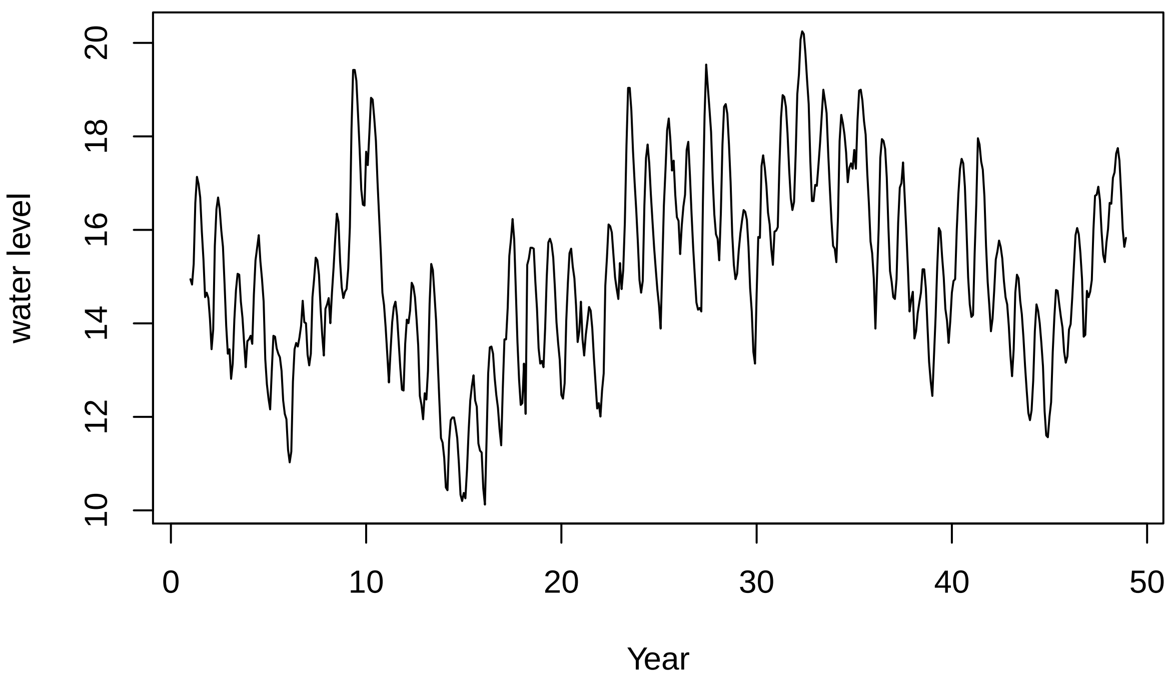
Scenario 1

Nipun

20/04/2021

```
library("fpp2")
library("forecast")
library("astsa")
library("tseries")
library("fGarch")
```

```
hyd_post <- read.table("hyd_post.txt",header = TRUE, stringsAsFactors = FALSE, sep=",")
hyd_ts <- ts(hyd_post$x, start=1, frequency=12)
plot(hyd_ts, ylab="water level", xlab="Year")
```



From the time series plot of hyd_post, we observe that there is a strong seasonal pattern within each year and somewhat cyclic behavior with period about 30-35 years. Since we are only given data for about 48 years, it is hard to be certain of a strong cyclic behavior in the time series.

We will attempt to fit SARIMA, ETS and NNAR models in order to forecast the time series 24 steps ahead. Then we will use perform cross-validation to evaluate the performance of each model and compare the significance of predictive accuracy of SARIMA, ETS and NNAR models by using Diebold-Marino Test.

```r
fets <- function(x, h) {
  forecast(ets(x), h = h)
}

farima <- function(x, h) {
  forecast(auto.arima(x), h = h)
}

fnn <- function(x, h) {
  forecast(nnetar(x, lambda = 0), h = h)
}

cvETS <- tsCV(hyd_ts, fets, h = 24)
cvNNAR <- tsCV(hyd_ts, fnn, h = 24)
cvAUTOAR <- tsCV(hyd_ts, farima, h = 24)

cvETS = na.remove(cvETS)
cvNNAR = na.remove(cvNNAR)
cvAUTOAR = na.remove(cvAUTOAR)

B = round(0.25*length(hyd_ts))

ETS.test= cvETS[(length(cvETS)-B):length(cvETS)]
NNAR.test= cvNNAR[ (length(cvNNAR)-B):length(cvNNAR)]
AR.test= cvAUTOAR[(length(cvAUTOAR)-B):length(cvAUTOAR)]



mean(ETS.test^2)
```

```
## [1] 3.512451
```

```r
mean(NNAR.test^2)
```

```
## [1] 13.60325
```

