# POW: Forecasting Snow Water Equivalent

Alex Ojemann and Andrew Floyd

## Abstract

We use daily time-series measurements of snow water equivalent (SWE) to predict future SWE values as a proxy for drought and flood risk. As a metric, SWE is used by hydrologists, policy-makers, and others to estimate the availability of water resources, as an important input in river and flood planning, reservoir management, and much more. We use the ARIMA time-series forecasting method to predict the future SWE values given the time-series for each of 629 stations across the western United States.

## Approach

Our data is sequential and only contains SWE values so we will use time-series forecasting to predict future SWE values. Time-series forecasting is a subfield of machine learning that deals with predicting future values of a time-dependent variable based on its past behavior. Time-series data can come in many different forms, such as stock prices, weather patterns, or website traffic, and the goal of forecasting is to identify patterns and trends in this data that can be used to make accurate predictions.

One of the most commonly used methods for time-series forecasting is Autoregressive Integrated Moving Average (ARIMA), which is a type of linear regression model. ARIMA models make use of past values of the variable to predict future values, with a focus on capturing the trends, seasonality, and irregularities in the data. ARIMA models have been around for several decades and have proven to be highly effective in many applications.

Exponential smoothing is another popular approach for time-series forecasting. Exponential smoothing models use a weighted average of past observations, with more recent observations receiving more weight than older ones. This approach is useful for capturing short-term changes in the data and is especially effective when the data is relatively smooth.

Both ARIMA and exponential smoothing have their strengths and weaknesses, and choosing between them often depends on the specific characteristics of the data being analyzed. For example, ARIMA is often better suited for data with clear seasonality, while exponential smoothing may be more effective for smoother data.

We trained these two models on each of the stations in our data set and compared the distributions of their accuracies by station. But first we needed to ensure that the data is stationary as this is a requirement for ARIMA. It isn't a requirement for exponential smoothing but we will train both models on the stationary data to ensure that the accuracies are based on the same data.

Stationarity refers to the property of a time-series where the statistical properties of the series do not change over time. In a stationary time-series, the mean, variance, and autocorrelation structure of the series are constant across time. This means that the behavior of the time-series is predictable and that any future values are likely to be similar to past values.

We checked for stationarity using the Augmented Dickey-Fuller (ADF) test. The ADF test is based on the Augmented Dickey-Fuller regression model, which is a variation of the standard linear regression model that includes lagged values of the dependent variable as regressors. The null hypothesis of the ADF test is that the time-series is non-stationary. The alternative hypothesis is that the time-series is stationary, meaning it does not have a unit root. We chose a level of significance of 0.05, so if the p value of the ADF test was greater than 0.05, we fail to reject the null hypothesis that the time-series is non-stationary. We did this test for each individual station in our data set.

For each station where we failed to reject that the time-series isn't stationary, we made it stationary using differencing. Differencing is a method used to transform non-stationary time-series data into stationary time-series data. The basic idea behind differencing is to take the difference between consecutive observations in the time-series. By taking the difference between consecutive observations, we can remove the trend and/or seasonality that may be present in the data. This can make the data stationary, which means that the mean and variance of the data remain constant over time.

Onced the data was made stationary, we trained each model for each station and used rolling window cross validation to evaluate its RMSE. Rolling window cross validation is a way to evaluate a time-series model where you choose an evaluation period, train a model based on that period, predict future values on a shorter period that immediately follows the training period and evaluate its RMSE relative to the true values, then move the periods forward one day and repeat the process until the end of the test period reaches the end of the data set. Once this is complete you have RMSE values for each testing period and can take the average to get a well validated RMSE for the model as a whole. One limitation of rolling window cross validation is that it can't use the first training set for evaluation because it's never part of the test set.

The ARIMA model takes in two significant parameters, p and q. The parameter p represents the number of previous days' values to incorporate in the model and q represents the number of previous days' error terms to incorporate into the model. They are commonly set to values of 2 or less so we set them both to 1. There are also two terms P and Q for seasonal ARIMA which are the same as p and q but for the previous season's values and errors, which were also both set to 1. We did not have time to tune these hyperparameters because of the large volume of data we are using .

## Data

We obtained ten years of daily snow water equivalent readings from 837 SNOTEL sites across the western United States. SNOTEL is an automated system of snowpack and other related climate sensors operated by the National Resources Conservation Service (NRCS) of the United States Department of Agriculture (USDA). Snow water equivalent is the depth of water that would cover the ground if all of the snow cover were suddenly in the liquid state. Snow water equivalent is just one measure of the snowpack and climate that SNOTEL sites capture, but it is an input to other important models, for example in river forecasting models.

