Teton River Water Supply Forecast Documentation

Montana Department of Natural Resources and Conservation
Water Management Bureau
Helena, MT
March 2019

Introduction

In 2019, Montana DNRC's Water Management Bureau developed a forecast of streamflow generated from snowmelt to aid water commissioners, water users and other parties interested in the Teton River Distribution Project. This document describes the data and methods used to develop the streamflow forecast model. The goal of this effort is to provide forecasts of runoff, generated from snowmelt in the mountainous headwaters of the North and South Forks of the Teton River, to aid in local decision-making and water distribution.

Available Data

Streamflow Forecast

The Natural Resources Conservation Service (NRCS) forecasts a streamflow (April 1 to July 31) volume at the United State Geological Survey (USGS) gage (06108000) Teton River near Dutton, which is below most diversions on the Teton River. The DNRC recognized that a forecast of the amount of water potentially available in the Teton River above the Dutton gage may also be useful to water users.

Snowpack

Snow-accumulating mountainous areas of the Teton River comprise approximately seven percent of the watershed. This small area, ranging from 5,500 to 9,000 feet in elevation, generates most of the water available in the Teton River.

Mountain snowpack is measured by the NRCS to forecast water supply. A common measure of snowpack is Snow Water Equivalent (SWE), which is the amount of water contained (in inches) in the snowpack. The NRCS maintains two types of sites in the mountains in and around the Teton River:

- SNOTEL (Snowpack Telemetry) stations remotely measures both Snow Water Equivalent (SWE) and precipitation in near real-time.
- Snow Course sites involve a manual measurement of SWE at a specific location once a month.

NRCS currently maintains two SNOTEL stations and one snow course site in the North Fork of the Teton drainage. Other nearby sites include one SNOTEL site in Birch Creek and two snow course sites in the North Fork of the Sun River drainage (Table 1).

Station Name	Site Type	ID	Watershed	Lat	Long	Elevation (ft)
Wrong Creek	Snow Course	12B04	N.F Sun	47.87	-112.93	5,700
Wrong Ridge	Snow Course	12B04	N.F Sun	47.91	-112.92	6,800
Mount Lockhart	SNOTEL	649	N.F Teton	47.92	-112.82	6,400
Waldron	SNOTEL	847	N.F Teton	47.92	-112.79	5,600
Freight Creek	Snow Course	12A01	N.F Teton	48.01	-112.82	6,000
Badger Pass	SNOTEL	607	N.F Birch Cr	48.13	-113.02	6,900

Table 1: List of sites providing mountain SWE and precipitation data in and near the Teton Watershed.

The NRCS SNOTEL sites provide data back to 1979, while Snow Course sites date back to 1948 in certain locations. The normal peak snowpack in the Teton occurs in mid-April and the snowpack is normally melted at or below the elevations of the snow monitoring sites by June 1.

