Water Level at Lake Mead

Peter Trubey

May 16, 2013

Abstract

In this paper we discuss the purpose and creation of Lake Mead, in regards to its water level and down-river obligations. We then take a look at the historical plot of the water level, and make observations about it. We recognize a substantial difference between before Lake Powell was constructed and after, causing us to truncate the data we use for subsequent analysis. We then fit models to the data and observe the diagnostic plots of those models to choose a best model. From there, we omit and predict the last 3 years of water level elevations, and then forecast up to 6 years into the future.

Introduction

Lake Mead is the largest man-made lake in the United States, located in the American Southwest near Las Vegas on the border between Nevada and Arizona. It began filling in 1936 after the construction of Hoover Dam. Its main tributary is the Colorado River, which passes through other managed reservoirs along with the Grand Canyon on its journey to Lake Mead. From Lake Mead, the Colorado heads south, forming the boundary between California and Arizona. It terminates into the Gulf of California. [3]

The water level at Lake Mead is governed by the rainfall and snowfall for the watershed that fills it. If there is little precipitation, then the water level necessarily drops. The water level is also governed by the outflow of the reservoir. It loses an estimated 800,000 acre-feet of water per year due to evaporation.[5] The water it stores provides drinking and irrigation water for communities including Los Angeles, San Diego, Phoenix, and all in between. [7] These communities rely on the water to be flowing consistently from Lake Mead. The agreement between states and Mexico that governs this water usage is called the Colorado River Compact, drawn up in 1922. [2]

Hoover Dam is also a hydro-electric power plant, generating at maximum 2080 megawatts. Of course, this generation capacity is tied directly to the amount of water the reservoir

management is willing or required to let through. As the reservoir goes down, so does the ability to generate electricity. [3]

Data

The data used in this analysis is elevation levels in feet of the water's surface, taken monthly from the time Hoover Dam was built in 1936, till April 2013. [4] In 1967, engineers began diverting water that was due for Lake Mead and downstream to fill Lake Powell, which had just been completed. [1] Due to this, there exists a massive difference in variability of seasonal water levels from before 1967, and after. We can see that a dam upstream will limit the variability of the water level downstream. We can observe this effect in figure 1.