We obtained the data from SNOTEL using the `snotelr` R package (Hufkens 2022). The time period for our observations began on Jan 1st, 2013 and it ended on Dec 31st, 2022. We chose a 10-year period to ensure we would have a sufficient quantity of data, captured over a long enough period for the effects of various climatic patterns (e.g. El Niño/La Niña, Polar Vortex(s), Bomb Cyclones, etc, etc.) to be observed. We also developed tools in Python to allow us to map stations to locations, to extract other metadata about the sites, and to collate the data by station and by year. We used these tools to compile the data into time-series, which could be admitted to the ADF testing for stationarity (adjusted as indicated) and fed into the ARIMA model.

When preprocessing the data, we ensured that each station that we used had data for every day in the specified 10 year period and had no NA values for SWE. This filtered down the 837 original stations to 629.

## Results

For the ARIMA models, the average RMSE across the 629 stations was 6.411. A histogram and boxplot of the RMSE for each station is shown below.

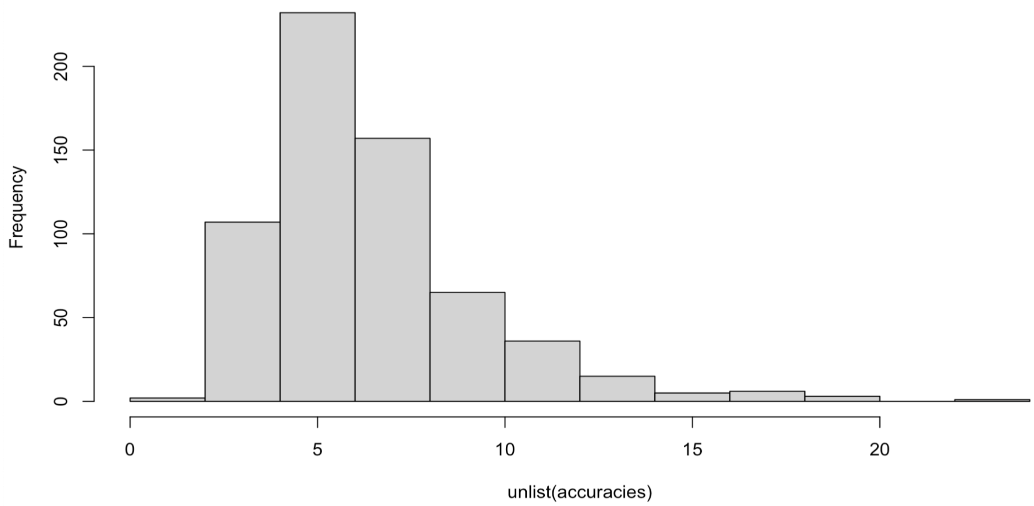


**Figure 1. Relative Frequencies of Station Accuracy for ARIMA**

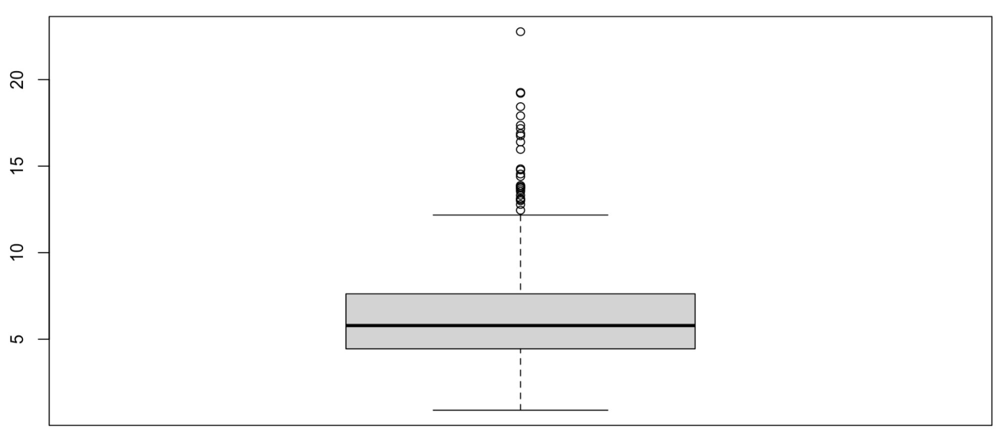


**Figure 2. Boxplot of Station Accuracy for ARIMA**

We can see that the results are centered around 5-10 with significant right skew and many outliers on the right but none on the left because the bottom whisker goes almost to 0.

For the exponential smoothing models, the average RMSE across the 629 stations was 6.672. A histogram and boxplot of the RMSE for each station is shown below.