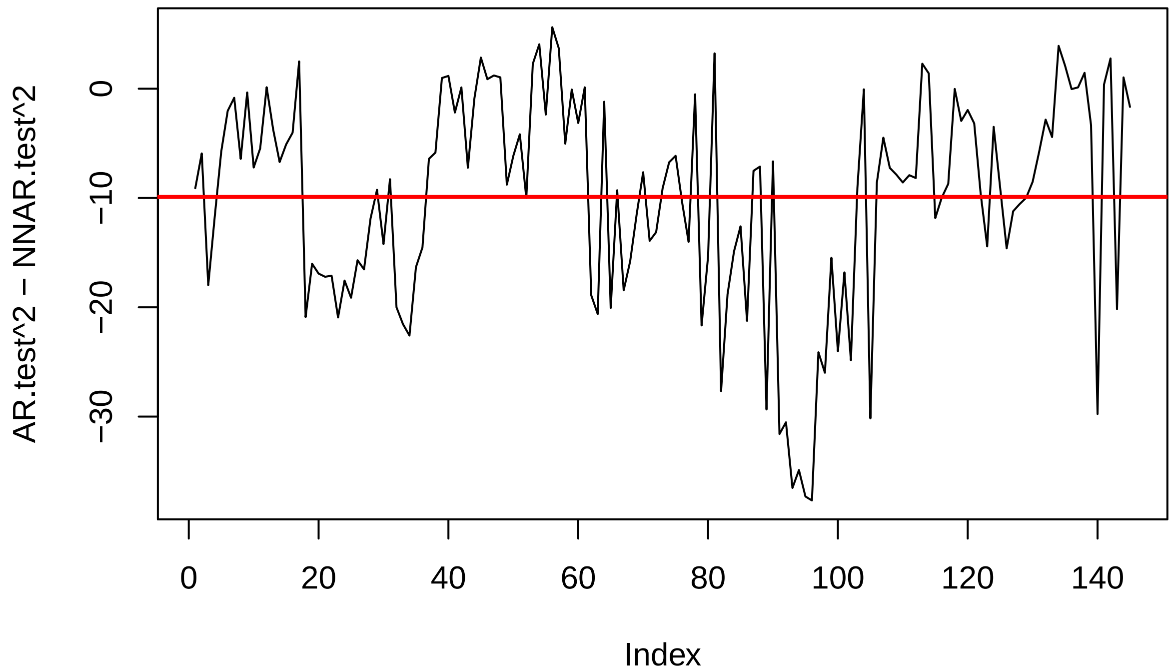
```r
mean(AR.test^2)
```

```
## [1] 3.698144
```