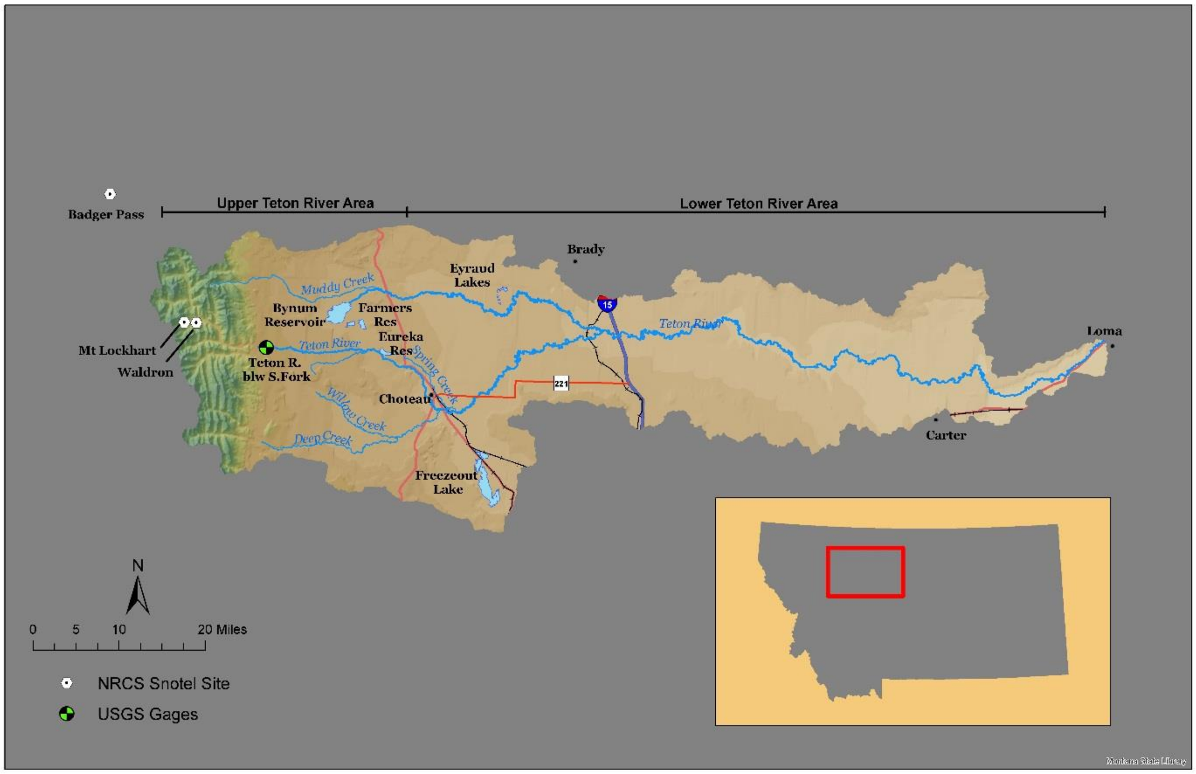

Figure 1: Map of SNOTELl sites and the USGS gage used in the forecast

Precipitation

Precipitation is recorded at the Mt. Lockhart, Waldron and Badger Pass SNOTEL sites (Figure 1). Late spring and early summer precipitation can enhance water supply and flooding. Large rain events in May of 2008 and June of 2018 markedly boosted runoff and the annual water supply. Precipitation data from the National Weather Service (NWS) Choteau Airport weather station (NOAA 2019) was used to help determine the start and end dates of the 2008 and 2018 rain events.

Flow

The USGS gage (06102500) Teton River below South Fork (TRSF) was used as the site for developing the streamflow forecast. The TRSF gage is located above most diversion on the Teton River. The period of record for the gage is 1947-1954 (year-round), 1998-2015 (April 1 to November 1), 2015-Present (year-round).

Climate Teleconnection

The accumulation of snowpack along the Rocky Mountain front is influenced by global patterns of atmospheric pressure and circulation, which influence the strength and position of the jet stream (CPC 2019). The Climate Prediction Center (CPC) of the Nation Weather Service uses the term "teleconnection pattern" to describe these global atmospheric systems. The CPC produces ten climate teleconnection indices monthly, which can improve streamflow forecasts.

Streamflow Volume Forecast Development

Methods

DNRC staff researched methods used by the NRCS, NWS, Army Corps of Engineers, and the Bureau of Reclamation to develop streamflow volume forecasts. The principal components analysis (PCA) and regression method described by Garen (1992) and NRCS (2011) was selected for use in the Teton Basin. Principle components regression is a statistical method which allows the use of highly correlated data as explanatory variables in a forecast model. For example, snowpack (SWE) and precipitation data collected from several stations in a small geographic area will be correlated with each other (they all experience similar storm systems). This "self" or autocorrelation must be accounted for in statistical forecasting or the results will appear to have much better predictive capability (i.e. lower uncertainty in the forecast) than in reality.

Principle Component Analysis restructures intercorrelated sets of data into equal sets of uncorrelated data (Garen 1992). The uncorrelated principle components can then be used to develop a linear regression between the response variable (Flow Volume, April 1 to July 31 for March and April forecasts, May 1 to July 31 for May forecast) and the explanatory variables (SWE, precipitation, antecedent flow conditions, climate teleconnection index). The open source statistical software program R Studio (Version 1.1.463) and Microsoft Excel were used for the computation and analysis of the streamflow forecast.

Forecast Issuance

Streamflow forecast certainty increases with time; a forecast issued close to the peak accumulation of snowpack in Mid-April is more likely to accurately forecast the observed streamflow than a forecast issued in March. The DNRC will issue three mid-month forecasts (Table 2). This schedule should provide time for water users and managers to incorporate streamflow forecasts into their planning for the upcoming season.

Forecast Issuance	Latest Data Used	Data Types	Forecasted Flow Volume
March 15	March 1	SWE, PPT, PNA	April 1 to July 31
April 15	April 1	SWE, PPT, PNA	April 1 to July 31
May 15	May 1	SWE, PPT, PNA	May 1 to July 31

Table 2: Forecast issuance dates and data used. PPT= Precipitation, PNA = Pacific North American Climate Teleconnection Index.

Streamflow Data Adjustments

Storms in May and June have, at times, dropped several inches of rain along the Rocky Mountain Front and in the adjacent mountains, adding significant volumes of water to the Teton River. Over the recent period of record at the TRSF gage (1998-2019), significant rain events occurred in 2008 and 2018. The unpredictability and infrequency of rain events led DNRC hydrologist to solely focus on the flow volume derived from snowmelt and to not address these large precipitation events.