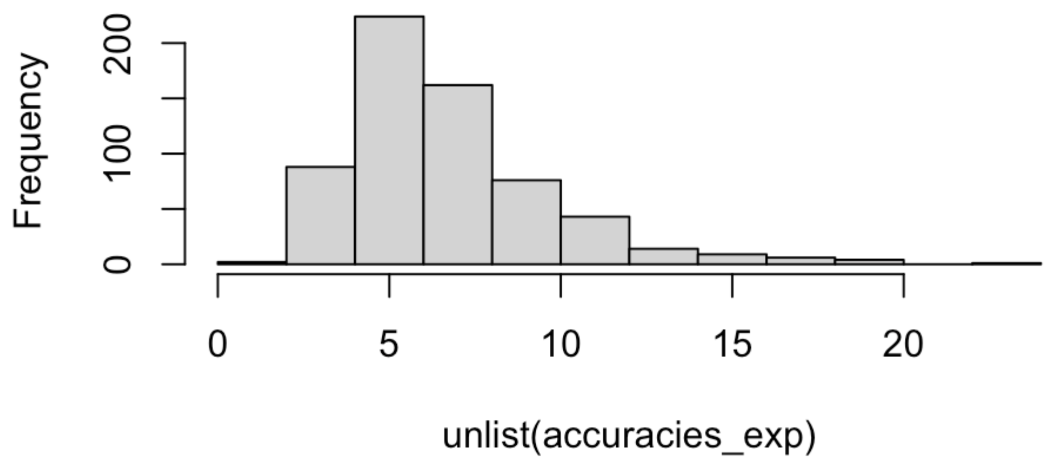


**Figure 3. Relative Frequencies of Station Accuracy for Exponential Smoothing**

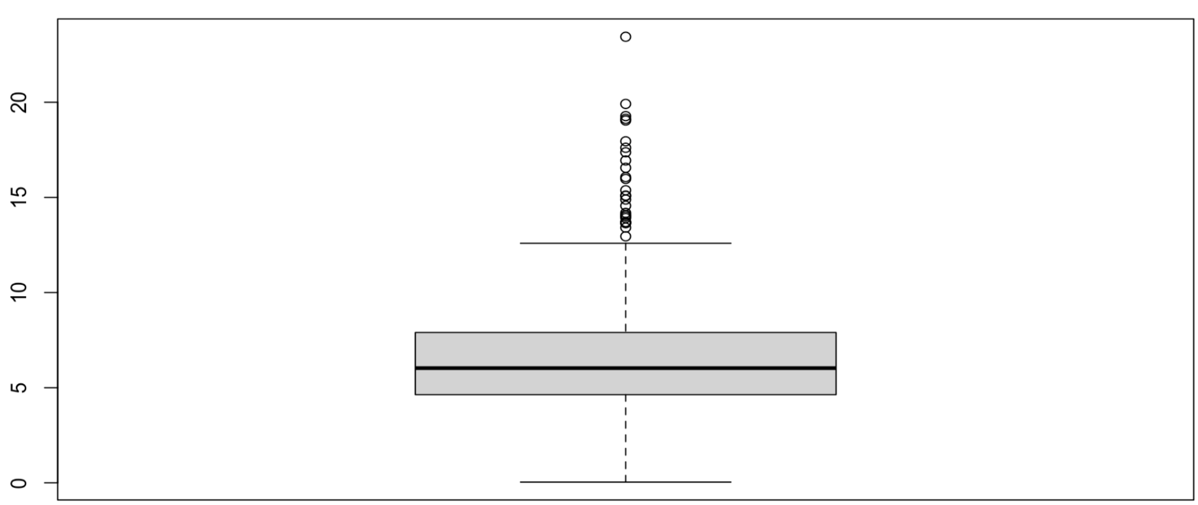


**Figure 4. Boxplot of Station Accuracy for Exponential Smoothing**

Very similarly to the ARIMA model, we can see that the results are centered around 5-10 with significant right skew and many outliers on the right but none on the left because the bottom whisker goes almost to 0.

## **Discussion**

Ultimately while the ARIMA and exponential smoothing models had very similar distributions of RMSE, the average RMSE was lower for the ARIMA models than the exponential smoothing models. ARIMA is believed to perform better with stationary data which has short term trends so this isn't surprising. These results might lead to more investigation of ARIMA models for predicting SWE using different hyperparameters or using multiple nearby stations in the same model so there is more data available. Future plans for work on this data could entail training other models that are useful for time series forecasting such as recurrent neural networks or long-short term memory neural networks and comparing their average RMSE to those we computed for ARIMA and exponential smoothing.

## **Bibliography**

Hufkens, K. (2022). snotelr: a toolbox to facilitate easy SNOTEL data exploration and downloads in R. Zenodo. https://doi.org/10.5281/zenodo.7012728.