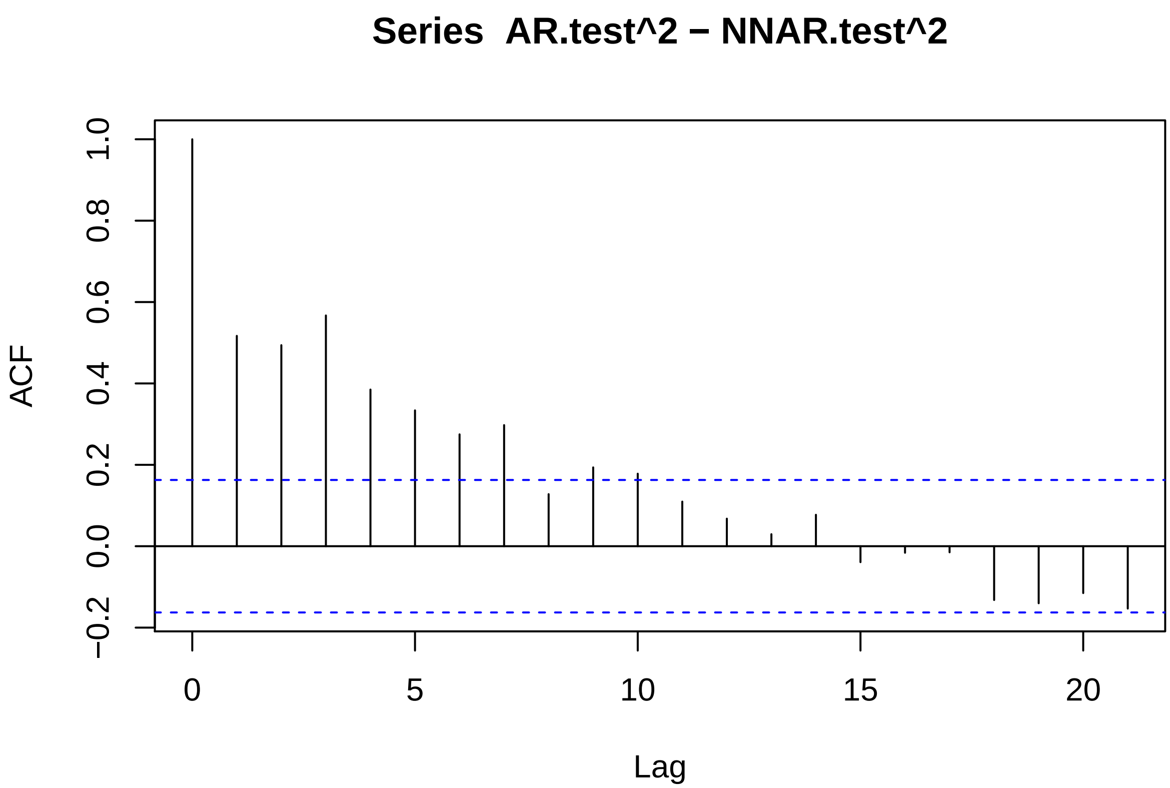
From the above results of cross-validation, we observe that ARIMA model did the best in terms of cross validated mean squared error, following by ETS and then NNAR.

Now, we will perform the DM test to compare the significance of difference in the predictive accuracy of ARIMA model to NNAR and ETS model.

```r
plot(AR.test^2 - NNAR.test^2,type='l')
abline(h=mean(AR.test^2 - NNAR.test^2),col=2,lwd=2)
```




```
acf(AR.test^2 - NNAR.test^2)
```



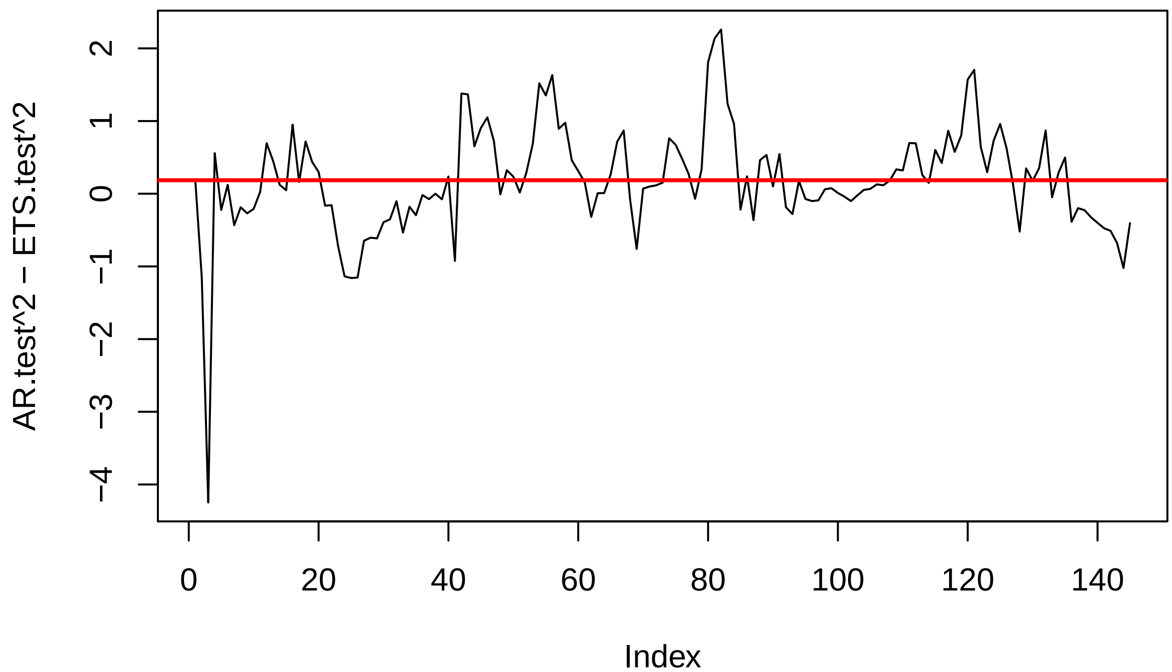

```
dm.test(AR.test,NNAR.test)
```

```
##
##  Diebold-Mariano Test
```