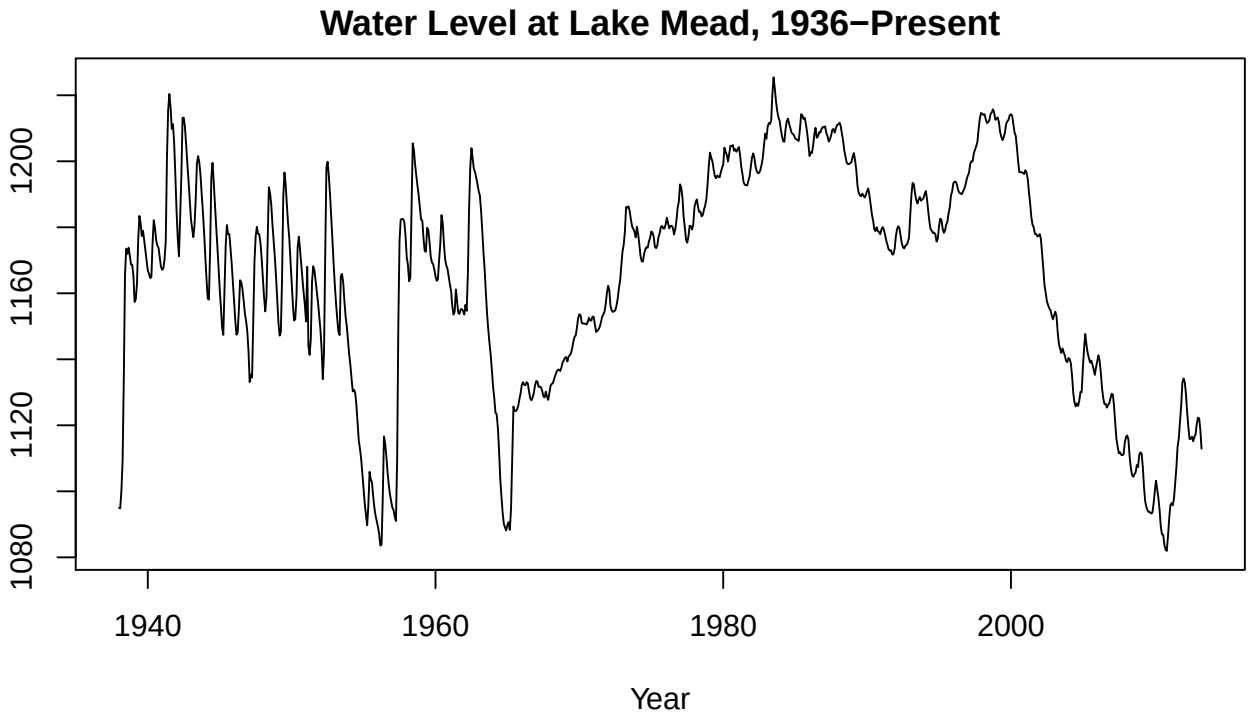


Figure 1: The effect of building Lake Powell upstream from Lake Mead is evident in the difference in variability between before it was built and after.

Method

Due to the significant difference in variance from before and after the building of Lake Powell, we will be truncating the data set to only include the portion after Lake Powell was finished.

We will be fitting a seasonal ARIMA model to this data.

$$\Phi_P(B^S)\phi_p(B)(1-B^S)^D(1-B)^d Z_t = \theta_q(B)\Theta_Q(B)a_t$$

Because the data is monthly, a year is 12 months long, and we would expect the season to have a significant effect on incoming water to lake Mead, we will begin by differencing by 12. From there, we will be trying various methods of fitting models to the data. After generating a working base model, we will omit some of the data, retrain the model, and attempt to predict the missing data. We will then evaluate our predictions, and attempt to forecast into the future.

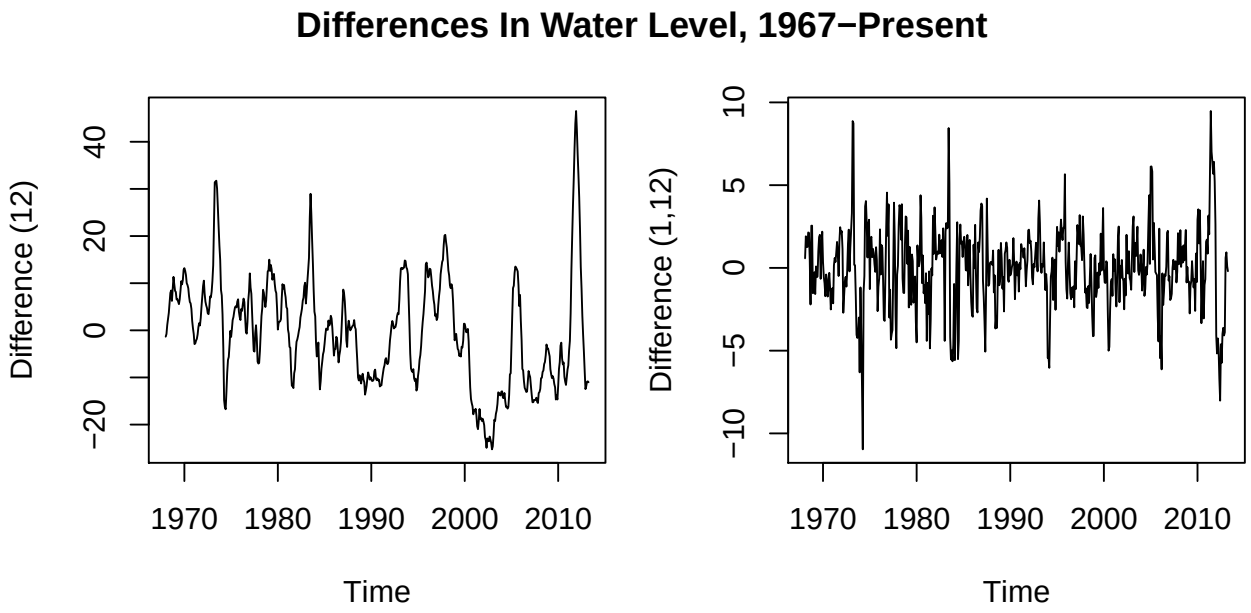


Figure 2: On the Left we've differenced the series by 12 steps. On the Right, we've further differenced the series again by a single step.

