**An analysis of Reservoir Levels in California**

*Jeremy Chan, Fourth-Year Economics and Statistics Major*

**Table of Contents**

**1 Summary 2**

**2 Introduction 2**

**3 Methods 3**

3.1 Initial Data Cleaning and Inspection 3

3.2 Fitting a Time Series Model 3

3.3 Forecasting 5

**4 Results and conclusions 6**

4.1 Initial Data Cleaning and Inspection Results 6

* 1. Fitting a Time Series Model Results 7

4.3 Forecasting Results 9

**5 Code Appendix**  12

**1 Summary**

This report contains an analysis of water reservoir levels in 11 selected California reservoirs. The ultimate goal of the report was to provide a summary of the current reservoir levels and provide a forecast for the year of 2016. As the report was written in May of 2016, there are already some data points from 2016 which will be compared with the forecasted points. To accomplish the task, first the data was plotted and inspected. A stationary time series was then modeled and predictions made from the residuals of that model.

**2 Introduction**

Since 2012, California has faced severe drought conditions, the likes of which the state has not seen since the 1970s. As not only the most populous state but also the state which produces the most food[[1]](#footnote-1), California faces unique challenges during a drought. It is, therefore, of great interest to track the level of drought which California faces and to make predictions on if and when the state will return to normal water levels.

According to the United States Geological Service, “a drought is a period of drier-than-normal conditions that results in water-related problems” resulting from “less than normal [rainfall] for several weeks, months, or years.”[[2]](#footnote-2) One key indicator of drought conditions is reservoir levels. The California Department of Water, provides reservoir information on 199 different reservoirs all throughout the state[[3]](#footnote-3). Of these, 199 reservoirs, eleven were selected for analysis in this report. The 11 that were selected were part of a group of 12 which the California Department of Water uses for their daily report on the conditions for “major reservoirs.”[[4]](#footnote-4) The reservoirs used are as follows: Trinity Lake, Lake Shasta, Lake Oroville, New Melones, Folsom Lake, Don Pedro, San Luis, Millerton Lake, Pine Flat, Castaic Lake, and Lake Perris. Exchequer reservoir is the one reservoir that was excluded because of an error in retrieving the data from the California Data Exchange Center (CDEC). These 11 reservoirs create a fairly representative picture of the overall reservoir levels in California and have been a steady indicator of drought levels. Thus, it is of interest to analyze and predict the levels in these reservoirs.

All of the data from the remaining reservoirs was retrieved through a query function from the California Department of Water Resources through the California Data Exchange Center. The query function is part of an R package called “sharpshootR” and automatically retrieves reservoir level readings from each of the reservoirs selected for observation. The data was queried at a monthly level, from the first recorded measurement to the most recent measurement of April 2016 and measured in acre-feet, with one acre-foot being equivalent to 325,851 gallons of water. The total capacity of each reservoir was also retrieved from the CDEC and each reservoir’s current level reading was divided by its total capacity to produce a reservoir capacity percentage, our observation of interest. This method was chosen for ease of interpreting the results by readers of this report.

The ultimate goal of this report is to provide an overview of the selected reservoirs’ current levels, provide a forecast for the 2016 calendar year for each reservoir, and to compare each reservoir’s forecast with the actual observations from 2016, up through the present time. To accomplish this, the first step is to retrieve, inspect, and clean the data which has been retrieved. Any outliers will be removed if they are deemed as incorrect readings and any non-stabilized variance will be corrected for. Additionally, if there are any reservoir readings which appear clearly non-linear, they will be adjusted for on a case-by-case basis.

The next step is to remove any deterministic component from each time series. Trend and seasonality will be removed and once stationarity has been established, a model will be fitted to the remaining residuals. From that model, a 12-month forecast for the calendar year of 2016 will be created. If the residuals from the model are normal, then a prediction interval will be provided in addition to the point forecasts.

**3 Methods and Results**

In this section, the methods used for analysis and the results of aforementioned analysis will be detailed. The analysis will be split into the following sections: Data Retrieval, Data Cleaning, Data Visualization, Removing Deterministic Components, Fitting a Time Series Model, and Forecasting. Due to the number of reservoirs that are being analyzed, all of the above tasks were combined into a single function in R which simplifies the actual coding analysis aspect of this report.

**3.1 Data Retrieval**

All of the data contained in this report was retrieved from the California Department of Water through the California Data Exchange Center (CDEC). To automate the retrieval of this data, the package “sharpshootR” was used. The package contains a function called CDECquery which retrieves the reservoir levels for a given reservoir between a certain date range and at a certain frequency. For this report, the following 11 reservoirs were used as a representative sample of California’s overall reservoir levels: Trinity Lake, Lake Shasta, Lake Oroville, New Melones, Folsom Lake, Don Pedro, San Luis, Millerton Lake, Pine Flat, Castaic Lake, and Lake Perris. The data was retrieved on a monthly basis spanning from when each reservoir was first opened up through April 2016. After retrieving the monthly capacity readings for each of these reservoirs, the capacity reading was divided by each reservoir’s overall capacity, resulting in a percentage of capacity for each reservoir. This was used as the primary unit of observation.

1. http://www.ers.usda.gov/faqs.aspx [↑](#footnote-ref-1)
2. http://ca.water.usgs.gov/data/drought/ [↑](#footnote-ref-2)
3. http://cdec.water.ca.gov/misc/resinfo.html [↑](#footnote-ref-3)
4. http://cdec.water.ca.gov/cdecapp/resapp/getResGraphsMain.action [↑](#footnote-ref-4)