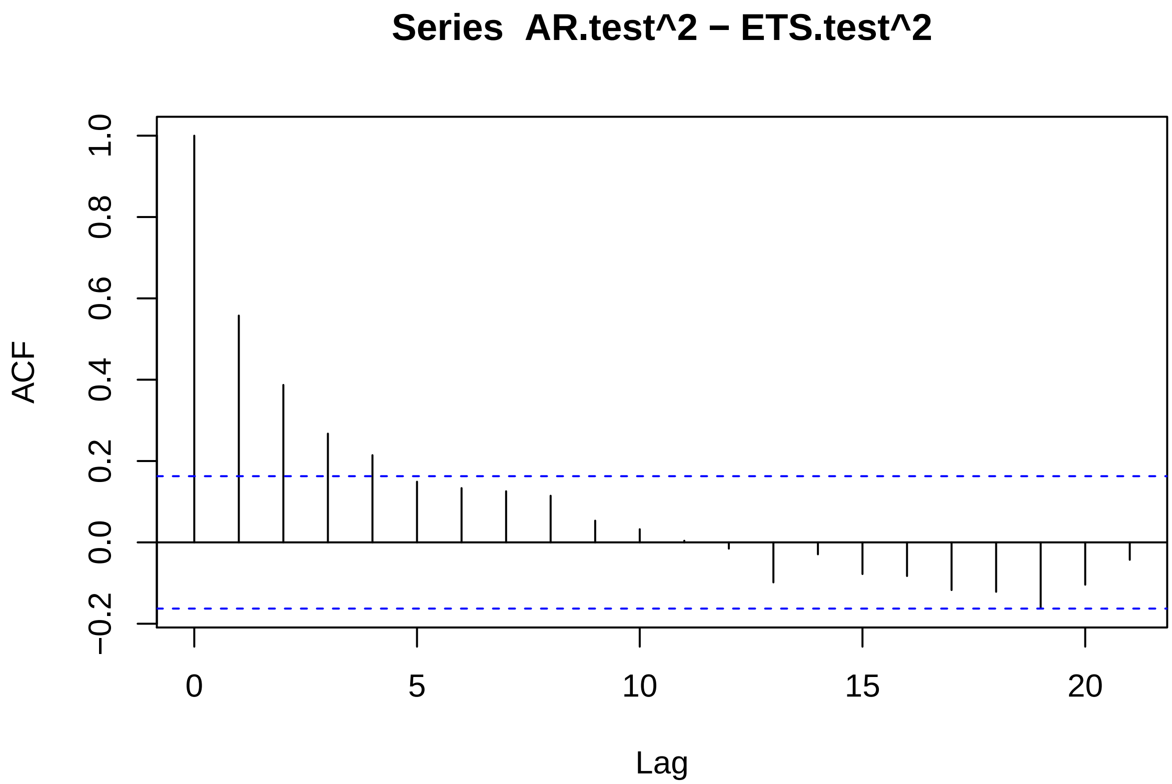
```
##
## data:  AR.testNNAR.test
## DM = -12.247, Forecast horizon = 1, Loss function power = 2, p-value <
## 2.2e-16
## alternative hypothesis: two.sided
```

From the plot of $AR.test^2 - NNAR.test^2$, we observe that the average of $AR.test^2 - NNAR.test^2$ tends to favor the ARIMA model that is mean squared errors of NNAR model are larger and therefore, pulling the values down. Additionally, from the ACF plot, we observe that $AR.test^2 - NNAR.test^2$ seems reasonably stationary after lag 3. The DM test gives us a very small p-value which tells us that there is a significant evidence against mean loss difference equal to zero. Hence, ARIMA model is performing better than NNAR mode in terms of 24 steps prediction.

```r
plot(AR.test^2 - ETS.test^2,type='l')
abline(h=mean(AR.test^2 - ETS.test^2),col=2,lwd=2)
```




```r
acf(AR.test^2 - ETS.test^2)
```