Analysis

$$(1-B^{12})(1-B)Z_t = \theta_4(B)\Theta_4(B)a_t \tag{1}$$

$$\Phi_1(B)\phi_1(B)(1-B^12)(1-B)Z_t = \Theta_2(B)a_t \tag{2}$$

We have fit two models to the data. For model 1, we fit an ARIMA(0,1,4)x(0,1,4) model. This generated a model that fit the data relatively well. It generates an Akaike's

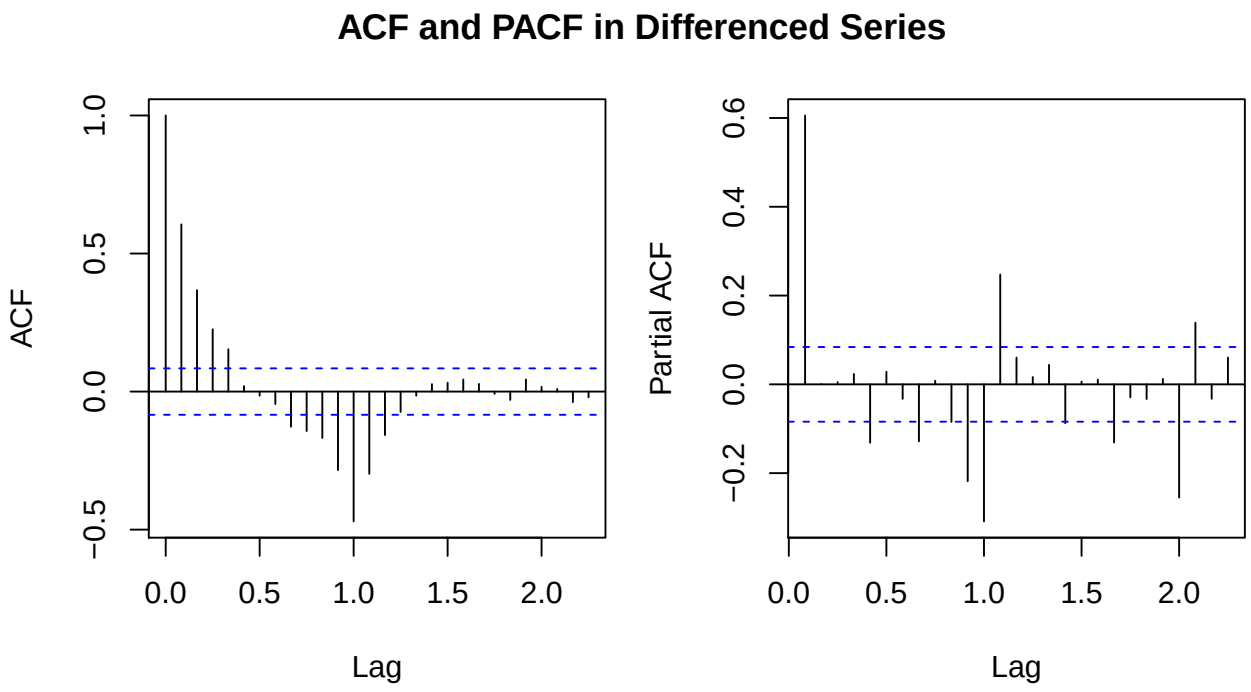


Figure 3: Here we plot the ACF and PACF of the series differenced by (1,12). We see a distinct seasonal component.