Flow data at the TRSF gage were adjusted in 2008 and 2018 to remove high flows generated by rain. The period of adjustment was identified by comparing precipitation records at the NWS Choteau and NRCS Mt. Lockhart sites to flows at the TRSF gage.

Flow at the TRSF gage determined the start and end of the rain-affected flow. The data were adjusted until the pre-rain flow was observed again at the TRSF gage (14 days in 2008 and 11 days in 2018). Careful attention was paid to determine if the rise in streamflow was strictly due to rain or a combination of rain and snowmelt. The data were adjusted by reverting to the period of record daily mean flow for that time period.

Selected Time-Period

Similar NRCS flow- forecasting procedures typically use a near-recent, 30-year period of record (1981-2010) of streamflow and climate data for flow forecasting. Streamflow data is limited to the current period of record 1998-present at the USGS TRSF gage. DNRC determined that the 20-year period (1998-2018) contained sufficient data to develop a robust streamflow forecast model in the Teton.

Explanatory Variable Selection

Explanatory variables available for the principle component regression equation include:

- First of the month precipitation values at three SNOTEL Sites (inches)
- First of the month SWE values at three SNOTEL sites (inches)
- Monthly SWE values at three Snow Course sites (inches)
- Antecedent flow conditions (October flow in acre-feet)
- Climate Teleconnection Indices (non-dimensional)

Prior to beginning the Principal Component Analysis and Regression, some of the variables were eliminated from consideration. Specifically, antecedent flow conditions and nine of the ten teleconnection indices were discarded due to their low correlation with the response variable. An explanatory variable was discarded if the correlation coeffect (Equation 1) was less than 0.30 (NRCS 2011).

$$Correlation\ Coefficient = \frac{Covarience\ (x, y)}{\sqrt{Variance\ (x) Variance\ (y)}} \quad \textbf{(1)}$$

X= explanatory variable, *y* = response variable

Principle Components Analysis

Principle components analysis (PCA) of the explanatory variables (first of the month values for SWE and precipitation and the PNA Climate Teleconnection Index) was completed using R. As stated previously PCA involves transforming correlated data into an orthogonal, uncorrelated dataset of Principal Components (PC)s. The general PCA process is:

1. Center and Scale the data $= \frac{Mean\ (X_1)}{Standard\ Deviation(X_1)}$ X_1= individual explanatory variable
2. Calculated a Covariance Matrix of [n x n] dimensions
   Covariance(a,b,c,...) = $\sum_{i=1}^{n} \frac{(ai-\overline{a})\ (bi-\overline{b})(....)}{(n-1)}$
   n = number of samples, a,b,c = explanatory variables , $\overline{a}$ = mean
3. Calculate the Eigenvalue of the Covariance Matrix $\lambda = \det(\mathrm{A} - \lambda\mathrm{I}) = 0$
   λ = Eigenvalue, A = covariance matrix, I = Identity matrix, det = determinant
4. Calculate the Eigenvector of the Covariance Matrix $\omega\ = (\mathrm{A} - \lambda\mathrm{I})\omega = 0$
   ω = Eigenvector, A = covariance matrix, I = identity matrix,
5. Take the Eigen decomposition of the Covariance Matrix $= (X^T X) = S\Lambda S^{-1}$
   X = data matrix, X^T = transposed data matrix, S = diagonal matrix of eigenvectors, Λ = diagonal matrix of eigenvalues
6. Determine the proportion of variance that each Principle Component explains.
7. Decide how many Principle Components to retain.
8. Multiply the Eigen decomposition Matrix [n x n] by the original centered and scaled data frame [n x m] to get a data frame of the original dimension [n x m] with the principle components values for each explanatory variable. m = number of measurements (i.e. years of data)

PCA analysis found that the first Principle Component (PC) explained 72-74% of variance of the explanatory variable, with PC1, PC2 and PC3 explaining more than 95% of the variance.

Principle Components Regression

Principal Components Regression consists of performing a linear regression of the original explanatory variable against the transformed response variables (PCs). The general process is:

1. Introduce the response variable into the [n x m] matrix of Principle Components
2. Linearly regress April 1 – July 31 flow against the first PC.
3. If the estimated coefficient is statistically significant, add the next PC to the regression.
4. Continue until the last PC added does not result in a statistically significant coefficient.
5. Once the initial number of PC regressors has been determined, change (or back-transform) the linear regression coefficients back to their original terms (centered and scaled)
6. Retrieve the Beta coefficients from the linear regression.
7. Finally, check the algebraic signs of the back-transformed Beta coefficients against their correlations with the response variable. If they are not the same, reject the current PC regression model and reduce the number of PCs until the signs agree.

