library(haven)

library(forecast)

library(fma)

library(tseries)

library(expsmooth)

library(lmtest)

library(zoo)

library(ggplot2)

library(lubridate)

library(dplyr)

#Read in well water depth data

well <- read.csv("C:\\Users\\aalmq\\OneDrive\\Documents\\Time Series II\\ts\_well.csv", header=T, stringsAsFactors = F)

#Create Date Time Object

well$datetime <- paste(well$date, well$time)

well$datetime <- as.POSIXct(well$datetime, format="%m/%d/%Y %H:%M")

#Aggregate Well Data by hour

well\_agg <- aggregate(list(Corrected = well$Corrected),

list(datetime = cut(well$datetime, "1 hour")),

mean)

well\_agg$datetime <- as.POSIXct(well\_agg$datetime)

time\_seq <- seq(as.POSIXct("2007-10-01 01:00:00"), as.POSIXct("2018-06-12 23:00:00"), "hour")

time\_seq <- as.data.frame(time\_seq)

names(time\_seq) <- c("datetime")

str(time\_seq)

#Merge Left Outer join of time sequence on well data

well\_merge <- merge(time\_seq, well\_agg, by = "datetime", all.x=TRUE)

#Find number of hourly entries with missing well data

sum(is.na(well\_merge$Corrected))

#Find which hourly entries are missing

well\_merge %>% filter(is.na(well\_merge$Corrected))

#Create time series object

Well = ts(well\_merge$Corrected, frequency=24\*365.25)

Well = tsclean(Well)

#Using a holdout data set for last week of data

training=subset(Well,end=length(Well)-24\*7)

test=subset(Well,start=length(Well)-(24\*7-1))

# Time Series Decomposition ...STL#

decomp\_stl <- stl(training, s.window = 7)

plot(decomp\_stl)

#Well Data with Seasonal Component removed

deseasonal\_well = seasadj(decomp\_stl)

plot(deseasonal\_well)

plot(Well, col = "grey", main = "Well Water Depth - Trend/Cycle", xlab = "", ylab = "Water Depth (m)", lwd = 2)

lines(decomp\_stl$time.series[,2], col = "red", lwd = 2)

seas\_well=Well-decomp\_stl$time.series[,1]

plot(Well, col = "grey", main = "Well Water Depth - Seasonally Adjusted", xlab = "", ylab = "Water Depth (m)", lwd = 2)

lines(seas\_well, col = "red", lwd = 2)

# Augmented Dickey-Fuller Testing #

adf.test(training, k=0, alternative = "stationary")

count\_well = diff(deseasonal\_well, differences = 1)

plot(count\_well)

adf.test(count\_well, alternative = "stationary", k=0)

plot(count\_well)

Acf(count\_well, main='ACF for Differenced Series', lag = 10)

Pacf(count\_well, main='PACF for Differenced Series', lag = 10)

fit<-auto.arima(deseasonal\_well, seasonal=FALSE)

summary(fit)

tsdisplay(residuals(fit), lag.max=10, main='(2,1,2) Model Residuals')

#Fit Using Seasonal Model Methods

arima\_seas<-Arima(training,order=c(2,1,2),xreg=fourier(training,K=1))

summary(arima\_seas)

fcast <- forecast(arima\_seas,xreg=fourier(training, K=1, h=24\*7))

error=test-fcast$mean

MAE=mean(abs(error))

MAPE=mean(abs(error)/abs(test))

#Plot Forecase vs Actual

plot(fcast$mean, ylim=c(7.5, 10.5), xlab="Hourly Data Over a Week", ylab="Water Depth", xaxt="n")

lines(test, col="red")