Information Criterion of 2063. For model 2, we fit an ARIMA(1,1,0)x(1,1,2) model. This also fits the data, and eliminates enough variables that it generates a lower AIC score of 2061. However, if we look at the p-values for the Ljung-Box test in figures 6 and 7, model 1 exhibits much higher p-values across the board as compared to model 2. While it is true that model 2 does not display a significant lack of fit here, it comes dangerously close. Due to this, I'm going to pick model 1 as the better model.

Forecasts and Results

in figure 4, we omit the data for the last 3 years from training the model, and then attempt to predict those 3 years based on the trained model. What we find is that the sharp spike in 2011 is difficult to predict, without knowing what else was going on in the world. For the previous 10 years, the American Southwest had been in a very serious drought. 2011 was a relatively wet year in a long streak of dry years, resulting in a not unprecedented, but unexpected increase in the stored volume of water. [6] However, Lake Mead is still significantly below capacity.

In figure 5, we attempt to predict into the future by up to 6 years. The model still has a downward trend, and owing to the fact it's a non-stationary model, the error lines

diverge. The model predicts that in April 2019, the water level will be somewhere between 1000 and 1181 feet, which seems is a pretty safe bet.

Conclusion

The model seemed to do a very good job of predicting the seasonal variation. However, it could not account for the large jumps between dry and wet years. This is to be expected, the model can only make predictions based on what came previously. If we wanted to account for those large variations, we would need to incorporate more sources of data in the analysis. Rainfall and snowfall patterns, the water levels at the reservoirs preceding and following Lake Mead, temperature levels, and water usage patterns of the communities served by Lake Mead all have a role to play. Beyond that, we realize Lake Mead is just a part of a managed system, with water engineers actively determining whether and where to raise or lower the water level to best serve those communities that rely on it. As such, without accounting for that, this sort of analysis for this subject is rather limited.

References

[1] Glen canyon history. http://www.lakepowell.com/glen-canyon-history.aspx.

[2] History - central arizona project. http://www.cap-az.com/index.php/cap-history.

[3] Hoover dam: Lake mead (faqs). http://www.usbr.gov/lc/hooverdam/faqs/lakefaqs.html.

[4] Lake mead at hoover dam, elevation. http://www.usbr.gov/lc/region/g4000/hourly/mead-elv.html.

[5] Lake mead online. http://www.lakemeadonline.com/Home.html.

[6] Wet year to boost level of lake mead. http://www.reviewjournal.com/news/water-environment/wet-year-boost-level-lake-mead.

[7] Richard F. Pourade. The rising tide: Chapter 9: Making a river. http://www.sandiegohistory.org/books/pourade/rising/risingchapter9.htm.