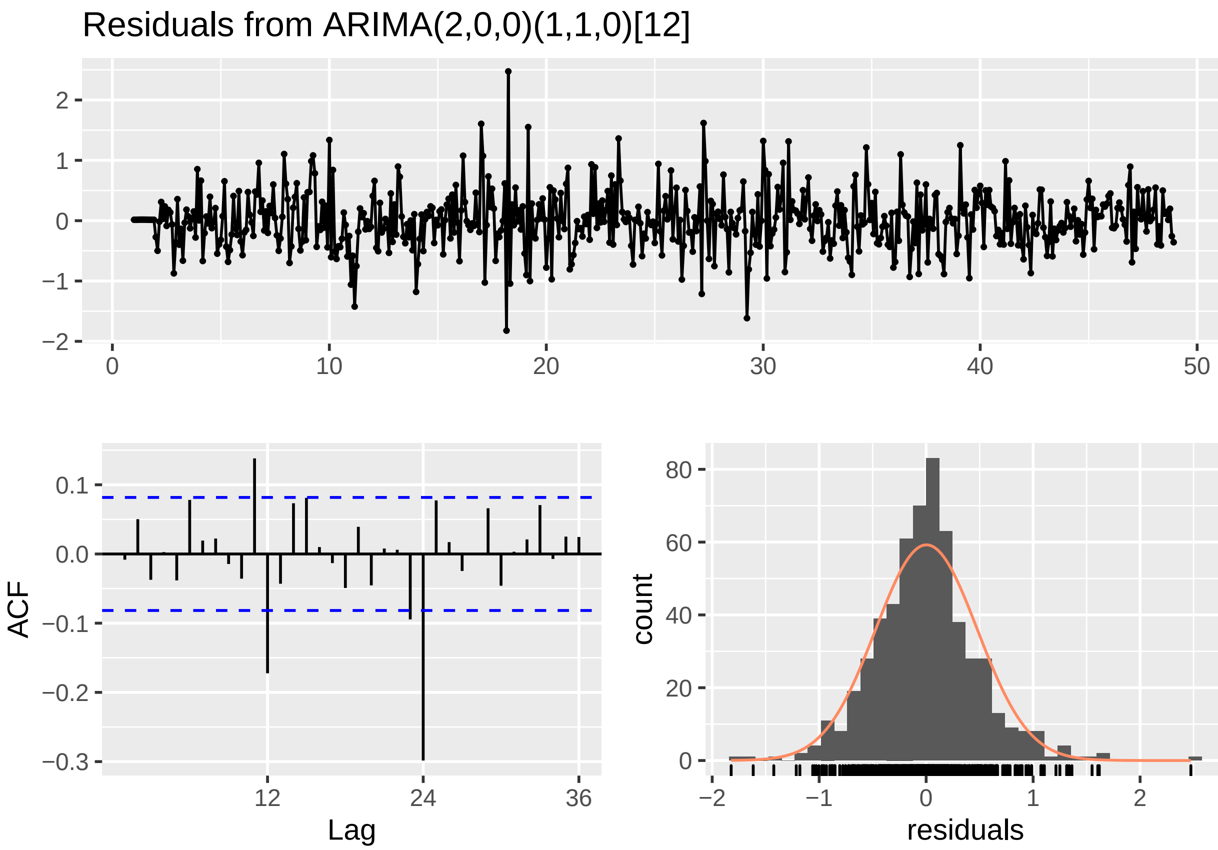

```
dm.test(AR.test, ETS.test)
```

```
##
##  Diebold-Mariano Test
##
## data:  AR.testETS.test
## DM = 3.0599, Forecast horizon = 1, Loss function power = 2, p-value =
## 0.002641
## alternative hypothesis: two.sided
```

From the plot of $AR.test^2 - ETS.test^2$, we observe that the average of $AR.test^2 - ETS.test^2$ tends to favor the ARIMA model that is mean squared errors of NNAR model are larger and therefore, pulling the values down. Additionally, from the ACF plot, we observe that $AR.test^2 - NNAR.test^2$ doesn't seem to be reasonably stationary. The DM test gives us a very small p-value which tells us that there is a significant evidence against mean loss difference equal to zero. Hence, ARIMA model is performing better than ETS model in terms of 24 steps prediction.