During the regression process it was found that first PC was statistically significant for all monthly forecast models using the p-values and, at least initially, two PCs were found to be significant in the March model. However, the addition of the second PC into the March equation created a negative regression coefficient, which is the opposite sign of the correlation between the explanatory variables and the response.

The additional PC was rejected based on NRCS documentation (NRCS). PCA reduced the dimensionality of the regression equation from eight principle components to one for all three models. A stepwise process was then executed to find the optimal combination of explanatory variables to obtain the smallest root mean squared error (RMSE) (Equation 2) (Table 3) for the model prediction.

$$RMSE = \sum_{i=1}^{n} \sqrt{\frac{\left(\hat{Y}i - Yi\right)^2}{n}} \qquad \textbf{(2)}$$

$\hat{Y}i$ =Observed value, *Yi* = Model predicted value, *N* = number of observations

	March	April	May
R^2	0.46	0.57	0.78
Root Mean Square Error	10,051	9,831	7,307
Standard Error Prediction	16,339	14,764	12,000

Table 3: Model results from R-Studio for the March, April and May models.

This stepwise process was undertaken for two reasons: 1) to re-examine and confirm the earlier decision on discarding variables, and 2) to explore the possibility of removing the snow course sites from the analysis. The snow courses, while providing useful information, are highly correlated with the SNOTEL sites but not as timely for forecasting. Given the relatively small forecast area, and relatively good SNOTEL coverage, the snow courses were ultimately dropped.

The final list of explanatory variables used in the model includes:

- Mt. Lockhart - SWE
- Mt. Lockhart - Precipitation
- Waldron - SWE
- Waldron - Precipitation
- Badger Pass - SWE
- PNA - Climate Teleconnection Index

The PC analysis and regression process was repeated with the final PC regression models (March, April, and May) regressing April 1 to July 31 (or May 1 to July 31) flow volume against the first PC for the final six explanatory variables. The PC values from the linear regression were multiplied by the PC rotation matrix to return the values to their original terms of centered and scaled.

Cross Validation

The ability of the model to predict streamflow was assed using a jackknife, cross-validation technique (NRCS 2011). Jackknife analysis gives the hydrologist a better understanding of the model's prediction error (standard error), rather than using validation parameters such as RMSE or R^2.

The Jackknife technique is a process where the response and explanatory variables are left out for one year. The model is then recalculated without the omitted data and the model is used to predict response variable using the omitted explanatory values for that year. This was repeated for each year of a 20-year period. The Jackknife predicted values were then compared to the observed values. The standard error (Equation 3) (NRCS 2019) of the sum of the Jackknife predictions were calculated for each model (Table 3).

$$SE = \sum_{i=1}^{n} \sqrt{\frac{\left(\hat{Y}i - Yi\right)^2}{N - \Phi}} \qquad \textbf{(3)}$$

$\hat{Y}i$ = Observed value, *Yi* = Jackknife predicted value, *N* = number of observations, Φ = number of constraints on the model (# of PCs plus the slope).

The standard error of the Jackknife cross validation was used to determine the prediction intervals of the model (e.g. prediction interval = median prediction value +/- number of standard deviations of interest * jackknife standard error). Note that this assumes a normal, or at least symmetric distribution whereby the median and mean are equivalent.

Probability of Exceedance	Number of Standard Deviations
10%	-1.282
30%	-0.524
50%	0
70%	0.524
90%	1.282

Uncertainty Associated with the Forecast

The DNRC streamflow runoff estimation technique follows NRCS methodology using statistical best practices and professional judgment. However, like the NRCS, DNRC-generated, or any streamflow forecasts contain uncertainty. Please consider the stated standard error, the R^2 and Root Mean Squared Error associated with each model when using the predicted flow in your decision-making process. An example would be: an April 1-July 31 forecast volume of 50,000 acre-feet is issued with a standard error of +/-15,000 Acre-feet. A forecast user should expect that the range of the forecast value could be anywhere from 35,000 to 65,000 acre-feet based on the error associated with the forecast.

References:

CPC (2019)- National Weather Service, Climate Prediction Center- Northern Hemisphere Teleconnection patterns. http://www.cpc.ncep.noaa.gov/data/teledoc/teleintro.shtml

Garen, David,C,. (1992). Improved Techniques in Regression-Based Streamflow Forecasting, Journal of Water Resources Planning and Management, Vol. 118, No.6

NOAA (2019), NOAA National Centers for Environmental Information https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USC00241737/detail

NRCS, (2011)., Natural Resources Conservation Service, Chapter 7, Water Supply Forecasting, Part 622 Snow Survey and Water Supply Forecasting National Engineering Handbook.

NRCS, (2019)., Natural Resources Conservation Service, Electronic communication with National Water and Climate Center Staff.

NRCS., Technical Note: Statistical Techniques in VIPER

NOAA (2019), NOAA National Centers for Environmental Information https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USC00241737/detail