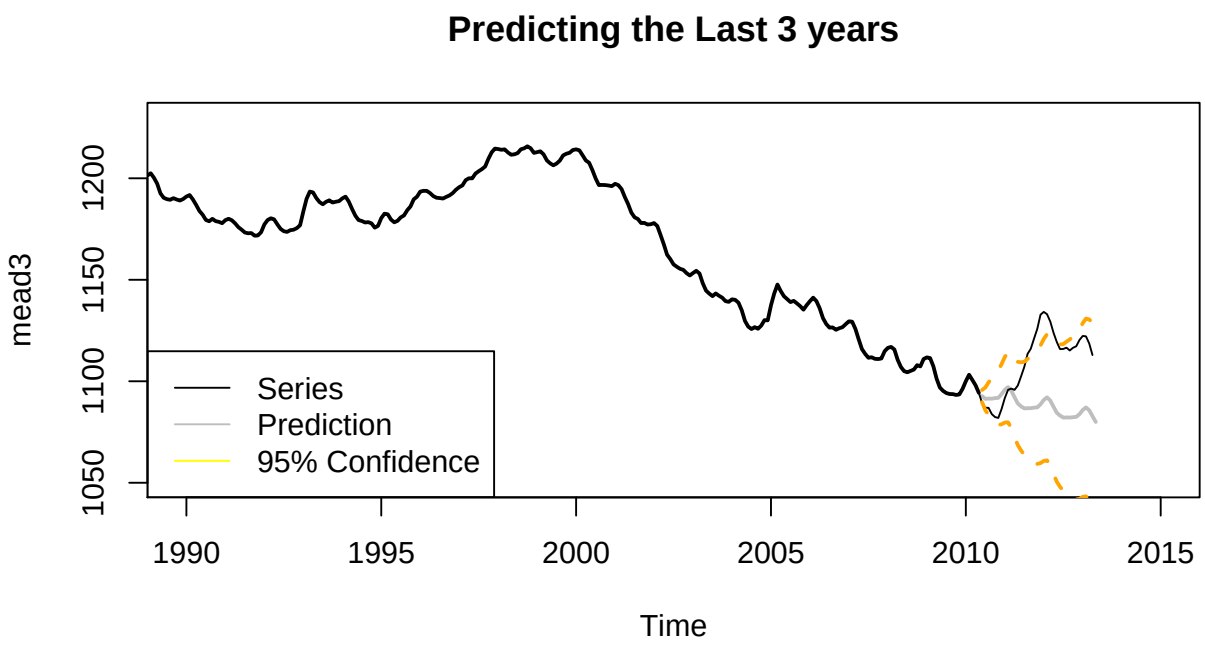


Figure 4: Attempting to Predict the last 3 years of Data

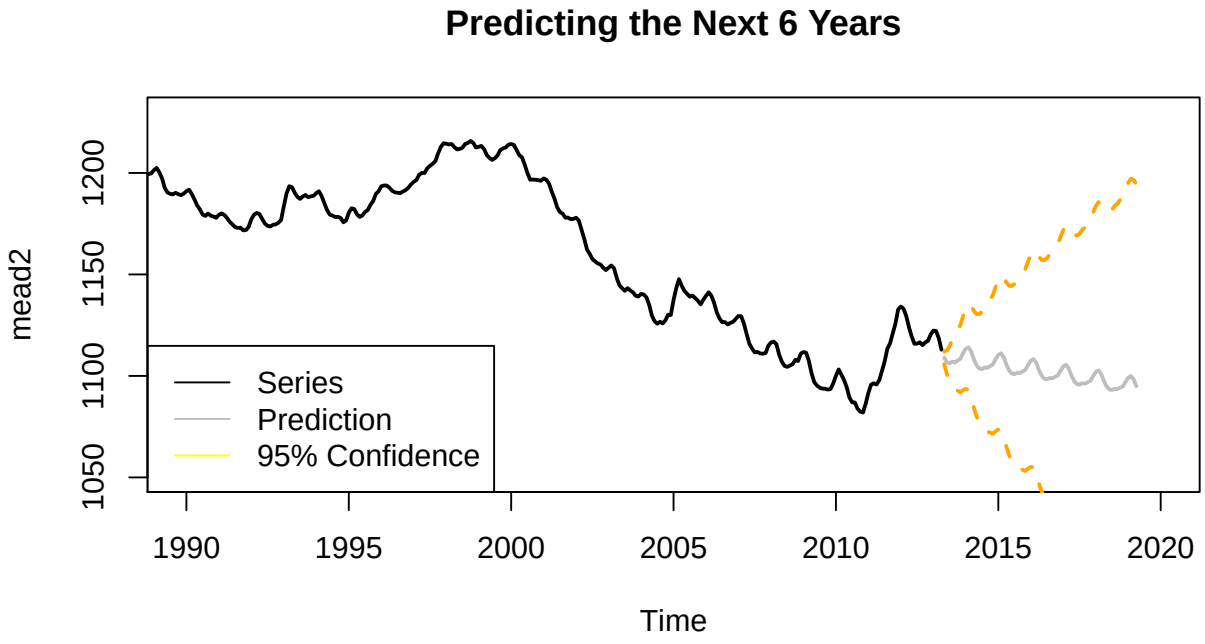