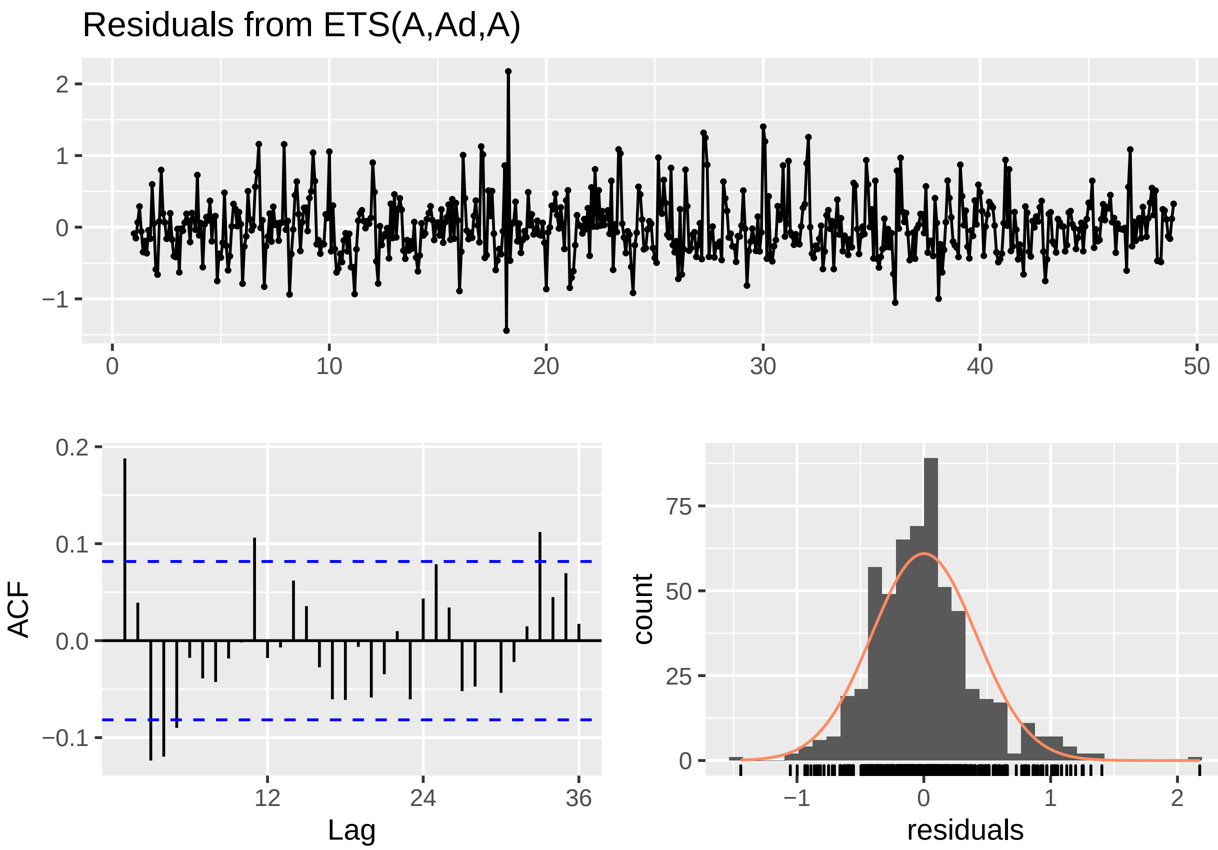
Since, NNAR model has the largest mean squared error, we rule it out. WE compare ETS and ARIMA model by running diagnostic tests for eg. we look at the stationary of their residuals and residuals ACF plot

```
par(mfrow=c(2,1))
# ARIMA model
AR.model <- auto.arima(hyd_ts)
ETS.model <- ets(hyd_ts)
# we get SARIMA(4,1,1,0,0,13) model after running auto.arima
par(mfrow=c(2,1))
checkresiduals(AR.model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,0)(1,1,0)[12]
## Q* = 107.98, df = 21, p-value = 1.089e-13
##
## Model df: 3.   Total lags used: 24
```

```
checkresiduals(ETS.model)
```

