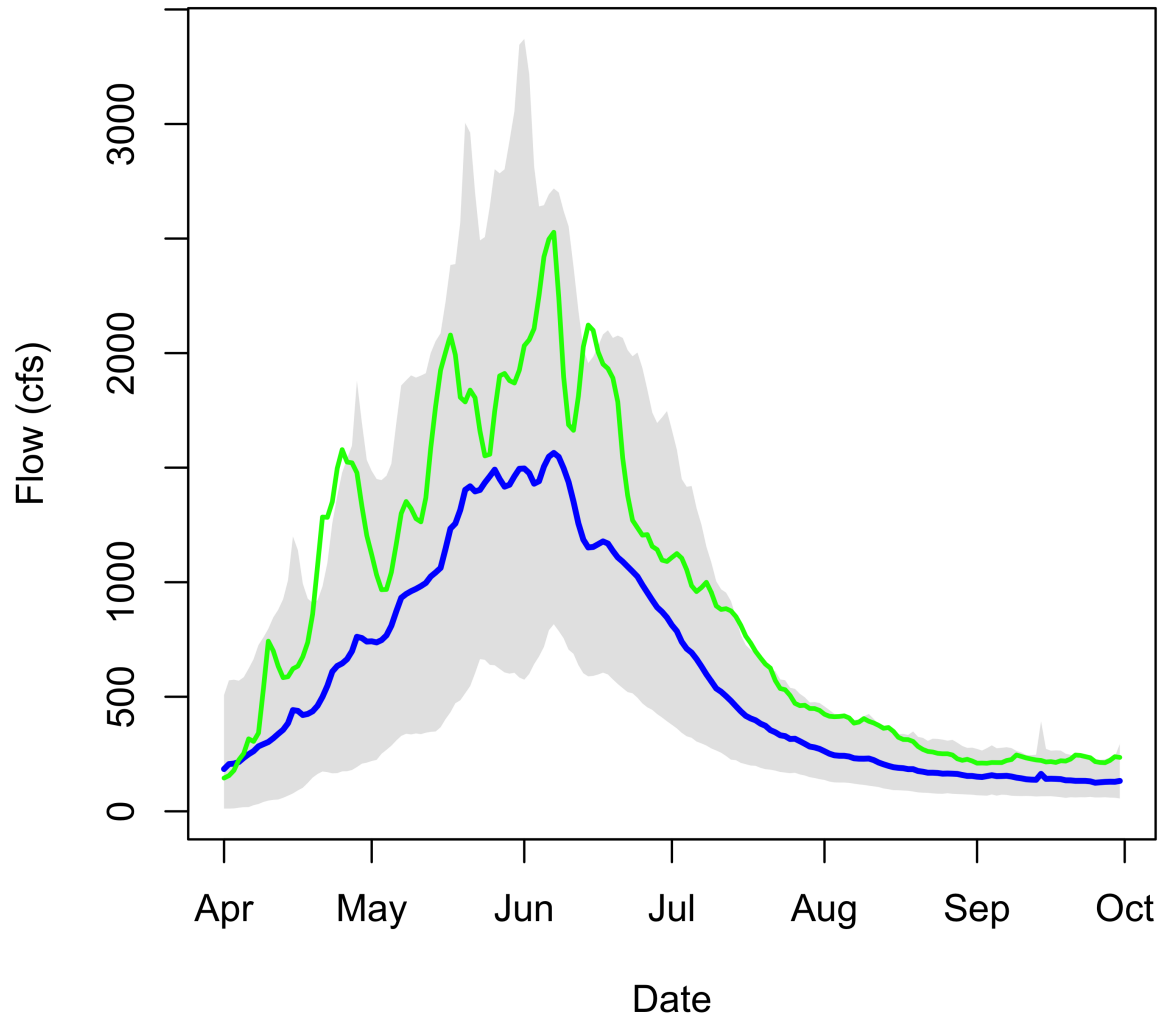


Figure 5: Simulated flows on the Big Wood River at Hailey

2.5 Curtailment Model and Simulation

3. Overview of modeling results

4. Recommendations