Figure 5: Forecasting 6 Years Into the Future

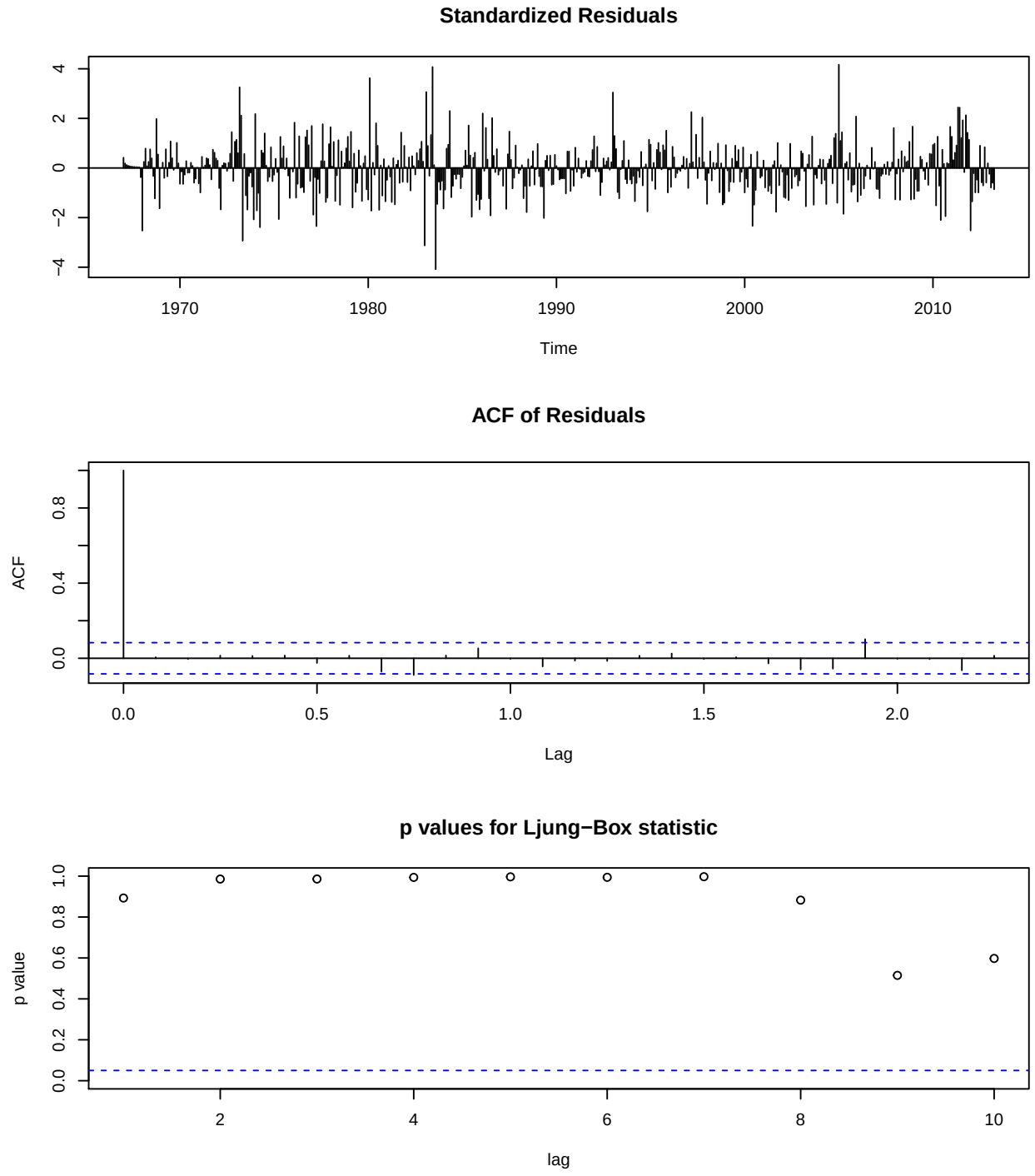


Figure 6: Diagnostic Plots for ARIMA(0,1,4)x(0,1,4)