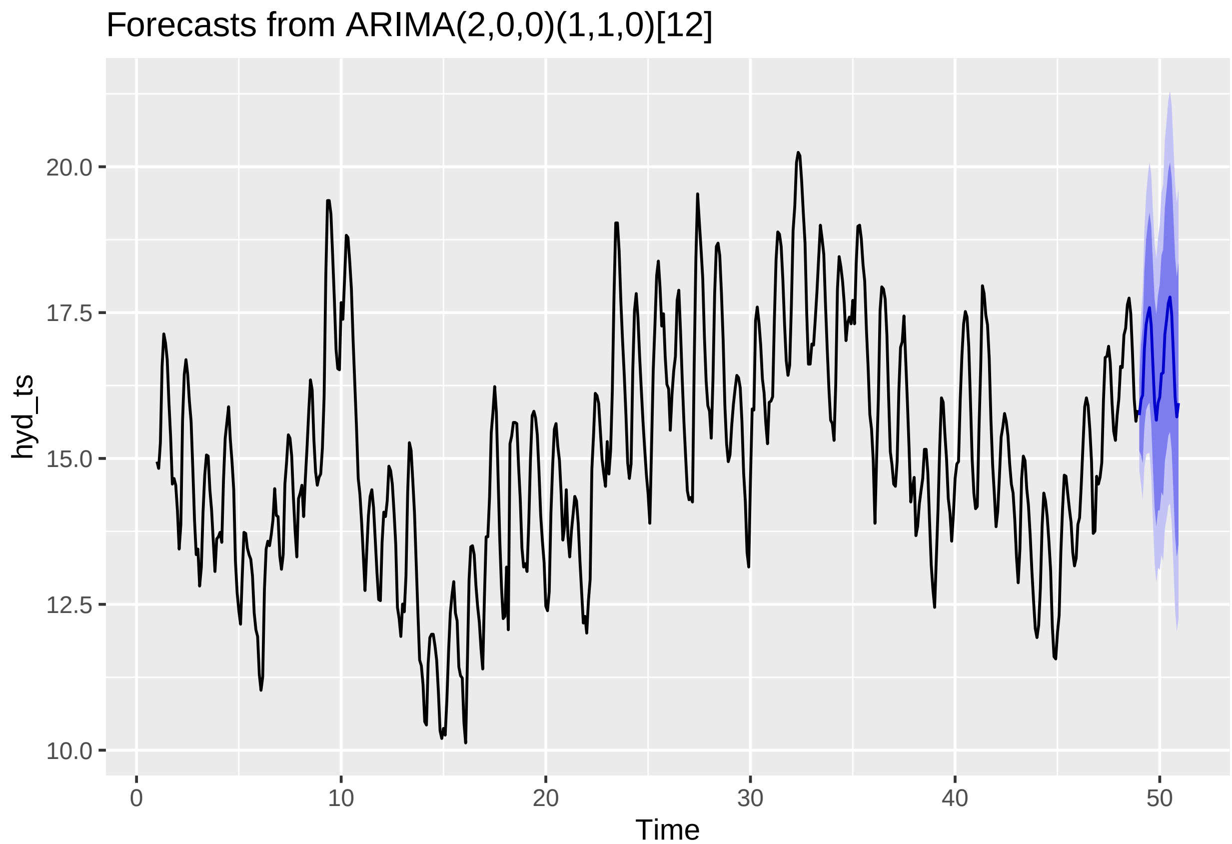



```
##
##  Ljung-Box test
##
## data:  Residuals from ETS(A,Ad,A)
## Q* = 66.546, df = 7, p-value = 7.349e-12
##
## Model df: 17.   Total lags used: 24
```

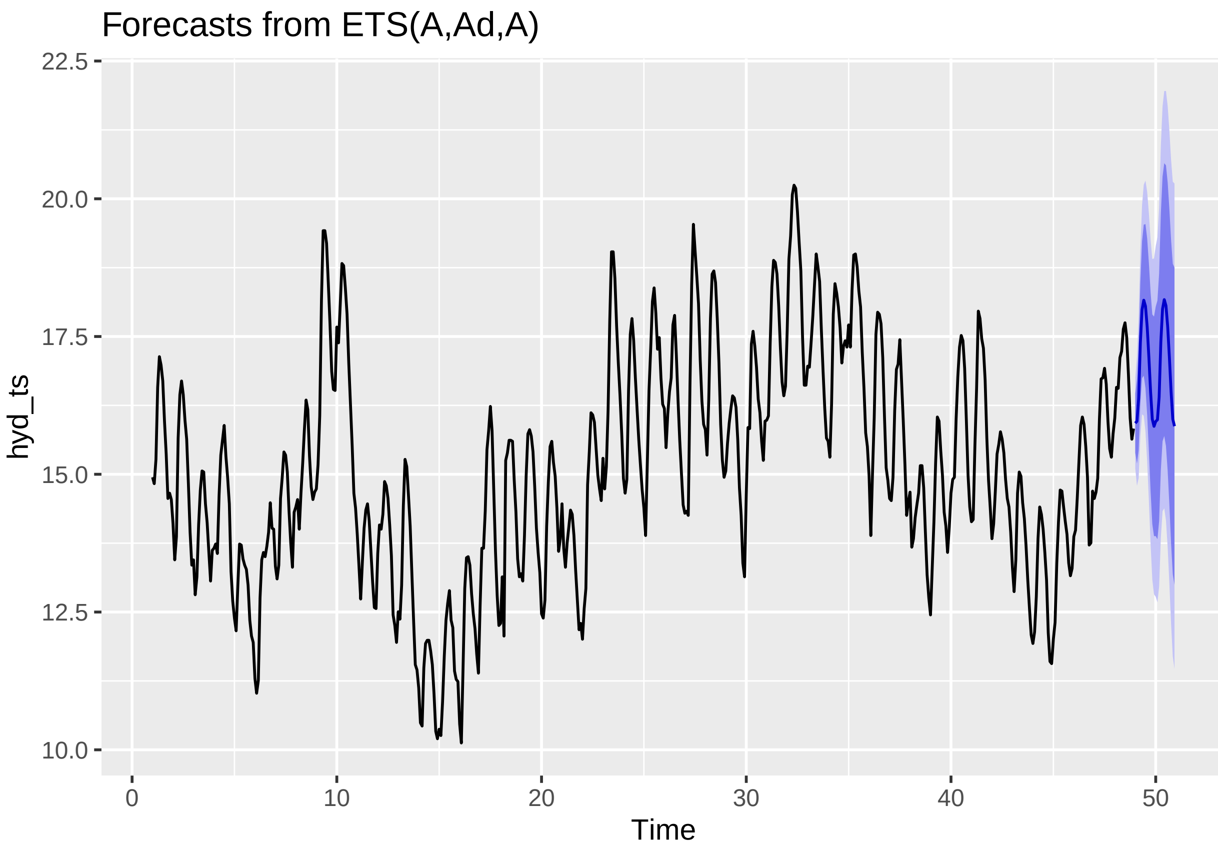
We observe from the ACF plot of residuals for SARIMA and ETS model that residuals of SARIMA model have smaller ACF values in comparison to residuals of ETS model. Additionally, the histogram plot of residuals looks more symmetric around 0 in SARIMA model than ETS model. Therefore, residual diagnostic test suggests that SARIMA model fits the time series data better

Now, we compare whether the forecasts and prediction intervals of ETS model or SARIMA model look more plausible.

```
par(mfrow=c(2,1))
AR.fc <- forecast(auto.arima(hyd_ts), 24)
ETS.fc <- forecast(ets(hyd_ts), 24)
autoplot(AR.fc)
```




```
autoplot(ETS.fc)
```



We observe from the forecast plots of both SARIMA and ETS model that SARIMA 24 step ahead forecasts fit the time series better than forecasts of ETS model. Additionally, in SARIMA model, the second-bounded interval(light purple), which is the 95% prediction interval seem to be capturing the seasonality in the data and they are wide enough to accommodate for the variability in the data, given the past observations. Whereas, the 95% prediction intervals of ETS model don't seem to be capturing the seasonality in the data and they are two wide to be considered plausible.

Hence, based on cross validation, residual diagnostics and forecast results, we conclude that SARIMA(4,1,1,0,0,13) fits the time series data the best

```
forecast1 <- AR.fc$mean
write(forecast1, file = paste("Scenario1_","Lamba","20751692",".txt", sep = ""), ncolumns = 1 )
```