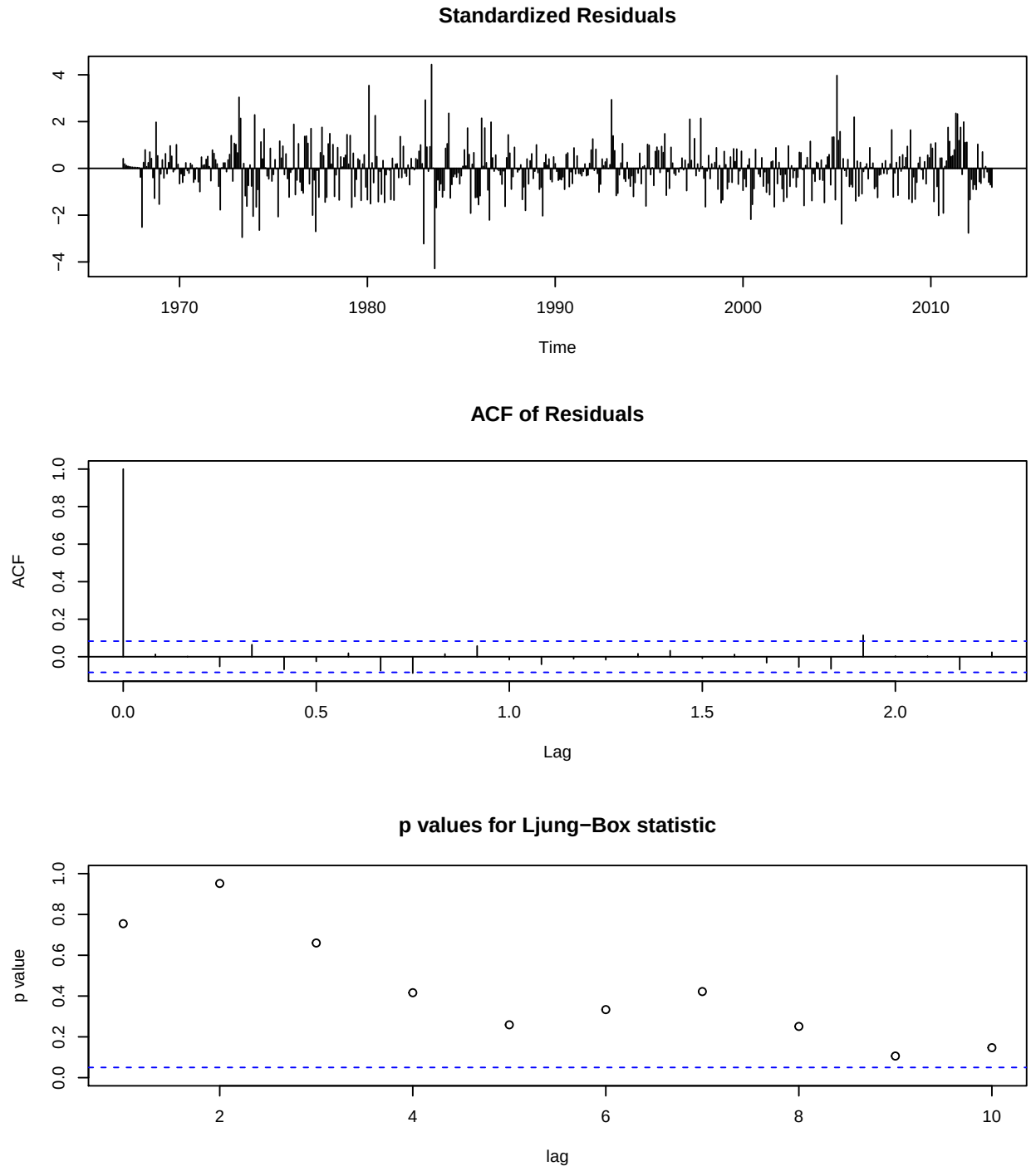


Figure 7: Diagnostic Plots for ARIMA(1,1,0)x(1,1,2)

Code Block 1: Structure of the Data Set, after stripping from website [4]

```
mead4 = scan("mead.csv", skip = 0, sep = ",", dec = ".", what = "numeric")
mead4 = mead[1:48]
print(mead4)

##  [1] 1095 1095 1100 1109 1134 1166 1174 1172 1174 1171 1169 1169
## [13] 1165 1157 1158 1163 1176 1183 1181 1177 1179 1176 1173 1170
## [25] 1167 1166 1165 1165 1175 1182 1180 1176 1174 1174 1171 1168
## [37] 1167 1168 1170 1176 1201 1216 1220 1216 1210 1211 1206 1196
```

Code Block 2: Code Used in Generating this Document, Part I

```
mead = scan("mead.csv", skip = 0, sep = ",", dec = ".", what = "numeric")
mead = mead[37:940]
mead = as.numeric(mead)
mead = ts(mead, frequency = 12, start = c(1938, 1))

# Truncating Series
mead2 = scan("mead.csv", skip = 32, sep = ",", dec = ".", what = "numeric")
mead2 = mead2[1:556]
mead2 = as.numeric(mead2)
mead2 = ts(mead2, frequency = 12, start = c(1967, 1))

# Model Fitting
m1 = arima(mead2, order = c(0, 1, 4), seasonal = list(order = c(0,
    1, 4), period = 12))
m2 = arima(mead2, order = c(1, 1, 0), seasonal = list(order = c(1,
    1, 2), period = 12))

# Plotting Original Series
par(mfrow = c(1, 1), mar = c(4, 2, 2, 1))
plot(mead, ylab = "Elevation", xlab = "Year", main = "Water Level at Lake Mead, 1936-Present")
```

Code Block 3: Code used in Generating this Document, Part II

```
# Differencing the Series
par(mfrow = c(1, 2), oma = c(2, 0, 2, 0), mar = c(5, 4, 1, 1))
plot(diff(mead2, lag = 12), ylab = "Difference (12)")
plot(diff(diff(mead2, lag = 12)), ylab = "Difference (1,12)")
title("Differences In Water Level, 1967-Present", outer = TRUE)

# Diagnostic Plots of Differenced Series
par(mfrow = c(1, 2), oma = c(0, 0, 2, 0), mar = c(5, 4, 1, 1))
acf(diff(diff(mead2, lag = 12)), main = "")
pacf(diff(diff(mead2, lag = 12)), main = "")
title("ACF and PACF in Differenced Series", outer = TRUE)

# Predictions for Last 3 Years
l = length(mead2) - 35
mead3 = ts(mead2[1:l], frequency = 12, start = c(1967, 1))
m3 = arima(mead3, order = c(0, 1, 4), seasonal = list(order = c(0,
    1, 4), period = 12))
m3.pred = predict(m3, n.ahead = 36)
plot(mead3, xlim = c(1990, 2015), ylim = c(1050, 1230), main = "Predicting the Last 3 years",
    lty = 1, lwd = 2, col = "black")
lines(m3.pred$pred, lty = 1, lwd = 2, col = "gray")
lines(mead2, lwd = 1, lty = 1)
lines(m3.pred$pred - 1.96 * m3.pred$se, lty = 2, lwd = 2, col = "yellow")
lines(m3.pred$pred + 1.96 * m3.pred$se, lty = 2, lwd = 2, col = "yellow")
legend("bottomleft", col = c("black", "gray", "yellow"), lty = 1,
    legend = c("Series", "Prediction", "95% Confidence"))

# Predictions for Next 6 Years
m2.pred = predict(m2, n.ahead = 72)
plot(mead2, xlim = c(1990, 2020), ylim = c(1050, 1230), main = "Predicting the Next 6 Years",
    lty = 1, lwd = 2, col = "black")
lines(m2.pred$pred, lwd = 2, lty = 1, col = "gray")
lines(m2.pred$pred - 1.96 * m2.pred$se, lty = 2, lwd = 2, col = "yellow")
lines(m2.pred$pred + 1.96 * m2.pred$se, lty = 2, lwd = 2, col = "yellow")
legend("bottomleft", col = c("black", "gray", "yellow"), lty = 1,
    legend = c("Series", "Prediction", "95% Confidence"))
```